Abstract: Proteasomes are the principal molecular machines for the regulated degradation of intracellular proteins. These self-compartmentalized macromolecular assemblies selectively degrade misfolded, mistranslated, damaged or otherwise unwanted proteins, and play a pivotal role in the maintenance of cellular proteostasis, in stress response, and numerous other processes of vital importance. Whereas the molecular architecture of the proteasome core particle (CP) is universally conserved, the unfoldase modules vary in overall structure, subunit complexity, and regulatory principles. Proteasomal unfoldases are AAA+ ATPases (ATPases associated with a variety of cellular activities) that unfold protein substrates, and translocate them into the CP for degradation. In this review, we summarize the current state of knowledge about proteasome–unfoldase systems in bacteria, archaea, and eukaryotes, the three domains of life.

Keywords: 19S RP; 20S proteasome; MPA; PAN; proteasome activators; VAT.

Introduction

Selective protein degradation plays a crucial role in almost every aspect of cellular physiology. The degradation of mistranslated, misfolded, damaged or otherwise malfunctioning proteins is an essential element of protein quality control (Chen et al., 2011). Likewise, the timely removal of regulatory proteins is of vital importance in housekeeping functions such as cell cycle control, signal transduction, transcription, DNA repair and translation (Alves dos Santos et al., 2001; Goldberg, 2007; Bader and Steller, 2009; Koepp, 2014). Consequently, any disruption of selective protein degradation pathways leads to a broad array of pathological states, including cancer, neurodegeneration, immune-related disorders, cardiomyopathies, liver and gastrointestinal disorders, and ageing (Dahlmann, 2007; Motegi et al., 2009; Dantuma and Bott, 2014; Schmidt and Finley, 2014).

In eukaryotes, two major pathways have been identified for the selective removal of unwanted proteins – the ubiquitin-proteasome-system (UPS), and the autophagy-lysosome pathway (Ciechanover, 2005; Dikic, 2017). UPS constitutes the principal degradation route for intracellular proteins, whereas cellular organelles, cell-surface proteins, and invading pathogens are mostly degraded via autophagy. The two pathways significantly differ with respect to their substrates, and the machinery involved. However, the elements of control are carefully orchestrated, and jointly they ensure cellular protein homeostasis (Korolchuk et al., 2010).

The unifying element that links both UPS and selective autophagy is a small protein, ubiquitin, which marks substrates for destruction. Via an enzymatic cascade, ubiquitin is covalently bound to substrate proteins by an isopeptide bond between the carboxyl group of its C-terminal glycine residue, and the ε-NH₂ group of substrate lysines. Iterations of the same reaction generate a ‘ubiquitin code’ of signals with different topologies and lengths (Ravid and Hochstrasser, 2008; Komander and Rape, 2012). Depending on the type of ubiquitin modification, a protein substrate is degraded either by the UPS or by autophagy (Kim et al., 2008; Korolchuk et al., 2010). Hence, ubiquitin labeling imparts specificity to a degradation process that is executed by an essentially non-specific molecular machine.

In bacteria and archaea, the existence of UPS or autophagy-lysosome pathways have not been reported. However, the occurrence of proteasomes, and ubiquitin-like small modifier proteins (SAMP and Pup) has been confirmed for all three domains of life.
Occurrence of proteasomes

The first description of proteasomes dates back to the late nineteen sixties, when a ‘cylinder shaped’ particle of unknown function was observed on electron micrographs of erythrocyte ghosts (Harris, 1968). More than a decade later, a cation-sensitive high-molecular-mass endopeptidase was discovered in bovine pituitaries, and designated the ‘multicatalytic protease complex’ (Wilk and Orlowski, 1980). Thereafter, the particle was rediscovered several times, and shown to exist ubiquitously in all eukaryotic cells. However, a lack of consensus over its biochemical nature, and physiological role, resulted in a plethora of names for the same protein complex. Eventually, the name ‘proteasome’ was coined (Arrigo et al., 1988), to highlight its character as a complex molecular machine of names for the same protein complex. For a review of the early history of the field, see Coux et al., 1996; Baumeister et al., 1997.

The first non-eukaryotic proteasomes were discovered in Thermoplasma acidophilum (Dahlmann et al., 1989), paving the way for a detailed analysis of proteasome structure and function in the other domains of life. Eventually, homologous protein complexes were isolated from Pyrococcus furiosus (Bauer et al., 1997), Methanosarcina thermophila (Maupin-Furlow et al., 1998), and several other archaea. Similarly in bacteria, the discovery of proteasomes in Rhodococcus erythropolis (Tamura et al., 1995) lead to its purification from Mycobacterium smegmatis (Knipfer and Shrader, 1997) and other related actinobacteria. So far, proteasomes have been characterized from at least 75 different genera, spanning all three domains of life (Figure 1).

In eukaryotes, proteasomes are the only known soluble ATP-dependent proteases present in the cytoplasm and nucleus. In bacteria however, a more diverse set of ATP-dependent proteases (ClpXP, ClpAP, ClpCP, HslUV, Lon and FtsH systems) are known to exist, whose molecular architecture are analogous to the proteasome (Gur et al., 2011). Among these proteases, Lon is widespread, while an analogue of ClpP shows an infrequent presence in archaea (Maupin-Furlow, 2018). In eukaryotes, homologues of Lon, ClpP, and FtsH are only found in the cellular organelles of bacterial descent, such as mitochondria, and chloroplasts (Adam et al., 2001).

