Stony Brook University

Structural Insights into Proteasome Activators in *Mycobacterium tuberculosis*

A Dissertation Presented

by

Kuan Hu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

August 2017

Stony Brook University

The Graduate School

Kuan Hu

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Name – Dr. Huilin Li
Professor, Van Andel Research Institute

Name – Dr. Miguel Garcia-Diaz
Associate Professor, Department of Pharmacological Sciences

Name – Dr. Markus Seeliger
Assistant Professor, Department of Pharmacological Sciences

Name – Dr. Steven E. Glynn
Assistant Professor, Department of Biochemistry and Cell Biology

Name – Dr. K. Heran Darwin
Professor, New York University School of Medicine

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Structural Insights into Proteasome Activators in *Mycobacterium tuberculosis*

by

Kuan Hu

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

2017

The proteasome is a large molecular machine for regulated proteolysis. It has been found in all archaea, eukaryotes, and several bacterial species of the order Actinomycetales. *Mycobacterium tuberculosis* (*Mtb*), an actinomycete pathogenic to humans, possesses a proteasome system that is essential for its virulence. Two proteasome activators, Mpa and PafE, mediate ATP-dependent and ATP-independent degradation, respectively, in *Mtb*. The molecular mechanisms of *Mtb* proteasome activation are still unknown due to the lack of structural information of these two activators. I have utilized several structural biology techniques, complemented with biochemical experiments to gain insights into the structure and function of Mpa and PafE. The crystal structure of PafE revealed a unique dodecameric ring structure. The cryo-EM structure of $PafE_{\Delta 155\text{-}166}$ in complex with 20S proteasome showed that the C-terminal GQYL interacts with α-subunits and facilitates gate opening. Crystal structures of the Mpa hexamer without N-terminal coiled-coils revealed a unique β-grasp fold near the C-terminus of Mpa, which buries the 20S-binding motif in the central channel and interferes with the interaction between Mpa and 20S proteasome. In sum, my results provide novel structural insights into the mechanisms of ATP-dependent and -independent proteasome activators of *Mtb* and further the understanding of proteolytic regulation in human pathogen.

Table of Contents

List of Figures/Tables

List of Abbreviations

AAA	ATPase associated with various cellular activities
ADP	adenosine diphosphate
ATP	adenosine triphosphate
CP	core particle
CTF	contrast transfer function
Dop	deamidase of Pup
EM	electron microscopy
HspR	heat shock protein repressor
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
MALS	multi-angle light scattering
Mpa	mycobacterial proteasome ATPase
Mtb	*Mycobacterium tuberculosis*
OB	oligonucleotide/oligosaccharide-binding
PafE	proteasome accessory factor E
PAN	proteasome-activating nucleotidase
Pup	prokaryotic ubiquitin-like protein
RP	regulatory particle
SAD	single-wavelength anomalous dispersion
SEC	size exclusion chromatography

Chapter 1: Introduction to proteasome systems

1.1 The significance of intracellular proteolysis

Proteostasis is required for the long-term stability of any cell type (Labbadia and Morimoto, 2015). A healthy proteome is the basis for nearly all the biological processes, such as cell metabolism, signal transduction, organelle biogenesis, and stress response. Proteostasis not only depends on robust protein synthesis, but also requires efficient removal of disused, misfolded, or damaged proteins. Why is proteolysis so important? First, rapid elimination of key regulatory proteins (e.g. cyclins) and rate-limited enzymes is a key requirement for regulation of cell cycle and metabolic pathways, respectively. Second, cell proteins may be synthesized with abnormal conformation. Accumulation of defective proteins severely undermines cellular functions and is closely associated with neurodegeneration, autoimmune diseases, metabolic disorder, and cancer, most of which are currently incurable. For example, Alzheimer disease, a chromic neurodegenerative disease, is partially attributable to defective clearance of amyloid β-protein. Due to the destructive nature of protein degradation, intracellular proteolysis must be tightly controlled. Otherwise, excessive digest of cellular proteins would cause irreversible damages to cells. For instance, premature activation of trypsinogen leads to apoptosis at the cellular level and pancreatitis at the level of whole organism.

Cells have evolved two major degradation systems to carry out intracellular proteolysis. One is the autophagy-lysosome pathway, which performs non-selective proteolysis of large and potential detrimental cellular components such as protein aggregates. The other is the ubiquitin-proteasome system (UPS). As the primary approach for degradation of disused, misfolded, or damaged proteins, the UPS is involved in a variety of cellular functions, including cell cycle,

proliferation, apoptosis, and other cellular pathways (Hershko and Ciechanover, 1998). Degradation products can be used to recycle amino acids or loaded on to major histocompatibility class I molecules to communicate the intracellular protein to immune system. The 2004 Nobel Prize in Chemistry was awarded jointly to Irwin Rose, Aaron Ciechanover, and Avram Hershko for their contribution to the discovery of the ubiquitin-mediated protein degradation.

Owing to the fundamental importance on cell differentiation, proliferation, transcriptional activation, and angiogenesis, protein degradation systems, particularly, proteasomes become crucial targets for treatment of several diseases such as neurodegenerative diseases, diabetes, and cancer. For instance, the UPS has a crucial role in carcinogenesis and cancer development since it is involved in regulation of many cancer related proteins including cyclins, P53, proapoptotic Bid and Bax proteins, and NF-κB (Kaplan et al., 2017). During the last two decades, proteasome inhibitors have been tested in the lab to highlight their mechanisms of actions as well as in the clinic to treat different types of cancer. A highly selective inhibitor of proteasome, the peptide boronate bortezomib has been approved for treating relapsed or refractory multiple myeloma, and mantle lymphoma. Proteasome inhibitors are mainly used as adjuvant therapy because they can sensitize cancer cells to classical chemo-, radio-, and immune-therapy. In spite of significant effects, proteasome inhibitors show severe side effects such as peripheral neuropathy. Further studies are needed to consolidate the biological foundation of proteasome-dependent proteolysis and overcome side effects of proteasome inhibitors.

1.2 Catalytic center of a proteasome – 20S core particle

A proteasome is a large compartmentalized protease complex that process ~80% of all cellular proteins into peptides of 3-22 amino acids in length (Rock et al., 1994). Typically, a proteasome centers on a 20S core particle (CP) which catalyzes the proteolysis of unfolded substrates. The 20S CP can be found in eukaryotes, archaea, and some bacteria. Crystal structures of the 20S CP reveal a conserved barrel-like architecture made up of four stacked rings: two heptameric rings of β-subunits sandwiched between two heptameric rings of α-subunits (Groll et al., 1997; Lowe et al., 1995). The seven α- or β-subunits are the same in archaea and bacteria but different in eukaryotes (**Fig 1-1**). Whereas all β-subunits are active proteases in archaea and bacteria, eukaryotes have only three catalytically active β-subunits (β1, β2, and β5) possessing caspase, trypsin, and chymotrypsin-like activities, which cleave after acidic, basic, and hydrophobic amino acids, respectively. In immune cells, three different β-subunits ($\beta1_i$, $\beta2_i$, and $\beta5_i$) are constitutively expressed and preferentially incorporated into a specialized type of proteasome called immunoproteasome. The access to the degradation chamber in 20S CP is strictly controlled. The gates composed by N-terminal region of α-subunits are generally closed in free 20S CPs and become open upon docking of proteasome activators (Forster et al., 2005; Groll et al., 2000; Sadre-Bazzaz et al., 2010; Stadtmueller and Hill, 2011; Whitby et al., 2000).

In eukaryotes, de novo chaperone-assisted biogenesis of 20S CPs begins with assembly of α-rings, which provides a platform for subsequent attachment of β rings. Newly synthesized β-subunits has propeptides making themselves inactive proforms to prevent unnecessary proteolytic activity before being incorporated into mature 20S CPs. Final dimerization of two half-proteasomes goes along with autocatalytic removal of the propeptides of β-subunits. Five

chaperones aid in the assembly and maturation of eukaryotic 20S CP. PAC1-PAC2 forge an α-ring out of seven different α-units, while PAC3-PAC4 and POMP order the seven β-subunits onto the α-ring (Mayor et al., 2016). Unlike eukaryotic proteasomes, prokaryotic proteasomes start the assembly process with heterodimers of α- and β-subunits, which form half-proteasomes. In both prokaryotic and eukaryotic 20S CP, the propeptide promotes the proteasome assembly. However, the propeptide in the mycobacterial proteasome appears to inhibit rather than promotes proteasome assembly (Li et al., 2010).

The active β-subunits of the 20S CPs belong to the Ntn (N-terminal nucleophile) hydrolases. They follow a classical two-step mechanism to hydrolyze peptide bonds. After autolysis, mature β-subunits have the N-terminal threonine (Thr1) attach the carbonyl carbon of the protein substrate's scissile bond, yielding a tetrahedral intermediate (Groll et al., 2002). In the second step, a water molecule activated by deprotonated Thr1 amino group then hydrolyzes the acyl-enzyme and releases Thr1 hydroxyl group for a subsequent catalytic cycle.

As the highly active catalytic site of the proteasome, β5-subunit is used as a major target in development of proteasome inhibitors. Crystallographic description of the active site of the proteasome results in an exponent growth in the number of studies on proteasome inhibitors over the last two decades (Genin et al., 2010). Generally, proteasome inhibitors can be classified into two types based on their mechanism of inhibition. Covalent inhibitors, such as peptide aldehydes, peptide boronates, epoxyketones, β-lactones, polyphenols, terpenoids, and peptide vinyl derivatives, possess an electrophilic reactive group that can react with the catalytic hydroxyl Thr1 group. Non-covalent inhibitors, such as aminobenzylstatin derivatives and peptides/pseudopeptides, bind to the proteasome active sites via weak bonds.

1.3 Eukaryotic proteasome systems

The UPS, as mentioned above, is the primary proteolytic route for short-lived proteins in the nucleus and cytosol of eukaryotic cells. Ubiquitinated proteins are recognized and degraded by the 26S proteasome in an ATP-dependent manner. There are also ATP-independent pathways led by two 11S types of proteasome regulators, PA28α/β and PA28γ, as well as two monomeric regulators, Blm10 and PI31.

1.3.1 Ubiquitin-proteasome system

Ubiquitination relies on a complex protein cascade that typically lead to the formation of an isopeptide bond between a lysine residue of the substrate and the C terminus of ubiquitin, a 76 amino acid protein in a β grasp fold. Generally, the ATP-dependent E1 activating enzyme transfers an activated ubiquitin to the E2 conjugating enzyme, which then attach the ubiquitin to the substrate with the help of E3 ligase. Substrate specificity is controlled by E3 ligase, which has a significantly larger population than E1 or E2 enzymes in eukaryotic cells. Deubiquitinases can reverse the modification and rescue the substrate from degradation.

In the UPS, the 20S CP is activated by the 19S RP. The 19S RP is a protein complex of about 900 kDa capping one or two ends of 20S CP to form the 26S proteasome. As core components of the base of 19S RP, regulatory particle triple-A (Rpt) 1-6 form a hexameric AAA ATPase that has direct interaction with an α-ring of 20S CP (**Figure 1-2**). The order of the unique ATPase subunits in the hexameric ring is Rpt1-2-6-3-4-5 (Forster et al., 2009; Tomko et al., 2010). The six subunits play specific and non-redundant roles in the gating of 20S CP and even in protein translocation (Erales et al., 2012; Kohler et al., 2001; Rabl et al., 2008; Rubin et

al., 1998; Smith et al., 2007). Mutagenesis assays demonstrate that only the C-terminal hydrophobic-tyrosine-X (HbYX) motifs of Rpt2, 3, and 5 are essential for gate opening in the 26S proteasome (Tian et al., 2011). Recently, two studies reported near-atomic-resolution cryo-EM structures of the human 26S proteasome that revealed specific docking of C-terminal tails of Rpt3 and Rpt5 to 20S CPs in spite of a closed CP gate (Huang et al., 2016; Schweitzer et al., 2016); Chen et al. reported an alternative state of human 26S proteasome with an open substrate-conducting channel by cryo-EM (Chen et al., 2016). High-resolution structures of proteasome ATPase hexamer were first reported in a crystallographic study of archaeal proteasome-activating nucleotidase (PAN) (**Figure 1-2**). The crystal structure of the OB domain and part of the CC domain of PAN revealed that the CC domain form dimer in a trimer symmetry, unlike the six-fold symmetry of the OB domain (Zhang et al., 2009a). The CC domain was reported to promote protein unfolding (Djuranovic et al., 2009; Zhang et al., 2009b). Moreover, the HbYX motifs of PAN can by themselves bind and trigger gate opening in the archaeal proteasome (Smith et al., 2007).

In addition to the hexameric ATPase, the base of the 19S RP also contains three scaffold subunits, Rpn1, Rpn2, and Rpn13. Rpn1 and Rpn13, together with Rpn10 in the lid, are responsible for substrate recognition and anchoring during the unfolding process (Shi et al., 2016; Husnjak et al., 2008). In the lid of 19S RP, six subunits, Rpn3, 5, 6, 7, 9, and 12, form a scaffold, which holds the Rpn8-Rpn11heterodimer on its inner side. Rpn11 serves as a deubiquitinase that cleaves ubiquitin chains from substrates (Verma et al., 2002).

1.3.2 ATP-independent proteasome system

The 11S activators, such as PA28α/β and PA28γ, form heptameric ring structures on one or two ends of 20S CP (Knowlton et al., 1997). The increased abundance of PA28α/β can be induced by interferon γ (Raule et al., 2014). In complex with 20S CP, PA28α/β is speculated to improve major histocompatibility class I antigen processing (Groettrup et al., 1996; Murata et al., 2001). PA28γ can only be found in nucleus, which probably contributes to cell cycle regulation (Realini et al., 1997). The crystal structures of 11S in complex with 20S CP revealed a different interacting interface from those of proteasome ATPases (**Fig 1-3**). The distinct gating process of 11S activators contains binding of C-termini to pockets between α-subunits and replacement of 20S CP Pro17 by the internal activation loops of 11S activators (Forster et al., 2005; Whitby et al., 2000). Generally, eukaryotic 20S CPs can be activated by docking of 19S RP or 11S activators at one or both ends. A hybrid with a 19S RP at one end and an 11S at the other has also been identified (Coux et al., 1996).

Contrary to the multi-subunit architecture of the above activators, Blm10 is formed with a single-polypeptide activator of about 250 kDa that can cap the end of a 20S CP like a dome (Schmidt et al., 2005). In the crystal structure of Blm10-20S complex, the C-terminal three residues are superimposable with that of 11S between α5 and α6 subunits of the 20S CP (**Fig 1-3**) (Sadre-Bazzaz et al., 2010). However, Blm10 inhibits the full opening of the gate by incomplete repositioning of the other α-subunits, resulting in a relatively low level of peptidase stimulation by Blm10 compared with 11S (Iwanczyk et al., 2006). So far, the biological function of the two monomeric regulators, Blm10 and PI31, is not well understood.

1.4 Mycobacterial proteasome systems

1.4.1 Pup-proteasome system

Tuberculosis (TB) is a disease caused by *Mtb*. According to *Global Tuberculosis Report 2016* by World Health Organization, nine million people developed TB in 2013 and 1.5 million died (http://www.who.int/tb/en/). The emergence of multidrug-resistant and extensively drug-resistant *Mtb* strains makes it difficult to cure TB patients and eliminate the global epidemic. Therefore, the research and development of the novel strategies to cure TB is urgently required.

Previous data showed that nitric oxide (NO), a host-produced antimicrobial molecule, plays a key role in controlling *Mtb* growth after infection (Nathan and Shiloh, 2000). NO can penetrate bacterial membrane to form reactive nitrogen intermediates and reactive oxygen intermediates, which can induce cytotoxic effects such as DNA and protein damage (Alvarez and Radi, 2003; Szabo, 2003). However, the protection is imperfect and *Mtb* can still survive under the attack of NO (MacMicking et al., 1997), indicating an unknown biological pathway responsible for overturning those cytotoxic effects. Drugs targeting the pathway could weaken bacterial defense against host-produced NO and help to completely eliminate *Mtb*.

To identify *Mtb* genes required for NO resistance, Darwin et al. screened over 10,000 transposon mutants for hypersusceptibility to nitric oxide. The results showed that 5 out of 12 mutants had mutation in bacterial proteasome-associated genes (Darwin et al., 2003). It was also later shown that the 20S CP was required for NO resistance (Gandotra et al., 2010; Gandotra et al., 2007). To reveal the molecular mechanism of NO resistance by *Mtb* proteasome system and design drugs targeting this system, the structure and function of proteasomes in *Mtb* has been extensively studied.

The architecture of 20S CP in *Mtb* is closely similar to that in the archaea (Hu et al., 2006). The protease activity of 20S CP is provided by the N-terminal threonine of the β-subunit (Seemuller et al., 1996). Given that β-subunits are synthesized with a 56-residue N-terminal pro-peptide, the catalytic threonine is not exposed until the pro-peptide is auto-catalytically removed. Mutation of the catalytic residue to alanine (T1A) prevents the auto-cleavage and poses a barrier to assembly of the mature CP (Li et al., 2010). The gating mechanism is quite important for proteasomes in controlling the entrance of substrate. Like the eukaryotic 20S CP, the *Mtb* 20S CP tightly seals two gates, as shown in a low-resolution cryo-EM structure (Li et al., 2010). Furthermore, a high-resolution crystal structure of the *Mtb* 20S T1A mutant revealed that seven identical N-terminal loops of α-subunits with three different conformations seal the gate. These observations give a possible reason why the *Mtb* 20S CP has a relatively low peptidase activity compared with other prokaryotic proteasome. An open-gate (OG) conformation by removing eight residues from N-terminus of α-subunits results in an increase in peptidolytic activity towards small peptides *in vitro* (Lin et al., 2006).

Extensive studies of the proteasome system in *Mtb* during the last decade demonstrated that the pathogen adopts a unique cascade for targeted protein degradation that is evolutionarily unrelated to the ubiquitin-proteasome system in eukaryotes (**Fig 1-4**). The protein substrates targeted for degradation in *Mtb* undergo a post-translational modification, specifically covalently conjugation of a modifier to a lysine residue. The small protein modifier, Pup, has a C-terminal Gly-Gly-Gln motif (Pearce et al., 2008). The C-terminal Gln is deamidated to Glu by Dop before conjugation to protein substrates by PafA (Striebel et al., 2010). Mpa is an AAA ATPase homologous to PAN in archaea and Rpt proteins in eukaryotes (Darwin et al., 2005). Mpa is

responsible for the recognition, unfolding, and translocation of pupylated protein substrates into the degradation chamber of the 20S CP for cleavage.

Identification of pupylated proteins in *Mtb*, as called "pupylome", demonstrates several important proteasome-dependent regulation mechanisms that are associated with normal growth and pathogenesis of *Mtb*. PafA and Mpa are pupylated proteasome substrates, which reveals a reverse control of pupylation and prevents excessive protein degradation. Pupylation inhibits the activity of Mpa by interrupting its oligomerization state (Delley et al., 2012). A recent screen of suppressor of NO hypersensitivity in Mpa mutant revealed that Rv1205, a pupylated proteasome substrate, is likely responsible for the sensitization of *Mtb* proteasome-associated mutants to NO (Samanovic et al., 2015). Rv1205 catalyzes the synthesis of cytokinins, a type of hormones secreted by *Mtb*. In Mpa-deficient strain, accumulation of Rv1205 results in deposition of cytokinin breakdown products, which leads to NO sensitivity. Beside pup-proteasome system, pupylation may also participate in other biological pathways since not all pupylated proteins accumulated in degradation mutants (Festa et al., 2010).

1.4.2 ATP-independent proteasome system

Aside from the Mpa-mediated ATP-dependent Pup-proteasome system, there is also an ATP-independent proteasome pathway in *Mtb*. PafE enhances degradation of disordered proteins without Pup or ATP (Delley et al., 2014; Jastrab et al., 2015). Heat shock protein repressor (HspR) is one of the most abundant proteins in a PafE-deficient *Mtb* strain, indicating a correlation between PafE-mediated proteasome degradation and the heat shock response in *Mtb* (Jastrab et al., 2015). PafE probably contributes to normal growth, heat shock resistance, and full

virulence primarily by promoting the expression of protein chaperone genes that are necessary for surviving proteotoxic stress (Jastrab et al., 2017).

1.4.3 Mycobacterial proteasome is a promising anti-TB target

Because pup-proteasome system is essential for lethal *Mtb* infections for numerous reasons, the *Mtb* proteasome is considered as a promising anti-TB target. TB Alliance, a non-profit organization working on TB therapy, has listed the proteasome system as one of five top anti-TB targets. Due to the extensively conserved structure of the proteasome, most inhibitors block both eukaryotic and prokaryotic proteasomes. Only those inhibitors substantially more potent against the mycobacterial proteasome than the human proteasome are usable for eliminating TB. The oxathiazol-2-one compounds have been identified as selective and irreversible inhibitors of *Mtb* proteasome by cyclocarbonylating the catalytic threonine (Lin et al., 2009). Another type of mycobacterial proteasome inhibitor is based on dipeptide compounds, which have already been identified as eukaryotic proteasome inhibitors. Library screening and structure-based optimization of dipeptide compounds confer selectivity towards the mycobacterial proteasome over the human proteasome (Hsu et al., 2017; Lin et al., 2013).

1.5 Rationale of structural studies of proteasome activators in *Mtb*

Mpa and PafE activate the mycobacterial proteasome for degradation of pupylated and unfolded protein substrates, respectively, and contribute to NO resistance, copper resistance, and proteotoxic stress response of *Mtb* (Becker and Darwin, 2017). Interruption of proteasome

pathways by inhibition of proteasomal activators could be a potential approach for *Mtb* virulence attenuation. Structural studies of Mpa and PafE will help to understand the molecular mechanisms for proteasome activation and substrate selection. Moreover, high-resolution structures of mycobacterial proteasome activators will provide hints for the development of inhibitors targeting the proteasome activation process.

In this dissertation, I describe novel structural and functional features of PafE and Mpa. This work highlights unique structural features of these two *Mtb* proteasomal activators relative to their homologues in archaea and eukaryotes. Based on the crystal structure of PafE, I identified key residues for oligomerization and 20S CP binding. Furthermore, cryo-EM structures of the $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complexes explain the molecular mechanism of proteasome activation by PafE. Structural and functional analysis of Mpa revealed the molecular mechanism of the weak association of Mpa to 20S CPs *in vitro*. In sum, this work provides structural insight into ATP-dependent and -independent mycobacterial proteasome activation processes.

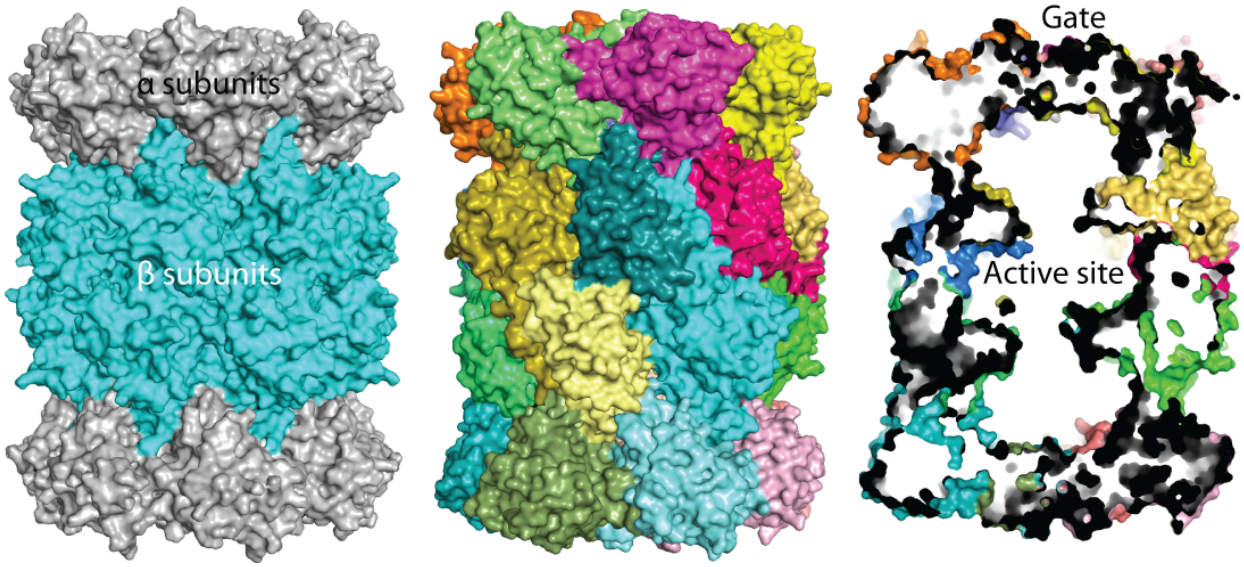


Figure 1-1. The architecture of proteasome CP.

The left panel shows the side view of the archaeal *Thermoplasma acidophilum* 20S CP (PDB: 1PMA). Two middle rings comprising seven identical β subunits (cyan) are sandwiched by two rings comprising seven identical α subunits (gray). The middle panel shows the side view of the eukaryotic *Saccharomyces cerevisiae* 20S CP (PDB: 1RYP). The α- and β-rings are composed of seven different subunits. The right panel shows the internal features of the *Saccharomyces cerevisiae* 20S CP. The gate in the α-ring and the active sites in the β-ring are labeled.

