Furthering the understanding of *Mtb*ClpS regulation of *Mtb*ClpC1 as it relates to structure and function

by

Justin D. Marsee

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of
Master of Science in Chemistry

Middle Tennessee State University

July 2020

Thesis Committee:

Dr. Justin M. Miller, Chair

Dr. Donald A. Burden

Dr. Kevin Bicker

While there are a great many people who have contributed to my life and career, there

are two people who certainly deserve the most credit and recognition: my parents.

While it is true that you cannot choose your family, I would certainly choose them if

presented with such a choice. I dedicate this body of work to them, as a symbol of

gratitude for their continual support and encouragement. All that I am, I owe to them.

Thank you.

ACKNOWLEDGEMENTS

First and foremost, thank you to Dr. Miller for asking me to research with him almost 4 years ago. I owe my career path and discovered love of biophysics to him. In addition to him, I would like to thank Amy Ridings and Dr. Yu at Tennessee Tech for their *in silico* contributions to the work. Along with those mentioned, I would not be where I am without all past and present members of the Miller Lab for making research such a pleasant experience.

ABSTRACT

Mycobacterium tuberculosis (*Mtb*) has developed significant drug-resistance in recent years, thereby necessitating novel drug targets for pathogenic bacteria. The AAA+ (ATPase Associated with various cellular Activities) chaperone of *Mtb*ClpC1 could serve as a target for drug development due to its critical role in the maintenance of protein homeostasis. *Mtb*ClpC1 hexamers couple the energy of ATP hydrolysis to unfold and translocate protein substrates into an associated *Mtb*ClpP protease. The substrate specificity of these unfoldases is regulated by the ATPase component and potential adaptor protein association. No adaptor protein-mediated control has been reported for *Mtb*ClpC1. Using in vitro and in silico methods, we report data demonstrating that *Mtb*ClpC1 catalyzed unfolding of SsrA-tagged protein is negatively impacted by *Mtb*ClpS association. This does not require the expected ClpC1 N-terminal domain (NTD), but instead requires the presence of an interaction located in the Middle Domain (MD).

TABLE OF CONTENTS

LIST OF TABLES & FIGURES

LIST OF ABBREVIATIONS

AAA+ – ATPases associated with various cellular activities

Hsp – Heat shock proteins

ATP – Adenosine Triphosphate

NTP – Nucleotide Triphosphate

Clp – Casienolytic Protein

NTD – N-terminal domain

MD – Middle Domain/Molecular Dynamics

CD – Computation Docking

Mtb – *Mycobacterium tuberculosis*

Ec – *Escherichia coli*

Bs – *Bacillus subtilis*

ADEPs – acyldepsipeptides

COM – Center of Mass

SMD – Steered Molecular Dynamics

SPR – Surface Plasmon Resonance

SRH – Second region of homology

ZBD – zinc binding domain

CHAPTER I: INTRODUCTION

1.   *Proteostasis*

Regulated proteolysis is the fundamental process used by all organisms to adapt to changing conditions involved in stress responses, growth, cell division, cell differentiation, pathogenesis, biofilm formation, and protein secretion [1]. Such a biological balancing act involves the maintaining of a network of enzymes in thousands of cellular reactions occurring at any given time in interconnected and competing regulatory pathways [2]. Proteostasis is the steady-state, or non-equilibrium, process of balancing protein synthesis and degradation. Maintenance of proteostasis ensures healthy aging, resistance to those external stressors (including temperature stress, halophilic stress, osmotic stress, toxins, oxidative stress, etc.) and can even help mitigate the perturbations to the broader cellular homeostasis caused by pathogenic invasions [3-4]. The cellular details of proteostasis are dynamic, where temporal and permanent shifts in the equilibrium of protein synthesis and degradation occur during different stages of development and throughout aging, in part to act as a buffer to stress conditions. Such responses can vary depending upon the nature of the stressor, in addition to the specific tissue and cell type involved [5, 6]. One specific response is the control of clpC gene expression by the transcriptional repressor ClgR in *Bifidobacterium breve* [7]. The expression of ClpC in gram-positive bacteria, rich in G-C content, is dually regulated by transcriptional activator ClgR, sigma factor, and protein-denaturing stressors (eg. heat shock and osmotic stress) []. In other words, we observe intracellular responses (regulation of gene expression) as a downstream effect of an

external stimulus (heat shock/osmotic stress). Understanding this process and the machines involved is an important field of research.

A central feature of proteostatic regulation is the need to control the availability and/or function of cellular proteins that may be misfolded or critical regulatory or metabolic proteins [8]. Protein folding can be described by a free energy landscape describing an infinite number of conformational states that can be adopted by the macromolecule. Figure 1 is a folding funnel, a schematic used to represent the energy landscape of the various conformational states achieved during protein folding. Native states of proteins are the most thermodynamically favorable conformational state, this is represented by the absolute minima of the funnel. Local minima, shown by slight dips on
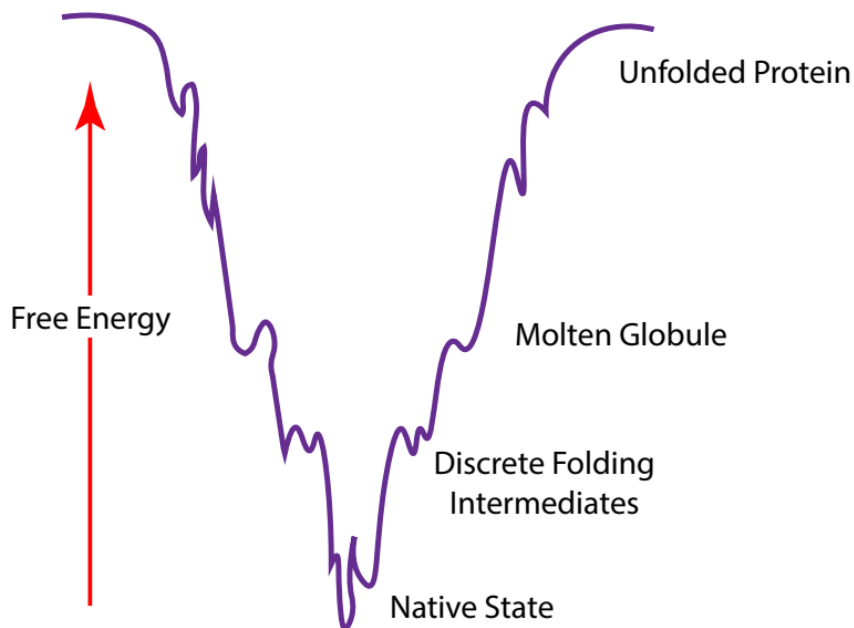
Figure 1 – **Representative "folding funnel"** as introduced by Ken A. Dill in 1987 in discussion of molten globules [116]. As indicated on the figure, the y-axis indicates the internal free energy, and every point along the x-axis represents one of an infinite possible conformational states of protein folding. In turn, this means that each time a single protein is folded, there exists an infinite number of ways it can be folded into its native structure, each with extraordinarily minute paths.

the sides of the funnel, are pockets where the free energy is low enough to stabilize a partially folded protein. More specifically, these "pockets" are representations of energy barriers preventing easy departure from the local minima of a given conformational state. Decreasing energy down the y-axis allows for depiction of the complex's thermodynamic stability in that state. This is to say that increased area within the interior of the funnel means a greater number of conformational possibilities for the folding protein. Ultimately, this folding funnel theory allows for a protein system specific number of folding pathways from primary structure to the native conformation of a protein. Such possibility provides an incredible resilience to mutations to mRNA sequence as a mutation that might interrupt one path to the native state, would be rendered ineffectual by another pathway (provided it is still thermodynamically possible to reach the native conformation).

2. Regulation of Proteostasis

2.1. Protein Folding

Proteins must achieve a highly specific three-dimensional conformation, or shape, to be active. The process of reaching this 'native' state is referred to as protein folding. Proteins start as a linear sequence of amino acids before randomly folding to reach a conformation with the lowest free energy. Folding begins with mRNA translation, where the ribosome decodes mRNA sequences of codons into a polypeptide chain. The ribosome being a multi-subunit complex composed of a species-specific amount of proteins and rRNA. Codons are sequences of 3 nucleotide bases that encode for specific amino acids, where one amino acid can be represented by multiple sequences. Recognition of a codon triggers tRNA transfer of the requisite amino acid to a nascent and growing polypeptide

chain bound by the ribosome [9]. Ribosomal stalling can occur because of mRNA containing rare codons, or infrequently translated codons. If the mRNA strand contains a consecutive sequence of rare codons, translation is aborted, and the nascent polypeptide chain is tagged for degradation [10]. In addition to this error, with the intracellular environment being crowded, proteins are often prevented from properly folding [9]. To address these, there exist a class of proteins called chaperones. Most chaperones are ATPases, meaning they couple ATP hydrolysis to generate the force necessary to unfold these improperly folded proteins. To avoid the erroneous unfolding of correctly folded proteins, chaperones have individual specificity. Foldases function to accelerate the rate limiting steps of the folding pathway, an example being the GroEL/ES chaperone and cochaperone complex. GroEL is a chaperone that works in tandem with the cochaperone GroES to mediate protein folding [11,12]. In contrast, as an unfoldase, GroEL achieves the same goal of correctly folded proteins by functioning in the opposite manner, reverting a folded protein to its simplest, primary structure. Once unfolded, a protein will refold through a series of trial and error random events in a second attempt to reach the native state. Such an event would be visualized as restarting from the top of the funnel after each unfolding event.

2.2. Machinery of Protein Folding

Unfoldases, however, often work in tandem with proteases to degrade those protein substrates into single amino acids or small peptide fragments. It is important that these two protein types work in concert as this conserves cellular energy. The chaperones relevant to this body of work are hexameric, or six-membered complexes, ATPase grabs a substrate

and threads it into the associated protease for degradation. In the compartmentalization of the proteolytic and unfoldase components, the protease activity is limited by the ATPase's substrate recognition. Such structural arrangement protects against non-specific degradation by the enzyme. As a proteolytic complex, the unfoldase component works as a motor feeding substrate into the protease component. Rather than requiring separate ATP hydrolysis events for both unfolding/translocation and degradation of the substrate, as a complex the coupling of hydrolysis to mechanical unfolding drives proteolytic function. These complexes can be thought of as paper shredders, wherein material is fed into an interior compartment and shredded [13-15,16-18]. These proteolytic complexes are included as a part of the broader AAA+ (ATPases Associated with Various Cellular Activities) superfamily which is characterized by a conserved ring-shaped ATPase domain that associates with a barrel-shaped protease [14,19].

2.2.1. AAA+ Architecture

Most ATPase domains in the AAA+ superfamily assemble as oligomeric ring-shaped complexes (frequently, but not exclusively, as homo-hexamers) with a large central pore for substrate processing [14,19]. ATPase sites are located at subunit interfaces and undergo conformational changes that facilitate the mechanical, motor-like activity of all AAA+ complexes. The Walker A and B motifs are critical for binding and catalytic activity of the ATPase. Walker A is identified by the following primary sequence, GXXXXGK[T/S], where X is any amino acid and the C-terminal residue is either threonine or serine. Walker A is also referred to as the P-loop for its Lysine interaction with the phosphate groups of ATP to coordinate ATP binding [3,5,6]. Walker B consists of

hhhhD[D/E], where h can be any hydrophobic residue and the C-terminal residue can be either Aspartate or Glutamate. The Walker B motif acts as a catalytic base in activation of a water molecule for nucleophilic attack on the γ-phosphate of ATP. Mutation of residues within the Walker B motif prevent ATP hydrolysis, even though ATP binding still occurs [19]. The second region of homology (SRH) is present to the C-terminal side of the Walker B motif. Containing approximately 20 residues, the SRH houses the sensor 1 and Arginine finger motifs which are both necessary for ATP hydrolysis [20]. In structure, sensor 1 is located between the Walker A and Walker B motifs, functioning in concert with Walker B to hydrolyze ATP by orienting the activated water molecule [19-20]. When in oligomeric complex, the Arginine finger protrudes into the ATP binding/hydrolysis site of the adjacent subunit and forms a bond with the γ-phosphate of ATP that is thought to stabilize the increased negative charge during ATP hydrolysis [14]. Mutation of this motif has provided evidence that it is not required for ATP binding, but solely for ATP hydrolysis [14,19-20].

3. *Physiological Importance*

Since the identification of Penicillin in 1941, pathogens have been developing resistance to each drug used in a clinical and research setting. A CDC report from 2019 estimated that at least 2.8 million antibiotic-resistant infections occur every year and leave over 35,000 dead [21]. With these figures steadily rising each year, the necessity for novel treatment methods has become more apparent. As proteostatic regulation is crucial to cellular health, the machinery involved make attractive drug targets. A couple of examples of such antibiotics are ADEPs (acyldepsipeptides), a class of antibiotics that occur as natural byproducts of aerobic fermentation in *Streptomyces hawaiiensis* that promote toxic

dysregulation of an ATP-dependent protease, and lassomycin, a ribosomally synthesized cyclic peptide deadly to *Mycobacterium tuberculosis* [22-25]. The complexes in question are each ATPases associated with a protease, ClpP, that has homologous proteins across prokaryotes and eukaryotes alike. While many of the associating ATPases/unfoldases are hexameric (when oligomerized, they are categorized into two groups: ClpA/B like and ClpX like. The ClpA/B ATPases assemble into a dual ATPase domain structure, wherein two ATPase rings are stacked onto each other [22]. ClpX ATPases do not assemble in this way, rather they have been shown as only assembling into a single ATPase domain [26]. Regardless of the differences in domain number, both motor families associate with the ClpP peptidase. Herein lies the physiological importance, ClpA/B unfoldases are not found in eukaryotic cells, but ClpX unfoldases are present in the mitochondria [27]. Because of this, targeting of ClpP activity by drug treatment could prove detrimental to proteostatic regulation of the mitochondria.

In addition to novel drug targets, phage therapy is another potential solution to the antibiotic resistance problem. Phage therapy is the process of introducing pathogen specific bacteriophages, as these viruses are more effective than current antibiotics are at penetrating the biofilm when covered by a polysaccharide layer [28]. The specificity of these phages for individual bacteria has two main benefits in that it would not attack human eukaryotic cells, nor should it attack the normal microbiota within humans [28]. The introduced phages work to hijack the machinery of the cell to produce more phages that cause the cell to lyse, releasing more phages that seek out more pathogenic bacteria in the host. The commonality with these multiple drug treatments is the toxic interference of

proteostatic regulation, whether through focus on specific machinery (ADEPs and lassomycin) or the hijacking of the cell by viral DNA. Thus, research serving to fill knowledge gaps in any proteostatic pathways is of great importance for multi-faceted applications. Provided within this body of work is a study of an important enzyme implicated within proteostatic regulation of *Mtb,* in hopes to lay the foundation for such goals as anti-tuberculosis treatments. Herein we report the novel and unique allosteric regulation of ClpC1 by ClpS that might aid in anti-tubercular drug targeting of ClpC1.

CHAPTER II: HSP100/CLP PROTEIN PHYLOGENETIC ANALYSIS

1. Introduction

1.1. Heat shock proteins

Since many ATPases serve to help correct the effects of extracellular stressors, it follows that they must be able to survive and function in extreme conditions. Such chaperones are heat shock proteins (Hsp), or proteins that are induced under stress conditions. Although named for their response to heat shock, the production of Hsp is also induced under infection, inflammation, exposure to toxins, nitrogen deprivation, oxygen deprivation, starvation, dehydration, etc. To identify subclasses of Hsp, they are grouped and named based on their approximate molecular weight in kilodaltons (KDa). Hsp100, for example, are such ATPases weighing approximately 100 KDa as monomeric subunits. These Hsp100 are further categorized as Clp (caseinolytic proteins) based on the endopeptidase activity of ClpP, the associating protease, that cleaves casein (among other proteins). The Hsp100 (or Clp) family of molecular motors can be distinctly separated based on the number of ATPase domains present in the complex. Class 1 proteins contain two canonical ATP binding/hydrolysis sites per monomer, while Class 2 members contain only one site. Class 1 family members include ClpA, ClpB, ClpC, ClpE, ClpK, and ClpL, whereas Class 2 includes ClpX and HslU [29-30].

1.2. Hsp100/Clp family

In the active state, Clp/Hsp100 ATPases form ring-shaped hexamers assembled from six copies of the same protein [31-32]. This topological arrangement provides the structural basis for enzyme catalyzed protein unfolding. Protein substrates bind at a central channel in the hexamer that is too small for folded proteins to traverse. Using NTP as fuel, the

hexamer applies a tugging force to the folded protein substrate that induces denaturation, thereby yielding a linear molecule that can be threaded into ClpP for degradation. Motor proteins from each group are structurally similar and contain the sequence motifs necessary to couple the energy of nucleotide binding and hydrolysis to the performance of mechanical work required for protein remodeling or translocation [14, 30].

