TECHNISCHE UNIVERSITÄT MÜNCHEN

Max-Planck-Institut für Biochemie

Abteilung für Molekulare Strukturbiologie

Structure of the 26S proteasome from *Schizosaccharomyces pombe* at subnanometer resolution

Stefan Bohn

Vollständiger Abdruck der von der Fakultät für Chemie der Technichen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender		Univ. – Prof. Dr. Chr. F. W. Becker
Prüfer der Dissertation:		
	1.	Hon. – Prof. Dr. W. Baumeister
	2.	Univ. – Prof. Dr. S. Weinkauf
	3.	Univ. – Prof. Dr. M. Sattler

Die Dissertation wurde am 26.01.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 26.05.2011 angenommen.

„ What is life? “

A Table of Contents

B Summary

The 26S proteasome is a macromolecular complex composed of two copies of 35 – 36 canonical subunits, which form the 20S core particle and the 19S regulatory particles. Here, cryo electron microscopy and single particle analysis in conjunction with chemical crosslinking and several labeling methods have been used to study the quaternary structure of the 26S holocomplex.

Firstly, a purification protocol for proteolytically active 26S holocomplexes in *Schizosaccharomyces pombe* has been established. Analysis by mass spectrometry has shown that the 26S proteasome from *S.pombe* is constituted of the canonical subunits α1-7, β1-7, Rpt1-6, Rpn1-9, and Rpn11-12. Furthermore, the subunits Rpn10, Rpn13a/b, Uch2 and Ubp6 are present in significant, yet varying amounts in preparations of 26S proteasomes.

The structure of the 26S proteasome from *S.pombe* has been determined to a resolution of $(9.1\ Å)^{-1}$ by cryo electron microscopy and single particle analysis. Patches with the characteristic features of short α-helical repeats are discernable, reminiscent of α-solenoidal folds predicted for RP subunits. The gate of the core particle adopts the closed conformation. Additional densities are present in the pockets between α4 – α3 and α2 – α1, presumably presenting C-terminal residues of Rpt2 and Rpt5, shown to be involved in gate opening.

Image classification and variance analysis have revealed a belt of high 'activity' at the base-lid interface surrounding the AAA-ATPase module, which is tentatively assigned to transiently bound proteasome-interacting proteins, ubiquitin or substrate and the conformational heterogeneity among the particles. The distal part of the lid complex, however, is rather invariable, thus providing a rigid frame or roof.

In order to purify the particle assembly *in silico*, maximum-likelihood classification has been used to separate 26S holocomplexes from partially (dis)assembled 26S proteasomes. 26S holocomplexes have been separated further according to the abundance of a distinct mass in the cap region ('extra mass'). In *D.melanogaster* and in *S.pombe*, the occupancy of the extra mass derived from the class averages correlates to the abundance of Rpn10 as determined by mass spectrometry. Classes of partially (dis)assembled 26S proteasomes always show densities resembling the AAA-ATPase hexamer. Moreover, some partially (dis)assembled

26S proteasomes show additional densities colocalizing to a stretch of low variance at the distal part of the regulatory particle, tentatively assigned to proteins with scaffolding function.

To further analyze the quaternary structure of the regulatory particle, different labeling techniques have been applied. Labeling experiments in *D.melanogaster* have demonstrated a high affinity of protein-gold conjugates to the base-lid interface of the regulatory particle. Due to this unspecific interaction the protein-gold conjugates are not suitable to label subunits of the regulatory particle.

The analysis of Uch2 deletion and GFP-fusion *S.pombe* mutants indicates a flexible binding of Uch2 in the 'mouth' region of the RP and further support the previously described interaction with Rpn13. Furthermore, the amount of identified assembly intermediates in Uch2-GFP and Uch2-deletion preparations suggests a role of the deubiquitinating enzyme Uch2 in the stability of the 26S holocomplex.

In addition, chemical crosslinking in conjunction with mass spectrometry has been used to identify numerous lysine pairs in close proximity to each other, providing an array of spatial restraints. These data clarify the topology of the AAA-ATPase module in the 19S regulatory particle and its spatial relationship to the α-ring of the 20S core particle. Antibody labeling and crosslinking data suggest that the C-terminal domain of Rpn11 is placed near the opening of the AAA-ATPase cavity, allowing efficient access to its substrate and release of ubiquitin.

An integrated model is presented which sheds light on the early steps of protein degradation by the 26S complex.

C Zusammenfassung

Als Grundlage für die Strukturbestimmung und der Analyse der biochemischen Zusammensetzung wurde zunächst eine Reinigungsmethode von proteolytisch aktiven 26S Proteasomen aus *S.pombe* etabliert. Analysen mittels Massenspektromie haben gezeigt, dass das *S.pombe* 26S Proteasom aus den kanonischen Untereinheiten α1-7, β1-7, Rpt1-6, Rpn1-9 und Rpn11-12 besteht. Des weiteren zeigte sich, dass die Untereinheiten Rpn10, Rpn13a/b, Uch2 und Ubp6 in stöchiometrisch unterschiedlichen Mengen an das 26S Proteasom gebunden sind.

Mit Hilfe der Kryoelektronenmikroskopie und Einzelpartikelanalyse wurde die Struktur des *S.pombe* 26S Proteasoms mit einer Auflösung von $(9.1\ Å)^{-1}$ rekonstruiert. In einigen Bereichen der Struktur sind α-helikale Elemente erkennbar, welche der typischen Anordnung α-solenoidaler Faltungen entsprechen und für Untereinheiten des regulatorischen Komplexes vorhergesagt wurden. Die Rekonstruktion zeigte eine geschlossene Konformation der Öffnung („gate“) des Kernkomplexes. Zusätzliche Dichten sind in den α4 – α3 und α2 – α1 Zwischenräumen erkennbar, welche den C-terminalen Aminosäureresten von Rpt2 und Rpt5 zugeordnet werden können und eine wichtige Rolle bei der Öffnung des gates spielen.

Mittels Bildklassifizierung und Varianzanalyse wurde ein Bereich hoher Aktivität zwischen der Basis und dem „Deckel“ des regulatorischen Komplexes, welcher das AAA-ATPase-Modul umgibt, lokalisiert. Diese Aktivität ist wahrscheinlich auf vorrübergehende Bindungen von Proteasom-interagierenden Proteinen, Ubiquitin oder Substrat und der strukturellen Heterogenität der Partikel zurückzuführen. Der weiter entfernte Teil des Deckels ist vergleichsweise stabil und bildet eine Art Dach oder festen Rahmen.

Mittels einer Maximum-Likelihood Klassifikation wurden vollständig assemblierte 26S Proteasomen von teilweise (dis)assemblierten Proteasomen getrennt. Vollständig assemblierte 26S Proteasomen wurden in einem zweiten Klassifikationsschritt anhand der Präsenz einer spezifischen Masse im regulatorischen Komplex (‚Extramasse‘) weiter separiert. In *D.melanogaster* und in *S.pombe* korreliert die durch die Klassifizierung nachgewiesene Stöchiometrie der Extramasse mit der Menge durch Massenspektrometrie detektiertes Rpn10.

In allen Klassen teilweise (dis)assemblierter 26S Proteasomen sind Dichten erkennbar, die dem AAA-ATPase Hexamer zuzuordnen sind. Weitere Dichten kolokalisieren mit den Bereichen hoher Aktivität im entfernteren Teil des Deckels des regulatorischen Komplexes. Diese Bereiche sind Proteinen zuzuordnen, die unter anderem für die strukturelle Integrität des 26S Proteasms verantwortlich sind.

Um die Quartärstruktur des regulatorischen Komplexes näher zu untersuchen, wurden verschiedene Proteinmarkierungsmethoden verglichen. Experimente mit *D.melanogaster* 26S Proteasomen haben gezeigt, dass Protein-Gold-Konjugate eine sehr hohe Affinität zum regulatorischen Komplex aufweisen. Aufgrund dieser unspezifischen Interaktion sind Protein-Gold-Konjugate ungeeignet, um Untereinheiten im regulatorischen Komplex zu identifizieren.

Die strukturelle und massenspektrometrische Analyse der 26S Proteasomen der Uch2 Deletions- und GFP-Fusionsmutanten hat gezeigt, dass das De-ubiquitilierungsenzym Uch2 flexibel im Bereich zwischen Basis und Deckel des regulatorischen Komplexes bindet. Des weiteren legt sie eine enge Interaktion von Uch2 und Rpn13 nahe. Darüber hinaus deutet die detektierte Menge an identifizierten (dis)assemblierten 26S Proteasomen darauf hin, dass Uch2 eine Rolle in der strukturellen Integrität des 26S Proteasoms spielt.

Durch chemische Vernetzung („crosslinking“) von Lysinseitenketten untereinander und deren massenspektrometrischen Analyse konnte eine Reihe von Lysinpaaren identifiziert werden, die sich in räumlicher Nähe zueinander befinden. Mit Hilfe dieser Daten war es möglich, die Topologie des AAA-ATPase-Moduls im regulatorischen Komplex und dessen räumliche Beziehung in Relation zum α-Ring des 20S Kernkomplexes aufzuklären. Eine Kombination von Markierungsexperimenten mit Antikörpern und Daten der chemischen Vernetzung legen nahe, dass sich die C-terminale Domäne von Rpn11 in der Nähe der oberen Öffnung des AAA-ATPase-Moduls befindet. Dies gewährleistet einen guten Zugang zum Substrat und die effiziente Freisetzung von Ubiquitin.

Abschließend wurde ein integriertes Model entwickelt, welches Einsichten in die ersten Schritte des Proteinabbaus durch das 26S Proteasom gibt.

1. Introduction

1.1. The Ubiquitin Proteasome Pathway

The ubiquitin-proteasome pathway (UPP) is the most important route for regulated proteolysis in the cytosol and nucleus of eukaryotic cells. The control of the lifespan of proteins and the degradation of misfolded proteins is essential to the function of living cells. Most proteins are affected by the UPP (Mayor et al., 2007) and malfunctions result in a wide variety of diseases (Glickman et al., 2002). Proteins targeted for degradation are modified by the attachment of ubiquitin (Ub) and are proteolytically cleaved in the most downstream element of the UPP, the 26S proteasome (Fig. 1).

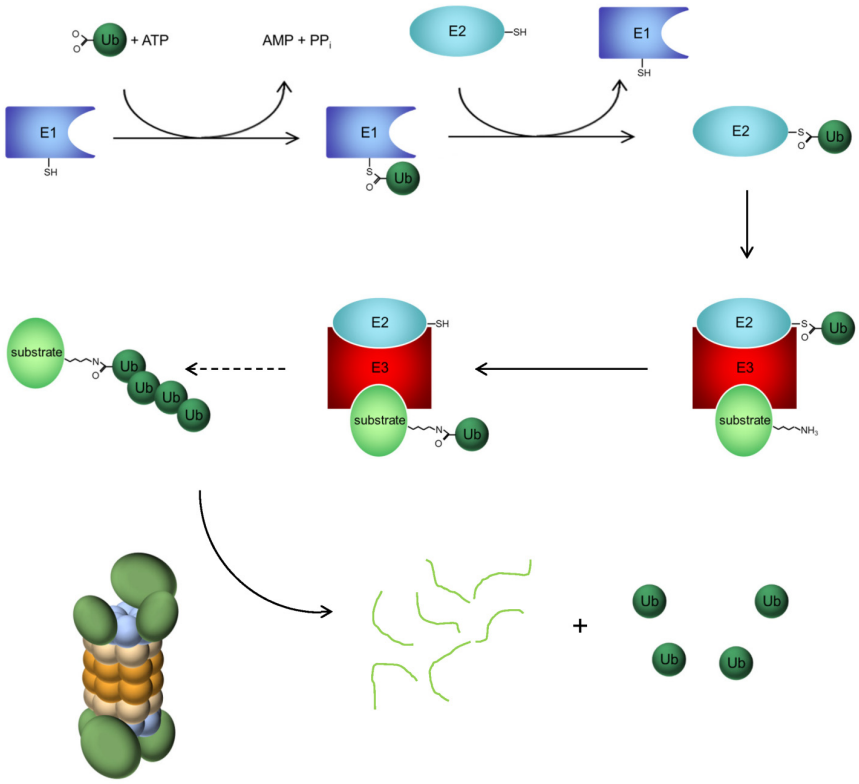

Figure 1: Protein degradation by the Ubiquitin Proteasome Pathway. E1 enzymes bind ubiquitin in an ATP-dependent reaction and transfer ubiquitin to a member of the E2 family. E3 enzymes conjugate the activated ubiquitin to a lysine residue of the target proteins, forming eventually polyubiquitin chains, which are recognized by ubiquitin receptors residing in the 26S proteasome. By cleaving off the ubiquitins, the pool of free ubiquitin is refilled, while the substrate is hydrolyzed into polypeptides typically seven to nine residues long. Figure adapted from (www.wikipedia.org, 2010).

1.2. Ubiquitylation of proteins

Essential for the UPP are the small protein ubiquitin (Ub) and three enzyme families: Ub activating enzymes (E1), Ub conjugating enzymes (E2) and Ub ligases (E3). E1 activates Ub in an ATP-dependent reaction by forming an E1~Ub intermediate, where Ub is linked via one of its C-terminal glycines to a cysteine in the active site of E1. This activated Ub is then transferred to a cysteine of E2, which either catalyzes the covalent attachment of Ub to target proteins or transfers the activated Ub to E3, forming a high energy E3-Ub intermediate. Finally, Ub is conjugated to a lysine of the target protein in the presence of E3. The E3s, which are responsible for the specific recognition of substrates, display the greatest variety among its family (Table 1).

Class	*Homo sapiens*	*Mus musculus*	*Drosophila melanogaster*	*Caenorhabditis elegans*	*Schizosaccharomyces pombe*	*Saccharomyces cerevisiae*
E1	16	16	11	9	9	8
E2	53	61	32	25	14	15
E3	527	442	189	657	82	68
DUB	74	78	27	39	20	20

Table 1: Ubiquitylation-associated proteins in eukaryotic genomes. Proteins associated to the (de)ubiquitylation of proteasomal substrates show highest diversity in mouse and human (except for the large number of F-box proteins among E3 ligases in *C.elegans*, (Kipreos et al., 2000)). The need of specific substrate recognition has led to the great variety among E3 ligases. Table adapted from (Semple, 2003).

1.3. The 26S proteasome

The 26S proteasome is a macromolecular complex composed of two copies of 34 – 35 polypeptide chains (Fig. 2), which form the 20S core particle (CP) and the 19S regulatory particles (RPs). The RPs bind to either one or both ends of the CP forming the 26S holocomplex. The proteasome is the most downstream element of the UPP, responsible for protein degradation (Hershko et al., 1998; Voges et al., 1999; Hershko et al., 2000). Unlike constitutively active proteases, the proteasome has the capacity to degrade almost any protein, yet it acts with exquisite specificity. The key stratagem is self-compartmentalization: the active sites of the proteolytic CPs are sequestered from the cellular environment in the interior of this barrel-shaped subcomplex (Baumeister et al., 1998).

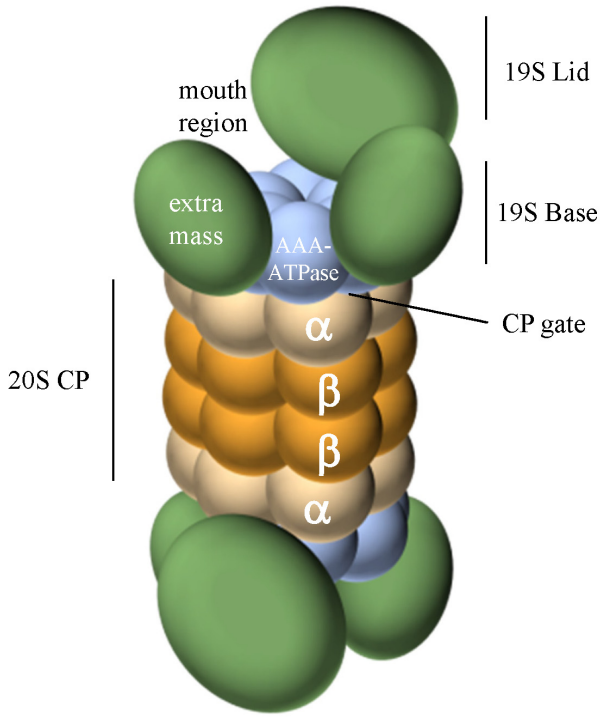

Figure 2: Schematic representation of the 26S proteasome. The 20S CP consists of four stacked heptameric rings (displaying α-β-β-α topology), sequestering the proteolytic chamber and two ante-chambers (between β-β and β-α, respectively) from the cellular environment. The CP gate is formed by the α-subunits. The AAA-ATPase hexamer is located in the RP base, stacked on top of the CP α-rings and binds the 'extra mass'. A region in close proximity to the AAA-ATPase hexamer between the 'extra mass' and the 19S lid is termed 'pharynx' or mouth region.

1.3.1. The 20S core particle

The 20S proteasome can be found in all three domains of life and consists of four stacked concentric heptameric rings, forming an elongated cylinder with three cavities and narrow constrictions between them. In eukaryotes, these rings consist of seven α and β subunits respectively, displaying an $\alpha_{1\text{-}7}\beta_{1\text{-}7}\beta_{1\text{-}7}\alpha_{1\text{-}7}$ topology. The structure of the CP has been elucidated in great detail by X-ray crystallography (Lowe et al., 1995; Groll et al., 1997).

The inner rings, formed by the β-type subunits, form the proteolytic chamber where the proteolytic active sites of β1, β2 and β5 are contained. These proteolytic subunits belong to the N-terminal nucleophile hydrolase family, with their N-terminal threonine acting both as catalytic nucleophile (on its side chain) and primary proton acceptor (the main chain N terminus). The active sites can cleave a broad range of peptides, and can be classified due to their cleavage preferences: caspase like (β1), trypsin like (β2) and chemotrypsin like (β5). In

mammals, these three subunits can be replaced upon γ-interferon signaling by βi1, βi2 and βi5, forming the so-called immuno-proteasome, exhibiting a higher efficiency to generate antigenic peptides.

The outer rings, formed by the α-type subunits act as an assembly platform for the 19S RP subunits and form a gate to the interior of the CP. At the α-α interfaces, seven α-pockets are formed on the RP-facing surface, providing binding sites for RP subunits. The N-terminal residues of the alpha subunits can interact with each other, thus forming the closed state of the gate. Upon occupancy of the α-pockets via interaction with some of the RP subunits, the gate of the CP can be opened.

1.3.2. The 19S regulatory particle

The 19S RP is composed of at least 19 different subunits (Fig. 3), six AAA-ATPases (Rpt1-6) and 13 non-ATPases (Rpn1-3, Rpn5-13, and Rpn15/Sem1p). The RP easily disassociates into two sub-complexes: the 'base', consisting of Rpn1, Rpn2, and the AAA-ATPases, and the 'lid', consisting of the remaining non-ATPases (Glickman et al., 1998). Whereas the structure of the RP is only dimly understood at present, it has been shown to be essential in a variety of regulatory functions. Polyubiquitylated proteins, i.e. proteins marked for degradation, are recognized by the RPs via their Ub receptors (e.g. Rpn10 or Rpn13). Subsequently, the polyubiquitin chains are cleaved and thus recycled by deubiquitylating enzymes (e.g. Rpn11), while the substrate is unfolded and translocated into the interior of the CP upon gate opening by AAA-ATPases. Thus, the RPs (*i*) recognize the polyubiquitylated substrates, (*ii*) trim and recycle the polyubiquitin chains, (*iii*) unfold substrates to be degraded, (*iv*) open the gate to the CP and (*v*) assist in substrate translocation into the interior of the CP.

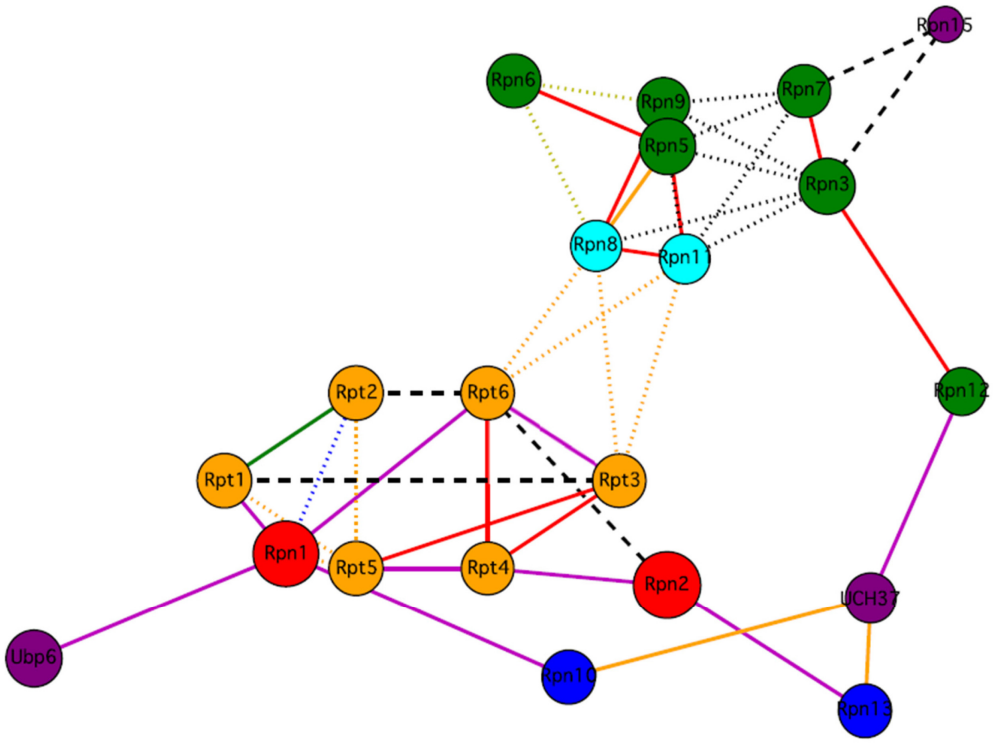

Figure 3: RP topology based on protein-protein interaction data. The network shows the protein-protein interactions determined by two-hybrid assays (red), *in vivo* pulldown experiments (orange), chemical cross-linking (dashed black), *in vitro* binding (magenta), and co-expression (green). Dotted lines indicate interactions involving more than two proteins. Figure adapted from (Förster et al., 2010).

1.3.3. AAA-ATPases

Out of all RP subunits, the AAA-ATPases are best characterized. They form a heterohexameric subcomplex situated at the base of the RP in close proximity to the α-rings of the CP (Nickell et al., 2009; Förster et al., 2010). The C-terminal residues of Rpt2 and Rpt5 were shown to be involved in opening the gate in the α-rings, allowing substrates to enter the CP. A similar mechanism has been shown for PAN, the archaeal homohexameric homolog of the eukaryotic AAA-ATPase module (Smith et al., 2007). Crystal structures of the two major fragments of PAN suggest that the N- and C-terminal domains form two stacked concentric rings (N-ring and AAA-ring, Fig. 4) (Djuranovic et al., 2009; Zhang et al., 2009a); the N-ring is implicated in substrate unfolding and the AAA-ring in gate opening and substrate translocation.

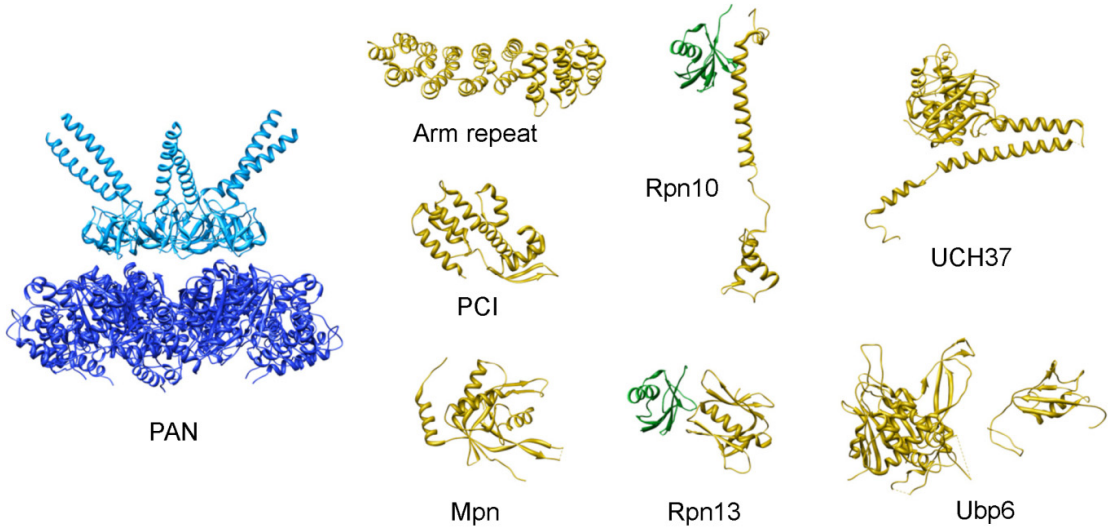

Figure 4: Structures of predicted domain folds of RP subunits. The AAA-ATPases Rpt1-Rpt6 form a heterohexameric ring and show high structural homology to the homohexameric **PAN** (crystal structures of the **light blue** N-terminal (Protein Data Bank code 2wg6) and **dark blue** AAA fold (3h4m) were determined separately). The PC repeats of Rpn1 and Rpn2 are predicted to adopt a solenoid-fold as in the **ARM repeat** (residues 282-584 of 1B3U) of the protein phosphatase PP2. The **PCI** domain fold (3chm) can be found in numerous RP subunits, as well as in the COP9 signalosome, which is structurally related to the RP. Rpn8 forms dimers in solution, with contacting **MPN** domains (2o95). The Ub receptors Rpn10 and Rpn13 contain distinct Ub binding motifs. Whereas **Rpn10** (1yx5) binds Ub (**green**) *via* helices of UIM, **Rpn13** interacts *via* the globular PRU domain (2z59) with Ub (**green**). The proteasome-associated DUBs **Uch37** (3ihr) and Ubp6 are members of the UCH family. The UCH fold of **Ubp6** (2ayn, left) is preceded by an N-terminal Ub-like domain (1wgg, right). Figure adapted from (Förster et al., 2010).