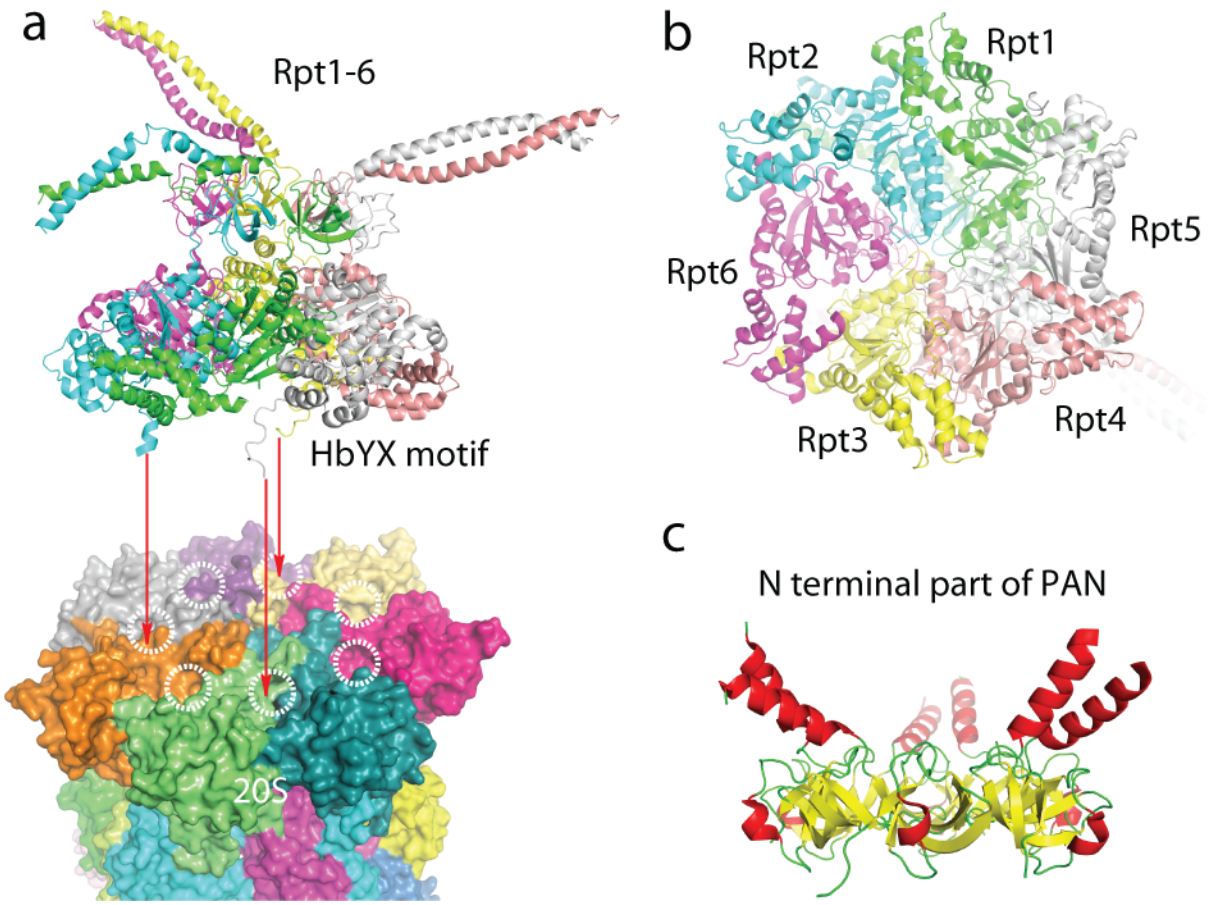


Figure 1-2. ATP-dependent proteasome activators in eukaryotes and archaea.

(a) The asymmetric association between Rpt1-6 (rainbow, upper) and 20S CP α ring (rainbow, below) is mediated by the C-terminal HbYX motifs of Rpt2, 3, and 5. **(b)** Bottom view of hexameric ring of Rpt1-6. **(c)** N terminal part of PAN comprises coiled-coil and OB domain.

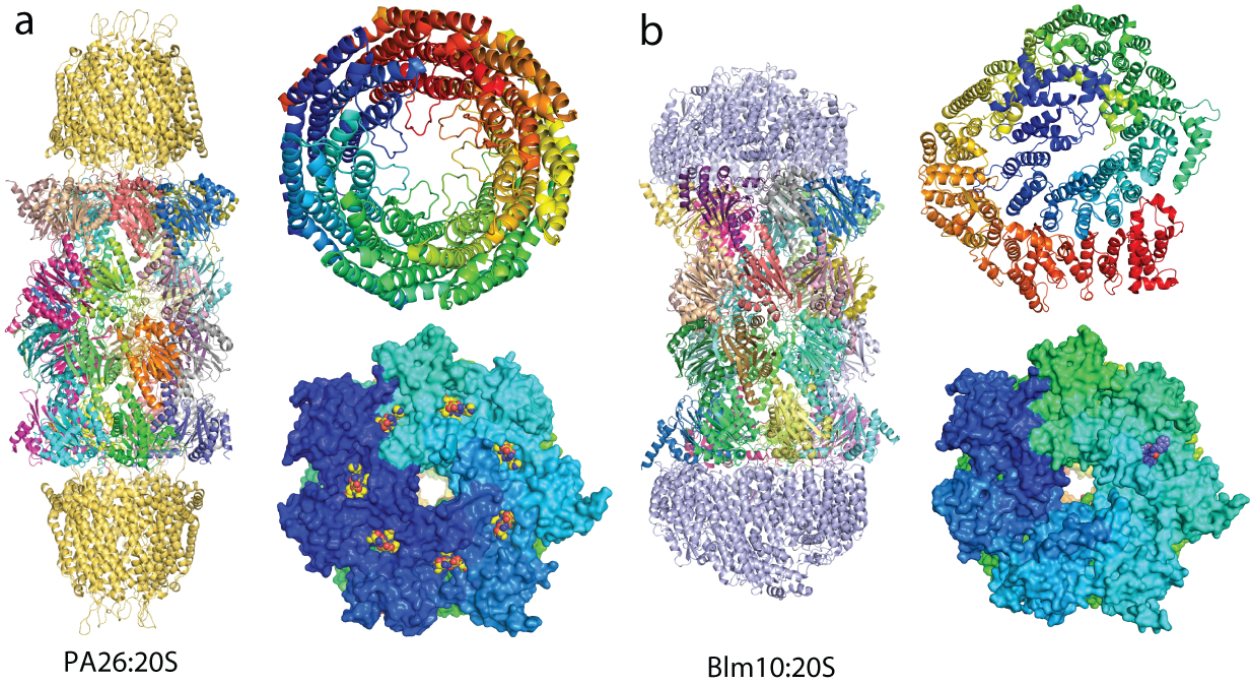


Figure 1-3. ATP-independent proteasome activators in eukaryotes.

(a) Crystal structure of *Trypanosoma brucei* PA26 heptamer (yellow) in complex with the *Saccharomyces cerevisiae* 20S CP (PDB: 1Z7Q). Top view of PA26 is shown in right upper panel. The open conformation induced by PA26 and seven ordered PA26 C-termini are shown in the right lower panel. **(b)** Crystal structure of the *Saccharomyces cerevisiae* Blm10 (light blue) in complex with the 20S CP complex (PDB: 1VSY). Top view of Blm10 colored from N-terminus (blue) to C-terminus (red) is shown in the right upper panel. The partially opened gate induced by Blm10 and Blm10 C-termini are shown in the right lower panel.

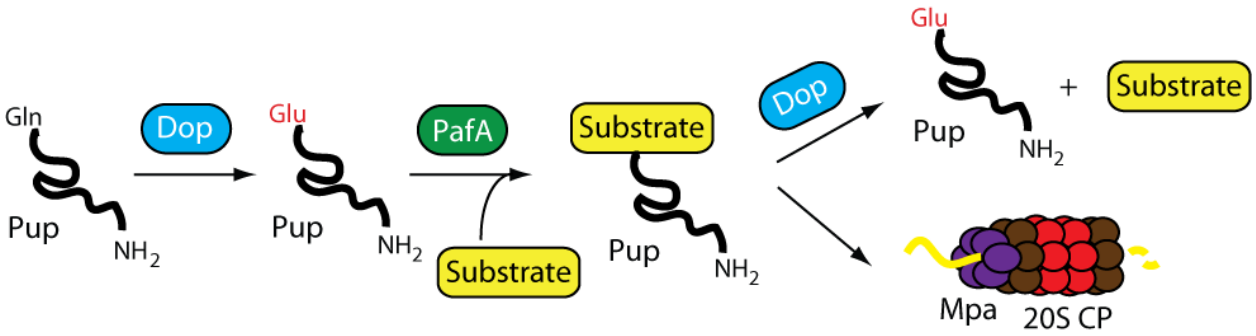


Figure 1-4. Pupylation-dependent degradation in Mycobacteria.

From left to right, the disordered Pup has its C-terminal Gln deamidated by Dop. PafA then covalently links the activated Pup_{Glu} to a substrate. The pupylated substrate is recognized and unfolded by Mpa, and finally degraded by 20S proteasome. The pupylated substrate can also be depupylated to escape from degradation.

Chapter 2: Methods and materials

2.1 Bacteria strain, growth conditions, and primers

E. coli strains were grown in Luria-Bertani broth medium (Teknova) at 37 °C, and Luria-Bertani agar (1.5%) was used for solid medium. For selection, media were supplemented with 30 μg/ml kanamycin or 100 μg/ml ampicillin. Primers were purchased from Integrated DNA Technologies (**Table 1**), and Stony Brook DNA Sequencing Facility performed plasmid sequencing.

2.2 Plasmid construction and protein purification

Gene fragments of *pafE* and *mpa* were amplified from *Mtb* chromosomal DNA using Pfu Ultra II Fusion HS DNA Polymerase (Agilent). The plasmid for co-expression of *pafE* and *parP* was provided by Dr. K. Heran Darwin. The Stratagene QuickChange Mutagenesis Kit was used to produce single or double point mutations in PafE or Mpa.

For PafE crystallization, a His_6 tag was fused to the C-terminus of $PafE_{15\text{-}154}$ or $PafE_{44\text{-}154}$. For *in vitro* binding assay and degradation assay, the hybrid protein $PafE_{\Delta155\text{-}174}$-$Mpa_{602\text{-}609}$ was prepared by linking $Mpa_{602\text{-}609}$ to $PafE_{\Delta155\text{-}174}$ by PCR; a His_6 tag was fused to the N-terminus of $PafE_{\Delta155\text{-}174}$, $PafE_{\Delta155\text{-}174}$-$Mpa_{602\text{-}609}$, $PafE_{\Delta155\text{-}159}$, $PafE_{\Delta155\text{-}164}$, $PafE_{\Delta155\text{-}166}$, and $PafE_{\Delta155\text{-}169}$. All gene fragments were cloned into pET24b (+) vector (Novagen) between the NdeI and NotI restriction sites. The above constructs were confirmed by sequencing and used to transform competent BL21(DE3) strain. Bacterial cells were cultured at 37 °C until OD_{600} reached 0.6-0.8. IPTG at a final concentration of 0.5 mM was added into medium to induce protein production.

Cells grew for another 4 h and collected by centrifugation at 4,000 g for 10 min at 4 °C. Bacterial pellet was resuspended in buffer containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, and 10 mM imidazole and passed through a French press twice at 1000 bar. Cell debris was removed by centrifugation at 24,000 rpm at 4 °C. Supernatant was loaded to a 5-ml Hitrap HP Ni^{2+}-nitrilotriacetate acid agarose (Ni-NTA) column (GE healthcare) pre-equilibrated with buffer containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, and 10 mM imidazole. Target proteins were eluted with buffer containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, and a linear gradient concentration of imidazole ranging from 20 mM to 300 mM. Elution was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Fractions containing target proteins were pooled, concentrated with Amicon Ultra centrifugal filters (Millipore), and further purified either in a HiLoad 16/60 Superdex 200 PG or a Superose 6 10/300 GL gel filtration column. For crystallization, $PafE_{15-154}$ was concentrated to 20 mg/ml in buffer containing 10 mM Tris-HCl pH 8.0 and 50 mM NaCl, whereas $PafE_{44-154}$ at 20 mg/ml was in 10 mM Hepes pH 7.4 and 50 mM NaCl. For 20S binding assays, PafE variants were concentrated to 100 mg/ml in buffer containing 20 mM Tris-HCl pH 8.0 and 150 mM NaCl. Selenomethionine-substituted proteins were produced as for the native proteins, except that the bacteria were grown in M9 medium. Dithiothreitol at a concentration of 5 mM was used in whole purification process of selenomethionine-substituted proteins to protect selenium from oxidation.

For Mpa crystallization in ADP-bound form, N-terminal 94 residues and C-terminal 7 residues were removed and a flexible loop outside the OB2 domain (194-210) was replaced with triple glycine linker. The Mpa variant was overproduced in *E. coli* with a N-terminal His_3 tag. IPTG with a final concentration of 0.2 mM was used for protein induction at 30 °C. Bacterial cells were collected after 6 h and homogenized in buffer containing 25 mM Tris-HCl, pH 8.2,

500 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM ATP, 10 mM imidazole. Cell debris was removed by centrifugation at 30,000 g at 4 °C for 60 min. Supernatant was loaded to a 5-ml Ni-NTA cartridge (Qiagen). Target proteins were eluted with an imidazole step gradient. The elution fractions containing Mpa proteins were immediately loaded onto a Superdex 200 gel filtration column equilibrated with 10 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM ATP. Fractions containing Mpa proteins were used directly for crystallization. Purification and crystallization were done within 24 h to prevent nonspecific degradation and aggregation.

For Mpa crystallization in apo form, N-terminal 97 residues were removed. The Mpa variant was overproduced in *E. coli* with a C-terminal His_6 tag preceded by a thrombin cleavage site. Overexpression of target proteins was induced by 0.4 mM IPTG at 20 °C. Induced cells were cultured for 16-20 h. Bacterial cells were lysed with microfluidics in buffer containing 25 sodium phosphate pH 7.4, 500 mM NaCl, and 10% glycerol. Supernatant was loaded onto a 5-ml Hitrap HP Ni-NTA column and washed with lysis buffer containing 60 mM imidazole and elution with a linear imidazole gradient. Fractions containing target proteins were pooled and further purified using a HiLoad 16/60 Superdex 200 PG column in buffer containing 10 mM Hepes pH 7.4 and 200 mM NaCl.

$Mpa_{C\text{-}ext}$ used for 20S CP binding study was prepared by inserting GGGGS between Thr-601 and Glu-602 in $Mpa_{98\text{-}609}$ with an internal His_6 tag preceding residue 199. $Mpa_{98\text{-}609}$ and $Mpa_{C\text{-}ext}$ were overproduced in BL21(DE3) by IPTG induction at 16 °C. Bacterial cells were lysed in buffer containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, and 10% glycerol. Target proteins were purified using a 5-ml Hitrap HP Ni-NTA column and a Superose 6 10/300 GL gel filtration column. Purified proteins were concentrated to 20 mg/ml in buffer containing 20 mM Tris-HCl pH 8.0, 150 mM NaCl, and 10% glycerol. The purity of protein was examined by SDS-

PAGE. Protein concentration was determined using a Nanodrop ND-1000 Spectrophotometer (A280).

The plasmid for co-expression of *pafE* and *parP* was transformed into E. coli BL21(DE3) and expression was induced by 0.2 mM IPTG at 16 °C. Bacteria were collected after over-night induction and homogenized in buffer containing 20 mM Tris-HCl pH 8.0 and 300 mM NaCl. The cell lysates were loaded to 5 ml Ni-NTA column and target proteins were eluted with a linear gradient concentration of imidazole. Further purification was performed using a Superose 6 10/300 GL gel-filtration column.

2.3 Crystallization, data collection, and structure determination

Crystallization of PafE variants was performed using the sitting-drop vapor-diffusion method at 20 °C. For $PafE_{15\text{-}154}$, crystals appeared after 2 d from drops consisting of 1 μl protein solution and 1 μl reservoir solution containing 0.2 M ammonium acetate, 0.1 M sodium citrate pH 5.9, and 29% (v/v) 2-methyl-2, 4-pentanediol. Selenomethionine-incorporated crystals were obtained under the same condition. The SAD data were collected from flash-cooled crystals at 100 K at 31-ID-D, Advanced Photon Source, Argonne National Laboratory, and processed using iMOSFLM (Battye et al., 2011) and SCALA (Winn et al., 2011). The positions of Se were determined using phenix.hyss (Adams et al., 2010). The initial phase and structural model were calculated using phenix.phaser and phenix.autobuild, respectively. COOT (Emsley and Cowtan, 2004), phenix.refine, and CCP4 were used in subsequent iterative model building and refinement until R_{work} = 19.6% and R_{free} = 21.5% were obtained.

For $PafE_{44-154}$, crystals were observed in drops consisting 1 µl protein solution and 1 µl reservoir solution containing 0.1 M Tris–HCl pH 8.3 and 2 M ammonium sulfate. Selenomethionine-incorporated crystals were produced using the same condition. The crystals were transferred to cryoprotectant containing 1.5 M lithium sulfate before flash freezing in liquid nitrogen. The SAD and native diffraction data were collected at X25/X29, National Synchrotron Light Source, Brookhaven National Laboratory, and processed using HKL2000 (Otwinowski and Minor, 1997). The initial phase and partial structure were calculated using PHENIX and improved manually using COOT. The phase was improved by molecular replacement using the initial model and the native data. The iterative refinement was carried out using PHENIX and manual building using COOT. Further model refinement was carried out using Refmac5 of CCP4 suite and PHENIX until R_{work} = 25.8% and R_{free} = 28.2% were obtained.

For crystallization of Mpa in ADP-bound form, 8-10 mg/ml purified proteins were crystallized in sitting drops containing 0.1 M Tris-HCl pH 8.2, 20% PEG 400, 0.2 M $MgCl_2$, and 5 mM ATPγS at 15 °C. Streak seeding and macro seeding were used to improve crystal quality. Crystals were flash-frozen in mother liquor containing 25% glycerol as a cryo-protectant. Diffraction data were collected at 0.9791 Å in the BL17U beamline, Shanghai Synchrotron Radiation Facility. Phase problem was solved by an advanced molecular replacement approach as implemented in the Phenix_Phaser-Rossetta program using a powerful computing cluster (DiMaio et al., 2013). Models of Mpa_{1-234} hexamer (PDB: 3M9B) and PAN-Rpt5C (PDB: 3WHK) were used as starting searching models. The initial model was largely built by ROSETTA and completed manually in COOT. ATPγS used in co-crystallization was hydrolyzed to ADP in the crystal. The structure was refined in PHENIX with Cartesian simulated annealing

and TLS refinement with a space group of P321 at a resolution of 2.9 Å, with R_{work} = 23.0% and R_{free} = 26.6%.

Hanging drop vapor diffusion was used for crystallization of Mpa in apo form. In brief, purified proteins were concentrated to 10 mg/ml in buffer containing 10 mM Hepes pH 7.4, 200 mM NaCl. A 0.5 µl droplet of protein sample was mixed with 0.5 µl of reservoir solution containing 100 mM Bis-Tris propane pH 7.0, 0.475 M NaCl, 18.2% PEG 4000, 20% glycerol. Crystals appeared in 7 days. Crystals were dehydrated for three weeks by replacing the mother liquor with 100 mM Bis-Tris propane pH 7.0, 0.6 M NaCl, 27-30% PEG 4000, 20% glycerol. Dehydrated crystals were directly flash frozen in liquid nitrogen. Crystal screening and data collection were carried out at the X25 and X29 beamlines of the National Synchrotron Light Source, Brookhaven National Laboratory. Diffraction images were indexed by IMOSFLM and scaled with SCALA. Most of the crystals were plate-like and highly anisotropic. The apo-form structure was solved by molecular replacement using the ADP-bound Mpa model in PHENIX. The initial model was manually adjusted in COOT and refined by PHENIX and Refmac5 in CCP4.

2.4 *In vitro* 20S CP binding assays

WT 20S, $20S_{OG}$, and $20S_{T1A}$ CPs were purified as described previously (Li et al., 2010). Purified $20S_{OG}$ or $20S_{T1A}$ CPs were incubated with WT PafE, internal-deletion PafE mutants, or PafE-Mpa chimeric protein at molar ratios of ~1:3 at 4 °C for 2-3 h. The mixture was loaded to a Superose 6 10/300 gel-filtration column in a buffer containing 20 mM Tris-HCl pH 8.0 and 150 mM NaCl.

2.5 Isothermal titration calorimetry

ITC experiments were performed in 20 mM Tris-HCl pH 8.0 and 150 mM NaCl at 25 °C using a VP-ITC microcalorimeter (MicroCal). WT PafE dodecamers (480 μM) were titrated into 33 μM WT 20S CPs; 95 μM PafE $_{\Delta 155\text{-}166}$ dodecamers were titrated into 4.5 μM WT 20S CPs and 5 μM $20S_{OG}$ CPs. In the absence and presence of 2 mM ADP or 2 mM AMPPNP, 30 μM $Mpa_{98\text{-}609}$ hexamers were titrated into 2 μM $20S_{OG}$ CPs; 40 μM $Mpa_{C\text{-}ext}$ hexamers without any exdogenous nucleotides were titrated into 2 μM $20S_{OG}$ CPs. The Origin 7.0 software was used to analyze the results.

2.6 *In vitro* degradation assay

For peptide degradation by PafE, 500 ng of purified *Mtb* 20S CPs were used and triplicate reactions were set up for each condition. At room temperature, 20S CP proteins were incubated in reaction buffer containing 50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, and 20 μM fluorogenic nonapeptide LF-2. A 10-fold molar excess of purified PafE was added; 20S CP and PafE:20S CP complexes were pre-incubated at 37 °C for 30 min and then allowed to return to room temperature before addition of substrate. Peptide degradation was monitored by measuring the change in fluorescence over time at excitation wavelength of 340 nm and emission wavelength of 405 nm. The reaction rates were calculated by determining relative fluorescence units generated per minute.

For the ParP degradation assay, 15 μg of purified 20SWT was incubated at room temperature in reaction buffer containing 50 mM Tris-HCl pH 8.0 and 5 mM $MgCl_2$. The reaction was initiated by addition of a 4-fold molar excess of purified PafE:ParP or $PafE_{\Delta 155\text{-}}$

$_{174}$:ParP. Reaction products were collected every 30 min and mixed with SDS loading buffer to stop reaction. The degradation of ParP was assessed by SDS-PAGE.

For the HspR degradation assay, 15 μg of purified $20S_{WT}$ CPs were incubated with 10 μg PafE or PafE mutants at room temperature in reaction buffer. The reaction was initiated by addition of a 30 μg HspR. Reaction products were collected every 30 min and mixed with SDS loading buffer to stop reaction.

For substrate degradation by Mpa, each reaction contained 6.9 μg 20S CP, 7.76 μg Mpa or Mpa mutant, 2 μg Myc-Pup~FabD-his, 2.5 mM ATP, 20 mM $MgCl_2$, 1 mM DTT, 50 mM NaCl, and 50 mM Tris-HCl, pH 8.0. Samples were collected and added to SDS loading buffer to stop reaction at indicated time points. Substrate amount was analyzed by SDS-PAGE, followed by immunoblotting with polyclonal antibodies to *Mtb* Fab-His_6.

2.7 *Mtb* growth conditions and heat-stress sensitivity assay

Mtb cultures were grown to an OD_{580} of about 1.0 in fresh Middlebrook 7H9 medium (Difco). For growth curves, cultures were diluted to OD_{580} = 0.025 and grown in triplicate, with OD_{580} monitored daily. For heat-stress sensitivity assays, cultures were diluted to OD_{580} = 0.08, a total of 1 mL was transferred to a 2-mL O-ringed tube and incubated for 24 h at 45 °C, and bacterial survival was assessed by plating serial dilutions on Middlebrook 7H11 agar to enumerate colony forming units. Statistical analysis was performed by using a non-parametric Student's two-tailed *t* test.

2.8 Negative staining EM and initial model

For negative staining EM of PafE alone and $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex, 5 µl protein solution (100 µg/ml) was applied on a glow-discharged carbon grid and incubated for 1 min. Subsequently, the grid was washed twice with water and stained with uranyl acetate. Images were recorded on a JEOL JEM-2010F transmission electron microscope operating at 200 kV and a magnification of 50,000× on a Gatan UltraScan 4000 CCD camera (4,096×4,096), resulting in a pixel size on the specimen level of 2.12 Å. Raw particle images were selected in a semiautomatic manner with e2boxer.py in EMAN2 (Tang et al., 2007). The CTF was first determined with raw images and corrected by phase-flipping the Fourier coefficients in the negative transfer regions. Phase-flipped data were then subject to reference-free 2D classification to generate a set of high-contrast class averages that were representative of views present in the raw particle images.

The 3D starting model of the $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex was calculated by e2initialmodel.py using reference-free 2D averages as input, and was refined against the phase-flipped dataset. D7 symmetry was applied during iterative refinement. The final 3D density map was then low-pass-filtered to 60 Å and used as an initial model in the following processing of cryo-EM datasets of $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex.

For negative stain EM analysis of ParP:PafE complex, 5 µl of protein sample were applied to a glow-discharged 300-mesh copper EM grid. The grid was then washed by water and stained twice for 45s by adding 5 µl 2% (w/v) uranyl acetate solution. The grid was then load to JEOL JEM-2010F transmission electron microscope operating at 200 kV equipped with a Gatan UltraScan 4000 CCD camera. Micrographs were recorded in low-dose mode (15 electrons/ $Å^2$) at

x50,000 magnification (4,096x4,096 pixel), which corresponded to 2.12-Å/pixel sampling at the specimen level. Particle selection and image processing were performed using EMAN2 software packages. PafE particles were selected in a semiautomatic manner with e2boxer.py. The particles were then subjected to 2D classification.