2. Results and Discussion

Class 1 Clp/Hsp100 ATPases share a common domain architecture that includes an N-terminal domain (NTD) and two AAA+ ATPase domains termed Domain 1 (D1) and Domain 2 (D2) that each contain the canonical Walker A and B motifs (Figure 2) [20, 27, 30, 32-36, 40]. Deletion analysis using Mycobacterium tuberculosis (*M. tuberculosis*) ClpC1 mutants has demonstrated that the NTD is dispensable for hexamerization, ATP hydrolysis, and ATP-dependent protein remodeling*,* though it may be necessary for degradation of specific protein substrates [41]. A similar study has shown that Escherichia coli (*E. coli*) ClpA and ClpX both retain the ability to form hexamers upon deletion of the NTD [42]. Structural studies of ClpA and ClpC reveal that each contains AAA+ ATPase domains with a highly conserved α-β-α fold flanked by a helical lid domain that likely mediates subunit interactions in assembled hexamers (Figure 2 A and B) [33, 34, 27, 39] ClpC also contains a distinctive coiled-coil Middle Domain (MD) inserted in the D1 lid domain (Figure 2A) that distinguishes it from other AAA+ hexameric translocases such as ClpA, ClpX, MCM, E1, Vps4, etc [33, 34, 42 -46].

4.1 Using structural features of the AAA+ cassette to differentiate ClpA and ClpC protein sequences

Primary sequence analysis indicates that Class 1 Clp/Hsp100 proteins can be grouped based upon three features that include an intact Middle Domain (MD), a zinc-binding domain (ZBD), and the combination of a MD/ZBD. A ZBD is not commonly found in either ClpA, ClpC, or ClpL proteins, but is instead found in the ClpE/ClpK protein families. Proteins that contain the MD feature may or may not possess a ZBD with examples that include ClpE/ClpK or ClpL, respectively. However, no atomic resolution structural data currently exist for any ClpE/ClpK/ClpL protein. As such, we utilized available X-ray structures for ClpA and ClpC alongside multiple sequence alignment analyses to identify unique features that may aid in differentiation between these protein groups.

To elucidate differentiating structural features of ClpA versus ClpC, we first examined available X-ray crystal structures of each protein [32-35]. The two ATP binding and hydrolysis sites found in each ClpA or ClpC monomer belong to the AAA+ superfamily [28, 21, 40] (Figure 2A and B). Each domain conforms to the AAA+ domain architecture with a central βsheet that is flanked on both sides by α-helices to form a three-tiered αβ-α sandwich (Fig. 2A and B). This domain architecture positions conserved active site elements for ATP hydrolysis and includes the canonical Walker A (GXXXXGK[T/S], where X is any residue) and Walker B (hhhhD[D/E], where h is any hydrophobic residue) motifs [21, 40]. The two ATP binding and hydrolysis sites, D1 and D2, differ from one another due to α-β-α core insertions that subdivide the AAA+ family into smaller clades [28, 40]. Specifically, D1 contains a short α-helical insertion between α2 and β2 (Clade 3, Fig. 1A), whereas D2 includes a β-turn insertion between α3 and β4 (Clade 5, Fig. 1B).
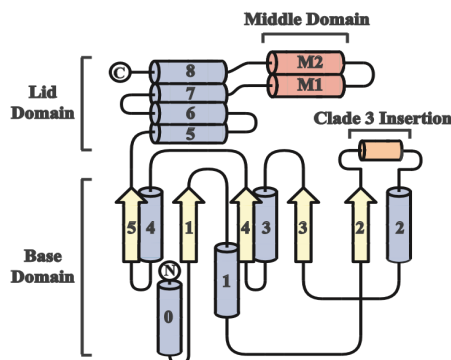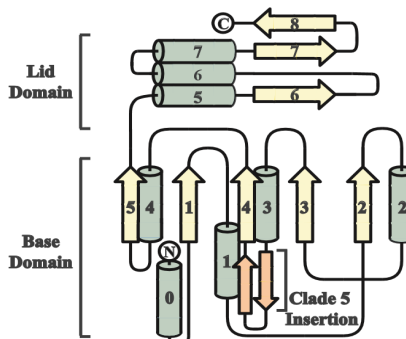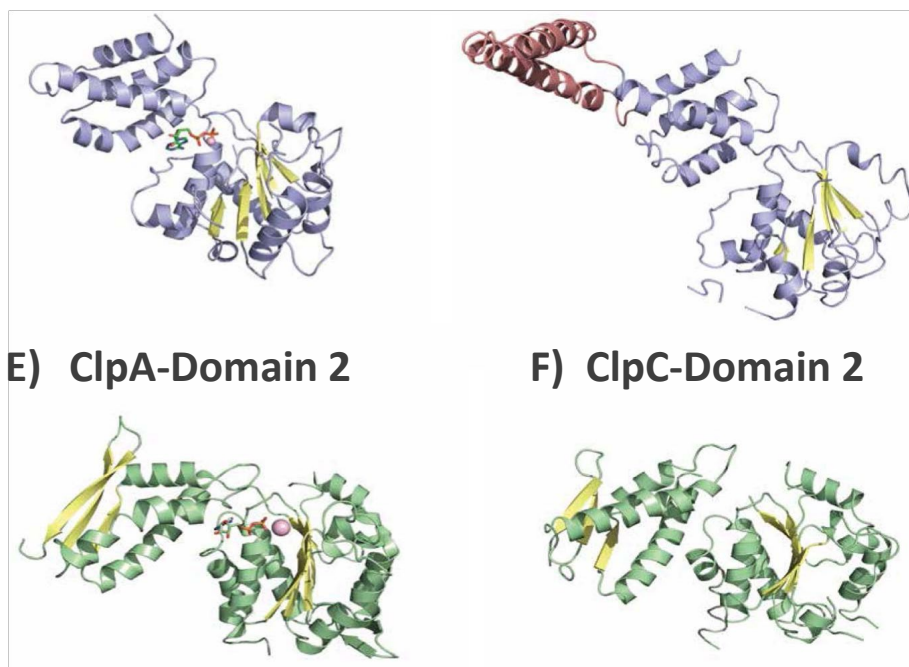
Figure 2 - Structural Comparison of ClpA/C AAA + ATPase Domains. A) Domain 1 of ClpA and ClpC both possess the AAA+ α–β–α fold topology. Helices and strands within the core α/β fold are colored in blue and yellow, respectively. Domain 1 contains a short α-helical insertion between α2 and β2 that is colored orange. The Middle Domain insertion between α8 and α7 is specific to ClpC and is colored salmon. B) Like Domain 1, Domain 2 of ClpA and ClpC both possess the AAA+ α–β–α fold topology. Helices and strands within the core α/β fold are colored in green and yellow, respectively. Domain 2 contains a β-hairpin insertion between α3 and β4 that is colored orange. C-D) The Middle Domain differentiates Domain 1 of ClpA (C) versus ClpC (D). Helices and strands within the base and lid domains are shown as blue ribbons and yellow arrows, respectively, and Middle Domain helices are represented by salmon ribbons. E-F) Domain 2 in ClpA and ClpC are structurally homologous. Helices and strands within the base and lid domains are shown as green ribbons and yellow arrows, respectively. Domains 1 and 2 for ClpA each contain a bound nucleotide molecule that is shown in stick. The protein topology cartoons in (A) and (B) were prepared using the TopDraw software package. All structure representations in Figure 1 were prepared with the PyMol software package and PDB accession codes 1R6B (C,E) (35), 3PXG (D), and 3PXI (F) [33].

The D2 insertion represents the presensor 1 β-hairpin (PS1 β), which is an element that is positioned near the central hexameric channel in the AAA+ PS1 β superclade [28, 40]. The PS1 β hairpin feature has been reported to mediate D1-D2 communication in Hsp104 oligomers [91]. Multiple sequence alignments indicate that each of these structural features is conserved in all of the Class 1 Clp/Hsp100 proteins examined here.

In addition to the α-β-α core, D1 and D2 each have a C-terminal αhelical bundle referred to as the lid domain (Fig. 2A and B). The functional roles of this domain may include the formation of a lid over the nucleotide binding site and facilitation of subunit interactions within the assembled hexamer [45]. The D2 lid domain is observed to possess a three-stranded β-sheet insertion, where β6 originates from the C-terminal end of α5 and β7-β8 form at the extreme C-terminus of the D2 AAA+ fold (Fig. 2B). Structural data suggests that variations of this insertion occur in both protein unfoldases and disaggregases with characterized examples including ClpA (Fig. 2E), ClpB [112], ClpC (Fig. 2F), Hsp104 [91], and HslU [111].

Given the high degree of structural homology between Class 1 Clp/ Hsp100 proteins, it is tempting to extrapolate experimental observations interchangeably between proteins. This approach, however, neglects the potential for functional divergence between proteins. One example centers on the hG[L/F] ClpP-interaction motif (h is any hydrophobic residue) located between β4 and α4 in D2 (Fig. 2B). Variations of this motif are present in all Clp/Hsp100 proteins known to interact with ClpP, but are not found in the ClpB/Hsp104 disaggregases or the ClpL aggregation-preventing unfoldase [47-48]. Incorporation of the ClpP-interaction loop into either ClpB or Hsp104 can support ClpP-interactions [49-51]. Another example is the relatively weak K_M of M. tuberculosis ClpC1 for ATP during

hydrolysis (3–5 mM), which contrasts with the reported values for E. coli ClpA equal to ~ 400 μM [37, 42, 52-53]. The weak Michaelis constant observed for M. tuberculosis ClpC1 binding to ATP has been proposed as a mechanism to sense the growth state of the bacteria since cellular [ATP] varies between dormant and actively proliferating states [53]. Such functional differences may be driven by structural changes that are the result of evolutionary adaptation of these proteins to specific environmental pressures.

The reported differences between Class 1 Clp/Hsp100 proteins from different bacterial species led us to ask: are divergences in protein behavior due to evolutionary changes in primary sequence? To address this question, we utilized phylogenetic methods to determine the evolutionary relationship between Class 1 Clp/Hsp100 proteins ClpA, ClpC, ClpE, ClpK, and ClpL proteins from a variety of bacteria. We have structured this analysis to examine Class 1 Clp/Hsp100 proteins with similar predicted overall domain architecture since this grouping yields a protein family with members that are highly similar in domain architecture and likely in function. Our analysis suggests that proteobacterial ClpC evolved via lateral gene transfer, but that this clade of proteins are not equally functional due to structural diversifications that may prevent ClpP-association. Furthermore, this analysis predicts that the first ClpA protein appeared in flavobacteria without an intact ClpP-interaction loop, which is a feature that later evolved in ClpA proteins from epsilonproteobacteria. The data reported here directly advance our knowledge regarding the origin and evolution of an important group of proteins required for ATP-dependent proteolysis in bacteria.
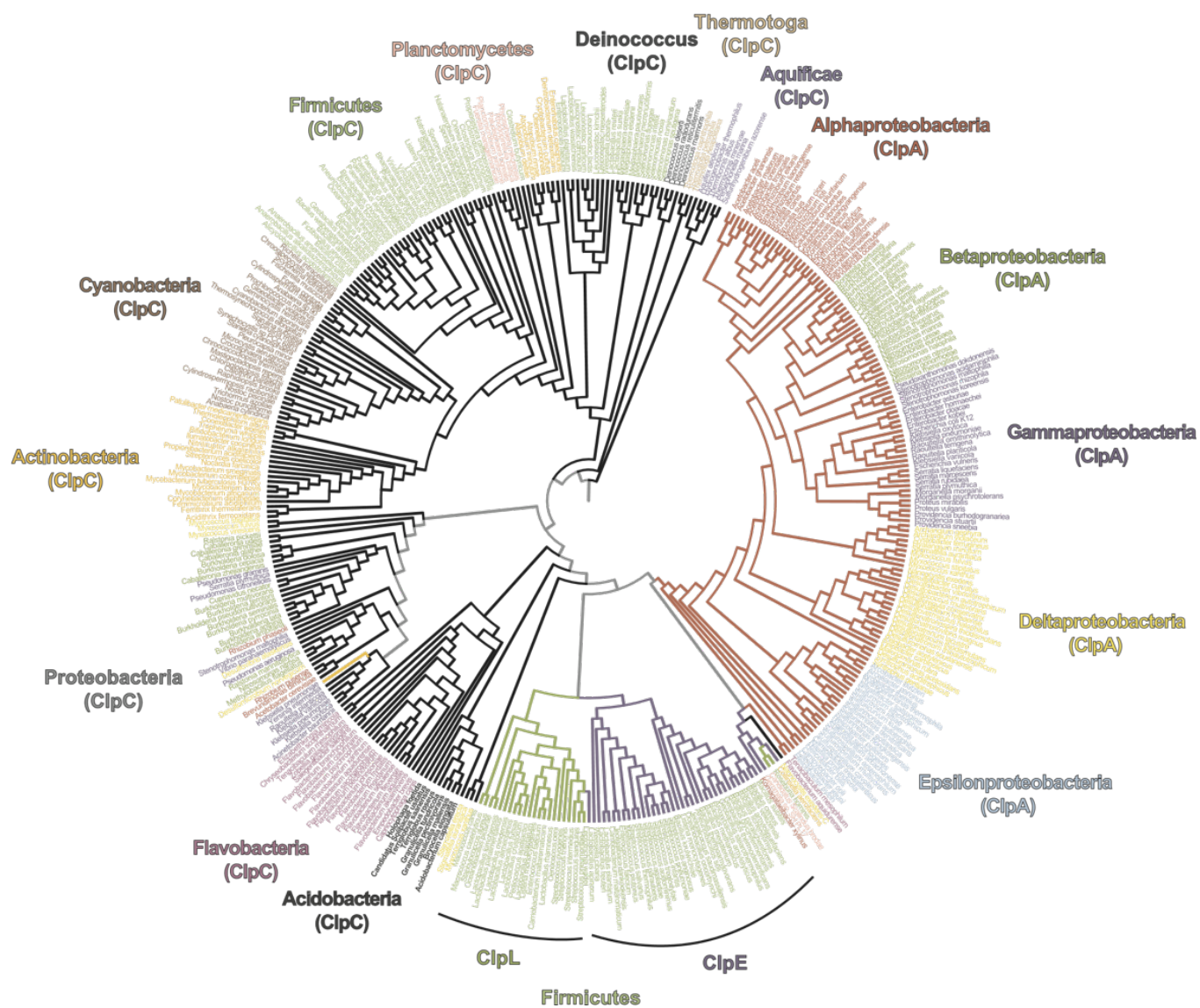
Figure 3 - Molecular Phylogeny of ClpA and ClpC protein sequences. Distance methods were used to generate a phylogenetic model to describe the evolutionary relationship between 422 Clp/Hsp100 primary sequences. All nodes are supported strongly by bootstrap analysis performed from 655 replicates. The majority-consensus tree representation shown in Fig. 2 was created and pruned using the Mesquite 3.2 software package. Branches are colored as red, black, blue, orange, or green to identify primary sequences as ClpA, ClpC, ClpE, ClpK, or ClpL, respectively.

Given the ambiguity resulting from our qualitative comparison of ClpC/ClpE/ClpL protein sequences discussed above, we endeavored further to clearly establish the relationship between Class 1 Clp/Hsp100 proteins. Global phylogenetic analysis of protein sequences identified as ClpA, ClpC, ClpE, ClpK, or ClpL revealed the existence of distinct clades. The tree shown in Fig. 3 was obtained using distance methods. Complimentary analyses using Bayesian and maximum parsimony (MP) methods correspond closely as well. The shortest tree was obtained by distance methods and roughly reflects the branch order expected from 16S rRNA-based bacterial phylogenies (Fig. 3) [54]. Aquificae was used as the outgroup for this analysis due to it being the most deeply branching bacterial phylum [55]. The assumption that ClpC speciation began in Aquificae is supported by the observation of long branches that are attracted to the base of the unrooted tree (data not shown) [56-57]. We do not expect that this is a consequence of long branch attraction (LBA) since all species represented within this cluster belong to the Aquificae phylum [58]. Despite the noticeably long Thermotogae branches, we also do not interpret this to be consistent with LBA given a shared composition of hyperthermophilic bacteria in Thermotogae and Aquificae as well as previous phylogenetic reconstructions that place the two phyla in close proximity.