1.3.4. Ubiquitin receptors

Ub receptors associated with the proteasome bind Ub *via* their Ubiquitin-associated (UBA), ubiquitin-interacting motif (UIM) or pleckstrin-like receptor for ubiquitin (Pru) domain (Fig. 4). As of today, five proteasome-associated Ub receptors are known: Rad23, Dsk2, Ddi1 and the proteasomal subunits Rpn10 and Rpn13 (Finley, 2009). Whereas none of these subunits are essential in yeast, at least Rpn10 and Rad23 were found to be essential in higher eukaryotes (Girod et al., 1999; Ng et al., 2003; Smalle et al., 2003; Szlanka et al., 2003; Hamazaki et al., 2007). Rad23, Dsk2 and Ddi1 belong to the UBA/UBL protein family, binding Ub with their UBA domain and the proteasome with their N-terminal ubiquitin-like (UBL) domain. These receptors are functioning as a shuttle for Ub: they are not integral proteasome subunits, bind the proteasome only weakly and are typically present in substoichiometric amounts. Rpn10 recognizes polyubiquitin via its C-terminal UIM, while its

N-terminally located von Willebrand A (VWA) domain, seems to be important for proteasome structure (Glickman et al., 1998).

Rpn13 binds polyubiquitin chains with its Pru domain, whereas its C-terminal domain is known to interact with the Ubiquitin C-terminal Hydrolase Uch37/Uch2 (Fig. 5) (Qiu et al., 2006). Thus, proteasome-associated Ub receptors recruit polyubiquitylated proteins to the proteasome, allow access of proteasome-associated deubiquitylating enzymes to the polyubiquitin chain, and deliver the substrate to the degradation machinery.

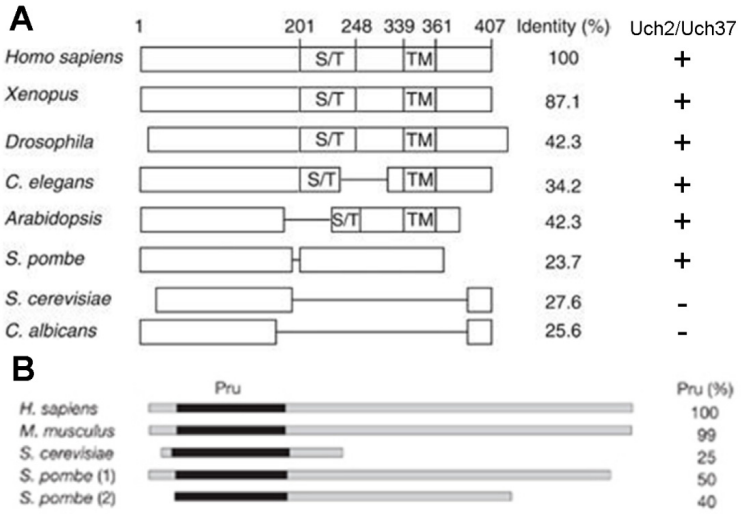

Figure 5: Architecture of Rpn13 from various species and its relation to Uch2/Uch37. A: hRpn13, especially its N-terminal half, is highly conserved throughout evolution. Homologs in several organisms show a high sequence identity to hRpn13 (S/T: Ser/Thr-rich region, TM: putative transmembrane domain). In *S.cerevisiae* and *C.albicans*, in which no homolog of Uch2/Uch37 is present, the C-terminal domain architecture differs strongly. **B:** In *S.pombe*, two orthologs of Rpn13 are present. Only Rpn13-1 (also termed Rpn13b) is capable of binding Uch2. A and B adapted from (Qiu et al., 2006) and (Husnjak et al., 2008).

1.3.5. Deubiquitylating enzymes

Deubiquitylating enzymes (DUBs) are responsible for Ub removal from substrates and Ub recycling (Verma et al., 2002; Yao et al., 2002), opposing the function of E3 ligases. In humans, ~ 80 DUBs are predicted to be active and have been categorized into five families (Fig. 6): ubiquitin C-terminal hydrolases (UCHs), ubiquitin-specific proteases (USPs), ovarian tumor proteases (OTUs), Josephins and JAB1/MPN/MOV34 metalloenzymes (JAMMs; also known as MPN+) (Nijman et al., 2005). They are either required for

proteasomal degradation (Rpn11) or can rescue substrates by deubiquitylation (Uch37/Uch2 and Usp14/Ubp6), thus regulating proteasomal degradation, i.e., they function as proteasome inhibitors (Lee et al., 2010).

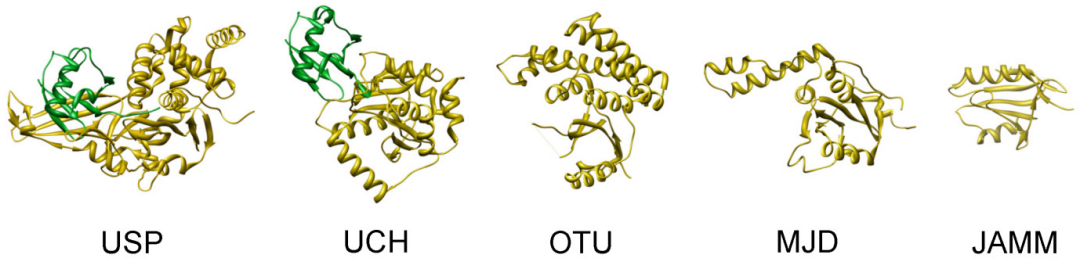

Figure 6: Domains of Deubiquitylating Enzymes. Based on sequence homology, DUBs are categorized into five families, showing a remarkable variability in secondary structure, while residues in the catalytic center are conserved (except Otu2) (Amerik et al., 2004). The DUB domains are colored in yellow, associated ubiquitin in green (PDB codes: USP7, 1nbf; UCH-L3, 1xd3; OTU2, 1tff; Ataxin-3, 1yzb; JAMM, 1r5x). Figure adapted from (Nijman et al., 2005).

The majority of DUBs encoded by the human genome belongs to the USP subclass (e.g. Usp14/Ubp6). Their number increased in a similar way during evolution as the number of E3 ligases, suggesting a close relationship between the two (Semple, 2003). The catalytic domain of USP contains two short and well-conserved motifs, called Cys and His boxes, which include the residues critical for catalysis.

Members of the UCH subclass have been among the first described DUBs, and are thought to mainly act in the recycling of Ub when Ub is inappropriately conjugated to intracellular nucleophiles (Nijman et al., 2005). UCHs prefer to cleave relatively small protein substrates (up to 20–30 amino acids) from Ub (Amerik et al., 2004), but larger substrates can also be cleaved (Misaghi et al., 2005).

Rpn11 is a constitutive subunit of the 26S holocomplex and belongs, together with Rpn8, to the Mpr1/Pad1 N-terminal (MPN) domain containing 26S subunits. Several studies have demonstrated the interactions between the MPN-containing subunits Rpn8 and Rpn11 (Cagney et al., 2001; Fu et al., 2001; Giot et al., 2003; Tanaka, 2009) which are likely to interact *via* their MPN domains, as observed in dimers formed in solution by Rpn8 (Sanches et al., 2007). The zinc-binding JAMM motif, which is only present in the MPN domain of Rpn11, is responsible for its deubiquitylating function (Verma et al., 2002; Yao et al., 2002).

Uch37/Uch2 and Usp14/Ubp6 are typically present in substoichiometric amounts in 26S proteasome preparations. Uch37 has been found to interact with the Ub receptors Rpn10 (Stone et al., 2004) and Rpn13 (Qiu et al., 2006). In most organisms, Uch37 binds via its C-terminal coiled-coil domain (Fig. 4) to the C-terminal domain of Rpn13, whereas eukaryotes such as *Saccharomyces cerevisiae* – in which no Uch37 homolog can be found – Rpn13 lacks this UCH-interacting domain.

1.3.6. Scaffolding proteins of the 19S RP

Rpn1 and Rpn2 are the largest subunits of the RP (~100 kDa and ~110 kDa, respectively) and function as scaffolds for the assembly of the AAA-ATPases (Funakoshi et al., 2009; Kaneko et al., 2009; Roelofs et al., 2009; Saeki et al., 2009). Both proteins contain large repetitive elements, proteasome cyclosome (PC) repeats, which are predicted to adopt α-soleonid folds (Kajava, 2002). In large protein assemblies, e.g. the nuclear pore complex, α-helical repeats are common scaffolds (Brohawn et al., 2009), forming highly diverse interfaces (Kobe et al., 1999).

1.3.7. PCI module-containing non-ATPases and Rpn15

The PCI module (Proteasome, COP9 signalosome, eIF3) can be found in a number of RP subunits (Rpn3, Rpn5, Rpn6, Rpn7, Rpn9 and Rpn12) as well as in subunits of the structurally related COP9 signalosome and the translation initiation factor eIF3. PCI modules of subunits eIF5 and Csn7 of eIF3 and COP9, respectively, have been analyzed by X-ray crystallography (Wei et al., 2006; Dessau et al., 2008): they are composed of a N-terminal helical bundle (HB) and a C-terminal winged-helix (WH). Thus, contrary to initial expectations, the PCI module is not a single structural domain, but rather two domains. PCI module-containing subunits of the RP are predicted to be preceded by repetitive bihelical blocks, similar to tetratrico-peptide repeats (TPRs), forming α-solenoids together with the HB-folds. WH domains can interact with nucleic acids (Brennan, 1993), which suggests that the WH domains in the PCI-module containing RP subunits are important for the role of the proteasome in DNA repair. For example, Rpn15, in complex with the PCI-module containing subunits Rpn3 and Rpn7, mediates proteasome association to DNA double-strand breaks (Krogan et al., 2004) and may be pivotal for regulation of proteins involved in DNA repair (Pick et al., 2009).

1.4. Cryo electron microscopy and Single Particle Analysis

Cryo electron microscopy (cryoEM) enables the structural analysis of thin (~100 nm or less) biological specimen in their close to native environment at cryogenic temperatures utilizing transmission electron microscopes (TEMs). In contrast to other preparation methods, such as plastic embedding or the 'negative stain' preparation, the samples are not dried, i.e., the macromolecules remain in their physiological hydrated environment. Structures of biological molecules are 'fixed' in amorphous (vitreous) ice. Vitrification is achieved by rapid (plunge) freezing of thin (< 1 μm) samples in liquid ethane near liquid nitrogen temperatures (~ -170 °C). The rapid cooling avoids the formation of crystalline ice, which otherwise would destroy the native structure of macromolecules or cells. The frozen-hydrated specimen is then transferred into the high-vacuum environment of a TEM for high resolution imaging. Since cryoEM has been introduced in the 1970's (Taylor et al., 1974; Dubochet et al., 1988), it was successfully applied to a wide variety of fundamental biological questions, while all aspects of it – from specimen preparation to image acquisition and data analysis – are still a subject of constant improvement.

Micrographs acquired by cryoEM are 2D projections of 3D objects. Single Particle Analysis (SPA) allows recovering the 3D information of an object from a set of 2D projection averages. SPA assumes the presence of many structural identical copies of macromolecular assemblies, which generate identical 2D projections when imaged in the same orientation. When macromolecular complexes exhibit structural variability, flexible structural elements or substoichiometrically bound subunits are less resolved then invariable regions. However, advanced image processing techniques allow classifying projection data according to encountered structural variability. Taken together, cryoEM in conjunction with SPA is a powerful tool to analyze the structural arrangement of one of the largest macromolecular assemblies in eukaryotic cells, the 26S proteasome.

2. Aim of this work

A precise localization of proteasomal subunits would help to elucidate the sequence of events from the initial binding of substrates to their translocation into the CP. High-resolution structural analysis by X-ray crystallography would allow the precise assignment of polypeptide chains but it depends on the availability of suitable crystals. Crystallization attempts have failed to date, due to the complexity of the 26S proteasome and its fragile nature, its dynamics and the association-dissociation of proteasome interacting proteins resulting in their presence in variable amounts. For cryoEM and for protein-protein interaction studies the requirements for sample homogeneity are less stringent. A ~ 25 Å resolution structure of the *D.melanogaster* 26S proteasome has been reported recently (Nickell et al., 2009). The localization and the boundaries of the AAA-ATPase module could be approximated. The primary aim of this work is to further explore the possibilities of cryoEM in terms of determining the structural organization of the 26S holocomplex.

Genetic engineering allows the fusion of high-affinity tags to the N- or C-terminus of proteins, which in turn can function as suitable labels for cryoEM ('extra density'). Purification of 26S holocomplexes using a Flag affinity tag has been recently established in the easily modifiable *S.cerevisiae* (Saeki et al., 2005). Since not all human proteasome-associated subunits are conserved in *S.cerevisiae*, the first aim of this study was to establish 26S proteasome preparations of sufficient purity for cryoEM in the model organism *S.pombe*. The 3D reconstruction of the 26S holocomplex isolated from wildtype *S.pombe* is supposed to yield a reference of medium resolution for comparison to labeled complexes.

Several labeling methods suitable for analysis by cryoEM are assessed in this study in order to localize RP subunits (Fig. 7): (*i*) binding of electron dense gold particles conjugated to specific inhibitors of RP subunits (e.g. ubiquitin-aldehyde, targeting the ubiquitin C-terminal hydrolase Uch37 in *D.melanogaster*); (*ii*) fusion of polypeptides to the termini of RP subunits, creating an extra density (e.g. GFP fusion to the C-terminus of Uch2 in *S.pombe*); (*iii*) deletion of RP subunits, creating a 'missing density'; and (*iv*) binding of monoclonal anti-bodies (e.g. monoclonal anti-Flag antibody, targeting the Flag epitope used for proteasome purification).

In order to gain further insights into the structural organization of the 26S holocomplex, chemical crosslinking in conjunction with mass spectrometry (Leitner et al., 2010; Rinner et al., 2008) is used to identify numerous residue pairs (crosslinked lysines) in close proximity to each other. Furthermore, the cross-linking data can be used to cross-validate results from labeling experiments.

The integration of these data and the EM maps allow generating a structural model providing new insights into the organization of the 26S holocomplex.

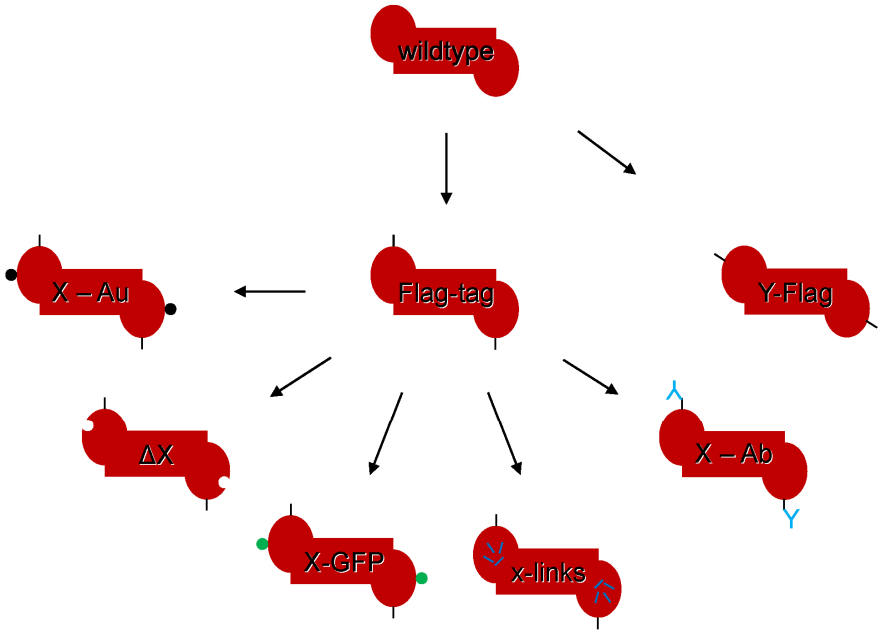

Figure 7: RP labeling techniques employed in this study. For purification of 26S holocomplexes, a Flag-tag can be fused to different RP subunits (**Flag-tag, Y-Flag**), allowing antibody-labeling of different subunits with the same high-affinity monoclonal anti-Flag antibody (**X-Ab**). By addition of protein densities (**X-GFP**) or electron dense, modified gold particles (**X – Au**) or deletion of RP subunits (**ΔX**) differences in the EM reconstructions allow mapping of specific subunits. Chemical cross-linking in conjunction with MS (**x-links**) provides further spatial constraints of RP subunits.

3. Materials and Methods

Chemicals and media were purchased from Difco (BD Biosciences, Franklin Drive, NJ, USA), Fluka (Buchs, Switzerland), Merck (Whitehouse Station, NJ, USA), Roth (Karlsruhe, Germany), Serva (Heidelberg, Germany) or Sigma-Aldrich (St. Louis, MO, USA), unless noted otherwise. A list of apparatuses can be found elsewhere (Bopp, 2008).

3.1. Primers, Plasmids and Strains

3.1.1. Primers

Primers were purchased from MWG Biotech, Ebersberg, Germany and diluted to a concentration of 100 µM.

GFP – Forward

ATGCGTCGACCATCATCACCATCACCACAAAGGAGAAGAACTC

GFP – Reverse

CTGAAGATCTTTACCCAGCAGCTGTTAC

Rpn11 – 3xFLAG – HIS3 – Forward

GTAAACAAGATCCCAAGAAGCATCTTGACGCTGAGGTTCAGAAATGTATAGACAACAATATTTCTTCCA
TGTTAGCTTGTATGCTTGATTCCGTTGCCTTCGGATCAGGATCACTGGAAGTTC

Rpn11 – 3xFLAG – HIS3 – Reverse

CAACAGGATTATTCAAACAGTTCATAAATGCTTTGAGTAAGGAAAGAAAGTAGCTCATAAAAAACGGA
TGATGGAAAAGAGGTATTACCAAAATAGATGTATCTGTTTAGCTTGCCTCGTC

ΔUch2 – Forward

TTAGCGAATTTTAAAGTAATACCATCATTTGATCTCTCATAAAAGTCAGCTTGCATAAGTGCTGTCAAAT
TCCATTCTTTCGTTCTTCTGTAACACACAAGCGTACGCTGCAGGTCGAC

ΔUch2 – Reverse

ATAAGTCCAACAAGGTAATCTAGAGTCATAACAGATTTTTATCCTGGAAATGAAAATTCCTTCATAGTG
GAATATTTTTTGAGAAAGAACGCCGTAATATAATCGATGAATTCGAGCTCG

Uch2 – GFP – Forward

AACTTTCCAAACTTCTGGTAAAAGATCGCATTGACAAGAATACTTGGAACTCTACCCTTGAAACTGCAA
AAGCGAAATACGCATCTCAAAAGCGACCGCGTACGCTGCAGGTCGAC

Uch2 – GFP – Reverse

ATAAGTCCAACAAGGTAATCTAGAGTCATAACAGATTTTTATCCTGGAAATGAAAATTCCTTCATAGTG
GAATATTTTTTGAGAAAGAACGCCGTAATATAATCGATGAATTCGAGCTCG

3.1.2. Plasmids and Strains

Plasmid pBlueScript 3xFlag-His-KanMx has been kindly provided by Y. Saeki (Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan). Eggs of *D.melanogaster* and *S.pombe* wildtype strain 972 h+ and plasmid pYM17-Nat were kindly provided by Oana Mihalache (MPI of Biochemistry, Martinsried) and M. Shravan (MPI of Biochemistry, Martinsried). The 3xFlag-His-KanMx cassette was introduced to *S.pombe* wildtype to generate 972 h+ rpn11::rpn11-3xFLAG-HIS3 (Rpn11-Flag, wildtype). The pYM17-Nat and pYM17-GFP-Nat cassette were introduced to Rpn11-Flag to generate 972 h+ rpn11::rpn11-3xFLAG-HIS3 Δuch2 (ΔUch2) and 972 h+ rpn11::rpn11-3xFLAG-HIS3 uch2::UCH2-GFP (Uch2-GFP).

3.2. Media, buffers and solutions

3.2.1. Antibiotics

In LB medium and plates ampicillin was used in a final concentration of 100 µg/ml. Strains carrying the KanMx or Nat resistance were selected on YES with 200 mg/l G418 or 100 mg/l Nurseothricin.

3.2.2. Buffers and Solutions

DEAE: 20 mM Hepes pH 7.6, 5 % Glycerol, 15 % Saccharose, 5 mM $MgCl_2$, 5 mM DTT, 2 mM ATP

Dialysis buffer: 5.6 g/l Hepes, 57.4 ml/l Glycerol, 1 g/l $MgCl_2$, 4.7 g/l NaCl, 0.76 g/l DTT, 1.1 g/l ATP

HA: 25 mM K-PO_4 (pH 6.7), 5 % Glycerol, 5 mM $MgCl_2$, 5 mM DTT, 2 mM ATP

HO: 20 mM Tris-HCl (pH 7.5), 0,25 M Saccharose, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 2 mM ATP

Lysis buffer: 50 mM Tris-HCl (pH 7.1), 10 mM NaCl, 10 mM $MgCl_2$, 4 mM ATP, 5 mM DTT, 10 mM Creatinephosphate, 3 U/ml Creatinephosphate kinase, 10% Glycerol

Sample buffer: 20 mM HEPES (pH 7.1), 20 mM NaCl, 10 mM $MgCl_2$, 4 mM ATP, 5 mM DTT, 10 mM Creatinephosphate, 3 U/ml Creatinephosphate kinase

TBE (5x): 445 mM Tris-HCl pH 8.0, 445 mM boric acid, 10 mM EDTA

TfbI: 100 mM RbCl, 50 mM $MnCl_2$, 30 mM Potassium acetate, 10 mM $CaCl_2$, 15 w% Glycerol, adjusted with HAc to pH 5.8, sterile filtered

TfbII: 10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15 w% Glycerol, adjusted with KOH to pH 6.8, sterile filtered

3.2.3. Media

LB: 10 g/l Bactotryptone, 10 g/l NaCl, 5 g/l Yeast Extract

LB plates: LB with 15 g/l Agar

SOC: 20 g/l Bactotryptone, 5 g/l Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 20 mM Glucose and 20 mM $MaCl_2$, adjusted with NaOH to pH 7.0

SORB: 100 mM LiAc, 10 mM Tris pH 8.5, 1 mM EDTA, 1 M Sorbitol, adjusted to pH 8, sterile filtered

YEL: 5 g/l Yeast Extract, 2 g/l Casamino Acids, 30 g/l Glucose, 0.1 g/l L-Histidine, 0.1 g/l L-Leucine, 0.1 g/l Uridine, 0.1/l g Adeninehemisulfate

YES: YEL with 18 g/l Agar

3.3. Genetic Modifications

3.3.1. Agarose gel electrophoresis

0.75 g (0.75 %) agarose was dissolved in 100 ml TBE buffer by heating and 5 µl of a 1% ethidiumbromide solution was added. Gels were run at 100 V. Samples were mixed with 6x loading dye and run with a GeneRulerTM 1 kb DNA ladder (MBI Fermentas) as standard marker.

3.3.2. Polymerase chain reaction

Standard polymerase chain reaction (PCR) was performed to amplify GFP and for testing genomic insertions (Table 1). Modified PCR was used to amplify DNA for transformation into *S.pombe* cells (Janke et al., 2004) (Table 2) using the PCR-Thermocycler GeneAmp PCR System 2400 (Perkin Elmer).

Reaction composition	
Water	38,7 µl
DNA template	1,0 µl
Pfu-Polymerase (50 U/µl)	0,5 µl
dNTPs (10 mM)	1,6 µl
Forward Primer (10 µM)	1,6 µl
Reverse Primer (10 µM)	1,6 µl
10x reaction buffer	5,0 µl
Final Volume	50,0 µl

Thermal cycles			
Denaturation	5 min	94 °C	
Denaturation	30 s	94 °C	
Annealing	30 s	54 °C	30 x
Polymerisation	60 s	72 °C	
Final Polymerisation	10 min	72 °C	
	∞	4 °C	

Table 2: PCR reaction conditions and thermal cycle setup for standard PCR.

Reaction composition	
Water	35,85 µl
DNA template	1,0 µl
Taq-Polymerase (50 U/µl)	0,4 µl
Vent-Polymerase (50 U/µl)	0,2 µl
dNTPs (10 mM)	1,75 µl
Forward Primer (10 µM)	3,2 µl
Reverse Primer (10 µM)	3,2 µl
10x reaction buffer	5,0 µl
Final Volume	50,0 µl

Thermal cycles			
Denaturation	5 min	95 °C	
Denaturation	30 s	95 °C	
Annealing	30 s	54 °C	10 x
Polymerisation	160 s	68 °C	
Denaturation	30 s	95 °C	
Annealing	30 s	54 °C	20 x
Polymerisation	160s + 20s/cycle	68 °C	
Final Polymerisation	5 min	68 °C	
	∞	4 °C	

Table 3: Modified PCR reaction conditions and thermal cycle setup for transformation of *S.pombe* cells.

3.3.3. PCR product purification

PCR products from standard PCR were purified with the QIAquick PCR Purification Kit (QIAGEN). For modified PCR, one tenth volume of 3 M NaOAc, pH 5.2 was added to DNA. Two volumes of cold 95% Ethanol were added and samples were placed at -80 °C for 30 min or at -20 °C overnight. DNA was pelleted (20 min, 21.000 g) and the supernatant discarded. The pellet was dried at 40°C and re-dissolved in 20 µl EB-buffer (QIAGEN).

3.3.4. DNA digestion and ligation

GFP PCR product and pYM17 plasmid were digested with restriction enzymes BglII and SalI for 2h at 37 °C (3µl 10x reaction buffer, 1 µl SalI, 1 µl BglII and 25 µl PCR product or 5 µl plasmid DNA and 20 µl H_2O respectively). Digestion reaction was analyzed by agarose gel electrophoresis and purified using the QIAquick Gel Extraction Kit (QIAGEN). Ligation was carried out over night at 16 °C (3µl 10x reaction buffer, 12.5 µl digested PCR product, 1 µl digested plasmid DNA, 1 µl T_4 Ligase, 12.5 µl H_2O).

3.3.5. Competent *E.coli* cells

E.coli cells were grown in LB medium at 37 °C to an OD_{600nm} of 0.7. Cells were cooled down on ice for 15 min and harvested (10 min, 1000 g, 4 °C). Pellets were resuspended in 100 ml TfbI solution and incubated for 15 min. Cells were pelleted (5 min, 2000 g, 4 °C). Pellets were gently resuspended in 5 ml TfbII solution and incubated for 45 min on ice. Cells were shock frozen in liquid Nitrogen and stored in 100 µl aliquots at -80 °C.

3.3.6. Transformation of *E.coli* cells

Competent cells (100 µl aliquots) were thawed on ice, incubated with ~ 100 ng plasmid DNA for 15 min on ice, heat-shocked at 42 °C for one minute and cooled on ice for two minutes. 100 µl SOC were added and cells were incubated for 30 min at 37°. Cells were plated on LB-Amp plates and grown over night at 37 °C. *E.coli* clones were selected from plates and shaken in 3 ml LB-Amp medium at 37 °C overnight. Positive clones were confirmed by plasmid restriction and analysis by agarose gel electrophoresis or by sequencing (MWG, Ebersberg, Germany).