Structure of the proteasome core is universally conserved

Across the domains of life, proteasomes vary in subunit composition. However, the central component of all proteasomes, the core particle (CP) (also known as the 20S proteasome) exhibits a high level of architectural conservation (Figure 1). The CP is a barrel-shaped, self-compartmentalized, protein complex, composed of four stacked seven-membered rings (Baumeister et al., 1998). The outer rings comprise α-type subunits, whereas β-type subunits form the inner rings. The co-axial stacking of four rings (α,β,β,α) creates three internal cavities, bound by four narrow constrictions, as originally observed in three-dimensional (3D) reconstructions from electron micrographs (Hegerl et al., 1991). The central cavity, formed at the junction of two β-rings, is the catalytic chamber, where protein degradation takes place. The two outer cavities, formed at α-β junctions serve as antechambers, where substrate proteins can be stored in an unfolded state prior to degradation (Sharon et al., 2006; Ruschak et al., 2010).

In archaea and actinobacteria, only one or two types of α and β-subunits exist, while in eukaryotes, at least 14 unique subunits (α1-7 and β1-7) build the constitutive CP. In addition, higher eukaryotes express specialized proteasomes such as the immunoproteasomes and thymoproteasomes, where β subunit variants (β1i, β2i, β5i, and β5t) substitute the house-keeping ones (Murata et al., 2018). The α- and β-type subunits are structurally related, differing only in their N-terminal regions. The N-terminal tails of α-subunits form a CP gate, which prevents access of folded proteins into the CP-axial channel (Wenzel and Baumeister 1995). Whereas the N-terminal regions of β subunits serve as pro-peptides that are autocatalytically cleaved off during proteasome maturation to expose the proteolytic active sites (Seemuller et al., 1996; Groll et al., 1999; Huber et al., 2016).

Proteosome active sites are sequestered within the CP

Active sites of the proteasome are located on the inner walls of the CP catalytic chamber, and comprise a triad of residues Thr1, Lys33, and Asp/Glu17 (Huber et al., 2016) (Figure 2). These residues are conserved among all proteolytically active β-type subunits in eukaryotic, bacterial and archaeal proteasomes, and hence the mechanism of peptide bond cleavage follows a universal principle among all CPs (Figure 2). The substrate cleavage preference is however distinct for every subunit, and depends on the chemical nature of the substrate-binding channel, especially the active site’s S1 specificity pocket (Groll and Huber, 2003). Thus, the CP in archaea and actinobacteria comprise 14 identical catalytic sites (Seemuller et al., 1995;
Proteasomes are classified as a family of N-terminal nucleophilic (Ntn) threonine proteases, where the free N-terminal Thr1 of a catalytically active β-subunit acts as a nucleophile in peptide bond hydrolysis (Seemuller et al., 1995). For mechanistic details, see Huber et al., 2016. Back in the day, the discovery of this unique mechanism immediately suggested a way of inhibiting the CP by capping the β-subunit N-termini, and in turn inspired the development of an important therapeutic agent bortezomib. For a brief history of the development of proteasome inhibitors, see Goldberg, 2012.
simultaneously prevent the irreversible inactivation of the Thr1 N terminus by N-acetylation (Arendt and Hochstrasser, 1999; Groll et al., 1999; Jager et al., 1999; Kwon et al., 2004). For a detailed account of the CP assembly process, and the chaperones involved, see Zuhl et al., 1997; Murata et al., 2009; Kunjappu and Hochstrasser, 2014.

Substrate processing requires proteasome activation

Protein degradation by the CP does not require ATP hydrolysis as such. However, the closed compartmentalized structure of the CP limits its constitutive proteolytic activity to only denatured or unstructured proteins and small peptides (Wenzel and Baumeister, 1995; Liu et al., 2003). Larger protein substrates are mostly degraded via ‘unfoldase-assisted-proteolysis’, whereby the CP recruits cofactors or activators to facilitate substrate access into its catalytic chamber.

Several proteasomal activators have thus far been identified, spanning both ATP-dependent and independent varieties (Stadtmueller and Hill, 2011). These activators cap the CP at one or both ends, and reposition the α-ring gating residues into an open conformation. For the processing of folded proteins, however, gate opening is not sufficient to allow access into the catalytic chamber. Additional steps of substrate unfolding and translocation are required, and can be achieved only at the cost of ATP hydrolysis. As a result, proteolysis is rendered an energy-dependent process.

Energy-dependent proteasome activators

In all domains of life, energy-dependent proteasome activation involves ATPases of the AAA+ variety (Yedidi et al., 2017). These are an extended group of ATPases associated with various cellular activities, among which AAA-ATPases form a sub-group. Members of the AAA+
superfamily are defined by the presence of AAA+ domains that use six conserved motifs for function; Walker A, Walker B, sensor-1, arginine finger (R-finger), sensor-2 and pore loops (Wendler et al., 2012; Miller and Enemark, 2016). The Walker A and Walker B motifs are required for ATP binding and hydrolysis. The sensor-1 and R-fingers coordinate nucleotide hydrolysis and propagate the associated conformational changes between subunits (Ogura et al., 2004). The sensor-2 region mediates conformational changes associated with a cycle of ATP binding and hydrolysis, whereas the pore loops make direct contact with substrate (Martin et al., 2008).