2.9 Three-dimensional cryo-EM of $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex

For low-resolution cryo-EM study of the $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex, grids were prepared in an FEI Vitrobot at 11 °C with the relative humidity set to 90% and the blotting pad height offset to -1.0 mm. Protein sample of 3.0 µl at about 0.4 mg/ml were pipetted onto a lacey carbon-coated 300-mesh copper grid that was freshly glow-discharged (Structure Probe Inc.). The sample solution was incubated on the EM grids for 30 s, blotted for 6 s before being plunged into liquid ethane that was precooled by liquid nitrogen. The cryo-EM grids were then transferred to a Gatan 626 cryo-specimen holder and then inserted into the JEOL JEM-2010F TEM operating at 200 kV. Cryo-EM images were recorded in the low-dose mode (15 $e^-/Å^2$) at 50, 000× microscope magnification on a Gatan UltraScan 4000 CCD camera (4,096 × 4,096 pixel), corresponding to a 2.12 Å/pixel sampling at the specimen level.

The CTF of raw micrographs was determined and corrected using ctffind3 (Mindell and Grigorieff, 2003) and particle selection was performed using semi-automated selection in RELION (Scheres, 2012). A total of 101,600 particles were extracted, normalized, and saved. Reference-free 2D classification gave several good 2D classes containing 77,100 particles for 3D classification and refinement. The low-pass-filtered 20S CP 3D map was used as the start model. Three main classes, 20S CP only (29,800 particles), 20S CP in complex with one PafE ring

(28,100 particles), and 20S CP in complex with two PafE rings (5,000 particles) were subjected to high-resolution 3D auto-refinement with or without a binary mask of 20S CP. D7, C7, and D7 symmetry were applied during the refinement of the above three classes, respectively. The resolution estimated by the gold-standard Fourier shell correlation method at the threshold of 0.143 (Scheres and Chen, 2012) was 9.4, 12, and 16 Å for the above three classes, respectively. The estimated resolution of the later two classes using binary masks was 7.9 and 12.6 Å, respectively. The difference map was calculated by subtracting the 20S CP from the 20S CP in complex with a PafE ring using the volume operation command *vop* in Chimera (Pettersen et al., 2004).

For high-resolution cryo-EM study of the $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex, 3 µl of protein sample were applied to a C-flat 2/1.2 holey carbon grid, which had been glow-discharged for 1 min by PELCO easiGlow (Ted Pella). Grids were plunge-frozen in liquid ethane using FEI Vitrobot with a blotting time of 7s. Data were collected using an FEI Titan Krios 300 kV transmission electron microscope equipped with a Gatan K2 Summit direct detector. Automated data collection was carried out in super-resolution mode at a dose rate of 11.5 electrons/physical pixel/s. For each movie, 30 frames were acquired in 6 s with a total accumulated exposure of 58 electrons/$Å^2$. A dataset of ~5,000 micrographs was collected using a defocus range between 1.0 and 3.0 µm. Beam-induced motion correction and CTF estimation was done by MotionCor2 (Zheng et al., 2017) and CTFFIND4 (Rohou and Grigorieff, 2015), respectively. Good micrographs were selected based on the extent and regularity of Thon rings. A total of about 180,000 particles were automatically picked and applied to reference-free 2D classification using RELION2 (Scheres, 2012) (**Fig 2-1a**). Classes with clear density of 20S proteasome, containing 141,275 particles, were selected for 3D classification. To make an initial reference, the EM

structure of *Mtb* 20S CP with both ends capped with $PafE_{\Delta 155\text{-}166}$ was low-pass filtered to 50 Å. The particles were subsequently classified into eight classes without imposing any symmetry (**Fig 2-1b**). Two out of the eight classes showed 20S CPs with one end capped with PafE. The two classes, containing 60,034 particles, were applied to high-resolution refinement without imposing any symmetry, which resulted in a map with an average resolution of 4.2 Å according to the FSC=0.143 criterion (**Fig 2-2**). Three classes containing 51,091 particles showed maps with good density of PafE on both ends of 20S CP. Subsequent refinement with the D7 symmetry was performed using those particles. A mask was then generated to mask the 20S CP for high-resolution refinement using the D7 symmetry. Finally, a map of 20S CP upon PafE binding was obtained with an average resolution of 3.4 Å according to the FSC=0.143 criterion (**Fig 2-3**). Local resolution was estimated for the reconstruction with Resmap (Kucukelbir et al., 2014). Maps were visualized using UCSF Chimera.

The model of *Mtb* 20S CP (PDB: 3MI0) was rigid body fitted into the 3.4-Å cryo-EM map and then iteratively refined using real-space refinement in PHENIX. The resulted model was visually inspected in COOT. Several secondary structures and side chain rotamers were manually adjusted to best fit the density. The last four residues of PafE were also manually added to the extra density between α subunits of 20S CP. The quality of the final map was analyzed with Molprobity (Chen et al., 2010).

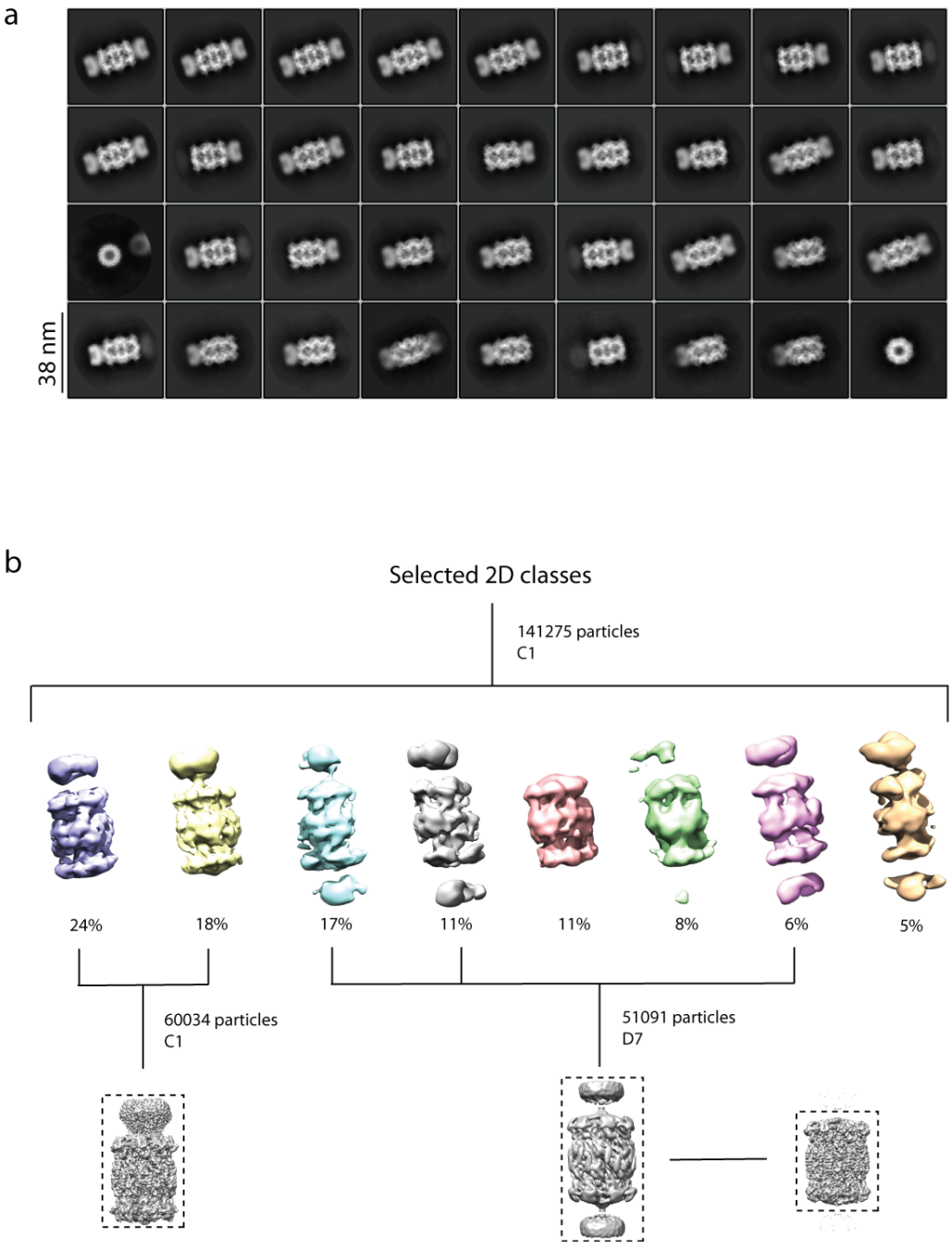


Figure 2-1. A cryo-EM study of the PafE$_{\Delta 155\text{-}166}$:20S$_{WT}$ complex.

(a) Representative 2D class averages of roughly 180,000 auto-picked particles. **(b)** The procedures outline to generate the 3D reconstructions of the PafE$_{\Delta 155\text{-}166}$:20S$_{WT}$ complex. The eight classes obtained from 3D classification are showed and the distribution is indicated by percentage below. The classes for high-resolution reconstructions are indicated. The dashed lines show the mask used in refinement.

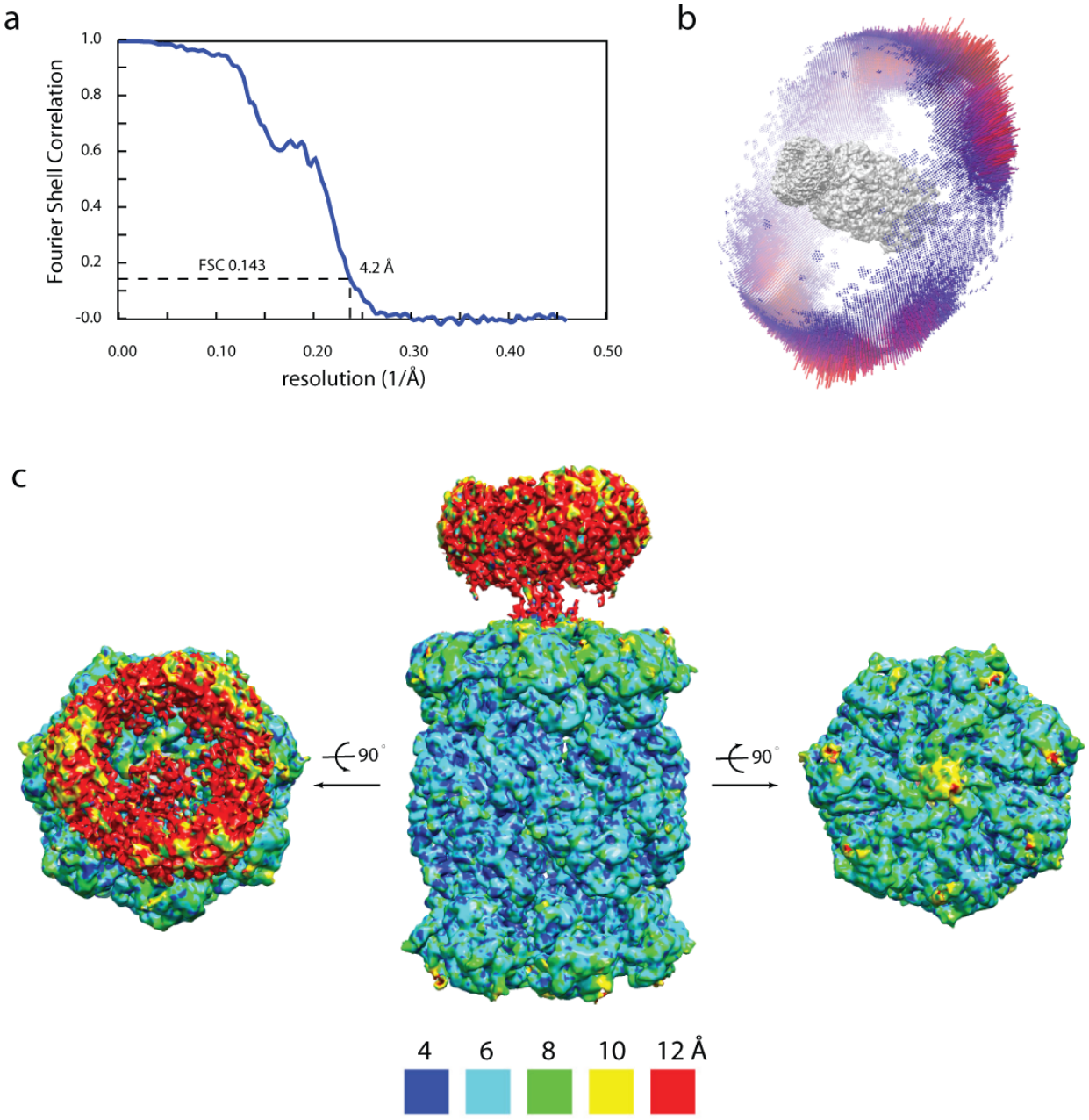


Figure 2-2. Fourier shell correlation and local resolution estimate for EM reconstruction of 20S CP with PafE$_{\Delta 155\text{-}166}$ bound at one end.

(a) Fourier shell correlation curve of 20S$_{WT}$ CP with PafE$_{\Delta 155\text{-}166}$ bound at one end. No symmetry is used in the reconstruction. **(b)** Euler angle distribution of the particles used in the reconstruction. **(c)** Surface representation of the EM map colored according to the local resolution estimate by ResMap.

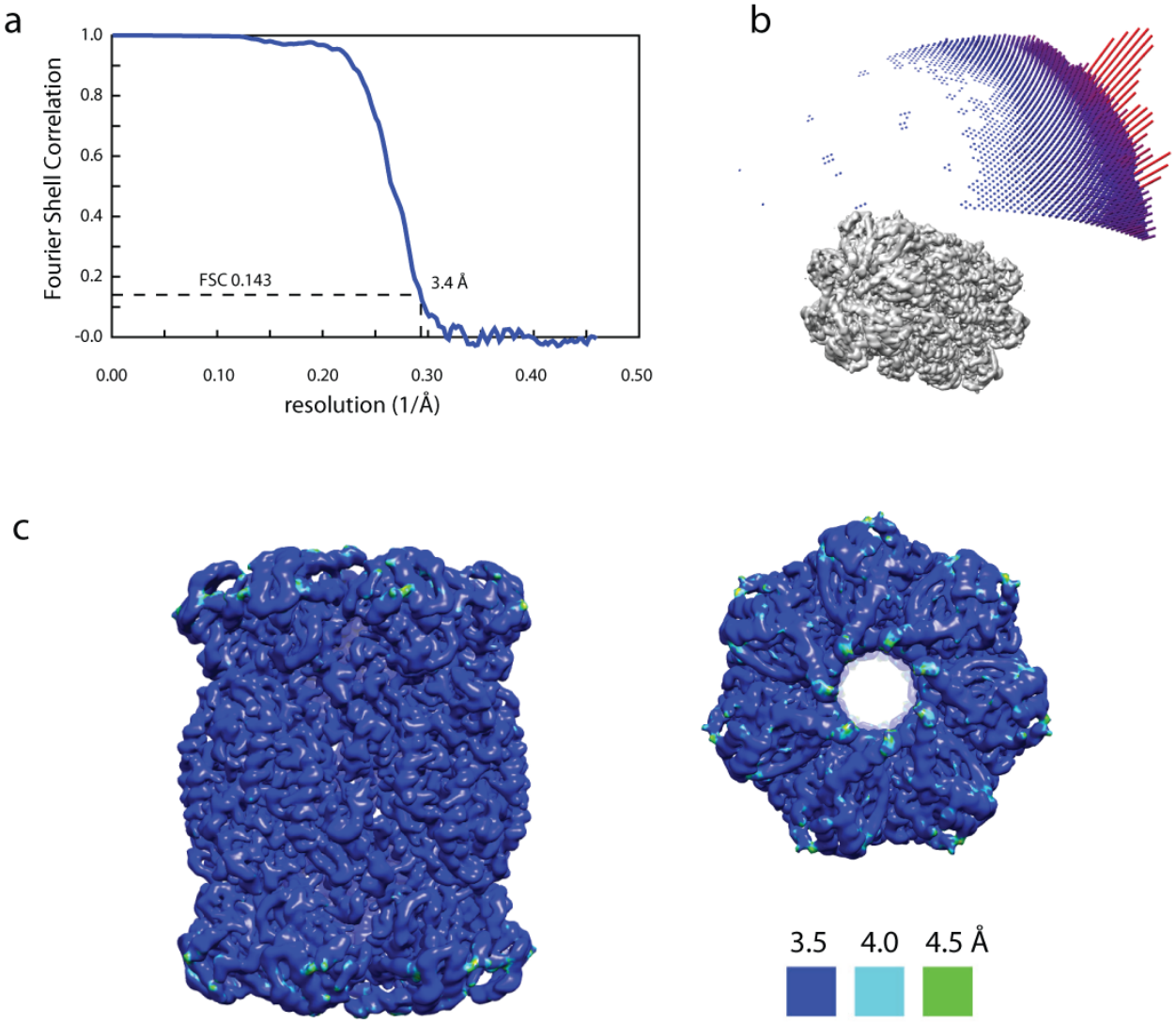


Figure 2-3. Fourier shell correlation and local resolution estimate for EM reconstruction of two-ends-capped 20S CP.

(a) Fourier shell correlation curve of $20S_{WT}$ CP with $PafE_{\Delta 155\text{-}166}$ bound at two ends. The PafE regions are masked out in the refinement. The D7 symmetry is used in the reconstruction. **(b)** Euler angle distribution of the particles used in the reconstruction. **(c)** Surface representation of the EM map colored according to the local resolution estimate by ResMap.

Table 2-1 Primers used in this study

Constructs	Primers
$PafE_{WT}$	Forward: GGAATTCCATATGGTGATTGGATTGAGTACCGG
	Reverse: ATAAGAATGCGGCCGCTTACAGGTATTGTCCGGTGC
$PafE_{35-174}$	Forward: GGAATTCCATATGTCGTCGCTGACCGAC
	Reverse: ATAAGAATGCGGCCGCTTACAGGTATTGTCCGGTGC
$PafE_{1-153}$	Forward: GGAATTCCATATGGTGATTGGATTGAGTACCGG
	Reverse: ATAAGAATGCGGCCGCTTACTGGCGCATTTGTTG
$PafE_{8-153}$	Forward: GGAATTCCATATGGGCAGCGACGACGACGA
	Reverse: ATAAGAATGCGGCCGCTTACTGGCGCATTTGTTG
$PafE_{15-153}$	Forward: GGAATTCCATATGGAGGTCATCGGCGG
	Reverse: ATAAGAATGCGGCCGCTTACTGGCGCATTTGTTG
$PafE_{22-153}$	Forward: GGAATTCCATATGCCGCGGCTGATAGC
	Reverse: ATAAGAATGCGGCCGCTTACTGGCGCATTTGTTG
$PafE_{28-153}$	Forward: GGAATTCCATATGCAGGAGAACGACTCCGAC
	Reverse: ATAAGAATGCGGCCGCTTACTGGCGCATTTGTTG
$PafE_{35-153}$	Forward: GGAATTCCATATGTCGTCGCTGACCGAC
	Reverse: ATAAGAATGCGGCCGCTTACTGGCGCATTTGTTG
$PafE_{44-153}$	Forward: GGAATTCCATATGCCCGCCAAGGTGATG
	Reverse: ATAAGAATGCGGCCGCTTACTGGCGCATTTGTTG
$PafE_{\Delta155-174}$-$Mpa_{602-609}$	Forward: GGAATTCCATATGCATCATCATCATCATCATGTGATTGGATTGAGTACCGG
	Reverse: ATAAGAATGCGGCCGCTTACAGGTACTGGCCGAGGTTGGACTCCTGGCGCATTTGTTGC

$PafE_{\Delta 155\text{-}174}$-$Mpa_{602\text{-}609/Y608A}$	Forward: GGAATTCCATATGCATCATCATCATCATCATGTGATTGGATTGAGTACCGG Reverse: ATAAGAATGCGGCCGCTTACAGGGCCTGGCCGAGGTTGGACTCCTGGCGCATTTGTTGC

Mutations	**Primers**
$PafE_{Q56I}$	Forward: GCACCATGATCAAG*ATA*CTGCTCGAGGAGG
	Reverse: CCTCCTCGAGCAG*TAT*CTTGATCATGGTGC
$PafE_{P65G}$	Forward: GAGGTTCGCGCCGCC*GG*ACTCGACGAAGCCAG
	Reverse: CTGGCTTCGTCGAGT*CC*GGCGGCGCGAACCTC
$PafE_{E96L}$	Forward: CGGAACTGCGCGAG*CAG*CTCGACCGGCTTA
	Reverse: TAAGCCGGTCGAG*CTG*CTCGCGCAGTTCCG
$PafE_{R99L}$	Forward: CGAGGAGCTCGAC*CTG*CTTACCCTGCCGTT
	Reverse: AACGGCAGGGTAAG*CAG*GTCGAGCTCCTCG
$PafE_{G124I}$	Forward: AGGCACAGCTGGTC*ATC*TGGCTGGAAGGGC
	Reverse: GCCCTTCCAGCCA*GAT*GACCAGCTGTGCCT
$PafE_{G128I}$	Forward: AGGCACAGCTGGTC*ATC*TGGCTGGAAGGGC
	Reverse: GCCCTTCCAGCCA*GAT*GACCAGCTGTGCCT
$PafE_{G132I}$	Forward: AAGGGCTGTTCCAC*ATC*ATCCAAACCGCGC
	Reverse: GCGCGGTTTGGAT*GAT*GTGGAACAGCCCTT
$PafE_{F138E}$	Forward: ATCCAAACCGCGCTA*GAA*GCTCAGCAAATGGCG
	Reverse: CGCCATTTGCTGAGC*TTC*TAGCGCGGTTTGGAT
$Mpa_{C\text{-}ext}$	Forward: *GCCACC*GAGGTTGGACTCGGTATCAAT
	Reverse: *GGAGGTAGT*GGCCAGTACCTGTAGGCG

Chapter 3 Structural analysis of the dodecameric proteasome activator PafE in *Mycobacterium tuberculosis*

This chapter is based on the following publications.

1) Lin Bai[1], Kuan Hu[1], Tong Wang[1], Jordan B. Jastrab, K. Heran Darwin, Huilin Li. (2016) Structural analysis of the dodecameric proteasome activator PafE in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 113: E1983–E1992. [1]Equal contribution.

Author contribution: L.B., K.H., T.W., and J.B.J. performed research; K.H.D. contributed new reagents/analytic tools; L.B., K.H., T.W., J.B.J., K.H.D., and H.L. analyzed data; and L.B., K.H.D., and H.L. wrote the paper.

2) Jordan B. Jastrab, Tong Wang, J. Patrick Murphy, Lin Bai, Kuan Hu, Remco Merkx, Jessica Huang, Champak Chatterjee, Huib Ovaa, Steven P. Gygi, Huilin Li, K. Heran Darwin. (2015) An adenosine triphosphate-independent proteasome activator contributes to the virulence of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 112(14): E1763–E1772.

Author contribution: J.B.J., T.W., J.P.M., R.M., C.C., H.O., S.P.G., H.L., and K.H.D. designed research; J.B.J., T.W., J.P.M., L.B., K.H., R.M., and J.H. performed research; R.M., J.H., C.C., and H.O. contributed new reagents/analytic tools; J.B.J., T.W., J.P.M., L.B., H.L., and K.H.D. analyzed data; and J.B.J. and K.H.D. wrote the paper.

3.1 Abstract

PafE (Rv3780, also known as Bpa) activates ATP-independent proteolysis by *Mtb* proteasomes. Moreover, PafE is required for the normal growth, heat-shock resistance, and full virulence of *Mtb*. We solved the crystal structure of PafE in the form of a dodecameric ring. PafE monomers form four-helix bundles similar to the eukaryotic ATP-independent proteasome activator, PA26 and PA28. A unique glycine-XXX-glycine-XXX-glycine motif is required for assembly of the PafE ring. Intriguingly, the truncation of PafE carboxyl-terminus results in the robust binding of PafE rings to native 20S CP and substantially increases proteasomal activity. The results suggest that the extended carboxyl-terminus of this cofactor confers suboptimal binding to 20S CP.

3.2 Introduction

Most bacteria do not have proteasome systems but use ATP-dependent proteases, such as ClpP, Lon, and HslUV, to degrade proteins (Butler et al., 2006; De Mot et al., 1999). However, the orders *Actinomycetales* and *Nitrospirales* possess proteasomes that structurally highly similar to eukaryotic and archaeal proteasomes (Burns and Darwin, 2010b). The human pathogen *Mtb*, an Actinomycete, requires proteasomes to cause lethal infections in mice (Darwin et al., 2003). The *Mtb* proteasome system has become a prioritized target for the development of anti-tuberculosis drugs due to the potential to target persistent or latent bacteria (Nathan, 2012; Nathan et al., 2008). In fact, *Mtb*-specific proteasome inhibitors have been identified as promising leads for new drugs to kill the pathogen (Lin et al., 2013; Lin et al., 2009).

The structure of 20S CP is highly conserved between prokaryotes and eukaryotes. However, its accessory factors diverge among the domains of life. To mark the substrate proteins, eukaryotes use the well-folded ubiquitin while *Mtb* uses the intrinsically disordered Pup (Burns and Darwin, 2010a; Wang et al., 2010). Furthermore, recognition and unfolding of ubiquitylated proteins by eukaryotic 26S proteasome relies on 19S RP that binds one or both ends of 20S CPs (Finley, 2009; Park et al., 2010). In contrast, the *Mtb* 20S CP uses Mpa for both recognition and unfolding of pupylated proteins (Darwin et al., 2005; Striebel et al., 2010; Wang et al., 2010).

In addition to the ATPase activators, proteolysis by proteasomes can also be stimulated by several ATP-independent factors, such as eukaryotic 11S activators PA26 and PA28. PafE has recently been discovered as an ATP-independent activator in *Mtb*, which stimulates the degradation of small peptides and β-casein *in vitro* (Delley et al., 2014; Jastrab et al., 2015). Both studies showed that a carboxyl-terminal glycine-glutamine-tyrosine-leucine (GQYL) motif is required for binding to and activating 20S CPs, and the penultimate tyrosine plays a similar role with tyrosine observed in the HbYX motif in other proteasome activators. Furthermore, PafE was reported to contribute to heat shock response of *Mtb* specifically by stimulating the degradation of HspR (Jastrab et al., 2017; Jastrab et al., 2015). PafE-mediated degradation does not require puplylation and ATP, which leads another path, aside from Mpa-mediated ATP dependent degradation, for targeting proteins to the mycobacterial proteasome. However, the oligomerization state of PafE and activation mechanism of the 20S CP by PafE are still unknown.