3.3.7. Plasmid DNA preparation

For plasmid preparation the QIAprep Spin Miniprep Kit (QIAGEN) was used.

3.3.8. Competent *S.pombe* cells

S.pombe cells were grown in 100 ml YEL medium at 30 °C to an OD_{600nm} of 0.7, pelleted (5 min, 500 g, 4 °C) and washed with H_2O and SORB medium. Cells were resuspended in 720 µl SORB and 80µl salmon sperm DNA and stored at -80 °C in 60 µl aliquots.

3.3.9. Transformation of competent *S.pombe* cells

50 µL competent cells were thawed on ice and mixed vigorously with 10µl purified PCR product. A six-fold volume of Polyethylenglycol (PEG 3350) was added and cells were mixed and incubated at 30 °C for 30 min. After heat-shock (42 °C, 10 min, sometimes mixing), 1 ml YEL medium was added and cells were pelleted (2 min, 13.000 rpm, Eppendorf centrifuge). Supernatant was removed, pellet re-dissolved in 2 ml YEL medium and incubated over night at 30 °C in 15 ml falcon tubes. Cells were sedimented (5 min, 4.500 rpm, Rotanta centrifuge). Supernatant was discarded, pellet resuspended in 200 µL YEL and plated on YES selection plates.

3.4. Protein Purification, Analysis and Labeling

3.4.1. Purification of 26S proteasomes from *D.melanogaster*

Intact 26S proteasomes from *D.melanogaster* were purified as described (Hölzl et al., 2000). Cells were homogenized in HO buffer and clarified by centrifugation (1h, 4 °C, 25.000 rpm, SW28). Nucleic acids were removed after stirring on ice for 10 min in 10% streptomycin sulfate by centrifugation (10min, 4 °C, 20.000 rpm, SS34). The supernatant was rotated on a hydroxylapathit column (15g Hydroxylapathit (BioRad Laboratories GmbH, Munich, Germany), 30 ml HA buffer) for 40 min at 4 °C and sedimented (10 min, 4 °C, 500 rpm, Rotanta). The column was washed three times with HA buffer and proteasomes were eluted with high salt HA buffer (HA buffer with 250 mM K-PO_4 pH 6.7). The elute was clarified (30 min, 4 °C, 20.000 rpm, SS34) and dialyzed over night at 4 °C in dialysis buffer. The dialyzed sample was washed on a DEAE column (15g DE52 from Whatman in 20 ml DEAE buffer) with DEAE buffer. Proteins were eluted on a FPLC over night (flow rate 1 ml/min, NaCl gradient of 80-400 mM in DEAE buffer with 40 % Saccharose). Fractions were concentrated to ~ 10mg/ml protein and subjected to a 15-40 % sucrose gradient (20 mM

Hepes 7.6, 80 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT, 2 mM ATP) for centrifugation (20 h, 4 °C, 28.000 rpm, SW28).

3.4.2. Purification of 26S proteasomes from *S.pombe*

Intact 26S proteasomes containing a 3xFLAG-tag at the C-terminus of Rpn11 from *S.pombe* were purified essentially as described (Saeki et al., 2005). Cells were cultured in YEL medium at 32 ˚C, and subsequently lysed in lysis buffer. Lysates were cleared by centrifugation (30 min, 4 °C, 55.000 g, Ti55) and incubated with anti-FLAG M2 agarose beads for 2h at 4 °C, which were then washed three times with lysis buffer. The bound proteins were eluted with 400 g/ml 3xFLAG peptide (Sigma) and subjected to a 15-40 % sucrose gradient (20 mM Hepes, pH 7.1, 20 mM NaCl, 10 mM $MgCl_2$, 4 mM ATP, 5 mM DTT, 10 mM Creatinephosphate, 3 U/ml Creatinephosphate kinase) for centrifugation (1000 min, 4 °C, 28.000 rpm, SW41).

3.4.3. Protease activity test

Each fraction from sucrose gradients was tested for their ability to hydrolyze the substrate Succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin (Suc-LLVY-AMC). 10 µl sample were mixed with activity test buffer (79 µl H_2O, 1 µl substrate, 10 µl 0.5 M Tris-HCl pH 8.0) and fluorescence of free AMC was measured at λ_{ex} = 360 nm / λ_{em} = 460 nm.

3.4.4. Mass Spectrometry

The *S.pombe* Rpn11-Flag and ΔUCH2 26S proteasome containing fraction with the highest proteolytic activity showed the highest 26S holocomplex concentration and was analyzed by mass spectrometry as described (Nickell et al., 2009). In short, 20 µl of the fraction with the highest amount of 26S holocomplexes were digested in solution by trypsin, separated by online reversed-phase nanoscale capillary liquid chromatography (nanoLC) and analysed by electrospray tandem mass spectrometry (ES-MS/MS). The experiments were performed with an Agilent 1100 nanoflow system connected to an LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). Peptide intensities for each protein were summed up and

normalized by the molecular weight of the protein. Resulting intensities were standardized by the average of all 20S protein intensities.

3.4.5. Colloidal gold particles

Colloidal gold particles were prepared essentially as described (Baschong et al., 1985). In short, 0.3 ml NaSCN were added to 50 ml H_2O with 0.5 ml of a 1 % $HAuCl_4$ and 0.75 ml 0.2 M K_2CO_3. The solution was stirred overnight in the dark at room temperature. Size of gold particles was determined by electron microscopy after applying 5 µl solution to glow discharged continuous carbon coated copper grids.

3.4.6. Conjugation of protein to colloidal gold

50 µg of ubiquitin aldehyde (Ub-Al) or bovine serum albumin (BSA) were solubilized in 50 µl 50 mM triethanolamine, pH 7.3 (TEA). To introduce sulfhydryl groups in ubiquitin for stronger binding of the protein to the surface of gold particles, 5 µl of UbAl or BSA (control reaction) solution were incubated with 4 µg of 2-Iminothiolane and 11 µl TEA at room temperature for six hours or overnight at 4 °C.

3.4.7. Gold-labeling of 26S proteasomes

For colloidal gold labeling, 1 µl of gold-bound protein solution was incubated with 4 µl purified 26S proteasomes from *D.melanogaster* for one hour on ice prior to negative staining. For antibody labeling, 20 µl purified proteasomes from *S.pombe* Rpn11-Flag were incubated with 20 µl, 2 µl or 0.2 µl of 1 mg/ml M2 Anti-FLAG antibody (Sigma) for one hour at 4 °C prior to vitrification.

3.4.8. Chemical crosslinking

Experiments were carried out at the laboratory of Rudi Aebershold (ETH Zurich, Zurich, Switzerland). In short, purified Rpn11-Flag proteasome samples were concentrated to 1 mg/ml using a Roti-Spin-MINI-30 cut-off filter device (Roth). 50 µl sample were crosslinked with 1 mM disuccinimidyl suberate d0/d12 (DSS, Creativemolecules Inc.) in

sample buffer at 35 °C for 30 min. The reaction was quenched by adding 50 mM ammonium bicarbonate for 10 min at 35 °C.

The protein samples were reduced with 1 mM Tris (2-Carboxyethyl) phosphine hydrochloride (TCEP, Pierce) at 37 °C for 30 minutes and subsequently alkylated with 2 mM iodoacetamide (Sigma-Aldrich) for 10 minutes at room temperature in the dark. For digestion 10% acetonitrile (ACN), urea (1M final concentration) and 2 % w/w trypsin (Promega) was added. Digestion was carried out at 37 °C overnight and stopped by acidification to 1% (w/v) trifluoroacetic acid (TFA). Peptides were purified with C18 MicroSpin columns (The Nest Group, Southborough, MA), according to the manufacturers protocol. Subsequently cross-linked peptides were enriched using a peptide size exclusion chromatography. LC-MS/MS analysis was carried out on an LTQ Orbitrap XL mass spectrometer (Thermo Electron, San Jose, CA).

Data were searched using xQuest (Rinner et al., 2008) in iontag mode against a database containing the protein sequences of previously identified proteasome proteins with a precursor mass tolerance of 10 ppm. For matching of fragment ions tolerances of 0.2 Da for common-ions and 0.3 Da for cross-link ions were used. Cross-linked peptides identified with a linear discriminant (ld) score > 25, corresponding to a false discovery rate is < 5 % (Leitner et al., 2010), were further analyzed manually. Thereby, matching ion series on both cross-linked peptide chains were reviewed by visual inspection in order to match the most abundant peaks.

3.5. Electron Microscopy

3.5.1. Negative Staining procedure

Carbon coated copper grids (PLANO W. Plannet GmbH, 100x400 or 100x100 mesh) were glow-discharged for 15-30 s with a plasma cleaner (Harrick Scientific Corporation, NY, USA). 5 µl sample were applied, incubated at room temperature for 1 min, washed 2x with H_2O and stained with 2 % uranylacetate for 1 min.

3.5.2. Data acquisition of negatively stained specimen

Electron micrographs of negatively stained specimen were acquired at a CM20 transmission electron microscope (Philips, Eindhoven, Netherlands) operated at an acceleration voltage of 160 kV at room temperature. Images were recorded on a 4096 x 4096 CCD camera (TVIPS, Gauting, Germany) with a nominal underfocus of 2 µm (final magnification: 54.000×, object pixel size: 2.78 Å).

3.5.3. Vitrification

Lacey carbon copper grids (Plano GmbH, Wetzlar, Germany) were glow discharged for 30 s and 4.5 µl sample was applied. Excess fluid was blotted using filter paper from the side after 45 s of incubation time and the sample was washed twice with H_2O and blotted finally from the front before vitrification in liquid ethane. The grids were stored in liquid nitrogen.

3.5.4. Data acquisition of vitrified specimen

For high-resolution single particle reconstruction, a large amount of particles is required to increase the signal to noise ratio (SNR) at different underfocus values (1-3 µm). Manual particle selection is not possible at very low defocus values (< 1.5 µm), whereas automated particle selection procedures rely on identical acquisition parameters for optimal use. Particles were acquired at two defocus values ('focal pairs'): a first image was acquired at a low underfocus value (~ 1-2.5 µm) for high-resolution reconstruction and a second image was acquired at the same position at a high underfocus (~ 3 µm) for particle selection and classification. The constant underfocus of the second acquired image allowed to use their high contrast necessary for particle selection and avoided classification bias due to different defocus value.

For wildtype *S.pombe* and ΔUch2 *S.pombe* focal pairs with a nominal defocus of 1-3 µm were recorded on an Eagle CCD camera using a FEI Tecnai F20 microscope (FEI, Eindhoven, Netherlands) operated at 200 kV (final magnification: 63,500×; object pixel size: 2.21 Å) in a semi-automated manner using the SerialEM software package (Mastronarde, 2005) (Table 4). For *S.pombe* wildtype-Ab and Uch2-GFP *S.pombe*, micrographs with a nominal defocus of 2.5 ± 0.5 µm were recorded. The final dose per micrograph was

10 – 20 e^- / $Å^2$. Particles were preselected with an automated picking routine of the TOM toolbox (Nickell et al., 2005) and datasets were further refined manually. In the case of focal pairs, particles were selected on the high defocus images. Even minute changes in stage or beam position occurring during the acquisition of focal pairs, will lead to a significant offset of the high and low defocus image. Image shifts of focal pairs were determined by cross-correlation and coordinates from high defocus images were transferred to low defocus images and checked interactively. From ~30,000 micrographs, ~380,000 particles were selected for 3D reconstruction (Table 4).

	wildtype	wildtype-Ab	ΔUch2	Uch2-GFP
Acquisition days	18	7	5	2
Number of focal pairs	7,933	-	2,517	-
Number of total images	17,416	5,659	5,034	2,026
Number of particles	270,061	51,057	44,139	28,590
Particles per image	27	9	12	14

Table 4: Single Particle datasets acquired in this work. For wildtype and ΔUch2 proteasomes, focal pairs (low defocus: ~ 1-2.5 µm, high defocus: ~ 3 µm) and single images (defocus: ~ 3 µm) have been acquired. For wildtype proteasomes incubated with anti-Flag antibody (Ab) and Uch2-GFP images were acquired at a constant defocus suitable for particle selection, classification and intermediate resolution reconstruction (defocus: ~ 2.5±0.5 µm).

3.6. Image processing

3.6.1. Image formation in the electron microscope

In a TEM, electrons emitted from a point source (e.g. field emission gun) are accelerated by a high potential (typically 120 – 300 keV) and form a coherent electron beam that interacts with the specimen. The structure of biological molecules can be altered by the high energy of the electron beam, which limits the applied electron dose for image acquisition to 10-20 $e^-/Å^2$ (Henderson, 1995). As a consequence, single 2D projections have a very low signal-to-noise ratio (SNR), which can be overcome by averaging many projections of the same orientation. The majority of the image contrast is caused by phase contrast, caused by the interference of single elastic scattering of electrons with the primary beam (Dubochet et al., 1988; Toyoshima et al., 1988; Costello, 2006). In linear approximation of the imaging process, the resulting wavefunction $\Psi_{scatt}(r)$ is convoluted by a point spread function (PSF) c, describing the imaging characteristics of the optical system:

$$\psi(\mathbf{r}) = \Psi_{scatt}(\mathbf{r}) \otimes c(\mathbf{r})$$

In a TEM, the PSF is the Fourier transform of the Contrast Transfer Function (CTF). The CTF describes the convolution of an object in frequency space q:

$$\psi(\mathbf{r}) = \mathcal{F}^{-1}\{\mathcal{F}\{\Psi_{scatt}\}(\mathbf{q}) \cdot CTF(\mathbf{q})\}$$

The spherical aberration C_s and defocusing of the objective lens Δz introduce a phase shift of the traversing electrons of wavelength λ:

$$CTF(\mathbf{q}) = sin\,[\frac{\pi}{2} C_s \lambda^3 \mathbf{q}^4 - 2\Delta z \lambda \mathbf{q}^2]$$

The CTF causes oscillation of the contrast between 1 and -1 in frequency space, with no contrast transfer at the zeros of the CTF (Fig. 8).

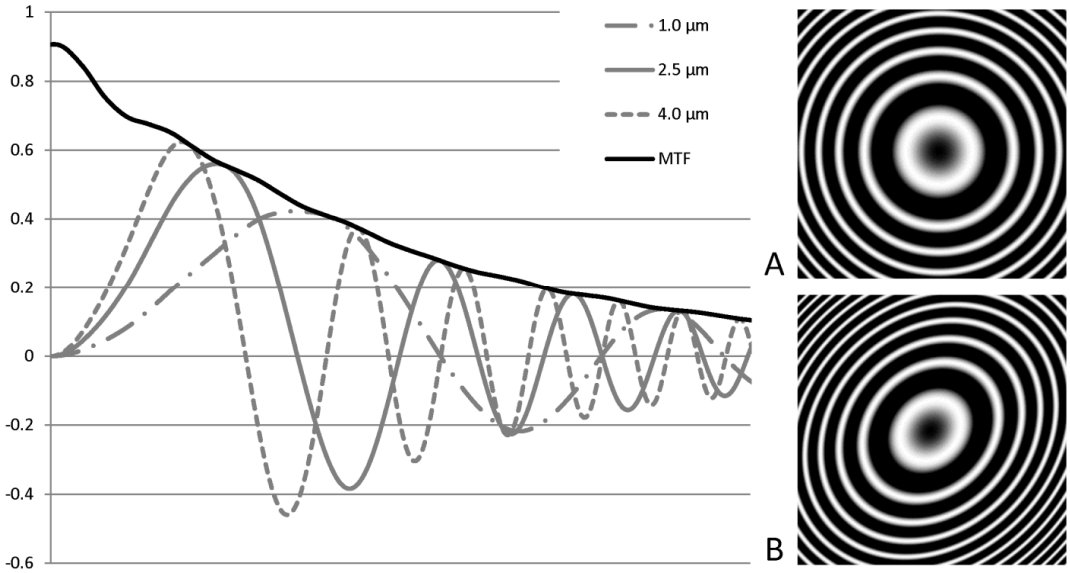

Figure 8: Representations of the contrast transfer function and the effects of the modulation transfer function (MTF) and astigmatism. The graph shows models of the CTF for different defocus values (acceleration voltage of 200 kV, spherical aberration C_s of 2 mm and objectpixelsize of 4.42 Å). Dampening due to the envelope function (e.g. measured MTF of the Eagle CCD camera used in this study) significantly decreases the signal at higher frequencies (x-Axis). **A and B:** Representation of the contrast transfer function in Fourier space. Thon rings (black rings in a CTF in Fourier space, left) relate to frequencies with very low or no contrast transfer. Errors in alignment of the microscope can lead to deformations of the concentric rings to ellipsoids (astigmatism, **B**). The size of the astigmatism is given by the difference of the major and the minor axis of the CTF's ellipsoid. For concentric CTFs, this value is zero (stigmatic, **A**).

Notably, the contrast at certain frequencies is flipped to negative values. The finite spatial and temporal coherence of the electron source and the chromatic aberration C_s damp the amplitude of the CTF in the higher frequency range, described by the envelope function (Zhu

et al., 1997). High frequencies are further dampened to substantial extent by the recording device (e.g., photographic film or charge coupled device (CCD) camera), which can be described by the modulation transfer function (MTF). Given an approximation of CTF, envelope function and MTF a TEM image can be corrected to recover the 'true' information of the object.

3.6.2. Deconvolution of electron micrographs

The contrast transfer function (CTF) was determined and micrographs were deconvoluted by phase-flipping and compensation for the modulation transfer function of the CCD camera (Nickell et al., 2005).

3.6.3. Reconstruction and ML3D Classification

To sort out assembly intermediates, i.e. to classify heterogeneous projection data, several strategies can be applied. Supervised classification (Gao et al., 2004; Heymann et al., 2004) separates the data according to their similarity to known reference structures, which requires *a priori* structural knowledge about the sample heterogeneity. Similarly, normal-mode analysis from an existing density map allows prediction of the conformational variability of a molecule and was used to generate starting models for classification (Brink et al., 2004). Other methods iteratively refine two-dimensional classification and three-dimensional reconstruction, by alternating the orientational or class membership assignment of the particles (Frank, 2006).

To avoid bias of the classification process by *a priori* generated reference models, a maximum-likelihood approach was chosen (Scheres et al., 2007). Here, the dataset is split randomly to generate the reference models. The discrete assignment of projection orientation and class membership is replaced by probability-weighted functions over all possible assignments. Thus, during refinement, each particle projection belongs to different classes and different models at the same time, weighted according to its respective cross-correlation coefficients. The algorithm converges when the change of projection assignment falls below a previously defined threshold value (i.e. the models remain unchanged from one iteration to the next).

The initial 3D model for refinement was the *D. melanogaster* 26S proteasome density filtered to a resolution of 4 nm (Nickell et al., 2009). Data were sorted using ML3D classification (Scheres et al., 2007) and densities were further refined as described using Xmipp (Scheres et al., 2008). In this study, a first classification step was used to sort out assembly intermediates. The pre-aligned wildtype particles (angular increment of 5°) were split into two groups (three groups for Uch2-GFP and ΔUch2), which were refined iteratively until the number of particles changing angular class or group were below 1 %. For the wildtype dataset, a second ML3D classification step was used to further separate the (dis)assembly intermediates (five groups) and the particles resembling 26S holocomplexes (two groups).

For the wildtype reconstructions reaching to subnanometer resolution, the temperature factor was determined (Rosenthal et al., 2003) and applied to the final reconstruction (~ -250 $Å^2$). The temperature factor before MTF correcting was ~ -1,000 $Å^2$. For Uch2-GFP, ΔUch2 and wildtype-antibody reconstructions, deconvoluted micrographs were binned once to speed up image processing, resulting in medium resolution reconstructions (~ 25 Å).

3.6.4. 3D-Variance map

The variance map was calculated according to (Penczek et al., 2006b). In this study, 120,000 low signal-to-noise models were generated from 300 randomly selected particles each.

3.6.5. Modeling

Comparative models of the *S.pombe* 20S proteasome and the AAA-ATPases were built as described previously (Förster et al., 2009). The positions of CP and AAA-ATPase hexamer were locally refined in the experimental map using UCSF Chimera (Pettersen et al., 2004).

3.6.6. Focused classification

For focused classification, angularly refined particles were C2 symmetrised, masked and split into 30 iteratively optimized groups (4.3.6 and Fig. 9). The position and radius of the mask was adapted by evaluating the variance map.

To sort the single particles according to their conformational states the particles were grouped into *N* classes according to their structural differences within a spherical mask 'm' positioned at an area of interest in the 3D map (Fig. 9). For classification of wildtype *S.pombe* particles, the mask was positioned at the 'mouth' region of the RP (Fig. 28). The algorithm chosen is similar to the focused classification suggested by (Penczek et al., 2006a), but it differs in the assignment of the conformational classes (by F. Beck, in preparation).

In the first step a 3D structure from the entire dataset was determined by projection matching using Xmipp (Scheres et al., 2008). For subsequent classification of particles according to conformations each particle keeps this angular assignment. Both RPs capping the CP were simultaneously classified by including each particle twice into the analyzed dataset: one copy with the original projection angles and translation vectors and one copy with the C2 transformed rotation angles.

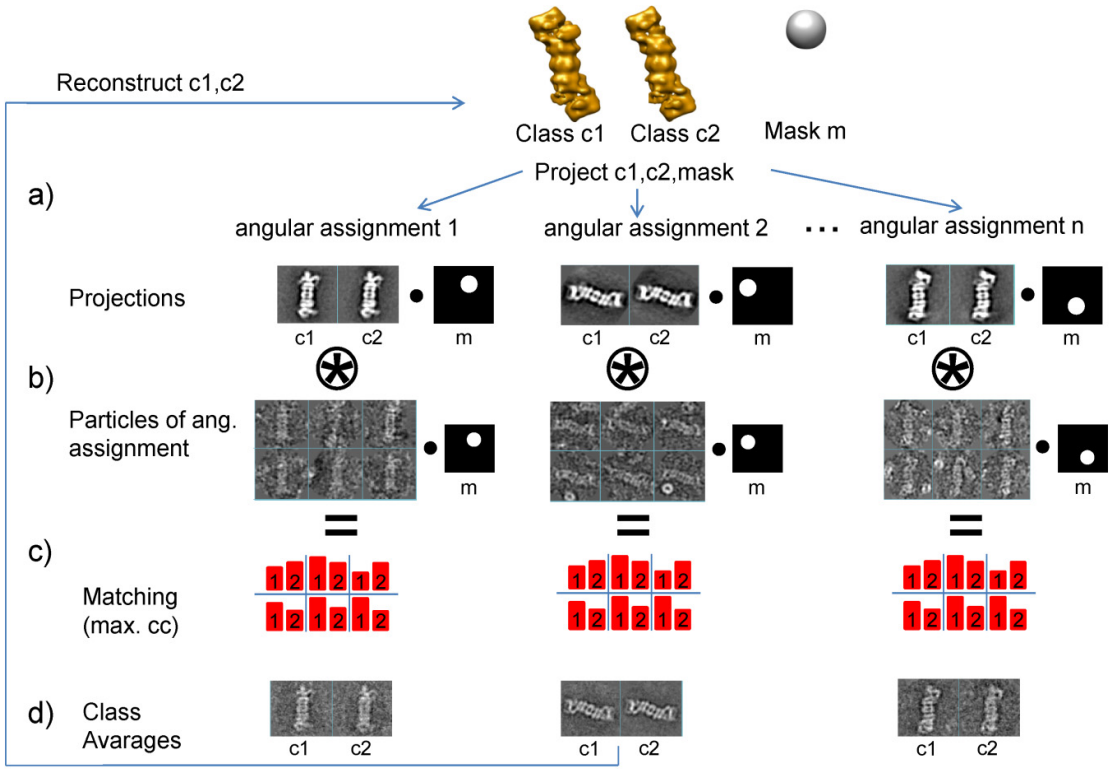

Figure 9: Principle of focused classification. Given a known angular assignment, particles are divided randomly among a chosen number (here: two) of identical starting models. All particles are masked and matched with masked projections of each model according to their highest cross correlation value. From the resulting class averages, new models are reconstructed. Iteratively this process is repeated, until the number of particles switching class drops below a certain threshold (e.g. < 10 %). Figure adapted from (Bohn et al., 2010).

Aim of the classification process was to divide the dataset into *N* different classes, such that their correlation to the corresponding 3D model is maximized. For optimization of the particle assignments the following expectation maximization approach was chosen: To initialize the actual classification particles were randomly split into *N* equally populated subsets, the corresponding *N* 3D models were reconstructed ('seeds'), and projections of these *N* models were calculated (Fig. 9a). The projections as well as all individual particles were multiplied with a projection of the 3D mask (alignment parameters from projection matching) (Fig. 9b). Within every angular class, each particle was correlated to the corresponding *N* model projections (Fig. 9c). Each particle was assigned to the model with the highest correlation value. Again, *N* models were calculated using this new particle assignment (Fig. 9d). The process was iterated until the number of particles changing their class was below 1%.

4. Results and Discussion

At first, the purification of *S.pombe* 26S proteasomes will be discussed in terms of their proteolytic activity, abundance of 26S holocomplexes and suitability for structural analysis by cryoEM and single particle analysis (sections 6.1 and 6.2). The structure of the 26S proteasome and the structural heterogeneity among 26S proteasomes will be analyzed in detail (section 6.3).

To obtain further insights into the structural organization of the proteasomal RP, several labeling techniques will be evaluated: (*i*) chemical crosslinking in conjunction with mass spectrometry (section 6.4); (*ii*) binding of gold-labeled ubiquitin-aldehyde, targeting the ubiquitin C-terminal hydrolase Uch37 in *D.melanogaster* (section 6.5); (*iii*) fusion of GFP to the C-terminus of Uch2 in *S.pombe*, creating an extra density (section 6.7); and (*iv*) creation of an *S.pombe* knockout strain lacking Uch2, creating a 'missing density' (section 6.7); (*v*) binding of the monoclonal anti-Flag antibody, targeting the Flag epitope of Rpn11 used for proteasome purification in yeast (section 6.8).

4.1. Purification of 26S proteasomes

The quaternary structure of the 26S proteasome is very sensitive to subtle changes in pH, salt or ATP concentrations. Several proteasome-interacting proteins are involved in 26S proteasome assembly (Saeki et al., 2009) and so far it has not been possible to reconstitute the 26S holocomplex *in vitro*. Successful 26S preparations rely on gentle, e.g., structure preserving, purification methods. In the late 1990's, a protocol for the purification of 26S holocomplexes from *D.melanogaster* eggs was established (Hölzl et al., 2000). For efficient, high resolution structural analysis by cryoEM the purity, concentration and stability of the complex are decisive, but not as demanding as for NMR or X-ray crystallography.

