IDENTIFICATION OF CLPS MEDIATED N-END RULE DEGRADATION IN *MYCOLICIBACTERIUM SMEGMATIS*

by

Christopher James Presloid

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Sciences

Fall 2023

IDENTIFICATION OF CLPS MEDIATED N-END RULE DEGRADATION

IN *MYCOLICIBACTERIUM SMEGMATIS*

by

Christopher James Presloid

Approved: _____

Velia M. Fowler, Ph.D.
Chair of the Department of Biological Sciences

Approved: _____

Deborah H. Norris
Interim Dean of the College of Arts and Sciences

Approved: _____

Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Karl R. Schmitz, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Ramona Neunuebel, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Melinda K. Duncan, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Brian Bahnson, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Carlton Cooper, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to start this document by expressing my appreciation to my advisor Dr. Karl Schmitz. When I first came to UD the lab was only just getting started and having the opportunity to help get the lab on its feet is something I will always cherish. Karl has been patient and supportive throughout my time working on this degree. He has helped spark an interest in proteins and infectious disease that I did not think I would ever have. I would not be where I am today without him.

I would also like to take this chance to thank all my committee members who guided me in my graduate training. Dr. Carlton Cooper has been incredibly supportive over the years and helped me get the opportunities to improve as a presenter and educator and kept me positive as much as possible during this process. Dr. Melinda Duncan has always been honest and an absolute wealth of knowledge during my training. She is tough and I thank her so much for being so. Dr. Bahnson has also been of great support offered unique perspectives to my committee. Finally, Dr. Neunuebel has been incredibly patient and helpful during my training especially these last few years. All of them have been incredible mentors and I am so very grateful for everything they have done.

My fellow lab members and past lab members also need to be thanked. Dr. Emmanuel Ogbonna and Dr. Jialiu Jiang, I do not think I can begin to thank you both enough for all the help and support you both have given over the years. Thank you, Patrick, for all the help you have offered and all the questions you have helped me with. Thank you, Monika for being such a positive and helpful presence in the lab.

Thank you also to Thomas Swayne who helped get this project started with me and was always such a wonderful person to talk with. Thank you also to Henry, Cady, Gaury, Jenn, and Pratista, you have all been so wonderful and I wish you all great success. You are all such fantastic scientists and colleagues I could not have asked for a better group of people to work with over these years.

A special thanks is also needed for Dr. Erica Selva and Betty Cowgill. Dr. Selva has made herself available time and again and has been such a supportive person. I can confidently say I would not have made it through this program without her support. Betty Cowgill is the other reason I've made it this far. She was always able to answer my questions, help me with paperwork, and anything else in between. Betty is an absolute treasure and deserves nothing but the best.

Lastly, I would like to thank my sister Alexis, and my friend Emily Zimmerman DVM. Alexis, thank you for keeping me grounded and helping keep me optimistic. As for Emily, you're the closest thing I have to a person who went through a similar field/ pursued a similar career path. You've been such a great friend and have made such a positive impact on me. I cannot begin to thank you both enough.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Tuberculosis (TB) has been and remains one of the most significant human infectious diseases, and currently causes more global death than any other single infectious agent aside from SARS-Cov-2. Advancements in public health have had a marked impact on prognosis and on the overall number of TB cases, however TB remains a persistent challenge in areas of the world lacking strong public health infrastructure or where public health programs have been disrupted by conflict. Adding to this burden, multidrug resistant strains (MDR-TB) of the causative bacterium, *Mycobacterium tuberculosis* (*Mtb*), have emerged, demanding the development of novel drugs and the exploration of novel drug targets.

One promising target is the strictly essential caseinolytic peptidase complex, ClpC1P1P2 protease, which uses chemical energy from ATP to degrade proteins in the mycobacterial cytosol. The ClpC1 unfoldase component of this protease has emerged as a particularly attractive molecular target, in part because no direct orthologs exists in humans. However, few ClpC1-specific substrates, proteolytic adaptors or proteolytic programs have been identified to date, hampering efforts to develop targeted screens for novel ClpC1 inhibitors. The discovery of additional ClpC1-specific degradation pathways would improve our ability to discover and characterize compounds that selectively inhibit ClpC1P1P2.

In the model bacterium *Escherichia coli* (*E. coli*), the closest homolog to ClpC1P1P2 is the protease ClpAP, which cooperates with a unique proteolytic adaptor: ClpS. *E. coli* ClpS facilitates a degradation pathway called the N-end rule in

which polypeptides that display the amino acids leucine, phenylalanine, tyrosine, or tryptophan at their N-terminus are recognized by ClpS and delivered to the unfoldase ClpA. ClpA then unfolds the polypeptide and passes it into the associated ClpP peptidase for destruction. *Mtb* and the surrogate model mycobacterium *Mycolicibacterium smegmatis (Msm)* both possess an ortholog of ClpS, but it has not yet been reported whether an analogous N-end rule pathways exists in these organisms.

In Chapter 2, we tested the hypothesis that ClpS cooperates with ClpC1P1P2 to carry out N-end rule proteolysis in mycobacteria. To address this, we looked at three features of the Clp protease system; accessing the conservation of mycobacterial ClpS to the well-studied *E. coli* ClpS, the interaction between mycobacterial ClpS and the unfoldase ClpC1, and finally the profile of N-end amino acids prioritized for degradation by ClpC1P1P2. Sequence and structural analysis demonstrate that $_{Eco}$ClpS and $_{Msm}$ClpS share pronounced similarities, including key features associated with N-end rule substrate binding. Biophysical studies with purified $_{Msm}$ClpS demonstrate its ability to bind model peptides bearing canonical N-end rule amino acids, and these interactions are confirmed by co-crystal structures with model N-end rule peptides. Finally, a cell based proteolytic reporter assay implemented in *Msm* establishes that model substrates displaying the four canonical N-end rule amino acids are expressed at lower steady state levels than substrates displaying any of the remaining 16 amino acids. Together, the data indicates that ClpS binds substrates presenting canonical N-end rule residues in *Msm*, and mediates an N-end rule degradation pathway in a similar pattern to *E. coli*. By identifying a novel substrate degradation pathway, we open a new avenue for drug screening that targets against

this essential protease. Furthermore, the methodology and techniques used in this project can serve as a baseline for future substrate screens of the mycobacterial Clp protease. Additional work will be required to understand the interaction of ClpS to ClpC1 as well as the role this pathway plays in mycobacterial protein homeostasis.

In Chapter 3, we report proof-of-principle studies that demonstrate the feasibility of using APEX2-facilitated proximity labeling to identify novel substrates and interaction partners of Clp proteases within mycobacteria. Mycobacterial Clp proteases, including ClpC1P1P2, are known to be essential for cell viability, but the specific protease function responsible are not well understood. Identifying novel substrates, adaptors, and regulators would expand our understanding of protease function and regulation in these cells. APEX2 proximity labeling is a method of covalently tagging nearby interaction partners of a protein of interest with biotin or other chemical handles. Here, we engineer fusion constructs of the APEX2 peroxidase with several proteins of interest. Furthermore, we establish that APEX2-mediated labeling can be carried out in mycobacterial cytosol and intracellular transport is not fundamentally impeded by the cell wall. Finding novel interaction partners of Clp protease will ultimately help determine how this essential protease is regulated and help to inform drug discovery through exploitation of these interaction surfaces.

Chapter 1

INTRODUCTION

1.1 Pathogenesis of *Mycobacterium tuberculosis*

Even as the world recovers from the COVID-19 pandemic, tuberculosis continues to be one of the leading causes of mortality worldwide. Based on the World Health Organization (WHO) reports, the COVID-19 pandemic caused substantial disruption to tuberculosis identification and treatment efforts. Tuberculosis is estimated to have caused 1.6 million deaths in 2021, a notable increase after years of decline [1]. As tuberculosis is also highly communicable, 10.6 million people are estimated to have been infected in 2021, an increase of 4.5% from 2020 [1,2]. This increase breaks another trend in reduced rates from 2005-2019 where ~2% reductions in infections were seen year over year [1]. These disruptions coupled with the existing estimates that up to one quarter of the world's population is infected with *Mycobacterium tuberculosis* (*Mtb*) warrants the allocation of significant resources to reverse these disruptions, and a continued investment in new treatment methods [1,3,4].

Infection by *Mtb* leads to the disease tuberculosis, which is characterized by infection of lung macrophages leading to the development of granulomas. These structures damage the lungs causing coughing, bloody phlegm, fatigue, and in advanced cases wasting, anemia, and death. The colonization of macrophages by *Mtb* is a process the bacterium uses to circumvent the body's natural immune response. Typically, macrophages are responsible for engulfing and digesting foreign bacteria,

but *Mtb* has several genes that allow it to prevent the acidification of the phagosomes. This pathogen-modified phagosome then serves as a protected pocket for the bacterium to replicate as phagosome maturation is further disrupted by the presence of *Mtb* disrupting the PI3K signaling cascade [5,6].

As the body continues to respond to the infection infected macrophages are sequestered and isolated by the recruitment of other immune cells including lymphocytes, monocytes, and neutrophils. These cells work to prevent the spread of further *Mtb* bacterium as they try to escape from the macrophage. The formation of the granuloma involves these immune cells coupled with epithelial cells and fibroblasts [5–7]. The formation of the granuloma is meant to isolate the infection site and starve the isolated cells of nutrients. However, studies have demonstrated that the granuloma microenvironment activates the expression of several genes to avoid lysosomal degradation, as well as oxidative stress[8]. These genes enable the pathogen to survive/ persist in a more dormant or latent infection[8,9].

Given time, the dormant *Mtb* bacteria can reactivate, especially in response to disruption of the granuloma or should the immune system become compromised. One example of a protein that can reactivate infection are lytic transglycosylases also called resuscitation promoting factors (RPFs)[10]. As mentioned, disruption of the immune system can reactivate a latent infection, and as such HIV/AIDS or other immunocompromising conditions can contribute to reactivation [11].

Active infections of *Mtb* are primarily characterized by the replication of the bacterium within the alveoli of the lung, but infection is not limited to these sites. The infection of the lung exclusively by *Mtb* makes up most cases and is referred to as pulmonary tuberculosis. As infection progresses or becomes more severe

extrapulmonary tuberculosis can develop and is characterized by the infection of other organs including kidneys, larynx, bones, and brain. A final form of infection by *Mtb* is called “miliary” tuberculosis and involves the bacterium entering the bloodstream and infecting multiple sites or organs in the body. A feature of these more advanced forms of infection (extrapulmonary and miliary) is the development of tuberculosis meningitis, or an infection of the central nervous system, and can be fatal [1,5,6].

Figure 1.1) *Mycobacteria tuberculosis* is a human pathogen

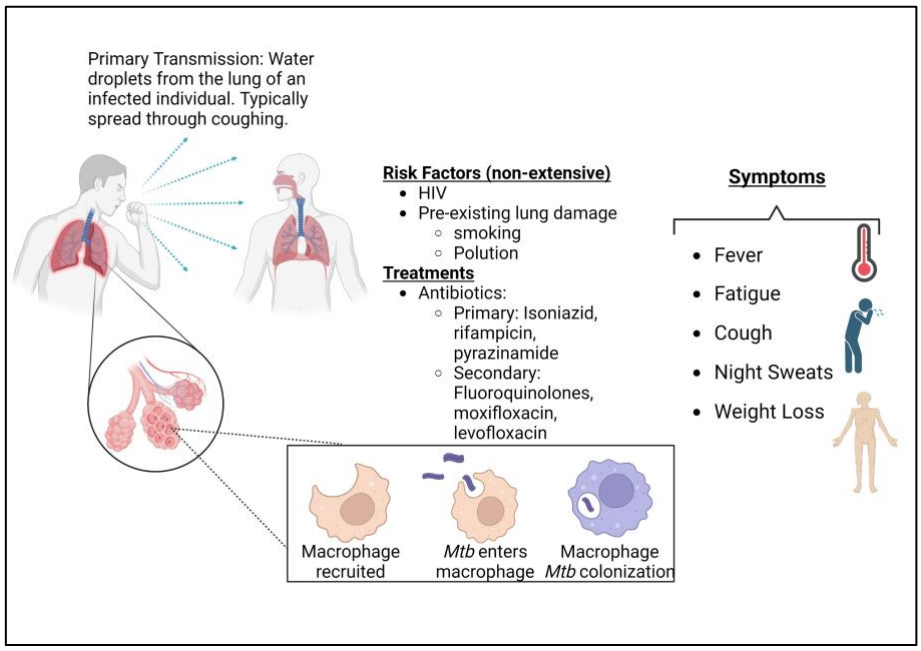


Fig 1.1: *Mycobacterium tuberculosis* is pathogenic bacteria that primarily affects the lungs and is spread through water droplets from the lungs of a person with an active infection. Those with compromised immune systems, and pre-existing lung damage, as well as the very old and young are at the highest risk of severe infection with 90% of total mycobacterial infections occurring in adults[1]. Antibiotic treatment is typically administered over several weeks and is divided based on severity of infection and potential resistances (primary for less severe and escalating to secondary for more severe. Symptoms of TB include but are not limited to coughing, fatigue/ weight loss, and fever. This figure was created with BioRender software.

1.2 TB drugs and drug resistance

Treatment of *Mtb* infection is typically carried out over several weeks using a limited number of antibiotics (Fig1.1)[12]. This is primarily because the cell wall architecture of *Mtb* is unique and restricts the movement of large compounds including antibiotics into the cell[13,14] The reduction of overall metabolic activity during latent infection also restricts the movement of compounds into the cell. The majority of currently available treatments target processes associated with cell wall synthesis and protein synthesis[15–17]. This overlap in targets means that as resistance develops for one drug, another may also lose effectiveness as they disrupt the same metabolic pathway. The ability of *Mtb* to persist within lung tissue, and the difficulty associated with clearing an infection entirely, further complicate treatment as incomplete treatment can lead to a more resistant secondary infection[12,18].

Following the disruptions caused by Covid-19, it appears that the overall number of cases has increased to 10.6 million from 10.1 the previous year while the rate of new resistant cases rose to 4.2% from a rate of 3.4% in 2020 based on 2022 statistics from the WHO[1]. These increases mark a reversal of trends over a decade in the making that saw infections, both conventional and resistant, falling or static. The ability of those dealing with rifampin-resistant (RR-TB) TB as well as multi-drug resistant (MDR-TB) was also disrupted increasing the number of people in need of escalation in treatment to about 1 in 3 based on reported data[1].

This increase in new resistant infections as well as disruption to the treatment of existing resistant infections means the need for new therapeutics and vaccines is even more urgent. In the past few years new targets for drug discovery have emerged that fall outside of existing treatment targets and may offer avenues of treatment for resistant cases of TB[19]. One of these new targets includes the focus of this project, the

ClpC1P1P2 Clp protease, which is an essential protein complex involved in protein degradation. ATP dependent proteases, including the Clp protease, are responsible for the removal of damaged or unneeded proteins[20,21]. They are also involved in the regulation of multiple cellular pathways[22,23]. Ecumicin is one example that targets this protease by increasing the ATP usage of ClpC1 but preventing it from actually proteolyzing substrates with ClpP1 and ClpP2[24,25] ADEPs are another class of cyclic compounds known to interfere with the activity of the Clp protease both in *Mtb* as well as other bacterial species[26–29].

1.3 The Basics of ATP-dependent Clp proteases

Protein degradation is an essential function in all cells. Within most mycobacteria there are five ATP-dependent proteases that work to maintain protein homeostasis. These are ClpC1P1P2, ClpXP1P2, FtsH, Lon, and the Mpa-20S proteasome [30]. While all carry out degradation of folded protein substrates, their overall makeup varies. ClpC1P1P2, ClpXP1P2, and the Mpa-20S proteases are composed of two separate oligomeric components: an unfoldase that recognizes and unfolds protein substrates, and a separate peptidase responsible for polypeptide degradation. By contrast, FtsH and Lon contain unfoldase and peptidase components within a single polypeptide chain [31–34]. Notably, all of these proteases are freely soluble in the cytosol except for FtsH, which contains a transmembrane region that anchors it to the plasma membrane[34,35].

Figure 1.2) Proteases of the mycobacterial cell.

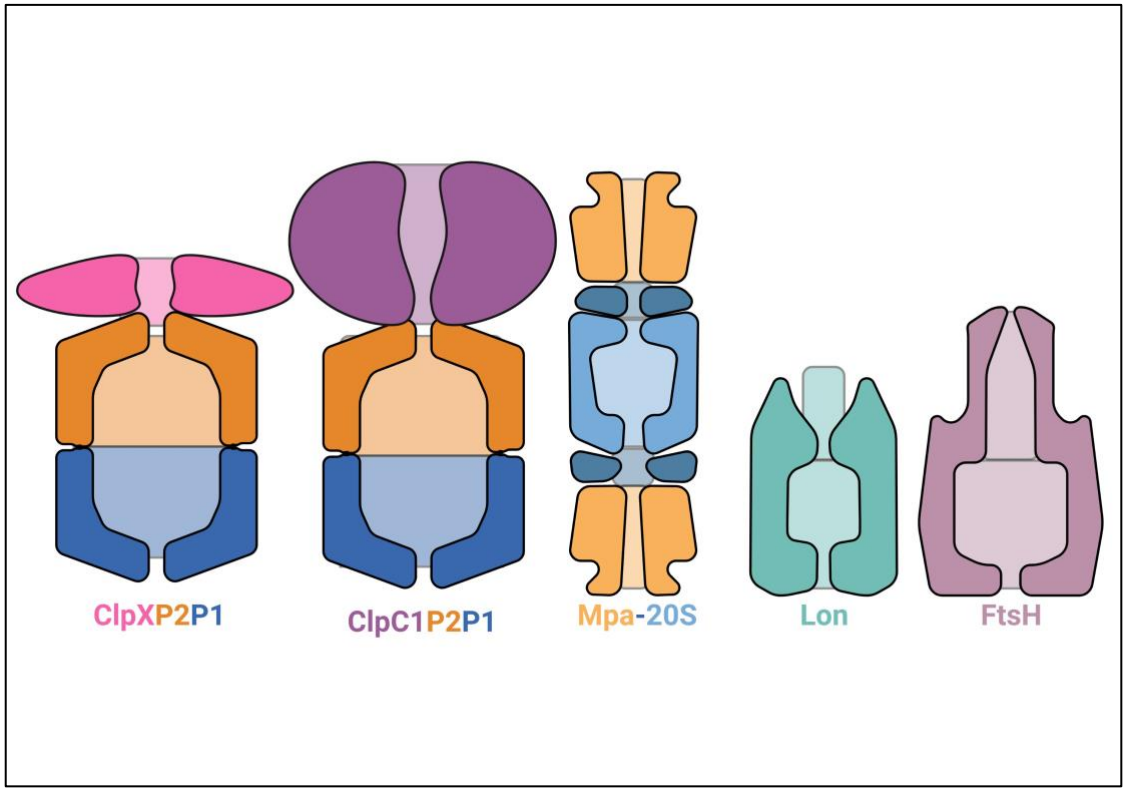


Fig 1.2: *Mtb* contains five protease complexes that help to maintain protein homeostasis by degrading proteins that are no longer needed or damaged. Of these two are essential: ClpXP2P1 and ClpC1P2P1. The Mpa-20S protease conditionally essential and controls the virulence of the bacteria. The remaining Lon and FtsH proteases are non-essential. This figure was created using BioRender software.

ClpXP1P2 and ClpC1P1P2 incorporate separate unfoldases (ClpX or ClpC1) that assemble atop a common peptidase, ClpP1P2. This barrel-shaped peptidase forms from separate heptameric rings of ClpP1 and ClpP2 that interact face-to-face. The interior of the peptidase barrel is a solvent-filled degradation chamber lined with 14 peptidase active sites. (Mycobacterial ClpP1P2 is more complex than the equivalent peptidase in many other bacteria, including *E. coli*, where a homo-tetradecameric ClpP barrel forms from a single ClpP isoform.) Both ClpX and ClpC1 form homomeric ring-shaped hexamers with a central axial channel, which dock co-axially atop the ClpP2 face of the peptidase[28,36]. To date, no interaction partners are yet known that bind to the surface of ClpP1.