Here we determine by crystallography that PafE have a four-helix bundle structure similar to the 11S activators, but assemble into an unprecedented dodecameric ring structure with

12-fold symmetry. We further used ITC and cryo-EM to reveal that PafE activates 20S CPs by binding to one or both ends of 20S CP and inducing the gate opening.

3.3 Results

3.3.1 PafE forms oligomeric rings

According to a BLAST homology search, there is no significant similarity in amino acid sequence between PafE and any protein of known function. In order to understand its potential function, PafE was made in *E. coli* and purified using an N-terminal 6×His tag. Preliminary assessment of PafE proteins using negative stain EM and computational image classification revealed that PafE formed homo-oligomeric rings. According to the reference-free 2D classification, the PafE rings had an approximate diameter of 112.0 ± 2.3 Å (**Fig 3-1**). A SEC-MALS analysis further indicated the molecular mass of a PafE oligomer as ~247 kDa, suggesting that PafE ring consists of 12 subunits (Jastrab et al., 2015).

3.3.2 Crystal structure of PafE dodecameric ring

PafE is predicted to possess four α-helices with disordered regions at both the N and C-terminus (**Fig 3-2a**). Crystallization with full-length PafE protein did not give crystals with good diffraction quality. In order to improve the crystal quality, I screened several truncation forms of PafE with different length of N or C-terminal loops. Finally, I identified two PafE truncation forms with diffraction at better than 3-Å resolution: $PafE_{15-153}$ and $PafE_{44-153}$ (**Fig 3-2c**). The crystal structure of $PafE_{15-153}$ was determined at 2.88 Å resolution with R_{work} = 19.6% and R_{free} =

21.5%. The experimental phases were derived from selenium anomalous signals (**Table 3-1**). In the crystal structure, PafE did form dodecameric rings with 12-fold symmetry as predicted by EM work and SEC-MALS analysis (**Fig 3-3a**). The core of PafE monomer was made up of a four-helix (H1-4) bundle similar to those in the eukaryotic 11S activators (**Fig 3-3c**). The central channel of the PafE ring has a diameter of 40 Å, which is larger than the channels observed in 11S activators. Several hydrophobic residues, including Leu58, Ala64, Pro65, Ala113, Leu137, and Phe 138, line the central channel, which probably facilitate the putative function for threading unfolded polypeptides into the 20S CP. Moreover, the distal and side faces of PafE are highly negatively charged (**Fig 3-3d**).

3.3.3 An N-terminal helix and a double-GXXXG motif are essential for PafE ring formation

Interestingly, $PafE_{44-153}$ proteins showed an oligomerization state different from that of $PafE_{15-153}$ according to the SEC profile (**Fig 3-2b**). The observation was further confirmed by the crystal structure of $PafE_{44-153}$. $PafE_{44-153}$, in lack of the H0, crystallized as tetramers, each consisting of a dimer of dimers (**Fig 3-4a**), suggesting the importance of the short N-terminal helix H0 (residue 36-43) in assembly of the dodecameric rings. In the $PafE_{15-153}$ dodecameric ring, H0 was located between H2, H3, and H4. In contrast, deletion of H0 resulted in an inter-subunit H4-H4 interaction, yielding an artificial dimer structure of $PafE_{44-153}$ (**Fig 3-4b, c**). Therefore, the PafE N-terminal H0 is critical for the assembly of the dodecameric rings.

The PafE ring consists of two concentric shells: the inner shell composed by H1 and H4 and the outer one by H2 and H3. The diameters of the inner and outer shells were 60.7 and 84.2

Å, respectively. For the inner shell, H1 and H4 are tightly packed together with the intra- and inter-subunit distance of 10.5 and 6.5 Å, respectively (**Fig 3-5a**). The unusually short distance between H1 and H4 was made possible by the presence of a GXXXGXXXG motif in H4. The GXXXG motif is known to be particularly important for tight α-helix packing in membrane proteins. Mutating any of the three glycines to isoleucine disrupted PafE ring formation (**Fig 3-5b**), indicating the importance of this motif in maintaining the dodecameric ring structure. In addition to the tight helix packing, the subunit interface is further stabilized by a hydrogen bond (Gln56-Glu127) of the inner shell and two salt bridges (Glu84-Arg99 and Arg49-Glu96) of the outer shell. Breaking either one of the two salt bridges by mutation of Glu96 or Arg99 to leucine resulted in dissociation of the ring structure (**Fig 3-5c**). Disrupting the hydrogen bond by mutating Gln56 to isoleucine made the ring much less stable than the wide type one. Therefore, both the electrostatic interactions between H2 and H3 and helix packing between H1 and H4 contribute to the assembly of PafE dodecameric rings (Bai et al., 2016).

3.3.4 The hydrophobic central channel is important for PafE-dependent degradation

To assess the role of the axial channel of PafE in substrate degradation, we preformed a negative staining EM study on the ParP:PafE complex. ParP is a potential substrate of PafE-dependent degradation and can be co-purified with PafE in *E. coli*. In 2D classes (**Fig 3-6a**), remarkably, most ParP:PafE particles revealed ring structures with the central channels occupied by dot-like density. Given that PafE alone showed a completely hollow ring structure in negative staining EM (Jastrab et al., 2015), the dots occupying the axial channels of PafE turned to be

ParP proteins. With a diameter of 40 Å, the central channel of PafE is large enough to accommodate proteins like ParP, whose molecular weight is predicted to be 28 kD.

According to the crystal structure of PafE, the inner channel of PafE is lined largely by some hydrophobic residues, which are supposed to facilitate PafE to engage unfolded polypeptides for proteasomal cleavage. We assessed the importance of the hydrophobic channel on *in vitro* degradation of HspR. We mutated two typical hydrophobic residues Pro65 and Phe138 to Gly and Glu, respectively (**Fig 3-6b**). Both PafE mutants possess the same oligomeric state with the wild-type protein. In the presence of 20S CP, remarkably, the capability of PafE to mediate the cleavage of HspR was significantly inhibited by either single mutation (**Fig 3-6c**). These data indicate that the hydrophobic central channel of PafE can accommodate protein substrate for degradation.

3.3.5 C-terminal truncation leads to enhanced function *in vitro* and *in vivo*

In order to study how PafE activates proteasome, we tried to produce stable PafE-$20S_{WT}$ complex in vitro but failed due to weak interaction between the two proteins. Preliminary EM studies showed that PafE ring appeared to float above the $20S_{WT}$ with a gap of ~20 Å. Conspicuously, PafE had a 21-residue unstructured region preceding the GQYL motif that is essential for gate opening. We speculated that the C-terminal spacer modulated the interaction between PafE and 20S CP. We produced a PafE mutant with shortened C-terminal spacer, $PafE_{\Delta 155\text{-}166}$. ITC experiments performed by Lin Bai showed that $PafE_{\Delta 155\text{-}166}$ had relatively higher affinity for $20S_{WT}$ CPs, with a K_d of 4.6 μM, than the WT PafE (**Fig 3-7a**). In accordance with the increased binding affinity to CPs, $PafE_{\Delta 155\text{-}166}$ stimulated proteolysis of a peptide

reporter by $20S_{WT}$ CPs by over an order-of-magnitude more than the WT PafE (**Fig 3-7b**) (Bai et al., 2016).

We also tested if the truncated PafE variants had increased activity *in vivo*. In our previous study, we found that complementing a *pafE* deletion-disruption mutant with a WT *pafE* allele only partially rescued the growth defect and heat-shock sensitivity phenotypes of this mutant. Jordan introduced a single copy of either of two truncated *pafE* alleles ($pafE_{\Delta 155\text{-}159}$ and $pafE_{\Delta 155\text{-}164}$) expressed from the native *pafE* promoter into the chromosome of an *Mtb pafE*-null mutant. Immunoblotting showed that in spite of lower protein levels than the parental protein, both of the truncated *pafE* alleles fully restore growth to normal levels (**Fig 3-7c, d**). Moreover, complementing the $pafE_{\Delta 155\text{-}159}$ allele, restored heat-shock resistance to normal levels (**Fig 3-7e**).

3.3.6 Cryo-EM structure of $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP

The discovery of a PafE mutant, $PafE_{\Delta 155\text{-}166}$, that had dramatically enhanced activity provided us an opportunity to investigate the molecular mechanism of proteasome activation. We studied the stable $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex by cryo-EM and single-particle 3D reconstruction. The first dataset of EM images were collected using 200 kV JEOL JEM-2010F TEM equipped with a CCD camera. Three major species were obtained in reference-free 2D classification and unsupervised 3D classification: $20S_{WT}$ CP only, $20S_{WT}$ CP with one $PafE_{\Delta 155\text{-}166}$ ring, and $20S_{WT}$ CP with two $PafE_{\Delta 155\text{-}166}$ rings (**Fig 3-8a**). We next determined if there were any structural changes in $20S_{WT}$ CP upon binding of $PafE_{\Delta 155\text{-}166}$. Notably, the gate of $20S_{WT}$ CPs was opened only at the $PafE_{\Delta 155\text{-}166}$–capped end, while the gate remained closed at the end free of $PafE_{\Delta 155\text{-}166}$ ring.

The difference map between $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP and $20S_{WT}$ CP alone showed extra density located between neighboring α-subunits (**Fig 3-8b**). This density probably resulted from the PafE C-terminal tail, conformational changes in CP, or both of them. Actually, the density is located where PA26 and PA28 CT activating peptides bind to the archaeal or eukaryotic 20S CP. We modeled the PA26 activating peptide in the corresponding pocket between neighboring α-subunits of *Mtb* 20S CP, and found that the modeled peptide overlapped with the difference density map.

A conserved lysine found in the α-subunit of eukaryotic proteasomes is essential for binding by PA26 and PA28 activators. The C-terminal carboxylates of those activators form salt bridges with the lysines to facilitate proteasomal activation. We mutated the equivalent lysine (K52) to alanine in PrcA and found that the mutant could not be activated by PafE to digest a peptide substrate. Taken together, these results suggest that PafE activates the *Mtb* 20S CP by docking its C-terminal tails to pockets between α-subunits that are similar to those found in the archaeal and eukaryotic 20S CPs (Bai et al., 2016).

The mechanism of proteasome activation by PafE remained unclear due to the low resolution of cryo-EM structures. Bolten et al. recently report a 3.5-Å cryo-EM structure of PafE:$20S_{OG}$ CP, which revealed the off-axis binding of PafE to 20S CP and molecular details of interaction between GQYL motifs and α-subunits (Bolten et al., 2016). However, the results did not fully explain the activation process by PafE due to the lack of N-terminal loops in α-subunits. In order to reveal the native structure of 20S CP upon PafE binding, we determined high-resolution cryo-EM structure of the $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex. Roughly 180,000 particles were automatically picked from 4,836 cryo-EM images and then submitted to reference-free 2D classification. Only 20S CP particles capped by PafE on either or both sides were selected for 3D

classification. Auto refinement of particles containing only CP particles associated with PafE on one side yielded a 4.2-Å map according to gold standard Fourier shell correlation (FSC). With a local resolution of 4-6 Å, 20S CP shows very clear secondary structural features. However, the local resolution of PafE and its interface with 20S only approach 10 Å, making it difficult to distinguish the core helix bundle of PafE protomers. The low resolution of PafE density may be caused by orientation heterogeneity of PafE upon binding to 20S CP (Bai et al., 2016; Bolten et al., 2016).

The EM map of the $PafE_{\Delta 155\text{-}166}$-bound 20S CP complex shares a similar architecture with that of the PafE:$20S_{OG}$ CP complex reported by Bolten et al (Bolten et al., 2016). Due to deletion of 12 amino acids from PafE C-terminal linker, $PafE_{\Delta 155\text{-}166}$ is 10-Å closer to 20S CP in our structure (**Fig 3-9a**). Under low threshold, discontinuous density can be observed between the C-terminus of PafE and the gate of PafE-capped α-ring (**Fig 3-9b**), which we suspect to be flexible N-terminal loops of α-subunits. In the structure under high threshold, the α-ring associated with PafE has an open gate with a diameter of 27 Å, perfectly consistent with our previous report (Bai et al., 2016). The non-capped end of the 20S, by contrast, has a gate tightly sealed by well-defined N-terminal loops of α-subunits (**Fig 3-9c**). When the capped α-ring is aligned to the non-capped one, a 3-Å translational shift away from central axis can be observed in H0 of α subunits upon PafE binding (**Fig 3-10**), which contributes to disordering of the N-terminal loops and facilitates gate opening.

3.3.7 Binding of the GQYL motifs to α-subunits induces gate opening

To further understand how C-tails of PafE interact with α-subunits of 20S CP, we masked out PafE from the particles with both ends capped by PafE and did high-resolution refinement towards 20S CP alone with D7 symmetry. We finally got a 3.4-Å map, which showed a much higher local resolution in 20S CP than the previous map. The last four residues of PafE, Gly171-Gln172-Tyr173-Leu174, can be well fit to the additional density showing up between neighboring α-subunits (**Fig 3-11a; Table 3-2**). The interaction between the C-tail of PafE and 20S α-subunit involves the salt bridge between carboxyl group of Leu174 and Lys52, as well as cation–π interaction between Tyr173 and Arg26. The contribution of Lys52 to interaction with PafE has already been confirmed by our previous mutagenesis study (Bai et al., 2016). To confirm the interaction between Tyr173 and Arg26, we mutated Arg26 to alanine and applied the mutant to the HspR degradation assay (**Fig 3-11b**). The result showed that the digestion of HspR by PafE and 20S is totally interrupted by R26A mutation, suggesting that the interaction is essential for the PafE-mediated degradation by 20S proteasome. When the PafE-capped 20S α-subunit model was superimposed to non-capped α-subunit EM density, substantial clash was found between carboxyl group of Leu174 and Lys52, as well as Tyr173 and Arg26 (**Fig 3-11c**). The observation indicated that the docking of PafE GQYL motifs is directly associated with the conformational changes in α-subunits, particularly, in H0.

3.4 Conclusions and discussion

In this study, we solved the crystal structure of an ATP-independent proteasomal activator, PafE. In spite of a four-helix bundle similar to eukaryotic 11S activators, PafE assembles into rings with 12-fold symmetry. A GXXXGXXXG motif, which has not been found

in proteasomal activators from other domains of life, is essential for the formation of this dodecameric ring structure. The hydrophobic residues lining the axial channel of PafE are important for PafE-dependent degradation. We determined that PafE has an extended C terminus, short deletion of which resulted in enhanced activity of PafE *in vitro* and *in vivo*. We also demonstrated that PafE induces conformational changes in α-subunits and opens the gate of *Mtb* 20S CP for substrate entry.

Given that previously characterized proteasomal activators are generally heptameric with the exception of hexameric ATP-dependent activators, monomeric Blm10, and tetrameric PbaB, the elucidation of PafE structure further confirms the notion that proteasomal activators can have different symmetries. If the seven PafE-binding sites of the heptameric 20S CP are fully occupied, there are five additional tails of PafE still remaining unbound. The excessive binding motifs in PafE probably increase the avidity for what is otherwise a weak binding event. It is also possible that those additional tails may contribute to other interactions important for either binding or degradation. The symmetric mismatch between PafE and 20S CP may result in the off-axis binding of PafE to 20S α-ring. However, the central channel of PafE is large enough to cover the entry gate in 20S α-ring. Owing to the low local resolution of PafE in cryo-EM structure of the $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex, it is difficult to know which seven out of twelve PafE tails participate in 20S binding or if full occupancy of seven binding sites is necessary for gate opening. More biochemical and structural data are needed to further understand the biological foundation of the symmetric mismatch between PafE and 20S CP.

The eukaryotic 19S RP and 11S activators are capable of forming stable complexes with 20S CPs even under stringent purification conditions. However, bacterial proteasome cofactors interact poorly with 20S CPs. The finding that shortened C-termini enhanced the affinity of PafE

to 20S CPs indicates that bacteria adopt suboptimal binding, possible to regulate proteolysis. The weak interaction between PafE and 20S CP may allow PafE to disengage more easily from 20S CPs so that degradation products can be released or Mpa can get access to 20S CPs. Further studies are required to understand the biological significance of the long C-terminal region.

The cryo-EM structure of the $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex reveals a unique gate opening mechanism in the mycobacterial system that is different from archaeal and eukaryotic systems. Typically, the conformational changes in N-terminal reverse turn in archaeal and eukaryotic 20S CPs are absent from *Mtb* 20S CPs. Our cryo-EM structure showed that a unique cation-π interaction formed between Y173 and R26 induces a 3-Å shift of H0 away from central axis that facilitates the opening of 20S gate. A similar rotation of α-subunits is also observed in the binding of PAN C-termini to 20S CPs (Rabl et al., 2008). Our structural study provides a complement for the current models of gate opening by ATP-independent proteasome activators.

This study provided the first structural analysis of a dodecameric proteasome activator and found that hyperactive PafE mutants will be useful for the exploration of the structural details of activator-20S CP interaction. The structural and biochemical studies of PafE-20S CP system indicate a potential mechanism of PafE-mediated degradation: disordered protein substrates are accommodated in the hydrophobic central channel of PafE and access to degradation chamber of 20S CP upon PafE binding to α-ring. The specfic interaction between PafE C-tails and 20S CP causes the conformational changes in α-subunits that induce disordering of the gate (**Fig 3-12**). Importantly, this study will further our understanding of *Mtb* physiology, and may provide some structural hints for designing drugs targeting the *Mtb* proteasomal system.

3.5 Accession codes

The coordinate and structure factor files of $PafE_{15\text{-}153}$ and $PafE_{44\text{-}153}$ are deposited in PDB with accession number 5IET and 5IEU, respectively.

3.6 Acknowledgement

I am grateful to Dr. Lin Bai for his efforts on SEC-MALS experiments, phase determination of PafE crystals, and ITC experiments between PafE and 20S CP. I thank Dr. Tong Wang for his support in negative stain EM of PafE and low-resolution cryo-EM studies of the $PafE_{\Delta 155\text{-}166}$:$20S_{WT}$ CP complex. I also thank Dr. Jordan B. Jastrab for his efforts on *Mtb* growth conditions and heat-stress sensitivity assay as well as *in vitro* peptide degradation assay of PafE with C-terminal truncations. I am grateful to Dr. Lin Bai, Dr. Huilin Li, and Dr. K. Heran Darwin for their efforts on data analysis and writing the paper.

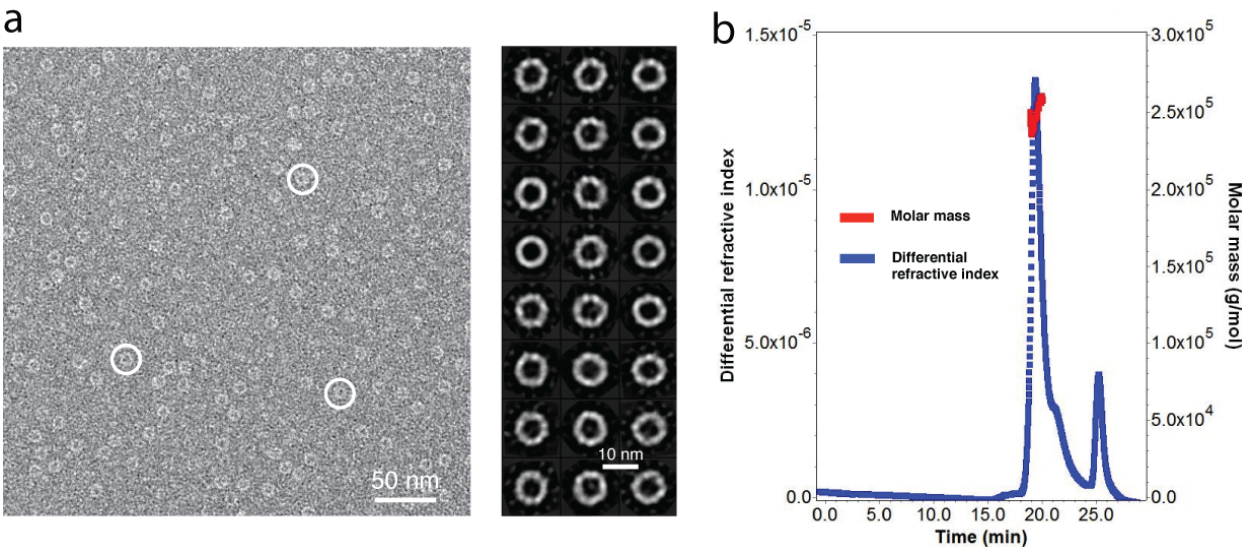
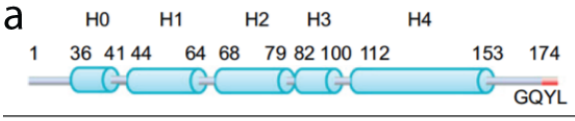


Figure 3-1. PafE forms oligomeric rings

(a) Right panel shows a raw EM image of negatively stained PafE particles. Three white circles mark typical PafE rings. The left panel shows 24 reference-free 2D class averages of PafE ring. **(b)** PafE proteins are characterized by SEC-MALS.

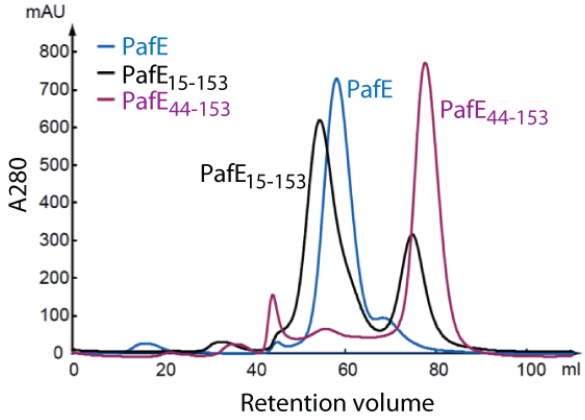
a



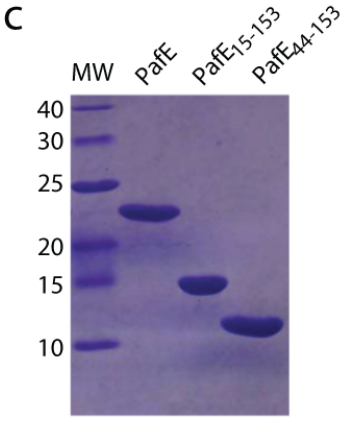
Construct	Yield	SEC*	Crystal
1-174	High	One peak	>7 Å
35-174	Low	-	-
1-153	High	Unstable	-
8-153	High	One peak	No crystal
15-153	High	Two peaks	Large peak <3 Å
22-153	Low	-	-
28-153	High	One peak	No crystal
35-153	Low	-	-
44-153	High	One peak	<3 Å

* SEC was performed using a Superdex 200 column

b



c



d

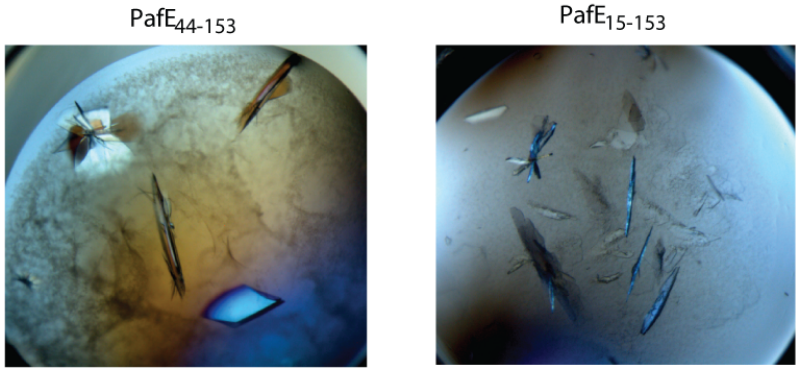


Figure 3-2. Crystallization screen of PafE truncations

(a) Top: PafE is predicted to contain four main α-helices. The C-terminal 21-residue peptide contains a GQYL motif (red). Bottom: expression, purification, and crystallization profiles of different PafE truncations. **(b)** SEC profiles of two PafE truncations used for crystallization. **(c)** SDS-PAGE analysis of two PafE truncations used for crystallization. **(d)** Crystals formed by two PafE truncations.