Whereas 26S holocomplex purifications from *D.melanogaster* fulfill all the criteria mentioned above, cell growth and consequently their genetic manipulation are lengthy procedures. However, most tags suitable for purification of macromolecular assemblies as well as labels suitable for analysis by cryoEM rely on genetic manipulations (e.g. His-tag, Flag-tag for purification or addition/deletion of protein densities for difference mapping).

Mutant construction and cell growth is comparably more efficient and well established in a variety of single-celled eukaryotes, such as yeast (Knop et al., 1999; Janke et al., 2004).

By introducing an affinity tag to the terminus of a constituent subunit of a macromolecular complex, the purification procedure can be significantly simplified and shortened compared to purifications from wildtype cells. Purification of 26S holocomplexes from *S.cerevisiae* carrying a Flag-tag at the C-terminus of Rpn11 has been established recently (Verma et al., 2000; Saeki et al., 2005). The tagging of Rpn11 does not seem to affect the structure or the function of the 26S proteasome (Sone et al., 2004). An aim of this study was to develop labeling techniques suitable to localize all known proteasome-associated subunits including the DUB Uch2/Uch37. As there exists no homolog of Uch2/Uch37 in *S.cerevisiae*, an Rpn11-C-term.-Flag-tag strain was generated in the closely related model organism *S.pombe*.

The purification protocol for *S.cerevisiae* 26S proteasomes was adapted and changed for 26S holocomplex purification in *S.pombe* (see 5.4.2). Sucrose gradient fractions of the last step of 26S proteasome preparations have a total protein concentration of ≤ 0.2 mg/ml. To assess the degree of purity and relative abundance of 26S holocomplexes (e.g., 20S CP with two bound 19S RP), samples were negatively stained and examined in a TEM (Fig. 10). Micrographs typically show primarily complete double-capped 26S proteasome particles; but some single-capped particles, as well as isolated CPs were also observed (Fig. 10).

Each fraction was tested for its ability to hydrolyze the substrate Suc-LLVY-AMC. While proteolytic activity was detected in a wide range of fractions, the highest activity was measured in the fraction with the highest concentration of 26S holocomplexes (Fig. 10) and was used for further analysis.

Purifications from Uch2-GFP and ΔUch2 *S.pombe* strains show a similar 26S holocomplex abundance in comparison to *S.pombe* Rpn11-Flag and *D.melanogaster* wildtype purifications (Fig. 10). The negative stain background for Flag-tag purified 26S proteasomes indicates a higher amount of impurities than in classical 26S preparations from *D.melanogaster*, which is mainly explained by the shorter protocol involving less purification steps.

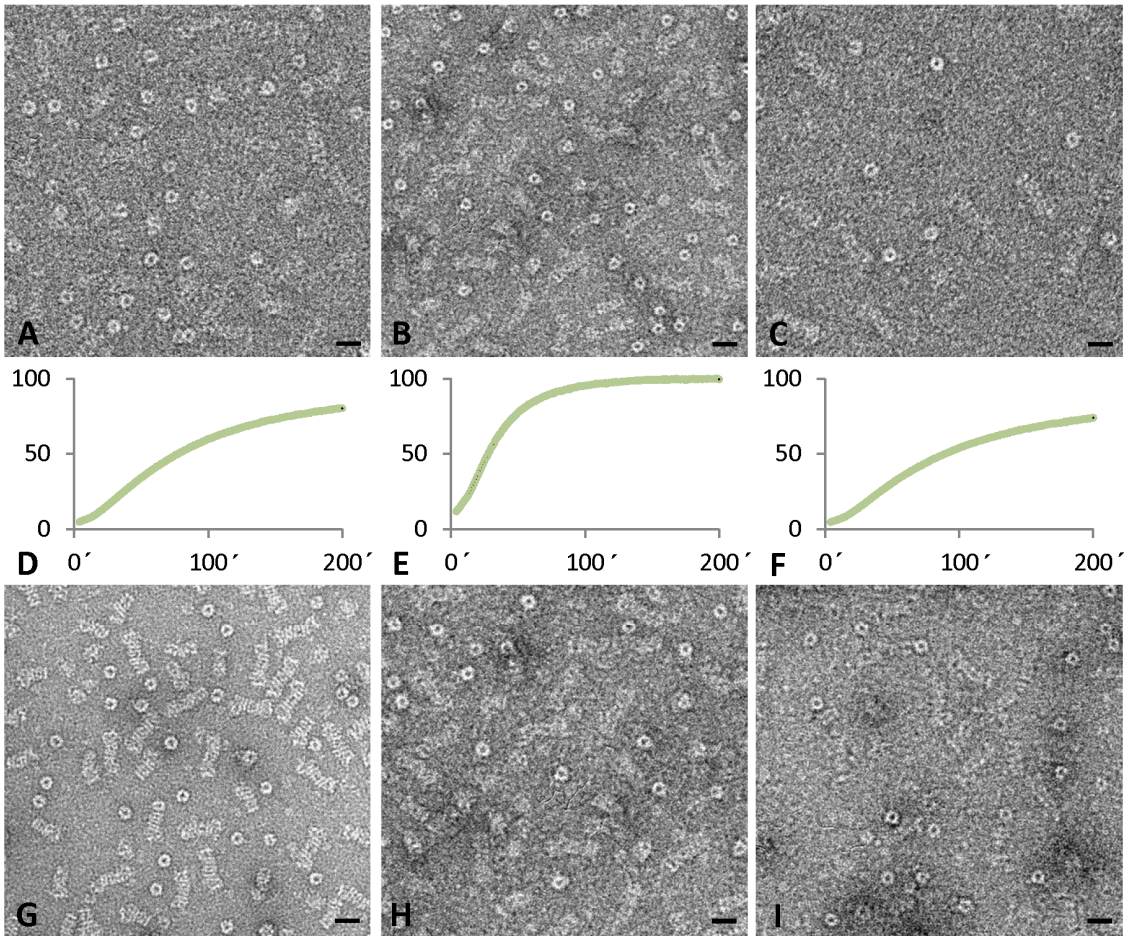

Figure 10: Stability and activity of purified 26S proteasomes from *S.pombe* and *D.melanogaster*. Fractions from sucrose gradients were negatively stained and imaged by EM (final magnification 54,000x at 2-3 µm defocus) to assess 26S holocomplex concentration. The highest concentration of holocomplexes is present in the upper half of the gradient (fractions 10-12 of 25 in total, **A-C**), whereas 'disassembly intermediates' (CP without or with one bound RP) are observed in all preparations. Proteolytic activity tests with Suc-LLVY correlate with the presence of 26S holocomplexes, i.e. the higher the proteolytic activity of the gradient fraction, the higher its holocomplex concentration (**D-F** show normalized (%) activity curves of the fractions 10-12 over a range of 200 min). Preparations from *D.melanogaster* (**G**, fraction 19) show less staining background, indicative of fewer impurities. *S.pombe* 26S holocomplex abundance in ΔUch2 (**H)** and Uch2-GFP strains (**I)** is comparable to wildtype levels (A-C). Scale bar is 10 nm.

In summary, the adapted Flag-tag purification protocol for *S.pombe* 26S proteasomes yields proteolytically active 26S holocomplexes in high concentrations suitable for further analysis by MS, labeling, cross-linking and cryoEM.

4.2. MS analysis of purified 26S proteasomes

The subunit composition of purified *S.pombe* 26S proteasomes was analyzed by quantitative MS and compared to previously published data (Nickell et al., 2009) from *D.melanogaster*

(Fig. 11). The canonical 26S proteasome subunits α1-7, β1-7, Rpt1-6, Rpn1-3 and Rpn5-12 are all present in approximately equimolar amounts. Rpn13 is present in two different orthologs, Rpn13a (Uniprot ID Q9Y7Y6) and Rpn13b (Q9USM1), each in a ~0.2:1 ratio compared to the constitutive subunits. Expression of organ-specific Rpn13 orthologs was previously reported for *D.melanogaster* testis (Belote et al., 2009), but simultaneous expression of different orthologs has not been described, yet. The abundance of the 9kDa peptide Rpn15 could only be determined with low accuracy by this MS approach since only one unique peptide was identified. The ubiquitin C-terminal hydrolases Uch2 and Ubp6, as well as Ub were detected in significant amounts.

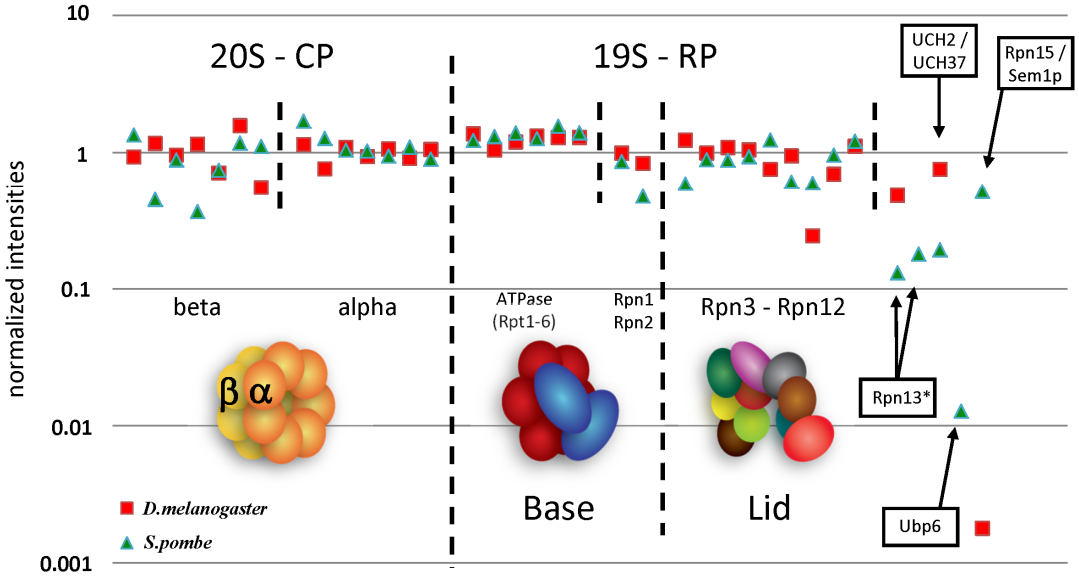

Figure 11: Subunit composition of 26S proteasomes from *S.pombe* compared to *D.melanogaster*. Sucrose gradient fraction showing highest activity in Suc-LLVY-AMC cleavage was trypsin digested and analyzed by MS. Peptide intensities were quantified four times independently, normalized (Nickell et al., 2009), and compared to data from *D.melanogaster* (Nickell et al., 2009). In *S.pombe* the canonical 26S proteasome subunits α1-7, β1-7, Rpt1-6, Rpn1-3, and Rpn5-12 are approximately present in an equimolar ratio. Rpn13 and the PIPs Uch2 and Ubp6 were found in sub-stoichiometric amounts. Rpn15/Sem1p was detected by a single peptide, while for all other subunits 23±11 peptides have been detected; thus, the determined Rpn15 stoichiometry is inaccurate. Notably, only the amounts of Rpn10, Rpn13 and Uch2/Uch37 differ significantly between *S.pombe* and *D.melanogaster* 26S proteasomes. Two orthologs of Rpn13 have been detected in *S.pombe* (* left arrow Rpn13b, right arrow Rpn13a). Figure adapted from (Bohn et al., 2010).

When comparing the relative stoichiometry of the *S.pombe* 26S proteasome to that determined for *Drosophila* (Nickell et al., 2009) the following differences were found: Rpn10 is substantially more abundant in *S.pombe* (~0.8:1) than in *Drosophila* (~0.25:1). Moreover, Rpn13a/Rpn13b and Uch2 are present in substoichiometric amounts in *S.pombe* whereas Rpn13 and Uch37 were found in approximately stoichiometric amounts in *Drosophila*

proteasomes. Interestingly, in 26S proteasomes purified from *Homo sapiens*, none of these subunits (Rpn10, Rpn13, Uch2/Uch37) are present in significant amounts (Isasa et al., 2010). Thus, quantitative analysis indicates that α1-7, β1-7, Rpt1-6, Rpn1-3, Rpn5-9 and Rpn11-12 constitute the canonical core of the 26S proteasome whereas the subunits Rpn10 and Rpn13, albeit always present, are found in variable amounts.

4.3. Structure of the *S.pombe* 26S proteasome

4.3.1. Data acquisition and refinement

Micrographs of vitrified samples of the wildtype *S.pombe* 26S preparation typically showed complete double-capped 26S proteasome particles; but some single capped particles (Fig. 12), as well as isolated CPs were also observed.

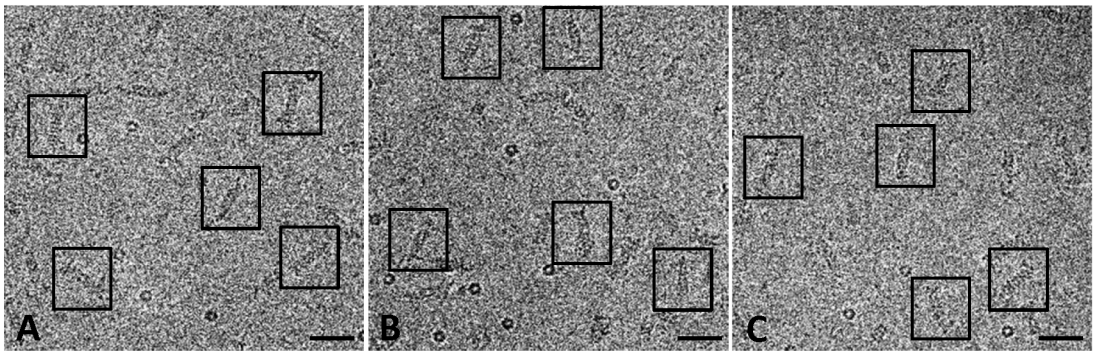

Figure 12: Vitrified 26S proteasomes from *S.pombe*. Samples were incubated on glow-discharged lacey carbon grids, washed and plunge-frozen in liquid ethane at liquid nitrogen temperatures. Micrographs (final magnification of 68,000x at 1-3 µm defocus; **A** – wildtype, **B** – ΔUch2, **C** – Uch2-GFP) show mostly singular 26S holocomplexes (**boxed particles:** 'side views') and almost no clustered protein aggregates, with ~ 5 - 60 particles per image. Scale bar is 40 nm.

For three-dimensional reconstructions from two-dimensional projections, a homogeneous angular distribution of vitrified particles is preferred. However, electrostatic interactions between the molecule's surface and the buffer solution or air-buffer interface just before vitrification mainly determine the particles' orientation in ice. Also, the ratio of ice thickness (50-300 nm) to proteasome length (~ 40 nm) influences the particles' orientation during vitrification: top views, i.e. 26S proteasomes projected along their longitudinal axis, are observed less frequently (Fig. 13). For such heterogeneous datasets, i.e. datasets with an uneven distribution of projection probability, more micrographs must be acquired to fill the missing gaps. The property of the 26S proteasome to adopt a preferred orientation in vitreous

ice becomes exceedingly problematic: the fine angular sampling necessary for high-resolution reconstruction requires tens of thousands of angular class averages with a high SNR, i.e. many particles of the same orientation.

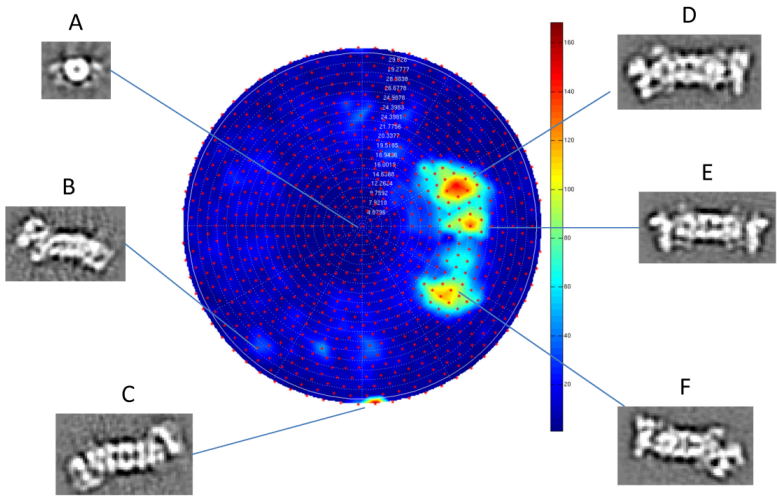

Figure 13: Angular distribution of vitrified *S.pombe* 26S proteasomes. Shown is the angular assignment of a randomly selected subset of 10,000 particles from the wildtype *S.pombe* dataset. Angular class averages with many particles (red, **C-F**) will have a significantly improved SNR, than class averages with very few particles (blue, **A, B**). Top views, i.e. 26S proteasomes projected along their longitudinal axis (**A**), are observed less frequently than most 'side views' (**C-F**).

In the course of this study, micrographs were acquired in a semi-automated fashion: user interactions were only required to choose areas for the acquisition of sets of micrographs, whereas focusing, acquisition and data storage relied on automated procedures implemented in SerialEM (Mastronarde, 2005). Due to drift, contamination, astigmatism or lack of particles only ~ 15 % of acquired images had to be discarded (Fig. 14). The semi-automated acquisition scheme proved to be reliable to generate a high quality dataset (e.g. low astigmatism, Fig. 14) of micrographs at the desired defocus range (Fig. 15). For high-resolution reconstructions, requiring hundreds of thousands of particles, an intermediate number of particles per image is desired to avoid (*i*) excessive acquisition and image processing time and (*ii*) particle crowding or overlapping. The adapted 26S proteasome purification and vitrification protocols in conjunction with the semi-automated acquisition scheme yielded on average ~ 30 particles per image (Fig. 14). From ~ 9,700 micrographs 270,000 particles resembling 26S holocomplexes were selected for single particle analysis.

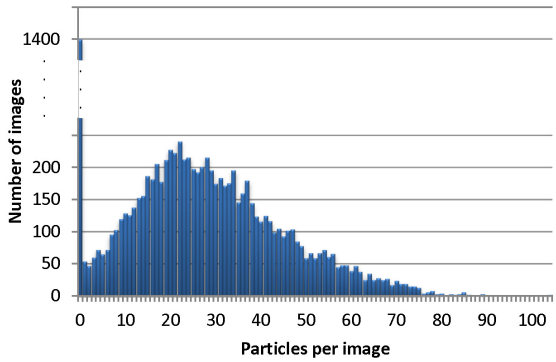

Figure 14: Distribution of vitrified wildtype *S.pombe* 26S proteasomes. Approximately 15 % (1,400 out of 9,700) of semi-automatically acquired micrographs had to be discarded due to drift, astigmatism, contamination or lack of particles. Micrographs acquired showed up to 104 particles resembling 26S holocomplexes, with on average 26 particles per image.

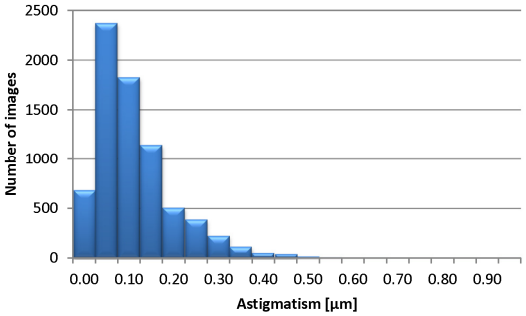

Figure 15: Astigmatism of semi-automatically acquired cryoEM images. Astigmatism occurs in all electron lenses, which can be corrected by proper alignment of condenser, objective and diffraction stigmator. Stage drift or an unstable beam can have an effect undistinguishable from astigmatism on the micrographs CTF. Less than four percent of acquired micrographs showed astigmatic values > 0.3 µm.

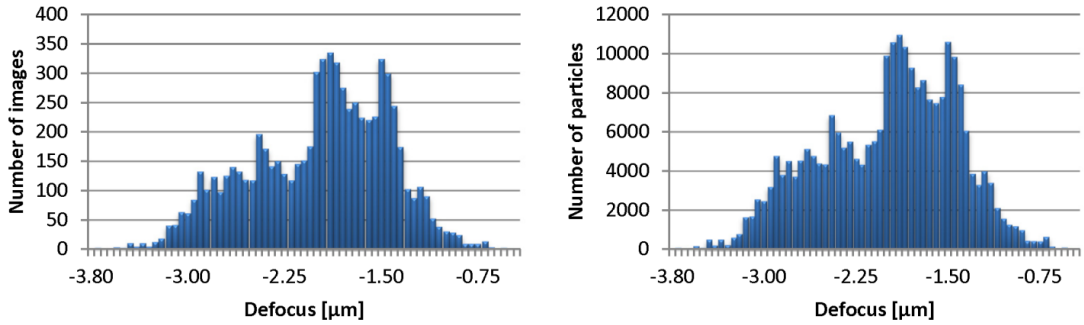

Figure 16: Defocus distribution of semi-automatically acquired cryoEM images. Around 9,700 unique micrographs of vitrified wildtype *S.pombe* 26S proteasomes were acquired in 18 sessions (Table 3). The nominal underfocus of 1-3 µm was adjusted with the autofocus routine implemented in the SerialEM software package (Mastronarde, 2005). Variations in ice thickness, daily microscope alignment and grid topology can lead to alterations of the actual defocus. Around 270,000 particles (right) were selected from micrographs (left), resulting in ~ 4,000 particles per defocus group (0.05 µm intervals) in the nominal defocus range of 1-3 µm.

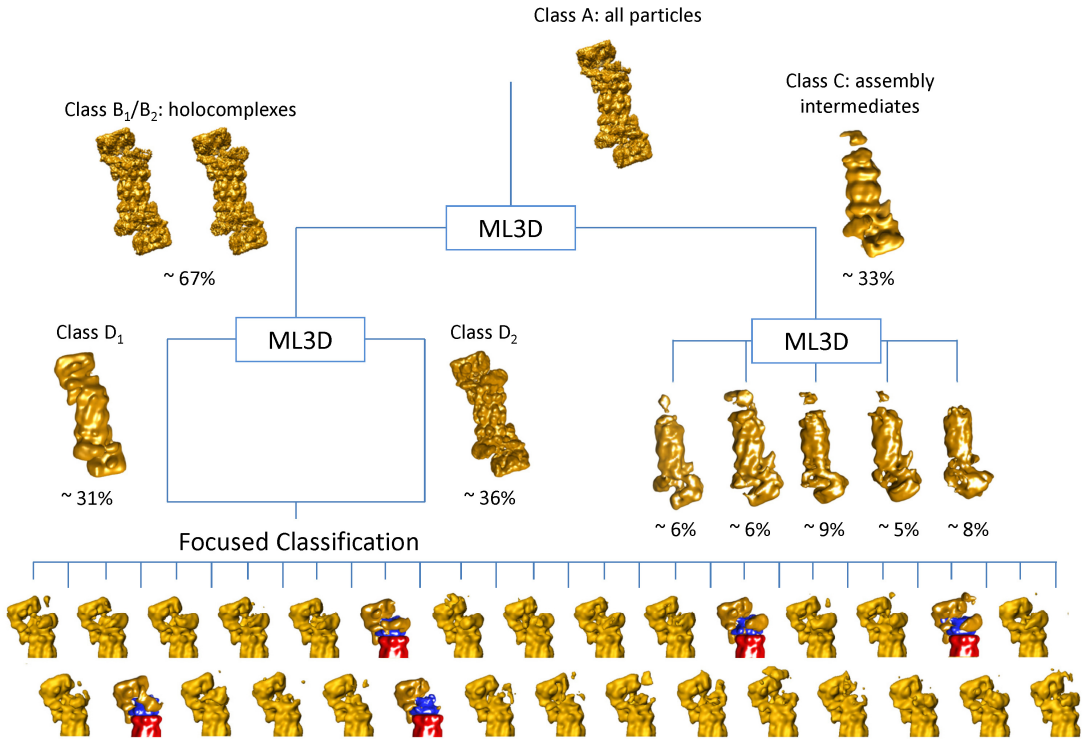

Figure 17: Classification of 26S proteasomes from *S.pombe*. From the initial ~9,700 micrographs a total number of ~270,000 particles have been selected. The initial reconstruction yielded a density map at 9.1 Å (class A). To sort out (dis)assembly-intermediates, ML3D-classification was performed. Particles were classified into two groups (class B and C), of which one showed all the features of 26S holocomplexes, while the other clearly corresponded to partially (dis)assembled 26S proteasomes. The reconstruction of fully assembled proteasomes (class B_2) yielded a map of 9.1 Å resolution (Fig. 18). In a second ML3D classification step, holocomplexes separated according to their difference of the distinct 'extra mass' in the cap region (class D_1 with one and D_2 with two 'extra masses') and partially (dis)assembled 26S proteasomes according to the remaining RP densities (Fig. 27). Using classification focused on the area with the highest variance (Fig. 9), particles belonging to class B were further classified into 30 classes (Fig. 28, colored classes are enlarged in Fig. 27). Figure adapted from (Bohn et al., 2010).

Manual particle selection is strongly biased, mainly because of the low SNR in micrographs of biological specimens acquired by cryoEM. Whereas projections of vitrified proteasomes usually resemble a cylinder-like shape ('side views'), it is not possible to distinguish actual 26S holocomplexes from (dis)assembly intermediates. However, a substantial structural variability among individual particles has to be anticipated due to the fragility and dynamics of the 26S proteasome. To 'purify' the particle ensemble *in silico* the maximum likelihood-based classification method ML3D was used to classify the particles into two different groups (Fig. 17). One class showed all the features of 26S holocomplexes, while the other clearly corresponded to partially (dis)assembled 26S proteasomes (~85,000 particles). A second ML3D classification step was applied for both classes. It was possible to separate the

holocomplexes further into 26S proteasomes which differed by a distinct mass in the cap region ('class D_1' with one and 'class D_2' with two 'extra masses', containing ~85,000 and ~100,000 particles, Fig. 16). Partially (dis)assembled 26S proteasomes were separated according to the remaining RP densities (see section 4.3.7).