Proteasomal ATPases share a common architecture and assemble into hexameric ring complexes, with N-terminal domains, AAA+ domains, and C-terminal HbYX motifs (where Hb is a hydrophobic residue, Y is tyrosine, and X is any amino acid). The N-terminal domain is believed to be involved in substrate capture. The HbYX motif is responsible for interaction with the CP, while the AAA+ domain unfolds protein substrates, and translocates them through the axial channel, into the degradation chamber. Such an association of hexameric ATPases with a heptameric CP α-ring gives rise to ‘symmetry-mismatched’ complexes.

In eukaryotes, the proteasomal ATPase is an assemblage of ‘regulatory particle triple A ATPases’ (Rpt1-6) that forms the core of the 19S regulatory particle (RP). In archaea however, the proteasome is believed to operate via a network of AAA+ ATPases (Forouzan et al., 2012), among which the ‘proteasome activating nucleotidase’ (PAN), and the Cdc48 homologue ‘VCP-like ATPase’ (VAT) have been characterized. For a long time only one proteasomal ATPase was identified in actinobacteria, the ‘AAA-ATPase forming ring-shaped complex’ (ARC) (Wolf et al., 1998), which is known as the ‘mycobacterial proteasome ATPase’ (Mpa) in Mycobacterium (Darwin et al., 2005). Recently however, a Cdc48 homologue ‘Cdc48-like protein of actinobacteria’ (Cpa) has also been observed to interact with the proteasome in a manner similar to other proteasomal ATPases (Ziems et al., 2018).

Similar to the CP, the subunit complexity of proteasomal ATPases varies in the different domains of life. Thus, the archaeal or actinobacterial ATPases are homohexameric, whereas the Rpts of the 26S proteasome form a heterohexameric complex, which is further associated with non-ATPase subunits (Rps) in the RP. Due to the absence of Rpn-equivalents, non-eukaryotic proteasomes presumably function as minimal complexes, where the ATPases alone are responsible for substrate capture, in addition to CP activation and substrate unfolding. However, these CP-ATPase complexes are rather unstable, and notoriously difficult to isolate from cell lysates. Hence, a question is often raised regarding the constitution and stability of these proteasomes, hinting at a fleeting interaction between CP and ATPase, or the involvement of additional stabilizing co-factors that are yet unknown.

**PAN**

PAN was the first proteasome-activating AAA-ATPase discovered in archaea (Zwickl et al., 1999). It exhibits ~41–45% sequence homology to the Rpts of the 26S proteasome (Zwickl et al., 1999), and is considered as its evolutionary precursor. In terms of domain architecture, it is the simplest among proteasomal ATPases, and comprises an N-terminal helix, an oligonucleotide/oligosaccharide binding (OB) domain, an ATPase domain, and a C-terminal HbYX motif. The association of six PAN protomers creates three two-stranded coiled-coils, an OB-ring and an AAA-ring, all of which are connected via flexible linkers (Djurunic et al., 2009; Zhang et al., 2009a).

Biochemical studies over the past two decades have confirmed a chaperone activity of PAN, by which it is able to prevent the *in vitro* aggregation of denatured proteins, and assist in their refolding (Benaroudj and Goldberg, 2000; Navon and Goldberg, 2001). There have also been speculations regarding the conformational switching of N-domains (Djurunic et al., 2009; Snoberger et al., 2018), nucleotide usage by the AAA-ring (Zhang et al., 2009b; Smith et al., 2011; Kim et al., 2015), and the involvement of HbYX tails in CP gate opening (Smith et al., 2007; Rabl et al., 2008). However, the physical association between PAN and CP has been shown for the first time in 3D in our recent cryo-EM structure of the *Archaeoglobus fulgidus* PAN-proteasome (Majumder et al., 2019).

Under *in vitro* conditions, the association of PAN and CP requires ATP. The interaction is short-lived, and PAN-proteasome holocomplexes are enriched only in the presence of the slowly hydrolyzable nucleotide analog ATPγS (Smith et al., 2005). Within a PAN-proteasome, PAN interacts with the CP via its C-terminus, while the N-terminal coiled-coils are free (Figure 3). A clear offset is visible between PAN and CP, and their axial channels are not perfectly aligned. However, the PAN AAA-ring adopts a uniform spiral staircase conformation, with helically arranged pore loops protruding into the axial channel.

Interestingly, the interaction between PAN and CP is bipartite. While the insertion of C-terminal HbYX tails into inter-α-subunit pockets establishes a classical contact between PAN and CP, the N-terminal residues
of the CP α-subunits extend outwards to engage in ‘gate contacts’ with the lowest protomer in the PAN AAA-staircase.

Extensive in silico classification allowed us to deconvolute five conformational states coexisting in the sample. The conformational states show the sequential rotation of the PAN AAA-staircase (Figure 3). In each state, PAN is positioned at an offset atop the CP axial channel, yet the CP exhibits an open gate with all six PAN HbYX tails inserted within inter α-subunit pockets. The PAN AAA-staircase features co-existing nucleotide states, and correlated nucleotide-dependent inter-subunit signaling contacts between adjacent protomers. The OB-ring and coiled-coils that seem to perch on top of the AAA-ring, also adjust with the conformational changes in the AAA-ring, and slant from the highest protomer.