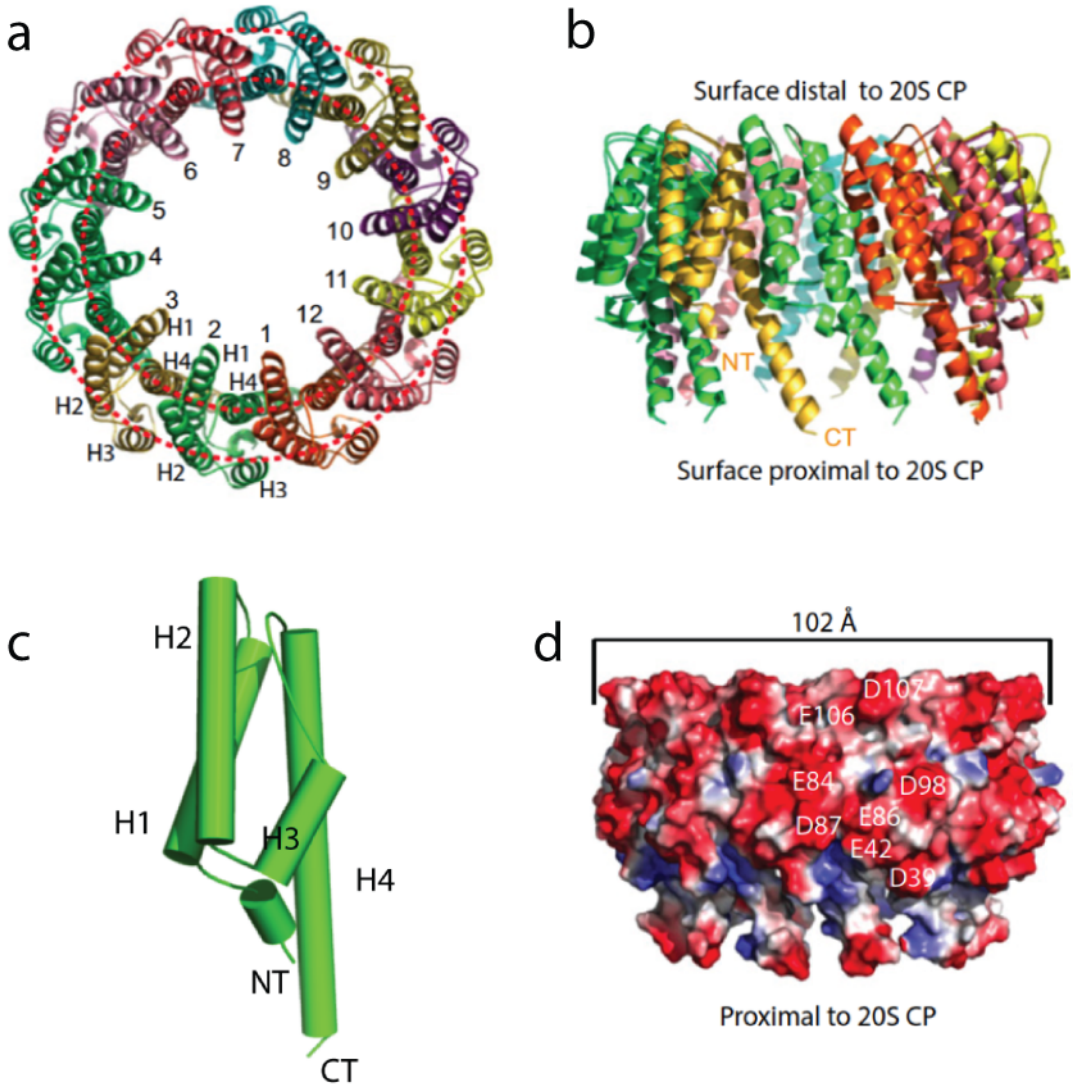


Figure 3-3. PafE shows dodecameric ring structure

(a) Top view of dodecameric PafE$_{15\text{-}153}$ crystal structure. **(b)** Side view of PafE$_{15\text{-}153}$ ring structure. **(c)** Structure of PafE$_{15\text{-}153}$ protomer in cartoon view. **(d)** Surface potential of PafE$_{15\text{-}153}$. The positive and negative charges are colored blue and red, respectively.

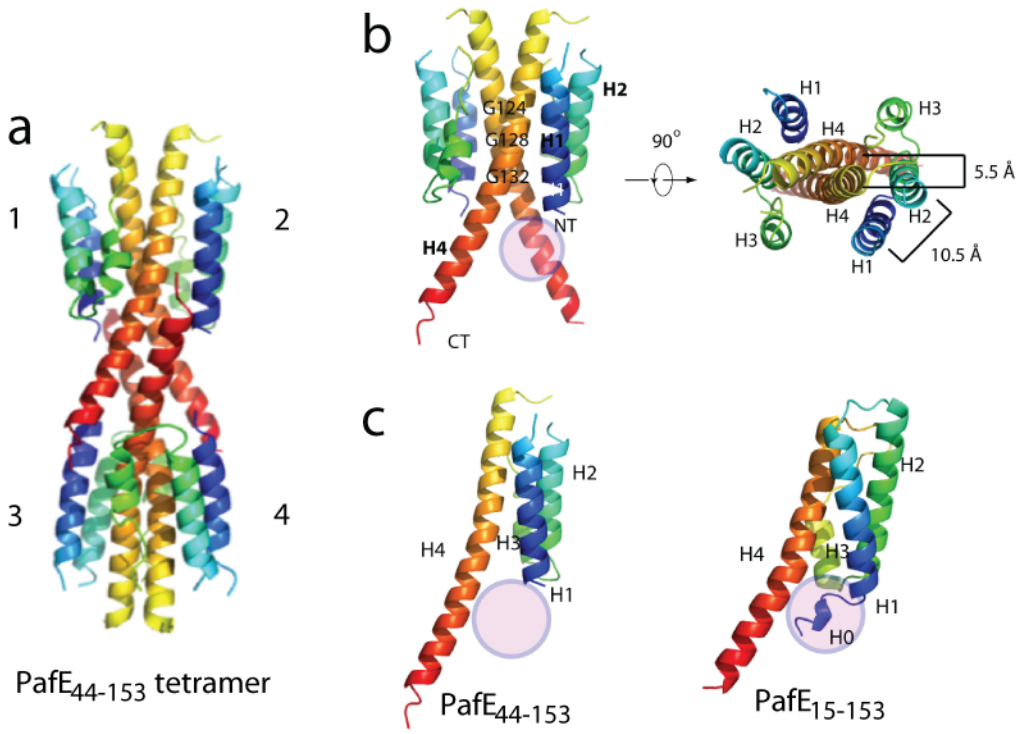


Figure 3-4. Crystal structure of PafE$_{44-153}$ is a tetramer

(a) The tetramer structure in rainbow display from N-terminal blue and C-terminal red. **(b)** Side and top views of a dimer. The tight packing of the two H4 helices is via the double GXXXG motif. **(c)** Comparison of monomer structures between PafE$_{44-153}$ and PafE$_{15-153}$. The H0 of PafE$_{15-153}$ highlighted in circle prevents the formation of the artificial dimer.

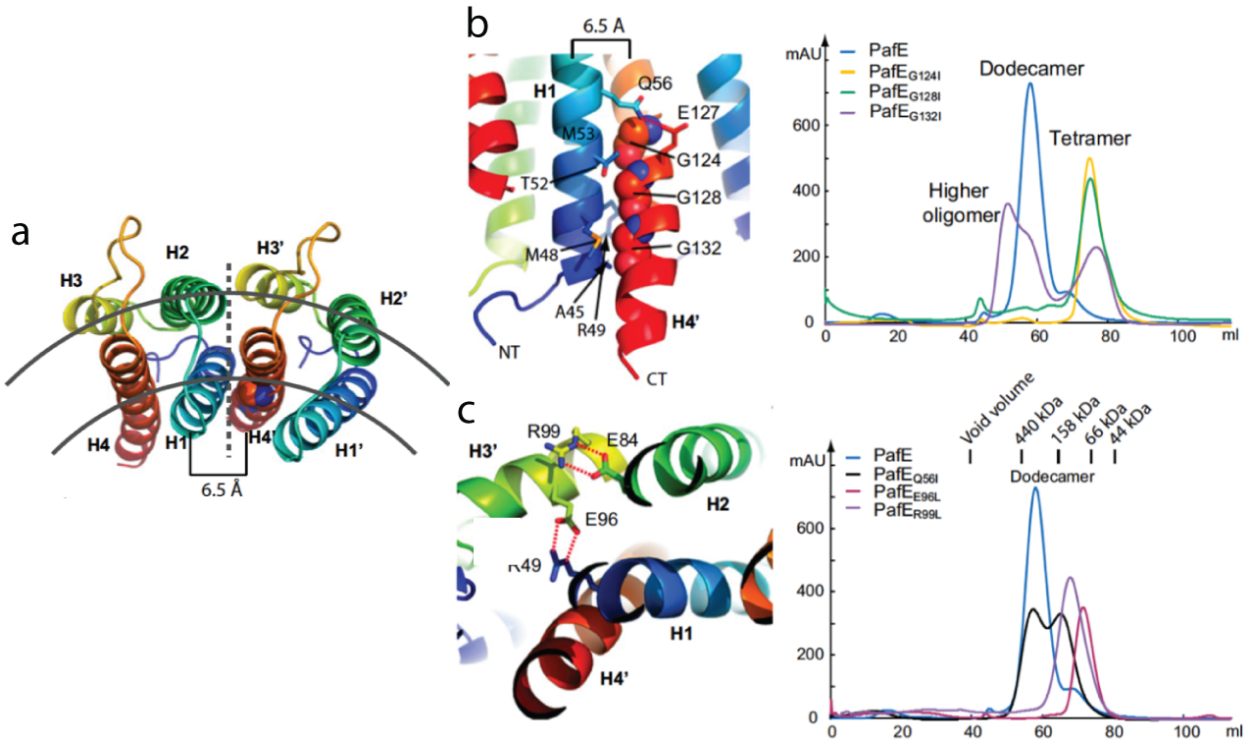


Figure 3-5. Integrity of PafE dodecamer depends on tight interactions between subunits

(a) Top view of two neighboring PafE monomers in the dodecameric ring in rainbow cartoon representation. Two concentric arcs line the inner and outer shells of the ring. The dashed line marks the subunit interface. **(b)** Left panel shows the tight H1/H4 packing via the double GXXXG motif. The three glycines are shown as spheres. Gln56 and Glu127 form H-bonds. Right panel shows the SEC profiles of WT and three Gly-mutated PafE proteins. **(c)** Left panel shows the inter-subunit interaction in the outer shell. Arg99 and Glu84 form two H-bonds, and Arg49 and Glu96 form two H-bonds. Right panel shows the SEC profiles of PafE with single site mutation in above residues in comparison with WT PafE. Size standards are marked between the two SEC panels.

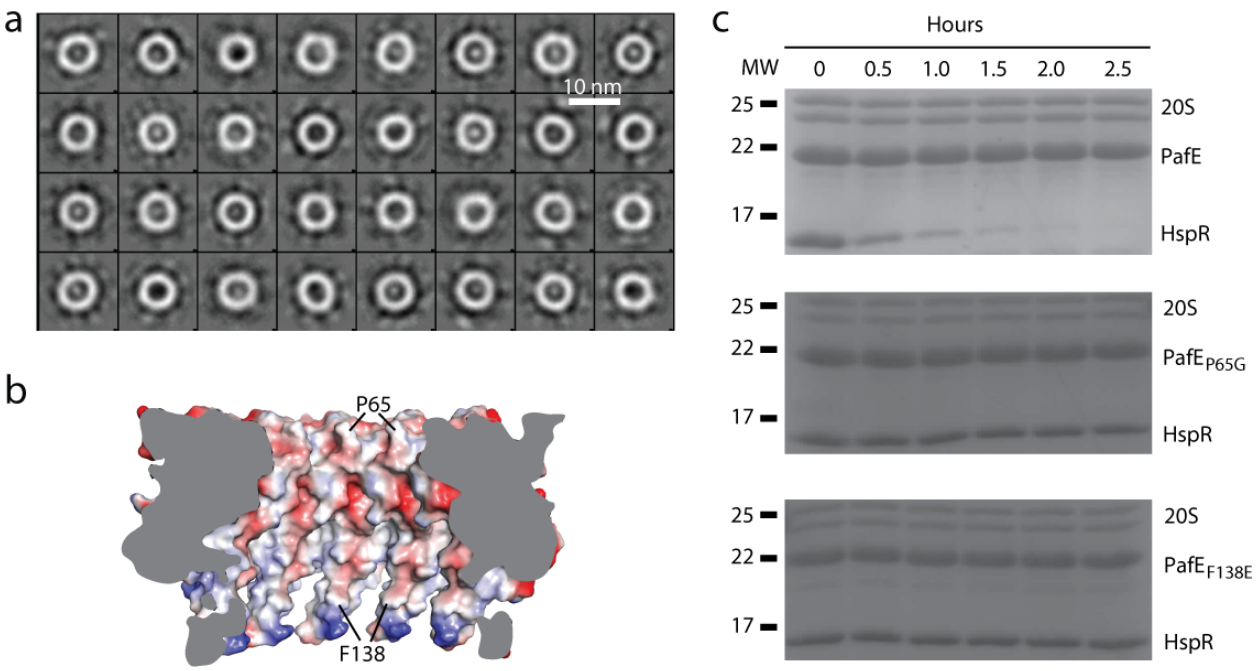


Figure 3-6. The hydrophobic central channel is important for PafE-dependent degradation

(a) Representative reference-free 2D class averages of negatively stained ParP:PafE particles expressed and purified from E coli. **(b)** Surface potential of central channel of PafE ring. The positive and negative charges are colored blue and red, respectively. **(c)** PafE proteins with P65G or F138E mutations losses its ability to promote HspR degradation by 20S CP. Recombinant 20S CP-His6, HspR-His6, and either wild-type His6-PafE or mutants were purified from E coli, mixed, and incubated in room temperature. Aliquotes were removed at indicated time pionts and analyzed by 15% (w/v) SDS-PAGE. The degradation of HspR was monitored by Coomassie Brilliant Blue staining.

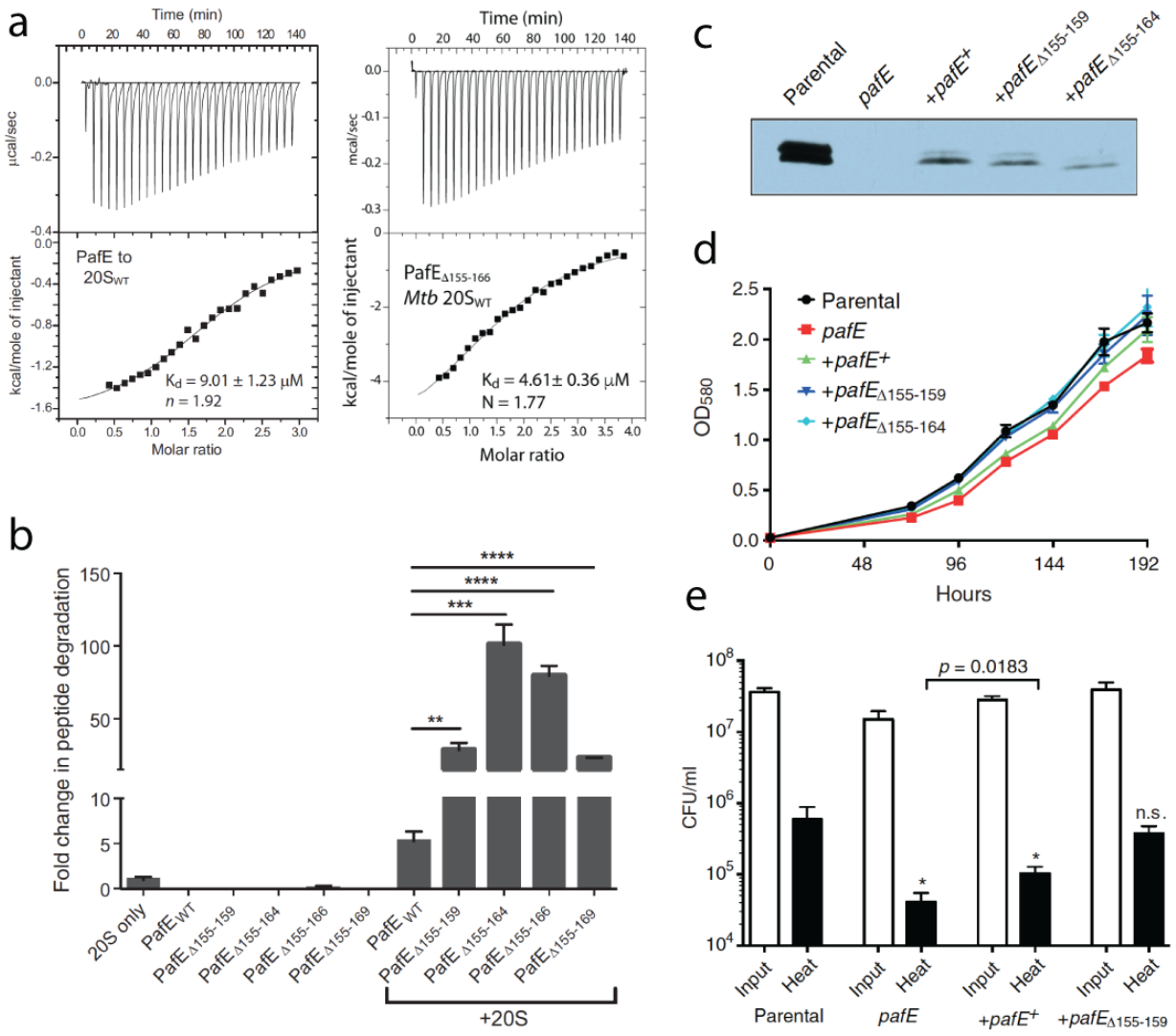


Figure 3-7. C-terminal truncation leads to enhanced PafE function *in vitro* and *in vivo*

(a) The left and right panels show raw calorimetry data (upper) and nonlinear least-square fitting of data points (lower) when PafE and PafE$_{Δ155-166}$ were injected at 25°C in WT 20S CPs, respectively. **(b)** PafE mutants with shortened C terminus have dramatically increased 20S activation capacity. Peptide degradation assays using PafE truncation mutants. **(c)** Complementation of an *Mtb pafE* mutant partially restores PafE protein levels. A single copy of WT or truncated *pafE* was introduced into chromosome. Total cell lysates were analyzed by immunoblotting using antibodies against PafE. **(d)** PafE variants with shortened C terminus complement the growth defect of a *pafE* mutant. The indicated strains were diluted in triplicate to OD_{580} = 0.025, and OD_{580} was measured at the indicated time points. **(e)** A truncated PafE variant fully complements the heat-shock resistance of a *Mtb pafE* mutant. The indicated strains were diluted in triplicate to OD_{580} = 0.08, incubated at 45°C for 24 h, and inoculated onto Middlebrook 7H11 agar to enumerate surviving bacteria. Statistical analysis was done by nonparametric Student's *t* test. **P* <0.05; ***P* <0.005; ****P* <0.0005; *****P* <0.0001.

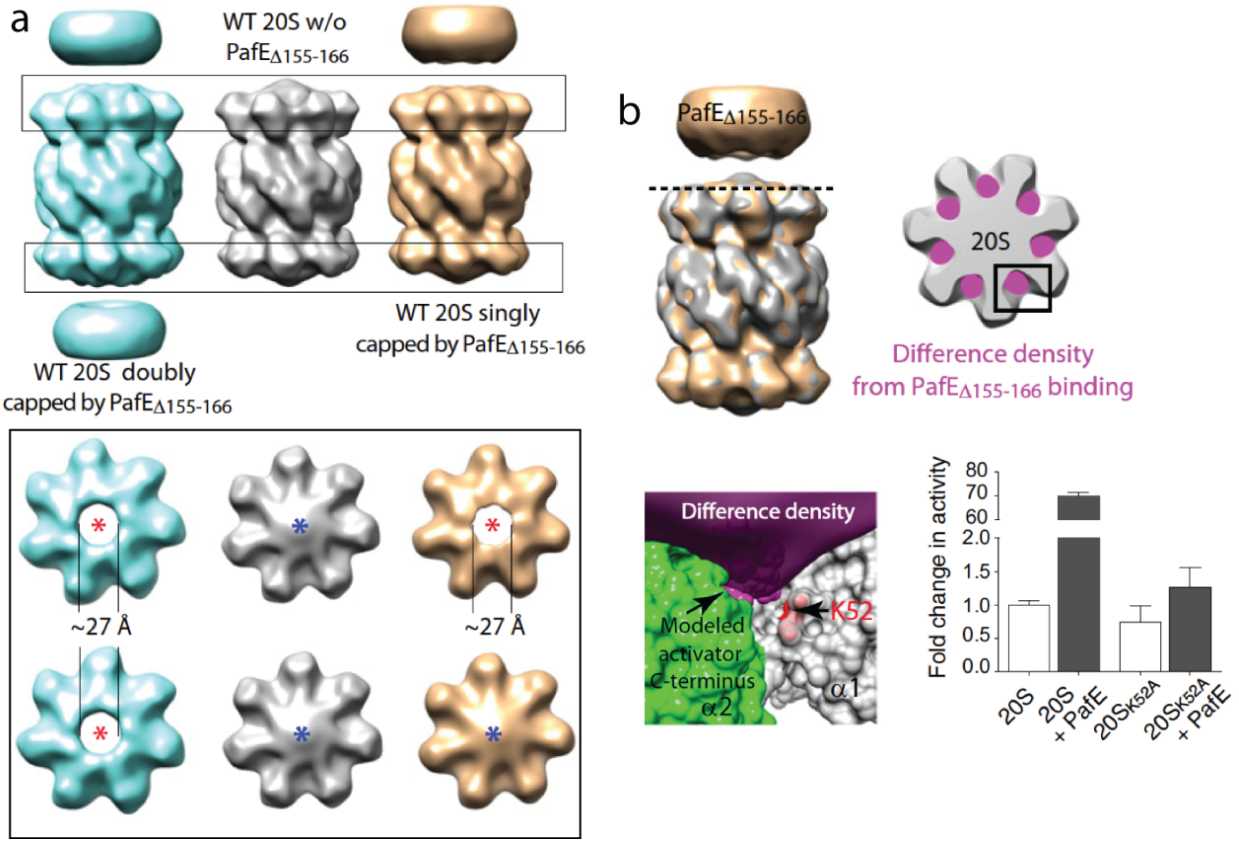


Figure 3-8. Cryo-EM structures of *Mtb* PafE$_{\Delta 155\text{-}166}$:20S$_{WT}$ complexes

(a) Top: surface-rendered side views of the three complexes that coexisted in solution: *Mtb* 20S$_{WT}$ CP capped by PafE$_{\Delta 155\text{-}166}$ at both ends (left, light green); *Mtb* 20S$_{WT}$ CP alone (center, gray); and *Mtb* 20S$_{WT}$ CP capped by PafE$_{\Delta 155\text{-}166}$ at one end (right, brown). The maps are low-pass-filtered to ~12 Å. Bottom: end-on views of α-rings of proteasome complexes. The substrate entrance in the α-ring is open with a diameter of ~27 Å at the PafE-capped end (left and upper right) and closed at the capped end (center and lower right). **(b)** Top left: superimposition of 3D map of *Mtb* 20S$_{WT}$ CP alone (gray) to that of 20S$_{WT}$ CP capped by PafE at one end (brown). Top right: a section of 3D difference map between 20S with and without PafE (magenta) in top view. The dashed line marks the sectioning plane in the left panel. Bottom left: The PA26 C-terminal motif (PDB 1FNT, magenta spheres), modeled in the predicted *Mtb* 20S CP activation pocket between two α-subunits (PDB 3MI0, green and gray surface). The difference density, as shown in C, is rendered semitransparent (magenta) for clarity. K52 in the right α-subunit is shown as red sphere. Bottom right: PafE-activated peptide degradation activities of *Mtb* 20S CPs with and without PrcA$_{K52A}$ mutation.

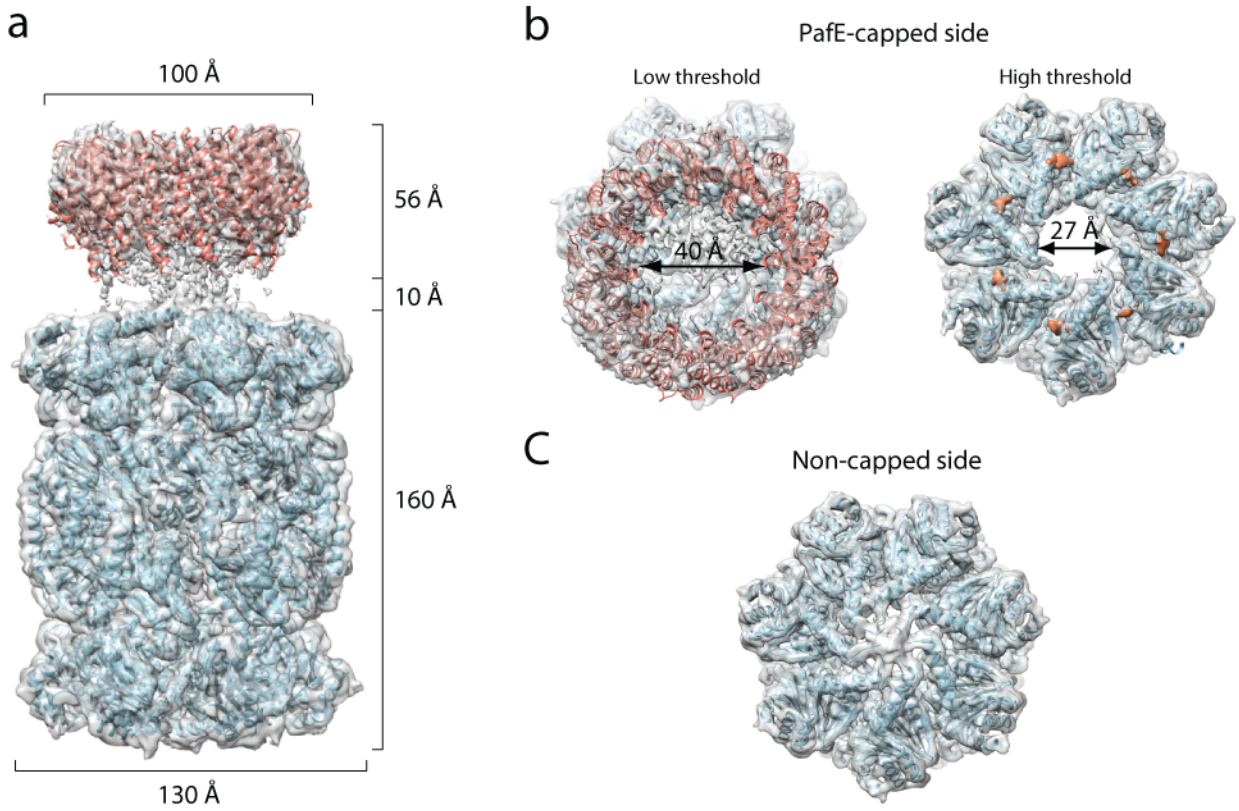


Figure 3-9. High-resolution cryo-EM structure of the PafE$_{\Delta 155\text{-}166}$:20S$_{WT}$ CP complexes

(a) Side view of the PafE$_{\Delta 155\text{-}166}$:20S$_{WT}$ CP complex. The crystal structure of PafE (salmon, PDB: 5iet) and refined *Mtb* 20S CP model (cyan, PDB: 3mi0) were docked into the reconstruction map. **(b)** Top view of the PafE$_{\Delta 155\text{-}166}$:20S$_{WT}$ CP complex in **(a)**. Left: under low threshold, PafE density and discontinuous density between PafE and 20S gate were observed; Right: under high threshold, the gate of 20S CP upon PafE binding was definitely open. The density of PafE C-tails were colored in salmon. **(c)** The bottom view of the map in **(a)** under high threshold.