Clp unfoldases are responsible for recognizing protein substrates, either through direct interactions, post-translational modifications, or via adaptor proteins [36,37]. Axial loops allow unfoldases to grip substrate polypeptides. ATP hydrolysis events within unfoldase subunits are coupled to loop motions that apply a mechanical unfolding force to protein substrates. Denatured proteins are translocated through the unfoldase pore and into the interior chamber of ClpP1P2 for hydrolysis. The interaction between the unfoldase and peptidase barrel is facilitated by LGF-loops that descend from the main structure of the unfoldase ring and dock within grooves between subunits of the ClpP2 ring [38]. Similar ClpXP proteases are found in virtually all bacteria and in eukaryotic mitochondria. Mycobacterial ClpX has 62% sequence identity to *E. coli* ClpX and 50% sequence identity to human mitochondrial CLPX[39–41]. *Mtb* ClpP2 has 44% (45% for ClpP1) sequence identity to human and 49% (56% for ClpP1) to *E. coli* ClpP[42–44]. By contrast, ClpC1 orthologs are absent in many

clades. Mycobacterial ClpC1 has distant homology to the *E. coli* ClpA unfoldase (38% sequence identity), but has no direct homolog in humans[40,45].

Once an unfolded polypeptide substrate has entered the interior of the ClpP1P2 barrel, the fourteen serine peptidase active sites work to break it down[46,47]. Hydrophobic grooves near the active sites, along with a hydrophobic S1 pocket, position the scissile peptide bond for nucleophilic attack by a trypsin-like Ser-His-Asp catalytic triad[28,48,49]. The resulting fragments are typically between seven and eight residues in length. These short fragments are hypothesized to exit the ClpP1P2 barrel through the axial pore of the barrel.

The interaction between the unfoldase ring and the ClpP2 ring stabilizes the overall protease complex, but also allosterically stabilizes assembly between ClpP1 and ClpP2 rings[27,36,46]. In the absence of proteolytic substrates, mycobacterial Clp proteases disassemble into constituent rings, and the ClpP1 and ClpP2 rings adopt an inactive resting state[36,49,50]. As a result, ClpXP1P2 and ClpC1P1P2 appear to assemble and disassemble dynamically, according to the availability of proteolytic substrates[36,51]. The overall stability of mycobacterial Clp proteases is much lower than those of *E. coli*, which remain stably associated even when substrates are absent[36,52,53]. This instability is surprising given the essentiality of both ClpXP1P2 and ClpC1P1P2 in mycobacteria, whereas the homologous proteases (ClpXP and ClpAP) in *E. coli* are non-essential[52,54–57].

The peptidase active sites in the ClpP1P2 barrel can cleave most polypeptide sequences that enter the interior, and full Clp proteases have a generic ability to unfold and destroy diverse protein substrates[49]. Consequently, bacteria rely on several regulatory paradigms to protect cells from unintentional destruction of cellular

proteins. The barrel-shaped architecture of ClpP1P2 prevents folded proteins from diffusing into the degradation chamber[48,49,58]. Proteins must first be unfolded by ClpX or ClpC1 and translocated through the ClpP2 axial pore. Additionally, ClpX and ClpC1 can stringently recognize valid substrates and ignore non-substrate proteins in the cytosol[36]. Unfoldases thus have a critical "gatekeeping" role. However, only a handful of mycobacterial Clp protease substrates have been identified to date, and the full repertoire of physiological substrates remains unknown.

1.4 N-end rule proteolytic pathways

The N-end rule pathway is a protein degradation pathway, first identified by Varshavsky *et al* in 1991 in *E. coli*[59]. This adaptor-mediated pathway involves the targeted binding and degradation of cytosolic proteins based on the identity of their N-terminal amino acid. N-end rule pathways have been identified in all domains of life, although the adaptors, proteases, and rules of amino acid selectivity differ. In all cases, the N-end rule links protein half-life to the identity of the (exposed) N-terminal residue. N-end rule pathways are tied to a broad range of cellular functions across organisms, including oxygen and nitric oxide sensing, bacterial virulence, cellular signaling, and in eukaryotes neurogenesis, cardiovascular development, and pancreas and brain function[21].

Figure 1.3) Basics of *E. coli* N-end rule degradation

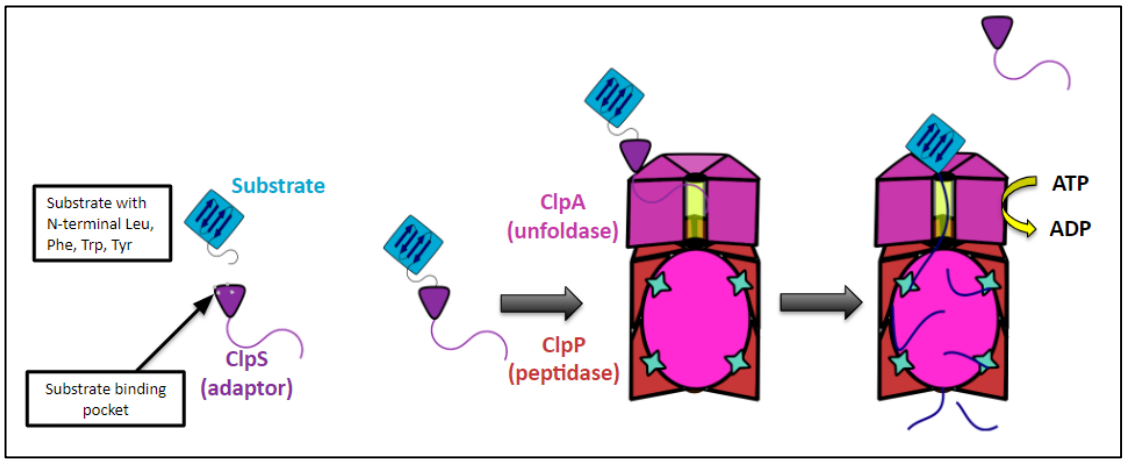


Fig 1.3: The *E. coli* N-end rule pathway is carried out with the ClpAP protease using a ClpS adaptor protein. After being either damaged or modified through aminotransferases, substrates have a revealed N-end rule degron. This single amino acid (Tyr, Trp, Leu, or Phe) is sufficient to reduce the half-life of this protein *in vivo*. ClpS first interacts with substrates by binding the N-terminus in a hydrophobic binding pocket located at the" top" of this heart shaped adaptor. After binding substrate, ClpS first interacts with ClpA with its unstructured tail interacting with the central pore of this unfoldase. Substrate is then exchanged with this unstructured tail, allowing ClpS to escape the ClpAP protease, while the substrate is first unfolded and then cleaved by ClpA and ClpP respectively.

The adaptor ClpS is responsible for N-end rule degradation in *E. coli*, through its ability to directly bind N-terminal tyrosine, tryptophan, phenylalanine, and leucine amino acids on substrates and deliver those proteins to the ClpAP protease[59–62]. As originally reported by Varshavsky and colleagues in *E. coli*, exposure of these four amino acids on a model protein is associated with a dramatic reduction in proteolytic half-life[59]. These amino acids, which are directly recognized by ClpS, are termed primary N-end rule degrons. Yet, nascent polypeptides do not start with these residues, so how do primary N-end rule degrons appear in *E. coli*?

Nascent polypeptide transcripts in bacteria begin with formyl methionine, which is subsequently converted to methionine (a non N-end rule amino acid) by the enzyme methionine deformylase[60,63,64]. In *E. coli*, the N-terminal methionine may be removed by methionine-aminopeptidase (MAP). Interestingly, methionine removal by MAP is dependent on the identity of the residue in the second position: Met removal is disfavored if the second amino acid is bulky and hydrophobic, thus preventing the constitutive exposure of a primary N-end rule degron.

How then are valid N-end rule substrates created in the cell? One way is by cleavage of mature proteins by endopeptidases to reveal new N-end rule degrons. Another path is through the activity of aminotransferases that add a primary N-end rule degron in front of an existing amino acid, termed a secondary N-end rule degron. Whereas primary degrons reduce protein half-life by direct ClpS recognition, secondary degrons lead to proteolysis through the constitutive addition of a primary degron. For example, the *E. coli* L/F^{K,R}-transferase adds a Leu or Phe residue in front of proteins displaying an N-terminal Lys or Arg[65]. Lys and Arg are thus secondary N-

end rule degrons in *E. coli*. [60] The existence of secondary degrons is not uniform across bacteria. Using *E. coli* and *V. vulnificus* as two examples, *E. coli* has two secondary degrons (Arg/Lys) while *V. vulnificus* has four secondary degrons (Arg/Lys/Asp/Glu)[59,60,62,66]. This unique profile of secondary degrons means the degradation profile of different organisms (both eukaryotic and prokaryotic) can vary across species. This is a potential boon for drug discovery as the unique profile of N-end degrons in bacteria could be targeted in a manner that do not interfere with native N-end rule degradation.

1.5 Mycobacterial Clp proteases as drug targets

Protein degradation is a critical process in all bacteria. However, as most bacteria process multiple ATP-dependent proteases, inhibition of one protease is often overcome or compensated for by other proteases. In the case of mycobacteria however, the Clp protease ClpC1P1P2, is strictly essential for the cell's viability: inhibition or deletion of any component results in cell death in culture and in models of infection. Why this protease is essential has not been determined, but this phenomenon can be exploited for drug development. Indeed, multiple classes of compounds have been shown to kill *Mtb* cell by targeting these enzymes.

Acyldepsipeptides (ADEPs) are one class of antibiotics that target the Clp protease that specifically bind to ClpP2 [26]. Binding of ADEPs to ClpP2 competitively hinders its interaction with Clp unfoldases, and also widens the axial pore of the ClpP1P2 peptidase barrel structure. The disruption of unfoldase binding removes the regulatory mechanism for the Clp protease and inhibits targeted degradation of substrates. Widening of the axial pore allows unregulated degradation of peptides, unstructured proteins, and nascent polypeptides that diffuse into the peptidase barrel.

While ADEPs can kill *Mtb,* they have relatively poor potency and are susceptible to active export by mycobacterial efflux pumps, thus further development of these compounds are needed[67]. They are also known to cross-target mitochondrial ClpP, disrupting mitochondrial protein homeostasis, which may limit their clinical utility[41,68].

Disruption of Clp protease function can also take place at the unfoldase. The non-ribosomally encoded cyclic peptide cyclomarin A binds to the N-terminal domain of ClpC1[69]. This interaction causes unregulated and indiscriminate degradation of proteins. A structurally related peptide antibiotic, rufomycin, binds in the same location but instead inhibits proteolysis and destabilizes ClpC1 hexamer assembly[70]. Both peptides kill *Mtb*, but it remains uncertain how these structurally related compounds exert different mechanistic effects on the unfoldase through the same binding site[15,69–71]. Ecumicin is structurally a different non-ribosomally produced cyclic peptide antibiotic that also binds to ClpC1NTD, but through a distinct binding site [24,25]. Ecumicin stimulates ATPase activity and leads to disruption of normal ClpC1-dependent proteolysis. The existence of multiple naturally occurring antibiotics that target ClpC1 highlights its potential utility as a therapeutic target. Unlike ClpP1P2, ClpC1 lacks a direct ortholog in humans, reducing the concern of off target interactions.

Table 1.1) Clp protease targeting antibiotic compounds.

Compound Name	**Effect on Clp Protease**
ADEPs[26,68,72]	1. Stimulation of ClpP1P2 for unregulated degradation. 2. Inhibition of unfoldase/peptidase binding interface.
Ecumicin[24,25,73]	Increase ClpC1 ATPase activity/ uncouple ATPase of ClpC1 from proteolysis by P1/P2
Cyclomarin[69,71,74]	Binds N-terminus of ClpC1 (domain responsible for substrate recognition) opening the unfoldase channel and allowing unregulated proteolysis
Lassomycin[75]	Increase ClpC1 ATPase activity/ uncouple ATPase of ClpC1 from proteolysis by P1/P2

1.6 Identification of Degradation Pathways to Improve Drug Screening

When looking at the development of new drugs targeting bacterial proteins, it is necessary to understand known functions of the targets. In the case of Clp proteases the mechanochemical process of substrate degradation is well understood, but the specific substrates or pathways targeted by these enzymes are not well characterized. The essentiality of mycobacterial Clp proteases makes them appealing drug targets, but also complicates efforts to identify associated substrates and pathways.

Studies in other bacterial species indicate that Clp proteases contribute to general protein homeostasis, as well as some stress responses, virulence regulation , and cell division [76,77]. However, only a handful of specific substrates and roles for ClpC1P1P2 have yet been identified in mycobacteria [36,48,54]. Known substrates include the protein RseA (an anti-sigma factor), heat shock protein Hsp20, protein fragments that have been labeled with an ssrA degradation sequence appended during the process of ribosomal rescue, and aspartate decarboxylases[28,56,78,79]. Recently, Dr. Emmanuel Ogbonna in the Schmitz lab identified two additional substrates: proteins

carrying phosphoarginine posttranslational modifications, and a *Msm* pseudogene annotated as 5'-oxoprolinase[80,81].

In many organisms, Clp protease are known to carry out N-end rule degradation. N-end rule pathways have been identified in several different bacterial and eukaryotic organisms but has not yet been definitively characterized in mycobacteria including *Mtb* or *Msm*[21,59,62,82]. What work that has been done shows a model GFP substrate bearing a "canonical" N-end rule amino acid, Phe, is degraded by ClpC1P1P2 in the presence of mycobacterial ClpS[54]. However, the breadth of N-end rule degrons, and the existence of this proteolytic pathway in mycobacteria, has not been demonstrated. Interestingly, ClpS also appears to be capable of facilitating non-N-end rule degradation through its targeting of Mrf (Mpy recruitment factor) in the presence of zinc[83].

The interaction between substrates and protease is important to understand, especially when it comes to drug screening. By finding which pathways the essential *Mtb*Clp protease interacts with, and what substrates are targeted, we can record how new drugs effect these pathways. Drug development can often look to existing substrates or interactions of a target protein for new options at disruption.

1.7 Project Aims and Dissertation Work

Clp proteases are essential for viability in mycobacteria, including the human pathogen *Mycobacterium tuberculosis* (*Mtb*) and the model mycobacterium *Mycolicibacterium smegmatis* (*Msm*)[54,84–86]. These ATP-dependent proteases have been the focus of multiple efforts to develop novel drugs for the treatment of drug resistant *Mtb* infections. The ClpC1P1P2 protease is a particularly attractive drug target because the ClpC1 unfoldase component lacks a direct ortholog in humans, thus

reducing the risk of off target interactions within the host. However, one of the major limiting factors in developing ClpC1 targeting therapies has been the dearth of ClpC1P1P2 specific substrates [54]. Our overarching objective is to identify novel ClpC1 specific substrates and proteolytic pathways, to support drug discovery efforts. Indeed, Chapters 2 and 3 address the global hypothesis that ClpC1P1P2 has multiple unidentified substrates, adaptors, and interaction partners that modulate its' proteolytic activity.

In Chapter 2, we test the hypothesis that mycobacterial ClpC1P1P2 works together with the adaptor ClpS to carry out an N-end rule degradation pathway. In *E. coli,* ClpS makes critical interactions with both ClpA and with N-end rule degrons that facilitate N-end rule degradation. Sequence analysis of mycobacterial ClpS identified conservation of residues associated with the substrate binding pocket and the unfoldase interaction region. This analysis showed robust similarity in the binding pocket of $_{Msm}$ClpS and $_{Eco}$ClpS, but more modest similarity in the unfoldase interaction region. The structure of $_{Msm}$ClpS was solved by X-ray crystallography, confirming strong structural conservation with $_{Eco}$ClpS.

Biochemical assays using the purified $_{Msm}$ClpS confirm that the protein binds canonical N-end rule degrons *in vitro*. Additionally, co-crystal structures of $_{Msm}$ClpS bound to short model N-end peptides show that all four canonical primary N-end rule amino acids bind to ClpS. Finally, a cell-based proteolytic reporter assay in *Msm* confirms that model substrates bearing N-end rule degrons are expressed at lower levels that substrates bearing any of the other 16 amino acids. Interestingly, our results point to the absence of secondary N-end rule degrons in *Msm*, at least under the conditions tested.

Chapter 3 addresses the question of substrates and interaction partners more broadly, by implementing an APEX2 based proximity biotinylation approach to identify novel interaction partners of mycobacterial Clp proteases in *Msm*.

Chapter 2

IDENTIFICATION AND CHARACTERIZATION OF CLP PROTEASE MEDIATED N-END RULE DEGRADATION IN *MYCOLICIBACTERIUM SMEGMATIS*.

2.1 Introduction

Despite modern medical advances, tuberculosis continues to pose a major challenge to global public health. It is estimated that a quarter of the global population is infected with tuberculosis, and that tuberculosis causes more deaths worldwide than any other bacterial disease. Recent data suggests that significant setbacks to the management of TB have occurred over the course of the COVID-19 pandemic[1]. Even as we continue to grapple with the aftermath of the pandemic some trends have remained stubbornly consistent. The percentage of antibiotic resistant cases of tuberculosis has remained at 3.5-4% (with a marked increase post Covid-19) since 2015, and the association between lower GDP per capital and undernourishment to TB infection cannot be ignored [1,2]. These trends highlight the necessity of new therapeutics that can address antibiotic resistance while also being stable, and low cost to be as universally available as possible.

One prospective group of novel therapeutic targets in the causative agent, *Mycobacterium tuberculosis,* are the Clp proteases, and particularly ClpC1P1P2. Clp proteases degrade folded protein substrates in an ATP dependent manner[20,36,37,46,56]. These complexes consist of a ATP-dependent unfoldase (ClpX or ClpC1 in *Mtb*) that caps an oligomeric peptidase barrel (ClpP1P2 in *Mtb* – formed from separate

heptameric rings of ClpP1 and ClpP2 peptidase isoforms)[27,56]. The Clp unfoldase selectively recognizes protein substrates, unfolds them using energy from ATP hydrolysis, and passes the resulting denatured polypeptides into the degradation chamber of ClpP1P2 or destruction[22,36,50,87]. All components of these enzymes are strictly essential for viability in mycobacteria, and compounds that inhibit Clp protease activity are known to kill *Mtb* in culture and in models of infection. However, the physiological reasons for the essentiality remain unknown, and few physiological substrates have been identified to date. These gaps in our knowledge constrain efforts to identify and develop prospective therapeutics that target these enzymes.

The mycobacterial ClpC1P1P2 protease bears strong structural and sequence similarity to the well-studied ClpAP protease from the model bacterium *Escherichia coli*[36,88,89]. *E. coli* ClpAP is known to interact with the small proteolytic adaptor ClpS to carry out a substrate degradation termed the N-end rule[21,59]. ClpS interacts with the N-terminal domains of ClpA, and with substrates bearing a leading N-terminal Phe, Trp, Tyr, or Leu, thereby delivering these substrates to ClpAP for destruction. Within mycobacteria there also exists a ClpS protein. Prior bacterial two-hybrid studies have shown that mycobacterial ClpS interacts with ClpC1, and that ClpS mediate proteolysis by ClpC1P1P2 of model substates bearing an N-terminal Phe[54]. However, to date the existence of a functional N-end rule pathway in mycobacterial has not been reported. As our understanding of what substrates and processes the mycobacterial Clp protease is responsible for degrading is still limited, there is a need to identify more roles of this essential protease. Additionally, by having a better understanding of the interactions of this protein complex, any drug development targeting this complex, can

then either be accessed based on disruption of these substrate degradation pathways, or can mimic and exploit the interactions identified.

In this study, we first looked at the adaptor ClpS to identify substrates interactions and to assess the effects ClpS has on the ClpC1P1P2 Clp protease. Using an expression system designed for *Mycolicibacterium smegmatis (Msm)*, we expressed mCherry substrates with variable N-termini to determine variations in degradation because of N-end rule degradation. Using this model substrate we are confident we have identified a pattern of N-end rule degradation in *Mycobacterium smegmatis*.

(Part of the work presented in this chapter was completed by or in collaboration with other members of the Schmitz Lab and will be denoted as such in both the section and the remainder of this work.)