Guided by the structural information from the five conformational states, it appears that progress of the ATP hydrolysis cycle is accompanied by the making and breaking of inter-subunit contacts, which in turn causes the rotation of the AAA-staircase. Thus, a sequential ATP hydrolysis event leads to a harmonic up and down motion of the pore helices and pore loops around the AAA-ring. This motion is communicated via the linkers...
to the OB ring and coiled coils, and helps drive the gripping, unfolding, and translocation of substrate proteins into the CP.

VAT

VAT is an archaeal homologue of the widely occurring Cdc48/p97/VCP ATPase (Golbik et al., 1999; Gerega et al., 2005), which is ubiquitous in both archaea and eukaryotes. Cdc48 plays a crucial role in a wide range of cellular activities (Meyer et al., 2012), including the unfolding of ubiquitin (Twomey et al., 2019). VAT was recognized as a proteasome activator when PAN was found not to exist in some archaea, and not required for the viability of some others, where the CP was essential (Zhou et al., 2008). Interestingly, the involvement of Cdc48-type proteins in proteasome activation has thus far been reported only for archaea and actinobacteria (Barthelme and Sauer, 2012; Forouzan et al., 2012; Ziemski et al., 2018). In eukaryotes, a collaboration between Cdc48/p97 and 26S proteasomes has been implicated, particularly in the context of endoplasmic-reticulum-associated protein degradation (ERAD) (Bodnar and Rapoport, 2017). However, compelling evidence for a direct physical interaction between Cdc48 and proteasomes is still lacking. The prevailing view is that p97, together with Ufd1/Npl4, extracts polyubiquitinated proteins from the ER, and hands them over to the proteasome (Wu and Rapoport, 2018). However, in situ studies have not shown a direct or stable interaction between Cdc48 and 26S proteasomes.

VAT is a Janus-faced chaperone occurring in the archaeon *T. acidophilum* (Pamnani et al., 1997; Golbik et al., 1999). Depending on the concentration of Mg²⁺ present, VAT assumes either a low activity state that is capable of refolding proteins or a high activity one, which accelerates the unfolding of the same (Golbik et al., 1999). It is a member of the Type II family of AAA+ ATPases, in which each protomer comprises three domains: an N-terminal domain that is divergent between AAA+ proteins, and two tandem nucleotide binding domains. In presence of the slowly hydrolyzable nucleotide analogue ATPγS, full length VAT exists either as a six-fold symmetric stack of rings, or as helical split-rings, the latter conformation being predominant also in presence of ADP (Huang et al., 2016) (Figure 4). Typical of proteasomal ATPases, the VAT C-terminus contains an HbYX motif, which is involved in an interaction with the CP. The

![Figure 4: Substrate processing by the *T. acidophilum* VATΔN (Ripstein et al., 2017). (A) Cryo-EM densities and superimposed models of the ATPγS bound VATΔN (PDB: 5VC7) is shown from two different angles. It has a characteristic six-fold symmetry, with a central pore/axial channel running through the middle of the complex. (B) The structure of the substrate-bound VAT complex (PDB: 5VCA). The substrate density is visible only in filtered versions of the density map (EMDB: EMD-8659), and its trajectory through the VAT complex is emphasized by the red dashed line. In each case, the N-terminal nucleotide binding domain (NBD1) is shown in sand, and the C-terminal nucleotide binding domain (NBD2) is colored dark green. Upon substrate binding, a conformational change in the NBD rings leads to a helical architecture.](image-url)
nucleotide binding domains coordinate ATP hydrolysis and substrate unfolding. However, the role of the N-terminal domain is unclear, and as suggested by mutational studies, it might partially inhibit the unfolding activity of VAT (Gerega et al., 2005). In fact, the N-terminal deletion mutants of VAT are more stable, and show increased CP binding affinity compared to full-length ones (Barthelme and Sauer 2012).

Pull-down assays indicate a short-lived interaction between VAT and CP (Barthelme and Sauer, 2012). In fact, the visualization of VAT-CP complexes has been possible only after site-specific crosslinking (Barthelme et al., 2014), although VAT has a higher affinity for the CP \( K_d(\text{VAT-CP}) = 160 \text{ nm} \) compared to PAN \( K_d(\text{PAN-CP}) = \sim 300 \text{ nm} \) (Barthelme and Sauer, 2012). Nevertheless, cryo-EM and NMR studies of substrate bound VAT have provided invaluable insights into its mechanism of unfolding and translocation (Ripstein et al., 2017; Augustyniak and Kay, 2018).