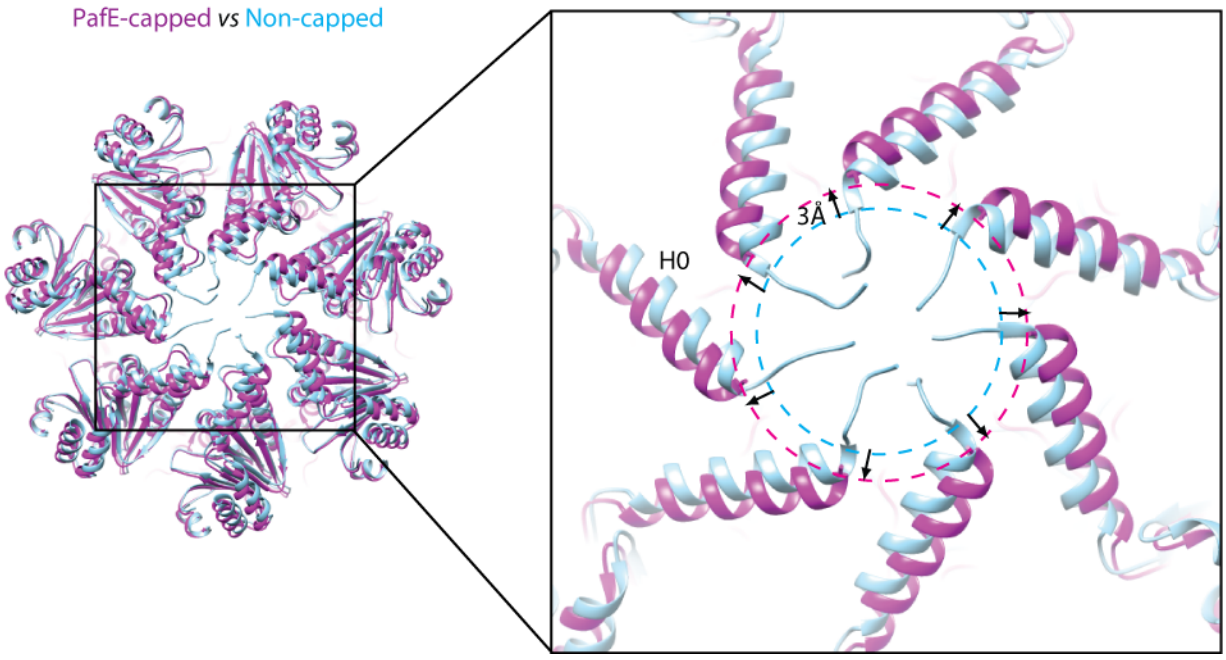


Figure 3-10. Conformational changes in 20S α-ring upon PafE binding

Superposition of PafE-capped α-ring (top side, in magenta) to non-capped α-ring (bottom side, in cyan). Translational shift of H0 away from the central axis is shown in the inset window.

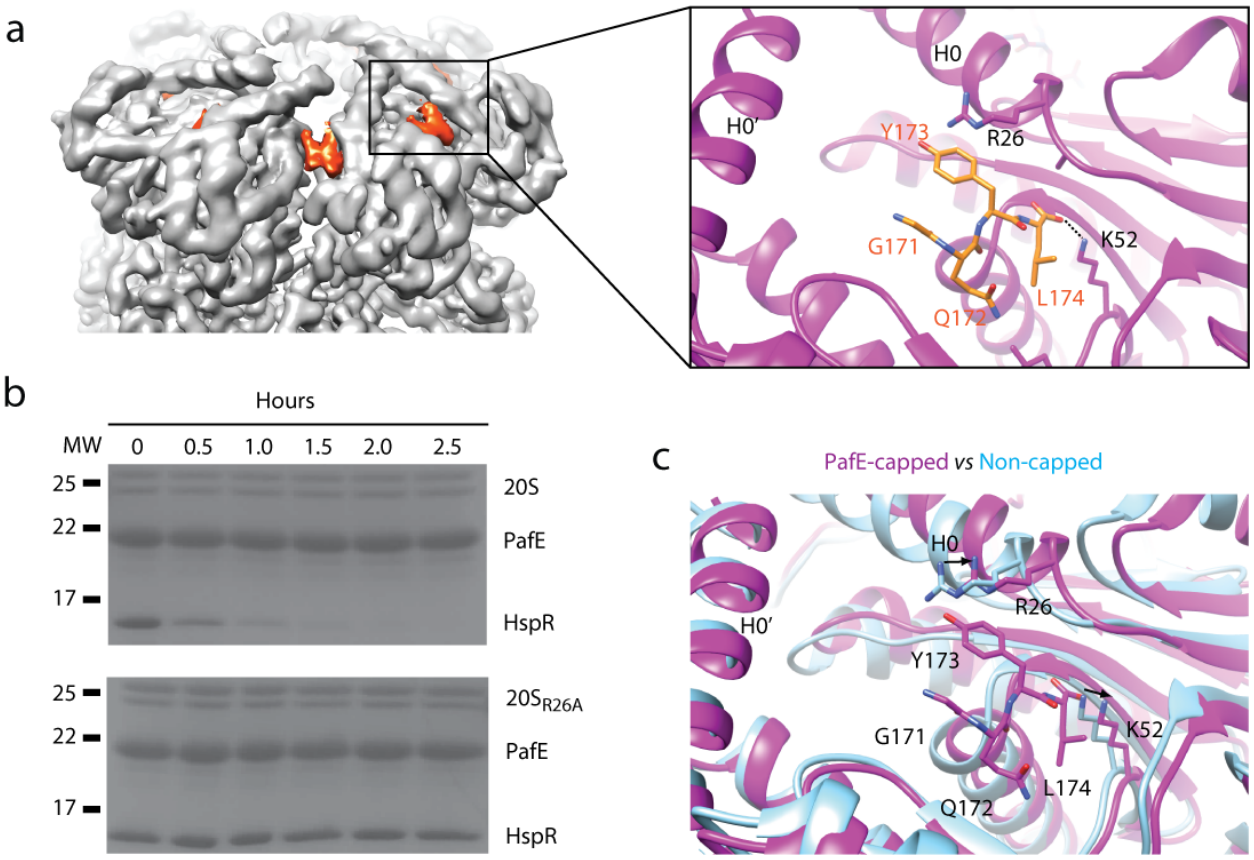


Figure 3-11. Docking of the GQYL motifs induces conformational changes in α-ring

(a) Left: α-ring region of 3.4-Å EM reconstruction of PafE-bound 20S CP with D7 symmetry. The density of PafE GQYL motifs is colored in salmon. Right: zoom-in view of the interaction between PafE GQYL motif (salmon) and 20S α-subunit (purple). The side chain of Y173 and the carbonxyl group of L174 in PafE C-tail interact with side chain of R26 and K52 in 20S α-subunit, respectively. **(b)** PafE can promote the degradation of HspR degradation by wild-type 20S CP rather than 20SR26A. Recombinant His_6-PafE, HspR-His_6, and either wild-type 20S CP-His_6 or $20S_{R26A}$ CP-His_6 were purified from *E. coli*, mixed, and incubated in room temperature. Aliquotes were removed at indicated time pionts and analyzed by 15% (w/v) SDS-PAGE. Degradation of HspR was monitored by Coomassie brilliant blue staining. **(c)** The both-end-capped 20S CP model (purple) was superimposed to the one-end-capped 20S CP model (cyan) shown in **Figure 3-9**. Alignment of one α-subunit in the non-capped end was shown.

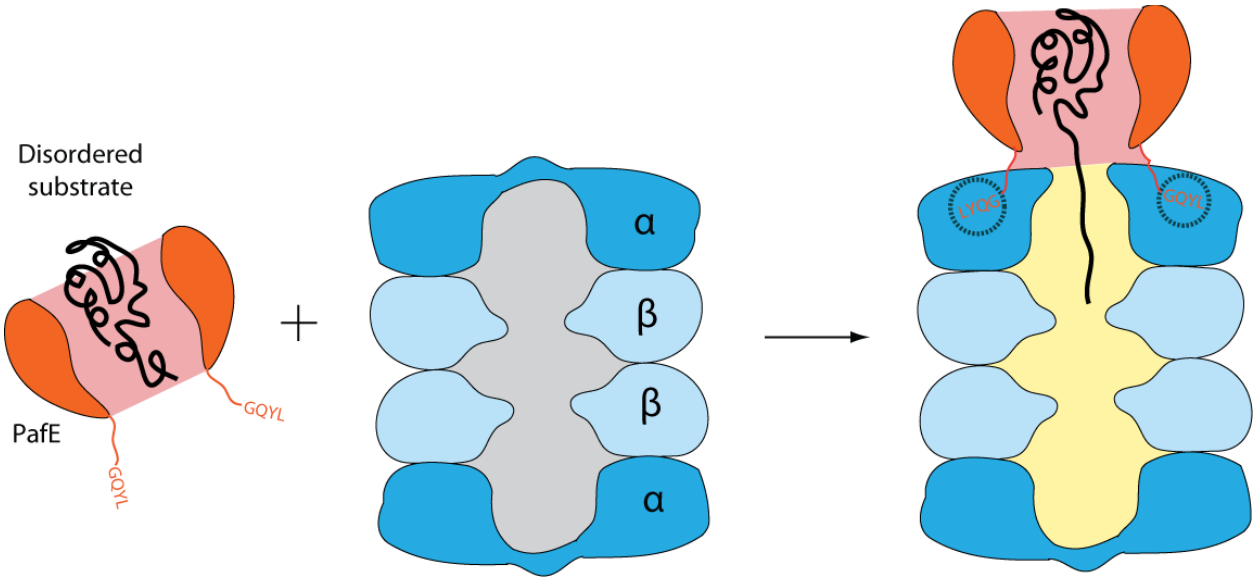


Figure 3-12. A proposed model of PafE-mediated proteasome degradation.

PafE possesses a central channel with a diameter of 40 Å that accommodates disordered protein substrates for degradation. The C-terminal GQYL motifs of PafE dock to α-ring and induce the gate opening. The disordered substrate in PafE can access to the proteasomal chamber for cleavage.

Table 3-1 Data collection and refinement statistics for $PafE_{15\text{-}153}$ and $PafE_{44\text{-}153}$

	$PafE_{\Delta N14\Delta C21}$ (dodecamers)		$PafE_{\Delta N43\Delta C21}$ (tetramer)	
	Pt^{2+} derivative	Se-Met	Se-Met	Native
Crystal form				
Space group	P622	P622	$P2_1$	$C222_1$
Unit cell dimensions (Å)				
a	100.75	100.52	36.85	47.01
b	100.75	100.52	128.62	63.36
c	215.67	229.19	38.63	129.3
α	90°	90°	90°	90°
β	90°	90°	104.94°	90°
γ	120°	120°	90°	90°
Data collection				
Wavelength (Å)	1.0697	0.9793	0.9791	1.1
Temperature (K)	100	100	100	100
Resolution (Å)	42.0 - 3.60 (3.79 - 3.60)	42.0 - 2.88 (3.03 - 2.88)	38.0 - 3.32 (3.42 - 3.32)	37.75 - 2.8 (2.86 - 2.8)
Unique reflections	8081(1115)	16296(2291)	5,159(397)	5,953(291)
Redundancy	18.2(16.6)	19.3(18.4)	6.0(5.2)	6.0(4.3)
*R*merge (%)	21.9(42.5)	17.6(43.4)	8.3(21.9)	7.6(39.2)
Completeness (%)	99.9(99.9)	99.9(99.9)	96.5 (95.0)	99.7(96.7)
I/σ	11.8(7.2)	15.7(7.4)	13.8 (4.6)	6.1(4.8)
Model statistics				
Model composition				
Non-hydrogen atoms		1866		1562
Protein residues		222		195
Refinement				
Resolution (Å)		42.0 - 2.88		37.75 - 2.8
R_{cryst} / R_{free}		0.196 / 0.215		0.258 / 0.282
Test reflections (%)		10		10
Average B-factors		25		83
R.m.s deviations:				
Bond lengths (Å)		0.006		0.013
Bond angles (°)		1.048		1.4
Ramachandran plot				
Favored (%)		98.62		96.17
Allowed (%)		1.38		3.83
Outlier (%)		0		0

Table 3-2 EM data collection and refinement statistics

	$PafE_{\Delta 155\text{-}166}$-bound 20S CP
Data Collection	
Pixel size (Å)	1.09
Defocus range (μM)	1.0-3.0
Voltage (kV)	300
Electron dose ($e^-/Å^2$)	49
Refinement	
Particles	51,091
Resolution range (Å)	171.5-3.42
Symmetry	D7
Model composition	
Non-hydrogen atoms	46,618
Protein residues	6,174
RMSD	
Bonds	0.007
Angles	1.210
Dihedrals	11.833
Ramachandran plot (%)	
Favored	97.21
Allowed	0.90
Outliers	0.00
Rotamer outliers	0.90
MolProbity validation	
Clash score	14.48
Average B factor	81.31

Chapter 4 Structural analysis of proteasomal ATPase Mpa in *Mycobacterium tuberculosis*

This chapter has been published as:

Yujie Wu[1], Kuan Hu[1], Defeng Li, Lin Bai, Shaoqing Yang, Jordan B. Jastrab, Shuhao Xiao, Yonglin Hu, Susan Zhang, K. Heran Darwin, Tao Wang, Huilin Li. (2017) *Mycobacterium tuberculosis* proteasomal ATPase Mpa has a β-grasp domain that hinders docking with proteasomal core protease. *Mol Microbiol.* [1]Equal contribution.

Author contribution: K.H.D, J.B.J, T.W., and H.L. designed research. Y.W., D.L, Y.G, Y.H., S.Y., and T.W. solved the Mpa structure. K.H. and L.B. performed in vitro biochemical studies. J.B.J. and S.Z. did the protein accumulation assays and data analysis. S.Z. did the native gel analysis of Mpa mutants. K.H., L.B, K.H.D., T.W., and H.L. analyzed the data and wrote the paper with input from all authors.

4.1 Abstract

The proteasome system is essential for the virulence of *Mtb*. A key component of this system is the proteasomal ATPase Mpa, which recognizes, unfolds, and translocates protein substrates into the proteasome CP for degradation. The C-terminal GQYL motif of Mpa, similar to the HbYX motif in eukaryotic proteasomal ATPase, is able to independently interact with 20S CPs. However, Mpa hardly binds to *Mtb* 20S CPs *in vitro*. Here, we determined crystal structures of C-terminal part of Mpa in apo and ADP-bound forms. Surprisingly, the structures revealed a ubiquitin-like β-grasp domain that precedes the GQYL motif. This domain, which is unique in

bacterial proteasomal ATPases, buries the carboxyl terminus in the central channel and hinders the interaction of Mpa with 20S CP.

4.2 Introduction

Like other proteasomal ATPases, Mpa proteins form hexamers with several distinct domains (Finley et al., 2016). First, the N-termini of a Mpa protomer forms α-helix, which dimerizes with the helix of an adjacent protomer to form a coiled-coil structure. This pairing of adjacent helices results in the formation of three extended dimers, breaking the six-fold symmetry at the N-termini (Djuranovic et al., 2009; Wang et al., 2009; Zhang et al., 2009a). The coiled-coil domain is responsible for recognition of Pup: upon binding to Mpa, Pup, which is mostly disordered in free form, is induced to form helix (Wang et al., 2010). Deletion of the coiled-coil region abolishes the ability of Mpa to unfold a model substrate (Striebel et al., 2010).

Following the coiled-coil domain are two OB folds, which form a double-ring structure. In contrast, the archaeal PAN and the eukaryotic Rpt1-6 have only one single OB fold. The OB domain alone is sufficient to maintain the hexameric state of Mpa in the absence of the N-terminal coiled-coil (Wang et al., 2009).

The C-terminal part of Mpa is an AAA domain, which hydrolyzes ATP to drive the unfolding and translocation of substrates. The GQYL motif is located at the C-terminus, which is functional equivalent of the HbYX motif found in the archaeal and eukaryotic proteasomal ATPases (Darwin et al., 2005; Jastrab et al., 2015). Mutagenesis studies showed that G, Y, or L, but not Q of the motif is associated with the proteasome activation (Delley et al., 2014; Jastrab et al., 2015). Unlike Rpt1-6 hexamer that is co-purified robustly with 20S CPs, Mpa binds weakly

to 20S CPs (Smith et al., 2011; Wang et al., 2009). Sequence alignment showed that Mpa has a extra region preceding the GQYL motif compared with PAN and Rpt proteins. The structure and function of this uncharacterized region are still unknown.

Here, we determined the crystal structure of the C-terminal part of Mpa hexamer, containing the OB domain and the AAA domain. The structure reveal a three-ring architecture composed of a double OB ring on top of an ATPase ring. Furthermore, the uncharacterized region forms a ubiquitin-like β-grasp domain prior to the C-terminus that obstructs the GQYL motif from docking with the 20S CP. Our study revealed an important structural feature that distinguishes bacterial proteasome activators from those found in other domains of life.

4.3 Results

4.3.1 Crystal structure of Mpa shows a three-ring architecture

Tao and Shaoqing determined Mpa structures in apo and ADP-bound forms at 3.6 and 2.9 Å, respectively. The apo Mpa crystal structure was determined in $P2_12_12_1$ space group with two distinct hexamers in an asymmetric unit, while the ADP-bound Mpa crystal structure was solved in P321 space group with two protomers in an asymmetric unit (**Table 4-1**). The apo structure was largely superimposable with the ADP-bound structure, with an RMSD value for each protomer ranging from 0.75 to 1.95 Å (Wu et al., 2017).

Mpa hexamer without the coiled-coil domain is about 100 Å high and 130 Å wide. The double OB ring on the distal face and the AAA ring on the proximal face are co-axial. As reported before, the double OB ring has an outer diameter of 70 Å and an axial channel of 25 Å,

with each OB domain composed of a five-stranded β-barrel. The loops between A194 and R210 as well as between E316 and A325 are disordered (**Fig 4-1a,b**). Extensive electrostatic interactions were formed between the negatively charged patch (E166, D181, E182, and E183) on the proximal face of the OB2 ring and the positively charged patch (R347, R350, R355, and R357) on the distal face of the ATPase ring (**Fig 4-1c**).

The AAA domain is composed of a large α/β region and a small helical region. The α/β domain consists of a five-stranded β-sheet core flanked by five α-helices on one side and three α-helices on the other, while the small helical domain contains five α-helices (**Fig 4-2a**). The pore loops 1 and 2, which are important for substrate translocation, are disordered. The two AAA subdomains closely resemble those of PAN and Rpt1, with RMSD of 1.77 Å for 153 Cα atoms and 1.60 Å for 152 Cα atoms, respectively.

4.3.2 Mpa has a unique β-grasp-like domain near the C-terminus

Sequence alignment shows that Mpa has an extra region near the C-terminus compared with PAN and Rpt1-6 (**Fig 4-2b**). The extra amino acids, surprisingly, form a β-grasp-like fold, comprising an α-helix and a three-strand β-sheet. The first two β-strands are between α10 and α11 of the small helical domain. The third β-strand and the α-helix are appended to the end (**Fig 4-2c**). The β-grasp domain contributes to the assembly of Mpa hexamer via five H-bonds between α13 of one protomer and β18 of the neighbor: two between S574 and Y583, one between G575 and R585, one between R580 and V582, and one between A571 and A598. The side chains of R580 in the loop between α13 and β18 line the constriction at the exit port (**Fig 4-2d**) (Wu et al., 2017).

4.3.3 Charged axial channel are important for substrate degradation

Unfolding and translocation of substrate occur in the axial channel of ATPase hexamer. Mpa has a vase-like axial channel, with a diameter that varies a lot from the top to bottom (**Fig 4-3a**). The narrowest potion is located in the lower region of OB ring, which reveals a diameter of 15 Å. The widest portion, with a diameter of 37 Å, is in the central part of ATPase domain. Disordered pore loops, which are important for interaction with substrates, are located between the narrowest and widest points of the axial channel.

Many charged residues line the axial channel of Mpa hexamer (**Fig 4-3b**). In order to reveal the importance of these residues on substrate degradation, we carried out the degradation of two proteasome substrates in *Mtb*, malonyl coA-acyl carrier protein transacylase (FabD) and inositol 1-phosphate synthetase (Ino1) *in vivo*. Single or double mutations of selected charged residues were introduced into complementation plasmids with *mpa* expressed form its native promoter. The resulting plasmids were incorporated into the *Mtb* chromosome of an *mpa*-null mutant to determine whether they could restore the degradation of FabD and Ino1. The results showed that most of the mutant proteins failed to restore the degradation of FabD and Ino1, suggesting that these residues are important for translocation of unfolded substrates. Importantly, only one mutation R580L showed less degradation of two substrates, indicating that this residue may be not directly involved in substrate translocation (**Fig 4-3c**).

4.3.4 Nucleotide binding site in the Mpa crystal structure

Every Mpa protomer bound to an ADP molecule in the crystal structure, yielding a stoichiometry of six nucleotides per Mpa hexamer (**Fig 4-4a**). As expected, the nucleotide-

binding site was located between the large α/β subdomain and the small helical domain (**Fig 4-4b**). The phosphate groups of ADP was stabilized in a highly basic cavity formed by backbone amides of G296 and G298, and K299 of the Walker A motif in the α/β subdomain, which explains the previous finding that mutating K299 to Q reduces the affinity of nucleotide to Mpa. I448 and Y452 in the small helical subdomain interact with the adenine group, while L301 and Q520 stabilize the ribose group. The neighboring protomer contributes two closely spaced arginines, R427 and R430, to the nucleotide-binding site. R427 is probably responsible for sensing the γ-phosphate of ATP, while R430 forms a salt bridge with E372 in Walker B motif (**Fig 4-4c**) (Wu et al., 2017).

4.3.5 C-terminal GQYL motif is concealed in Mpa

Physical interaction with 20S CP is required for Mpa to deliver unfolded proteins to cleavage chamber. Eukaryotic proteasome ATPase hexamer, Rpt1-6, has concave surface proximal to 20S CP and three fully exposed HbYX motifs that are radially located at the binding pockets in the α-ring of the 20S CP (**Fig 4-5a**). In Mpa, C-terminal GQYL motifs are essential for activating degradation by the *Mtb* 20S CP. The structure of the C-terminal eight residues that include the GQYL motif is unknown in our Mpa structure. However, the last resolved residue T601 is found located inside the axial channel. Furthermore, a modeled eight-residue peptide extending from T601 would barely emerge from the exit port, making it difficult to reach the binding pocket of *Mtb* 20S CP α-ring (**Fig 4-5a**). Actually, our previous work showed that Mpa interacts with 20S CPs very weakly. In current work, we attempted to quantify the interaction between Mpa and $20S_{OG}$ CPs in the present and absence of ADP or AMPPNP, but detected no

binding signal (**Fig 4-5c**). Therefore, we hypothesize that the weak interaction between Mpa and 20S CP *in vitro* was due to the recessed position of the GQYL motif.

To examine if fully exposed Mpa C-termini bind to 20S CP robustly, we replaced the C-terminal 20 residues of PafE with the last eight amino acids of Mpa to produce a hybrid protein, $PafE_{\Delta 155\text{-}174}$-$Mpa_{602\text{-}609}$. Gel filtration and SDS-PAGE demonstrated that the hybrid protein bound to $20S_{OG}$ CPs well, while a PafE allele either lacking the 20S CP-binding peptide ($PafE_{\Delta 155\text{-}174}$) or with a substitution in the essential tyrosine of the GQYL motif ($PafE_{\Delta 155\text{-}174}$-$Mpa_{602\text{-}609/Y608A}$) did not bind $20S_{OG}$ CP at all (**Fig 4-5b**). This result demonstrated that the Mpa C-termini are capable of mediating a strong interaction with $20S_{OG}$ CPs when they are fully exposed.

We further reasoned that extension of the Mpa C-terminus would expose the GQYL motifs and facilitate binding to 20S CPs. We inserted five residues (GGGGS) preceding the GQYL motif to produce an Mpa variant with 13 unstructured C-terminal residues ($Mpa_{C\text{-}ext}$). ITC demonstrated that $Mpa_{C\text{-}ext}$ was bound to *Mtb* $20S_{OG}$ CPs with an estimated K_d of 2.2 μM (**Fig 4-5c**). These results support our hypothesis that the β-grasp fold domains near the C-termini of the *Mtb* proteasomal ATPase hexamer prevent Mpa binding to *Mtb* 20S CPs by burying the C-terminal GQYL motif.