2.2 Materials and Methods

2.2.1 Expression and purification of Clp proteins

DNA encoding ClpS was amplified from *M. smegmatis* MC^2 155 gDNA (ATCC) by PCR and introduced into a pET-22b-derived plasmid (EMD) in frame with an upstream 6xHis tag, MBP (maltose binding protein) tag, and SUMO (yeast Smt3) sequence. DNA encoding the unfoldase ClpC1 was amplified from *Msm* cDNA and cloned in frame with an N-terminal H_7-SUMO tag within pET-22b-derived parent vector. Both ClpP1 and ClpP2 were amplified from *Mtb* H37Rv gDNA (ATCC) in frame with a downstream H_6 tag and cloned into a pET22b vector. Genes were introduced to their parent vectors by Gibson Assembly [90]. All constructs were expressed using *E. coli* strain ER2566 (NEB) and were grown in 1.5xYT media at 37°C. Cultures were grown to exponential phase (OD_{600}: 0.8-1.0) and then induced

with 0.5 mM IPTG (Isopropyl-β-D-thiogalactoside) at 18°C overnight (8-10 hours), 25°C for 5 hours, or 30°C for 4 hours. Cultures were collected and spun down at 4000×g for 30 minutes at 4°C. After spinning down the remaining media was removed and the pellet resuspended in 25 mL/1 L pelleted culture in Lysis Buffer (25 mM HEPES pH 7.5, 300 mM NaCl, 10 mM imidazole pH 7.5, 10%glycerol). For expression of ClpS 1 mM PMSF was introduced prior to lysis. Lysis was done through sonication and the resulting samples clarified by centrifugation at 15000×g for 30 minutes. Lysate was introduced to Ni-NTA beads, which were incubated shaking for at least 1 hour (max overnight shaking at 4°C). Incubated beads were introduced to a column and washed with 25 mM imidazole wash buffer and eluted with 300 mM imidazole elution buffer. Verification of elution and following purification steps was done through SDS-PAGE on a 15% acrylamide gel and visualized by Coomassie staining. Following confirmation of protein containing elution, samples were spin concentrated at 4000×g (3000 or 10000 MWCO Amicon Spin concentrator; Millipore Sigma). Further purification (excluding ClpC1) was accomplished by ion exchange chromatography (Source 15Q 10/100) using low salt Buffer A (25 mM HEPES pH 7.5, 50 mM NaCl, 10% glycerol, 0.1 mM EDTA) and high salt Buffer B (25 mM HEPES pH 7.5, 750 mM NaCl, 10% glycerol, 0.1 mM EDTA) and confirmed lysate was then spin concentrated. H7-SUMO tags were removed using Ulp1 SUMO protease and incubating samples at 37°C for a minimum of 1 hour[91,92]. Protein was purified by size exclusion chromatography (Superdex 200/75) into desired buffer (typically CPD buffer: 25 mM HEPES pH 7.5, 200 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA) and protein-containing fractions were combined and spin concentrated to a

final volume. Protein concentration was measured through absorbance of sample at 280nm using NanoDrop1000 (ThermoScientific).

2.2.2 *In vitro* degradation assays

In vitro assays were conducted in CPD buffer with 1 µM ClpC1 hexamer, 1 µM ClpP1 and ClpP2 tetradecamer, and 10 µM GFP-based substrate. ClpS adaptor when present was at a concentration of 10 µM. Agonist peptide of 50 µM Z-Leu-Leu-Nva-CHO (benzyloxycarbonyl-L-leucyl-L-leucyl-L-norvalinal) or Bz-Leu-Leu with 20 mM ATP along with an ATP regeneration of 187.5 U mL^{-1} pyruvate kinase with 50 mM phosphoenolpyruvate (Sigma)[28,50]. Samples were loaded into 384 well flat black plates for assays. Samples were measured for emission at 510 nm with an excitation of 450 nm (based on GFP based substrates).

2.2.3 Anisotropy Assay

Fluorescent model peptides were ordered from Biomatik with the sequence X-Leu-Phe-Val-Gln-Leu-Ala-Ser-Lys-Cys-Ser-Gly-Ser-Lys^{TAMRA}, where X was varied to the canonical N-end rule amino acids Phe, Tyr, Trp, Leu or the non-N-end rule amino acid Ser[21,93]; TAMRA denotes a C-terminal tetramethylrhodamine fluorophore. Samples consisting of 0.1 µM fluorescent peptide and varying concentrations of purified ClpS in 50 µM total CPD were prepared in 96 well flat black plates (Corning) and fluorescence anisotropy was measured with a TECAN Spark plate reader. Assays featured a 2-fold dilution series of ClpS starting at 100 µM over 11 rows of the plate with a 0µM control lane. Data were plotted in Prism (Graphpad) and fit to a single site binding model by nonlinear regression.

2.2.4 *M. smegmatis in vivo* N-end degradation of fluorescent substrates

Mycobacterial pTH shuttle vectors were constructed by Gibson Assembly[90] from the following synthetic DNA fragments (Twist Biosciences): a ColE1 *E. coli* origin of replication[94], a pAL5000 mycobacterial origin of replication [95], a hygromycin resistance cassette, a TetR tetracycline expression cassette driven by the Pimyc promoter, and a SUMO-X-mCherry^{myc} cassette driven by the tetracycline-inducible Pmyc1-tetOn promoter (Fig 3.2A). An parallel set of pTHU vectors were also constructed that additionally contain an expression cassette for the *Chaetomium thermophilum* Ulp1 SUMO protease [96] driven by the constitutive *M. tuberculosis* Rv0005p promoter [97]. The junction between SUMO and mCherry encoded the 12-residue sequence XLRVQSGTASGT[98] [99], versions of the plasmid were constructed where the initial "X" position was varied to each of the 20 amino acids. Plasmid sequences were verified by Sanger (Genewiz) or whole plasmid sequencing (Plasmidsaurus).[100,101]

The pTH plasmids express intact SUMO-tagged mCherry constructs, in which SUMO effectively shields the "X" position. Expression the same reporter constructs from pTHU plasmids results in removal of the shielding SUMO by the co-expressed Ulp1 enzyme, revealing the new N-terminal "X" position (except for Pro, which prevents cleavage by Ulp1).

Plasmids were electroporated into *M. smegmatis* (MC^2155) at 2.5 kV using a MicroPulser (BioRad). Liquid cultures (6 mL) were grown in Middlebrook media (HiMedia) supplemented with 0.2% glycerol, 0.05% Tween80 and 50 µg/mL hygromycin at 37°C for 48-72 hours. Growth was then measured at an absorbance of 600 nm, and starter cultures of 400 µL at 0.05 OD_{600} in fresh Middlebrook media with 50 µg/mL hygromycin were generated for loading into 96 well black assay plates with

a clear bottom (CORNING). Individual wells had a final volume of 100 µL. Included in the plate was a media blank and a mCherry substrate reference at a concentration of 250 nM, which was used to calibrate detection sensitivity. Although mCherry reporter constructs were created under the control of an anhydrotetracycline (aTc) sensitive promoter, we found substantial leaky expression in the absence of aTc, and ~50% increase in the presence of 100 ng/µL aTc; for consistency, all experiments were carried out with inducer. Plates were preincubated at 37°C for one hour shaking prior to loading into the plate reader. Fluorescence of mCherry was measured by excitation at 580 nm and emission at 610 nm, and cell density was measured 600 nm absorbance every 15 minutes over the course of 16 hours. Plates were agitated through orbital shaking prior to each measurement.

2.2.4.1 Immunoblot verification of mCherry expression and cleavage

Initiating starter cultures were prepared as detailed. Following initial growth, starter cultures of 40 mL were grown for 10-12 hours at 37°C at which point cultures were transferred to 50 mL falcon tubes and spun down at 4000×g for 15 minutes. Supernatant was then removed and sterilized prior to disposal and pellet resuspended in 3 mL lysis buffer (25 mM HEPES pH 7.5, 300 mM NaCl, 10 mM imidazole pH 7.5, and 10% glycerol). This resuspended pellet was stored at -80°C until samples were needed. Once ready 650 µL of pellet slurry was then mixed with 325 µL 0.1 mm zirconia/silica beads (BioSpec) into 1 mL microcentrifuge tubes. Samples were lysed by bead beating using Disruptor Genie (Scientific Industries) with 1 min shaking followed by 30 seconds of rest for samples on ice. This cycle of lyse and rest was repeated a total of five times. Samples were then centrifuged at 4000×g for five minutes and the resulting supernatant collected and stored at -20°C. Supernatant was

run on SDS-PAGE gels and then transferred to a polyvinylidene difluoride (PVDF) membrane. Blotting was done in 5% skim milk rehydrated in TBS (20 mM Tris, 150 mM NaCl, pH 7.5) for one hour and then washed three times in TBST (TBS supplemented with 0.1% Tween20). Both primary and secondary antibodies were generated at 1:4000 dilutions of stocks using a 3% skim milk in TBS. Following blocking and washing, the membrane was then incubated in primary antibody (mouse anti-cMyc antibody 9E10 (Invitrogen)) solution for 1 hour. Another TBST 3X washing regiment is done prior to incubation in secondary antibody (goat anti-mouse IgG-HRP conjugate (Invitrogen)). Blots were treated with SuperSignal West Pico PLUS chemiluminescent substrate (Thermo) and imaged using BioRad Chemidoc.

2.3 Results

2.3.1 Sequence and Structural comparison of mycobacterial ClpS to *E. coli* ClpS

. To start to understand the roll of mycobacterial ClpS, we examined the occurrence of ClpS orthologs among families in the phylum Actinomycetota (synonym: Actinobacteria). Using *E. coli* ClpS as a query, we used NCBI BLAST to search for similar sequences in individual actinobacterial families and found clear ClpS orthologs in 13 out of 15 families, suggesting that ClpS sequences are widespread and perform functions that drive evolutionary conservation. We next constructed the pattern of sequence conservation overall across actinobacteria, and particularly in ClpS orthologs from *Mtb, Msm,* and the proteobacterium *E. coli. E. coli* ClpS is known to consist of a conserved globular ~70 residue C-terminal core domain preceded by a hypervariable ~20 residue N-terminal extension[89]. This extension is hypothesized to be involved in substrate handoff to ClpA, but for the

purposes of this work we focused on the well characterized substrate binding pocket, and unfoldase binding regions of the core domain (Fig 2.1). Based on the alignment, the 75 residues following the first 20-25 are well conserved among actinobacteria and roughly 30% identical to those of ^{Eco}ClpS. Of this 30% the area of the substrate binding pocket showed high similarity while the unfoldase binding region was variable, even across the actinobacterial phylum.

Figure 2.1) Phylogenetic and sequence analysis of ClpS orthologs

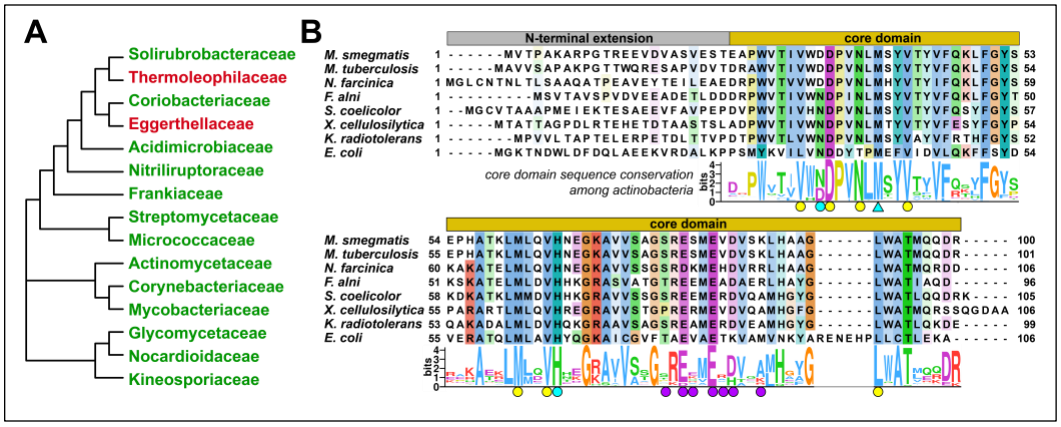


Fig 2.1: The presence of the protein ClpS within multiple species of bacteria has been extensively reported and can be visualized in part A) with a phylogenetic map between actinobacterial species containing ClpS (map based on ClpS sequence homology). The level of sequence conservation of ClpS (part B) between actinobacteria is high, with 97% homology between our working model (*M. smegmatis*) and disease focus (*M. tuberculosis*). Between *E. coli* and *Msm* there is 30% homology across the whole sequence but increases at points of significant function for the ClpS protein. Sequence similarity at key points of interest including the binding pocket (yellow), the amine residue binding region (cyan) and unfoldase binding region (purple) was also high even when comparing actinobacteria to *E. coli.* Conservation of residues is visualized with the cartoon below the alignment.

Figure 2.2) Structural Analysis of *Eco*ClpS (PDB:1MG9[102]) and *Msm*ClpS

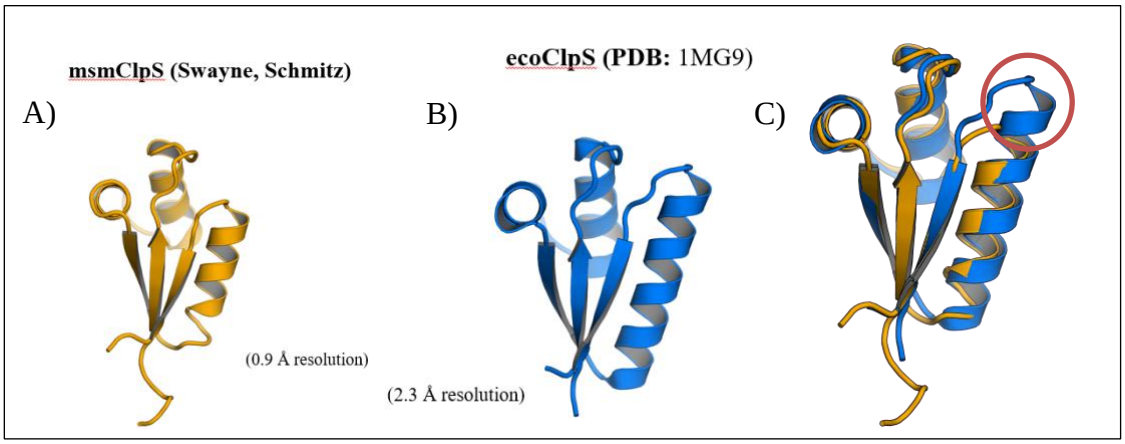


Fig 2.2: A) Solved crystal structure of *Msm*ClpS at 0.9Å resolution, obtained from Dr. Karl Schmitz and Thomas Swayne in collaboration with this project. B) Known (PDB:1MG9) crystal structure of *Eco*ClpS to a resolution of 2.3Å. C) An overlay of structure from part (A) and (B). The red ciricle denotes the only point of significant deviation of the two structures (an extra helical turn on *Eco*ClpS).

2.3.2 Protein Crystallography of *Msm*ClpS

With this conservation in mind, a truncated ClpS variant lacking the N-terminal extension was crystallized and its structure was determined by X-ray crystallography to a resolution of 0.97 Å. The structure was solved through molecular replacement using *Eco*ClpS (PDB:3O2O) as a model[93]. During the screening for optimal cystallization conditions it was revealed that divalent metals are required to stabilize the protein crystals. Attempts to remove Ni^{2+} from the buffer condition (0.1 M HEPES, 12% PEG3350, 5 mM $NiCl_2$, pH 7.5) using EDTA as a chelator resulted in destruction of the crystal. Supplementation of other divalent metals (Co^{2+}, Zn^{2+}, Mg^{2+}) seemed to restore crystal stability. While work exists showing a relationship between ClpS and Zinc, the role of other divalent metals as it relates to ClpS function has not been extensively explored[83].

Three-dimensional alignment of the *Msm*ClpS structure with published *Eco*ClpS structures revealed a highly similar fold, with a very low backbone RMSD of 0.9Å. The crystal structure of *Msm*ClpS compared to *Eco*ClpS showed only one noticeable difference. In the structures was a 6-residue gap that is present in many actinobacterial sequences including *Msm*ClpS, which corresponds to a loss of about one turn in the last α-helix and a shorter preceding connector to the prior β-strand. This position, however, does not appear to be involved in either substrate or unfoldase binding.

To test the ability of *Msm*ClpS to bind to N-end rule amino acids, apo crystals of *Msm*ClpS were soaked with peptide ligands (Phe-Ala, Tyr-Arg, Trp-Ser, and Leu-Thr) in conditions containing either Ni_2Cl or Mg_2Cl. In the resulting crystal structures, peptides were observed to bind similarly as to *Eco*ClpS in that the α-carbon of N-end rule substrates rests at the edge of the binding pocket and the hydrophobic N-end rule

side chain packs neatly in the hydrophobic pocket[82,93,103]. Interestingly, both Ni^{2+} and Mg^{2+} were able to stabilize binding of peptide N-terminus to the lip of the binding pocket. Of these, free Ni^{2+} is likely present in vanishingly small concentrations within mycobacteria, and, if they function equivalently in stabilizing N-end degron binding, Mg^{2+} is more likely to play this role in cells. Other divalent metals including calcium, magnesium, and zinc are present in the cell, and are important to the cells survival, though information on the specific roles of these divalent metals in bacteria, especially mycobacteria is rather limited[104,105]. Additionally, the possibility of this position being occupied by a water molecule, in vivo, is even more reasonable. Curiously, ClpS has been reported to deliver the substrate CobW to ClpC1P1P2 for destruction in a Zn^{2+}-dependent manner, but it is not yet clear if this involves binding of metal to the same site. [106]

Figure 2.3) Crystal structure of *Msm*ClpS co-crystallization with N-end rule substrates

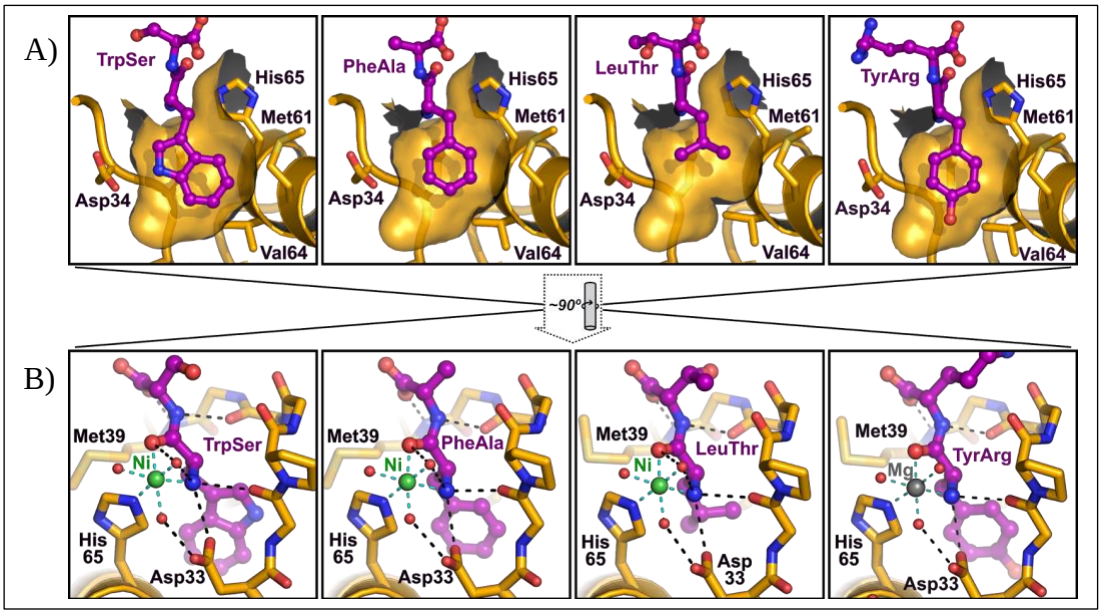


Fig 2.3: The ability of ClpS to bind N-end substrates is essential for N-end rule degradation and was tested (A) by co-crystalizing *Msm*ClpS with canonical N-end rule substrates (Tyr, Trp, Phe, Leu). All substrates appear to occupy the binding pocket of ClpS. To stabilize the crystals, divalent metals were required and appeared within the binding pocket shown in part B. The physiological significance of divalent metals facilitating substrate binding by ClpS remains an ongoing question within mycobacteria but was required for the crystallization of ClpS. Structures were generated by Dr. Karl Schmitz and Thomas Swayne in collaboration with this project.