Trapping calmodulin bound VAT via engineered inter-molecular disulfide bridges, it was shown that VAT initiates substrate translocation by pulling it onto intrinsically unstructured regions, irrespective of amino acid sequence, or location (Augustyniak and Kay, 2018). Substrate binding is accompanied by a conformational change in each protomer of N-terminal deleted VAT, so as to rearrange the nucleotide binding domains from a stacked-ring conformation into a split lockwasher (Figure 4). The transition occurs via the horizontal displacement and rotation of one nucleotide binding domain with respect to the other, and results in a decreased distance between their pore loops. The pore loops of five of six protomers interact with the substrate, wrapping \( \sim 80 \text{ Å} \) length of substrate in a hydrophobic helical sheath. The sixth (non-interacting) protomer, occupies the top and bottom of the helix, forming a seam subunit with dramatically distinct protomer-protomer interactions. Thus, by a processive ‘hand-over-hand’ mechanism, substrates are translocated as each VAT protomer releases its grip on the substrate, and re-engages further along the target (Ripstein et al., 2017).

A new addition to the list of Cdc48-type proteasome activators is Cpa, which by phylogenetic analysis appears to descend from a common Cdc48 ancestor (Ziemski et al., 2018; Müller and Weber-Ban, 2019). The interaction of Cpa with the proteasome has been recently recognized, and remains to be explored in detail. Like the other proteasomal AAA+ ATPases, Cpa forms hexameric ring complexes. However, its oligomerization occurs only presence of nucleotide. Interestingly, Cpa lacks a classical HbYX motif, and yet binds the CP with an affinity \( K_d(\text{Cpa-CP}) = 360 \text{ nm} \) comparable to a PAN-CP interaction \( K_d(\text{PAN-CP}) = \sim 300 \text{ nm} \).

MPA

In actinobacteria, proteolysis by the CP requires assistance of the AAA-ATPase ARC/MPA (Darwin et al., 2005). The MPA-proteasome system renders Mycobacterium tuberculosis resistant to killing by host macrophages, and plays a crucial role in tuberculosis (Samanovic and Darwin, 2016). Consequently, the MPA-proteasome system serves as an attractive target for the development of antituberculosis therapeutics (Lupoli et al., 2018).

ARC/MPA shares structural similarity with the archaeal PAN or the eukaryotic Rpts. Each MPA protomer comprises an N-terminal extended \( \alpha \)-helical region, followed by two tandem oligonucleotide/oligosaccharide binding (OB) domains, an AAA-domain, a \( \beta \)-grasp domain and a C-terminal GQYL motif (Wu et al., 2017). Upon oligomerization, adjacent \( \alpha \)-helical domains form three pairs of two stranded coiled-coils, while the OB and AAA domains assemble into co-axially stacked hexameric rings.

The N-terminal coiled-coils in MPA serve as a specific receptor for the small ‘prokaryotic ubiquitin-like protein’ (Pup) (Sutter et al., 2009). Pup, which is an otherwise unstructured protein, interacts with the coiled-coils in MPA via its N-terminus and assumes a helical structure (Wang et al., 2010). It is believed that MPA-proteasome complexes engage Pup-tagged substrates, and unfolds and drops them into the CP (Striebel et al., 2010). The precise role of the OB-domains is unclear. However, they have been found sufficient to maintain the hexameric state of MPA (Wang et al., 2009). The C-terminal GQYL motif is a functional equivalent of the HbYX motif found in other proteasomal ATPases, and is believed to be responsible for activator-CP interaction (Bolten et al., 2016; Hu et al., 2018). However, the \( \beta \)-grasp domain buries the C-terminus of each protomer in the central channel of the MPA hexamer, and hinders interaction with the CP (Wu et al., 2017).

Incidentally, a 3D structure of an MPA-proteasome complex is still lacking. Only a loose physical association between MPA and CP has thus far been observed, and is limited to gate-deleted CP constructs (Wang et al., 2009). In fact, MPA alone cannot facilitate robust in vitro degradation by the CP, and requires the presence of additional factors such as the proteasome accessory factor (PafA or PafE), a Pup ligase, a depupylase (Dop), and possibly others that are yet unknown (Pearce et al., 2008; Striebel et al., 2009; Burns et al., 2010; Hu et al., 2018).
19S RP

Among the proteasome activators known thus far, the 19S RP is the most complex macromolecular assemblage. It forms the regulatory component of the 26S proteasome, and is composed of at least 19 canonical subunits, along with an array of substoichiometric ‘proteasome interacting proteins’ (Pips). The RP can be biochemically separated into the ‘base’ and ‘lid’ subcomplexes (Glickman et al., 1998). Such a separation assigns 10 subunits to the base, among which are six distinct AAA+ ATPases (Rpt1-6), and four non-ATPase subunits (Rpn1, Rpn2, Rpn10 and Rpn13). The remaining nine subunits (Rpn3, Rpn5, Rpn6, Rpn7, Rpn9, Rpn12, Rpn8, Rpn11, and Sem1) constitute the lid. For a detailed structural description of the 19S RP, see Beck et al., 2012; Lander et al., 2012; Lasker et al., 2012 and Schweitzer et al., 2016 (Figure 5).

At the core of the 19S RP is a hexameric assemblage of Rpt1-2-6-3-4-5 ATPases (Tomko et al., 2010), which serves as the molecular motor of the 26S proteasome. It resembles the archaeal PAN in overall structure, and comprises an AAA-ring, an oligosaccharide/oligonucleotide binding (OB) ring (also known as the N-ring), and coiled-coils between pairs of neighboring subunits (Rpt1/Rpt2, Rpt6/Rpt3, and Rpt4/Rpt5). The Rpt1-6 motor is primarily responsible for substrate unfolding and translocation into the CP. In addition, it regulates CP gate opening by insertion of its C-terminal HbYX tails into the CP inter-α-subunit pockets. The remaining subunits, which include ubiquitin receptors, deubiquitylases (DUBs), and scaffolding proteins, partially embrace the Rpt1-6 hexamer and the CP. The resulting physical interaction between ATPase and non-ATPase subunits helps to connect substrate recruitment and ubiquitin processing to the unfolding machinery.