Our analysis predicts a diversification event that resulted in the formation of proteobacterial ClpA from an ancestral ClpC protein. All ClpA protein sequences are found in a well-defined cluster that subdivides clearly into clades corresponding to the major proteobacterial divisions: epsilonproteobacteria, deltaproteobacteria, gammaproteobacteria, betaproteobacteria, and alphaproteobacteria (Figure 3) Our analysis demonstrates a node that yielded ClpE/ClpL proteins and all proteobacterial ClpA proteins

(Figure 3, branches are colored as black, red, blue, orange, or green to represent ClpC, ClpA, ClpE, ClpK, or ClpL, respectively). This analysis suggests that ClpE/ClpL proteins represent evolutionary intermediates between ClpC and ClpA. However, we predict that the diversification event that produced ClpA occurred in flavobacteria via lateral gene transfer since copies of the clpC and clpA genes are present in Tenacibaculum mesophilum and Flavobacterium aquidurense. Because these organisms do not harbor copies of the clpE or clpL genes, we predict that flavobacterial ClpA evolved directly from ClpC. However, both of these flavobacterial clpA genes lack a critical tripeptide motif necessary for ClpP-association, which is a feature that is fully conserved in proteobacterial ClpA proteins. From this, we propose that ClpA proteins competent for involvement in regulated proteolysis occurred first in epsilonproteobacteria.
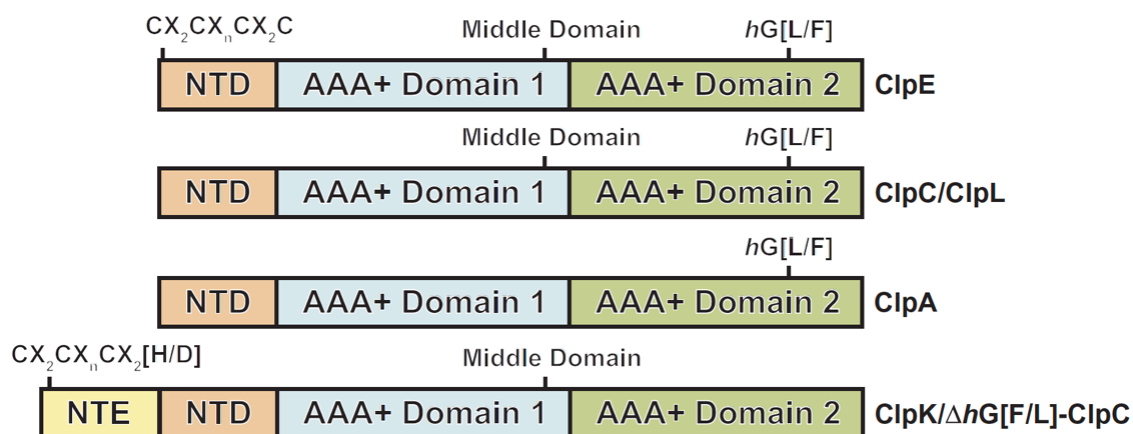
Figure 4 - General architecture of Class 1 Clp/Hsp100 proteins. Class 1 Clp/Hsp100 proteins contain an N-terminal domain (NTD) and two distinct AAA+ ATP binding and hydrolysis sites, Domain 1 (D1) and Domain 2 (D2). Insertions, specific primary sequence motifs, and the presence of an N-terminal extension (NTE) differentiate these proteins into unique families.Specific features include the presence of an N-terminal zinc binding motif, a coiled-coil MiddleDomain, and a ClpP-interaction loop (hG[L/F]).

From an evolutionary perspective, our data suggest that a common ancestor first evolved the zinc binding domain and then later lost the hG[L/F] ClpP-interaction motif. Interestingly, ClpK also contains a putative zinc binding domain (Figure 4) [107]. Given the zinc binding feature, primary sequence homology patterns, and the disappearance of a clearly identifiable hG[L/F] motif, it is highly likely that proteobacterial ClpC proteins evolved into ClpK. For this reason, we classify the ΔhG[L/F] group of ClpC proteins as belonging to a clade of ClpK-like proteins that likely share a high degree of structural and functional similarity (Figure 4). Our classification results in distinct clades for ClpA, ClpC/ClpL, ClpE, and ClpK (including ΔhG[L/F] proteobacterial ClpC proteins) proteins since each group is phylogenetically distinct. Based on existing biochemical data for ClpK, this suggests a divergence in function for proteobacterial ClpC proteins [59-60].

This analysis of the evolutionary relationship between the ClpP-interacting subset of Class 1 Clp/Hsp100 proteins has advanced our understanding of the functional diversity found within this protein family. The molecular phylogeny reported here strongly predicts a divergent function for proteobacterial ClpC proteins as a result of evolution-based structural diversifications. Because of these divergences, we hold that it would be erroneous to assume complete conservation of structure and function among all ClpC proteins. Due to the importance of regulated proteolysis in maintaining cellular homeostasis, the basic mechanistic details describing ClpC function in varied bacteria will likely provide novel routes to the development of therapeutic treatments designed to eliminate infections by pathogenic bacteria.

3. Materials and Methods

Primary sequence searches were made using the National Center for Biotechnology Information (NCBI) Protein Database, which includes protein sequence records from multiple sources that include Genpept, RefSeq, Swiss-Prot, PIR, PRF, and PDB. Protein sequences were obtained in the FASTA format and aligned using the MUSCLE algorithm within the MEGA7 software suite or MegAlign Pro (DNASTAR). Sequences were classified by comparison of primary sequence features with characterized protein sequences of known identity as discussed below [61]. Percent Identity Matrices were calculated using MegAlign Pro (DNASTAR). Molecular phylogenies were computed from full-length protein sequences using MEGA7, PAUP 4.0, and Mr. Bayes [61-64]. The shortest tree was identified using distance methods within PAUP 4.0. Nodal support was obtained by 655 bootstrapping replicates using heuristic, step-wise addition search methods. Circular phylogenetic tree representations were created and pruned using the Mesquite 3.2 software package [65]. All structural representations were created using the Pymol Software Package.

4. Conclusion

The active Clp protease forms through the association of a hexameric Clp/Hsp100 ATPase with the oligomeric serine protease ClpP. A comprehensive search of protein sequences deposited in the NCBI Protein Database reveals that ClpX proteins are present in nearly all bacterial phyla. In contrast, a similar search for Class 1 Clp/Hsp100 protein sequences suggests phylum-specific groupings. For example, ClpC proteins are present in bacterial phyla ranging from Aquificae to Actinobacteria, but are not significantly represented in proteobacteria. However, proteobacterial species lacking ClpC proteins

typically possess a copy of the clpA gene. The structurally similar ClpE and ClpL proteins are present only in firmicutes. Given a common biological function and highly similar domain architecture, we asked: what structural features differentiate Class 1 Clp/Hsp100 proteins and can these differences be utilized to establish the evolutionary relationship between them? To address these questions, we used protein sequences and structural models available from public databases to examine this relationship in an evolutionary context.

This analysis of the evolutionary relationship between the ClpP-interacting subset of Class 1 Clp/Hsp100 proteins has advanced our understanding of the functional diversity found within this protein family. The molecular phylogeny reported here strongly predicts a divergent function for proteobacterial ClpC proteins as a result of evolution-based structural diversifications. We have identified a group of ClpC proteins that lack an intact ClpP interaction loop (hG[L/F]) and harbor a putative zinc binding domain, which positions this group as potential thermoprotection factors similar to *K. pneumoniae* ClpK. This is clinically significant due to the fact that sterilization of specific hospital equipment by heat treatment has been shown to sometimes be ineffective as a result of thermotolerance conferred to *K. pneumoniae* by ClpK and associated small heat-shock proteins [61]. By comparison, we have identified additional pathogens such as *A. baumannii and S. maltophilia* that contain ClpK-like proteins, which suggests the potential to evolve similar thermotolerance properties. For this reason, drug development efforts using ClpC as a target in proteobacterial pathogens such as *K. pneumoniae*, *A. baumannii, S. maltophilia*, etc. may require novel approaches given the discussed structural divergence

between ClpC proteins. Due to the importance of regulated proteolysis in maintaining cellular homeostasis, the basic mechanistic details describing ClpC function in varied bacteria will likely provide novel routes to the development of therapeutic treatments designed to eliminate infections by pathogenic bacteria.

CHAPTER III: MtbClpC1 N-terminal domain is dispensable for allosteric regulation by adaptor protein

1. Introduction

ATP-dependent proteases represent a family of molecular machines responsible for the regulated turnover of misfolded, aggregated, or degradation-tagged cellular proteins [8,14,15]. In mycobacteria, the regulated removal of protein substrates in the cytoplasm is mediated by at least four different proteolytic complexes that are broadly divided into two groups that include the bacterial-like proteases (FtsH, Lon, ClpXP and ClpC1P) and the eukaryotic-like proteasome [19]. These proteases share a common architecture in which a ring-shaped AAA+ (ATPases Associated with various cellular Activities) ATPase associates with one or both ends of a barrel-shaped peptidase that contains active sites inaccessible to bulk solvent [14,19]. In the active state, the hexameric ATPase couples the energy of ATP hydrolysis to unfold and thread protein substrates into the associated protease for degradation.

Substrate recognition by AAA+ proteases is driven by the ATPase component, which functions as a gate-keeper to selectively determine protein substrates to be degraded. However, the specificity of AAA+ unfoldases for protein substrates is often modulated by an associated adaptor protein. For example, *Escherichia coli* (*E. coli*) ClpA specifically recognizes the SsrA degradation sequence, which is a C-terminal degradation tag that is co-translationally added by the tmRNA tagging system to nascent polypeptide chains of stalled ribosomes [69]. ClpA has been reported to bind the SsrA sequence with an affinity equal to ~200 nM, which drives specific degradation of SsrA-tagged proteins by the ClpAP complex [70]. Association with the ClpS adaptor protein results in a substantial decrease

in the affinity of ClpA towards SsrA-tagged proteins such that SsrA-tagged substrates are not degraded by the ClpAPS complex [71–73]. Deletion of the N-terminal domain of ClpA reverses this observation such that ClpAP catalyzed degradation of an SsrA-tagged Green Fluorescent protein occurs independent of *E. coli* ClpS [71]. Complimentary work has demonstrated that a single ClpS molecule associates with ClpA hexamers with an affinity of ~40 nM but that an additional one to two ClpS molecules may associate with a decreased affinity equal to >700 nM [73,74]. Taken together, saturation of ClpS binding to ClpA hexamers as indicated by a 6:1 ratio of ClpS:ClpA$_6$ is not necessary for significant control of ClpA function but an intact N-terminal domain (NTD) is required for ClpS-dependent inhibition of ClpAP catalyzed degradation of SsrA-tagged protein substrates.

Until recently, ClpC protein function was widely thought to depend on adaptor protein association. *Bacillus subtilis* (*B. subtilis*) ClpC oligomerization and subsequent chaperone activity was previously reported to depend on the association of the MecA adaptor protein [75,76]. High-resolution structures for *Staphylococcus aureus* (*S. aureus*) ClpC in the presence and absence of MecA have recently revealed a more complicated picture. Association of the MecA adaptor protein with *S. aureus* ClpC promotes formation of enzymatically active ClpC hexamers via transition from an inactive helical assembly [77]. *B. subtilis* ClpCP can also bind and degrade phosphoarginine substrates independent of any adaptor protein [78], thereby demonstrating that MecA association is not always required for ClpCP function as once thought. In contrast, ClpC proteins from cyanobacteria and actinobacteria are not widely known to possess chaperone activities that are adaptor protein-dependent [42, 53, 79–81]. All available data describing actinobacterial ClpC protein function have been collected using *Mycobacterium tuberculosis* (*M. tuberculosis*)

ClpC1 (ClpC2 lacks identifiable motifs associated with ATPase activity) and demonstrate adaptor protein-independent activity. However, the protein degradation activity of *Synechoccus elongatus* ClpCP3/R is regulated by two ClpS isoforms, ClpS1 and ClpS2, where ClpS1 promotes the binding of N-degron protein substrates bearing N-terminal Phe and Tyr residues and ClpS2 blocks degradation of α-casein substrates [80]. No data has been reported regarding adaptor protein-dependent regulation of Clp protease complexes in actinobacteria.

Given the close phylogenetic relationship between mycobacterial and cyanobacterial ClpC proteins [82], it is likely that the former may be subject to ClpS-mediated control. However, it is currently unclear whether mycobacterial ClpC1 is subject to regulation by any adaptor protein. For this reason, we set out to determine the functional relationship between *M. tuberculosis* H37Rv ClpC1 (Accession Number: *Rv3596c*) and ClpS (Accession Number: *Rv1331*). That is to say, if a physical ClpC1:ClpS interaction occurs, does it impact ClpC1 function? Here, we demonstrate that a physical association occurs to form a stable complex detectable by pulldown methods. From stopped-flow fluorescence experiments reporting on ClpC1 catalyzed unfolding of an SsrA-tagged fluorescent protein, we report a [ClpS]-dependent two-state transition between a fully catalytic state and a partially-inhibited ClpC1 state. Our data suggest that specific residues located in the Middle Domain are necessary for allosteric control by ClpS independent of the ClpC1-NTD. These results demonstrate that ClpS allosterically impacts ClpC1 catalyzed unfolding of an SsrA-tagged protein and concurrently represent the first report of adaptor protein-mediated regulation of a mycobacterial ClpC protein.

2. Results

2.1. Dependence of Apparent Unfolding Rate Constant on [ClpS]

To determine whether mycobacterial ClpS is a functional adaptor of ClpC1, we performed stopped-flow fluorescence experiments using a method that reports on ClpC1 catalyzed unfolding of an SsrA-tagged protein. Figure 5A illustrates the experimental design as described in Materials and Methods. In our experiments, syringe A of the stopped-flow fluorometer contains a solution of 1 µM ClpC1 incubated in the presence or absence of ClpS. Syringe B is loaded with a solution containing 9.5 mM ATP and 100 nM photoactivated SsrA-Kaede (SsrA-Kaede$_{Red}$). *Trachyphyllia geoffroyi* (*T. geoffroyi*) Kaede belongs to a family of fluorescent proteins that are structurally homologous to the green fluorescent protein (GFP) from *Aequorea Victoria* [83]. Kaede contrasts GFP through photoactivation-dependent excitation and emission properties, where irradiation by
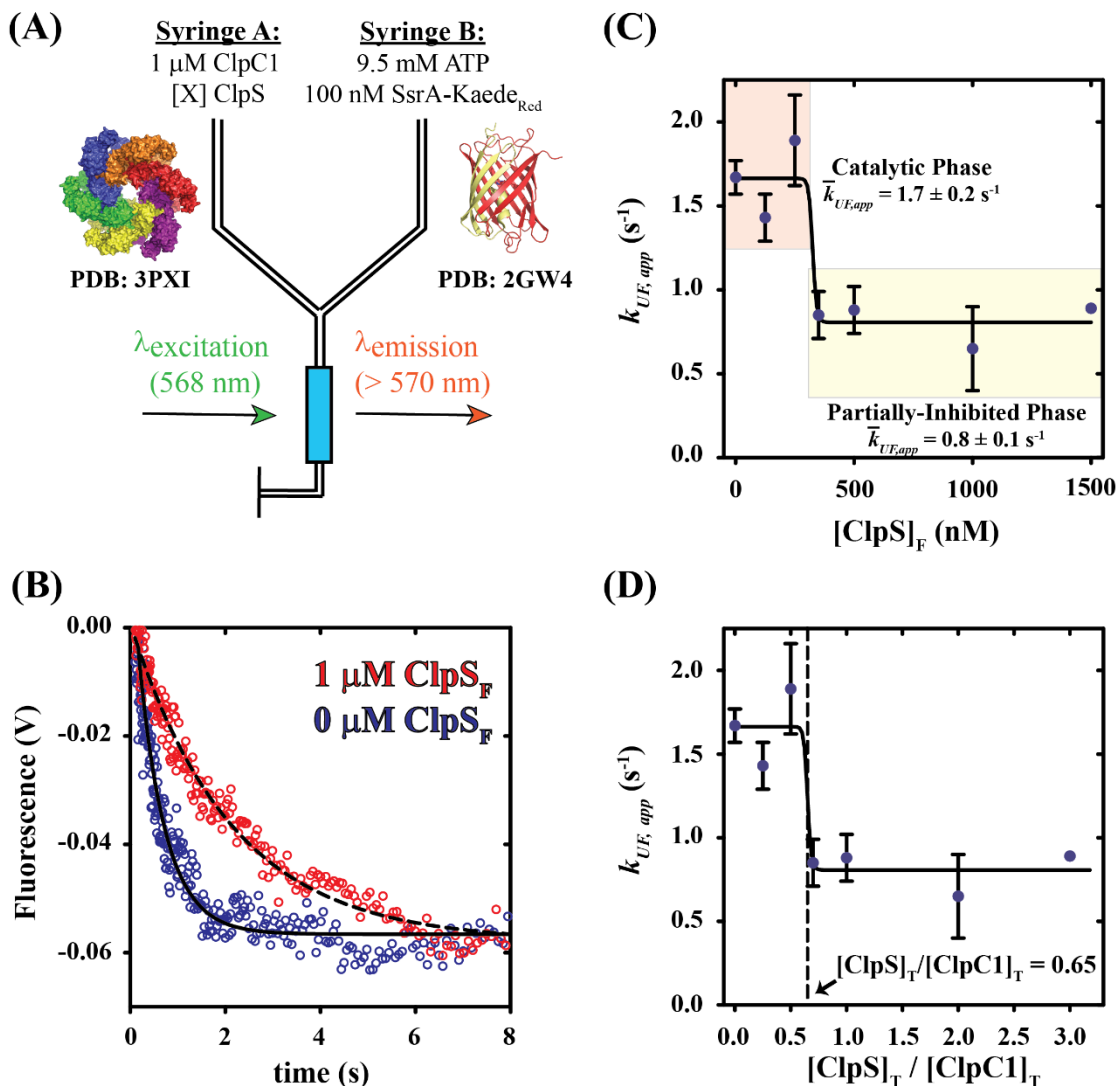
Figure 5 - **Examining the impact of [ClpS] on ClpC1 catalyzed unfolding of an SsrA-tagged protein**. (A) Schematic representation of stopped-flow fluorescence protein unfolding experiments. Syringe A contains the indicated reagents, 1 µM ClpC1 and varied concentrations of ClpS (indicated in text). Syringe B contains 9.5 mM ATP to fuel protein unfolding and 100 nM photoactivated Kaede bearing a C-terminal SsrA-degradation tag (SsrA-KaedeRed). Fluorescence is observed using an excitation wavelength equal to 568 nm and emissions are observed above 570 nm with a 570-nm-long pass filter. Upon mixing, the concentrations are two-fold lower than in the preincubation syringe. (B) Representative fluorescence time courses for ClpC1 catalyzed SsrA-KaedeRed unfolding. Time courses represent 1 µM ClpC1 incubated with (Red Circles) or without (Blue Circles) 2 µM ClpS prior to mixing with 9.5 mM ATP and 100 nM SsrA-KaedeRed. The dashed and solid lines represent nonlinear least squares (NLLS) fits using a single-exponential function for time courses collected in the presence or absence of ClpS, respectively. All apparent unfolding rate constants are summarized in Table 1. (C) Dependence of the apparent unfolding rate constant, kUF,app, on [ClpS]F, where [ClpS]F represents the final mixing concentration of ClpS. Average unfolding rate constants for the Catalytic and Partially-Inhibited phases are equal to 1.7 ± 0.2 and 0.8 ± 0.1 s−1, respectively. (D) Replotting the data shown in Figure 1C as kUF,app versus [ClpS]T/[ClpC1]T indicates that the transition from Catalytic phase to Partially-Inhibited phase occurs when the ratio of total ClpS concentration to total ClpC1 monomer concentration equals 0.65. All data shown are from independent experiments and error bars indicate ± standard deviation.

elimination reaction [23–25]. Cleavage by photoactivation yields Kaede$_{Red}$, which exhibits red-shifted emissions and a chromophore structure that is distinct from the green (non-photoactivated) form [83]. In our experiments, we include Kaede$_{Red}$ bearing a C-terminal SsrA degradation tag (SsrA-Kaede$_{Red}$). As illustrated by Glynn and coworkers with *E. coli* ClpX, unfolding of SsrA-Kaede$_{Red}$ irreversibly displaces an N-terminal Kaede fragment and quenches native Kaede$_{Red}$ emissions [84]. From this, we expect that mixing the contents of the two syringes depicted in Figure 1A will result in ClpC1 catalyzed unfolding of SsrA-Kaede$_{Red}$ and subsequent quenching of fluorescence observed at wavelengths longer than 570 nm.