2.3.3 *In vitro* biochemical analysis of ClpS interactions

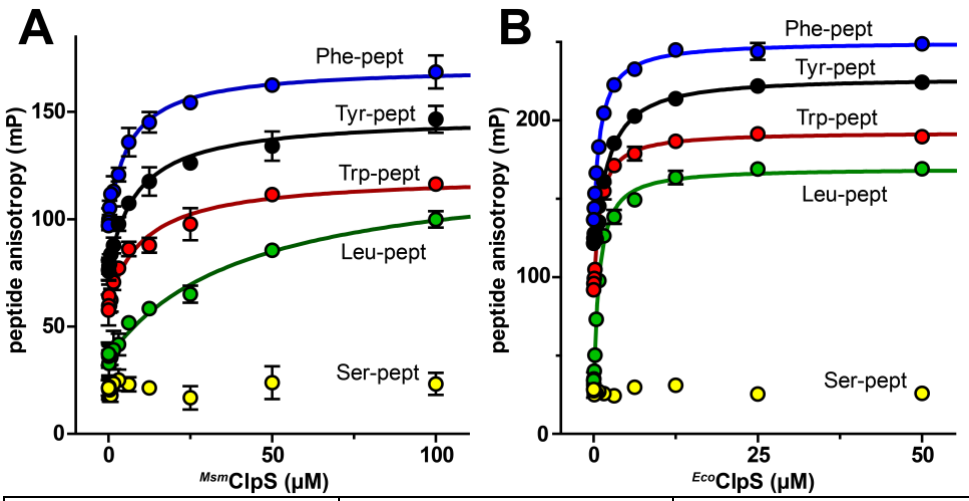
Canonical N-end degrons were observed to bind ClpS in the context of crystal structures but given the high concentrations within protein crystals it remained uncertain whether these bind with physiologically meaningful affinities. To understand the affinity of ^{Msm}ClpS interactions with N-end rule degrons we use fluorescence anisotropy to monitor binding of TAMRA-labeled peptides (Tyr, Trp, Leu, Phe, and Ser [a non-N-end rule control]) to ^{Msm}ClpS. In the anisotropy experiment an increase in anisotropy corresponds to a reduction in tumbling of the fluorescent substrate due to binding to the larger protein. Those fluorescent substrates that featured a canonical N-end rule amino acid at their N-terminus showed increasing anisotropy as the concentration of ^{Msm}ClpS increased, while the control, Ser, maintained a fixed anisotropy value regardless of the concentration of ^{Msm}ClpS (Fig 2.3). This initial result matches those results of other studies looking at ClpS[78,107].

For comparison, another round of anisotropy experiments was carried out to measure binding of the same peptides to ^{Eco}ClpS. (These experiments were carried out by Thomas Swain, a former undergrad in the Schmitz lab.) The results of this *E. coli* anisotropy assay roughly match the results of the mycobacterial anisotropy assays in terms of binding profile, but not in terms of binding strength. Canonical N-end rule amino acids bound to both ClpS variants, while the Ser-bearing peptide did not bind to either. However, the K_d values of N-end rule peptides binding to *E. coli* were ≥10-fold lower (low micromolar) than to mycobacterial ClpS (10 – 50μM; Fig 2.3). The K_d values observed to *Msm* ClpS protein are notably weak and is unlikely that either ClpS or N-end rule substrates are present at high enough concentrations within the cell to saturate binding. However, other factors could affect binding affinity including

upstream sequences altering binding, or even the interaction of ClpS with ClpC1 may increase the affinity of ClpS with substrates[88,93,102,106,108].

To better understand the interaction between ClpS and ClpC1 in mycobacteria a collaboration was done with another Schmitz lab member, Henry Anderson, to use microscale thermophoresis (MST). This technique measures the interaction strength of two proteins based on a fluorescence signal change from one of the interaction partners. This change in fluorescence is induced by altering the temperature of the microenvironment, and can be quantified to interpret binding strength[109]. Recently, Henry has generated data that suggests an interaction of *Msm*ClpS with *Msm*ClpC1 (SuppFig1.1).

Figure 2.4) Fluorescence Anisotropy of N-end substrates in the presence of ClpS adaptor proteins



N-end Residue	Affinity to *Msm*ClpS	Affinity to *Eco*ClpS
Tyrosine	K_D=8.8 +/- 1μM	K_D=2.5 +/- 0.2 μM
Tryptophan	K_D=11 +/- 2 μM	K_D=0.95 +/- 0.08 μM
Leucine	K_D=42 +/- 9 μM	K_D=0.88 +/- 0.07 μM
Phenylalanine	K_D=6.5 +/- 0.9 μM	K_D=1.0 +/-0.08 μM

Fig 2.4: Anisotropy assays of *Msm*ClpS and *Eco*ClpS show interaction with N-end rule substrates. A) *Msm*ClpS bound to canonical N-end rule peptides (Phe, Leu, Tyr, Trp) while showing no binding to a Non-N-end rule control (Serine peptide). B) *Eco*ClpS bound to N-end rule peptides (Phe, Leu, Tyr, Trp) with no binding to non-N-end rule control peptide (Serine peptide). *E. coli* data (B) was collected in collaboration with Thomas Swayne.

2.3.4 *In vivo* analysis of N-end rule degradation in mycobacteria

The interaction of ClpS with N-end rule substrates suggests that N-end rule degradation occurs in the mycobacterial cell, despite the relatively weak binding affinity of ClpS to model N-end rule peptides. Prior studies have also suggested the

existence of an N-end rule in mycobacteria[78,107]. Due to variance in the activity of ClpC1 purified in our lab, we ultimately decided test N-end rule degradation in mycobacteria in a manner similar to the original Varshavsky experiment[21,59]. This experiment was a collaborative effort by Jialiu Jiang and Christopher J. Presloid where contribution of work was equal.

Plasmid construction for this project was led by Drs. Jialiu Jiang and Patrick Beardslee from the Schmitz lab as part of a collaborative effort to improve expression of fluorescent proteins in *Msm*. Substrates with varying N-terminal amino acids and a removable SUMO shield were expressed *in Msm* (Fig 2.4). The general reporter construct design, from N-terminus to C-terminus, incorporated a SUMO tag, followed by a variable amino acid, a constant RSKGEELVTGT[82] sequence, mCherry, and finally a C- terminal myc tag. SUMO-X-mCherry^{myc} expression was controlled via a P_{myc1}tetOn promoter[110], which for our purposes was found to express the target constitutively in the absence of inducer.

To reveal the amino acid of interest following the SUMO tag, the SUMO protease Ulp1 was expressed from a separate promoter in plasmid pTHU. As a control, each reporter construct was also expressed in parallel from a separate pTH plasmid lacking the Ulp1 cassette. Immunoblot of mycobacterial cell lysates expressing constructs from pTHU and pTH confirmed intact SUMO-X-mCherry^{myc} was expressed from pTH, while at least partial cleavage of SUMO was observed in all constructs expressed from pTHU, with the exception of SUMO-Pro-mCherry^{myc}. This failure to cleave this one substrate corresponds to the inability of Ulp1 to cleave a target when Proline is in the position following the cleavage site[111]. The use of a

shielding SUMO tag in conjunction of Ulp1 ensures that the revealed amino acid is the one intended at the variable X position in SUMO-X-mCherry^{myc}.

Figure 2.5) Plasmid design for mycobacterial expression of mCherry substrates with variable N-termini

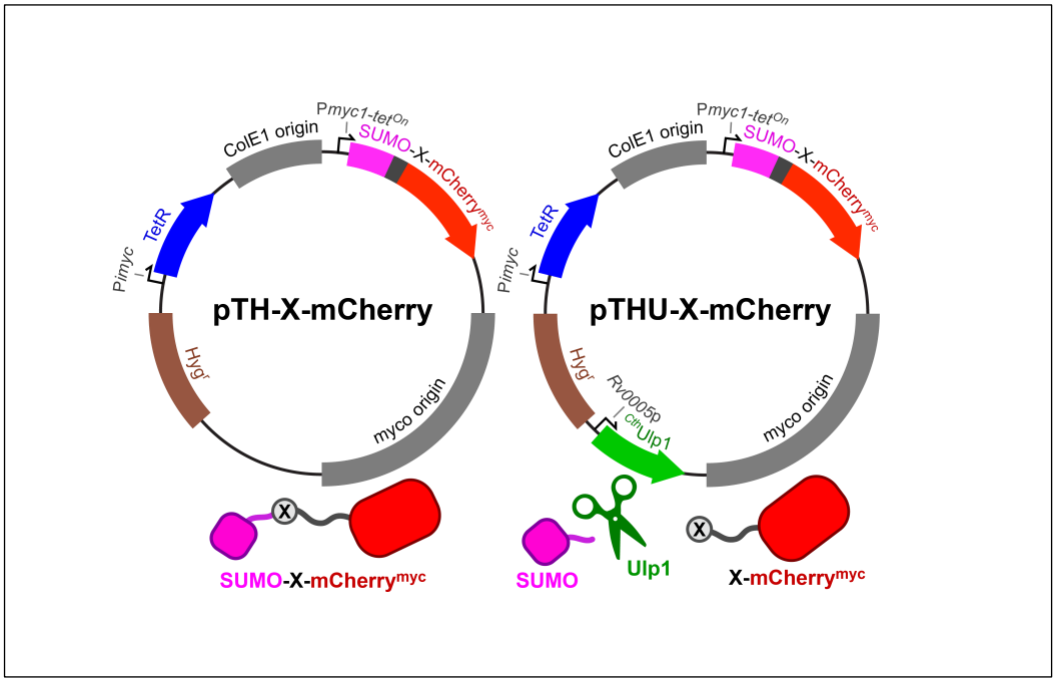


Fig 2.5: To test N-end rule degradation we opted to design two plasmids capable of expressing mCherry substrates. Our control plasmid pTH (left) features an mCherry substrate with a variable N-terminus shielded by a SUMO tag preventing recognition of this variable N-terminus. pTHU (right) features this same SUMO-shielded mCherry substrate with the addition of the cleavage protein ULP1, which removes the SUMO shield revealing the N-terminus of mCherry. A cartoon depicting the difference between the final expressed mCherry substrates is below the plasmid maps. mCherry substrates also feature a C-terminal myc tag to assist with western blot analysis. Both plasmids contain a hygromycin resistance cassette, mycobacterial and *E. coli* origin of replication and a Tetracycline inducible promotor.

Overall growth of transformed *Msm* cultures did not significantly deviate from non-transformed controls. Additionally, the introduction of Ulp1 into the plasmid did not affect the growth rate of the cells (Fig 2.6 A, B). During initial testing it was shown that the induction of both Ulp1 and mCherry appeared to be constitutive rather than inducible. While not ideal test assays showed that the level of expression of mCherry was largely unchanged whether aTc was introduced or not (not shown), but for consistency all samples were induced with aTc prior to reading. The same problem existed for Ulp1 however western blot analysis showed that regardless of induction mCherry substrates were completely cleaved (Fig 2.5). With these issues addressed we proceeded with testing.

Figure 2.6) Confirmation of mCherry substrate cleavage in *Msm*, using Western Blot Analysis of *Msm* lysate expressing pTH/ pTHU

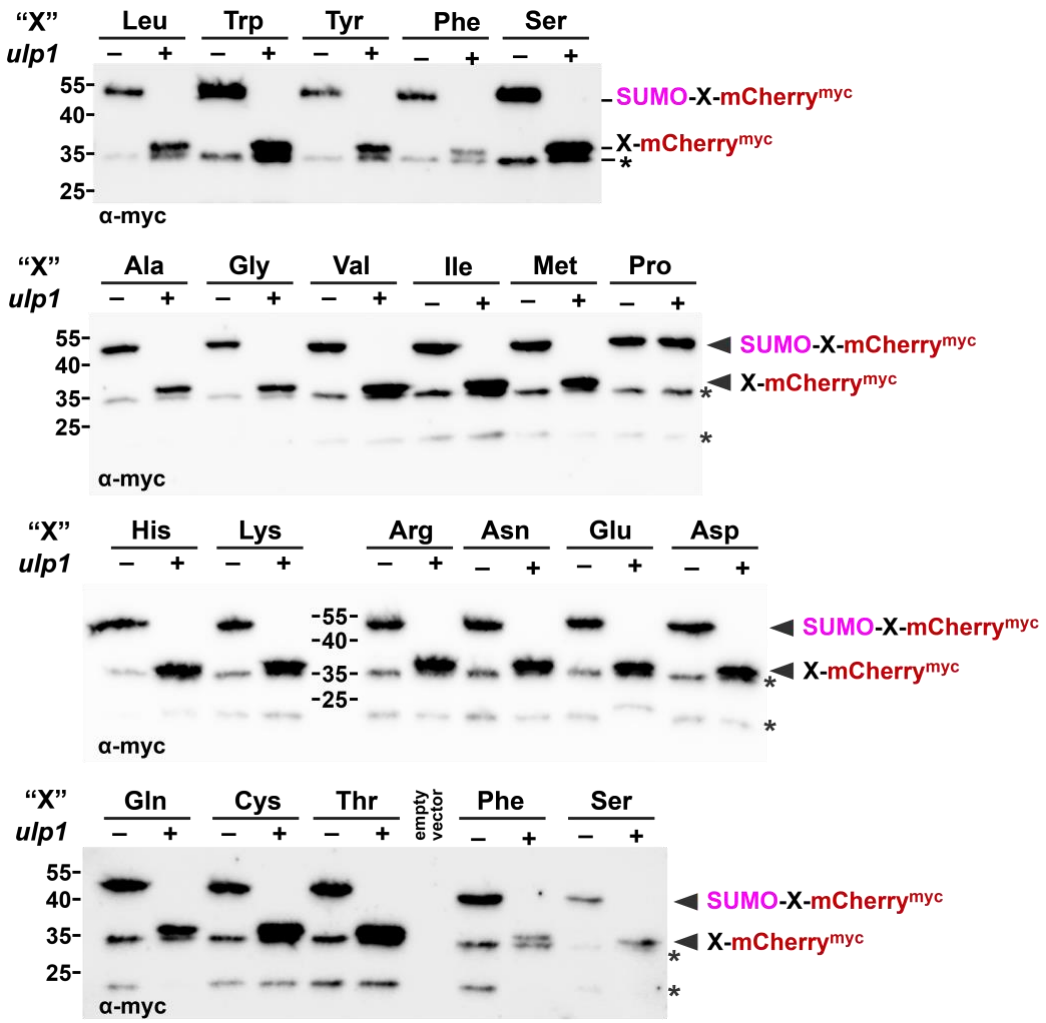


Fig 2.6: In vivo cleavage of SUMO-mCherry substrates was confirmed targeting the C-terminal Myc-tag of the substrates. *Msm* cultures expressing either pTH (-ULP1) or pTHU (+ULP1) were lysed and show that cleavage of SUMO-X-mCherry-Myc (50kDa) was occurring in those samples containing ULP1, resulting in a cleaved X-mCherry-Myc (~35kDA). Only substrates with Proline at the N-terminus of mCherry failed to result in near total cleavage.

Using a TECAN Plate reader, expression of mCherry was monitored by fluorescence at 580 nm excitation and 610 nm emission. This reading was then normalized to the measured cell density. The assay was run over 16 hours to account for any lag in induction or expression. Clear differences in RFU/OD emerged by 10 hours after the start of the assays. As shown in Figure 2.6 parts C and D, a reporter construct presenting an N-terminal Ser showed similar normalized fluorescence over the time course regardless of cleavage of SUMO, whereas a clear reduction was observed in a construct presenting an N-terminal Tyr in cells co-expressing.

Figure 2.7) Growth curves, and *in vivo* proteolysis assays of mycobacterial N-end rule substrates.

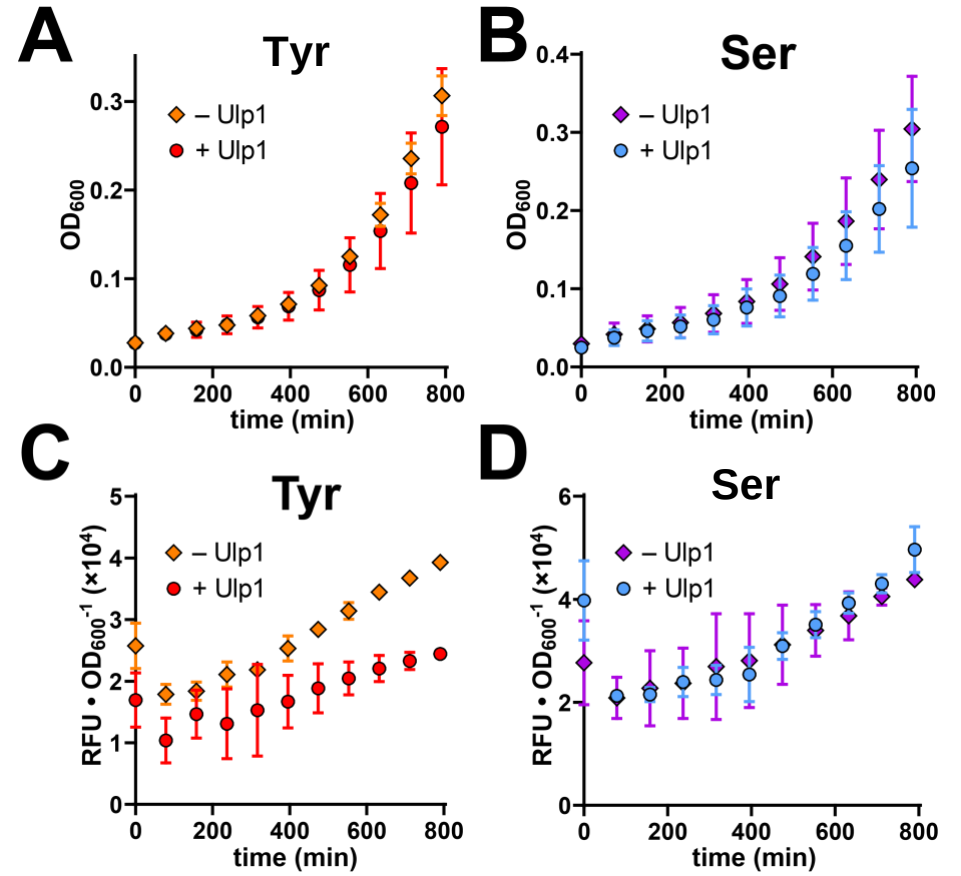


Fig 2.7: A) Growth curve of *Msm* expressing pTH (gold) and pTHU (red) with a revealable Tyr residue. B) Growth curve of *Msm* expression pTH (purple) or pTHU (blue) with a revealable Ser residue. C) RFU/OD_{600} of Tyr bearing substrates. D) RFU/OD_{600} of Ser bearing substrate over 16 hours.

.

Comparing readings at the 10-hour time point, we generally saw no significant difference in normalized fluorescence of cells expressing non-N-end rule mCherry constructs with or without the shielding SUMO tag. However, those constructs containing a canonical N-end rule amino acid (Leu, Phe, Tyr, Trp) showed normalized fluorescence between 44% and 64% lower when Ulp1 was co-expressed (Fig 2.7). This reduction in fluorescence is consistent with N-end rule degradation of substrates presenting canonical primary N-end rule degrons. Notably, due to the inability of Ulp1 to cleave before Proline, we cannot assess the susceptibility of the Pro-bearing construct to degradation in this assay. Furthermore, while significant differences were seen with both Gly and Asp residues (Fig 2.7), the magnitude of these differences were small, suggesting that these are due to anomalies in data collection rather than true N-end rule proteolysis.