Ubiquitin receptors are placed along the periphery of the RP, where they are accessible to protein substrates. Being highly flexible, they are the least well-resolved parts in 26S proteasome structures. Rpn10 binds polyubiquitin chains through its ubiquitin-interacting motif (UIM), while its N-terminal von Willebrand A (VWA) motif mediates its interaction with the proteasome (Deveraux et al., 1994; Verma et al., 2004; Wang et al., 2005). The

\[\text{Figure 5: Cryo-EM structure of a single-capped human 26S proteasome is shown in three rotated orientations (EMDB: EMD-4002; Schweitzer et al., 2016).}
\]

The CP is formed of four stacked hetero-heptameric rings (α1-7 β1-7 β1-7 α1-7), whereas the RP comprises a hetero-hexameric ATPase (Rpt1-6) core, surrounded by non-ATPase subunits (Rpn1-3, Rpn5-13, and Sem1). Highlighted labels indicate the RP subunits with specialized roles in substrate processing. The corresponding functions are enlisted in the color key on the right. Ubiquitin receptor subunits are peripherally located on the RP. Rpn10 and Rpn13 are the primary ubiquitin receptors, while Rpn1, Rpn2 and Sem1 recruit ubiquitinated substrates via shuttle factors. In the human 26S proteasome preparations, Rpn13 is present in substoichiometric amounts, and therefore missing in the density map. However, it is known to be recruited via Rpn2. The view in the middle exhibits the proteasome lid with a horseshoe-like arrangement of scaffolding subunits (Rpn12-Rpn3-Rpn7-Rpn6-Rpn5-Rpn9). These subunits presumably play structural roles, with Rpn6 clamping together the RP and the CP. The view on the right highlights the Rpt1-6 motor (in shades of gray). It comprises an AAA-ring, an OB-ring, and three coiled-coils. The Rpt1-6 AAA-ring docks onto the CP via insertion of C-terminal tails (of Rpt1, Rpt2, Rpt6, Rpt3, and Rpt5) into inter-α-subunit pockets. The insertion of Rpt1 and Rpt6 C-terminals lead to CP gate opening. Substrates are deubiquitylated by Rpn11, which forms a hetero-dimer with Rpn8, and is located close to the Rpt1-6 OB-ring. The other proteasomal DUBs, Ubp6 (recruited via Rpn1) and Uch37 (recruited via Rpn13), could not be localized in the density map.
yeast Rpn10 contains a single UIM. However, the number of UIMs increase in higher eukaryotes, and the additional UIMs can serve as binding sites for ubiquitin-like (UBL) domain containing shuttle factors. It has been shown that Rpn10 functions both when free, and in the proteasome bound form (Matiuhin et al., 2008). However, ubiquitylation of Rpn10 regulates its capacity to further bind ubiquitinated substrates (Isasa et al., 2010).

Rpn13 recognizes ubiquitinated substrates through its N-terminal pleckstrin-like receptor of ubiquitin (PRU) domain, while its opposite surface interacts with Rpn2 (Husnjak et al., 2008; Lu et al., 2017). In higher eukaryotes, Rpn13 contains an additional DEUBiquitinase adaptor (DEUBAD) domain, which interacts with a DUB Uch37 (Hamazaki et al., 2006; Yao et al., 2006). Like Rpn10, Rpn13 also binds ubiquitin-like (UBL) domains of UBL-ubiquitin-associated (UBA) proteins. The combination of ubiquitin binding and deubiquitinating functionalities renders Rpn13 as a ‘proof-reading’ machinery that enables substrates to either engage or escape from the proteasome.

Rpn1, a PC-repeat protein, serves as a ligand-binding hotspot, with two adjacent receptor sites for substrate binding and deubiquitination (Shi et al., 2016). With the capacity to recognize both ubiquitin, and UBL domains of substrate shuttling factors, Rpn1 recruits extrinsic ubiquitin shuttle receptors (Rad23, Ddi1 and Dsk2) and the non-essential DUB Ubp6 to the proteasome (Elsasser et al., 2002; Leggett et al., 2002; Gomez et al., 2011). Ubiquitin recognizing properties have also been associated with the scaffolding protein Sem1/Dss1 (Paraskevopoulos et al., 2014), and the proteasomal ATPase, Rpt5 (Lam et al., 2002), however these are less well characterized.

The 19S RP contains three DUBs – Rpn11, Ubp6/Usp14 and Uch37, among which only Rpn11 is essential. It is a Zn\(^2+\) containing JAMM metalloprotease, which forms an obligate heterodimer with its neighboring subunit Rpn8. Rpn11 exists in a self-inhibited state (Pathare et al., 2014; Worden et al., 2014), which is further stabilized by its neighboring lid subunit, Rpn5 (Dambacher et al., 2016). In the ground state of the proteasome, it is located above the Rpt1-6 motor, between Rpn2 and the ubiquitin receptor Rpn10. Upon substrate engagement, a conformational switching brings Rpn11 to the mouth of the substrate translocation channel (Matyskiela et al., 2013), and substrate-bound ubiquitin moieties are directly pulled into its catalytic groove (de la Pena et al., 2018).