Figure 1B shows representative fluorescence time courses collected by rapidly mixing the contents of syringes A and B, as schematized in Figure 1A. The representative time courses were collected with final mixing concentrations of each reactant equal to 0.5 µM ClpC1, 50 nM SsrA-Kaede$_{Red}$ and 4.75 mM ATP in the presence (Red trace in Figure 1B) and absence (Blue trace in Figure 1B) of 1 µM ClpS. As predicted, all time courses display a time-dependent decrease in observed emissions. The magnitude of the observed change in SsrA-Kaede$_{Red}$ emissions in this experimental design is low relative to previously reported translocation assays performed with other Clp/Hsp100 family members [85-86]. To insure that our observed emissions signal was reproducible, we repeated the experiments shown in Figure 1B six times using protein samples derived from independent preparations. As such, we are confident in asserting that the rate of ClpC1 catalyzed unfolding of SsrA-Kaede$_{Red}$ is dependent on [ClpS]. Nonlinear least squares (NLLS) analysis of each time course shown in Figure 1B using a single-exponential function yields an apparent unfolding rate constant, $k_{UF,app}$, equal to 1.7 ± 0.1 or 0.7 ± 0.3 s^{-1} for conditions

including 0 or 1 μM ClpS$_F$ (final mixing [ClpS]), respectively. Our observation of a nearly three-fold decrease in the apparent unfolding rate constant for conditions lacking ClpS relative to conditions with [ClpS]$_F$ in excess lead us to conclude that *M. tuberculosis* ClpC1 is inhibited by ClpS.

We next examined the ClpS concentration dependence of the apparent unfolding rate constant describing ClpC1 catalyzed the protein unfolding. Stopped-flow fluorescence experiments were performed as schematized in Figure 1A by varying the [ClpS] in syringe A. Time courses were collected at final mixing concentrations of ClpS equal to 0, 125, 250, 350, 500, 1000 and 1500 nM. Each data set was subjected to NLLS analysis using a single-exponential function to determine the apparent rate constant (Table 1). A plot of the apparent unfolding rate constant, $k_{UF,app}$, versus the final mixing concentration of ClpS, [ClpS]$_F$, displays a pronounced dependence on the molar concentration of ClpS$_F$ (Figure 1C). NLLS analysis of the data shown in Figure 1C using a Hill function (Equation (1)) suggests an affinity describing the ClpS:ClpC1 interaction as ~325 nM under the final mixing conditions examined here. Due to the steep slope, our analysis was unable to accurately estimate the Hill coefficient. The maximum binding stoichiometry can be determined from the breakpoint in a plot of the degree of binding versus the ratio of the concentrations of total ligand to total macromolecule, [X]$_T$/[M]$_T$ [87]. Since ClpC1 is incubated with ClpS prior to mixing with ATP and SsrA-Kaede$_{Red}$ in our experiments, we expect that the resulting apparent rate constants describing the [ClpS]$_F$-dependence of SsrA-tagged protein unfolding by ClpC1 are proportional to the degree of binding. Figure 5D shows a plot of the apparent unfolding rate constant versus the ratio of total ClpS$_F$ concentration to ClpC1 monomer concentration. This plot yields a curve with an inflection

point equal to 0.65. An inflection point of less than unity strongly suggests a deviation from a 1:1 binding stoichiometry between ClpS and ClpC1.Calculation of the ratio of $[ClpC1]_T$ to $[ClpS]_T$, instead of $[ClpS]_T{:}[ClpC1]_T$, using monomeric terms yields an estimate of 1–2 ClpC1 molecules associated per ClpS monomer ($1/0.65\ [ClpS]_F/[ClpC1]_T$) = 1.54 ($[ClpC1]_T/[ClpS]_T$)). Given that the data presented in Figure 5C, D are not hyperbolic, this observation likely indicates the average binding stoichiometry to be two ClpC1 subunits per single ClpS molecule.

Table 1. Apparent rate constant describing ClpC1 catalyzed SsrA-Kaede$_{Red}$ unfolding as a function of $[ClpS]_F$.

$$K_{UF,\ app}\ (s^{-1})$$

$[ClpS]_F$ (nM)	*Full-Length ClpC1*	*ΔNTD-ClpC1*
0	1.7 ± 0.1	2.4 ± 1.1
125	1.4 ± 0.1	
250	1.9 ± 0.3	1.4 ± 0.1
350	0.9 ± 0.1	
500	0.9 ± 0.1	
1000	0.7 ± 0.3	0.9 ± 0.2
1500	0.89 ± 0.03	

$K_{UF,\ app}$ is the apparent rate constant describing ClpC1 catalyzed unfolding of an SsrA-tagged protein. $[ClpS]_F$ represents the final reaction concentration of ClpS after mixing in the stopped-flow spectrophotometer.

2.2. N-Terminal Domain of ClpC1 Is Dispensable for ClpS-Mediated Inhibition of Protein Unfolding

Each time course shown in Figure 6A was subjected to NLLS analysis using a single-exponential function to determine the apparent rate constant (Table 1). The resulting apparent unfolding rate constants, $k_{UF,app}$, corresponding to both full-length ClpC1 and ΔNTD-ClpC1 are plotted in Figure 2C. In the absence or presence of 1 μM ClpS$_F$, the
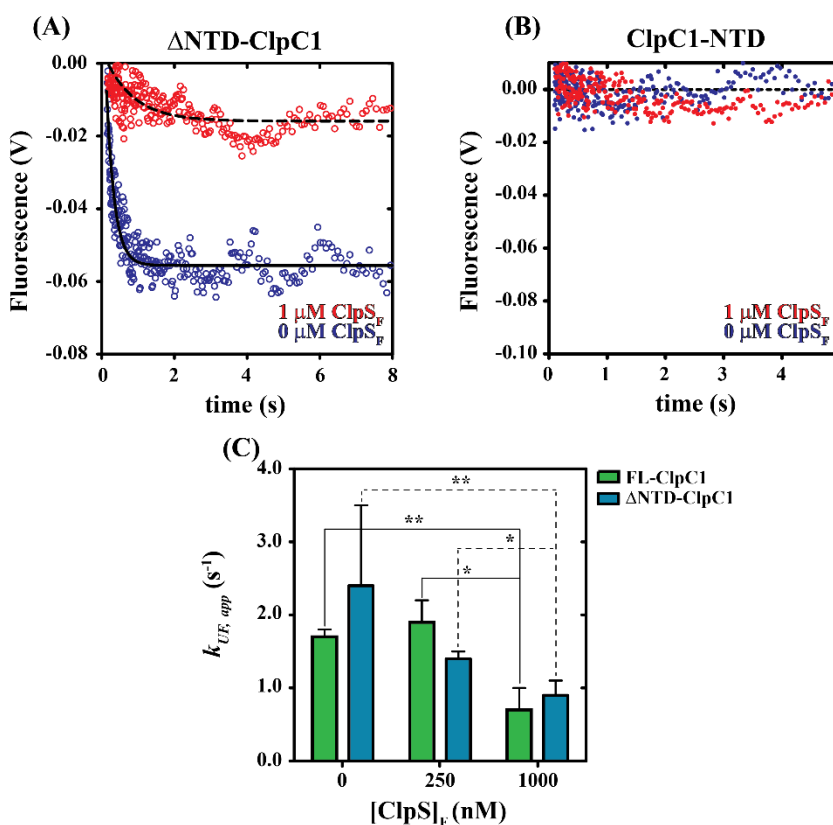
Figure 6 - The ClpC1 N-terminal domain is dispensable for protein unfolding. (A) Representative fluorescence time courses for ΔNTD-ClpC1 catalyzed SsrA-KaedeRed unfolding. Time courses represent 1 μM ΔNTD-ClpC1 incubated with (Red Circles) or without (Blue Circles) 2 μM ClpS prior to mixing with 9.5 mM ATP and 100 nM SsrA-KaedeRed. The dashed and solid lines represent NLLS fits using a single-exponential function for time courses collected in the presence or absence of ClpS, respectively. All apparent unfolding rate constants are summarized in Table 1. (B) Representative fluorescence time courses for ClpC1-NTD catalyzed SsrA-KaedeRed unfolding with coloring identical to Figure 1A. No significant unfolding is observed; all data points are observed to fluctuate about the baseline. (C) Comparison of the apparent unfolding rate constants observed for full-length ClpC1 (Green Bars) and ΔNTD-ClpC1 (Blue Bars) in the presence of 0, 250 and 1000 nM ClpS. * $p \leq 0.05$, ** $p \leq 0.001$ calculated from an unpaired Student's t-test (two-tailed). All data shown are from independent experiments and error bars indicate ± standard deviation.

apparent unfolding rate constant describing ΔNTD-ClpC1 catalyzed unfolding of SsrA-Kaede$_{Red}$ is 2.4 ± 1.1 or 0.9 ± 0.2 s^{-1}, respectively. Unfolding rate constants estimated in the presence of 0 or 250 nM ClpS$_F$ for either full-length ClpC1 or ΔNTD-ClpC1 are statistically indistinguishable. However, comparison of means by *t*-testing demonstrate that the mean unfolding rate constants describing ClpC1 or ΔNTD-ClpC1 catalyzed protein unfolding in the presence of 0 versus 1000 nM ClpS$_F$ or 250 versus 1000 nM ClpS$_F$ are statistically different (Figure 6C). Taken together, our data suggest a model wherein the NTD is dispensable for ClpS-dependent allosteric inhibition of ClpC1 catalyzed unfolding of SsrA-tagged Kaede.

2.3. Mycobacterial ClpS Primary Sequence Analysis Reveals ClpC1-MD (Middle Domain) Binding Features

Previous X-ray structures for *B. subtilis* ClpC in complex with MecA indicate that complex formation involves binding interactions at the NTD and middle domain (MD) [31]. This observation led us to ask; can this prior knowledge on interactions between MecA and the *B. subtilis* ClpC-MD be used to make predictions regarding the identity of the ClpC1:ClpS interaction surface? To address this question, we utilized multiple sequence alignments to identify putative MD-ClpS interacting residues between actinobacterial ClpC and ClpS proteins (Figure 7A).

Reference to the ClpC:MecA complex structure predicts two *B. subtilis* ClpC-MD residues, R443 and Q432, to be involved in complex formation (Figure 7B) [31]. Multiple sequence alignments comparing *E. coli* ClpA and ClpC proteins demonstrates a high degree of sequence conservation among MD-containing ClpC proteins for the R443 position but lesser conservation is observed for the Q432 position (Figure 7A).
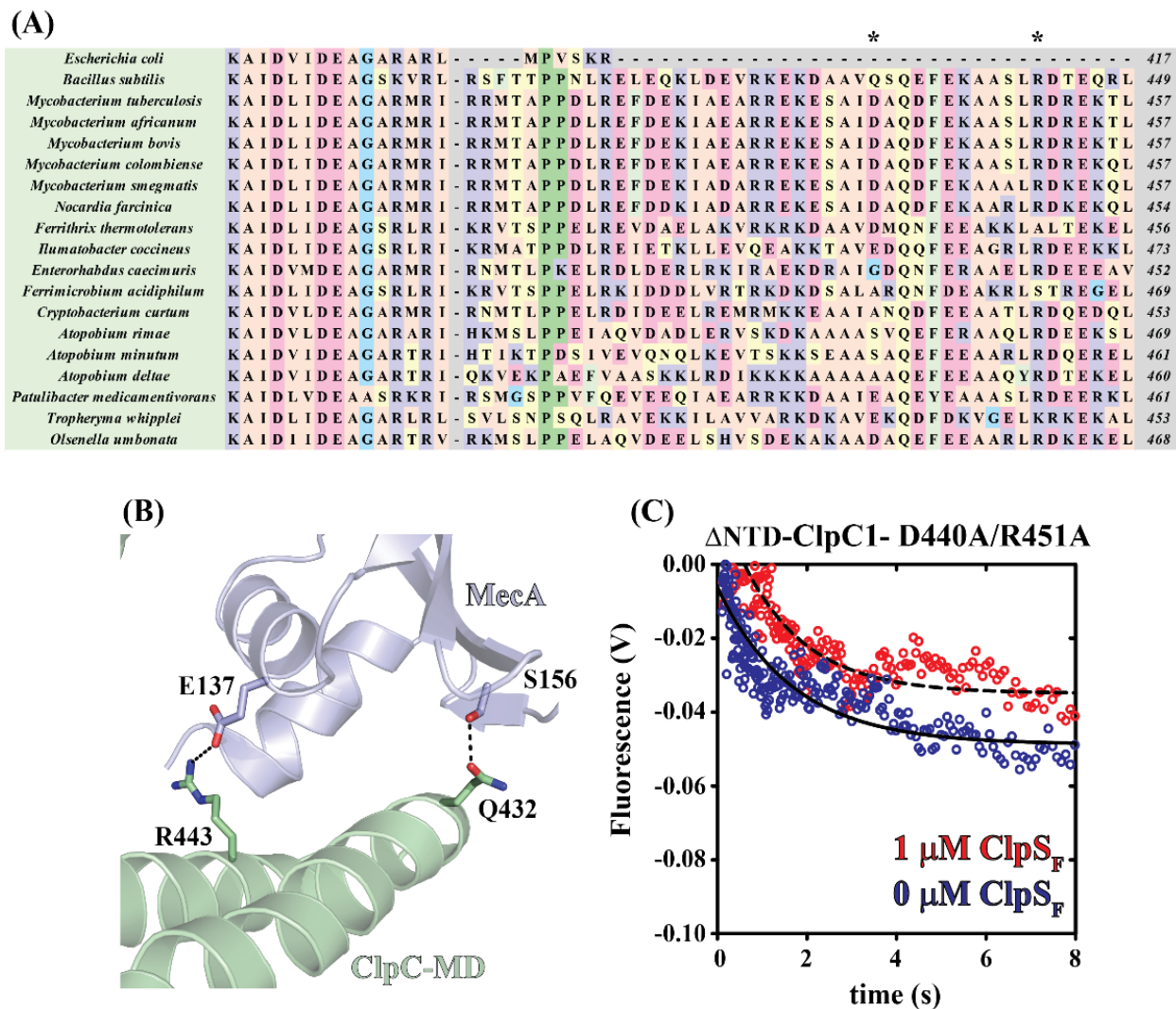
Figure 7 - **Primary sequence analyses predict the ClpC1-MD as a secondary binding surface for ClpS**. (A,B) Multiple sequence alignments reveal the conservation of two residues (indicated by asterisks *) in actinobacterial ClpC1 protein sequences that are involved in complex formation between B. subtilis MecA and the ClpC-MD. (C) Representative fluorescence time courses for ΔNTD-ClpC1-D440A/R451A catalyzed SsrA-KaedeRed unfolding. Time courses represent 1 μM ΔNTD-ClpC1-D440A/R451A incubated with (Red Circles) or without (Blue Circles) 2 μM ClpS prior to mixing with 9.5 mM ATP and 100 nM SsrA-KaedeRed. The dashed and solid lines represent NLLS fits using a single-exponential function for time courses collected in the presence or absence of ClpS, respectively. All apparent unfolding rate constants are summarized in Table 1. All structure representations in Figure 3 were prepared with the Pymol software package and Protein Databank (PDB) accession code 3PXG [31].

From the ClpC:MecA structure, we expect that a similar interaction would be possible if a residue is conserved at the *B. subtilis* residue position 432 that harbors a side-chain with a carbonyl functional group available for hydrogen bonding to the side-chain of nearby MecA S156. Consistent with this prediction, Figure 7A illustrates that mycobacterial ClpC proteins maintain a conserved aspartate in the corresponding residue position. Thus, the mycobacterial ClpC1-MD contains primary sequence features that may support interaction with *B. subtilis* MecA.