Figure 2.8) Normalized RFU/OD600 for N-terminal variants of mCherry show a pattern of N-end rule degradation

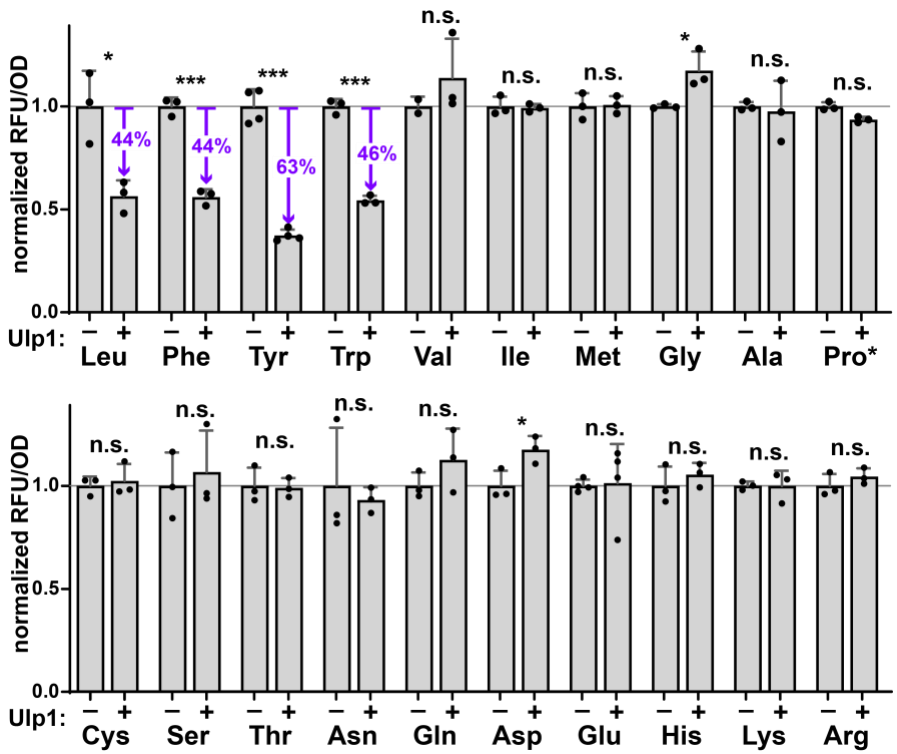


Fig 2.8: To identify whether the different mCherry substrates were being targeted for degradation the relative fluorescence of *Msm* cultures was quantified relative to the optical density of the culture. At 10 hours post induction this value of RFU/ OD_{600} was recoded and displayed above as a bar graph. Based on these values four residues displayed visually and statistically significant differences (Tyr, Trp, Phe, Leu). As the only difference between mCherry substrates was the residue at the N-terminus it appears these four residues comprise the primary amino acids for N-end rule degradation in *Msm.*

2.4 Discussion

2.4.1 Sequence and Structural analysis of ClpS

Based on sequencing searches and comparisons, actinobacteria appear to have a strongly conserved ClpS protein that shares a significant sequence similarity to proteobacterial (*E. coli*) ClpS. As the similarity in sequence is especially high in positions associated with the ClpS protein's adaptor function to facilitate N-end rule degradation we are confident that actinobacterial ClpS serves a similar function. Results from other groups have reached a similar conclusion using a range of methods including biochemical assays and independent sequence and structural analysis[107,112]. Of the two areas of interest to this study, the binding pocket and the unfoldase binding region, the binding pocket maintains a much greater level of sequence conservation than the overall sequence (37%) both across actinobacteria, but also in relation to the *E. coli* sequence. This (~95%) conservation at the binding pocket would support our hypothesis that binding of N-end rule substrate is possible with mycobacterial ClpS. This high conservation also extends to the residues surrounding the binding pocket that are responsible for interacting with help to coordinate the binding of substrate through interactions with the peptide backbone. The unfoldase binding region, however, does not maintain a strong level of conservation between actinobacteria and proteobacteria based on our search. Additionally, it appears that this region is fairly disordered across the actinobacteria phylum at large. The variance between ClpS sequences of actinobacteria like *M. smegmatis* and *M. tuberculosis* is not completely unexpected given this region is responsible for binding ClpA in *E. coli* and the mycobacteria of interest lack this unfoldase.

However, the lack of conservation in the putative unfoldase interaction surface across the larger actinobacterial phylum is somewhat surprising. While there is some level of conservation at two positions in the unfoldase binding region (Fig 2.1), most residues share little similarity. As ClpS appears to be able to bind N-end rule substrates, this adaptor should have a destination for these bound peptides which, if it functions as it does in the canonical *E. coli* N-end rule pathway, should be an unfoldase. It is interesting that the "destination unfoldase" may in fact vary across actinobacteria, or at least the interaction surface would vary among a class of unfoldases. While multiple ATP-dependent unfoldases occur in actinobacterial, ClpC (termed ClpC1 in mycobacteria) shares strongest similarity to *E. coli* ClpA [56,113]. Work by other groups has found that ClpS can function as an adaptor for ClpC1, and while some testing has shown some level of N-end rule degradation through this interaction the entire profile of the N-end rule has not been previously characterized[107]. It is important to note that the N-end rule pathway requires an interaction with ClpS to facilitate substrate delivery while also adjusting the "degradome" of the unfoldase ClpA to prioritize the degradation of N-end rule substrates. However, this adjustment in degradation profile has yet to be observed in the mycobacterial Clp protease.

Interestingly, ClpC is one of three groups of related actinobacterial unfoldases that all possess similar N-terminal domains (NTD): ClpC, ClpB, and ClpI[114]. ClpC is present in most actinobacterial, is known to cooperate with ClpP1P2 in proteolysis, and possesses a highly conserved (NTD). ClpB is also widely distributed across the phylum, and operates independently as an unfoldase and disaggregase during stress responses[76,114,115]. A third group of these unfoldases, termed ClpI, was recently

discovered by the Schmitz lab. ClpI proteins share features of ClpB and ClpC, and individual ClpI orthologs are thought to have separate roles as independent unfoldases or as protease components[114]. ClpB and ClpI have more variable NTD sequences than ClpC, but it is possible that these enzymes cooperate with ClpS in some species, and account for the apparent diversity in the putative interaction surface of ClpS.

When comparing the overall structures of the core domain of ClpS between *E. coli* (PDB:3O2O[113]) and *M. smegmatis* we saw a notably low backbone RMSD (<1 Å). This low level of difference between these proteins suggests that overall function should be maintained. Of the differences, only a small sequence resulting in an extension of an alpha helical structure (β1-α1 loop) outside of the regions hypothesized to be responsible for N-end rule degradation is significant and present in the *E. coli* structure (Fig1.3). In co-crystal structures with N-end rule peptides (Fig 2.2) we see binding of N-end rule amino acids within the ClpS binding pocket to a similar degree of those structures seen in other bacteria[89,116]. Overall orientation of the substrates has the α-carbon of the N-end residue positioned roughly in line with the entry of the binding pocket with all residues (Tyr, Phe, Leu, Trp). This interaction has the side chains of the substrate projecting into the binding pocket. The overall size of the side chains seems to roughly correspond with the strength of the interaction between ClpS and the substrate (Fig 2.3).

Uniquely, crystals of *Msm*ClpS only formed in our crystallization condition in the presence of Ni^{2+} and were only stable in the presence of similar divalent metal ions. The interaction between the N-terminus of the substrate and Asp33 of ClpS is facilitated by a salt bridge featuring a divalent metal stabilizing the interaction (Fig 2.2). While these ions may stabilize the binding of peptides, we note that their

occurrence here – and particularly the presence of the uncommon transition metal Ni – may be an artifact of the crystallization condition. Additional Ni^{2+} ions were observed at key positions in the asymmetric unit, bridging electrostatic interactions between copies of ClpS within the unit cell, suggesting that they play a role in stabilizing overall crystal packing. This would explain why chelation of these ions by EDTA caused these crystals to crack and dissolve. While significant for crystallization, the transition metal ions at the N-end rule binding site may only occur because of their presence in the condition.

The unique position of this divalent metal (Fig 2.2) in the substrate adaptor interaction raises a question of physiological relevance. In the cell, Mg2+ ions are far more abundant, and are more likely to have a role in stabilizing peptide binding. Initial testing by fluorescence anisotropy (Fig 2.8) suggests that the absence of divalent metals reduces the affinity of ClpS for N-end rule substrates. However, we also observe that Mg^{2+} supports peptide binding similar to Ni^{2+} (Fig 2.8), suggesting that trace metal ions are not required for N-end rule proteolysis in these bacteria. Further testing of this phenomenon falls outside of the scope of this final study but is a meaningful observation that was obtained during this thesis work.

Figure 2.9) Fluorescence anisotropy of Tyr-GFP in the presence *Msm*ClpS under restrictive divalent metal conditions

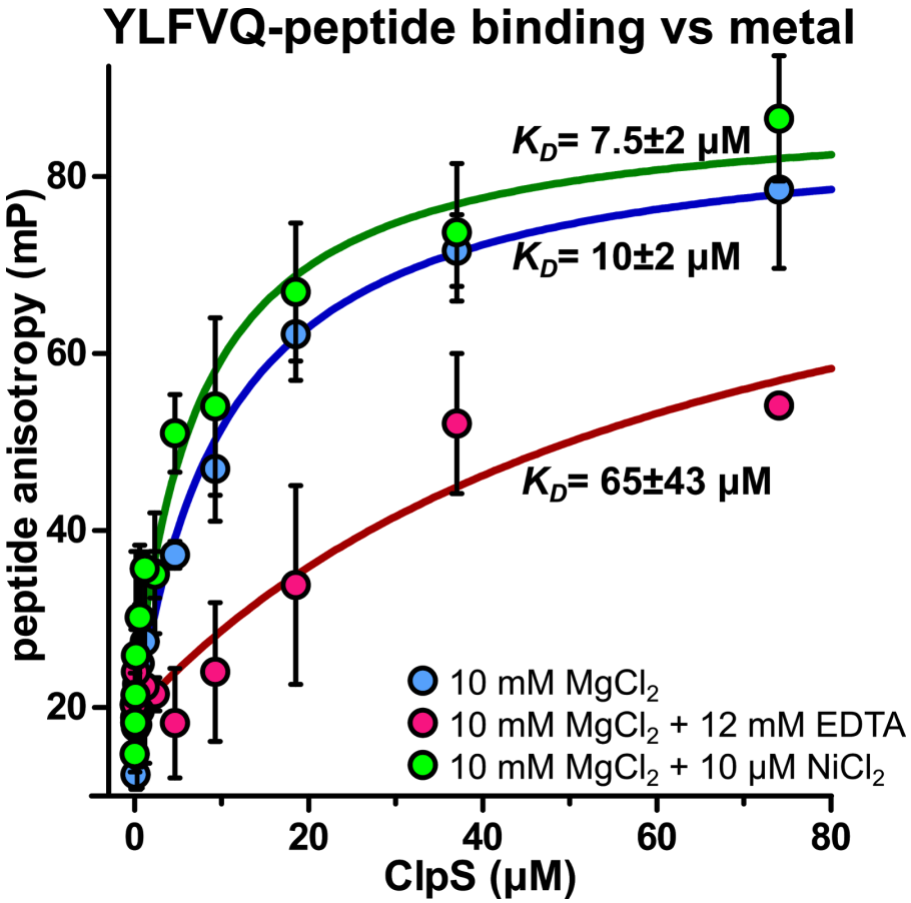


Fig 2.9: Comparison of standard $MgCl_2$ conditions for anisotropy assay (blue), to buffer conditions containing $NiCl_2$ (green) displayed slightly improved binding affinity (kDa). The chelation of metal through the addition of EDTA (red) reduced binding affinity significantly (>50µM).

In an effort to access the level of conservation along interaction surfaces the sequence conservation of actinobacterial ClpS was mapped onto the surface of *Msm*ClpS using Consurf (Fig 2.1)[117,118]. The conservation previously observed extends beyond structural elements and extends to surface interactions, especially within the hydrophobic substrate binding pocket. As was observed with the divalent metal positioning conservation of these residues seem to help stabilize the interaction with N-end degrons. Further conservation of negatively charged residues along the "ringed edge" of this binding pocket may further help in coordinating the interaction with the positive termini of N-end rule motifs.

As we continued to access the conservation using the Consurf map the interaction surface for unfoldases again came into question. Using co-crystal structures of *Eco*ClpS with *Eco*ClpA (PDB: 1MG9, 1R6O)[102,119] we found the corresponding surface residues on the map of *Msm*ClpS and accessed conservation. There was overall very little conservation of this section of α3 helix with only Glu78 and Glu81 being conserved. While conservation of more of this interaction surface would be ideal these equivalent positions in the *E. coli* ClpA/ClpS protein complex are essential stabilizing contacts (positions Glu79 and Glu82 of *Eco*ClpS). It is possible that these two conserved points are structurally essential for adaptor unfoldase interaction with the surrounding residues making up a "lock and key" profile of ClpS interaction partners. Given the variability of Clp proteins across actinobacteria as previously discussed with ClpB and ClpC a profile of ClpI's may require some ambiguity in this interaction surface. More analysis of ClpS interaction partners and co-crystallization utilizing *Msm*ClpS could further clarify these interaction surfaces.

2.4.2 In-vitro biochemical assays

While conducting biochemical assays the purification of proteins was a constant necessity. While $_{Msm}$ClpS can be purified with little deviation from standard purification protocol $_{msm}$ClpC1 appears not to be able to tolerate high salt concentration. Even when not precipitating it appears that high salt concentration appears to damage the overall activity of the ClpC1 sample leading to a reduction or halting of all degradation activity. This variability in protein preps did have an impact on our ability to get replicable data for *in vitro* biochemical assays and helped to drive us to conduct *in vivo* testing of N-end rule degradation.

2.4.2.1 Anisotropy

In vitro protein substrate interactions using $_{Eco}$ClpS with canonical N-end rule substrates (Leu, Phe, Trp, Tyr) show sub-micromolar levels of interaction while using a Ser substrate control. In the equivalent testing done using $_{Msm}$ClpS the interactions, while present for the same substrate residues were not as strongly bound by $_{Msm}$ClpS. While the strength of the interactions is not as strong, the related sequence and structural data paired with these interactions results still support the hypothesis that $_{Msm}$ClpS facilitates N-end rule degradation in mycobacteria.

Of the N-end rule substrates tested Leucine showed the lowest binding strength (42+/- 8.8 μM) to $_{Msm}$ClpS. This low affinity is also supported by structural analysis (FIG) using the crystal structure of $_{Msm}$ClpS bound to a Leu bearing substrate. Leu does not appear to occupy much of the binding pocket due to its smaller sidechain structure than other N-end rule substrate sidechains. While Leu may not be as well recognized or degraded in mycobacteria it should be noted that in *E. coli* the ClpS binding of N-end rule substrates is increased when ClpS associates with the unfoldase

ClpA. A similar improvement in substrate binding may occur in mycobacteria but requires further analysis.

2.4.2.2 Degradation assays

While *in vitro* degradation of a model N-end rule substrate by ClpC1P1P2 has been reported[36,107], several attempts in our hands to collect *in vitro* biochemical data of degradation patterns of the ClpC1P1P2 Clp protease were unsuccessful (not shown). Ultimately the data collected featured proteins of variable activity resulting in data that lacked significance and consistency. With the recent expression of consistently active *Msm*ClpC1 within the Schmitz lab these assays could be reattempted and should be to support the data collected in this thesis work. \

2.4.3 Cell based Screening.

Due to complications with purification of *msm*ClpC1 we decided to look to the initial identification of N-end rule degradation using beta-galactosidase testing and adapting a protocol for fluorescent proteins we have access to[21,59]. Using a SUMO-tag to protect the N-terminal residue of substrates, ULP1 protease cleavage was utilized to expose N-terminal residues (Fig 2.4/ Fig 2.5). We observed degradation that appears to match the N-terminal residues identified in *E. coli*[21,120]. These results paired with the structural and anisotropic data confirm that primary N-end rule degrons for mycobacteria include leucine, phenylalanine, tryptophan, and tyrosine.

N-end rule degradation is characterized by the targeted degradation of substrates based on the identity of their N-terminal residues. Another unique feature in some bacteria and eukaryotes is a wider range of N-terminal residues that can be destabilizing. This range of residues can include primary residues that are themselves

destabilizing, or secondary residues that are targeted by amino transferases to have primary amino acids added to substrates[21,103]. Based on the results of our mCherry expression we are confident that no pattern of secondary N-end amino acids is present in the mycobacterial N-end rule. This may suggest that mycobacteria lack the amino transferases necessary for secondary degrons to be present or added to substrates.

This expression model and the ability to track cell fluorescence as a measurement of cell culture growth has provided a unique opportunity to test candidate substrates within a live mycobacterial culture. This testing can be expanded to include other substrates, but most immediately should be utilized in conjunction with a mycobacterial cell strain that has had *msm*ClpS removed, or inactivated. Without ClpS the N-end rule pattern of substrate degradation should stop and degradation patterns should be uniform regardless of the N-terminal amino acids of the expressed mCherry substrates.

Chapter 3

DEVELOPMENT OF APEX2 PROXIMITY LABELING TO IDENTIFY CLP PROTEASE INTERACTION PARTNERS IN *MYCOLICIBACTERIUM SMEGMATIS*.

3.1 Introduction

Efforts to combat tuberculosis infections have traditionally relied on relatively older but still highly effective treatments of isoniazid and rifampin[12,15,17]. While the overall incidence rate of resistant infections remains relatively stable at 5% the rate of resistant infection for reemerging infections is significantly higher at 17%[1]. This increase in resistant infection especially following a pre-existing infection of *Mtb* is especially concerning as the aftermath of the covid-19 pandemic is addressed. Already data has emerged that suggests major disruptions of tuberculosis treatment and screening have occurred as the result of the pandemic[1]. To better treat both conventional and resistant tuberculosis infections new drug targets must be identified. The Clp protease is an essential class of proteases in mycobacteria that have emerged as one promising target[4,24,29,56,75].

Clp proteases have been extensively studied in many proteobacterial bacteria including *E. coli* but have only relatively recently begun to be characterized in actinobacteria like *M. smegmatis* and *M. tuberculosis*. While the overall structure and "basic" interactions are understood between the unfoldase and peptidase components of the Clp protease (Fig 2.9), the interactions that take place that are essential for the cell are not as well understood[28,37,56]. While general functions of the Clp protease in

other bacteria include protein quality control, stress response and virulence, the protein pathways that facilitate these functions in mycobacteria have yet to be identified completely[76,87,121–123].

Previous work in the lab has used a "trapping" mechanism to identify substrates for the Clp protease[81]. While successful, I wanted to explore a method of capturing interaction partners that both maintained the functionality of the protein complex and captured interaction partners that are not specially substrates. To do this I decided to make use of APEX2 proximity labeling. APEX2 is a modified ascorbate peroxidase from soybeans that has been used in the past to identify interaction partners in a variety of organisms[124–126]. APEX2 does this by facilitating a free radical reaction in the presence of hydrogen peroxide (H_2O_2)[127]. This allows APEX2 to oxidize biotin-phenol generating phenoxy radicals, while also labeling any protein with ~100Å with biotin. These labeled proteins can then be visualized through STREP-HRP staining, and isolated from cell lysate using streptavidin capture chromatography. The captured proteins can then be identified through a combination of biochemical analysis, and mass spectrometry[127].

This APEX2 biotinylation technique has been used in microorganisms but has a history of poor labeling efficiency. Fortunately, APEX2 can work with substrates other than biotin-phenol to carry out proximity labeling. One way to do this is using a Cu catalysis click reaction with either a tyramide-alkyne or tyramide-azide substrate[128,129]. In this reaction APEX2 facilitates a covalent linkage between an a protein of interest with a fluoresceine-labeled probe, in our case FAM[130], by making use of a Cu catalyst and a tyramide alkyne/azide substrate. The result is a profile of covalently labeled interaction partners that can be imaged by SDS-PAGE and

analyzed by Mass Spec[129,131,132]. This method has some advantages in that it does not require the use of biotin as a label, which can be present naturally in the organism (*Msm*) that labeling is taking place.

Figure 3.1) The surface of ClpP1 and ClpP2 show conservation suggesting function

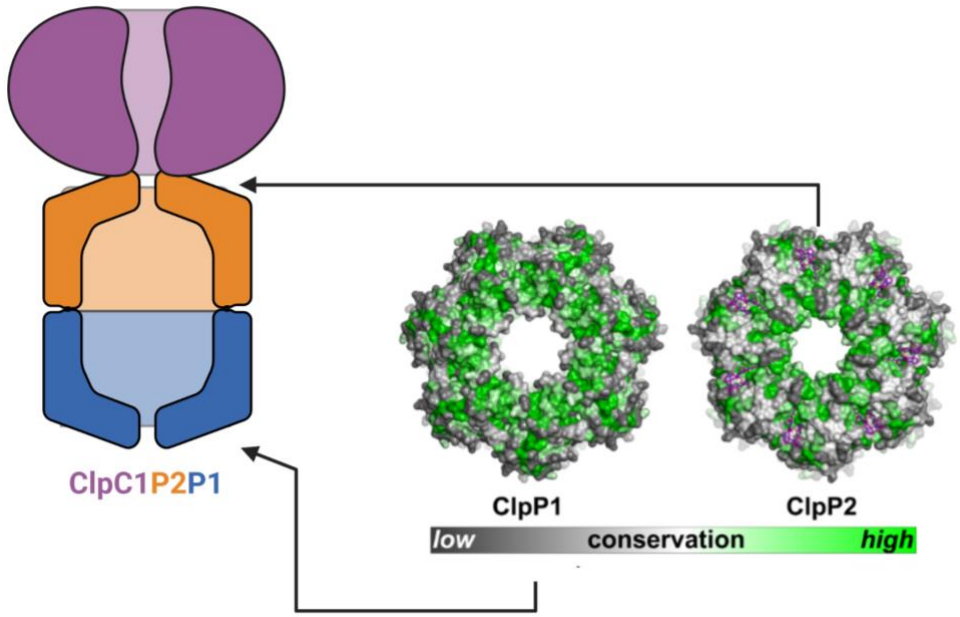


Fig 3.1: The assembly of the mycobacterial Clp protease (left) features a peptidase barrel with two ends. The "top" of this barrel features ClpP2 with a moderate level of conservation (right). ClpP1 makes up the bottom of the barrel and has higher surface conservation than even ClpP2. ClpP1 does not have any known interaction partners at this surface currently. Conservation map was generated using Consurf software to generate conservation values and mapping these values onto known ()structures of ClpP1 and ClpP2. Mapping sequence conservation map was generated by Dr. Karl Schmitz in collaboration with this project.