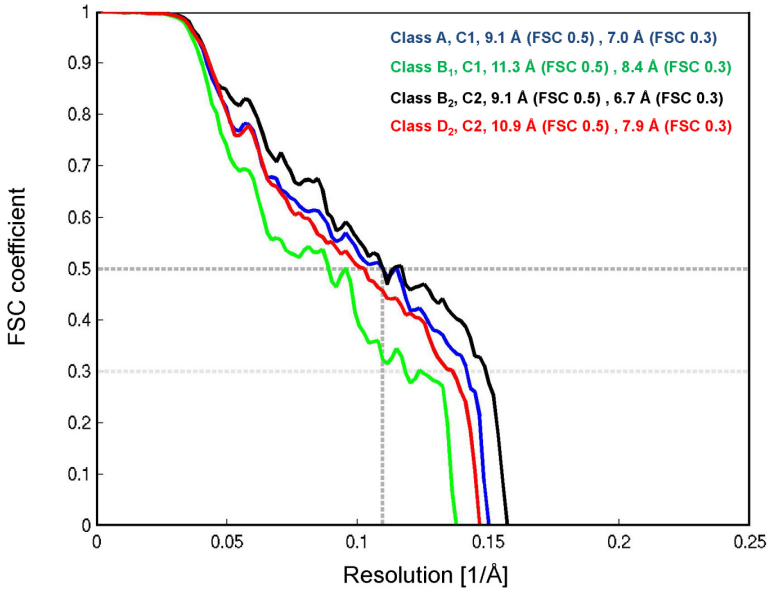

Figure 18: Fourier Shell Correlation (FSC) curves of reconstructed wildtype *S.pombe* models. Classes A, B and D_2 were refined with a final angular increment of 0.2° (data split into 39 CTF-groups), without imposed symmetry (A, B_1) or with imposed C2 symmetry (B_2, D_2). Imposed C2 symmetry improved the resolution to 9.1 Å / 6.7 Å at FSC of 0.5 / 0.3 respectively. When not explicitly noted, the FSC criterion of 0.5 is used. Figure adapted from (Bohn et al.).

The 3D densities corresponding to intact holocomplexes (class B_1/B_2, Fig. 18) displayed C2 symmetry (apart from the 'extra mass'). Therefore, to obtain higher resolution, C2 symmetry was imposed for the reconstruction of the 26S proteasome density from all holocomplexes (resolution of 9.1 Å / 6.7 Å at FSC of 0.5 / 0.3 respectively; Fig. 18). In the final symmetrical reconstruction several helical motifs became discernable (Fig. 19, 21).

4.3.2. Positioning of CP and AAA-ATPase

To further assess the quality of the high-resolution reconstruction (model B_2) a comparative model of the CP was fitted into the EM map (Fig. 19). The prominent N-terminal α-Helix, a hallmark of the α4 subunit, allows unambiguous positioning of the CP. The fit of the atomic model and the EM map is excellent, as indicated by the high cross-correlation coefficient

(CCC=0.78, both models filtered to 9 Å) and the visual correspondence of helices in model and map.

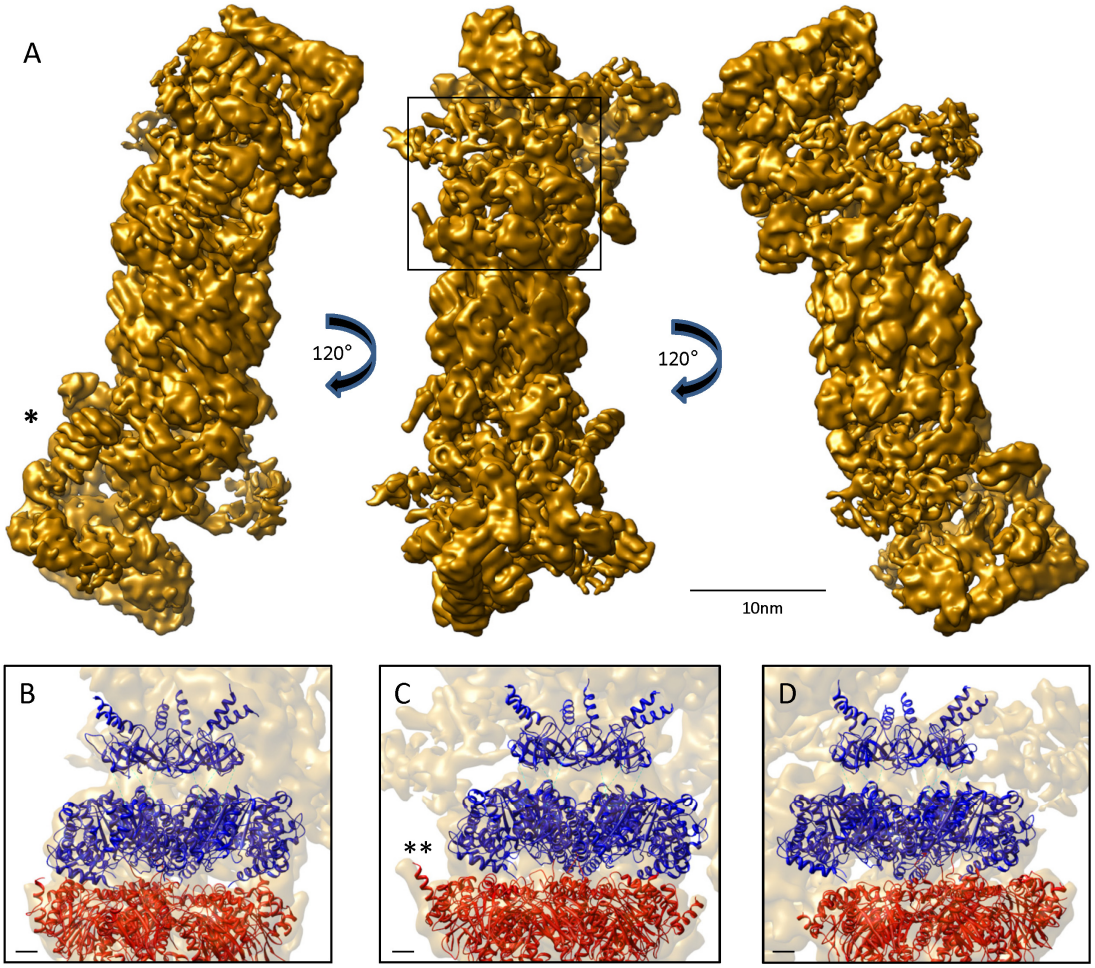

Figure 19: Structure of the 26S proteasome from *S.pombe*. Isosurface representation of the 26S proteasome reconstruction rotated around the pseudo-7-fold axis of the CP by 120º. The isosurface threshold was set to include a protein mass of 755 kDa for the CP. Secondary structure elements like bi-helical repeats can be clearly discerned (**A**, *). **B – D** Cut-out of the boxed area in A shows a transparent view through the reconstruction (rotations as in **A**) with the fitted atomic models of the CP (red, **: α-helix of subunit α4) and the AAA-ATPase (blue). Scale bar B – D is 1 nm.

Adjacent to the CP, a comparative model of the AAA-ATPases Rpt1-6 was positioned into the EM map as described previously (Förster et al., 2009) (Fig. 19). The C-terminal AAA-ATPase domains form the AAA-ring that binds to the CP, while the N-terminal segments form a smaller ring positioned atop of the AAA-ring (Djuranovic et al., 2009; Zhang et al., 2009a). In the EM map, density corresponding to the N-terminal coiled-coils is clearly discernable and correlates well to the atomic model. The atomic model of the

AAA-ring is in good agreement with the EM density, albeit some helical densities do not co-localize precisely with the helices in the atomic model indicative of dynamic conformation changes (see below). The center of the hexameric AAA-ATPase is shifted by ~20 Å from the pseudo 7-fold rotational symmetry axis of the CP, similar as observed for the *D.melanogaster* 26S proteasome (30 Å). The hexamer is less inclined with respect to the CP axis: whereas the pseudo 3-fold symmetry axis of the AAA-ATPase hexamer was tilted by ~10º with respect to the CP axis in the *D.melanogaster* density, it is tilted by ~4º in the map from *S.pombe*.

4.3.3. The CP gate in 26S holocomplexes

When comparing the EM density at the gate of the CP to the atomic model two major differences are visible: (*i*) extra densities appear in the pockets between α4 – α3 and α2 – α1 and (*ii*) the density in the central pore belonging to α3 in the atomic model is rotated by approximately 50° relative to a similar density observed in the central pore of the EM map (Fig. 20 B-D). The degree of rotation (~ one seventh of 360 °) indicates that the observed density in the central pore of the EM map belongs to a different alpha subunit.

The extra densities observed in the α-subunit pockets likely originate from the AAA-ATPases Rpt2 and Rpt3 (see section 4.5, Fig. 28, 29). The C-termini of Rpt2, Rpt3 and Rpt5 share a common HbYX motif (i.e. a hydrophobic and a tyrosine residue followed by a residue of any type), responsible for gate opening of the CP in PAN (Smith et al., 2007). However, of all RP AAA-ATPase subunits, only the C-termini of Rpt2 and Rpt5 are capable to induce CP gate opening alone, whereas point mutations in HbYX of Rpt2, Rpt3 and Rpt5 can inhibit gate opening (Smith et al., 2007). To compare the gate density to previously published conformational states (Rabl et al., 2008), D7 symmetry was applied to the CP EM density of the 26S proteasome reconstruction (Fig. 20 E). The density at the gate of the CP is in good agreement with the closed-state of the crystal structure (Fig. 20 F).

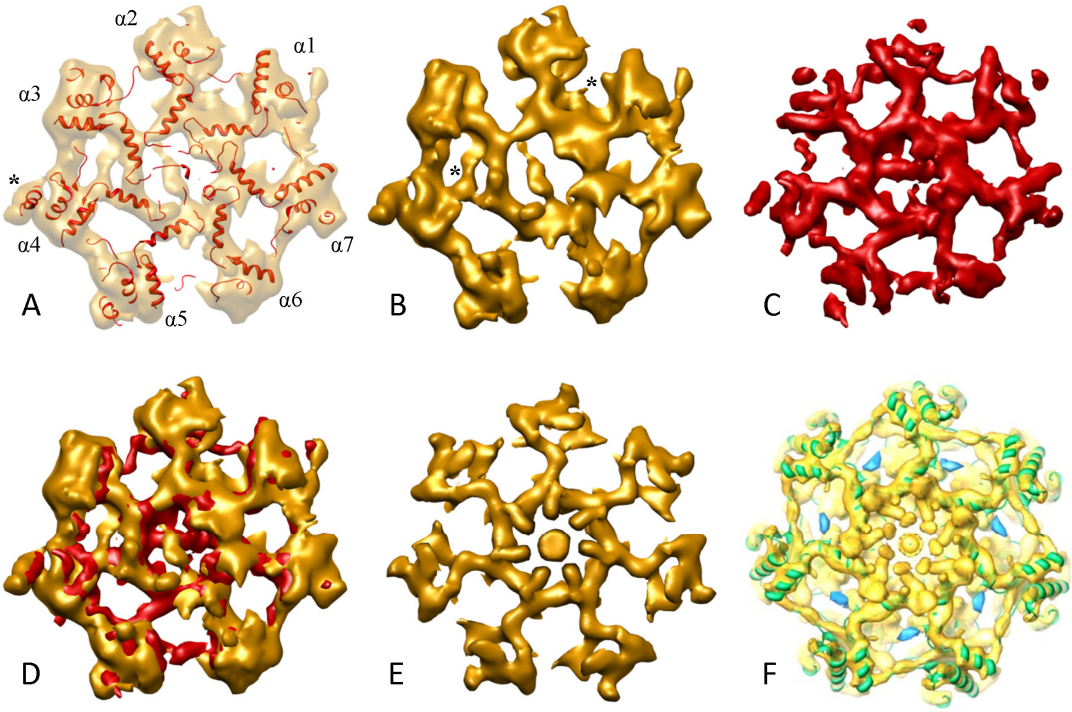

Figure 20: Gate of the 26S proteasome from *S.pombe*. A: The comparative atomic model of the CP (red, template: *Saccharomyces cerevisiae* 20S proteasome, PDB: 1ryp) and the EM density of the wildtype *S.pombe* 26S proteasome (transparent with a slab thickness of 10 Å) are in very good agreement. The prominent α-helix of subunit alpha4 (*) allows assignment of the remaining alpha subunits. **B:** EM density as in A. Notably, extra densities (*) are present in the pockets between α4 – α3 and α2 – α1. **C:** Simulated density map of the atomic model from A shows the central pore blocked by N-terminal residues of α3. **D:** Overlay of C and D shows the different orientations of the densities blocking the central pore. **E:** When applying D7 symmetry to the CP EM density of the 26S proteasome reconstruction, the resolution of identical rigid structural elements of the alpha subunits increases. **F:** The gate of the wildtype *S.pombe* 26S proteasome closely resembles the closed gate conformation of the CP proposed before (Rabl et al., 2008).

4.3.4. Secondary structure elements in the RP

In the remaining density of the RP, features reminiscent of short bi-helical repeats in supercoiled quaternary structures are recognizable (Fig. 19 A*). Indeed, such repetitive motifs, probably similar to tetratricopeptide repeats (TPRs), have been predicted before for Rpn1 and Rpn2, as well as the PCI-module containing subunits Rpn3, Rpn5, Rpn6, Rpn7, Rpn9, and Rpn12 (Kajava, 2002; Ciccarelli et al., 2003; Scheel et al., 2005; Enchev et al., 2010; Förster et al., 2010). The width of the helical ribbons is ~25 Å, which is consistent with the typical helix length observed in TPRs (Fig. 21).

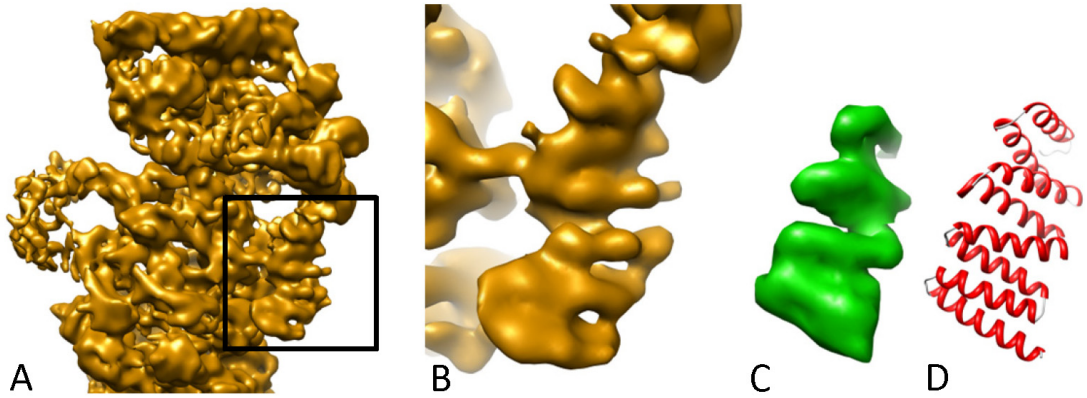

Figure 21: Bi-helical repeats in the RP density. In the 26S proteasome map (**A**) patches with apparent helical segments are discernible (rectangle; enlarged in **B**). The density of TPR repeats from Sec17 (PDB code: 1qqe) filtered to 9.7 Å resolution (**C**) resembles these patches. **D:** ribbon representation of 1qqe. Figure adapted from (Bohn et al., 2010).

4.3.5. 3D-Variance map of the 26S proteasome

The stoichiometries determined by quantitative MS analysis suggest conformational variations among individual particles beyond those being resolved by ML3D classification. To visualize the spatial distribution of the major variations in the 26S proteasome, a 3D variance map was calculated (Penczek et al., 2006b). As expected, the variance is relatively low in the density corresponding to the CP (Fig. 22). More surprisingly, the lid-region of the RP also shows a low variance level. The variance is highest in a belt surrounding the AAA-ATPase in the RP.

It is known from structural studies that AAA-ATPases typically undergo large conformational changes as part of their ATP cycle (Wang et al., 2001). However, the structural variation apparently extends to areas adjacent to the hexameric AAA-ATPase.

These structural changes can probably be mostly attributed to the sub-stoichiometrically bound proteins Rpn10, Rpn13a/b, Uch2 and Ubp6. Indeed, when comparing the reconstruction of 26S proteasomes from an Uch2 deletion strain (see below) to the wild-type map no significant differences could be observed; this is probably due to the fact that the region of Uch2 binding is highly variable; or because Uch2 is bound in substoichiometric amounts (Fig. 11) and hence invisible in the overall reconstruction.

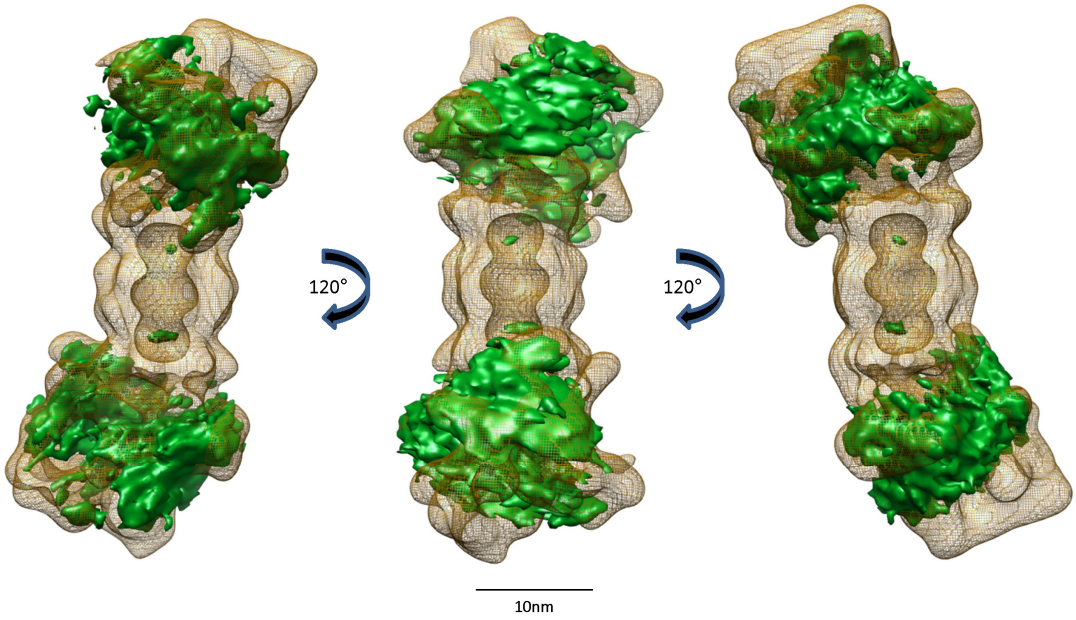

Figure 22: Variance map of the 26S proteasome from *S.pombe*. A mesh representation of the reconstruction with an overlay isosurface in green highlights the main variances among the particles. Notably, almost no variance is observed in the CP and an elongated stretch of the RP lid. Localized in a belt surrounding the AAA-ATPase hexamer is the majority of the variance, most likely caused by domain flexibility or the absence and presence of substrate or proteasome-interacting proteins.

4.3.6. Classification of the RP

To obtain further insights into the different 26S proteasome conformations the particles were grouped into different classes to localize structural variations specifically in the belt of high variance around the AAA-ATPases (Fig. 14). Classes derived from focused classification (3D) are consistent with results obtained from 2D analysis (Fig. 23). The analysis further resolves the 'extra mass': this mass adopts a variety of different shapes, e.g., it protrudes to the lid in one class (Fig. 24). This would be consistent with Rpn10 positioned in the extra mass because the ubiquitin interacting motif (UIM) is reported to be highly flexible

(Wang et al., 2005). Other classes show significant extra densities (of up to 30-40 kDa) in a region around the AAA-ATPases (e.g., Fig. 25).

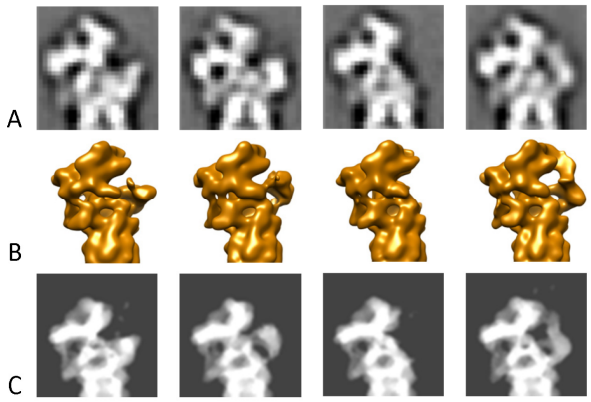

Figure 23: Comparison of classification methods. 2D class averages (**A**), derived from Principle Component Analysis (PCA) and K-means clustering, are qualitatively similar to projections of classes of the focused classification approach (**B**: 3D classes selected from Fig. 25; **C**: respective back-projection of models shown in B). Figure adapted from (Bohn et al., 2010).

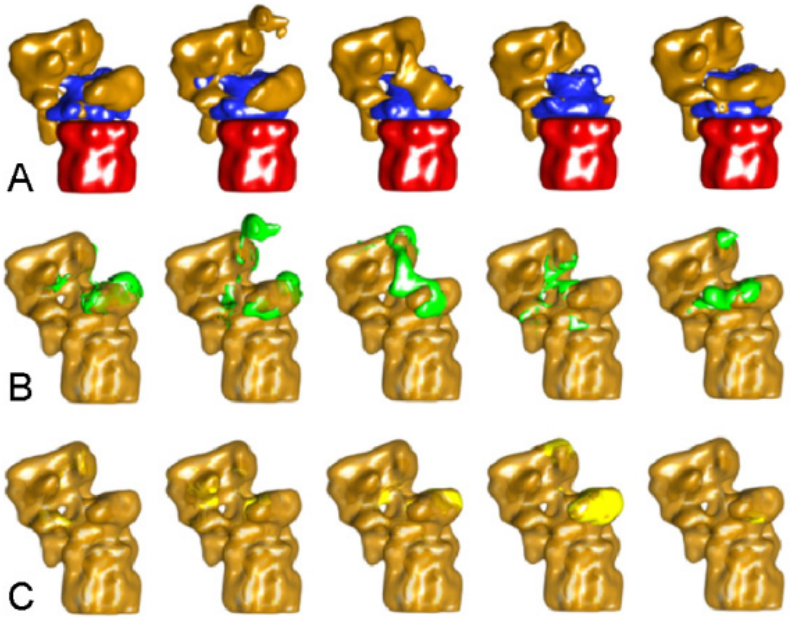

Figure 24: Classification of *S.pombe* 26S proteasomes reveals variations in the RP. Five class-averages obtained by focused classification (CP in red, AAA-ATPase-ring in blue, remaining RP in brown) are shown in **A**. Major differences of each class to the overall average correlate to the localization of the variance map shown in Fig. 22 (**B**: additional density in green, **C**: missing density in yellow). As expected from ML3D classification, a subset of 26S proteasomes is lacking the ‚extra-mass' (class 4). Further classes show additional densities in the region between base and lid, close to the upper AAA-ATPase-ring (classes 2, 3 and 5). Figure adapted from (Bohn et al., 2010).

Taken together, the variance map and the classification results suggest that the distal part of the lid provides a mostly invariable frame while the base and the base–lid interface are sites

of high variability. The structural variation ('activity') is caused by both structural flexibility of constitutive subunits as well as the reversible association of additional densities, tentatively assigned to PIPs. The most likely candidates for these extra-densities are the sub-stoichiometrically bound proteins Rpn10, Rpn13a/b, Ub, Uch2 and Ubp6.

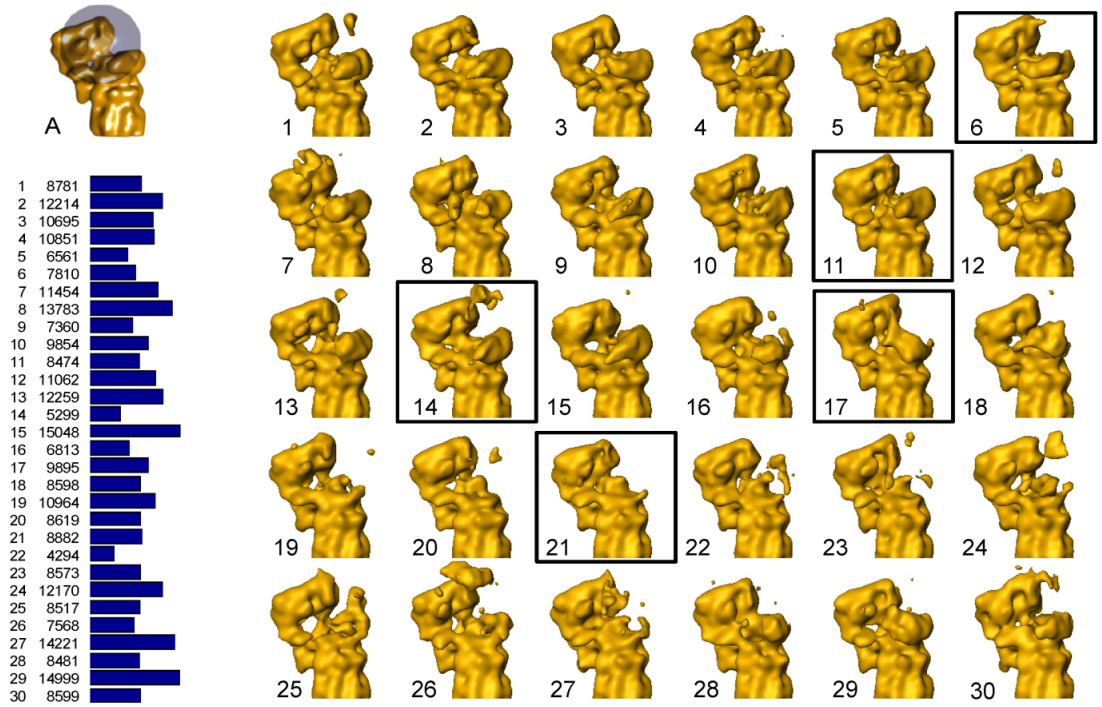

Figure 25: Classes of *S.pombe* 26S proteasomes. From a dataset of ~ 185,000 26S holocomplexes 30 classes is the maximum number which provides an amount of particles per model (~ 4,000-15,000) yielding reconstructions of sufficient resolution (3-4nm). The classification of the RP resulted in models with varying densities in the region surrounding the AAA-ATPase, between lid, base and 'extra mass' (boxed classes were chosen for detailed visualization in Fig. 24). The mask used for focused classification is shown in **A** and was chosen according to the variance map. The occupation of each class is depicted in the diagram and respective numbers of particles per class are given. Figure adapted from (Bohn et al., 2010).

4.3.7. Classification of assembly intermediates

A second ML3D classification step allowed to further separate partially (dis)assembled proteasomes (Fig. 26). Partially (dis)assembled 26S proteasomes show a high structural heterogeneity, but all classes show a ring-like protrusion adjacent to the CP, resembling the density attributed to the AAA-ATPase hexamer. Furthermore, protrusions from the AAA-ATPase density colocalize with the 'extra mass' in 26S holocomplexes (Fig. 26 A, B). The remaining densities connect to the upper N-ring of the AAA-ATPase hexamer and strongly resemble parts of the RP density of 26S holocomplexes: an elongated stretch in the distal part of the RP lid (Fig. 26 B). Whereas it is not possible to assign these densities clearly to certain

in vivo assembly intermediates, their colocalization with rigid stretches of almost no variance is indicative of scaffolding proteins (e.g. Rpn1 and Rpn2), providing a backbone for RP subunit arrangement.