The two other DUBs, Ubp6/Usp14 and Uch37, associate with the RP in substoichiometric amounts. Ubp6 tethers to Rpn1 via an N-terminal UBL domain, whereas its C-terminal catalytic domain is positioned variably (Außerheide et al., 2015). Interestingly, ubiquitin docking to the active site of Ubp6 ‘non-catalytically’ inhibits protein degradation by the proteasome (Bashore et al., 2015; Kim and Goldberg 2017). The other DUB Uch37, associates with proteasomes via interaction of its C-terminal Uch37-like domain (ULD) and the DEUBAD of Rpn13 (Yao et al., 2006; Vander Linden et al., 2015). Like Rpn11, Uch37 exists in a self-inhibited state, and binding to RPN13 relieves the inhibitory effect (Yao et al., 2006; Jiao et al., 2014).

The remaining 19S RP subunits are believed to play structural roles. Rpn3, Rpn7, Rpn6, Rpn5, Rpn9 and Rpn12 are arranged in a horseshoe-shaped anchor from which their N-terminal domains extend radially. These subunits are characterized by a C-terminal PCI (proteasome – cop9 signalosome – initiation factor 3) domain, and likely act as scaffolds (Finley 2009). Rpn2, which is structurally similar to Rpn1, also functions as a scaffold, binding Rpn13 and ubiquitin shuttles (Rosenzweig et al., 2012). Due to the presence of these structural subunits, especially the RP-CP clamping subunit Rpn6 (Pathare et al., 2012), the 26S proteasome is significantly more stable than the proteasomal ‘minimal complexes’ found in the other domains of life. In fact, intact endogenous 26S proteasomes are routinely purified from eukaryotic cell lysates, and have also been successfully visualized in situ (Asano et al., 2015; Albert et al., 2017; Guo et al., 2018).

**Mechanism of unfoldase-assisted-proteolysis**

Thanks to the numerous cryo-EM studies on yeast and human 26S proteasomes, the proteasome conformational landscape has been significantly expanded in recent years (Matyskiela et al., 2013; Sledz et al., 2013; Ding et al., 2017; Wehmer et al., 2017; de la Pena et al., 2018; Eisele et al., 2018; Zhu et al., 2018; Ding et al., 2019; Dong et al., 2019). However, the increasing array of conformational states makes it challenging to predict the actual sequence of events. The conformations observed reflect large scale motions of the proteasome lid that lead to the relocation of certain RP subunits, and a concomitant restructuring of the inter-subunit interaction network. The latter is believed to play a vital role in ubiquitin recognition and processing, the properties unique to eukaryotic proteasomes. However, the unfolding and translocation of engaged substrates is controlled largely by the Rpt1-6.

Both yeast and human 26S proteasomes exhibit different planar and rotated spiral staircase conformations of the Rpt1-6 ring. This indicates the co-existence of multiple
nucleotide states, or an ATPase cycle in progress. With the parallel discovery of rotated spiral staircases in the archaeal PAN, the fundamental role of an around-the-ring ATPase cycle is further established. Although the sequence of steps in an ATPase cycle is yet to be determined, the existence of inter-subunit signaling motifs in both PAN and Rpt1-6 indicate an underlying communication between adjacent protomers during the progression of the cycle. Another notable feature conserved between PAN and the Rpt1-6 is the motion of the OB ring, which is coordinated with the ATPase domain transitions. In the case of PAN, such motions lead to the wagging of the exposed N-terminal coiled-coils, like ‘swinging arms’, and are possibly responsible for substrate capture. However, in the case of Rpt1-6, motions of the OB-ring and coiled-coils results in varied interactions with lid subunits. As non-eukaryotic proteasomes require the assistance of AAA+ unfoldases, one would expect that such transitions in the Rpt1-6 motor in turn coordinate the conformational changes in the 26S proteasome lid.

The functional cycle of mycobacterial proteasomes however, seem to be distinct from the eukaryotic or archaeal counterparts. As the presence of a β-grasp domain in MPA hinders its HbYX motifs from physically interacting with the CP, the CP is presumably activated with the help of accessory factors PafA or PafE. It would be interesting to note how gate opening, substrate unfolding, and translocation are coordinated in such systems.

ATP-independent activators

In addition to the ATPase activators of the CP, several non-ATPase activators have been identified. They are less widely conserved, and their substrates and biological functions are less clear. However, in most cases, the interaction between CP and non-ATPase activators have been successfully characterized from a structural perspective (Forster et al., 2005; Sadre-Bazzaz et al., 2010). In eukaryotes, ATP-independent proteasome activators belong to two major categories: PA200 and PA28 (11S). The PA200 activators in human, and their homologous Blm10 activators in yeast, are composed of a single polypeptide chain with an array of HEAT repeats. This elongated solenoid gives rise to a dome like architecture (Sadre-Bazzaz et al., 2010), which forms complexes with the CP (Schmidt et al., 2005; Toste Rego and da Fonseca, 2019). The C terminus of the Blm10 chain contains a YYX motif (analogous to the HbYX motif of ATP-dependent activators) that binds between the CP α5 and α6 subunits, and induces partial opening of the CP gate (Dange et al., 2011). These activators are believed to participate in a broad variety of processes, including CP assembly (Fehlker et al., 2003), DNA repair (Ustrell et al., 2002), genomic stability (Blickwedehl et al., 2008), proteasome inhibition (Lehmann et al., 2008), spermatogenesis (Khor et al., 2006), and mitochondrial function (Sadre-Bazzaz et al., 2010).