In this study, we sought to use APEX2 proximity labeling to label interaction partners of the Clp protease in *M. smegmatis*. The development of APEX2 Clp protein fusions were designed and expressed within the desired pTH expression system and preliminary testing suggests viability of the system for in-vivo labeling of interacting proteins. Continuation of this project would see labeled proteins captured and identified through mass spectrometry, and biochemically characterized primarily in the context of the interaction with the Clp protease.

3.2 Materials and Methods

3.2.1 Design and expression of APEX2 fusion constructs

A complete list of constructs can be seen in Table 1. Constructs were first cloned into the *E. coli* expression plasmid pET-22b through Gibson Assembly.[90] All constructs incorporated a FLAG-APEX2 gene fragment derived from plasmid "pcDNA3 APEX2-NES" obtained from the Ting lab[124,126] (via Addgene), and which was placed in frame before or after the protein of interest. Constructs were test expressed in *E. coli* strain ER2566 (NEB) harboring a deletion of ClpP. Following successful generation of *E. coli* constructs in pET-22b, the entire construct was excised by restriction digestion with SpeI and NdeI restriction endonucleases (Thermo) and ligated into mycobacterial shuttle vector pUV15tetORm (Ehrt Lab; Addgene) for mycobacterial expression.

3.2.2 Western Blot Analysis

M. smegmatis was grown in Middlebrook 7H10 media following transformation with pUV plasmids endoging APEX2 fusions by electroporation. This growth was carried out for 48 hours at 37°C. Cells were then subcultured into 15 mL cultures with a starting OD_{600} value of 0.25 for 6 hours at 37°C. Following growth, cells were pelleted through centrifugation (4000 RPM for 10 minutes in 15 mL conical tubes), media was removed and replaced with 3mL of lysis buffer (25 mM HEPES pH 7.5, 300 mM NaCl, 10 mM imidazole pH 7.5, and 10% glycerol). In a 1.7 mL Eppendorf tube, 625 µL of pellet slurry with 325 µL zirconia/silica beads (BioSpec[133]) were combined and lysed by bead beating using a Disruptor Genie (Scientific Industries). Lysis pulses featured 1 minute of shaking (samples) with 30 second rests on ice to reduce heat because of beating. Five total cycles (shake/rest) were conducted followed by centrifugation of sample at 4000xg for 5 minutes. Supernatant was collected and e run immediately on SDS-PAGE gels (20 µL) or stored in 1.7mL Eppendorf tubes at -20°C (or -80°C for longer term storage). Samples were then transferred to PVDF membranes and blocked using a 5% skim milk solution made from powdered skim milk and TBS (20mM Tris, 150mM NaCl, pH 7.5) for 1-2 hours. The membrane was washed with 10mL of TBST (TBS supplemented with 0.1% Tween20) 3 times, and then introduced to 15 mL of 1:1000 primary[134] rabbit anti-Flag antibody solution for 1-2 hours. After washing 3 times in TBST again the PVDF membrane was incubated in 1:2000 secondary[135] goat-anti-rabbit antibody solution for 1-2 hours. Both primary and secondary antibody solutions were prepared in 3% skim milk solution. Blots were then treated with SuperSignal West Pico PLUS chemiluminescent substrate (Thermo) and imaged using a BioRad Chemidoc.

3.2.3 APEX2 Biotinylation

Labeling protocol was adapted from the proximity labeling of membrane bound periplasmic proteins by Ganapathy et.al. [132]. Transformed *M. smegmatis* MC^2155 cells with pTH plasmids containing APEX2 fusion constructs were initially grown for 48 hours in Middlebrook H710 media with 100 ng/µL hygromycin. Starter culture was then sub-cultured into 15 mL samples with a starting OD_{600} value of 0.25. Induced (100 ng/µL anhydrotetracycline) and uninduced samples were prepared for each fusion construct. Samples were incubated at 37°C for 6 hours. Samples were then pelleted at 4000xg for 5 minutes with supernatant removed and 1 mL fresh media introduced to pelleted cells and sample transferred to microcentrifuge tube (2 mL). Following this step samples (cell containing samples) were kept on ice as much as possible. The following steps involve a 3000xg spin for 5 minutes unless otherwise specified and supernatant removed. Biotin phenol at 1 mM was added to samples and incubated shaking at 37°C for 30 minutes. (Optionally, another sample lacking biotin phenol introduction can be included in this workflow to compare labeling efficiency). Hydrogen peroxide at 1 mM was then added to these samples for 1 minute and the resulting reaction was quenched using 1mL of 2x quenching solution (20mM sodium ascorbate, 20mM sodium azide, 10mM Trolox, in PBS buffer). There is no spin between the introduction of hydrogen peroxide and 2x quench but is spun following the addition of 2x quench. Following spin pellets were resuspended in 1x quench (2x quench diluted with PBS buffer), spun down again, and washed a second time in 1x quench solution. Another washing step with 1ml PBS w/ Tween80 (0.05% Tween), followed by a final wash in 1x quench solution, a final spin and final samples are finally resuspended in 2 mL of PBS w/ Tween80. Samples were stored at -80°C or lysed immediately through bead beating and run on SDS-PAGE gel (10%-15%).

3.2.4 Click Chemistry Labeling Reaction

Labeling protocol was adapted from the proximity labeling of membrane bound periplasmic proteins by Ganapathy et.al. [132]. Initial growth and culturing of *M. smegmatis* bacteria transformed with APEX2 fusion constructs (Table 1) is the same as the biotinylation protocol described in section 3.2.3. Change in protocol occurs at the point of biotin phenol introduction. Following initial pelleting of cells and introduction of 1 mL new fresh H710 Middlebrook media, 1 mL of 1 mM tyramide alkyne was added to samples and incubated for 15 minutes at room temperature. Sample was then spun down at 4000xg for 5 minutes, supernatant was removed, and 1 mL of 2x quench solution added to halt labeling. Sample was spun down again as previously described, supernatant removed, and cells are first washed once with 1 mL 1x quench solution and again (following another spin) with 1 mL PBS w/ 0.05% Tween80. Samples were washed a second time in 1 mL 1x quench and spun down before being washed a final time in 1.5 mL of 20 mM TRIS buffer pH 7.4. Samples were then spun down at 4000xg for 10 minutes and pellets resuspended in click chemistry lysis buffer (20 mM TRIS, 150 mM NaCl, 1% Triton x-100, 1% SDS, pH 7.4. Cells are then lysed by bead beating using zirconium-silica beads and Disruptor Genie (see early ref). Lysed samples were then spun at 10000xg for 5 minutes to remove beads and cellular debris. Cell lysate should be immediately labeled to maximize protein capture but can be stored at -80°C if necessary. From the cell lysate, 500 µL of sample were transferred to a fresh 2 mL microcentrifuge tube with 1 mL of labeling solution (5 mM sodium ascorbate, 25 µM 5-fluorescein amidite (FAM) azide (Lumiprobe[136]), 1 µM $CuSO_4$). At this point samples were shielded from light as much as possible, and incubated shaking at 22°C for 1 hour. Finally, samples were

quenched using EDTA to a final concentration of 10 mM based on volume of the sample. Samples were kept in the dark as much as possible and stored at -80°C.

3.2.5 Imaging biotinylated and labeled protein samples.

Labeled protein samples were thawed on ice and 20 µL samples taken from samples for testing. The samples had 3X sample buffer with BME (15% glycerol, 6% SDS, 200 mM Tris-HCl pH 6.8, 0.03% bromophenol blue, 10 % BME) added and were boiled at 95°C for 5 minutes and were loaded into a 12.5% SDS-PAGE gel. Transfer to PVDF membrane along with blocking steps were the same as those detailed in section 2.2.4.1. Following initial 3x wash with TBST, membranes were stained using Strep-HRP at 0.15 µg/mL and incubated shaking for 1 hour at room. Imaging was then done using BioRad Chemidoc.

Imaging of Cu Click chemistry samples was done by running samples on 15% SDS-PAGE gels and imaged using BioRad Chemidoc. Gel should be shielded from light while running as much as possible and kept dark until imaging is complete.

As necessary, confirmation of APEX2 fusion expression was confirmed by western blot analysis using rabbit anti-Flag primary antibody (solution consists of antibody with 3% skim milk (from powder) rehydrated in 1xTBS (200 mM Tris, 1.5 M NaCl)), and goat anti-rabbit secondary antibody (same solution makeup as primary). Blocking solution for western blots consisted of 5% skim milk solution (powder milk) rehydrated with 1x TBS.

3.3 Results

3.3.1 Design and testing of APEX2 protein fusion constructs in *E. coli*.

Our goal was to develop a method of capturing interaction partners of the ATP dependent Clp protease in a manner that does not interfere with the native function of this protein complex. By maintaining the normal function of the complex, we felt confident that we would be able to expand the profile of interaction partners for this complex that has only a handful of identified substrates and even fewer regulatory partners and adaptors.

We opted to generate fusion constructs of different portions of the Clp protease with C-terminal or N-terminal additions of a FLAG-APEX2 gene fragment (Table 1.2). APEX2 is an engineered soybean ascorbate peroxidase adapted from its wild-type gene for use in proximity labeling and microscopy [124]. This APEX2 protein is able to be utilized both for biotin-phenol/hydrogen peroxide labeling as well as labeling through tyramide compounds[126,137]. Gibson assembly was used to create N- or C-terminal fusions of APEX2-flag with protease components of interest in an *E. coli* expression vector. The resulting fusion constructs were test expressed in *E. coli* strain ER2566[138]. In all cases expression levels of constructs appeared to be low, as judged by SDS-PAGE. Nevertheless, we sought to assess the ability of these APEX2 fusions to catalyze proximity labeling in the context of *E. coli* lysate. Strep-HRP of *E. coli* lysate showed faint but present bands (Fig 2.9) representative of labeling. Follow up western blot analysis targeting the FLAG tag present in the fusion constructs confirmed expression of at least partial full-size POI-APEX2 fusions. With this initial validation we decided to proceed with construction of plasmids capable of mycobacterial expression.

Table 3.1) APEX2 Fusion Constructs and Validation Status:

Fusion Construct	**Sequence Verified in pET22b (Eco)**	**Sequence Verified in pTH (Myco)**	**Undergone Biotinylation protocol**	**Undergone Cu-Click Chemistry protocol**
FLAG-APEX2 (control)	Yes	Yes	Yes	Yes
*Msm*ClpC1-FLAG-APEX2	Yes	Yes	Yes	Yes
FLAG-APEX2-*Msm*ClpC1	Yes	No	NO	No
*Msm*ClpS-FLAG-APEX2	Yes	No	No	No
FLAG-APEX2-*Msm*ClpS	Yes	No	No	No
ClpP1-APEX2	Yes	Yes	Yes	Yes

Figure 3.2) Plasmid design for expression of APEX2 fusion protein constructs in *Msm*

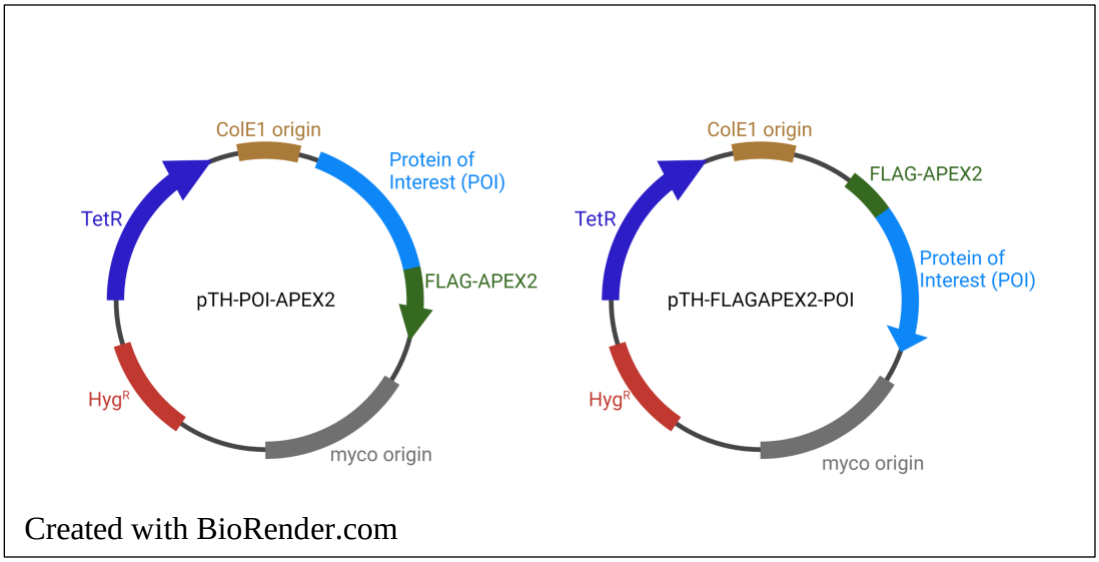


Fig 3.2: With the successful application of pTH and pTHU in the identification of N-end rule substrates in chapter 2 of this thesis work, pTH was modified for the expression of an APEX2-POI fusion construct. The original SUMO-X-mCherry substrate was removed and replaced with the fusion construct of interest (light blue) with either a C-terminal (left) or N-terminal (right) FLAG tag to assist in western blot analysis. Previous key features including mycobacterial origin of replication, hygromycin resistance cassette, and Tet promotor were maintained.

The generation of the mycobacterial plasmids of these fusion constructs (Table 3.1) was done using the *E. coli* plasmids as a template to amplify and excise the POI (protein of interest)-APEX2 fusions. This PCR amplified fragment was digested with Spe1, and Nde1 along with pTH[139] mycobacterial expression vector. These resulting constructs were then sequence verified. Verification of POI-APEX2 fusion expression needs to be confirmed through western blot analysis targeting the FLAG tag of the fusion constructs. Previous attempts to validate have shown inconclusive banding or overblown sample loading (not shown)

3.3.2 Biotinylation optimization of *E. coli* and *Msm* cell lysate

To test the ability the APEX2 to label proteins *in vivo* we decided to first test labeling in *E. coli* (ER2566). While any labeling seen in this initial screen is unlikely to represent meaningful interactions with the fusion constructs containing *Msm* proteins it still offers validity to the functionality of the APEX2 protein in our fusion constructs and offers initial assurance that the fusion constructs are not being cleaved. We transformed *E. coli* (ER2566) with FLAG-APEX2 containing fusion constructs (Table 1) through electroporation and grew samples under antibiotic pressure corresponding to the parent pET22b plasmid.

Figure 3.3) SDS-PAGE and Strep-HRP of biotinylated E. coli lysate

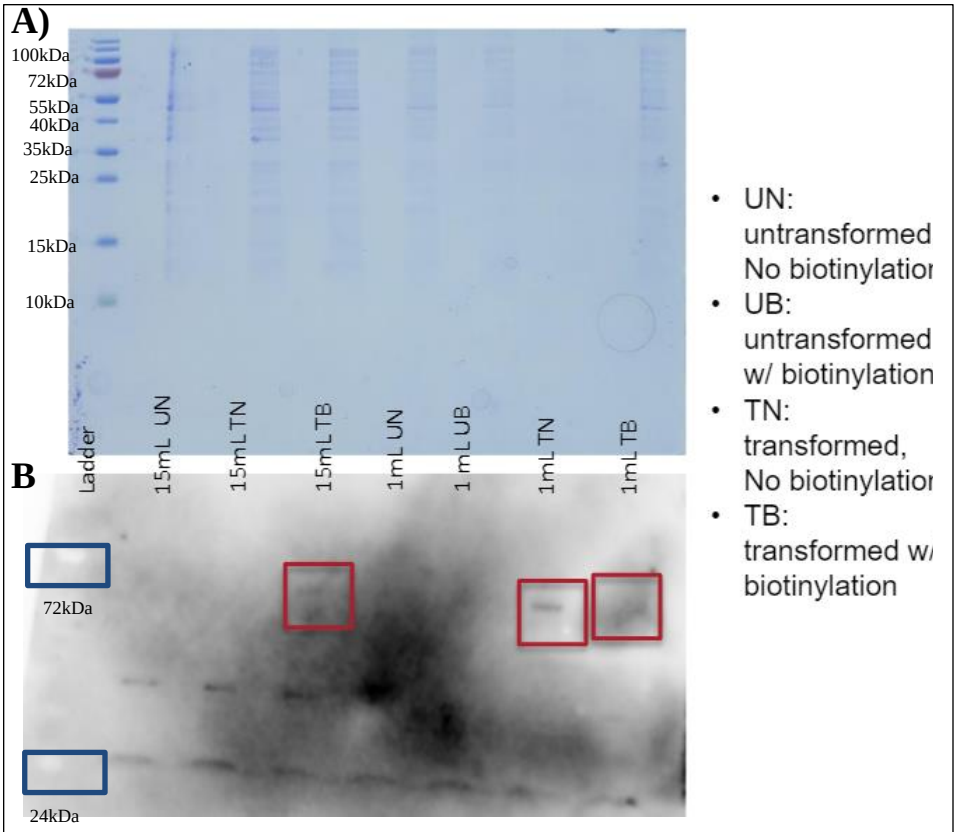


Fig 3.3: To test the ability of our APEX2 fusion constructs we first transformed and expressed these pTH plasmids in *E. coli.* Total cell lysate was first run on a brilliant blue stained SDS-PAGE gel (A) to determine the banding pattern of this total cell lysate. Samples included untransformed (U_) and transformed (T_) *E. coli,* as well as tested with the biotinylating protocol (_B) or not (_N). In the Strep-HRP blot (B) we see individual bands suggesting labeling is possible with this construct, including bands that match our construct size (RED-50kDA).

Initial optimization of protocol involved determination of incubation time and concentration of biotin-phenol, as well as optimization cell lysis methodology (see materials and methods). Following optimization and STREP-HRP imaging of samples it was found that APEX2 was functional within these.

E. coli lysate was optimized for general labeling with biotin-phenol labeling based on labs that initially developed this expression model as well as a group that had utilized the APEX2 proximity labeling outside of *E. coli*[126,137,140]. The resulting labeling protocol (see 3.2.3) was determined to be most optimal for our proteins of interest. The results of this initial labeling reaction (Fig 3.1) revealed distinct bands on PVDF following Strep-HRP staining. Further optimization of Strep-HRP concentration significantly improved the overall clarity of the final imaged blot, with an ideal concentration of 0.15µg/ml This banding pattern, while simply a proof of principle, does show clear bands in a manner similar to those seen by other groups following successful labeling including auto labeling of the fusion construct[132].

The labeling of *Msm* cell lysate was hindered due to incomplete lysis of mycobacterial cells containing our fusion constructs. Our first method of lysing cells for imaging was done using a microfluidizer and seemed to work adequately in the context of *E. coli*. Attempts to grow cells from *E. coli* lysate failed to yield colonies, however lysate from *Msm* did yield colonies. This failure to achieve near total lysis was concerning but resolved by changing the lysis protocol from microfluidics to bead beating using zirconium/ silica beads[133,141]. Subsequent attempts to grow bacteria from bead beating lysate yielded few/no colonies for both *E. coli* and *Msm* (not shown).

With lysis optimized labeling of mycobacterial lysate could continue. Initial labeling using biotin-phenol was carried out using a ClpP1-APEX2 fusion construct in a manner identical to the pilot labeling done in *E. coli*. The resulting image (Fig 3.2) was light and required higher exposure to identify banding but showed clear variable banding in response to treatment with biotin phenol. However, banding was also seen in unlabeled samples. Upon further review we concluded this is likely the result of native biotinylating by *Msm*. While this can be screened out, this native biotinylation gave us reasonable justification to explore APEX2 proximity labeling that does not require biotinylation.