Next, we wanted to ask if extension of Mpa C-terminus could enhance the degradation of a pupylated substrate by the *Mtb* proteasome. We used pupylated FabD (Pup~FabD) as the substrate in *in vitro* degradation assay. Consistent with the previous study using PanB as the substrate (Striebel et al., 2010), we detected only very low in vitro activity for the WT proteasome in the presence of Mpa, irrespective of a C-extension in Mpa (**Fig 4-6a**). When we used $20S_{OG}$ CPs, we observed degradation of Pup~FabD with WT Mpa (**Fig 4-6a**). Importantly,

$Mpa_{C\text{-}ext}$, but not a Walker B mutant version of $Mpa_{C\text{-}ext}$, significantly accelerated degradation by $20S_{OG}$ CPs (**Fig. 4-6b**) (Wu et al., 2017).

4.4 Conclusions and discussion

To understand the atomic structure and function of ATP-driven proteasomal ATPase in bacteria, we determine crystal structures of a truncated Mpa hexamer in ADP-bound and apo forms. An important finding of this study is the identification of a β-grasp domain near the C-termini of Mpa. The β-grasp domain distinguishes Mpa from its homologues in archaea and eukaryotes and appears to have a negative impact on the interaction between Mpa and 20S CP.

The ITC and in vitro degradation results show that the recessed C-terminal GQYL motif is responsible for the weak interaction between Mpa and 20S CP. The low affinity of Mpa to 20S CP in the absence of pupylated substrates probably allows 20S CP to participate in other biological pathways such as PafE-mediated proteasomal degradation. Structures of Mpa in other functional states are needed to understand how the hidden GQYL motif is exposed for 20S CP binding.

The binding mode of nucleotides to proteasomal ATPases remains controversial. The archaeal PAN hexamer was reported to bind four nucleotides, with two neighboring protomers binding ATP, the next two protomers binding ADP, and the remaining two protomers unoccupied (Kim et al., 2015). A recent cryo-EM structure of human 26S proteasome at an average resolution of 3.5 Å showed Rpt1-6 hexamer fully occupied by nucleotides (Huang et al., 2016). Although the resolution is insufficient to distinguish ATP from ADP, it was suggested that both ATP and ADP are present in the structure, because three Rpts resemble an ATP-bound

conformation, and the other three are more similar to an ADP-bound state. In another cryo-EM structure at a 3.9-Å resolution, all six Rpts were also occupied by nucleotides, with Rpt6 bound by ADP and the other five all bound by ATP (Schweitzer et al., 2016). Different from those above hexamers with asymmetric nucleotide binding, Mpa hexamers in our study show nearly symmetric binding of ADP in all six protomers. It is still unknown if our symmetric structure represents an intermediate state during the nucleotide cycle or a crystallization artifact.

A full-length model of the Mpa hexamer can be built up by combining the previous Mpa_{1-234} hexamer (Wang et al., 2010) with this Mpa_{95-601} hexamer (**Fig. 4-7**). These two models share the same OB ring structure. Despite the unstructured loops at the N-terminus, Mpa has a dramatic height of over 160 Å, approaching the height of the 20S CP, which is 170 Å tall. The coiled-coil is significantly longer in Mpa than in PAN or Rpt1-6. Previous data showed that the coiled-coil is essential for recruiting pupylated proteins (Wang et al., 2010). Further studies are needed to understand the function of the extra OB domain and the β-grasp domain.

4.5 Accession codes

The coordinate and structure factor files of apo and ADP-bound Mpa are deposited in PDB with accession number 5KZF and 5KWA, respectively.

3.6 Acknowledgement

I am grateful to Dr. Yujie Wu, Dr. Tao Wang, and Dr. Defeng Li for his efforts on structure determination of ADP-bound Mpa. I thank Dr. Shaoqing Yang and Dr. Lin Bai for their support

in structure determination of apo-form Mpa. I also thank Dr. Jordan B. Jastrab for his efforts on *in vivo* degradation assay of Mpa mutants. I also thank Susan Zhang for her efforts on *in vitro* degradation assay of wild-type and C-terminus-extended Mpa. I am grateful to Dr. Huilin Li and Dr. K. Heran Darwin for their efforts on data analysis and editing the manuscript.

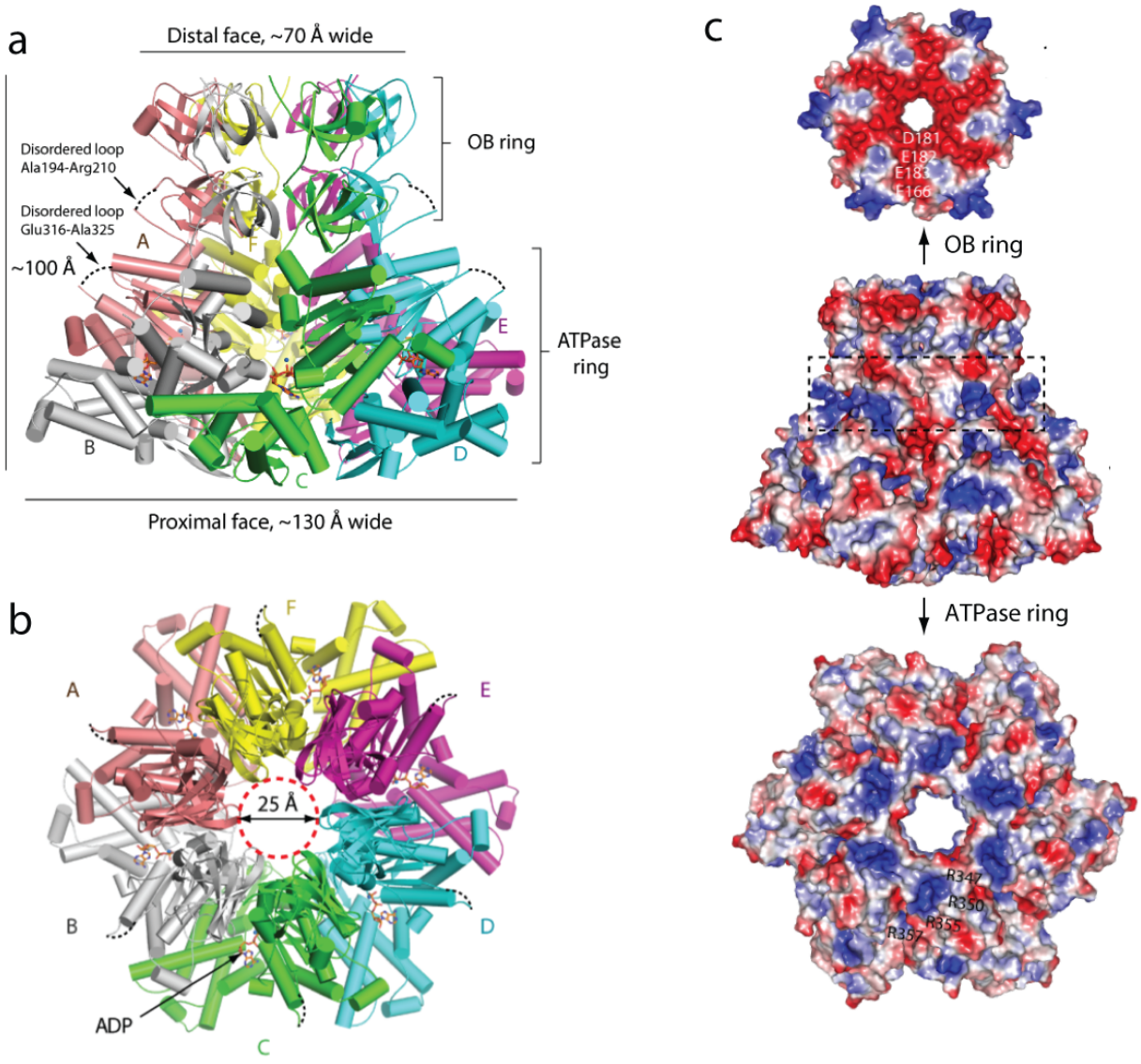


Figure 4-1. Crystal structure of an engineered Mpa hexamer

(a) Side view of ADP-bound hexameric Mpa. The hexamer can be divided into two OB rings and an ATPase ring. Six Mpa protomers labeled A through F are colored differently. ADP molecules are shown in stick form. Distal and proximal refer to relative orientation with respect to 20S CP. The disordered loops are shown as dashed black curves. **(b)** Top view of ADP-bound hexameric Mpa. **(c)** Surface potential of Mpa hexamer. The middle panel shows the side view of Mpa hexamer. The dashed rectangle outlines the interface between OB ring and ATPase ring. Top: the proximal (bottom) OB ring surface. Bottom: the distal (top) surface of ATP ring.

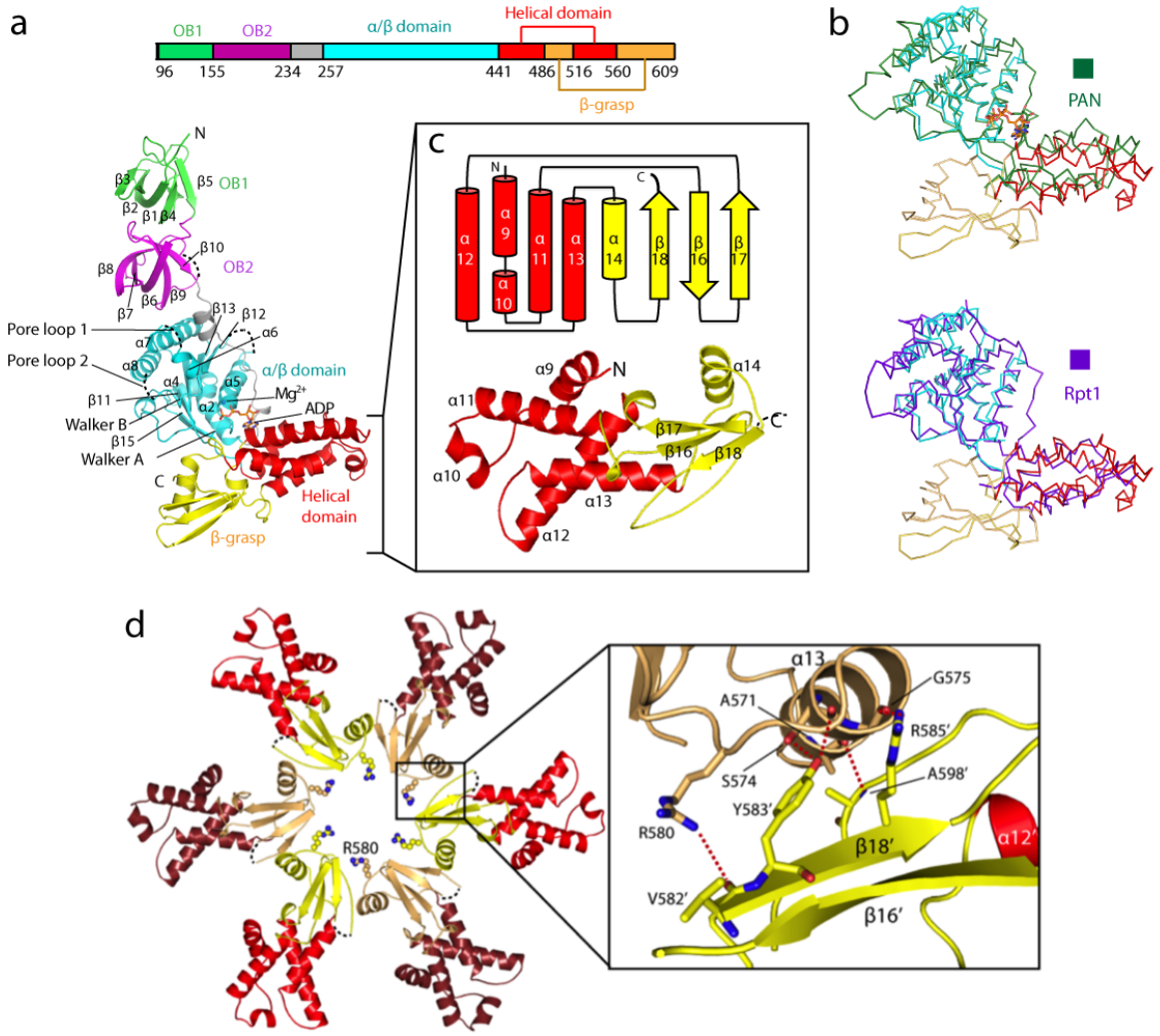


Figure 4-2. Mpa has a unique β-grasp fold near the C-terminus

(a) Top: the domains of Mpa. Bottom: structure of a single Mpa protomer, with OB1 domain in green, OB2 in purple, large AAA subdomain in cyan, small AAA subdomain in red, and β-grasp fold in yellow. ADP molecule is shown in stick form. Disordered loops are shown as dashed black curves. **(b)** Superimposition of ATPase region in Mpa with that in PAN (top, dark green ribbon) or in Rpt1 (bottom, purple ribbon). The β-grasp near C-terminus is found only in Mpa. **(c)** The structure (lower) and topology diagram (upper) of β-grasp fold in Mpa. **(d)** The β-grasp-like domains interact with each other to form the exit port lined by R580. Five hydrogen bonds at the interface between two neighboring β-grasp-like domains are shown as dashed red lines in the inset window.

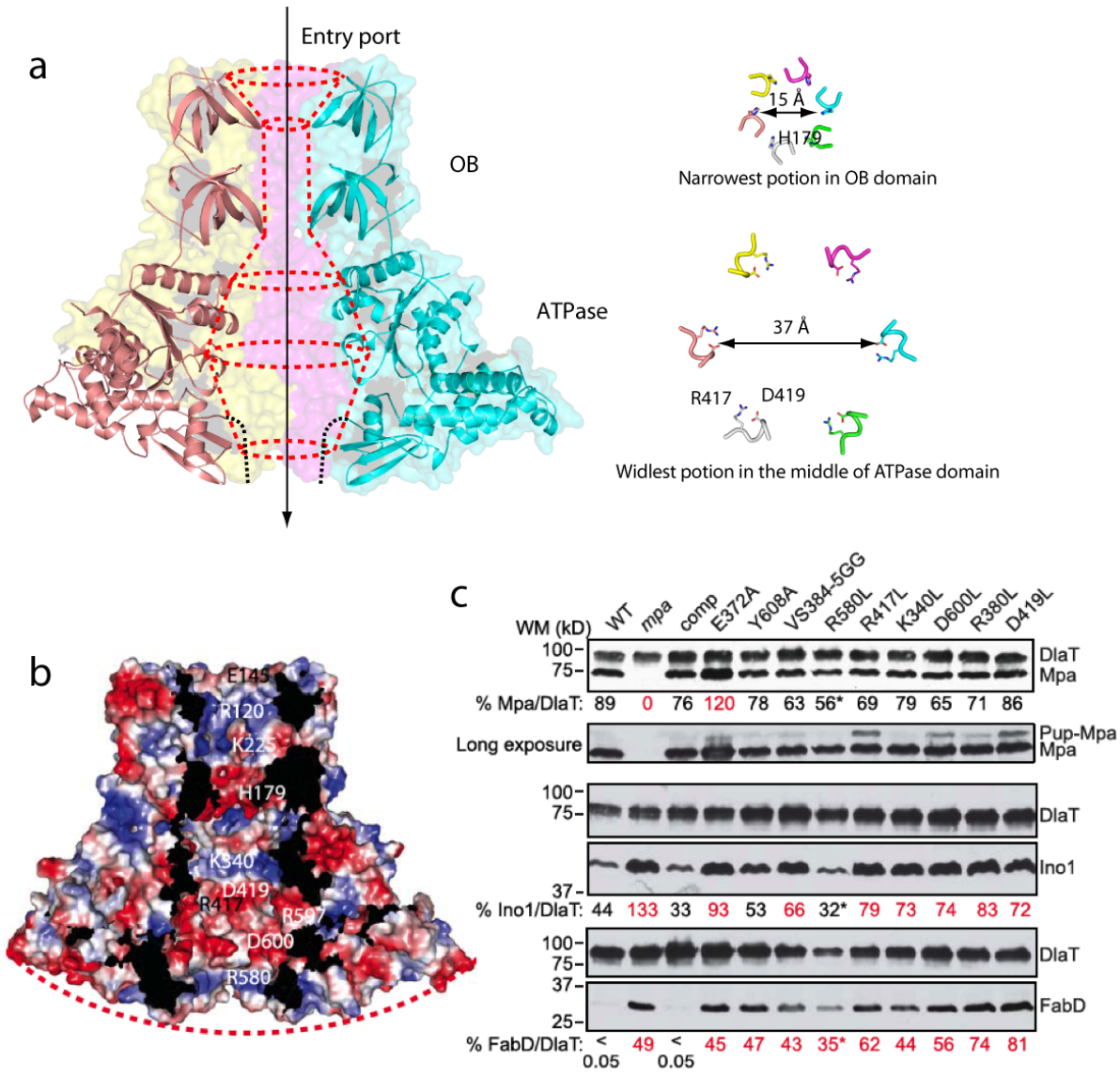


Figure 4-3. Mpa hexamer has a charged axial channel

(a) Mpa hexamer reveals a vase-like central chamber. The entry port and neck are formed by OB ring. ATPase ring forms the widest portion of the channel. The β-grasp fold defines the exit port. The dashed black curves indicate the disordered C-terminal eight residues. **(b)** Surface potential of the translocation chamber of Mpa hexamer. Charged residues are labeled. **(c)** Immunoblot analysis of two proteasome substrates, FabD and Ino1, in *Mtb* strains expressing mutant alleles of *mpa*. Charged residues lining the axial channel of Mpa hexamer were mutated to leucine or glycine. E372A is Walker B motif mutant, and Y608A has a mutation in the essential GQYL motif at the C-terminus of Mpa. DlaT was the loading control. A longer exposure of the immunoblot to Map is shown below the first panel, revealing a potential pupylated species of Mpa in several strains. The amount of Mpa, Ino1, and FabD is shown as a percentage of the DlaT loading control and was quantified from the same immunoblot using Image J software.

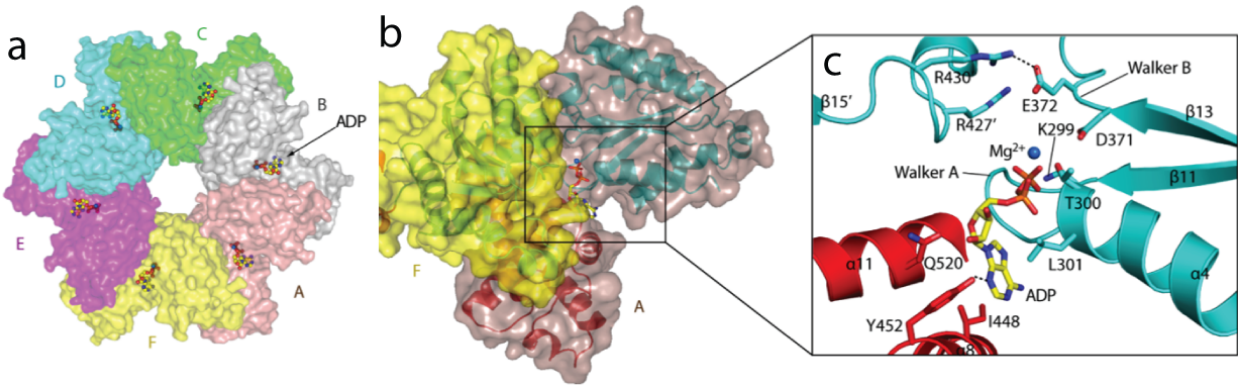


Figure 4-4. The nucleotide-binding site of Mpa hexamer

(a) Surface view of ATPase ring in Mpa hexamer, viewed from the distal face. Each of the six nucleotide-binding pockets is occupied by an ADP molecule. **(b)** A selected region of Mpa hexamer, showing that the nucleotide pocket is formed between the large and small AAA subdomains of one protomer (wheat) and the large AAA subdomain of a neighboring protomer (yellow). **(c)** Interaction between ADP and Mpa. ADP is shown in stick form and Mg^{2+} as a blue sphere.

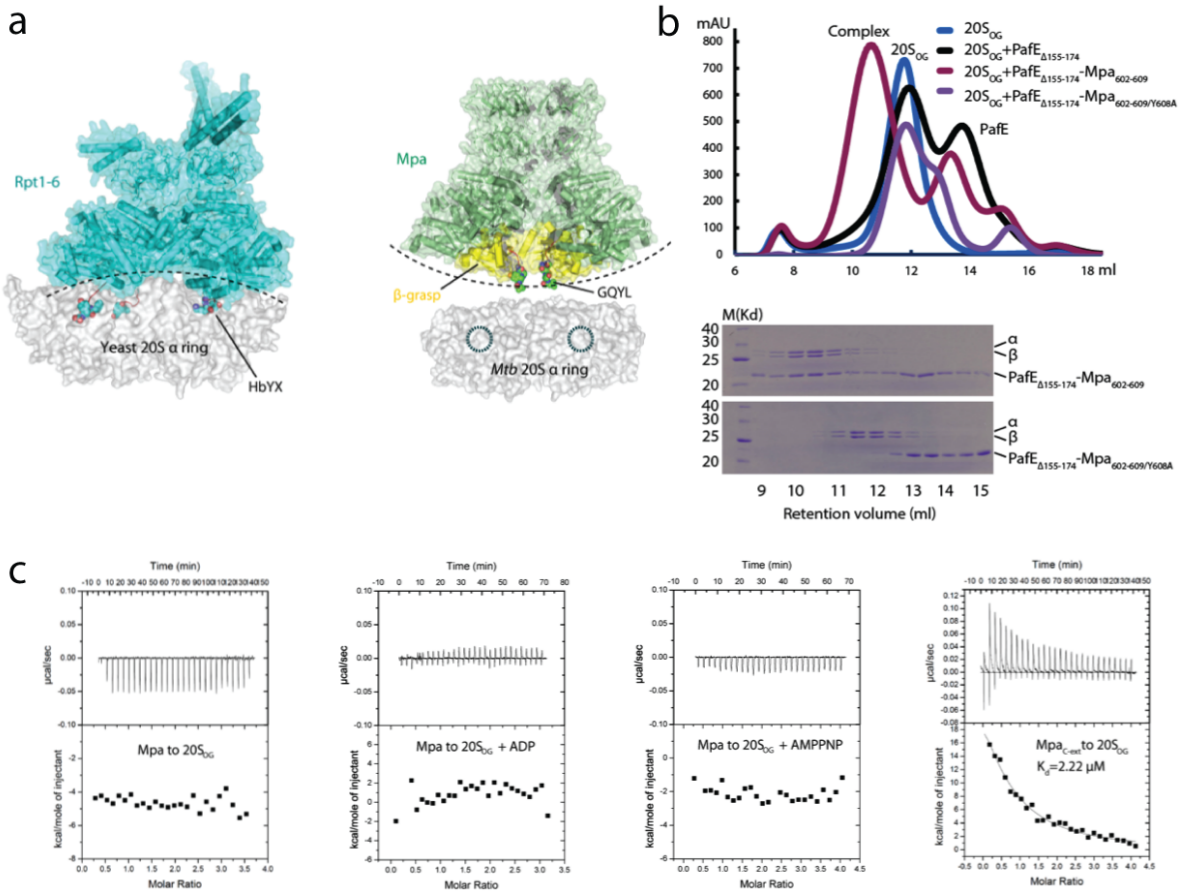


Figure 4-5. C-terminal GQYL motifs are partially hidden inside the axial channel of Mpa

(a) Left: C-terminal HbYX motifs (cyan spheres) in Rpt2, 3, and 5 are fully exposed, facilitating interaction with 20S CP (PDB 4CR2). Right: Last eight disordered residues of Mpa (E602 to L609, red) are modeled in Coot by extending downward from T601. The β-grasp-like domain is colored in yellow. The modeled GQYL is shown as green spheres. Dashed green circles in the *Mtb* 20S CP α-ring denote the binding pockets for Mpa GQYL motifs. **(b)** Gel filtration profiles (upper panel) of $20S_{OG}$, $20S_{OG}$+$PafE_{\Delta155\text{-}174}$, $20S_{OG}$+$PafE_{\Delta155\text{-}174}$-$Mpa_{602\text{-}609}$, and $20S_{OG}$+ $PafE_{\Delta155\text{-}174}$-$Mpa_{602\text{-}609/Y608A}$ and peak fractions in SDS PAGE (lower panel). **(c)** In vitro binding of WT Mpa to *Mtb* $20S_{OG}$ CP as measured by ITC at 25°C in the absence and presence of ADP and AMPPNP. In vitro binding of $Mpa_{C\text{-}ext}$ to *Mtb* $20S_{OG}$ CP was measured in the absence of nucleotide.

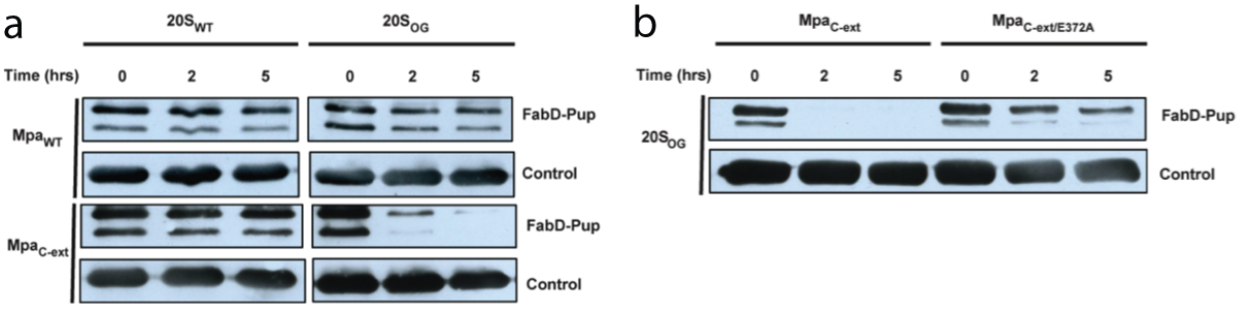


Figure 4-6. Mpa with an extended C-terminus enhances the degradation of pupylated FabD by open-gate proteasomes

(a) Left: in vitro degradation of the substrate Pup~FabD by $20S_{WT}$ or $20S_{OG}$ CPs in the presence of Mpa_{WT} or $MpaC_{C\text{-}ext}$ was analyzed by SDS-PAGE, followed by immunoblotting with polyclonal antibodies to *Mtb* FabD-His_6. Substrate content was monitored by removing aliquots at indicated time points and adding sample buffer to stop reaction. Right: a quantitative view of the degradation assays in the left panels by calculating the percentage of Pup~FabD compared with PrcB-His_6 (control) using Image J software. All data points were normalized to protein present at $t = 0$ h. **(b)** Degradation as described in (A) using $20S_{OG}$ CPs in the presence of $_{MpaC\text{-}ext}$ or $Mpa_{C\text{-}ext/E372A}$.