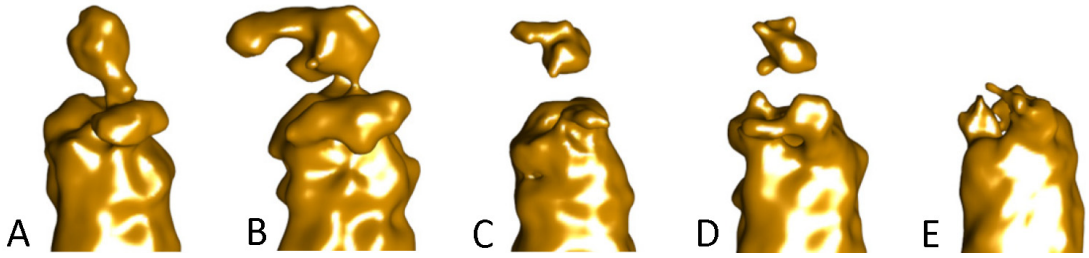

Figure 26: Classes of assembly-intermediates of *S.pombe* 26S proteasomes. Of ~ 270,000 particles from the *S.pombe* wildtype dataset approximately one third belongs to partially (dis)assembled proteasomes (Fig. 18). ML3D classification further separates this subset according to the densities in one of the RPs. Apart from densities in proximity of the CP α-ring found in all classes – resembling the density of the AAA-ATPase hexamer – the 'extra mass' becomes discernable as well (**A, B**). Interestingly, further densities belonging to the distal lid region of the RP (**B**) colocalize to the elongated stretch of almost no variance in holocomplexes (Fig. 24). These densities form a contact to the base of the RP via the N-ring of the AAA-ATPase hexamer.

4.4. Putative localization of Rpn10

The overall structural arrangement of *D.melanogaster* and *S.pombe* proteasomes is very similar (Fig. 27). Notably, the variable 'extra mass' was at approximately the same position where a variable mass was observed before with the *D.melanogaster* proteasome (Nickell et al., 2009). However, in *S.pombe* occupancy with the 'extra mass' was much higher than in *D.melanogaster*. Whereas in *S.pombe* the extra mass is present at either one or both RPs, the extra mass is observed only at one RP or not at all in the *D.melanogaster* proteasome (Fig. 27). In *D.melanogaster*, the relative abundance of the extra mass compared to the remaining RP density (~ 25%) correlated to the abundance of Rpn10 as determined by MS. Therefore, Rpn10 was tentatively localized to the extra mass. Interestingly, the relative abundance of the extra mass (~ 75%) also correlates to the Rpn10 stoichiometry (0.8:1). Thus, the EM and MS data support the putative localization of Rpn10.

A further density difference can be distinguished at the RP base ('hinge region', Fig. 27). This difference is located in an area of high variance in both, *D.melanogaster* and *S.pombe* (Fig. 23), suggesting a high flexibility of this domain. The structural features of this 'hinge region' are not as well defined as the short bi-helical repeats described earlier (Fig. 21 A, B).

However, with its configuration (e.g. width of helical ribbons) being in good agreement with the TPR fold, the assignment of Rpn10 to the 'hinge region' is unlikely.

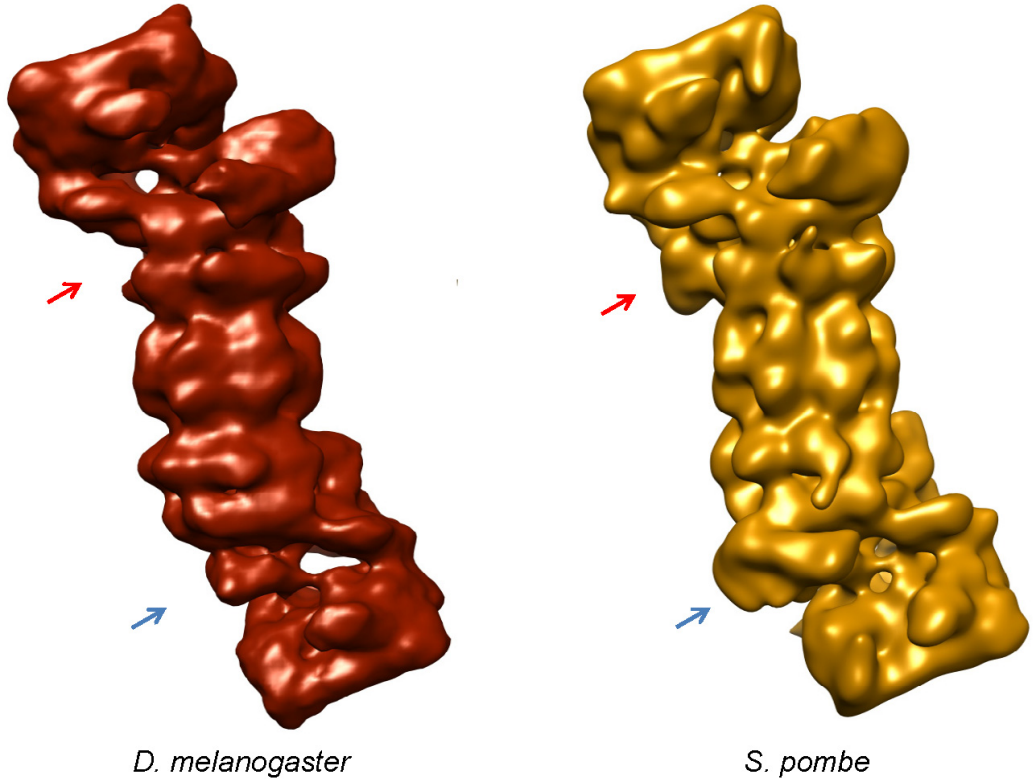

Figure 27: Model comparison of *S.pombe* and *D.melanogaster* 26S proteasomes. The map of *D.melanogaster* with a nominal resolution of ~ 25 Å (Nickell et al., 2009) (**left**) was compared to the model of *S.pombe,* bandpass filtered to the same resolution (**right**). The isosurface levels of both models were chosen such that they encapsulate the volume expected for the CP. At this resolution, both reconstructions are very similar. However, a clear difference in the occupancy of the 'extra mass' can be seen (blue arrows). A further density difference at the base of the RP ('hinge' region, red arrows) is located in an area that is highly variable in both, *S.pombe* and *D.melanogaster* 26S proteasomes. Figure adapted from (Bohn et al., 2010).

4.5. Quaternary structure of the CP/AAA-ATPase sub-complex

Previously a model for the quaternary structure of the AAA-ATPase hexamer and its position on the CP based on the cryoEM map of the *D.melanogaster* 26S proteasome and protein-protein interactions from the literature was suggested (Förster et al., 2009). The predicted AAA-ATPase topology, Rpt1/Rpt2/Rpt6/Rpt3/Rpt4/Rpt5 was recently confirmed by disulfide engineering (Tomko et al., 2010). The suggested positioning of the AAA-ATPase hexamer is however not in accordance with (Gillette et al., 2008; Yu et al., 2009): from interactions of the isolated C-terminal peptides of Rpt2 and Rpt5 with the CP subunits a

different CP-AAA-ATPase topology was suggested, which is not compatible with the model from (Förster et al., 2009) and the protein-protein interactions underlying it.

In order to address the question of the correct topology of the AAA-ATPase hexamer and possibly its relation to other RP subunits, chemical crosslinking in conjunction with MS was conducted. Disuccinimidyl suberate was used to crosslink lysine residues of proteasomal subunits in close proximity to each other (~ 20 Å). Subsequently, the crosslinked peptides were analyzed by LC-MS/MS and refined (Rinner et al., 2008). The crosslinking/MS data revealed several interactions within the AAA-ATPase and between CP and the AAA-ATPase subunits (Table 5).

Residue 1	Residue 2	C_α-C_α (Å)
Intra ATPase		
Rpt4:71	Rpt5:124	15
Rpt4:318	Rpt5:217	9
Rpt5:77	Rpt4:38	15
Rpt3:299	Rpt6:357	16
Rpt4:71	Rpt5:124	15
Rpt3:186	Rpt4:164	24
Rpt5:72	Rpt4:41	13
Rpt4:37	Rpt4:41	6
Rpt3:242	Rpt3:299	17
Rpt5:72	Rpt4:38	10
Intra ATPase cavity		
Rpt5:124	Rpt3:207	34
Rpt5:124	Rpt4:205	31
Rpt6:84	Rpt6:218	33
Rpt1:120	Rpt4:205	35
Rpt1:124	Rpt3:207	39
ATPase – 20S		
alpha4:168	Rpt1:422	17
alpha4:168	Rpt1:426	22
alpha2:176	Rpt6:390	17
alpha2:176	Rpt6:347	16
ATPase – lid		
Rpt3:49	Rpn11:281	NA

Table 5: Identified crosslinked lysines involving the AAA-ATPase hexamer. The data were classified according to the positions of the crosslinked residues. For each crosslink, both identified lysine residues and their distance in our CP-ATPase model are shown. The Intra ATPase crosslinks are consistent with the experimentally confirmed Rpt1-Rpt2-Rpt6-Rpt3-Rpt4-Rpt5 topology. Table adapted from (Bohn et al., 2010).

Four different crosslinks between Rpt and CP subunits were identified: two different residue pairs link Rpt1-α4 and Rpt6-α2, respectively (Fig. 28 A – C). All four crosslinks are in

excellent agreement with the model from (Förster et al., 2009). In fact, no other AAA-ATPase rotation can fulfill these restraints (Fig. 29). Thus, the crosslinking data derived from the fully-assembled 26S proteasome comply with the published protein-protein interaction data, such as Rpt2 binding to α4 (Zhang et al., 2000; Chen et al., 2008), as well as genetic data suggesting Rpt2-α3 interaction (Bajorek et al., 2004), but they do conflict with the interactions reported for the synthetic C-terminal Rpt peptides (e.g., Rpt5-α3 and Rpt5-α4) (Gillette et al., 2008; Yu et al., 2009). This discrepancy might be explained by different binding specificities of the full-length proteins as compared to the short C-terminal peptides.

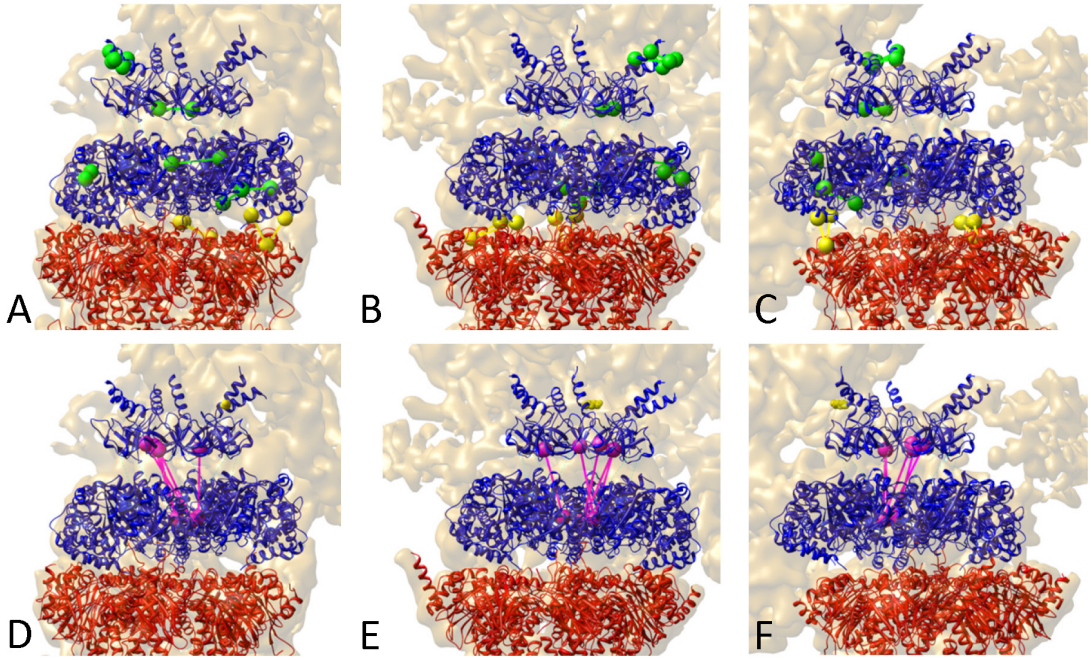

Figure 28: Crosslinks identified in the Regulatory Particle. Rpt1-α4 and Rpt6-α2 crosslinks (**A-C**, yellow) between AAA-ATPase (blue) and CP (red) corroborate the previously suggested CP-AAA topology (Tomko et al., 2010). The crosslinks between AAA-ATPase subunits (**A-C**, green) confirm the Rpt1/Rpt2/Rpt6/Rpt3/Rpt4/Rpt5 topology of the AAA-ATPase hexamer. Crosslinks in the AAA-cavity are indicative of inducible conformational changes of central rings (**D-F**, magenta). Depicted in yellow is Lys:49 of Rpt3, which has been crosslinked with Lys:281 of Rpn11 (**D-F**). The viewing direction was chosen as in Fig. 19 and Fig. 24.

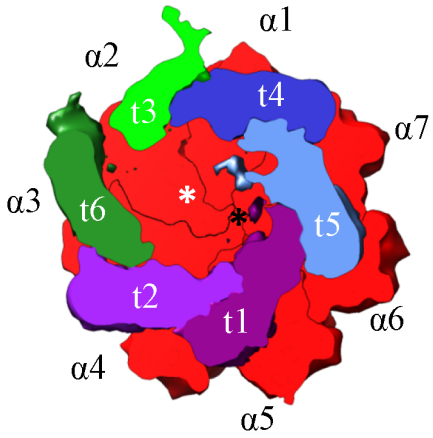

Figure 29: Topology of the AAA-ATPase hexamer and its relation to the CP. Top-view visualizing the positioning of the AAA-ATPase ring (t1-t6, center of AAA-ring: white *) relative to the α-ring of the CP (α1-α7, center: black *). Figure adapted from (Bohn et al., 2010).

Five crosslinks connecting residues located in the AAA-ATPase channel were identified. All of these crosslinks connected residues in the proximal ring of the N-terminal domains and the highly conserved Ar-Φ loop located in the AAA-ring (Fig. 28 D – F). The distances of these residues in the AAA-ATPase model (~35 Å, Table 4) exceeded the distance that is typically bridged by disuccinimidyl suberate (<20 Å). Indeed, it has been suggested that PAN and the proteasomal AAA-ATPases possess a translocation mechanism similar to that of HslU (Zhang et al., 2009b). In HslU, the Ar-Φ loop undergoes substantial conformational changes upon ATP hydrolysis, which is believed to enable it to pull the substrate through the pore (Wang et al., 2001). The identified crosslinks suggest that the Ar-Φ loop of the AAA-ATPase of the 26S proteasome also displays a structural flexibility similar to that observed with HslU.

4.6. Labeling of RP subunits with colloidal gold

The Ubiquitin C-terminal Hydrolase Uch37 is specifically inhibited by the non-hydrolysable substrate analogue ubiquitin aldehyde (Ub-Al) (Hölzl et al., 2000). Conjugation of Ub-Al to the surface of electron dense gold particles generates a suitable marker for cryoEM, assuming a specific binding of Ub-Al to the proteasome subunit Uch37 (Hölzl et al., 2000) (Fig. 30).

Protein – gold interactions are mainly based on hydrophobicity (e.g. tryptophan), SH-groups (cysteine) and electrostatics (e.g. lysine). To increase binding affinity of Ub-Al, sulfhydryl groups can be introduced to its lysine residues by reacting with 2-Iminothiolane. Purified 26S proteasomes from *D.melanogaster* were incubated with activated Ub-Al gold and show a

distinct labeling pattern at the RP of 26S proteasomes (Fig. 30), whereas no gold particles could be observed near or at the CP. This indicates a preferred binding of the marker gold to Uch37 via Ub-Al. However, BSA conjugated to gold or gold particles alone show a similar binding pattern (Fig. 31). Binding of colloidal gold is limited to a specific region near the base-lid interface of the RP, but does not seem to depend on the presence of a protein marker. Whereas the preparation of protein-gold conjugates is an established and widely used method, it does not guarantee full coverage of the gold particle surface by proteins. 26S preparation 'impurities', e.g. free proteasomal subunits or proteasome-interacting proteins, may bind to free surfaces of gold particles.

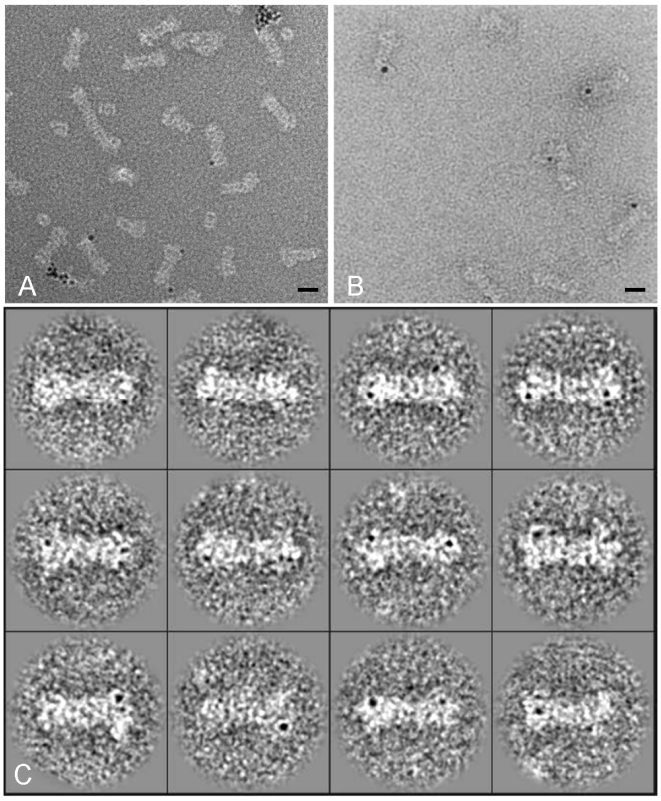

Figure 30: RP labeling with colloidal gold. Thiocyanate colloidal gold binds to the RP of *D.melanogaster* 26S proteasomes (**A**), similar to the binding of activated ubiquitin-aldehyde gold observed previously (**B**) (Hölzl et al., 2000). A gallery of single particles shows, that colloidal gold binds to the distal lid region (**C**). Scale bar is 10 nm.

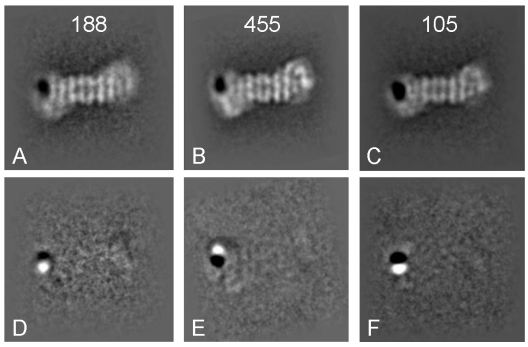

Figure 31: Colloidal gold binds RPs specifically. *D.melanogaster* 26S proteasomes were incubated with colloidal gold (control, **A** and **D**), BSA – colloidal gold (**B** and **E**) or ubiquitin-aldehyde – colloidal gold (**C** and **F**), negatively stained, classified and averaged (number of particles on top). Whereas no binding was observed in the central part (CP and RP base), a clear high density region is distinguishable in the distal RP lid region. Principal Component Analysis (PCA) allows distinguishing distinct sub-classes, indicative of a shift of the gold particle with respect to the remaining 26S (difference maps of the averages in **D-F**).

In summary, two explanations for the specific binding of colloidal gold seem most likely: (*i*) a subregion of the RP has a very high affinity for the electrostatic potential displayed on colloidal gold particles or (*ii*) proteins bound to colloidal gold particles may mimic proteasomal substrates, targeted for degradation. Given the efficiency and specificity of gold binding to RPs, this method was found to be unsuitable to label RP subunits (Uch37) specifically *via* substrate analogues or inhibitors (Ub-Al) bound to gold particles.

4.7. Putative localization of Uch2

Uch2, the *S.pombe* homolog of Uch37, is not essential in yeast (Li et al., 2000; Stone et al., 2004) and proteasomal subunits can assemble to active 26S holocomplexes in the absence of Uch2 (Fig. 10 H). Whereas the C-terminal part of Uch2 is required for binding to the 26S proteasome (Hamazaki et al., 2006; Yao et al., 2006), C-terminal tagging of Uch2 by GFP does not abolish its binding capability to 26S proteasomes (Fig. 32, 33). These properties of Uch2 allow assessment of two 'labeling' techniques based on genetic modifications, potentially suitable for analysis by single particle analysis: (*i*) deletion of Uch2 to create a missing density and (*ii*) C-terminal GFP fusion to create an additional density.

Purifications of 26S proteasomes from Uch2-GFP and ΔUch2 strains showed a similar abundance of 26S holocomplexes as the wildtype preparation (Fig. 10, 12). GFP fluorescence can be observed at the cellular level (Fig. 32), indicative of the correct, barrel-shaped domain fold of GFP (Fig. 39 F).

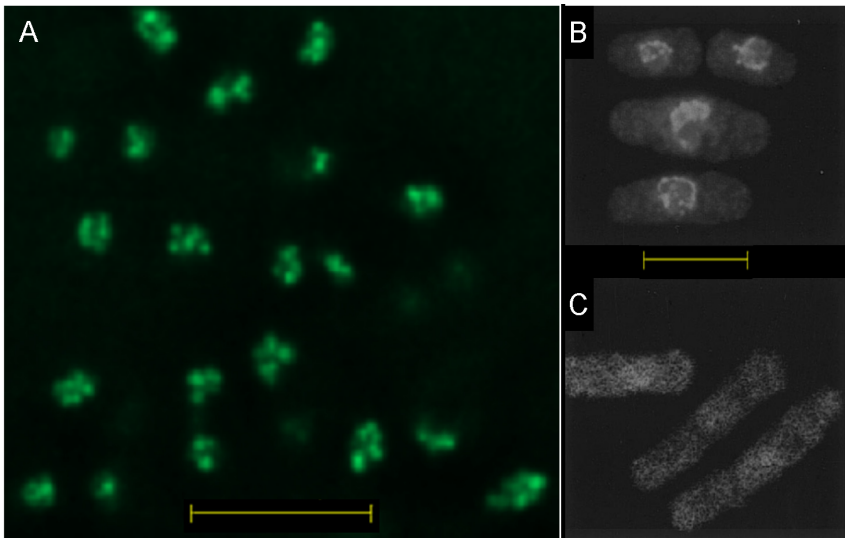

Figure 32: Uch2-GFP distribution in *S.pombe* cells. GFP is fluorescently active in *S.pombe* cells with GFP C-terminally tagged Uch2 (**A**). The distribution of the fluorescent signal is similar to the perinuclear distribution of Uch2-GFP observed before (**B**) (Li et al., 2000). In contrast, Uch2 lacking its C-terminal domain responsible for proteasome binding, shows a disperse cellular distribution (**C**). Figure B and C adapted from (Li et al., 2000). Scale bars are 10 µm.

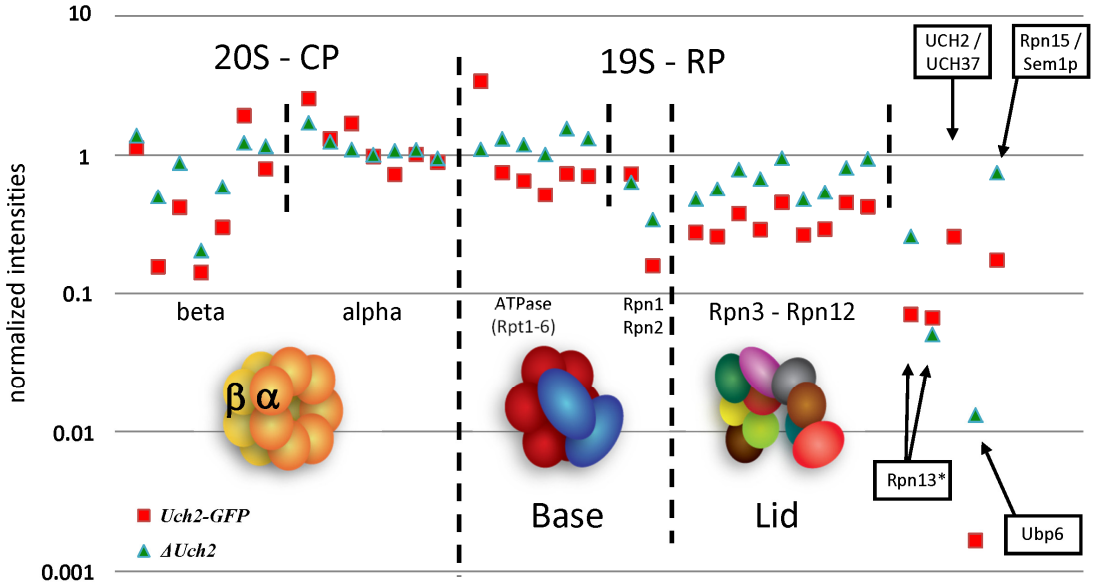

Figure 33: Subunit composition of 26S proteasomes from *S.pombe* Uch2-GFP and ΔUch2. Sucrose gradient fractions showing highest activity in Suc-LLVY-AMC cleavage were trypsin digested and analyzed by MS as described (Fig. 11 and (Nickell et al., 2009)). In both strains the canonical 26S proteasome subunits α1-7, β1-7, Rpt1-6, Rpn1-3, and Rpn5-12 are approximately present in an equimolar ratio. Rpn13 and Ubp6 were found in substoichiometric amounts. Rpn15/Sem1p was detected by a single peptide, while for all other subunits 19±10 peptides have been detected; thus, the determined Rpn15 stoichiometry is inaccurate. No peptides were detected for Uch2 in ΔUch2 preparations.

26S proteasome preparations from *S.pombe* Uch2-GFP and ΔUch2 strains show similar abundances of 26S proteasome-associated subunits when compared to wildtype *S.pombe* (Fig. 33). As expected, no peptide for Uch2 was detected in ΔUch2 preparations. In Uch2-GFP and ΔUch2 the canonical 26S proteasome subunits α1-7, β1-7, Rpt1-6, Rpn1-3 and Rpn5-12 are present in a 1:1 stoichiometry relative to their wildtype counterparts. Rpn13a (~ 0.3:1) and Rpn13b (~ 2:1) seem to have an altered binding affinity to the 26S proteasome in ΔUch2 (Fig. 34). Rpn13 binds to the C-terminal end of the proteasome scaffolding protein Rpn2 (Ito et al., 2001; Gandhi et al., 2006; Hamazaki et al., 2006; Schreiner et al., 2008). Both Rpn13 orthologs share high sequence similarity in their N-terminal domain, which is responsible for association with the proteasome (Chen et al., 2010). The results of the quantitative MS indicate, that in the absence of Uch2, the Uch2-binding ortholog of Rpn13, Rpn13b, has an increased binding affinity to the proteasome. Rpn13b has been shown to deliver di-ubiquitin bonds to the proximity of the catalytic centre of Uch2, a process that is stalled in the absence of Uch2.