Figure 3.4) Strep-HRP blot analysis of $_{Msm}$ClpP1-FLAG-APEX2 containing *Msm* lysate

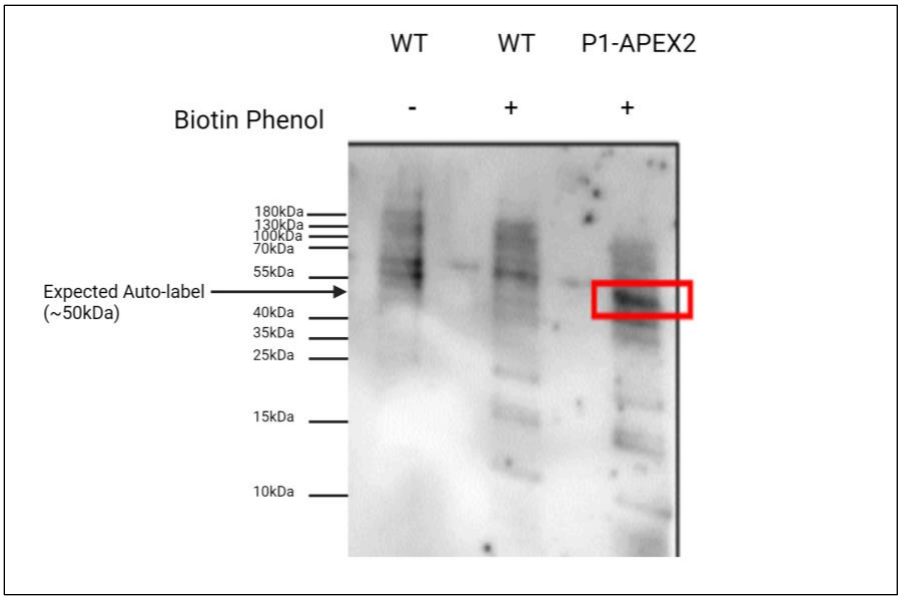


Fig 3.4: Initial testing of biotinylation protocol using biotin phenol resulted in banding suggesting successful auto labeling of fusion construct (ClpP1-APEX2). WT samples lacking pTH plasmid, along with *Msm* transformed with pTH were analyzed with a Strep-HRP blot. Native biotinylation was seen in both WT samples regardless of the introduction of biotin-phenol but lack the intensity of the auto labeled banding seen in the transformed P1-APEX2 sample (red).

3.3.3 Cu-Click Chemistry proximity labeling of *Msm* lysate.

As was shown in our own blot and mentioned in source material, *Msm* and many other bacteria utilize native biotinylation[132,142]. Cu mediated click chemistry is a viable alternative to biotinylation that also uses APEX2, allowing us to use our existing fusion constructs without another round of cloning. Initial protocol was adapted from Ganapathy et al. and piloted using the ClpP1-FLAG-APEX2 construct. The banding pattern observed (Fig 3.3) appears to show multiple illuminated bands (roughly matching those seen by another group[132]) suggesting multiple polypeptides had been labeled with the fluorescent tag (FAM). Bands with the most intensity should correspond to our fusion constructs via auto labeling, however, upon inspecting the sizes of the most prominent bands none appear to match the correct size of the *Msm*ClpP1-FlAG-APEX2 fusion construct. Further optimization of this labeling is required, and validation of the fusion constructs size and integrity need to be addressed immediately. That said the labeling protocol that has been developed appears to work well based on banding pattern and clarity of results (little overblown banding).

Figure 3.5) Cu Click Chemistry proximity labeling using *Msm*ClpP1-FLAG-APEX2 construct in *Msm*

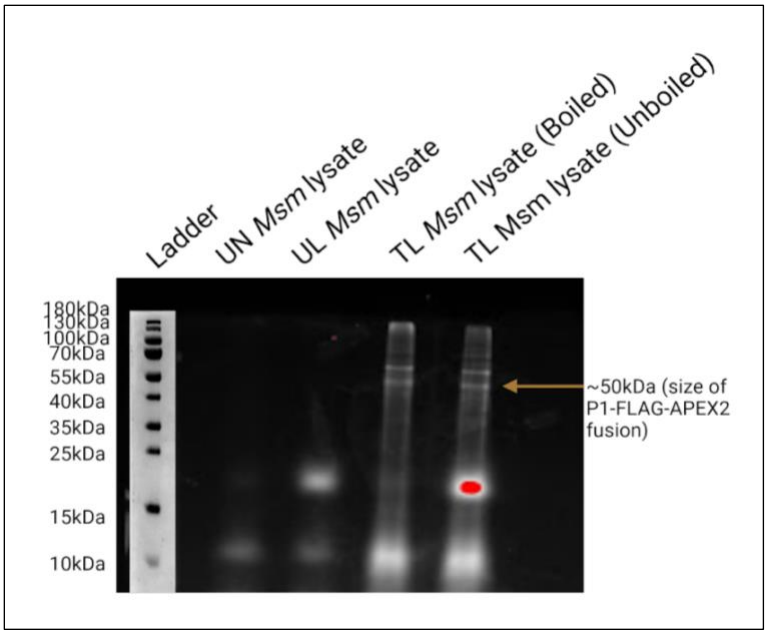


Fig 3.5: Initial testing of Cu-mediated Click chemistry to label the interaction partners of ClpP1 using the fusion construct (ClpP1-FLAG-APEX2) was done in a manner similar to the testing of biotin-phenol mediated labeling. Two WT construct lacking pTH and either lacking the fluorescent label FAM (UN-untransformed/ no label) or featuring it (UL- untransformed/ with label) were compared to *Msm* transformed with pTH (TL- transformed/ with label). A further comparison of transformed samples was done to determine what/ if any effect boiling of lysate with FAM would do to the final banding pattern. This result seems to suggest that the fusion construct is being auto labeled at the expected (50kDA) position and features other bands which make up the interactome of the ClpP1-FLAG-APEX2 construct tested.

3.4 Discussion

At the point of starting this project the usage of proximity labeling using APEX2 within mycobacteria was not common practice, with one group publishing data during our initial planning of this aspect of the project[132]. Fortunately, this gave us the opportunity to compare methodology and results much more effectively. Based on our results we are confident we have developed a workflow for proximity labeling in live *Msm* using both biotin-phenol and Cu mediate click chemistry.

Initially, developed the Ting Lab[143] APEX2 is a modified peroxidase originally present in soybeans. Its use as a tool in proximity labeling/ and electron microscopy in both eukaryotic and prokaryotic cells has gained some momentum but remains a niche technique. As such its usage in mycobacteria had not been widely utilized or explored. The first group that appears to have made use of this system in *Msm* was Ganapathy et al[132] and had used it to great effect to study membrane bound peroxidase proteins.

The cytosolic nature of the Clp protease meant that the cellular environment around any fusion construct we made would be different than existing data. As such we felt it was necessary to access total cell lysate of native bacteria, those with a free unbound APEX2, as well as our fusion constructs. The resulting initial analysis (Fig 3.3) was rough but a positive starting point as we transitioned from *E. coli* to *Msm*.

The optimization of the labeling protocol ultimately revealed the native biotinylation of *Msm* would emerge through Strep-HRP staining. While we felt confident that we could differentiate individual bands not a result of native biotinylation we wanted to improve the technique to ensure future application would be as clear and simple as possible. Cu click chemistry is a new technique that bypasses the need to biotinylate proteins and instead utilizes the peroxidase function

of APEX2 to instead fluorescently label interaction partners instead using a fluorescent tag, like FAM[125,129,132]. The results of this (Fig 3.5), suggest the system is functional and capable of labeling proteins in the cell. However, upon further analysis it appears that the auto labeling expected of our constructs does not match the size of our fusion construct. Further analysis is necessary before any assessment can be made on the integrity of the fusion construct, however the banding pattern observed is still a positive result in terms of development of a proximity labeling protocol for the Clp protease.

Chapter 4

FUTURE DIRECTIONS

4.1 Evaluation of N-end rule Degradation in *Msm* with a KO of $_{Msm}$ClpS through CRISPRi

The major goal of this thesis work was to characterize the N-end rule of *Msm* in live cells and compare that to existing N-end degradation pathways. In Chapter 2 we established the existence of an N-end rule degradation pathway in *Msm* that features 4 primary degrons (Tyr, Phe, Trp, Leu) that correspond to the primary residues of the well characterized *E. coli* N-end rule. The lack of any apparent secondary degrons is also a unique characteristic of *Msm*. *E. coli* make use of an Aat L/F-aminotransferase to facilitate the degradation of secondary degrons like Lys and Arg. *Mtb* and *Msm* do not have a homolog of this aminotransferase within their genome. Therefore, if *Msm* makes use of secondary degrons it makes use of a yet unknown aminotransferase, but our methodology does not support the existence of secondary degrons.

One missing feature of this *in vivo* model is confirming the role of ClpS as an adaptor for ClpC1 to facilitate N-end rule degradation. While we have supporting crystal data, and anisotropy data, the question remains whether ClpS is carrying out this process in cells. Based on MST data shared by Henry Anderson in the Schmitz lab it appears that ClpS can interact with the ClpC1 unfoldase based on his MST analysis. While a positive result it still does not answer the question of that's

happening in the cell completely. The role ClpS appears to be to serve in the N-end rule pathway of mycobacteria however this has not yet been analyzed in a living system like that one used in this thesis work. However, ClpS is known to be non-essential in both *Mtb* and *Msm* therefore the generation of cells lacking this protein is certainly possible and is being actively pursued.

While multiple methods exist pre-existing work in the Schmitz lab has sought to make use of CRISPRi to delete the non-essential ClpS protein from Mc^2155 cells among other proteins. As we have data and an observable trend that denotes an N-end rule is present we can now evaluate deviation from this baseline. As we hypothesize that ClpS is facilitating the delivery of these substrates for degradation we would expect that with the elimination of ClpS from these cells the N-end rule degradation pattern would be significantly altered/ reduced. Similar exploration into N-end rule degradation and the role of ClpS done in *E. coli* has shown the effects of ClpS reduction on this unique Clp degradation system[112,113,120,144].

To improve the impact of the findings of chapter 2 *in vitro* biochemical assays using ClpC1, ClpP1, ClpP2, ClpS, and N-end substrates should be completed as soon as possible. Previous groups have successfully been able to generate data with these proteins, and recent progress in the lab has improved the activity of $_{Msm}$ClpC1 protein preps generated through the lab's internal protein expression methodology. While capable of being published independent of this additional data, the impact of this finding would be increased with the addition of this biochemical analysis.

4.2 Utilization of APEX2 fusion constructs to explore interaction partners of the *Msm* Clp protease.

In Chapter 3, our goal was to develop a methodology for the capture of interaction partners of the *Msm* Clp protease. While previous work, even internally in the Schmitz lab, has made use of trapping mechanisms to identify substrates, this process still yields a non-functional peptidase, and does not account for non-substrate proteins interacting with the Clp protease[81]. Using previous work with membrane bound proteins as a framework I worked to adapt these methods for our cytosolic protein complex[126,127,132,143]. Initial development of the plasmid constructs for the expression of APEX2 fusion proteins has been successful for both ClpP1 and ClpC1. Optimization of labeling protocol is ongoing with the biotin-phenol specific labeling being the most representative of data obtained by others using this labeling protein[132]. Further optimization of click-chemistry labeling and imaging is required for further use in this project. Of the two methods developed the biotin-phenol labeling protocol is the more referenced of the two and shows imaged results (Fig 3.3) representative of successful labeling.

While initial testing through imaging of labeled proteins suggests that biotin-phenol labeling is functioning as previously tested by others, the click-chemistry labeling still offers significant benefits if optimized. Initial attempts at labeling and imaging have shown some positive results but raise questions of fusion complex integrity. Improved labeling and imaging protocols could reveal labeling is sufficient, but most immediately western-blot analysis to access the integrity of the fusion construct should be done to resolve the issues in banding seen in Figure 3.5. Auto labeling of the fusion construct should appear as one of the most prominent bands when imaged, and while a band matching the size of the fusion construct (70kDa)

appears on figure 3.5 it is not the most prominent with lower banding being much more pronounced. Should the construct be intact then improvements to imaging can proceed and initial piloting of substrate capture and assessment through mass spec can also proceed. Improvements of labeling should include variation to the time labeling takes place, the time and intensity of quenching, as well as the storage of lysate post sonication to maintain protein integrity.

Further work to expand the profile of fusion constructs shown in table 1.2 can proceed once western blot analysis accesses the integrity of the existing ClpP1 and ClpC1 APEX2 fusions that have made up this initial testing. Further testing of these fusion constructs should be conducted as has been done for the constructs shown in this thesis work. However, the ability to capture and access proteins using mass spec analysis remains the major barrier to further progress in this project aim. To date assessment of labeling has been limited to lysate imaging through Strep-HRP (Fig 3.3), or chemiluminescence (Fig 3.5). Stored lysates bearing label have been stored and are available for initial assessment, however these samples still bear unlabeled proteins and require chromatography to separate labeled proteins. This chromatography will likely require optimization including buffer makeup, and digestion protocol following capture of labeled substrates.

Overall, this portion of this thesis work has set up the basis for a labeling protocol primarily using *Msm*ClpP-FLAg-APEX2 as a pilot construct. Labeling assessment using biotin-phenol shows the most promise based on its similarity in profile to referenced works. Labeled lysates should proceed to chromatography to separate labeled proteins and piloting the sample prep and mass spec analysis, but as the capture and label was done with biotin-phenol, screening samples with free

APEX2 would be required. This screening can also be compared to interaction data presented by other groups targeting the Clp protease. Click-chemistry offers improvements to labeling and capture but has shown variable imaging data that may reveal flaws in the construct viability using this method. However, western blot analysis could resolve this problem most immediately and improve the data of the biotin-phenol labeled lysate as well.

REFERENCES

If you want to number your bibliographic entries, change the style of the items to *Bib Entry - numbered*.

1. 1. Global Tuberculosis Report 2022. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022.
2. 2. Global tuberculosis report 2021. https://www.who.int/publications-detail-redirect/9789240037021.
3. 3. Al-Saeedi, M. & Al-Hajoj, S. Diversity and evolution of drug resistance mechanisms in Mycobacterium tuberculosis. *Infect Drug Resist* **10**, 333–342 (2017).
4. 4. Shetye, G. S., Franzblau, S. G. & Cho, S. New tuberculosis drug targets, their inhibitors, and potential therapeutic impact. *Translational Research* **220**, 68–97 (2020).
5. 5. Cudahy, P. & Shenoi, S. V. Diagnostics for pulmonary tuberculosis. *Postgraduate Medical Journal* **92**, 187–93 (2016).
6. 6. Gagneux, S. Host–pathogen coevolution in human tuberculosis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 850 (2012).
7. 7. Lönnroth, K. & Raviglione, M. Global epidemiology of tuberculosis: Prospects for control. *Seminars in Respiratory and Critical Care Medicine* **29**, 481–491 (2008).

8. 8. Smith, I. Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence Downloaded from http://cmr.asm.org/ on March 30 , 2014 by WITWATERSRAND Health Sciences LIBRARY. *Society* **16**, 463–496 (2003).
9. 9. Koch, R. *Robert Koch and Tuberculosis: Koch's Famous Lecture*. (Nobel Foundation, 2008).
10. 10. Gupta, R. Kr. & Srivastava, R. Resuscitation Promoting Factors: a Family of Microbial Proteins in Survival and Resuscitation of Dormant Mycobacteria. *Indian J Microbiol* **52**, 114–121 (2012).
11. 11. Bruchfeld, J., Correia-Neves, M. & Källenius, G. Tuberculosis and HIV Coinfection. *Cold Spring Harb Perspect Med* **5**, a017871 (2015).
12. 12. Ambardekar, N. Tuberculosis (TB) Treatment After Exposure: Medications Used. (2021).
13. 13. Alderwick, L. J., Harrison, J., Lloyd, G. S. & Birch, H. L. The Mycobacterial Cell Wall—Peptidoglycan and Arabinogalactan. *Cold Spring Harbor Perspectives in Medicine* **5**, 1–16 (2015).
14. 14. Zhao, J. *et al.* Mycolic acid-specific T cells protect against Mycobacterium tuberculosis infection in a humanized transgenic mouse model. *eLife* **4**, e08525 (2015).
15. 15. Campbell, E. A. *et al.* Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* **104**, 901–912 (2001).

16. 16. Wehrli, W. Rifampin: Mechanisms of Action and Resistance. *Reviews of Infectious Diseases* **5**, S407–S411 (1983).
17. 17. Unissa, A. N., Subbian, S., Hanna, L. E. & Selvakumar, N. Overview on mechanisms of isoniazid action and resistance in Mycobacterium tuberculosis. *Infection, Genetics and Evolution* **45**, 474–492 (2016).
18. 18. Zavascki, A. P., Goldani, L. Z., Li, J. & Nation, R. L. Polymyxin B for the treatment of multidrug-resistant pathogens: a critical review. *Journal of Antimicrobial Chemotherapy* **60**, 1206–1215 (2007).
19. 19. Dong, M., Pfeiffer, B. & Altmann, K.-H. Recent developments in natural product-based drug discovery for tuberculosis. *Drug Discov Today* **22**, 585–591 (2017).
20. 20. Sauer, R. T. & Baker, T. A. AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. *Annual Review of Biochemistry* **80**, 587–612 (2011).
21. 21. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci* **20**, 1298–1345 (2011).
22. 22. Lunge, A., Gupta, R., Choudhary, E. & Agarwal, N. The unfoldase ClpC1 of Mycobacterium tuberculosis regulates the expression of a distinct subset of proteins having intrinsically disordered termini. (2020) doi:10.1074/jbc.RA120.013456.
23. 23. Kim, Y. I. *et al.* Molecular determinants of complex formation between Clp/Hsp 100 ATPases and the ClpP peptidase. *Nature Structural Biology* **8**, 230–233 (2001).

24. 24. Gao, W. *et al.* The cyclic peptide ecumicin targeting CLpC1 is active against Mycobacterium tuberculosis in vivo. *Antimicrobial Agents and Chemotherapy* **59**, 880–889 (2015).
25. 25. Wolf, N. M. *et al.* Structure of the N-terminal domain of ClpC1 in complex with the antituberculosis natural product ecumicin reveals unique binding interactions. *Acta Crystallographica Section D: Structural Biology* **76**, 458–471 (2020).
26. 26. Kirstein, J. *et al.* The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. *EMBO Mol Med* **1**, 37–49 (2009).
27. 27. Beardslee, P. C. *et al.* Enyzmes | Clp Proteases. *Encyclopedia of Biological Chemistry: Third Edition* **3**, 292–306 (2021).
28. 28. Schmitz, K. R., Carney, D. W., Sello, J. K. & Sauer, R. T. Crystal structure of Mycobacterium tuberculosis ClpP1P2 suggests a model for peptidase activation by AAA+ partner binding and substrate delivery. *Proceedings of the National Academy of Sciences* **111**, E4587–E4595 (2014).
29. 29. Compton, C. L., Schmitz, K. R., Sauer, R. T. & Sello, J. K. Antibacterial Activity of and Resistance to Small Molecule Inhibitors of the ClpP Peptidase. *ACS Chem Biol* **509**, 2669–2677 (2013).
30. 30. Alhuwaider, A. A. H. & Dougan, D. A. AAA+ Machines of Protein Destruction in Mycobacteria. *Frontiers in Molecular Biosciences* **4**, 1–14 (2017).

31. 31. Kirthika, P., Lloren, K. K. S., Jawalagatti, V. & Lee, J. H. Structure, Substrate Specificity and Role of Lon Protease in Bacterial Pathogenesis and Survival. *International Journal of Molecular Sciences* **24**, 3422 (2023).
32. 32. Roudiak, S. G., Seth, A., Knipfer, N. & Shrader, T. E. The Lon Protease from Mycobacterium smegmatis: Molecular Cloning, Sequence Analysis, Functional Expression, and Enzymatic Characterization. *Biochemistry* **37**, 377–386 (1998).
33. 33. Simmons, L. A., Grossman, A. D. & Walker, G. C. Clp and lon proteases occupy distinct subcellular positions in Bacillus subtilis. *Journal of Bacteriology* **190**, 6758–6768 (2008).
34. 34. Langklotz, S., Baumann, U. & Narberhaus, F. Structure and function of the bacterial AAA protease FtsH. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 40–48 (2012).
35. 35. Narberhaus, F., Obrist, M., Führer, F. & Langklotz, S. Degradation of cytoplasmic substrates by FtsH, a membrane-anchored protease with many talents. *Research in Microbiology* **160**, 652–659 (2009).
36. 36. Schmitz, K. R. & Sauer, R. T. Substrate delivery by the AAA+ ClpX and ClpC1 unfoldases activates the mycobacterial ClpP1P2 peptidase. *Mol Microbiol* **93**, 617–628 (2014).
37. 37. Baker, T. A. & Sauer, R. T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochimica et biophysica acta* **1823**, 15–28 (2012).