To determine the importance of ClpC1-MD residues for ClpS-dependent inhibition of ClpC1 catalyzed protein unfolding, we further modified ΔNTD-ClpC1 to harbor D440A and R451A mutations. These mutations are based on the *B. subtilis* R443 and Q432 residue positions as discussed above.

2.4. Identification of ClpC1 Surfaces Involved in Complex Formation

In order to resolve whether ClpC1:ClpS complex formation requires the ClpC1-NTD, pulldown experiments were performed using His_6-tagged ClpS as a reporter for ClpC1 binding. Varied concentrations of ClpC1 were incubated in the presence of 2 μM His_6-SUMO-ClpS and 1 mM ATPγS, a slowly-hydrolysable ATP analogue (Figure 4A). In this experimental design, all samples contain equimolar amounts of His_6-SUMO-ClpS and are treated with a constant volume of Ni-NTA slurry. Thus, we expect that the amount of His_6-SUMO-ClpS eluted from the Ni-NTA resin will remain constant across all conditions surveyed here. If His_6-SUMO-ClpS interacts physically with ClpC1, SDS-PAGE analysis will reveal a gel band that corresponds to ClpC1 with molecular weight equal to approximately 95 kDa.

The SDS-PAGE analysis illustrated in Figure 8A confirms that a ClpC1:ClpS interaction occurs. As expected, a gel band corresponding to the molecular weight of ClpC1 is not observed by Coomassie staining when ClpC1 is not present (Figure 4A, Lanes 2 and 4). We note the presence of a ClpC1 band when ClpC1 is incubated with 2 μM His$_6$-SUMO-ClpS in the presence and absence of ATPγS (Figure 8A, Lanes 3, 5–7). We are confident that this is not the result of contaminating His$_6$-SUMO-ClpC1 in our ClpC1 preparation since our purification protocol includes a Ni-NTA Immobilized Metal Affinity Chromatography (IMAC) step after His$_6$-SUMO-tag removal, thereby separating His$_6$-
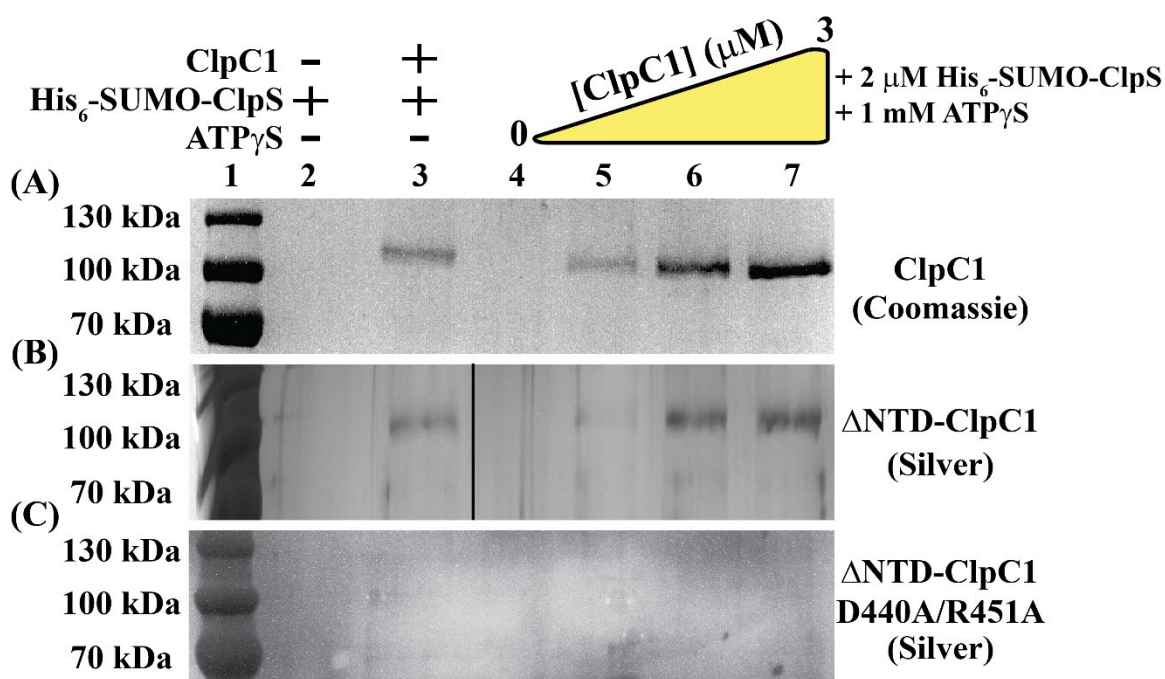
Figure 8 - Pulldown experiments reveal a physical ClpC1:ClpS interaction. All samples were prepared by incubating 2 μM His$_6$-ClpS with or without ClpC1/ATPγS. Following complex formation, His$_6$-ClpS was isolated using affinity pulldown methods based on the His$_6$-tag:Ni-NTA resin interaction. (**A**) *M. tuberculosis* ClpC1:ClpS protein complex formation is enhanced relative to background levels in the presence of 1 mM ATPγS. Lanes 2 and 3 were loaded with 2 μM His$_6$-ClpS previously incubated with or without 2 μM ClpC1, respectively. In addition to 2 μM His$_6$-SUMO-ClpS, Lanes 4–7 contained 0, 1, 2, or 3 μM ClpC1, respectively and 1 mM ATPγS. ClpC1:ClpS interactions were observed by Coomassie staining methods. (**B**) The ΔNTD-ClpC1 truncation shows diminished association with His$_6$-SUMO-ClpS when compared to wild-type *M. tuberculosis* ClpC1 as indicated by the need to visual band patterns using silver staining methods. All lanes were loaded as described in Figure 4A but with ΔNTD-ClpC1 substituted in place of full-length ClpC1. (**C**) Pulldown experiments described in Figure 4A were repeated by substituting ΔNTD-ClpC1-D440A/R451A in the place of full-length ClpC1.

SUMO-ClpC1 from cleaved ClpC1 (Section 4). We note the observation of non-specific interactions between full-length ClpC1 and Ni-NTA resin (Figure 9, Lane 1). However, the incubation of 2 µM His$_6$-SUMO-ClpS with 1 mM ATPγS and 2 µM ClpC1 yields qualitatively denser gel bands (Figure 9, Lane 3), which is an observation consistent with specific binding between His$_6$-SUMO-ClpS and ClpC1.

A) Full-Length ClpC1			**B) ΔNTD-ClpC1**		
ClpC1	+	−	+	+	−	+
His$_6$-SUMO-ClpS	−	+	+	−	+	+
ATPγS	+	+	+	+	+	+
	1	2	3	1	2	3
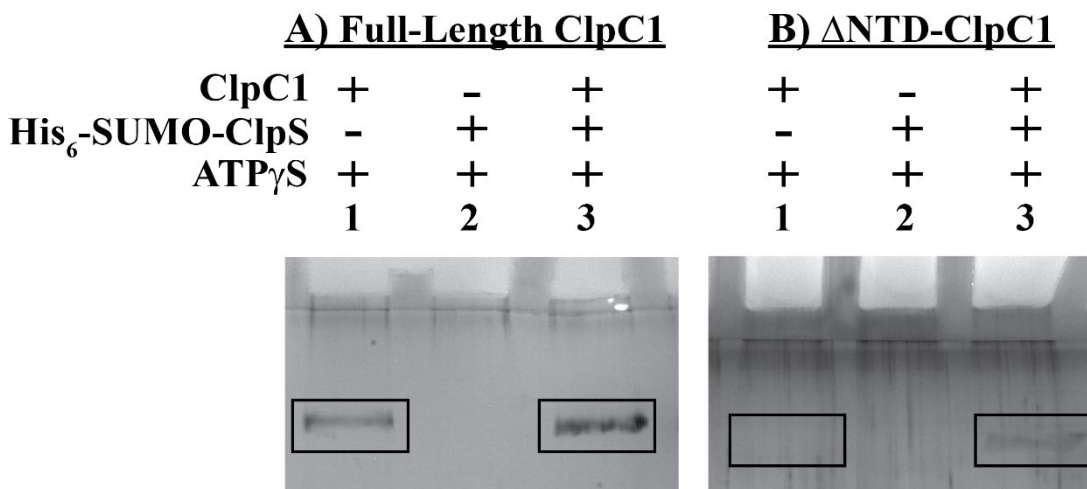
Figure 9 - Full-length ClpC1 Interacts Non-Specifically with Ni-NTA Resin. A) Control experiments performed as described in Figure 4 reveal non-specific binding between Ni-NTA and full-length ClpC1 in the absence of His6-SUMO-ClpS. B) Identical experiments performed with ΔNTD-ClpC1 do not support the same conclusion in the absence of an intact N-terminal domain. All SDS-PAGE gels shown in Figure S1 have been visualized by silver staining methods.

A common feature shared amongst characterized adaptor proteins is that the primary contact surfaces on the associated Clp/Hsp100 protein are located in the N-terminal domain [15,31,32]. Based on this and our stopped-flow fluorescence data presented in Figures 1 and 2, we performed additional affinity pulldown experiments with ΔNTD-ClpC1. Experiments were performed by incubation of 2 µM His$_6$-SUMO-ClpS with 0, 1, 2, or 3 µM ΔNTD-ClpC1 and 1 mM ATPγS (Figure 4B, Lanes 4–7). In order to resolve gel bands corresponding to ΔNTD-ClpC1 in these experiments, silver-stain

methods were required for SDS-PAGE gel visualization. Similar to the identical experiments performed with full-length ClpC1, ΔNTD-ClpC1 was observed to coelute in Ni-NTA pulldown experiments under conditions including both His_6-SUMO-ClpS and ΔNTD-ClpC1. Unlike full-length ClpC1, ΔNTD-ClpC1 was not observed to interact non-specifically with Ni-NTA resin (Figure 9). From this, we conclude that binding of ClpS to a ClpC1 construct lacking an intact N-terminal domain does occur. However, complete ablation of ClpC1:ClpS complex formation is observed when the same experiments are performed with the ΔNTD-ClpC1-D440A/R451A mutant (Figure 4C). Taken together, these data suggest that ClpC1:ClpS complex formation involves interactions between the ClpC1-MD and ClpS.

2.5. Molecular Dynamics Simulations Predict Unique ClpS Interface Involved in Complex Formation

Adaptor protein regulation of Hsp100 proteins commonly involves interactions with the N-terminal domain. Molecular Dynamics simulations were performed to investigate the structural arrangement necessary to drive ClpC1:ClpS assembly via the ClpC1-NTD surface. Figure 10 displays the atomistic details describing the ClpC1-NTD and ClpS binding structure obtained by molecular dynamics simulation. The contact residues in the binding interface were determined based on a separation distance less than
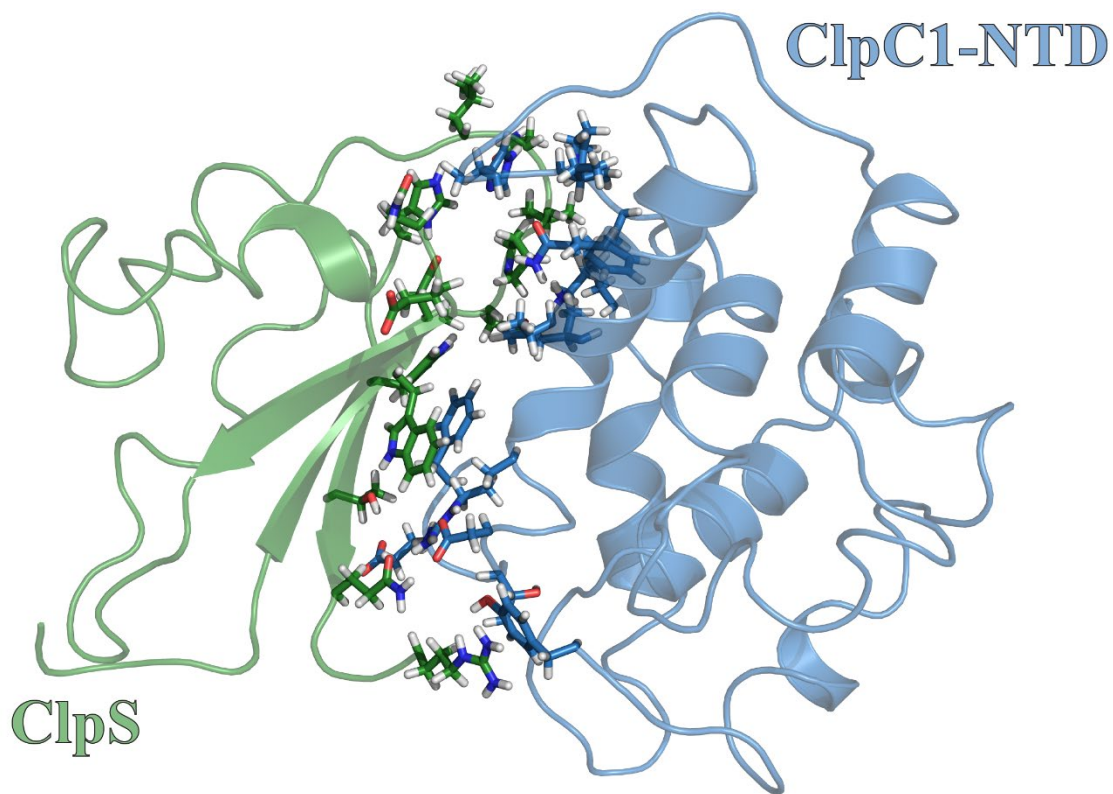
Figure 10 - **Molecular Dynamics simulations predict unique ClpS interface involved in complex formation**. The snapshot of the *M. tuberculosis* ClpC1-NTD:ClpS complex obtained by the Molecular Dynamics simulation. All contact residues are highlight in stick representation. Structures for ClpC1-NTD and ClpS are shown in cartoon representation and colored as blue and green, respectively. All structure representations in Figure 5 were prepared with the Pymol software package.

4.5 Å. Most contact residues were observed to be neutral but some charged residues were also involved in the binding interface. From this, we conclude that the underlying molecular driving forces that function to stabilize the ClpC1-NTD:ClpS interface represent contributions from electrostatic interactions, hydrogen bonding and hydrophobic interactions.

Table 2. List of the contact residues. The distance and interaction energy is in unit of Å and kcal/mol.

ClpC1	ClpS	Energy	Dist
Phe2	Trp33	−4.802	3.95
Phe2	Asp34	−0.118	4.26
Phe2	Trp94	−2.193	3.95
Phe2	Thr96	−0.877	4.19
Glu3	Thr96	−1.604	4.16
Thr6	Arg101	−0.128	4.21
Asp7	Trp33	−0.252	3.87
Asp7	Gln98	−3.051	3.99
Asp7	Arg101	0.018	4.20
Arg10	Trp33	−5.015	3.93
Arg10	Gln98	−1.263	3.52
Val13	Asp35	−0.257	4.34
Val13	Pro36	−1.65	4.09
Val14	Glu68	−0.466	4.35
Val14	Gly69	−0.605	4.02
Gln17	Pro36	1.15	4.19
Gly76	His66	−0.244	3.89
Gly76	Asn67	0.098	3.92
His77	Pro36	1.52	3.94
His77	Val37	0.391	4.21
His77	Asn38	−0.334	4.02
His77	Leu39	0.291	4.29
His77	His66	1.173	3.83
Ile78	Pro36	−0.065	3.94
Pro79	Val37	−0.877	3.71
Pro79	Asn38	0.5	4.47
Phe80	Pro36	−1.774	4.02
Phe80	Val37	−2.005	3.99
Lys85	Asp35	−22.528	3.47
Lys85	Pro36	−7.72	3.62
Lys85	Val37	−1.107	4.02
Tyr145	Arg101	−0.077	3.98

3. Discussion

ClpC1 Binding of ClpS May Involve Multiple Binding Surfaces

Recent Cryo-EM structures for the *S. aureus* ClpC:MecA [15] complex and *Saccharomyces cerevisiae* (*S. cerevisiae*) Hsp104 [88] have implicated the Middle Domain as a molecular switch in HSP100 proteins. Association of *S. aureus* ClpC with MecA induces a structural rearrangement that transitions the complex from an inactive resting state to a fully functional chaperone [77]. This switch is accompanied by a transformation from a helical assembly stabilized by head-to-head intermolecular MD contacts to the planar ring structure associated with active Clp/Hsp100 ATPases. The molecular basis for describing this process centers on the Uvr motif of the MD, a primary sequence motif characterized by a conserved motif, [E/D]φE, similar to the coiled-coil motifs of UvrB and UvrC (φ represents any aromatic residue) [89-90]. Previous data with *B. subtilis* ClpC has demonstrated Uvr motif residue F436 as critical for protein complex assembly and function [31]. The F436A mutation and outright MD deletion in *S. aureus* ClpC yields no detectable stimulation of ATPase activity in the presence of MecA [77]. Cryo-EM structures for *S. aureus* ClpC with and without MecA present reveal the interaction of this MD phenylalanine residue with MecA such that intermolecular MD-MD interactions are disrupted upon introduction of MecA, thereby reconfiguring the complex into an enzymatically active hexamer. However, in the absence of an intact Uvr motif, the inactive helical assembly is no longer stabilized via MD-MD intermolecular interactions and MecA is no longer necessary for ClpC activation.