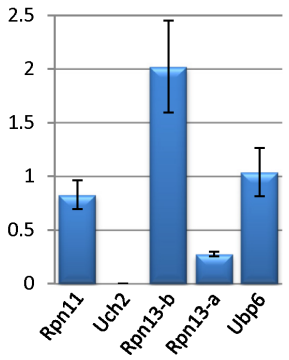

Figure 34: Relative abundance of proteasomal subunits in wildtype and ΔUch2 *S.pombe* strains. In ΔUch2 preparations the canonical 26S proteasome subunits α1-7, β1-7, Rpt1-6, Rpn1-3, and Rpn5-12 are approximately present in the same amount when compared to the wildtype preparation (Fig. 11). No peptides were detected for Uch2. Interestingly, the Uch2-interacting ubiquitin receptor Rpn13b shows a two-fold increased binding affinity to purified 26S proteasomes, while Rpn13a, which lacks the Uch2-interacting C-terminal domain, shows an approximately three-fold reduced binding affinity. The deubiquitinating enzymes Rpn11 and Ubp6 were detected in similar amounts as in wildtype preparations.

From the classification of the wildtype dataset a substantial amount of selected Uch2-GFP and ΔUch2 particles was anticipated to resemble partially (dis)assembled 26S proteasomes. As the smaller size of the datasets (29,000 and 44,000 particles for Uch2-GFP and ΔUch2) does not allow for two rounds of ML3D classification, particles resembling 26S holocomplexes were separated by ML3D into three classes for each dataset independently (Fig. 35).

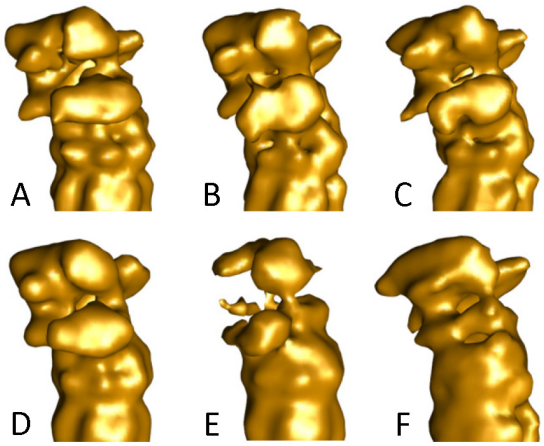

Figure 35: ML3D classification of Uch2-GFP and ΔUch2 particles. To purify the particle ensembles *in silico*, particles were classified with ML3D and separated into three groups. In the case of Uch2-GFP (~ 29,000 particles), all classes resembled fully assembled holocomplexes with only minor variances in the RP (**A-C,** with 39 %, 25 % and 36 % of all particles, respectively). Classification of ΔUch2 particles (~ 30,000 particles) reveals only one class resembling 26S holocomplexes (**D,** 48 %), whereas class **E** (34 %) resembles partially (dis)assembled 26S proteasomes as described for the wildtype (Fig. 24). In the third class (**F,** 18 %) most particles are aligned falsely. Consequently, the model shows artefacts such as elongation and gaps in the CP density. Notably, it was not possible to identify assembly intermediates for Uch2-GFP, indicative of a lower abundance of partially (dis)assembled proteasomes.

As in the *S.pombe* wildtype preparation, one ΔUch2 class (~ 34 % of all particles) resembles partially (dis)assembled 26S proteasomes (Fig. 34) and one class shows all the features of 26S holocomplexes. The third class (~ 18 %) consists mainly of falsely aligned 26S holocomplexes and assembly intermediates, generating a 3D model with artefacts such as holes in the CP density. Surprisingly, all three Uch2-GFP classes show all the features of 26S holocomplexes. The differences among the three classes are minor compared to the large variations observed between fully and partially assembled 26S proteasomes. This indicates, that the amount of partially (dis)assembled proteasomes in Uch2-GFP is less than in the class with the smallest occupancy (Fig. 35 B, ~ 25 %). Taken together, these data suggest, that the amount of partially (dis)assembled 26S holocomplexes in Uch2-GFP preparations is < 25 % compared to ~ 33 % in wildtype and > 34 % in ΔUch2. Whereas this indicates a possible role of Uch2 in the stability of the RP, it cannot be excluded, that subtle changes during the purification procedure could alter 26S holocomplex stability.

In order to localize Uch2 within the RP, reconstructions of Uch2-GFP and ΔUch2 were compared to the wildtype model. Given the molecular weight of Uch2 and GFP of ~ 30 kDa each, a difference in the EM density maps corresponding to ~ 60 kDa between Uch2-GFP

and ΔUch2 can be expected. However, when comparing 2D class averages of wildtype, Uch2-GFP and ΔUch2 no significant differences were found (Fig. 36).

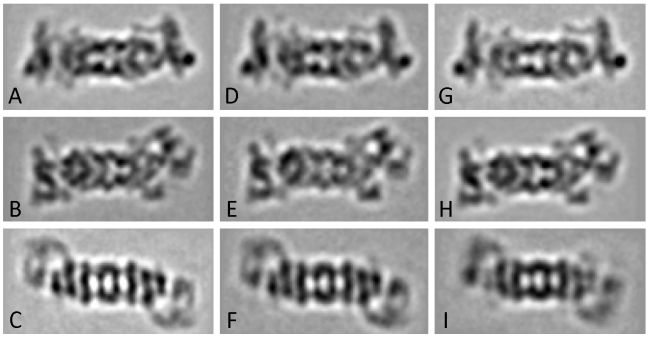

Figure 36: 2D class averages of vitrified *S.pombe* 26S proteasomes. Pictured are examples of 2D projection class averages of 26S holocomplexes (**A-C**: wildtype, **D-E**: ΔUch2, **G-I**: Uch2-GFP). No significant density differences can be distinguished between wildtype, knockout and fusion strain, albeit a molecular weight of ~ 30 kDa for both, Uch2 and GFP.

Interestingly, PCA classification of 2D projection averages reveals particles with an unusual RP structure in the ΔUch2 dataset (Fig. 37) that is not observed in the wildtype or the Uch2-GFP dataset. These 'strange caps' resemble projections of partially (dis)assembled proteasomes found in wildtype and ΔUch2 (Fig. 37 E, F). The inability to recover these projection averages by PCA classification in wildtype and Uch2-GFP indicates their lower abundance when compared to ΔUch2.

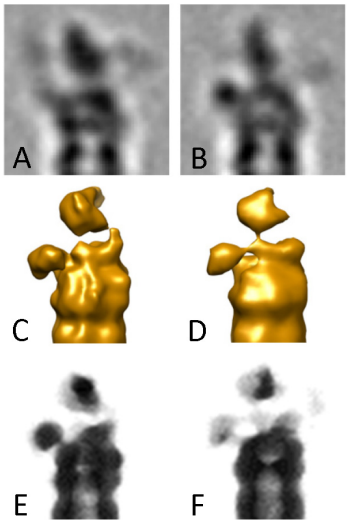

Figure 37: Assembly intermediates in wildtype and ΔUch2 proteasomes. Classification of projection averages by PCA identifies particles with an unusual RP structure in the ΔUch2 dataset (**A, B**) in addition to typical averages of this angular class (Fig. 36 A, D and G). These particles are reminiscent of partially (dis)assembled 26S proteasomes found in wildtype (**C:** wildtype class B from Fig. 26) and ΔUch2 (**D:** ΔUch2 class E from Fig. 35) datasets (**E, F** are projections of C and D, respectively). These 'strange caps' are not found by PCA classification in wildtype or Uch2-GFP datasets.

The *in silico* purified datasets of Uch2-GFP and ΔUch2 allowed 3D reconstructions reaching to intermediate resolution (~ 25 Å, Fig. 38). When comparing the 3D reconstructions of Uch2-GFP, ΔUch2 and wildtype, no significant density differences can be found (Fig. 39). This is consistent with the PCA analysis of 2D projection images (Fig. 36).

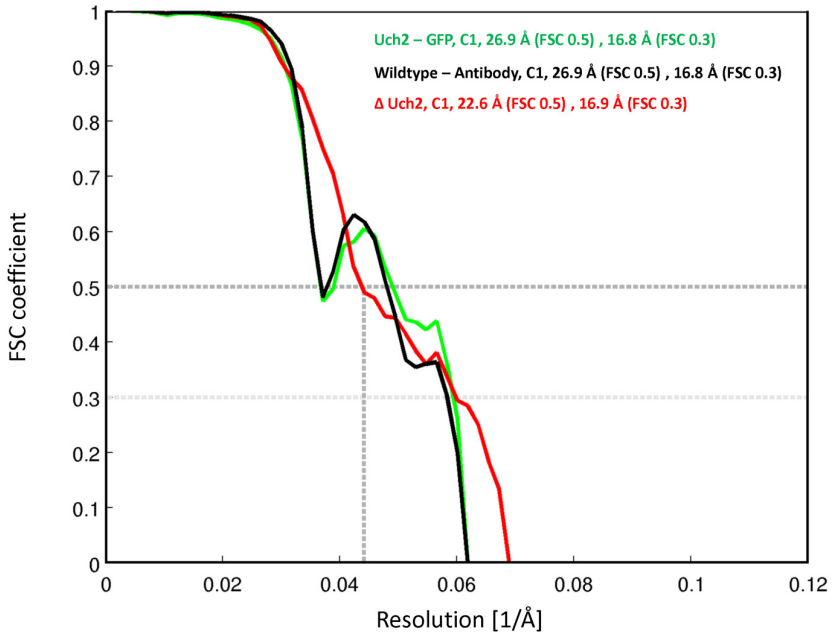

Figure 38: Fourier Shell Correlation (FSC) curves of labeled *S.pombe* 26S proteasomes. Datasets were binned once and refined with a final angular increment of 1°, without imposed symmetry. By covering a wider range of defocus (1 – 3 µm) for ΔUch2 datasets by acquiring focal pairs, more information from higher frequencies was recovered (first gap in FSC as in Uch2-GFP or wildtype-Ab is filled).

Whereas no novel proteins have been detected by MS in either Uch2-GFP or ΔUch2, Rpn13b was detected in an approximately twofold increased amount in ΔUch2 26S preparations (Fig. 34). The similarity of ΔUch2 and wildtype models could be explained by density replacement, due to their similar spherical fold and molecular weight (~ 35 kDa and ~ 40 kDa for Uch2 and Rpn13b). This however does not explain the similarity of Uch2-GFP and wildtype models: at a resolution of ~ 25 Å, densities corresponding to the size of GFP should be clearly discernable (Fig. 38 F). Whereas the twelve amino acid linker between Uch2 and GFP ensures correct folding and functionality of GFP, it also introduces a possible source of flexibility to the added GFP domain with respect to Uch2.

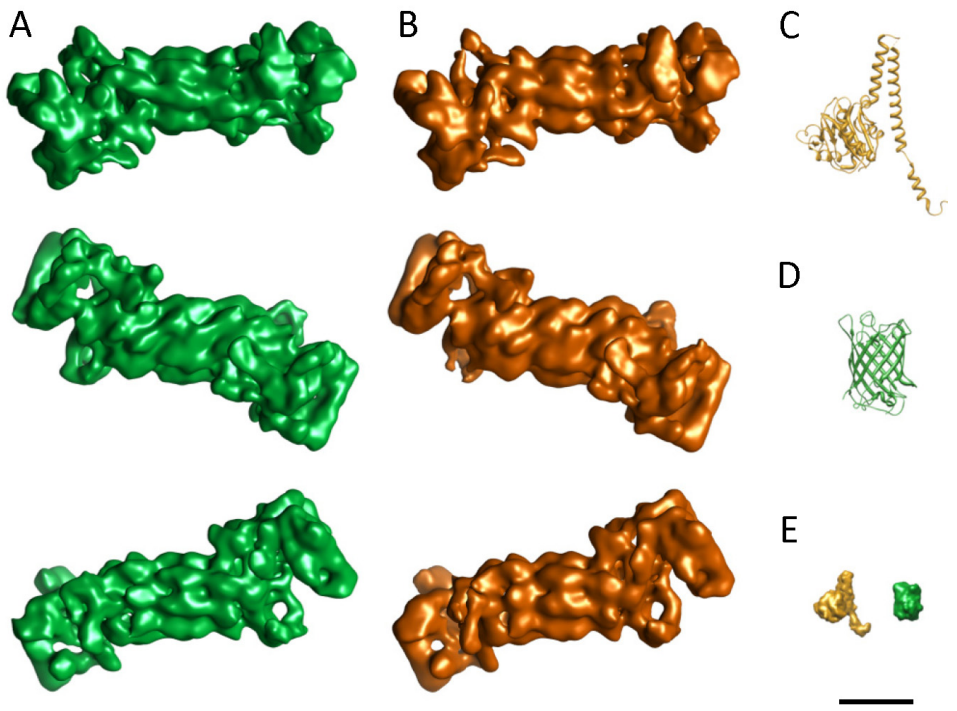

Figure 39: Comparison of 3D reconstruction from Uch2-GFP and ΔUch2. The models of 26S proteasomes from *S.pombe* Uch2-GFP (**A**) and ΔUch2 (**B**) show very high similarity. The expected density difference corresponds to approximately 65 kDa, contributed by Uch2 and GFP. For comparison, **E** shows the simulated EM densities of Uch2 and GFP, drawn to the same scale of the models in A and B (**C** and **D**: crystal structures of Uch2 and GFP, respectively; scale bar for A, B, E: 10 nm).

To test, whether Uch2 can be unambiguously assigned to the variable densities in the 'mouth' region observed in wildtype 26S proteasomes (Fig. 24), Uch2-GFP and ΔUch2 datasets were analyzed by focused classification as described (Fig. 13, 24). Both datasets of ~ 30,000 particles each were split into twelve classes comprising ~ 3,000 – 6,000 particles, yielding reconstructions at a resolution of ~ 4 nm (Fig. 39).

As among wildtype 26S proteasomes, some classes are missing the 'extra mass' (Fig. 40: classes 1 and 2), whereas others harbor extra densities in the 'mouth' region between 'extra mass' and the lid of the RP (classes 4 and 10-12). The absence of Uch2 in ΔUch2 indicates that Uch2 cannot be the only proteasome-interacting protein responsible for the variable densities. Due to the modest resolution of the reconstructions and the relatively small amounts of particles per model, it is not possible to assign these variable densities to specific subunits. However, the amount of additional densities detected in Uch2-GFP is higher than in ΔUch2: ten Uch2-GFP classes show more (classes 3, 4, 7, 10-12) or similar amounts (classes 2, 5, 6, 8) of additional densities compared to ΔUch2.

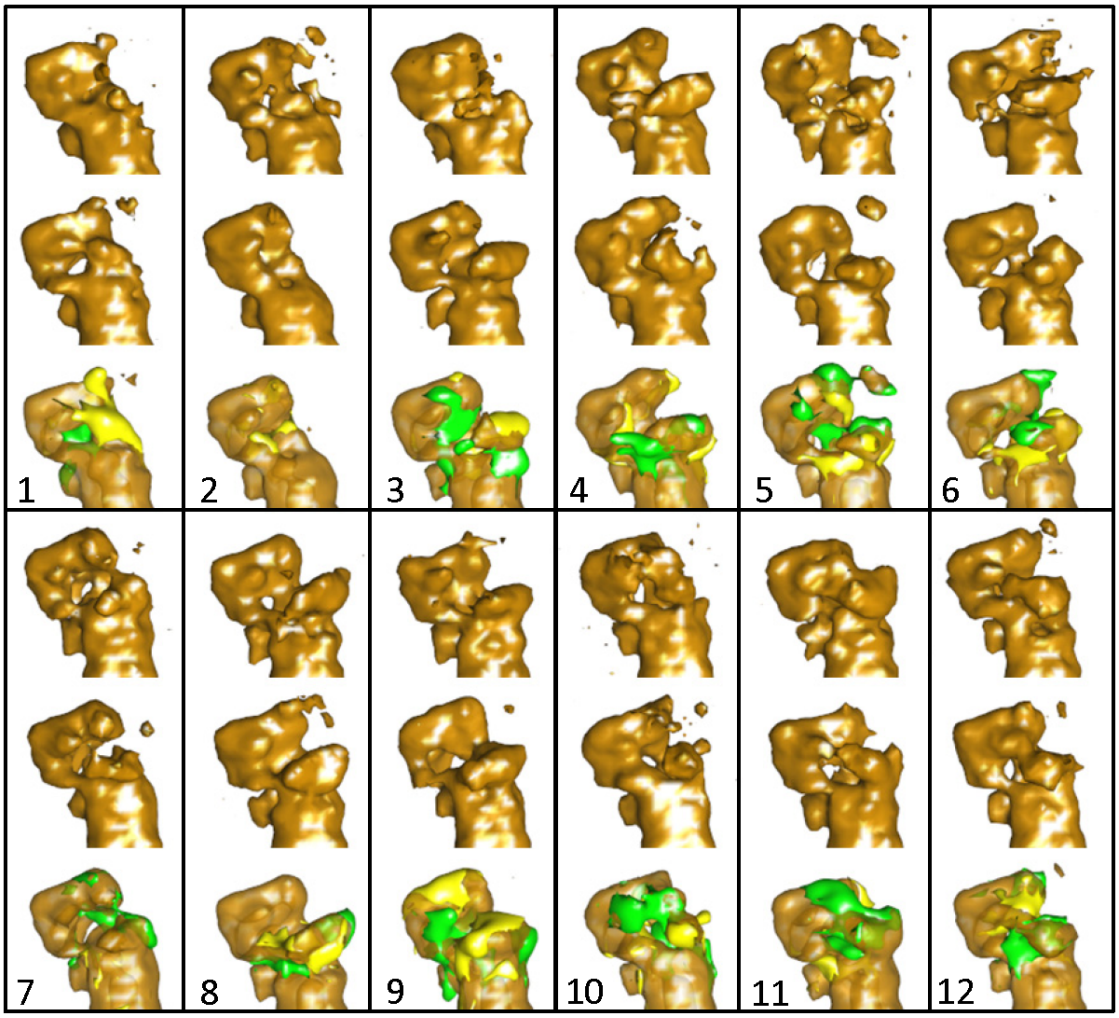

Figure 40: Classes of Uch2-GFP and ΔUch2 *S.pombe* 26S proteasomes. From datasets of ~ 30,000 26S holocomplexes twelve classes for Uch2-GFP (**top**) and ΔUch2 (**middle**) were iteratively refined by focused classification (~ 3,000 – 6,000 particles per model) yielding reconstructions of modest resolution (~ 4 nm, classification mask as in Fig. 25). The classification of the RP for both mutants resulted in variations similar as observed in the wildtype classification (Fig. 25). Classes were sorted according to their similarity and their differences were plotted (**bottom,** extra densities in Uch2-GFP in **green**, extra densities in ΔUch2 in **yellow**). When compared to the wildtype classification, both mutant strains show the same major variances among the classes, e.g. missing 'extra mass' (**1, 2**) or extra densities at the RP lid (**5**). Notably, extra densities between lid and 'extra mass' are observed in both strains (classes **9-12** and **9, 11, 12** for Uch2-GFP and ΔUch2, respectively), indicating that these densities cannot be attributed only to Uch2. However, the majority of Uch2-GFP classes (ten out of twelve) show more or similar amounts of extra densities compared to ΔUch2.

Taken together, the similarity of Uch2-GFP, ΔUch2 and wildtype models can be explained by one or a combination of the following: (*i*) multiple binding sites of Uch2 in the belt of high variance surrounding the AAA-ATPase hexamer attributed to proteasome-interacting proteins; (*ii*) abundance of Uch2 in sub-stoichiometric amounts; (*iii*) replacement of Uch2 by its binding partner Rpn13b.

4.8. Localization of Rpn11

Rpn11 was labeled in 26S proteasomes from wildtype *S.pombe* preparations with an anti-Flag antibody (Ab) using the C-terminal 3x-Flag-epitope of Rpn11 (Fig. 41 A – C). CryoEM images of 26S proteasome incubated with the anti-Flag Ab were acquired at a nominal defocus 2.5±0.5 µm. Micrographs displayed typical 26S proteasome 'twin' configurations (i.e. two holocomplexes aligned with their long axes), as well as isolated 26S proteasomes (Fig. 41 D).

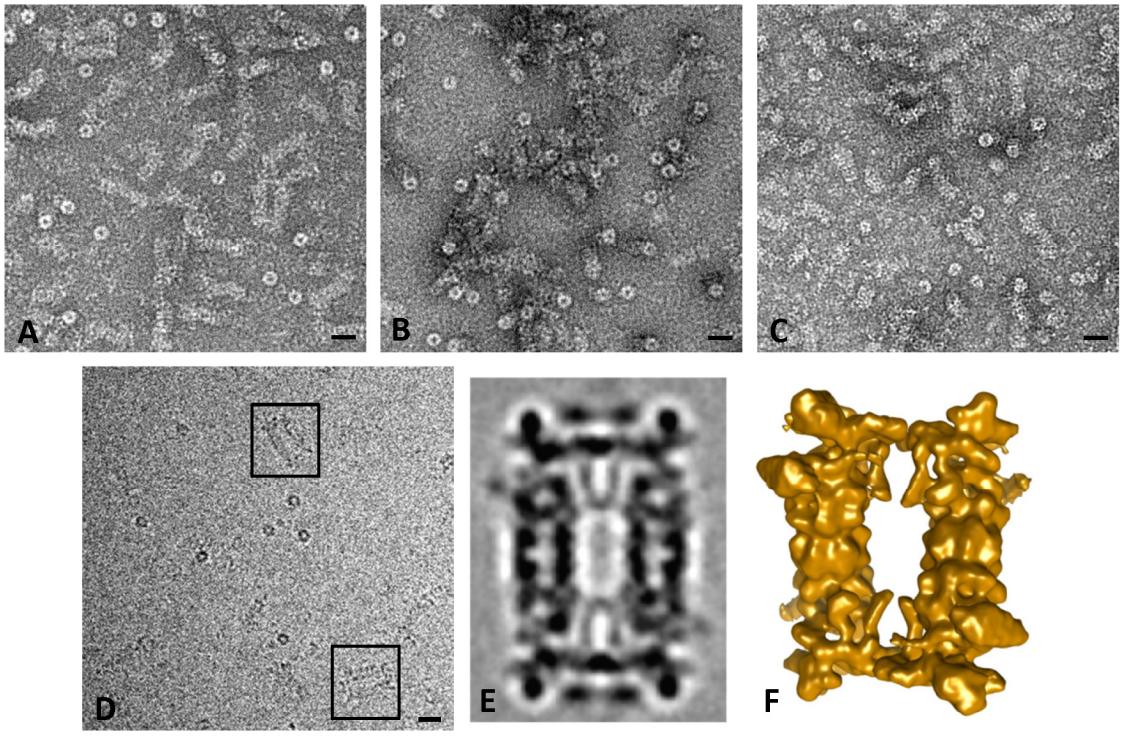

Figure 41: Binding of Anti-Flag antibodies to the Rpn11-C-terminal Flag-tag in intact 26S proteasomes. After incubation of proteasomes with monoclonal Anti-Flag-antibody, particles often seem to be aligned in pairs of two ('twins') or to form aggregates at higher Ab concentrations (**A**, Ab dilution 1:100; **B**, Ab dilution 1:10; **C**, undiluted; **D**, Ab dilution 1:100, cryoEM; **A-C**: negative stain). 'Twins' – two 26S proteasomes displaying four identical 3xFlag epitopes linked by two anti-Flag-Abs – typically show only one orientation when embedded in ice (**E**: 2D average of ~ 600 'twins'). The obtained class average can be used to model the relative orientation of two 3D density maps of the 26S proteasome (**F**). Scale bars for A – C: 10nm, D: 20 nm.

The 2D class averages clearly indicate a distinct density above the lid region (Fig. 41 E), which is not present in the unlabeled proteasome (Fig. 42 A). Since the twins adopt a single spatial orientation in the ice layer, single 26S proteasomes were reconstructed to map the Ab density in 3D. Comparison of the 17.4 Å resolution map with the map of the unlabeled 26S proteasome reveals a rather blurred additional density indicative of a high degree of

flexibility (Fig. 41 B). This density can be mainly attributed to the structural variability of the 17 residue C-terminal linker to the 3x-Flag epitope, the presence of three Flag epitopes, and the intrinsic variability of the Ab. Thus, the Ab density maps the C-terminus of Rpn11 with relative low resolution: it can be estimated that the Rpn11 C-terminus is located within a radius of ~ 80 Å around the center of the Ab mass (Fig. 41 C).

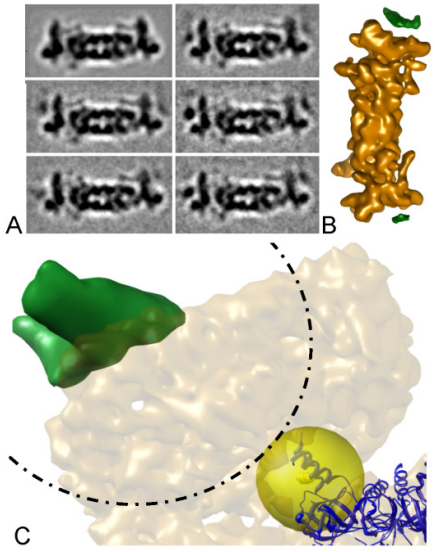

Figure 42: Mapping of the Rpn11-C-terminal 3xFlag-tag in intact 26S proteasomes. Extra densities are clearly visible in 2D class averages (**A**, upper left: control). After classification and 3D reconstruction (**B**) the center of mass of the antibody can be clearly seen on the top of the RP lid (green). The crosslink between Lys49 in the coiled-coil region of Rpt3 (**C**, yellow) and Lys281 of Rpn11 has an estimated length of 20 Å and allows mapping Rpn11:Lys281 within a sphere of 20 Å radius (yellow sphere). A sphere of ~ 80 Å radius (dotted line) with its central point in the center of mass of the antibody depicts the possible localization of the IgG binding site. Figure adapted from (Bohn et al., 2010).