38. 38. Amor, A. J., Schmitz, K. R., Baker, T. A. & Sauer, R. T. Roles of the ClpX IGF loops in ClpP association, dissociation, and protein degradation. *Protein Science : A Publication of the Protein Society* **28**, 756 (2019).
39. 39. Gottesman, S., Clark, W. P., de Crecy-Lagard, V. & Maurizi, M. R. ClpX, an alternative subunit for the ATP-dependent Clp protease of Escherichia coli. Sequence and in vivo activities. *J Biol Chem* **268**, 22618–22626 (1993).
40. 40. Zheng, H. *et al.* Genetic basis of virulence attenuation revealed by comparative genomic analysis of Mycobacterium tuberculosis strain H37Ra versus H37Rv. *PLoS One* **3**, e2375 (2008).
41. 41. Corydon, T. J. *et al.* Human and mouse mitochondrial orthologs of bacterial ClpX. *Mamm Genome* **11**, 899–905 (2000).
42. 42. Maurizi, M. R. *et al.* Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of Escherichia coli. *J Biol Chem* **265**, 12536–12545 (1990).
43. 43. Bross, P., Andresen, B. S., Knudsen, I., Kruse, T. A. & Gregersen, N. Human ClpP protease: cDNA sequence, tissue-specific expression and chromosomal assignment of the gene. *FEBS Lett* **377**, 249–252 (1995).
44. 44. Fleischmann, R. D. *et al.* Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. *J Bacteriol* **184**, 5479–5490 (2002).
45. 45. Chaudhuri, R. R. *et al.* Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative Escherichia coli strain 042. *PLoS One* **5**, e8801 (2010).

46. 46. Akopian, T. *et al.* The active ClpP protease from M. tuberculosis is a complex composed of a heptameric ClpP1 and a ClpP2 ring. *EMBO Journal* **31**, 1529–1541 (2012).

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

48. 48. Akopian, T. *et al.* Cleavage specificity of Mycobacterium tuberculosis ClpP1P2 protease and identification of novel peptide substrates and boronate inhibitors with anti-bacterial activity. *Journal of Biological Chemistry* **290**, 11008–11020 (2015).

49. 49. Benaroudj, N., Raynal, B., Miot, M. & Ortiz-Lombardia, M. Assembly and proteolytic processing of mycobacterial ClpP1 and ClpP2. *BMC Biochemistry* **12**, 61 (2011).

50. 50. Leodolter, J., Warweg, J. & Weber-Ban, E. The Mycobacterium tuberculosis ClpP1P2 Protease Interacts Asymmetrically with Its ATPase Partners ClpX and ClpC1. *PLoS ONE* **10**, e0125345 (2015).

51. 51. Amor, A. J., Schmitz, K. R., Sello, J. K., Baker, T. A. & Sauer, R. T. Highly Dynamic Interactions Maintain Kinetic Stability of the ClpXP Protease during the ATP-Fueled Mechanical Cycle. *ACS Chemical Biology* **11**, 1552–1560 (2016).

52. 52. d'Andrea, F. B. *et al.* The essential M. tuberculosis Clp protease is functionally asymmetric in vivo. *Sci Adv* **8**, eabn7943 (2022).

53. 53. Nagpal, J. *et al.* Molecular and structural insights into an asymmetric proteolytic complex (ClpP1P2) from Mycobacterium smegmatis. *Sci Rep* **9**, 18019 (2019).

54. 54. Ziemski, M., Leodolter, J., Taylor, G., Kerschenmeyer, A. & Weber-Ban, E. Genome-wide interaction screen for Mycobacterium tuberculosis ClpCP protease reveals toxin–antitoxin systems as a major substrate class. *The FEBS Journal* **288**, 99–114 (2021).

55. 55. Erbse, A. *et al.* ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439**, 753–756 (2006).

56. 56. Duran, E. C., Weaver, C. L. & Lucius, A. L. Comparative Analysis of the Structure and Function of AAA+ Motors ClpA, ClpB, and Hsp104: Common Threads and Disparate Functions. *Frontiers in Molecular Biosciences* **4**, (2017).

57. 57. Gottesmans, S. & Clark, P. The ATP-dependent Clp protease of Escherichia coli. Sequence of clpA and identification of a Clp-specific substrate. *Journal of Biological Chemistry* **265**, 7886–7893 (1990).

58. 58. Zolkiewski, M. A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. *Molecular microbiology* **61**, 1094 (2006).

59. 59. Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. The N-end rule in bacteria. *Science (New York, N.Y.)* **254**, 1374–1377 (1991).

60. 60. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Science* **20**, 1298–1345 (2011).

61. 61. Gao, X., Yeom, J. & Groisman, E. A. The expanded specificity and physiological role of a widespread N-degron recognin. (2019) doi:10.1073/pnas.1821060116.
62. 62. Dougan, D. A., Truscott, K. N. & Zeth, K. The bacterial N-end rule pathway: Expect the unexpected. *Molecular Microbiology* **76**, 545–558 (2010).
63. 63. Naor, N., Gadot, O., Meir, M. & Barkan, D. Peptide Deformylase (def) is essential in Mycobacterium smegmatis, but the essentiality is compensated by inactivation of methionine formylation. *BMC Microbiology* **19**, 232 (2019).
64. 64. Wingfield, P. N-Terminal Methionine Processing. *Curr Protoc Protein Sci* **88**, 6.14.1-6.14.3 (2017).
65. 65. Cha, H. J., Jeong, J.-H., Rojviriya, C. & Kim, Y.-G. Structure of Putrescine Aminotransferase from Escherichia coli Provides Insights into the Substrate Specificity among Class III Aminotransferases. *PLoS One* **9**, e113212 (2014).
66. 66. Dong, X. *et al.* The crystal structure of leucyl/phenylalanyl-tRNA-protein transferase from Escherichia coli. *Protein Sci* **16**, 528–534 (2007).
67. 67. Laws, M., Jin, P. & Rahman, K. M. Efflux pumps in Mycobacterium tuberculosis and their inhibition to tackle antimicrobial resistance. *Trends in Microbiology* **30**, 57–68 (2022).
68. 68. Wong, K. S. *et al.* Acyldepsipeptide Analogs Dysregulate Human Mitochondrial ClpP Protease Activity and Cause Apoptotic Cell Death. *Cell Chemical Biology* **25**, 1017-1030.e9 (2018).

69. 69. Vasudevan, D., Rao, S. P. S. & Noble, C. G. Structural Basis of Mycobacterial Inhibition by Cyclomarin A. *J Biol Chem* **288**, 30883–30891 (2013).

70. 70. Choules, M. P. *et al.* Rufomycin Targets ClpC1 Proteolysis in Mycobacterium tuberculosis and M. abscessus. *Antimicrobial Agents and Chemotherapy* **63**, (2019).

71. 71. Weinhäupl, K. *et al.* The antibiotic cyclomarin blocks arginine-phosphate-induced millisecond dynamics in the N-terminal domain of ClpC1 from Mycobacterium tuberculosis. *The Journal of biological chemistry* **293**, 8379–8393 (2018).

72. 72. Schmitz, K. R. *et al.* Acyldepsipeptide Antibiotics and a Bioactive Fragment Thereof Differentially Perturb Mycobacterium tuberculosis ClpXP1P2 Activity in vitro. *ACS Chem Biol* 10.1021/acschembio.9b00454 (2020) doi:10.1021/acschembio.9b00454.

73. 73. Gao, W. *et al.* Discovery and Characterization of the Tuberculosis Drug Lead Ecumicin. *Org. Lett.* **16**, 6044–6047 (2014).

74. 74. Schmitt, E. K. *et al.* The natural product cyclomarin kills mycobacterium tuberculosis by targeting the ClpC1 subunit of the caseinolytic protease. *Angewandte Chemie - International Edition* **50**, 5889–5891 (2011).

75. 75. Gavrish, E. *et al.* Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2. *Chemistry & biology* **21**, 509–518 (2014).

76. 76. Tripathi, P., Singh, L. K., Kumari, S., Hakiem, O. R. & Batra, J. K. ClpB is an essential stress regulator of Mycobacterium tuberculosis and endows survival advantage to dormant bacilli. *International journal of medical microbiology : IJMM* **310**, (2020).
77. 77. Zhao, B., Li, X., Zeng, Y. & Lu, Y. ClpP-deletion impairs the virulence of Legionella pneumophila and the optimal translocation of effector proteins. *BMC Microbiol* **16**, 174 (2016).
78. 78. Marsee, J. D., Ridings, A., Yu, T. & Miller, J. M. Mycobacterium tuberculosis ClpC1 N-terminal domain is dispensable for adaptor protein-dependent allosteric regulation. *International Journal of Molecular Sciences* **19**, (2018).
79. 79. Gopal, P. *et al.* Pyrazinamide triggers degradation of its target aspartate decarboxylase. *Nat Commun* **11**, 1661 (2020).
80. 80. Ogbonna, E. C. Identification and characterization of novel Clp protease interaction partners in Mycolicibacterium smegmatis. (University of Delaware, 2022).
81. 81. Ogbonna, E. C., Anderson, H. R. & Schmitz, K. R. Identification of Arginine Phosphorylation in Mycolicibacterium smegmatis. *Microbiol Spectr* **10**, e0204222 (2022).
82. 82. Erbse, A. *et al.* ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439**, 753–756 (2006).

83. 83. Li, Y., Corro, J. H., Palmer, C. D. & Ojha, A. K. Progression from remodeling to hibernation of ribosomes in zinc-starved mycobacteria. *Proceedings of the National Academy of Sciences* **117**, 19528–19537 (2020).

84. 84. Lupoli, T. J., Vaubourgeix, J., Burns-Huang, K. & Gold, B. Targeting the Proteostasis Network for Mycobacterial Drug Discovery. *ACS Infectious Diseases* **4**, 478–498 (2018).

85. 85. Ollinger, J., O'malley, T., Kesicki, E. A., Odingo, J. & Parish, T. Validation of the essential ClpP protease in Mycobacterium tuberculosis as a novel drug target. *Journal of Bacteriology* **194**, 663–668 (2012).

86. 86. Y, et al Y. H., Yamaguchi M, Igarashi. Mycolicibacterium smegmatis, Basonym Mycobacterium smegmatis, Expresses Morphological Phenotypes Much More Similar to Escherichia coli Than Mycobacterium tuberculosis in Quantitative Structome Analysis and CryoTEM Examination. *Frontiers in Microbiology* **9**, (2018).

87. 87. Raju, R. M. *et al.* Mycobacterium tuberculosis ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLoS Pathogens* **8**, (2012).

88. 88. De Donatis, G. M., Singh, S. K., Viswanathan, S. & Maurizi, M. R. A single ClpS monomer is sufficient to direct the activity of the ClpA hexamer. *Journal of Biological Chemistry* **285**, 8771–8781 (2010).

89. 89. Guo, F., Esser, L., Singh, S. K., Maurizi, M. R. & Xia, D. Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. *Journal of Biological Chemistry* **277**, 46753–46762 (2002).
90. 90. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343–345 (2009).
91. 91. Takahashi, Y., Mizoi, J., Toh-e, A. & Kikuchi, Y. Yeast Ulpl, an Smt3-Specific Protease, Associates with Nucleoporins1. *The Journal of Biochemistry* **128**, 723–725 (2000).
92. 92. Zhang, F. *et al.* Profiling Substrate Specificity of the SUMO Protease Ulp1 by the YESS–PSSC System to Advance the Conserved Mechanism for Substrate Cleavage. *Int J Mol Sci* **23**, 12188 (2022).
93. 93. Román-Hernández, G., Hou, J. Y., Grant, R. A., Sauer, R. T. & Baker, T. A. The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. *Molecular cell* **43**, 217–228 (2011).
94. 94. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. & Helinski, D. R. Plasmid ColE1 as a Molecular Vehicle for Cloning and Amplification of DNA. *Proceedings of the National Academy of Sciences* **71**, 3455–59 (1974).
95. 95. Rauzier, J., Moniz-Pereira, J. & Gicquel-Sanzey, B. Complete nucleotide sequence of pAL5OOO, a plasmid from Mycobacterium fortuitum. *Gene* **71**, 315–321 (1988).

96. 96. Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of Structural and Functional Genomics 2004 5:1* **5**, 75–86 (2004).

97. 97. Zhu, Y. *et al.* Characterization of a minimal type of promoter containing the -10 element and a guanine at the -14 or -13 position in mycobacteria. *Journal of Bacteriology* **199**, (2017).

98. 98. Giasson, B. I., Murray, I. V. J., Trojanowski, J. Q. & Lee, V. M. Y. A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly. *The Journal of biological chemistry* **276**, 2380–2386 (2001).

99. 99. Giasson, B. I., Murray, I. V. J., Trojanowski, J. Q. & Lee, V. M. Y. A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly. *The Journal of biological chemistry* **276**, 2380–2386 (2001).

100. 100. Oxford Nanopore Technologies. *Oxford Nanopore Technologies* https://nanoporetech.com.

101. 101. GENEWIZ from Azenta Life Sciences. https://www.genewiz.com/en.

102. 102. Zeth, K. *et al.* Structural analysis of the adaptor protein ClpS in complex with the N- terminal domain of ClpA. *Nature Structural & Molecular Biology* **9**, 906–911 (2002).

103. 103. Tasaki, T., Sriram, S. M., Park, K. S. & Kwon, Y. T. The N-End Rule Pathway. *Annu Rev Biochem* **81**, 261–289 (2012).

104. 104. Wagner, D. *et al.* Elemental Analysis of *Mycobacterium avium* -, *Mycobacterium tuberculosis* -, and *Mycobacterium smegmatis* -Containing Phagosomes Indicates Pathogen-Induced Microenvironments within the Host Cell's Endosomal System. *The Journal of Immunology* **174**, 1491–1500 (2005).

105. 105. Trimble, W. S. & Grinstein, S. TB or not TB: Calcium Regulation in Mycobacterial Survival. *Cell* **130**, 12–14 (2007).

106. 106. Edmonds, K. A., Jordan, M. R. & Giedroc, D. P. COG0523 proteins: a functionally diverse family of transition metal-regulated G3E P-loop GTP hydrolases from bacteria to man. *Metallomics* **13**, mfab046 (2021).

107. 107. Ziemski, M., Leodolter, J., Taylor, G., Kerschenmeyer, A. & Weber-Ban, E. Genome-wide interaction screen for Mycobacterium tuberculosis ClpCP protease reveals toxin–antitoxin systems as a major substrate class. *The FEBS Journal* **288**, 99–114 (2021).

108. 108. Velasco-García, R. & Vargas-Martínez, R. The study of protein–protein interactions in bacteria. *Canadian Journal of Microbiology* **58**, 1241–1257 (2012).

109. 109. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure* **1077**, 101–113 (2014).

110. 110. Ehrt, S. *et al.* Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic acids research* **33**, (2005).

111. 111. Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of Structural and Functional Genomics 2004 5:1* **5**, 75–86 (2004).

112. 112. Erbse, A. *et al.* ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439**, 753–756 (2006).

113. 113. Román-Hernández, G., Hou, J. Y., Grant, R. A., Sauer, R. T. & Baker, T. A. The ClpS adaptor mediates staged delivery of N-end-rule substrates to the AAA+ ClpAP protease. *Mol Cell* **43**, 217–228 (2011).

114. 114. Jiang, J. & Schmitz, K. R. Bioinformatic identification of ClpI, a distinct class of Clp unfoldases in Actinomycetota. *Frontiers in Microbiology* **14**, (2023).

115. 115. Zhang, T. *et al.* Flexible connection of the N-terminal domain in ClpB modulates substrate binding and the aggregate reactivation efficiency. *Proteins: Structure, Function and Bioinformatics* **80**, 2758–2768 (2012).

116. 116. Ahyoung, A. P., Koehl, A., Vizcarra, C. L., Cascio, D. & Egea, P. F. Structure of a putative ClpS N-end rule adaptor protein from the malaria pathogen Plasmodium falciparum. *Protein Science* **25**, 689–701 (2016).

117. 117. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research* **38**, W529–W533 (2010).

118. 118. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research* **44**, W344 (2016).

119. 119. Xia, D., Esser, L., Singh, S. K., Guo, F. & Maurizi, M. R. Crystallographic investigation of peptide binding sites in the N-domain of the ClpA chaperone. *Journal of Structural Biology* **146**, 166–179 (2004).

120. 120. Hou, J. Y., Sauer, R. T. & Baker, T. A. Distinct structural elements of the adaptor ClpS are required for regulating degradation by ClpAP. *Nature Structural & Molecular Biology 2008 15:3* **15**, 288–294 (2008).

121. 121. Snider, J., Thibault, G. & Houry, W. A. The AAA+ superfamily of functionally diverse proteins. *Genome Biology* **9**, 1–8 (2008).

122. 122. Bhandari, V. *et al.* The Role of ClpP Protease in Bacterial Pathogenesis and Human Diseases. *ACS Chemical Biology* **13**, 1413–1425 (2018).

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

124. 124. Ellisman, M. H. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods* **12**, 51–54 (2014).

125. 125. Dumrongprechachan, V. *et al.* Cell-type and subcellular compartment-specific APEX2 proximity labeling reveals activity-dependent nuclear proteome dynamics in the striatum. *Nat Commun* **12**, 4855 (2021).

126. 126. Hung, V. *et al.* Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat Protoc* **11**, 456–475 (2016).

127. 127. Nguyen, T. M. T., Kim, J., Doan, T. T., Lee, M.-W. & Lee, M. APEX Proximity Labeling as a Versatile Tool for Biological Research. *Biochemistry* **59**, 260–269 (2020).

128. 128. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **40**, 2004–2021 (2001).

129. 129. Gao, X. & Hannoush, R. N. A Decade of Click Chemistry in Protein Palmitoylation: Impact on Discovery and New Biology. *Cell Chemical Biology* **25**, 236–246 (2018).

130. 130. 5’ 6-FAM (Fluorescein) modification | IDT. https://www.idtdna.com/site/catalog/modifications/product/1108.

131. 131. Ahamed, M. *et al.* Chapter Twelve - Optimized APEX2 peroxidase-mediated proximity labeling in fast- and slow-growing mycobacteria. in *Methods in Enzymology* (ed. Carlson, E. E.) vol. 664 267–289 (Academic Press, 2022).

132. 132. Ganapathy, U. S. *et al.* Compartment-Specific Labeling of Bacterial Periplasmic Proteins by Peroxidase-Mediated Biotinylation. *ACS Infect Dis* **4**, 918–925 (2018).

133. 133. Zirconia/Silica Beads - BioSpec Products. https://www.biospec.com/products/zirconia-silica-beads.

134. 134. DYKDDDDK Tag Recombinant Monoclonal Antibody (8H8L17) (701629). https://www.thermofisher.com/antibody/product/DYKDDDDK-Tag-Antibody-clone-8H8L17-Recombinant-Monoclonal/701629.

135. 135. Goat anti-Rabbit IgG Antibodies | Invitrogen. https://www.thermofisher.com/antibody/secondary/target/goat%20anti-rabbit%20igg.

136. 136. Petrunina, N. A. *et al.* DNA Intercalated Motifs with Non-Nucleoside Inserts. *Russ J Bioorg Chem* **47**, 1341–1344 (2021).

137. 137. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**, 880–887 (2018).

138. 138. NEB® 10-beta Electrocompetent E. coli | NEB. https://www.neb.com/products/c3020-neb-10-beta-electrocompetent-e-coli#Product%20Information_Reagents%20Supplied.

139. 139. Guo, X. V. *et al.* Silencing Essential Protein Secretion in Mycobacterium smegmatis by Using Tetracycline Repressors. *Journal of Bacteriology* **189**, 4614–4623 (2007).

140. 140. Dickinson, M. S. *et al.* Proximity-dependent proteomics of the Chlamydia trachomatis inclusion membrane reveals functional interactions with endoplasmic reticulum exit sites. *PLOS Pathogens* **15**, e1007698 (2019).

141. 141. BioSpec Products 0.1 MM Zirconia/Silica Beads. https://www.fishersci.com/shop/products/0-1-mm-zirconia-silica-beads/NC0362415.

142. 142. Hu, Z. & Cronan, J. E. The primary step of biotin synthesis in mycobacteria. *Proceedings of the National Academy of Sciences* **117**, 23794–23801 (2020).

143. 143. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**, 51–54 (2015).

144. 144. Zeth, K. *et al.* Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nature Structural Biology* **9**, 906–911 (2002).