Cyro-EM structures for *S. cerevisiae* Hsp104 have led to the proposal that the MD adopts two nucleotide-specific conformations that correspond to the hydrolytic state of the

bound nucleotide (Figure 6) [88]. From this, the MD was proposed to stabilize the ATP-state of Hsp104 subunits with bound polypeptide substrate. Superimposition of the X-ray structure for *B. subtilis* ClpC-D1 [31] alongside the Cryo-EM structures for Hsp104 [88] bound to either ATPγS or ADP positions the ClpC-MD in an ATP-like conformation that is approximately orthogonal to the position predicted for the ADP-state (Figure 6). Unlike the Hsp104 structures, however, the model shown for ClpC is derived from a co-crystal structure with MecA bound, which suggests a role for adaptor protein in stabilization of the ATP-state.

Our data presented here suggest the presence of two distinct ClpC1 sites that participate in ClpS interactions. Based on pulldown data and comparisons with homologous Clp/Hsp100 ATPases, those sites are represented by the ClpC1-NTD and
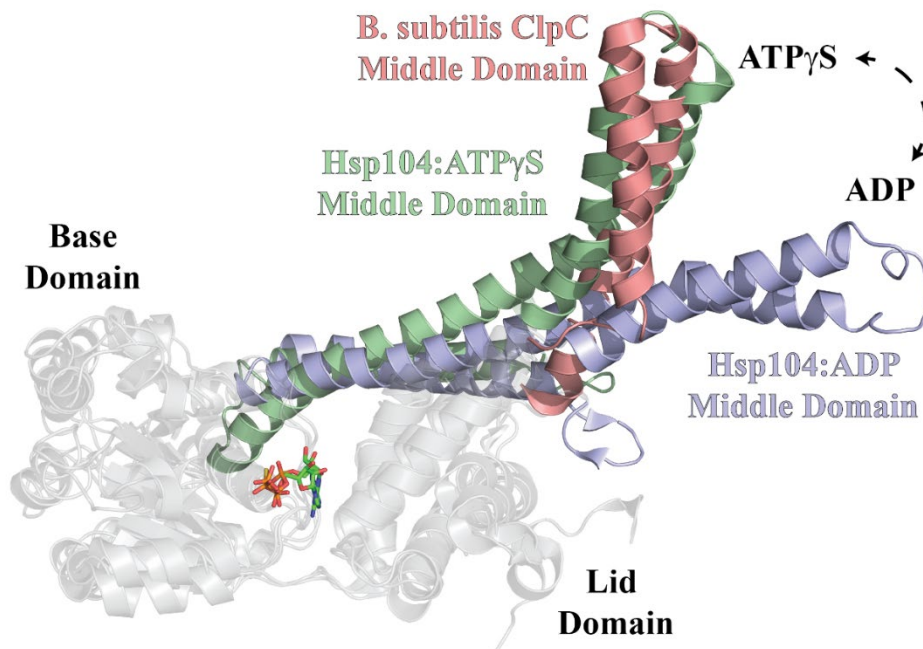

Figure 11 - **Structural Comparisons Among Middle Domain-Containing Hsp100 proteins**. Alignment of X-ray structures for Hsp104 bound to either ATPγS (green) or ADP (purple) alongside B. subtilis ClpC (salmon) reveal the Middle Domain conformation to be nucleotide-dependent. The PDB accession codes used to create Figure 6 were 5VY9, 5VY9 and 3PXG.

ClpC1-MD. Though the ClpC1-NTD is dispensable for ClpC1 catalyzed unfolding of an SsrA-tagged protein, it may contribute to complex stability. However, our molecular dynamics model indicates that the ClpS:ClpC1-NTD interaction may differ substantially from the models reported for the ClpA-NTD:ClpS complex [10,36]. In our model, much of the free energy associated with complex formation involves a single interaction between Lys85 of ClpC1-NTD and Asp158 of ClpS. Though an Asp is conserved at this position in all ClpS proteins, reference to comparable structures for ClpA-NTD:ClpS complexes indicate that this residue is positioned on the ClpS-face opposite to the ClpA-interaction site [72,91]. Furthermore, no interaction pair identified in our ClpC1-NTD:ClpS model would be expected to occur similar to those reported for ClpA. However, our stopped-flow fluorescence data represented in Figures 2 and 3 demonstrate that the ΔNTD-ClpC1 truncation mutant is subject to allosteric inhibition by ClpS, which requires that a physical interaction occur between the two proteins in the absence of an intact ClpC1-NTD. Further mutations to residue positions surrounding the Uvr motif, D440A and R451A, ablate ClpS-dependent inhibition of ClpC1 catalyzed SsrA-tagged protein unfolding. Taken together, our data support a model wherein ClpS interacts directly with the ClpC1-NTD and -MD in a configuration that may be more "MecA-like." Based on the structural data presented in Figure 6, we expect that this would necessarily need to involve the ClpS-dependent stabilization of a ClpC1-MD conformation other than the ATP-state. However, the structural details of these conformational dynamics are currently unclear.

Taken together, our data suggest that ClpS functions as binary regulator of ClpC1 catalyzed protein unfolding (Figure 1C). This statement is based on the observation in Figure 1C of two phases separated by a narrow range of [ClpS], which we term the

Catalytic and partially-Inhibited phases (colored orange and yellow in Figure 1C, respectively). The average apparent unfolding rates are 1.7 ± 0.2 or 0.8 ± 0.1 s^{-1} in the catalytic state or partially-inhibited state, respectively. This contrasts reports involving *E. coli* ClpAP catalyzed degradation of SsrA-tagged GFP in the presence of ClpS, where a continuum of kinetic behaviors are observed such that inhibition of ClpAP catalyzed protein degradation occurs over a broader range of [ClpS] [73].

Though Clp/Hsp100 proteins are highly homologous in primary sequence and overall architecture, regulation of this family of proteins is not necessarily identical across all bacterial species. The data reported in this study illustrate that regulation of mycobacterial ClpC proteins cannot be assumed as functionally equivalent to well-studied Clp/Hsp100 proteins such as *E. coli* ClpA. We demonstrate that ClpS-dependent inhibition of *M. tuberculosis* ClpC1 catalyzed unfolding of SsrA-proteins does not require an intact ClpC1-NTD and involves the ClpC1-MD. This contrasts reported data for *E. coli* ClpA, where the ClpA-NTD is necessary for the observation of ClpS-dependent inhibition of function. Our data qualitatively may suggest a role for the ClpC1-NTD in stabilization of the ClpC1:ClpS complex. This observation is of potential clinical significance since anti-tuberculosis drugs such as cyclomarin A [92-93] and lassomycin [38,39] are expected to bind to the ClpC1-NTD, thereby leading to competition for binding between ClpS and anti-tuberculosis drug. However, future work will be needed to clarify whether ClpS-overexpression in mycobacteria represents a viable resistance mechanism against these novel anti-tuberculosis drugs.

4. Materials and Methods

4.1. Materials

All solutions were prepared with reagent-grade chemicals in double-distilled water produced from a Purelab Ultra Genetic System (Siemens Water Technology, Munich, Germany). All genes were synthesized and each cloned into the pET-24a(+) vector commercially by Genscript (Piscataway, NJ, USA). Plasmids for ClpC1 proteins encode His$_6$-SUMO fusions with full-length ClpC1, ClpC1 truncation mutant lacking N-terminal residues M1-Y145 (ΔNTD-ClpC1), the isolated ClpC1 N-terminal domain (ClpC1-NTD), or ΔNTD-ClpC1 bearing two single-point mutations, D440A/R451A, in the ClpC1-MD (ΔNTD-ClpC1-D440A/R451A).

4.2. Protein Expression and Purification

All expression constructs were prepared as N-terminal His$_6$-SUMO fusions and overexpressed from the pET-24a(+) vector in BL21(DE3) competent cells. Bacterial cultures were initially grown in Lysogeny broth (LB) at 37 °C, followed by induction at OD$_{600}$ = 0.6 absorbance units with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, ThermoFisher Scientific, Waltham, MA, USA) and overnight incubation with shaking at 18 °C. The harvested cell paste was resuspended in chilled lysis buffer containing 50 mM Tris (pH = 8.3), 400 mM NaCl, 10% glycerol, 1 mM 2-mercaptoethanol, 10 mM imidazole (pH = 8) and a protease inhibitor cocktail tablet (ThermoFisher Scientific). The resulting suspension was subjected to sonication and clarified by centrifugation at ~50,000× g. Affinity chromatography was next applied such that the supernatant resulting from the previous centrifugation step was incubated at 4 °C for 2 h with Ni-nitriloacetic acid solid-

phase resin (Ni-NTA, G-Biosciences, St. Louis, MO, USA) previously equilibrated with lysis buffer. Following incubation, the Ni-NTA resin was subjected to centrifugation at ~250× g to isolate solid-phase resin. The supernatant was discarded and the Ni-NTA resin washed with fresh lysis buffer. This wash cycle was repeated five to seven times in order to remove any proteins not associated with the Ni-NTA resin. After the final Ni-NTA wash cycle, all His$_6$-SUMO-fusion proteins were dissociated from the resin by gravity flow using elution buffer containing 50 mM Tris pH = 8.3, 300 mM NaCl, 10% glycerol, 2 mM 2-mercaptoethanol and 500 mM imidazole (pH = 8). Unless otherwise stated, the His$_6$-SUMO tag was removed from all fusion proteins by overnight digestion with His$_6$-tagged Ulp1 protease in elution buffer [94]. Cleaved protein was separated from His$_6$-Ulp1 and any residual uncleaved protein by a second round of Ni-NTA binding. For *M. tuberculosis* His$_6$-SUMO-ClpS, His$_6$-SUMO-ClpC1-NTD and *T. geoffroyi* His$_6$-SUMO-SsrA-Kaede (expressed with C-terminal *M. tuberculosis* SsrA sequence, AADSHQRDYALAA [19]) fusion proteins, the resulting flow-through volume was dialyzed overnight against H200 buffer (25 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (pH = 7.6), 200 mM NaCl, 10 mM MgCl$_2$, 1 mM 2-mercaptoethanol and 10% glycerol), flash-frozen in liquid nitrogen and stored at −80 °C.

Additional purification steps were applied to isolate full-length ClpC1, ΔNTD-ClpC1 and ΔNTD-DR440AA. After removal of the His$_6$-SUMO tag, the resulting cleaved ClpC1 protein solution was diluted with lysis buffer lacking NaCl to a final [NaCl] = 100 mM and loaded onto a Hiprep Q FF 16/10 column (GE Healthcare, Piscataway, NJ, USA) previously equilibrated with 20 mM Tris (pH = 8.3), 10 mM NaCl, 1 mM 2-mercaptoethanol and 10% glycerol. The sample was eluted with a linear gradient from 10

mM NaCl to 1000 mM NaCl over 8 column volumes. Fractions derived from the linear gradient elution were subjected to analysis by SDS-PAGE to confirm the presence of ClpC1 protein and pooled accordingly. Pooled fractions were dialyzed overnight against storage buffer containing (50 mM Tris pH = 8.3, 400 mM NaCl, 5 mM 2-mercaptoethanol and 50% glycerol), flash-frozen in liquid nitrogen and stored at −80 ℃. Prior to storage, purity was judged to be >95% by Coomassie staining (Figure 12). Protein concentrations were determined spectrophotometrically in reaction buffer H200 using extinction coefficients $\varepsilon_{280} = 3.59 \times 10^4 \ M^{-1} \cdot cm^{-1}$, $\varepsilon_{280} = 3.14 \times 10^4 \ M^{-1} \cdot cm^{-1}$, $\varepsilon_{280} = 2.65 \times 104 \ M{-1} \cdot cm{-1}$, $\varepsilon_{280} = 4.47 \times 103 \ M{-1} \cdot cm{-1}$ and $\varepsilon_{280} = 2.74 \times 104 \ M{-1} \cdot cm{-1}$, respectively, for full-length *M. tuberculosis* ClpC1, ΔNTD-ClpC1/ΔNTD-ClpC1-D440A/R451A, *M. tuberculosis* ClpS, ClpC1-NTD and *T. geoffroyi* SsrA-Kaede. The concentration of His$_6$-SUMO-ClpS was determined spectrophotometrically in H200 buffer described above using an extinction coefficient of $\varepsilon_{280} = 2.79 \times 10^4 \ M^{-1} \cdot cm^{-1}$. *T. geoffroyi* SsrA-Kaede was subjected to photoactivation prior to storage using methods previously described [83–84].
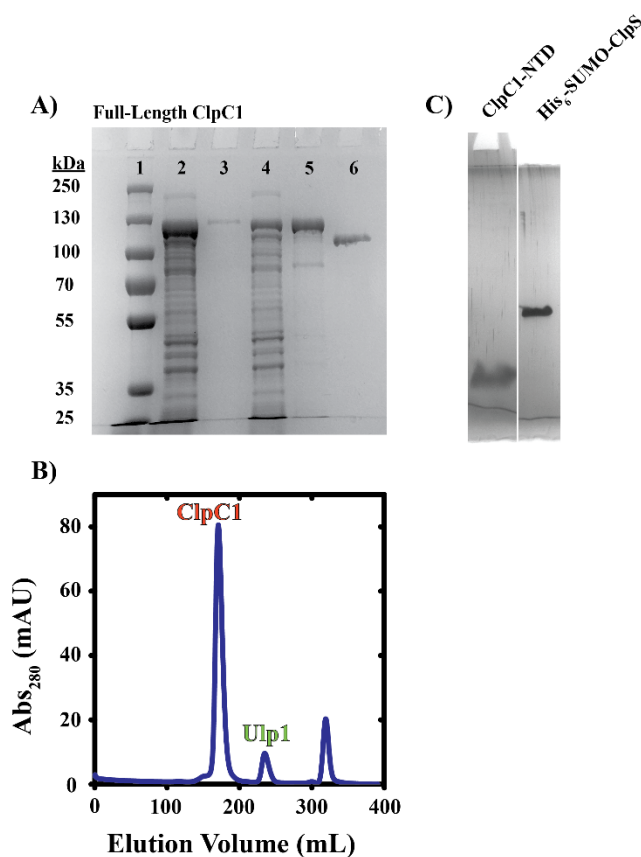
Figure 12 - *M. tuberculosis* **ClpC1 Purification**. A) Active ClpC1 can be isolated to greater than 95% purity as judged by Coomassie staining. Lanes on a 10 % acrylamide gel were loaded as follows: 1 – molecular weight standard, 2 – soluble lysate, 3 – Supernatant from Ni-NTA wash, 4 – Ni-NTA flow-through solution, 5 – Ni-NTA elution, and 6 – pooled fractions after His6-Ulp1 cleavage and ion exchange chromatography. B) Size-exclusion chromatography analysis has been performed as a final purification step in place of ion exchange chromatography. Fractions containing ClpC1 elute at a volume that approximately corresponds to the mass of a dimer.

4.3. Methods

4.3.1. Stopped-Flow Fluorescence Assay

Stopped-flow fluorescence experiments were performed using an Applied Photophysics SX.20 stopped-flow fluorometer (Letherhead, UK). All reactions were performed at 37 ˚C in buffer H200 (25 mM HEPES (pH = 7.6), 200 mM NaCl, 10 mM MgCl$_2$, 1 mM 2-mercaptoethanol and 10% (*v/v*) glycerol). Syringe A contained 1 μM
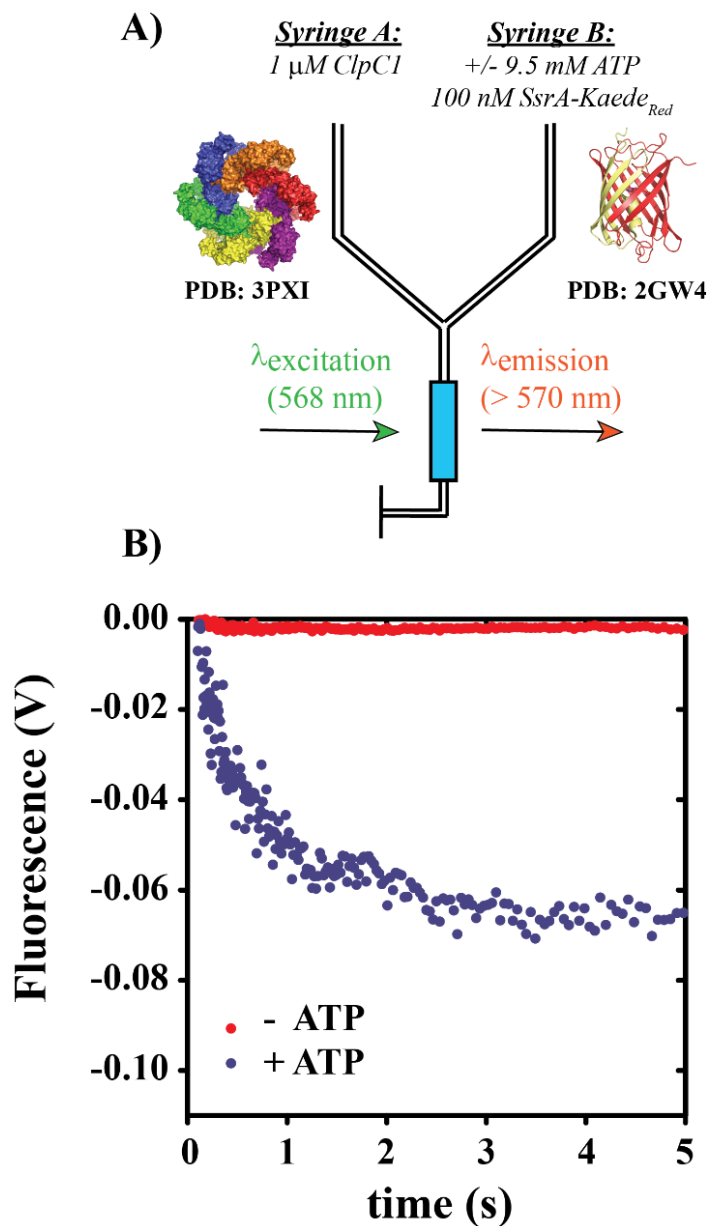
Figure 13 - **ClpC1 catalyzed protein unfolding requires ATP**. A) Schematic representation of stopped-flow fluorescence protein unfolding experiments. Syringe A contains 1 µM ClpC1. Syringe B contains 9.5 mM ATP to fuel protein unfolding and 100 nM photoactivated Kaede bearing a C-terminal SsrA-degradation tag (SsrA-KaedeRed). Fluorescence is observed using an excitation wavelength equal to 568 nm and emissions are observed above 570 nm with a 570-nm-long pass filter. B) Representative fluorescence timecourses for ClpC1 catalyzed SsrA-KaedeRed unfolding. Time courses were measured in the presence (Blue Circles) or absence (Red Circles) of 9.5 mM ATP.