Higher eukaryotes and certain unicellular eukaryotes (such as trypanosomes) additionally express a PA28/11S/REG family of proteasome activators. However, they are missing in yeast and plants. The PA28 activators are heptameric toroids that form hybrid complexes with the CP (Hendil et al., 1998). Unlike the PA200/Blm10 activators, here each subunit of the heptamer interacts with the CP via internal activation loops, and non-HbYX type insertions of their C-terminal residues (Forster et al., 2005). In mammals, three different isoforms of PA28 have been identified (PA28α, β, and γ), among which PA28γ has earned an attention due to its role in cancer biology (He et al., 2012). PA28 activators are believed to be involved in cell cycle control and apoptosis (Murata et al., 1999; Masson et al., 2003), MHC class-I antigen presentation (Groettrup et al., 2010), and a variety of cellular roles.

ATP-independent proteasome activators have also been found in bacteria and in archaea. PaFE (Bpa) is functionally similar but evolutionarily unrelated to the eukaryotic proteasome activators PA26 and PA28, and contributes to the virulence of M. tuberculosis (Jastrab et al., 2015). Cryo-EM studies have recently elucidated how a dodecameric assemblage of PaFE interacts with the CP via its C-terminal GOYLM motifs (Bolten et al., 2016; Hu et al., 2018). In archaea however, PbaA and PbaB are homologous to the eukaryotic CP assembly factors PAC1-PAC2. PbaB forms a homotetramer with tentacle-like C-terminal HbYX segments. It can function both as a molecular chaperone as well as an ATP-independent proteasome activator (Kumoi et al., 2013).

Concluding remarks

In the three domains of life, proteasomes and their activators show substantial structural and functional overlap, and yet there are key differences in their mechanisms of assembly, activation, and substrate targeting for degradation. Like the UPS in eukaryotes, the proteasomes in actinobacteria function as a part of the Pup-proteasome system, whereby substrates tagged with the small modifier protein Pup are recognized by MPA, and targeted to the proteasome (Pearce et al., 2008). Similarly, in archaea,
the enzymatic attachment of a ubiquitin-like modifier protein SAMP (Humbard et al., 2010), serves as a signal for proteasome mediated proteolysis (Fu et al., 2016). Not to mention, both archaea and actinobacteria possess elaborate machineries for the conjugation and removal of Pup and SAMP, respectively.

In evolutionary terms, the bacterial ATP-dependent protease, HslUV, is considered a prototype of the proteasome. This two-component, composed of AAA+ (HslU), and proteolytic (HslV) units, is related to the proteasome both structurally and functionally (Rohrwild et al., 1996). With sequence homology between HslV and proteasome β subunits, HslUV even demonstrates a threonine-dependent proteolytic mechanism, typical of proteasomes. It is therefore tempting to investigate in greater depth how the proteasomal systems have evolved. However, efforts in this direction have not yet yielded a hypothesis, largely due to the limited knowledge of evolutionary precursors.

The recent identification of two proteasomal evolutionary intermediates, ancestral β subunit (Anbu), and betaproteobacterial homolog (BPH) have allowed the analysis of how transitions in symmetry mark pivot points in the evolution of the proteasome family (Fuchs et al., 2017, 2018; Fuchs and Hartmann, 2019). Characterization of the PAN-proteasome has also provided invaluable insights into the functional cycle of archaeal proteasomes (Majumder et al., 2019). Certain structural features such as ‘gate contacts’ have been revealed, which resemble the bipartite communication modes present in bacterial ClpXP (Martin et al., 2007). These contacts appear to mark intermediate steps in the progression from prokaryotes to eukaryotes.

With regards to an evolutionary understanding, the ubiquitously present proteins can also provide useful information. The Cdc48 proteins VAT, Cpa, and p97 exhibit distinct functional modes in archaea, actinobacteria and eukaryotes, respectively. While VAT and Cpa dock directly onto the CP, the eukaryotic homologues operate upstream of the 26S proteasome, and physical interactions with the 26S have not yet been reported. At this juncture, the in situ as well as ex situ structural characterization of other proteasome-ATPase complexes are necessary. It is important to identify and characterize the entire proteasome interaction network in both eukaryotic and non-eukaryotic degradation pathways, their sequence of interaction and regulatory mechanisms. Such an analysis would provide deeper insights into how proteasomes operate in the different domains of life, and would help to put forward hypotheses regarding their evolutionary connection.

Acknowledgments: This work has been supported by Deutsche Forschungsgemeinschaft SFB-1035/Project A01, Funder Id: http://dx.doi.org/10.13039/501100001659 (to W.B.).

References