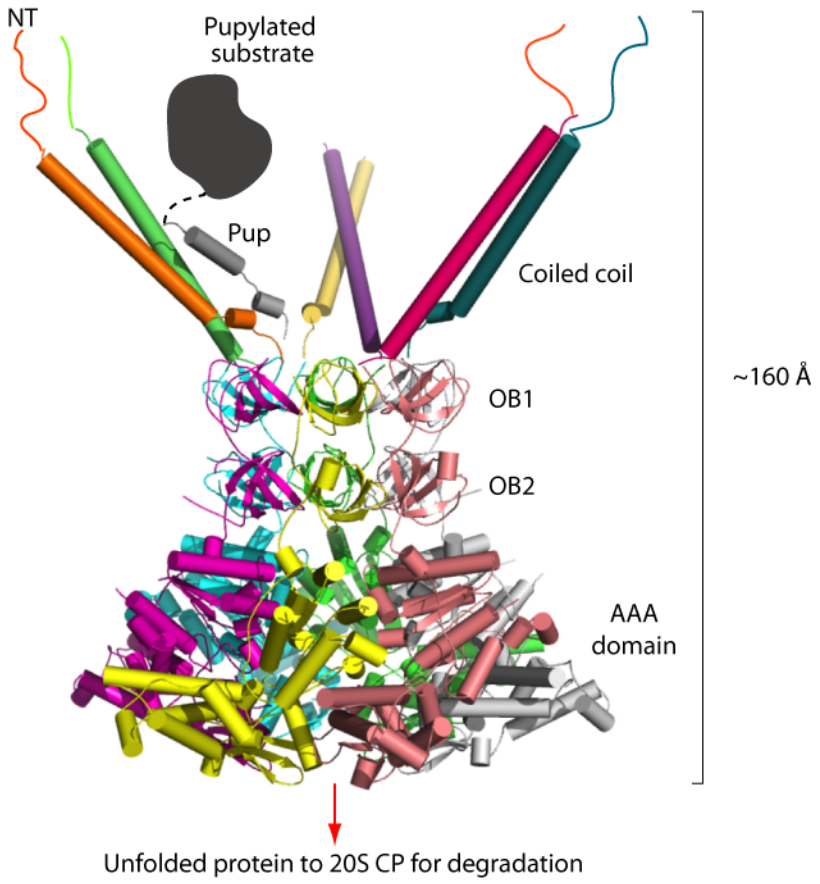


Figure 4-7. A proposed model of the full-length Mpa hexamer

An atomic model of the full-length Mpa hexamer in complex with Pup. The complete Mpa model is generated by overlapping the shared OB domain in the crystal structures of Mpa_{1-234} hexamer (PDB: 3M9D) and Mpa_{95-601} hexamer. The unstructured N-terminal peptides are shown as lines. Only one of the three pairs of coiled coils can bind to pupylated substrate at any given time.

Table 4-1. Data collection and refinement statistics for apo and ADP-bound Mpa

	Mpa-ADP	Mpa
Data collection		
Wave length (Å)	0.9791	1.100
Space group	P321	$P2_12_12_1$
Cell dimensions (Å)	a =b=111.93, c=196.07, α=β= 90°, γ=120°	a =115.27, b=202.59, c=303.01, α=β=γ= 90°
Resolution range (Å)	43.45-2.9 (3.06-2.9)	70.2-3.49 (3.58-3.49)
R_{merge} (%)	11.2 (70.3)	6.9(72.1)
I/σI	19.73 (1.93)	9.9(1.7)
Completeness (%)	100 (100)	98.7 (99.3)
Total reflections	698900(66742)	277696(13888)
Multiplicity	21.7(21.2)	3.1(3.1)
Refinement		
Resolution (Å)	43.45-2.90	70.2-3.49
Unique reflection	32237(3155)	89013(4499)
R_{work}/R_{free} [b]	0.230/0.266	0.237/0.282
Non-H atoms	7179	44299
Protein	7102	44299
Ligand	56	0
Average B-factors (Å^2)	62.47	130.08
Macromolecules	62.73	
Ligands	38.46	
Root-mean-square deviation		
Bond length (Å)	0.008	0.006
Bond angle (°)	1.31	1.134
Ramachandran statistics (%)		
Favored	97	91.5
Allowed	2.9	7.8
Outliers (%)	0	0.7
Molprobity score	1.55	2.93
Clashscore	3.14	15.12

Future perspectives

This work determined crystal structures of PafE and Mpa as well as cryo-EM structures of PafE-20S complex, providing a preliminary investigation into ATP-dependent and -independent proteasome activation processes. In spite of the observation in current work, there remain significant gaps in our understanding of mycobacterial proteasome system.

Given that both PafE and Mpa can access to 20S CPs, the relationship between the two activators becomes an interesting question. Do PafE compete Mpa for binding to 20S CPs? *In vitro* studies have already showed that PafE inhibits Pup-proteasome degradation process mediated by Mpa (Delley et al., 2014). Moreover, this work showed that PafE has a significantly higher affinity to 20S CP than Mpa *in vitro*. In the *mpa* mutant *Mtb*, however, PafE is not enriched by the 20S CP trap compared with the WT bacteria (Jastrab et al., 2015). The access of PafE and Mpa to 20S CP would be controlled in an unknown mechanism.

Asymmetric conformations have been widely identified among other ATP-driven unfoldase as important intermediates during dynamic translocation of substrates (Gates et al., 2017; Lander et al., 2013; Olivares et al., 2016). The crystal structures of Mpa hexamer in our work show the six-fold symmetry, which probably reveal a conformation at the resting stage. A priority goal is to determine Mpa structures in other nucleotide-binding states and substrate-engaged state. Moreover, further studies are needed to understand how the C-terminal GQYL motifs are exposed to engage the 20S CP for the degradation of substrates.

Reference

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica. Section D, Biological Crystallography *66*, 213-221.

Alvarez, B., and Radi, R. (2003). Peroxynitrite reactivity with amino acids and proteins. Amino Acids *25*, 295-311.

Bai, L., Hu, K., Wang, T., Jastrab, J.B., Darwin, K.H., and Li, H. (2016). Structural analysis of the dodecameric proteasome activator PafE in *Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences of the United States of America *113*, E1983-1992.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallographica. Section D, Biological Crystallography *67*, 271-281.

Becker, S.H., and Darwin, K.H. (2017). Bacterial proteasomes: mechanistic and functional insights. Microbiology and Molecular Biology Reviews *81*.

Bolten, M., Delley, C.L., Leibundgut, M., Boehringer, D., Ban, N., and Weber-Ban, E. (2016). Structural analysis of the bacterial proteasome activator Bpa in complex with the 20S proteasome. Structure 24, 2138-2151.

Burns, K.E., and Darwin, K.H. (2010a). Pupylation : a signal for proteasomal degradation in *Mycobacterium tuberculosis*. Subcellular Biochemistry *54*, 149-157.

Burns, K.E., and Darwin, K.H. (2010b). Pupylation versus ubiquitylation: tagging for proteasome-dependent degradation. Cell Microbiology *12*, 424-431.

Butler, S.M., Festa, R.A., Pearce, M.J., and Darwin, K.H. (2006). Self-compartmentalized bacterial proteases and pathogenesis. Molecular Microbiology *60*, 553-562.

Chen, S., Wu, J., Lu, Y., Ma, Y.B., Lee, B.H., Yu, Z., Ouyang, Q., Finley, D.J., Kirschner, M.W., and Mao, Y. (2016). Structural basis for dynamic regulation of the human 26S proteasome. Proceedings of the National Academy of Sciences of the United States of America *113*, 12991-12996.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom

structure validation for macromolecular crystallography. Acta Crystallographica. Section D, Biological Crystallography 66, 12-21.

Coux, O., Tanaka, K., and Goldberg, A.L. (1996). Structure and functions of the 20S and 26S proteasomes. Annual Review of Biochemistry *65*, 801-847.

Darwin, K.H., Ehrt, S., Gutierrez-Ramos, J.C., Weich, N., and Nathan, C.F. (2003). The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. Science *302*, 1963-1966.

Darwin, K.H., Lin, G., Chen, Z., Li, H., and Nathan, C.F. (2005). Characterization of a *Mycobacterium tuberculosis* proteasomal ATPase homologue. Molecular Microbiology *55*, 561-571.

De Mot, R., Nagy, I., Walz, J., and Baumeister, W. (1999). Proteasomes and other self-compartmentalizing proteases in prokaryotes. Trends in Microbiology *7*, 88-92.

Delley, C.L., Striebel, F., Heydenreich, F.M., Ozcelik, D., and Weber-Ban, E. (2012). Activity of the Mycobacterial Proteasomal ATPase Mpa Is Reversibly Regulated by Pupylation. Journal of Biological Chemistry *287*, 7907-7914.

Delley, C.L., Laederach, J., Ziemski, M., Bolten, M., Boehringer, D., and Weber-Ban, E. (2014). Bacterial proteasome activator bpa (rv3780) is a novel ring-shaped interactor of the mycobacterial proteasome. PloS One 9, e114348.

DiMaio, F., Echols, N., Headd, J.J., Terwilliger, T.C., Adams, P.D., and Baker, D. (2013). Improved low-resolution crystallographic refinement with Phenix and Rosetta. Nature Methods 10, 1102-1104.

Djuranovic, S., Hartmann, M.D., Habeck, M., Ursinus, A., Zwickl, P., Martin, J., Lupas, A.N., and Zeth, K. (2009). Structure and activity of the N-terminal substrate recognition domains in proteasomal ATPases. Molecular Cell *34*, 580-590.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallographica. Section D, Biological Crystallography *60*, 2126-2132.

Erales, J., Hoyt, M.A., Troll, F., and Coffino, P. (2012). Functional asymmetries of proteasome translocase pore. The Journal of Biological Chemistry *287*, 18535-18543.

Festa, R.A., McAllister, F., Pearce, M.J., Mintseris, J., Burns, K.E., Gygi, S.P., and Darwin, K.H. (2010). Prokayrotic Ubiquitin-Like Protein (Pup) Proteome of Mycobacterium tuberculosis. PloS one *5*.

Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. Annual Review of Biochemistry *78*, 477-513.

Finley, D., Chen, X., and Walters, K.J. (2016). Gates, channels, and switches: elements of the proteasome machine. Trends in Biochemical Sciences *41*, 77-93.

Forster, A., Masters, E.I., Whitby, F.G., Robinson, H., and Hill, C.P. (2005). The 1.9 A structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. Molecular Cell *18*, 589-599.

Forster, F., Lasker, K., Beck, F., Nickell, S., Sali, A., and Baumeister, W. (2009). An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. Biochemical and Biophysical Research Communications *388*, 228-233.

Gandotra, S., Lebron, M.B., and Ehrt, S. (2010). The *Mycobacterium tuberculosis* proteasome active site threonine is essential for persistence yet dispensable for replication and resistance to nitric oxide. PLoS Pathogens *6*, e1001040.

Gandotra, S., Schnappinger, D., Monteleone, M., Hillen, W., and Ehrt, S. (2007). *In vivo* gene silencing identifies the *Mycobacterium tuberculosis* proteasome as essential for the bacteria to persist in mice. Nature medicine *13*, 1515-1520.

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.* (2017). Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. Science.

Genin, E., Reboud-Ravaux, M., and Vidal, J. (2010). Proteasome Inhibitors: Recent Advances and New Perspectives In Medicinal Chemistry. Current Topics in Medicinal Chemistry 10, 232-256.

Groettrup, M., Soza, A., Eggers, M., Kuehn, L., Dick, T.P., Schild, H., Rammensee, H.G., Koszinowski, U.H., and Kloetzel, P.M. (1996). A role for the proteasome regulator PA28 alpha in antigen presentation. Nature *381*, 166-168.

Groll, M., Bajorek, M., Kohler, A., Moroder, L., Rubin, D.M., Huber, R., Glickman, M.H., and Finley, D. (2000). A gated channel into the proteasome core particle. Nature Structural Biology *7*, 1062-1067.

Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 A resolution. Nature *386*, 463-471.

Groll, M., Nazif, T., Huber, R., and Bogyo, M. (2002). Probing structural determinants distal to the site of hydrolysis that control substrate specificity of the 20S proteasome. Chemistry & Biology 9, 655-662.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. Annual review of biochemistry 67, 425-479.

Hsu, H.C., Singh, P.K., Fan, H., Wang, R., Sukenick, G., Nathan, C., Lin, G., and Li, H.L. (2017). Structural basis for the species-selective binding of N,C-capped dipeptides to the *Mycobacterium tuberculosis* proteasome. Biochemistry *56*, 324-333.

Hu, G., Lin, G., Wang, M., Dick, L., Xu, R.M., Nathan, C., and Li, H. (2006). Structure of the *Mycobacterium tuberculosis* proteasome and mechanism of inhibition by a peptidyl boronate. Molecular Microbiology *59*, 1417-1428.

Huang, X.L., Luan, B., Wu, J.P., and Shi, Y.G. (2016). An atomic structure of the human 26S proteasome. Nature Structural & Molecular Biology *23*, 778-785.

Husnjak, K., Elsasser, S., Zhang, N.X., Chen, X., Randles, L., Shi, Y., Hofmann, K., Walters, K.J., Finley, D., and Dikic, I. (2008). Proteasome subunit Rpn13 is a novel ubiquitin receptor. Nature *453*, 481-488.

Iwanczyk, J., Sadre-Bazzaz, K., Ferrell, K., Kondrashkina, E., Formosa, T., Hill, C.P., and Ortega, J. (2006). Structure of the Blm10-20 S proteasome complex by cryo-electron microscopy. Insights into the mechanism of activation of mature yeast proteasomes. Journal of Molecular Biology *363*, 648-659.

Jastrab, J.B., Samanovic, M.I., Copin, R., Shopsin, B., and Darwin, K.H. (2017). Loss-of-function mutations in HspR rescue the growth defect of a *Mycobacterium tuberculosis* proteasome activator E (pafE) mutant. Journal of Bacteriology *199*.

Jastrab, J.B., Wang, T., Murphy, J.P., Bai, L., Hu, K., Merkx, R., Huang, J., Chatterjee, C., Ovaa, H., Gygi, S.P.*, et al.* (2015). An adenosine triphosphate-independent proteasome activator contributes to the virulence of *Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences of the United States of America *112*, E1763-1772.

Kaplan, G.S., Torcun, C.C., Grune, T., Ozer, N.K., and Karademir, B. (2017). Proteasome inhibitors in cancer therapy: Treatment regimen and peripheral neuropathy as a side effect. Free Radical Biology and Medicine 103, 1-13.

Kim, Y.C., Snoberger, A., Schupp, J., and Smith, D.M. (2015). ATP binding to neighbouring subunits and intersubunit allosteric coupling underlie proteasomal ATPase function. Nature Communications *6*, 8520.

Knowlton, J.R., Johnston, S.C., Whitby, F.G., Realini, C., Zhang, Z., Rechsteiner, M., and Hill, C.P. (1997). Structure of the proteasome activator REGalpha (PA28alpha). Nature *390*, 639-643.

Kohler, A., Cascio, P., Leggett, D.S., Woo, K.M., Goldberg, A.L., and Finley, D. (2001). The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. Molecular Cell *7*, 1143-1152.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. Nature Methods 11, 63-65.

Labbadia, J., and Morimoto, R.I. (2015). The Biology of Proteostasis in Aging and Disease. Annual Review of Biochemistry, Vol 84 84, 435-464.

Lander, G.C., Martin, A., and Nogales, E. (2013). The proteasome under the microscope: the regulatory particle in focus. Current Opinion in Structural Biology *23*, 243-251.

Li, D., Li, H., Wang, T., Pan, H., Lin, G., and Li, H. (2010). Structural basis for the assembly and gate closure mechanisms of the *Mycobacterium tuberculosis* 20S proteasome. The EMBO Journal *29*, 2037-2047.

Lin, G., Chidawanyika, T., Tsu, C., Warrier, T., Vaubourgeix, J., Blackburn, C., Gigstad, K., Sintchak, M., Dick, L., and Nathan, C. (2013). N,C-capped dipeptides with selectivity for mycobacterial proteasome over human proteasomes: role of S3 and S1 binding pockets. Journal of the American Chemcal Society *135*, 9968-9971.

Lin, G., Hu, G., Tsu, C., Kunes, Y.Z., Li, H., Dick, L., Parsons, T., Li, P., Chen, Z., Zwickl, P.*, et al.* (2006). *Mycobacterium tuberculosis* prcBA genes encode a gated proteasome with broad oligopeptide specificity. Molecular Microbiology *59*, 1405-1416.

Lin, G., Li, D.Y., de Carvalho, L.P.S., Deng, H.T., Tao, H., Vogt, G., Wu, K.Y., Schneider, J., Chidawanyika, T., Warren, J.D.*, et al.* (2009). Inhibitors selective for mycobacterial versus human proteasomes. Nature *461*, 621-U663.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. Science *268*, 533-539.

MacMicking, J., Xie, Q.W., and Nathan, C. (1997). Nitric oxide and macrophage function. Annual Review of Immunology *15*, 323-350.

Mayor, T., Sharon, M., and Glickman, M.H. (2016). Tuning the proteasome to brighten the end of the journey. American Journal of Physiology-Cell Physiology 311, C793-C804.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. Journal Structural Biology *142*, 334-347.

Murata, S., Udono, H., Tanahashi, N., Hamada, N., Watanabe, K., Adachi, K., Yamano, T., Yui, K., Kobayashi, N., Kasahara, M.*, et al.* (2001). Immunoproteasome assembly and antigen presentation in mice lacking both PA28 alpha and PA28 beta. Embo Journal *20*, 5898-5907.

Nathan, C. (2012). Fresh approaches to anti-infective therapies. Sci Transl Med *4*.

Nathan, C., Gold, B., Lin, G., Stegman, M., de Carvalho, L.P.S., Vandal, O., Venugopal, A., and Bryk, R. (2008). A philosophy of anti-infectives as a guide in the search for new drugs for tuberculosis. Tuberculosis *88*, S25-S33.

Nathan, C., and Shiloh, M.U. (2000). Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. Proceedings of the National Academy of Sciences of the United States of America *97*, 8841-8848.

Olivares, A.O., Baker, T.A., and Sauer, R.T. (2016). Mechanistic insights into bacterial AAA plus proteases and protein-remodelling machines. Nature Reviews Microbiology *14*, 33-44.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, Part A *276*, 307-326.

Park, S., Tian, G., Roelofs, J., and Finley, D. (2010). Assembly manual for the proteasome regulatory particle: the first draft. Biochemical Society Transactions *38*, 6-13.

Pearce, M.J., Mintseris, J., Ferreyra, J., Gygi, S.P., and Darwin, K.H. (2008). Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. Science *322*, 1104-1107.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computional Chemistry 25, 1605-1612.

Rabl, J., Smith, D.M., Yu, Y., Chang, S.C., Goldberg, A.L., and Cheng, Y. (2008). Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. Molecular Cell *30*, 360-368.

Raule, M., Cerruti, F., Benaroudj, N., Migotti, R., Kikuchi, J., Bachi, A., Navon, A., Dittmar, G., and Cascio, P. (2014). PA28 alpha beta Reduces Size and Increases Hydrophilicity of 20S Immunoproteasome Peptide Products. Chemistry & Biology *21*, 470-480.

Realini, C., Jensen, C.C., Zhang, Z.G., Johnston, S.C., Knowlton, J.R., Hill, C.P., and Rechsteiner, M. (1997). Characterization of recombinant REG alpha, REG beta, and REG gamma proteasome activators. Journal of Biological Chemistry *272*, 25483-25492.

Rock, K.L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D., and Goldberg, A.L. (1994). Inhibitors of the Proteasome Block the Degradation of Most Cell-Proteins and the Generation of Peptides Presented on Mhc Class-I Molecules. Cell *78*, 761-771.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. Journal of Structural Biology 192, 216-221.

Rubin, D.M., Glickman, M.H., Larsen, C.N., Dhruvakumar, S., and Finley, D. (1998). Active site mutants in the six regulatory particle ATPases reveal multiple roles for ATP in the proteasome. The EMBO Journal *17*, 4909-4919.

Sadre-Bazzaz, K., Whitby, F.G., Robinson, H., Formosa, T., and Hill, C.P. (2010). Structure of a Blm10 complex reveals common mechanisms for proteasome binding and gate opening. Molecular Cell *37*, 728-735.

Samanovic, M.I., Tu, S.J., Novak, O., Iyer, L.M., McAllister, F.E., Aravind, L., Gygi, S.P., Hubbard, S.R., Strnad, M., and Darwin, K.H. (2015). Proteasomal Control of Cytokinin Synthesis Protects Mycobacterium tuberculosis against Nitric Oxide. Molecular cell *57*, 984-994.

Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. Journal of Structural Biology *180*, 519-530.

Scheres, S.H.W., and Chen, S.X. (2012). Prevention of overfitting in cryo-EM structure determination. Nature Methods *9*, 853-854.

Schmidt, M., Haas, W., Crosas, B., Santamaria, P.G., Gygi, S.P., Walz, T., and Finley, D. (2005). The HEAT repeat protein Blm10 regulates the yeast proteasome by capping the core particle. Nature Structural & Molecular Biology *12*, 294-303.

Schweitzer, A., Aufderheide, A., Rudack, T., Beck, F., Pfeifer, G., Plitzko, J.M., Sakata, E., Schulten, K., Forster, F., and Baumeister, W. (2016). Structure of the human 26S proteasome at a resolution of 3.9 angstrom. Proceedings of the National Academy of Sciences of the United States of America *113*, 7816-7821.

Seemuller, E., Lupas, A., and Baumeister, W. (1996). Autocatalytic processing of the 20S proteasome. Nature *382*, 468-471.

Shi, Y., Chen, X., Elsasser, S., Stocks, B.B., Tian, G., Lee, B.H., Shi, Y.H., Zhang, N.X., de Poot, S.A.H., Tuebing, F.*, et al.* (2016). Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. Science *351*.

Smith, D.M., Chang, S.C., Park, S., Finley, D., Cheng, Y., and Goldberg, A.L. (2007). Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. Molecular Cell *27*, 731-744.

Smith, D.M., Fraga, H., Reis, C., Kafri, G., and Goldberg, A.L. (2011). ATP binds to proteasomal ATPases in pairs with distinct functional effects, implying an ordered reaction cycle. Cell *144*, 526-538.

Stadtmueller, B.M., and Hill, C.P. (2011). Proteasome activators. Molecular cell *41*, 8-19.

Striebel, F., Hunkeler, M., Summer, H., and Weber-Ban, E. (2010). The mycobacterial Mpa-proteasome unfolds and degrades pupylated substrates by engaging Pup's N-terminus. The EMBO Journal *29*, 1262-1271.

Szabo, C. (2003). Multiple pathways of peroxynitrite cytotoxicity. Toxicology letters *140-141*, 105-112.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: An extensible image processing suite for electron microscopy. J Struct Biol *157*, 38-46.

Tian, G., Park, S., Lee, M.J., Huck, B., McAllister, F., Hill, C.P., Gygi, S.P., and Finley, D. (2011). An asymmetric interface between the regulatory and core particles of the proteasome. Nature Structural & Molecular Biology *18*, 1259-U1108.

Tomko, R.J., Jr., Funakoshi, M., Schneider, K., Wang, J., and Hochstrasser, M. (2010). Heterohexameric ring arrangement of the eukaryotic proteasomal ATPases: implications for proteasome structure and assembly. Molecular Cell *38*, 393-403.

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

Wang, T., Darwin, K.H., and Li, H. (2010). Binding-induced folding of prokaryotic ubiquitin-like protein on the *Mycobacterium* proteasomal ATPase targets substrates for degradation. Nature Structural & Molecular Biology *17*, 1352-1357.

Wang, T., Li, H., Lin, G., Tang, C., Li, D., Nathan, C., Darwin, K.H., and Li, H. (2009). Structural insights on the *Mycobacterium tuberculosis* proteasomal ATPase Mpa. Structure *17*, 1377-1385.

Whitby, F.G., Masters, E.I., Kramer, L., Knowlton, J.R., Yao, Y., Wang, C.C., and Hill, C.P. (2000). Structural basis for the activation of 20S proteasomes by 11S regulators. Nature *408*, 115-120.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A.*, et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallographica. Section D, Biological Crystallography *67*, 235-242.

Wu, Y., Hu, K., Li, D., Bai, L., Yang, S., Jastrab, J.B., Xiao, S., Hu, Y., Zhang, S., Darwin, K.H.*, et al.* (2017). *Mycobacterium tuberculosis* proteasomal ATPase Mpa has a beta-grasp domain that hinders docking with the proteasome core protease. Molecular microbiology *105*, 227-241.

Zhang, F., Hu, M., Tian, G., Zhang, P., Finley, D., Jeffrey, P.D., and Shi, Y. (2009a). Structural insights into the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. Molecular Cell *34*, 473-484.

Zhang, F., Wu, Z., Zhang, P., Tian, G., Finley, D., and Shi, Y. (2009b). Mechanism of substrate unfolding and translocation by the regulatory particle of the proteasome from *Methanocaldococcus jannaschii*. Molecular Cell *34*, 485-496.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nature Methods 14, 331-332.