ClpC1 and varied initial concentrations of ClpS from 0 to 3 μM. Syringe B contained 100 nM of photoactivated Kaede (SsrA-Kaede$_{Red}$) bearing a C-terminal *M. tuberculosis* SsrA degradation tag and 9.5 mM ATP. The observation of ClpC1 catalyzed protein unfolding requires the inclusion of ATP in Syringe B (Figure 13). Prior to mixing, both solutions were incubated for 15 min at 37 °C in the stopped-flow instrument to establish thermal equilibrium. Additional incubation of either solution had no effect on the observed fluorescence time courses. SsrA-Kaede$_{Red}$ was excited at λ_{ex} = 568 nm and emissions were observed above 570 nm using a 570-nm-long pass filter. All kinetic traces shown represent the average of at least seven individual determinations. Averaged time courses were subjected to non-linear least squares (NLLS) analysis using a single exponential function. The dependence of the apparent unfolding rate constant, $k_{UF,app}$, on the final mixing concentration of ClpS, $[ClpS]_F$, was subjected to NLLS analysis using a modified form of the Hill equation:

Equation (1)

$$k_{UF,\ app} = 1 - \left(k_{max} \cdot \frac{(K_{app}[ClpS]_F)^n}{1 + (K_{app}[ClpS]_F)^n} + b\right) \tag{1}$$

where k$_{max}$ represents the maximum apparent unfolding rate constant, K$_{app}$ approximates the association equilibrium constant, *n* is the Hill coefficient and *b* is the *y*-intercept term.

4.3.2. Ni-NTA Pulldown Experiments

Ni-NTA affinity pulldown experiments were performed by incubating 2 μM His$_6$-SUMO-ClpS with varied concentrations of ClpC1 and 1 mM ATPγS (as indicated) at 25 °C for 30 min to promote complex formation. All experiments were performed in 150 μL reaction volumes in buffer H200 (25 mM HEPES (pH = 7.6), 200 mM NaCl, 10 mM

MgCl$_2$, 1 mM 2-mercaptoethanol and 10% glycerol). After initial incubation, 100 µL of Ni-NTA slurry, in H200 buffer supplemented with 1 mM ATPγS and 10 mM imidazole (pH = 8), was added to each reaction solution, followed by incubation with agitation at 25 °C for 60 min to promote Ni-NTA binding by His$_6$-SUMO-ClpS. After 60 min, each reaction mixture was transferred to an empty polypropylene column (Bio-Rad Laboratories, Hercules, CA, USA), where solid resin was isolated from buffer by gravity flow. The isolated Ni-NTA resin was washed 3 times with an excess volume of H200 supplemented with 1 mM ATPγS and 10 mM imidazole (pH = 8). His$_6$-SUMO-ClpS was eluted from the Ni-NTA resin by the addition of 200 µL of H200 supplemented with 1 mM ATPγS and 250 mM imidazole (pH = 8). The resulting elution samples were analyzed by SDS-PAGE using either Coomassie- or silver-staining methods to visual gel bands.

4.3.3. Structure Preparation

The PDB file for the crystal structure of ClpC1 [92] was obtained from the Protein Data Bank [95] (PDB ID: 3WDB) and included an expression tag at N-terminus. The three-dimensional structure of *M. tuberculosis* ClpS was obtained from homology modeling with the SWISS-Model Server [96] using the crystal structure for ClpS from *E. coli* [97] from the protein databank (PDB ID: 2WA9) as a template. The 3D structures for ClpC1-NTD and ClpS from *M. tuberculosis* were prepared for docking using the MOE2018 [98] software suite (2018, Montreal, QC, Canada). MOE's Protonate 3D [99] utility was used to add the appropriate amount of hydrogens to each structure at a pH of 7.5, salt concentration of 0.15M and temperature of 310 K. The protonated structures were energy minimized using the AMBER14:EHT [100-101] force field.

4.3.4. Docking Process

The prepared structures of *M. tuberculosis* ClpC1-NTD and ClpS were submitted to the ClusPro 3.0 [102] server for rigid body docking calculations. The structure for ClpC1 was defined as the receptor and that of *M. tuberculosis* ClpS was defined as the ligand. No residues were designated to be attractive or repulsive.

4.3.5. Molecular Dynamics

Molecular dynamics simulations were performed using the NAMD2 platform [103]. The initial structure was from the docking result with the lowest binding energy. Explicit solvent was employed using the TIP3P water model [104]. The NPT ensemble was used with a constant temperature of 310 K. The damping coefficient used in the simulations was set as 5 ps. The MD pressure was set as 1 atm and kept as constant with Langevin piston method [105,106]. In the simulations, no constraint was applied to any atomic coordinates. To include long-range electrostatic interactions in the simulations, the particle-mesh Ewald (PME) method was used with a 1-angstrom grid width [107]. The nonbonded interactions were evaluated every 10 time-steps using a group-based cutoff with a switching function. The SHAKE algorithm [108] was employed to fix the covalent bonds involving hydrogen in the simulations [108]. The systems were equilibrated for 20 ns, followed by another 20 ns molecular dynamics runs. The last snapshot was selected used for the binding analysis.

CHAPTER IV: Conclusions and Future Directions

*Mtb*ClpC1 regulation by *Mtb*ClpS deviates from the previously characterized EcClpS regulation of EcClpA in both functional impact and residue interaction, instead favoring an interaction more akin to MecA regulation of *Bs*ClpC. The individualism of each of these homologous proteins suggests the need for species' specific characterization of Hsp100/Clp proteins. Additionally, the emergence of drug-resistant strains of *Mycobacterium tuberculosis* highlights a need for novel and species-specific drug development. The Hsp100 family of proteins are required for proteostatic control within the cell and thus exhibit high degrees of sequence conservation. They are therefore being increasingly recognized as desirable Mtb targets. Phylogenetic analysis identifies the possibility of functional differences of ClpC1 from closely related Hsp100 ATPases, this work provides valuable insight into the structural and functional identity of *Mtb*ClpC1.

Our data presented herein suggest the presence of two distinct ClpC1 sites that participate in ClpS interactions. Based on pulldown data and comparisons with homologous Clp/Hsp100 ATPases, those sites are represented by the ClpC1-NTD and ClpC1-MD. Additionally, our work determines that while the N-terminal domain of ClpC1 is dispensable for catalyzed unfolding of a SsrA-tagged protein, it may be implicated in complex stability. From examination of MD simulations, we predict the ClpC1:ClpS complex to resemble the ClpC:MecA complex, as opposed to the ClpS:ClpA complex [35, 72]. Analyses of these simulations indicates the ClpC1:ClpS complex is centered on a single interaction between Lys85 of ClpC1-NTD and Asp158 of ClpS. Additional mutations to residues surrounding the Uvr motif, D440A and R451A, prevent ClpS-

dependent inhibition of ClpC1 catalyzed SsrA-tagged protein unfolding. From this, we suggest a model wherein ClpS interacts directly with the ClpC1-NTD and -MD. Taken together, all of our data lead to a conclusion that although proteins may share great levels of homogeneity in primary sequence, and even folding, mechanistic details are not consistent across bacterial species.

This observation is of potential clinical significance since recent anti-tuberculosis drugs (cyclomarin A [92-93] and lassomycin [26,109]) are purported to target the ClpC1-NTD. Considering our findings, and the shared binding site between anti-TB drugs and the adaptor protein ClpS, we hypothesize that ClpS might be a competitor to these anti-TB drugs. Work from Gavrish and colleagues first identified lassomycin as binding to the N-terminal domain of ClpC1 and uncoupling the ATPase activity of ClpC1 and subsequently the proteolytic activity of ClpP [26]. This is reported as being accomplished by stimulation of ATPase activity such that proteolysis by ClpP is prevented [110]. Gavrish et al identified the following as interacting with lassomycin: Proline-79 (P79), Gultamine-17 (Q17), and Arginine-21 (R21) as their acidic nature complimented the basic charge of lassomycin. However, while they report on the in vivo effectiveness of lassomycin in cell cultures, the *in silico* docking of the peptide does not take an adaptor protein into account. When visualizing the residues reported from their *in silico* methods, they were obscured by ClpS in the complex from our own Docking data. Figure 12 is a visualization of these residues (shown in red) alongside those identified in our own work (shown in purple and blue). We propose that the combination of these two bodies of work in future projects would provide closure to knowledge gaps regarding how lassomycin affects ClpC1 and how ClpS might act as an immunity or in some other fashion to this bactericidal peptide. For instance,

completing complementary docking with both ClpS and lassomycin alongside fluorescent based in vitro efforts with both adaptor and peptide present would provide a clearer understanding towards possible in vivo mechanics of ClpC1.

Perhaps a larger obstacle to implementation of anti-tubercular treatment targeting ClpC1 would be mitochondrial ClpX. The Class II Hsp100 ATPase is present in human mitochondria and is structurally and functionally similar to ClpA, although lacking a second ATPase domain. However, since ClpS and homologs have been reported as allosteric regulators of ClpA and ClpC like enzymes, further study of ClpS impact on such proteins would be required. In addition to the possible regulation of ClpX by ClpS,
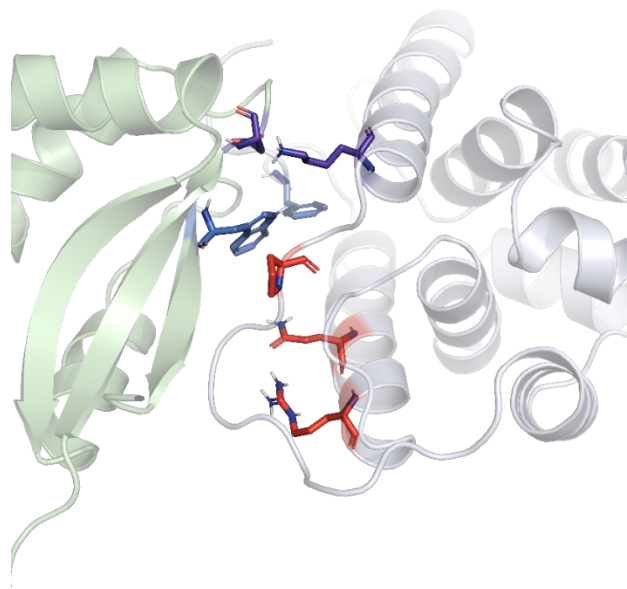
Figure 14 – ClpS competition with lassomycin? ClpC1-NTD (green) in complex with ClpS (gray). Highlighted are the residues driving complex formation, as identified by computational docking mentioned in chapter III. In blue are Phe2 and Lys85 of ClpC1 that interact with Trp94 and Asp35 of ClpS, respectively. In red are the three residues identified by Gavrish and colleagues that are likely binding sites for Lassomycin interaction with ClpC1. The image was prepared using the PyMol software package.

structural homogeneity within the NTDs of each unfoldase might create unwanted impact upon Human ClpXP by antibiotic targeting ClpC. Therefore, it would be necessary to

create a similar study centered on lassomycin impact on ClpX as well as the search for an allosteric regulator of mitochondrial ClpX. Ultimately, this novel reporting of allosteric regulation of ClpC1 confirms the earlier hypothesis that all AAA+ proteins cannot be assumed to function identically.

References

1. Konovalova, A., Sogaard-Andersen, L., Kroos, L., 2014. Regulated proteolysis in bacterial development. FEMS Microbiol. Rev. 38, 493–522.

2. Bartlett A, Radford S. 2009. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms.

3. Powers, E.T.; Morimoto, R.I.; Dillin, A.; Kelly, J.W.; Balch, W.E. "Biological and Chemical Approaches to Diseases of Proteostasis Deficiency". *Annu. Rev. Biochem.* **78**: 959–91 (2009).

4. Balch WE, Morimoto RI, Dillin A, Kelly JW (Feb 2008). "Adapting proteostasis for disease intervention". *Science*. **319** (5865): 916–919 (2008)

5. Hartl F, et al. 2011. Molecular chaperones in protein folding and proteostasis. (2011)

6. Morimoto R. I. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes & development*, *22*(11), 1427–1438. https://doi.org/10.1101/gad.1657108 (2008)

7. Ventura, M., Fitzgerald, G. F., & van Sinderen, D. Genetic and transcriptional organization of the clpC locus in Bifidobacterium breve UCC 2003. *Applied and environmental microbiology*, *71*(10), 6282–6291. https://doi.org/10.1128/AEM.71.10.6282-6291.2005 (2005)

8. Andréasson, C., Ott, M., & Büttner, S. (2019). Mitochondria orchestrate proteostatic and metabolic stress responses. *EMBO reports*, *20*(10), e47865. https://doi.org/10.15252/embr.201947865

9. Gottesman, S.; Maurizi, M.R. Regulation by proteolysis: Energy-dependent proteases and their targets. *Microbiol. Rev.* **1992**, *56*, 592–621. [PubMed]

10. Claudio A.P. Joazeiro. Ribosomal Stalling During Translation: Providing Substrates for Ribosome-Associated Protein Quality Control. Annual Review of Cell and Developmental Biology 2017 33:1, 343-368. (2017)

11. McGinness, K. E., Baker, T. A. & Sauer, R. T. Engineering controllable protein degradation. *Molecular cell* **22**, 701707, doi:10.1016/j.molcel.2006.04.027 (2006).

12. Samali A, Cai J, Zhivotovsky B, Jones DP, Orrenius S. "Presence of a pre-apoptotic complex of pro-caspase-3, Hsp60 and Hsp10 in the mitochondrial fraction of jurkat cells". *EMBO J*. **18** (8): 2040–8 (1999)

13. Lee KH, Kim HS, Jeong HS, Lee YS (October 2002). "Chaperonin GroESL mediates the protein folding of human liver mitochondrial aldehyde dehydrogenase in Escherichia coli". *Biochem. Biophys. Res. Commun*. **298** (2): 216–24 (2002)

14. Gottesman, S.; Wickner, S.; Maurizi, M.R. Protein quality control: Triage by chaperones and proteases. *Genes Dev*. **1997**, *11*, 815–823. [CrossRef] [PubMed]

15. Sauer, R.T.; Baker, T.A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **2011**, *80*, 587–612. [CrossRef] [PubMed]

16. Olivares, A.O.; Baker, T.A.; Sauer, R.T. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat. Rev. Microbiol.* **2016**, *14*, 33–44. [CrossRef] [PubMed]

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

18. Horwich, A.L., Weber-Ban, E.U., Finley, D., 1999. Chaperone rings in protein folding and degradation. Proc. Natl. Acad. Sci. U.S.A. 96, 11033–11040.

19. Matyskiela, M.E., Martin, A., 2013. Design principles of a universal protein degradation machine. J. Mol. Biol. 425, 199–213.

20. Alhuwaider, A.A.H.; Dougan, D.A. AAA+ machines of protein destruction in mycobacteria. *Front. Mol. Biosci.* **2017**, *4*, 49. [CrossRef] [PubMed]

21. Hanson, P.I., Whiteheart, S.W., 2005. AAA+ proteins: have engine, will work. Nat. Rev. Mol. Cell Biol. 6, 519–529.

22. *Centers for Disease Control and Prevention*. "ANTIBIOTIC RESISTANCE THREATSIN THE UNITED STATE.", www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf. (2019)

23. K. H. Michel, R. E. Kastner (Eli Lilly and Company), US 4492650, 1985 [Chem. Abstr, 102, 130459]. (1985)

24. Osada, Hiroyuki; Yano, Tatsuya; Koshino, Hiroyuki; Isono, Kiyoshi (1991). "Enopeptin A, a novel depsipeptide antibiotic with anti-bacteriophage activity". The Journal of Antibiotics. **44** (1991)

25. Carney, Daniel W.; Schmitz, Karl R.; Truong, Jonathan V.; Sauer, Robert T.; Sello, Jason K. "Restriction of the Conformational Dynamics of the Cyclic Acyldepsipeptide Antibiotics Improves Their Antibacterial Activity". JACS. **136**: 1922–1929 (2014)

26. Hinzen, Berthold; Labischinski, Harald; Brötz-Oesterhelt, Heike; Endermann, Rainer; Benet-Buchholz, Jordi; Hellwig, Veronica; Häbich, Dieter; Schumacher, Andreas; Lampe, Thomas; Paulsen, Holger; Raddatz, Siegfried (2006). "Medicinal Chemistry Optimization of Acyldepsipeptides of the Enopeptin Class Antibiotics". ChemMedChem. **1** (7): 689–693. (2006)

27. Gavrish, E.; Sit, C.S.; Cao, S.; Kandror, O.; Spoering, A.; Peoples, A.; Ling, L.; Fetterman, A.; Hughes, D.; Bissell, A.; et al. Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2. *Chem. Biol.* **2014**, *21*, 509–518. [CrossRef] [PubMed]

28. Erzberger, J.P., Berger, J.M., 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu. Rev. Biophys. Biomol. Struct. 35, 93–114.

29. Keen EC. "Phage therapy: concept to cure". *Frontiers in Microbiology*. **3**: 238. (2012)

30. Schlothauer, T.; Mogk, A.; Dougan, D.A.; Bukau, B.; Turgay, K. MecA, an adaptor protein necessary for ClpC chaperone activity. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2306–2311. [CrossRef] [PubMed]

31. Kirstein, J., Moliere, N., Dougan, D.A., Turgay, K., 2009. Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. Nat. Rev. Microbiol. 7, 589–599.

32. Wang, F.; Mei, Z.; Qi, Y.; Yan, C.; Hu, Q.; Wang, J.; Shi, Y. Structure and mechanism of the hexameric MecA-ClpC molecular machine. *Nature* **2011**, *471*, 331–335. [CrossRef] [PubMed]

33. Liu, J., Mei, Z., Li, N., Qi, Y., Xu, Y., Shi, Y., Wang, F., Lei, J., Gao, N., 2013. Structural dynamics of the MecA-ClpC complex: a type II AAA+ protein unfolding machine. J. Biol. Chem. 288, 17597–17608.

34. 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. *J. Biol. Chem.* **2002**, *277*, 46753–46762.

35. 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. *J. Struct. Biol.* **2004**, *146*, 166–179.

36. Kress, W., Maglica, Z., Weber-Ban, E., 2009a. Clp chaperone-proteases: structure and function. Res. Microbiol. 160, 618–628.

37. Kress, W., Mutschler, H., Weber-Ban, E., 2009b. Both ATPase domains of ClpA are critical for processing of stable protein structures. J. Biol. Chem. 284, 31441–31452.

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

39. Rose, P.W.; Prlic´, A.; Altunkaya, A.; Bi, C.; Bradley, A.R.; Christie, C.H.; Costanzo, L.D.; Duarte, J.M.; Dutta, S.; Feng, Z.; et al. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2017**, *45*, D271–D281.

40. Miller, J.M., Enemark, E.J., 2016. Fundamental characteristics of AAA+ protein family structure and function. Archaea 2016, 9294307.

41. Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J., 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.

42. Kar, N.P.; Sikriwal, D.; Rath, P.; Choudhary, R.K.; Batra, J.K. Mycobacterium tuberculosis ClpC1: Characterization and role of the N-terminal domain in its function. *FEBS J.* **2008**, *275*, 6149–6158. [CrossRef] [PubMed]

43. Singh, S.K., Rozycki, J., Ortega, J., Ishikawa, T., Lo, J., Steven, A.C., Maurizi, M.R., 2001. Functional domains of the ClpA and ClpX molecular chaperones identified by limited proteolysis and deletion analysis. J. Biol. Chem. 276, 29420–29429.

44. Enemark, E.J., Joshua-Tor, L., 2006. Mechanism of DNA translocation in a replicative hexameric helicase. Nature 442, 270–275.

45. Glynn, S.E., Martin, A., Nager, A.R., Baker, T.A., Sauer, R.T., Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. Cell 139, 744–756. (2009)

46. Miller, J.M., Arachea, B.T., Epling, L.B., Enemark, E.J., 2014. Analysis of the crystal structure of an active MCM hexamer. Elife 3, e03433.

47. Caillat, C., Macheboeuf, P., Wu, Y., McCarthy, A.A., Boeri-Erba, E., Effantin, G., Gottlinger, H.G., Weissenhorn, W., Renesto, P., 2015. Asymmetric ring structure of Vps4 required for ESCRT-III disassembly. Nat. Commun. 6, 8781.

48. Park, S.S., Kwon, H.Y., Tran, T.D., Choi, M.H., Jung, S.H., Lee, S., Briles, D.E., Rhee, D.K., 2015. ClpL is a chaperone without auxiliary factors. FEBS J. 282, 1352–1367.

49. Tao, L., Biswas, I., 2013. ClpL is required for folding of CtsR in Streptococcus mutans. J. Bacteriol. 195, 576–584.

50. Li, T., Weaver, C.L., Lin, J., Duran, E.C., Miller, J.M., Lucius, A.L., 2015. Escherichia coli ClpB is a non-processive polypeptide translocase. Biochem. J. 470, 39–52.

51. Tessarz, P., Mogk, A., Bukau, B., 2008. Substrate threading through the central pore of the Hsp104 chaperone as a common mechanism for protein disaggregation and prion propagation. Mol. Microbiol. 68, 87–97.

52. Weibezahn, J., Tessarz, P., Schlieker, C., Zahn, R., Maglica, Z., Lee, S., Zentgraf, H., Weber-Ban, E.U., Dougan, D.A., Tsai, F.T., Mogk, A., Bukau, B., 2004. Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. Cell 119, 653–665.

53. Miller, J.M., Lin, J., Li, T., Lucius, A.L., 2013. E. coli ClpA catalyzed polypeptide translocation is allosterically controlled by the protease ClpP. J. Mol. Biol. 425, 2795–2812.

54. Schmitz, K.R.; Sauer, R.T. Substrate delivery by the AAA+ ClpX and ClpC1 unfoldases activates the mycobacterial ClpP1P2 peptidase. *Mol. Microbiol.* **2014**, *93*, 617–628. [CrossRef] [PubMed]

55. Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221–271.

56. Lake, J.A., Skophammer, R.G., Herbold, C.W., Servin, J.A., Genome beginnings: rooting the tree of life. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 2177–2185. (2009)

57. Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T., Manuel, M., Worheide, G., Baurain, D., Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602. (2011)

58. Philippe, H., Laurent, J., How good are deep phylogenetic trees? Curr. Opin. Genet. Dev. 8, 616–623. (1998)

59. Felsenstein, J.,. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410. (1978)

60. Bojer, M.S., Struve, C., Ingmer, H., Hansen, D.S., Krogfelt, K.A., Heat resistance mediated by a new plasmid encoded Clp ATPase, ClpK, as a possible novel mechanism for nosocomial persistence of Klebsiella pneumoniae. *PLoS One* **5**, e15467 (2010)

61. Bojer, M.S., Struve, C., Ingmer, H., Krogfelt, K.A., 2013. ClpP-dependent and -independent activities encoded by the polycistronic clpK-encoding locus contribute to heat shock survival in Klebsiella pneumoniae. Res. Microbiol. **164**, 205–210 (2013)

62. Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. (2016)

63. Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts. (2002)

64. Huelsenbeck, J.P., Ronquist, F., MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755. (2001)

65. Ronquist, F., Huelsenbeck, J.P., MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574. (2003)

66. Maddison, W.P.a.D.R.M., Mesquite: a modular system for evolutionary analysis Version **3** (2017)

67. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R.,Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., Obradovic, Z.,. Intrinsically disordered protein. J. Mol. Graph. Model. 19, (2001)

68. Garner, E., Romero, P., Dunker, A.K., Brown, C., Obradovic, Z., 1999. Predicting Binding Regions within Disordered Proteins. Genome Inform. Ser. Workshop Genome Inform. 10, 41–50. (1999)

69. Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., Dunker, A.K., 2001. Sequence complexity of disordered protein. Proteins 42, 38–48. (2001)

70. Karzai, A.W.; Roche, E.D.; Sauer, R.T. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat. Struct. Biol.* **2000**, *7*, 449–455. [CrossRef] [PubMed]

71. Li, T.; Lucius, A.L. Examination of the polypeptide substrate specificity for Escherichia coli ClpA. *Biochemistry* **2013**, *52*, 4941–4954. [CrossRef] [PubMed]

72. Dougan, D.A.; Reid, B.G.; Horwich, A.L.; Bukau, B. ClpS, a substrate modulator of the ClpAP machine. *Mol. Cell* **2002**, *9*, 673–683. [CrossRef]

73. Zeth, K.; Ravelli, R.B.; Paal, K.; Cusack, S.; Bukau, B.; Dougan, D.A. Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat. Struct. Biol.* **2002**, *9*, 906–911. [CrossRef] [PubMed]

74. Hou, J.Y.; Sauer, R.T.; Baker, T.A. Distinct structural elements of the adaptor ClpS are required for regulating degradation by ClpAP. *Nat. Struct. Mol. Biol.* **2008**, *15*, 288–294. [CrossRef] [PubMed]

75. 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. *J. Biol. Chem.* **2010**, *285*, 8771–8781. [CrossRef] [PubMed]

76. Kirstein, J.; Schlothauer, T.; Dougan, D.A.; Lilie, H.; Tischendorf, G.; Mogk, A.; Bukau, B.; Turgay, K. Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. *EMBO J.* **2006**, *25*, 1481–1491. [CrossRef] [PubMed]

77. Carroni, M.; Franke, K.B.; Maurer, M.; Jäger, J.; Hantke, I.; Gloge, F.; Linder, D.; Gremer, S.; Turgay, K.; Bukau, B.; et al. Regulatory coiled-coil domains promote head-to-head assemblies of AAA+ chaperones essential for tunable activity control. *Elife* **2017**, *6*. [CrossRef] [PubMed]

78. Trentini, D.B.; Suskiewicz, M.J.; Heuck, A.; Kurzbauer, R.; Deszcz, L.; Mechtler, K.; Clausen, T. Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature* **2016**, *539*, 48–53. [CrossRef] [PubMed]

79. Tryggvesson, A.; Stahlberg, F.M.; Mogk, A.; Zeth, K.; Clarke, A.K. Interaction specificity between the chaperone and proteolytic components of the cyanobacterial Clp protease. *Biochem. J.* **2012**, *446*, 311–320. [CrossRef] [PubMed]

80. Tryggvesson, A.; Ståhlberg, F.M.; Töpel, M.; Tanabe, N.; Mogk, A.; Clarke, A.K. Characterization of ClpS2, an essential adaptor protein for the cyanobacterium Synechococcus elongatus. *FEBS Lett.* **2015**, *589*, 4039–4046. [CrossRef] [PubMed]

81. Leodolter, J.; Warweg, J.; Weber-Ban, E. The Mycobacterium tuberculosis ClpP1P2 Protease Interacts Asymmetrically with Its ATPase Partners ClpX and ClpC1. *PLoS ONE* **2015**, *10*, e0125345. [CrossRef] [PubMed]

82. Miller, J.M.; Chaudhary, H.; Marsee, J.D. Phylogenetic analysis predicts structural divergence for proteobacterial ClpC proteins. *J. Struct. Biol.* **2018**, *201*, 52–62. [CrossRef] [PubMed]

83. Hayashi, I.; Mizuno, H.; Tong, K.I.; Furuta, T.; Tanaka, F.; Yoshimura, M.; Miyawaki, A.; Ikura, M. Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein Kaede. *J. Mol. Biol.* **2007**, *372*, 918–926. [CrossRef] [PubMed]

84. Glynn, S.E.; Nager, A.R.; Baker, T.A.; Sauer, R.T. Dynamic and static components power unfolding in topologically closed rings of a AAA+ proteolytic machine. *Nat. Struct. Mol. Biol.* **2012**, *19*, 616–622. [CrossRef] [PubMed]

85. Rajendar, B.; Lucius, A.L. Molecular mechanism of polypeptide translocation catalyzed by the Escherichia coli ClpA protein translocase. *J. Mol. Biol.* **2010**, *399*, 665–679. [CrossRef] [PubMed]

86. Lucius, A.L.; Miller, J.M.; Rajendar, B. Application of the sequential n-step kinetic mechanism to polypeptide translocases. *Methods Enzymol.* **2011**, *488*, 239–264. [CrossRef] [PubMed]

87. Van Holde, K.E.; Johnson, W.C.; Ho, P.S. *Principles of Physical Biochemistry*; Prentice Hall: Upper Saddle River, NJ, USA, 1998; p. 657.

88. Gates, S.N.; Yokom, A.L.; Lin, J.; Jackrel, M.E.; Rizo, A.N.; Kendsersky, N.M.; Buell, C.E.; Sweeny, E.A.; Mack, K.L.; Chuang, E.; et al. Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science* **2017**, *357*, 273–279. [CrossRef] [PubMed]

89. Moolenaar, G.F.; Franken, K.L.; Dijkstra, D.M.; Thomas-Oates, J.E.; Visse, R.; van de Putte, P.; Goosen, N. The C-terminal region of the UvrB protein of Escherichia coli contains an important determinant for UvrC binding to the preincision complex but not the catalytic site for 3'-incision. *J. Biol. Chem.* **1995**, *270*, 30508–30515. [CrossRef] [PubMed]

90. Moolenaar, G.F.; Franken, K.L.; van de Putte, P.; Goosen, N. Function of the homologous regions of the Escherichia coli DNA excision repair proteins UvrB and UvrC in stabilization of the UvrBC-DNA complex and in 3'-incision. *Mutat. Res.* **1997**, *385*, 195–203. [CrossRef]

91. Heuck, A., Schitter-Sollner, S., Suskiewicz, M.J., Kurzbauer, R., Kley, J., Schleiffer, A., Rombaut, P., Herzog, F., Clausen, T.. Structural basis for the disaggregase activity and regulation of Hsp104. Elife 5. (2016)

92. Vasudevan, D.; Rao, S.P.; Noble, C.G. Structural basis of mycobacterial inhibition by cyclomarin A. *J. Biol. Chem.* **2013**, *288*, 30883–30891. [CrossRef] [PubMed]

93. Schmitt, E.K.; Riwanto, M.; Sambandamurthy, V.; Roggo, S.; Miault, C.; Zwingelstein, C.; Krastel, P.; Noble, C.;
Beer, D.; Rao, S.P.; et al. The natural product cyclomarin kills Mycobacterium tuberculosis by targeting the ClpC1 subunit of the caseinolytic protease. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 5889–5891. [CrossRef] [PubMed]

94. Bienert, S.; Waterhouse, A.; de Beer, T.A.P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res.* **2017**, *45*, D313–D319. [CrossRef] [PubMed]

95. Rose, P.W.; Prlic´, A.; Altunkaya, A.; Bi, C.; Bradley, A.R.; Christie, C.H.; Costanzo, L.D.; Duarte, J.M.; Dutta, S.; Feng, Z.; et al. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2017**, *45*, D271–D281. [CrossRef] [PubMed]

96. Molecular Operating Environment (MOE). Available online: https://www.chemcomp.com/MOEMolecular_Operating_Environment.htm (accessed on 13 November 2018).

97. Schuenemann, V.J.; Kralik, S.M.; Albrecht, R.; Spall, S.K.; Truscott, K.N.; Dougan, D.A.; Zeth, K. Structural basis of N-end rule substrate recognition in Escherichia coli by the ClpAP adaptor protein ClpS. *EMBO Rep.* **2009**, *10*, 508–514. [CrossRef] [PubMed]

98. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef] [PubMed]

99. Protonate 3D: Assignment of Macromolecular Protonation State and Geometry. Available online: https: //www.chemcomp.com/journal/proton.htm (accessed on 13 November 2018).

100. Kozakov, D.; Hall, D.R.; Xia, B.; Porter, K.A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein-protein docking. *Nat. Protoc.* **2017**, *12*, 255–278. [CrossRef] [PubMed]

101. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. [CrossRef] [PubMed]

102. Lamoureux, G.H.E.; Vorobyov, I.V.; Roux, B.; MacKerell, A.D. A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245–249. [CrossRef]

103. Kalé, L.S.R.; Bh, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312. [CrossRef]

104. Feller, S.E.Z.; Zhang, Y.; Pastor, R.W.; Brooks, B.R. Constant-pressure molecular-dynamics simulation: The langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613. [CrossRef]

105. Martyna, G.J.; Douglas, J.T.; Klein, M.L. Constant-pressure molecular-dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189. [CrossRef]

106. Andersen, H.C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. J.. *Comput. Phys.* **1983**, *52*, 24–34. [CrossRef]

107. Darden, T.; York, D.; Pedersen, L. Particle mesh ewald: An N.Log(N) method for ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]

108. Hartl F, et al. 2011. Molecular chaperones in protein folding and proteostasis.

109. Mossessova, E.; Lima, C.D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **2000**, *5*, 865–876. [CrossRef]

110. E. Gavrish, C.S. Sit, S. Cao, O. Kandror, A. Spoering, A. Peoples, L. Ling, A. Fetterman, D. Hughes, A. Bissell, H. Torrey, *et al.* Chem. Biol., **21**, pp. 509-518 (2014)

111. Bochtler, M., Hartmann, C., Song, H.K., Bourenkov, G.P., Bartunik, H.D., Huber, R., 2000. The structures of HsIU and the ATP-dependent protease HsIU-HsIV. Nature 403, 800–805.

112. Lee, S., Sowa, M.E., Watanabe, Y.H., Sigler, P.B., Chiu, W., Yoshida, M., Tsai, F.T., 2003. The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. Cell 115, 229–240.