Further data on the protein-protein proximity of Rpn11 to other RP subunits was obtained from crosslinking experiments described before (section 4.5). Among the identified crosslinks, one inter-subunit interaction involved Rpn11: Lys49 located in the coiled coil region of Rpt3 is in proximity to Lys281 in the C-terminal domain of Rpn11 (Table 5). Previous tandem affinity purification experiments revealed Rpt3/Rpt6/Rpn8/Rpn11 forming a sub-complex (Gavin et al., 2006), consistent with the proximity of Rpt3 and Rpn11. The placement of Rpn11 close to the AAA-ATPases is not unexpected: the deubiquitylating activity of Rpn11 is ATP dependent, but Rpn11 is not an ATPase itself, which implies cooperativity of substrate translocation and deubiquitylation (Verma et al., 2002; Yao et al., 2002). Thus, the crosslinking data suggest that part of the Rpn11 C-terminal domain is accessible from inside the cavity surrounding the AAA-ATPase whereas the C-terminus itself

projects to the periphery of the RP. There are 29 residues between Lys281 and the C-terminus, which are sufficient to bridge the ~ 30 Å space between the proteasome mouth and the periphery. Interestingly, it is exactly these ~ 30 C-terminal residues that have been shown to be essential for the role of Rpn11 in mitochondrial biogenesis (Rinaldi et al., 2008).

Based on the Ab-labeling and the crosslinking data the C-terminal domain of Rpn11 can be mapped approximately. From the crosslink Rpt3:Lys49-Rpn11:Lys281 one can conclude that these two lysine residues are ~ 20 Å apart from each other; the C-terminal domain of Rpn11 is placed at the intersection of the sphere centered on the Flag-Ab and a sphere around Lys49 of Rpt3 (Fig. 42 C). At this point the de-ubiquitylating site, which is located in the N-terminal MPN domain, cannot be mapped precisely. However, it appears reasonable that it is accessible from the 'pharynx' between base and lid, such that the placement of Rpn11 near the mouth of the AAA-ATPase is ideally positioned for Ub removal from substrates immediately prior to substrate translocation into the CP.

5. Outlook

In this study, cryoEM has been successfully applied to resolve the structure of the 26S proteasome to subnanometer resolution. Whereas at this resolution secondary structure elements become discernable, an unambiguous localization of proteasomal subunits requires a model at higher resolution. A model of the proteasome near the atomic level would allow gaining further insights into the structural organization of the 26S proteasome and the sequence of events preceding substrate degradation. Given the observed structural heterogeneity of 26S holocomplexes, the generation of crystals suitable for X-ray crystallography seems unlikely. In this study, the resolution was mainly limited by the size of the dataset and the chosen acquisition parameters (e.g. magnification). Given a larger dataset, single particle analysis should in principle allow 3D model reconstructions to near of atomic resolution (Fig. 43).

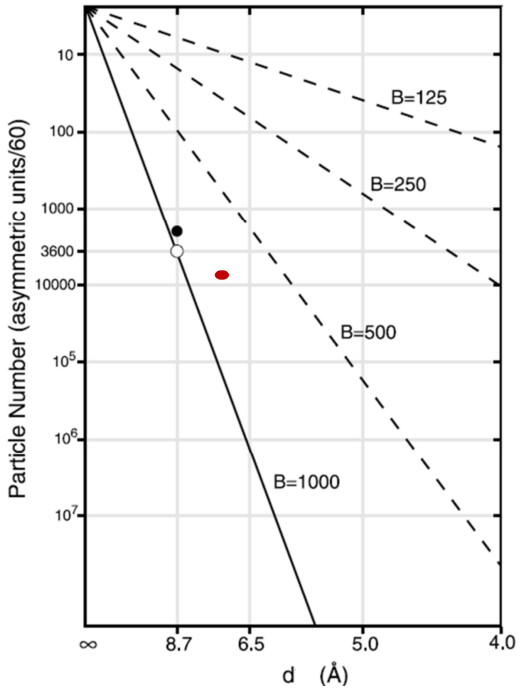

Figure 43: Particle number in relation to the expected single particle reconstruction resolution. The amount of particles necessary for high-resolution reconstruction by SPA increases drastically for each Å beyond subnanometer resolution and is dependent on the temperature factor B (model prediction black dot, experimental result white dot) (Rosenthal et al., 2003). The 26S proteasome models' resolution from this study (**red dot**: resolution of ~ 6.7 Å at FSC of 0.14, ~ 500,000 asymmetric units, temperature factor B ~ 900 before MTF correction) correlates well to the theoretical model described by (Rosenthal et al., 2003). Figure adapted from (Rosenthal et al., 2003).

However, image acquisition and procession of such datasets are extremely time consuming and only feasible with a higher degree of automation. In recent years, tremendous progress has been achieved due to, e.g., increased computing capabilities and better image digitalization hardware (i.e. larger CCD chips). Therefore it is promising to further explore the route of structural analysis of the 26S proteasome by cryoEM.

The availability of much larger datasets than used in this study will allow an exhaustive classification, promising to further analyze the structural heterogeneity among 26S proteasomes due to, e.g., the presence of (dis)assembly intermediates, proteasome-interacting proteins or bound proteasome substrates.

In this study important insights into the structural organization of the 26S proteasome have been gained by hybrid approaches: combining the cryoEM density maps with results from several different labeling techniques, such as chemical crosslinking, mass spectrometry and antibody-labeling. Here, chemical crosslinking proved a highly useful method to derive spatial restraints of this macromolecular assembly. It will be particularly interesting to expand the list of residue pairs in close proximity to each by the application of alternative crosslinkers (e.g. for acidic amino acid residues). Furthermore it seems plausible to further explore antibody-labeling of the affinity tag fused to N- or C-termini of other regulatory particle subunits, albeit the expected less stringent spatial restraints.

Whereas the mere localization of proteasomal subunits is highly valuable in context of the established interacting network of the regulatory particle, the ultimate goal of a structural study should be to shed some light on the mechanistic details underlying biological processes. It was possible to further develop the mechanistic model of proteasomal degradation in the course of this study (Fig. 44), but the majority of the regulatory particle and with it the details of substrate recognition, unfolding and translocation remain ‘uncharted territory’ – to be resolved in the future by the means mentioned above.

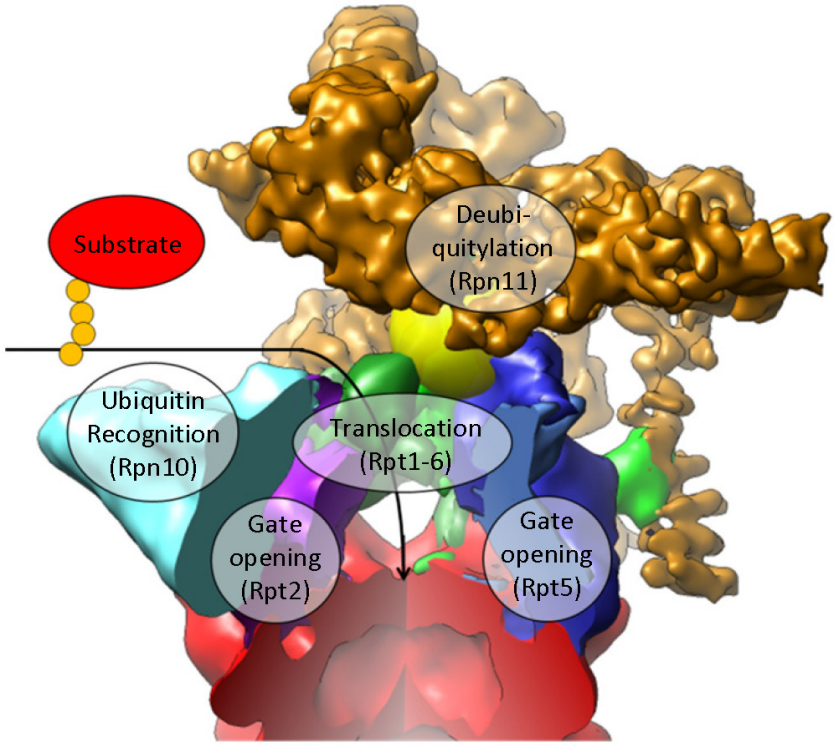

Figure 44: Integrated model of the RP and α-ring of the CP. Ubiquitylated substrates are recognized at the RP (e.g. Rpn10 or Rpn13) of the 26S proteasome, deubiquitylated (e.g. Rpn11 or Uch2) for Ub recycling and translocated (Rpt1-6) through the AAA-ATPase cavity and the opened gate (Rpt2, Rpt5) of the CP for degradation (extra mass: cyan, Rpt1-6: as in Fig. 29, Rpt3-Rpn11 cross link: as in Fig. 42 C). Figure adapted from (Bohn et al., 2010).

Due to the preservation of biological specimen in a close to native state by cryoEM, the development and analysis of precisely-defined biochemical states (e.g. substrates stalling the translocation or degradation process) would allow to study substrate recognition, unfolding or translocation by the regulatory particle. The combination of advanced classification methods such as ML3D or focused classification with appropriately sized datasets, will allow visualizing the order and organization of events preceding substrate degradation.

D Appendix

Index of Figures

Index of Tables

Abbreviations

aa	Amino acid
AAA	ATPases associated with diverse cellular activities
Ab	Antibody
ATP	Adenosine-5'-triphosphate
bp	Base pair
BSA	Bovine Serum Albumine
CCD	Charge coupled device
CP	Core particle
cryoEM	Cryo electron microscopy
CTF	Contrast transfer function
DTT	Dithiothreitol
DUB	Deubiquitylating enzymes
E1	Ubiquitin activating enzymes
E2	Ubiquitin conjugating enzymes
E3	Ubiquitin ligases
EDTA	Ethylenediaminetetraacetic acid
FEG	Field emission gun
FSC	Fourier shell correlation
GFP	Green fluorescent protein
HB	Helical bundle
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazine)-ethansulfonic acid
IgG	Immunoglobulin G
JAMM	Josephins and JAB1/MPN/MOV34 metalloenzymes
kb	Kilo base pairs
kDa	Kilo Dalton
keV	Kilo electron Volt
LB	Lysogeny broth
LC-MS	Liquid chromatography mass spectrometry
ML3D	Maximum Likelihood 3-Dimensional classification
MPN	Mpr1/Pad1 N-terminal

MS	Mass spectrometry
MTF	Modulation transfer function
MWCO	Molecular weight cutoff
m/z	Mass-to-charge ratio
OTU	Ovarian tumor proteases
PAGE	Poly acrylamide gelelectrophoresis
PAN	Proteasome-activating nucleosidase
PBS	Phosphate-buffered saline
PC	Proteasome cyclosome
PCA	Principal Component Analysis
PCI	Proteasome, COP9 signalosome, eIF3
PCR	Polymerase chain reaction
PIP	Proteasome interacting protein
Pru	Pleckstrin-like receptor for ubiquitin
PSF	Point spread function
RP	Regulatory particle
RT	Room temperature
SNR	Signal to noise ratio
SPA	Single particle analysis
TEM	Transmission electron microscope
TEMED	N,N,N',N'-Tetramethylethylendiaminetriphosphate
TPR	Tetratrico-peptide repeats
Ub	Ubiquitin
UbAl	Ubiquitin aldehyde
UBL	Ubiquitin-like
UBA	Ubiquitin-associated
UCH	Ubiquitin C-terminal Hydrolase
UIM	Ubiquitin-interacting motif
UPP	Ubiquitin proteasome pathway
USP	Ubiquitin-specific proteases
VWA	Von Willebrand A
WH	Winged-helix

References

Amerik, A. Y. and M. Hochstrasser (2004). Mechanism and function of deubiquitinating enzymes. *Biochim Biophys Acta* 1695:189-207.

Bajorek, M. and M. H. Glickman (2004). Keepers at the final gates: regulatory complexes and gating of the proteasome channel. *Cell Mol Life Sci* 61:1579-88.

Baschong, W., J. M. Lucocq, et al. (1985). "Thiocyanate gold": small (2-3 nm) colloidal gold for affinity cytochemical labeling in electron microscopy. *Histochemistry* 83:409-11.

Baumeister, W., J. Walz, et al. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell* 92:367-80.

Belote, J. M. and L. Zhong (2009). Duplicated proteasome subunit genes in Drosophila and their roles in spermatogenesis. *Heredity*.

Bohn, S., F. Beck, et al. (2010). Structure of the 26S proteasome from Schizosaccharomyces pombe at subnanometer resolution. *Proc Natl Acad Sci U S A* 107:20992-7.

Bopp, K. (2008). Kryo-Elektronentomographie an in vitro Translationssystemen, Technical University of Munich.

Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* 74:773-6.

Brink, J., S. J. Ludtke, et al. (2004). Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis. *Structure* 12:185-91.

Brohawn, S. G., J. R. Partridge, et al. (2009). The nuclear pore complex has entered the atomic age. *Structure* 17:1156-68.

Cagney, G., P. Uetz, et al. (2001). Two-hybrid analysis of the Saccharomyces cerevisiae 26S proteasome. *Physiol Genomics* 7:27-34.

Chen, C., C. Huang, et al. (2008). Subunit-subunit interactions in the human 26S proteasome. *Proteomics* 8:508-20.

Chen, X., B. H. Lee, et al. (2010). Structure of proteasome ubiquitin receptor hRpn13 and its activation by the scaffolding protein hRpn2. *Mol Cell* 38:404-15.

Ciccarelli, F. D., E. Izaurralde, et al. (2003). The PAM domain, a multi-protein complex-associated module with an all-alpha-helix fold. *BMC Bioinformatics* 4:64.

Costello, M. J. (2006). Cryo-electron microscopy of biological samples. *Ultrastruct Pathol* 30:361-71.

Dessau, M., Y. Halimi, et al. (2008). The Arabidopsis COP9 signalosome subunit 7 is a model PCI domain protein with subdomains involved in COP9 signalosome assembly. *Plant Cell* 20:2815-34.

Djuranovic, S., M. D. Hartmann, et al. (2009). Structure and activity of the N-terminal substrate recognition domains in proteasomal ATPases. *Mol Cell* 34:580-90.

Dubochet, J., M. Adrian, et al. (1988). Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* 21:129-228.

Enchev, R. I., A. Schreiber, et al. (2010). Structural insights into the COP9 signalosome and its common architecture with the 26S proteasome lid and eIF3. *Structure* 18:518-27.

Finley, D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem* 78:477-513.

Förster, F., K. Lasker, et al. (2009). An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. *Biochem Biophys Res Commun* 388:228-33.

Förster, F., K. Lasker, et al. (2010). Toward an integrated structural model of the 26S proteasome. *Mol Cell Proteomics* 9:1666-77.

Frank, J. (2006). Three-dimensional Electron Microscopy of Macromolecular Assemblies. New York, Oxford University Press.

Fu, H., N. Reis, et al. (2001). Subunit interaction maps for the regulatory particle of the 26S proteasome and the COP9 signalosome. *Embo J* 20:7096-107.

Funakoshi, M., R. J. Tomko, Jr., et al. (2009). Multiple Assembly Chaperones Govern Biogenesis of the Proteasome Regulatory Particle Base. *Cell* 137:887-99.

Gandhi, T. K., J. Zhong, et al. (2006). Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet* 38:285-93.

Gao, H., M. Valle, et al. (2004). Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *J Struct Biol* 147:283-90.

Gavin, A. C., P. Aloy, et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440:631-6.

Gillette, T. G., B. Kumar, et al. (2008). Differential roles of the COOH termini of AAA subunits of PA700 (19 S regulator) in asymmetric assembly and activation of the 26 S proteasome. *J Biol Chem* 283:31813-22.

Giot, L., J. S. Bader, et al. (2003). A protein interaction map of Drosophila melanogaster. *Science* 302:1727-36.

Girod, P. A., H. Fu, et al. (1999). Multiubiquitin chain binding subunit MCB1 (RPN10) of the 26S proteasome is essential for developmental progression in Physcomitrella patens. *Plant Cell* 11:1457-72.

Glickman, M. H. and A. Ciechanover (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* 82:373-428.

Glickman, M. H., D. M. Rubin, et al. (1998). A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* 94:615-23.

Groll, M., L. Ditzel, et al. (1997). Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* 386:463-71.

Hamazaki, J., S. Iemura, et al. (2006). A novel proteasome interacting protein recruits the deubiquitinating enzyme UCH37 to 26S proteasomes. *Embo J* 25:4524-36.

Hamazaki, J., K. Sasaki, et al. (2007). Rpn10-mediated degradation of ubiquitinated proteins is essential for mouse development. *Mol Cell Biol* 27:6629-38.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys* 28:171-93.

Hershko, A. and A. Ciechanover (1998). The ubiquitin system. *Annu Rev Biochem* 67:425-79.

Hershko, A., A. Ciechanover, et al. (2000). Basic Medical Research Award. The ubiquitin system. *Nat Med* 6:1073-81.

Heymann, J. B., J. F. Conway, et al. (2004). Molecular dynamics of protein complexes from four-dimensional cryo-electron microscopy. *J Struct Biol* 147:291-301.

Hölzl, H., B. Kapelari, et al. (2000). The regulatory complex of Drosophila melanogaster 26S proteasomes. Subunit composition and localization of a deubiquitylating enzyme. *J Cell Biol* 150:119-30.

Husnjak, K., S. Elsasser, et al. (2008). Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* 453:481-8.

Isasa, M., E. J. Katz, et al. (2010). Monoubiquitination of RPN10 regulates substrate recruitment to the proteasome. *Mol Cell* 38:733-45.

Ito, T., T. Chiba, et al. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98:4569-74.

Janke, C., M. M. Magiera, et al. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* 21:947-62.

Kajava, A. V. (2002). What curves alpha-solenoids? Evidence for an alpha-helical toroid structure of Rpn1 and Rpn2 proteins of the 26 S proteasome. *J Biol Chem* 277:49791-8.

Kaneko, T., J. Hamazaki, et al. (2009). Assembly pathway of the Mammalian proteasome base subcomplex is mediated by multiple specific chaperones. *Cell* 137:914-25.

Kipreos, E. T. and M. Pagano (2000). The F-box protein family. *Genome Biol* 1:REVIEWS3002.

Knop, M., K. Siegers, et al. (1999). Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* 15:963-72.

Kobe, B., T. Gleichmann, et al. (1999). Turn up the HEAT. *Structure* 7:R91-7.

Krogan, N. J., M. H. Lam, et al. (2004). Proteasome involvement in the repair of DNA double-strand breaks. *Mol Cell* 16:1027-34.

Lee, B. H., M. J. Lee, et al. (2010). Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* 467:179-84.

Leitner, A., T. Walzthoeni, et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics* 9:1634-49.

Leitner, A., T. Walzthoeni, et al. (2010). Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics* 9:1634-49.

Li, T., N. I. Naqvi, et al. (2000). Identification of a 26S proteasome-associated UCH in fission yeast. *Biochem Biophys Res Commun* 272:270-5.

Lowe, J., D. Stock, et al. (1995). Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. *Science* 268:533-9.

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol.* 152:36-51.

Mayor, T., J. Graumann, et al. (2007). Quantitative profiling of ubiquitylated proteins reveals proteasome substrates and the substrate repertoire influenced by the Rpn10 receptor pathway. *Mol Cell Proteomics* 6:1885-95.

Misaghi, S., P. J. Galardy, et al. (2005). Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate. *J Biol Chem* 280:1512-20.

Ng, J. M., W. Vermeulen, et al. (2003). A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. *Genes Dev* 17:1630-45.

Nickell, S., F. Beck, et al. (2009). Insights into the molecular architecture of the 26S proteasome. *Proc Natl Acad Sci U S A* 106:11943-7.

Nickell, S., F. Förster, et al. (2005). TOM toolbox Acquisition and Analysis for Electron Tomography. *J Struct Biol* 149:227-34.

Nijman, S. M., M. P. Luna-Vargas, et al. (2005). A genomic and functional inventory of deubiquitinating enzymes. *Cell* 123:773-86.

Penczek, P. A., J. Frank, et al. (2006a). A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *Journal of Structural Biology* 154:184-94.

Penczek, P. A., C. Yang, et al. (2006b). Estimation of variance in single-particle reconstruction using the bootstrap technique. *Journal of Structural Biology* 154:168-83.

Pettersen, E. F., T. D. Goddard, et al. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-12.

Pick, E., K. Hofmann, et al. (2009). PCI complexes: Beyond the proteasome, CSN, and eIF3 Troika. *Mol Cell* 35:260-4.

Qiu, X. B., S. Y. Ouyang, et al. (2006). hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37. *Embo J* 25:5742-53.

Rabl, J., D. M. Smith, et al. (2008). Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol Cell* 30:360-8.

Rinaldi, T., L. Hofmann, et al. (2008). Dissection of the carboxyl-terminal domain of the proteasomal subunit Rpn11 in maintenance of mitochondrial structure and function. *Mol Biol Cell* 19:1022-31.

Rinner, O., J. Seebacher, et al. (2008). Identification of cross-linked peptides from large sequence databases. *Nat Methods* 5:315-8.

Roelofs, J., S. Park, et al. (2009). Chaperone-mediated pathway of proteasome regulatory particle assembly. *Nature* 459:861-5.

Rosenthal, P. B. and R. Henderson (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology* 333:721-45.

Saeki, Y., E. Isono, et al. (2005). Preparation of ubiquitinated substrates by the PY motif-insertion method for monitoring 26S proteasome activity. *Methods Enzymol* 399:215-27.

Saeki, Y., E. A. Toh, et al. (2009). Multiple proteasome-interacting proteins assist the assembly of the yeast 19S regulatory particle. *Cell* 137:900-13.

Sanches, M., B. S. Alves, et al. (2007). The crystal structure of the human Mov34 MPN domain reveals a metal-free dimer. *J Mol Biol* 370:846-55.

Scheel, H. and K. Hofmann (2005). Prediction of a common structural scaffold for proteasome lid, COP9-signalosome and eIF3 complexes. *BMC Bioinformatics* 6:71.

Scheres, S. H., H. Gao, et al. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat Methods* 4:27-9.

Scheres, S. H., R. Nunez-Ramirez, et al. (2008). Image processing for electron microscopy single-particle analysis using XMIPP. *Nat Protoc* 3:977-90.

Schreiner, P., X. Chen, et al. (2008). Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* 453:548-52.

Semple, C. A. (2003). The comparative proteomics of ubiquitination in mouse. *Genome Res* 13:1389-94.

Smalle, J., J. Kurepa, et al. (2003). The pleiotropic role of the 26S proteasome subunit RPN10 in Arabidopsis growth and development supports a substrate-specific function in abscisic acid signaling. *Plant Cell* 15:965-80.

Smith, D. M., S. C. Chang, et al. (2007). Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. *Mol Cell* 27:731-44.

Sone, T., Y. Saeki, et al. (2004). Sem1p is a novel subunit of the 26 S proteasome from Saccharomyces cerevisiae. *J Biol Chem* 279:28807-16.

Stone, M., R. Hartmann-Petersen, et al. (2004). Uch2/Uch37 is the major deubiquitinating enzyme associated with the 26S proteasome in fission yeast. *J Mol Biol.* 344:697-706.

Szlanka, T., L. Haracska, et al. (2003). Deletion of proteasomal subunit S5a/Rpn10/p54 causes lethality, multiple mitotic defects and overexpression of proteasomal genes in Drosophila melanogaster. *J Cell Sci* 116:1023-33.

Tanaka, K. (2009). The proteasome: overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci* 85:12-36.

Taylor, K. A. and R. M. Glaeser (1974). Electron diffraction of frozen, hydrated protein crystals. *Science* 186:1036-7.

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

Toyoshima, C. and N. Unwin (1988). Contrast transfer for frozen-hydrated specimens: determination from pairs of defocused images. *Ultramicroscopy* 25:279-91.

Verma, R., L. Aravind, et al. (2002). Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* 298:611-5.

Verma, R., S. Chen, et al. (2000). Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol Biol Cell* 11:3425-39.

Voges, D., P. Zwickl, et al. (1999). The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem* 68:1015-68.

Wang, J., J. J. Song, et al. (2001). Nucleotide-dependent conformational changes in a protease-associated ATPase HsIU. *Structure* 9:1107-16.

Wang, Q., P. Young, et al. (2005). Structure of S5a bound to monoubiquitin provides a model for polyubiquitin recognition. *J Mol Biol* 348:727-39.

Wei, Z., Y. Xue, et al. (2006). Crystal structure of the C-terminal domain of S.cerevisiae eIF5. *J Mol Biol* 359:1-9.

www.wikipedia.org (2010). The Proteasome. http://en.wikipedia.org/wiki/Proteasome.

Yao, T. and R. E. Cohen (2002). A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* 419:403-7.

Yao, T., L. Song, et al. (2006). Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nat Cell Biol* 8:994-1002.

Yu, Y., D. M. Smith, et al. (2009). Interactions of PAN's C-termini with archaeal 20S proteasome and implications for the eukaryotic proteasome-ATPase interactions. *EMBO J*.

Zhang, F., M. Hu, et al. (2009a). Structural insights into the regulatory particle of the proteasome from Methanocaldococcus jannaschii. *Mol Cell* 34:473-84.

Zhang, F., Z. Wu, et al. (2009b). Mechanism of substrate unfolding and translocation by the regulatory particle of the proteasome from Methanocaldococcus jannaschii. *Mol Cell* 34:485-96.

Zhang, Z., N. Torii, et al. (2000). Structural and functional characterization of interaction between hepatitis B virus X protein and the proteasome complex. *J Biol Chem* 275:15157-65.

Zhu, J., P. A. Penczek, et al. (1997). Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. *J Struct Biol* 118:197-219.

Danksagung

This manuscript summarizes most of my scientific work of the recent years. It does not show, however, who was responsible for me being able to do this work. At first I was simply intrigued by the possibilities given by cryo electron microscopy, its ability to 'see life in progress'. I could not have imagined how difficult it might be to apply this technique in a meaningful way and to follow this project throughout the whole course of time. With this I would like to thank the people that have invested their time and energy to make this work come to an end. Above all I wish to thank my parents and sister, who always supported and helped me shape my decisions, even in difficult times.

Foremost I would like to thank Prof. Wolfgang Baumeister for providing the essence of scientific work: time, shelter and ideas. The outline has been further shaped mainly by the help of Stephan Nickell and Friedrich Förster, who focused my attention on the important details. Florian Beck has been invaluable in terms of exploring the possibilities provided. Eri Sakata came at the right time, sharing her valuable knowledge and experience in a selfless way. Thanks to Jürgen Plitzkos efforts in taming both microscopes and their company, the in the end highly efficient data acquisition was possible. Without Marek Cyrklaffs and Friedrich Frischknechts belief in me I could not have started on this work in the first place. Jana Sticht and Eva Gottwein have been great supervisors and mentors regarding which environment to look for to work on long-term project such as a PhD thesis.

Most of the time I was learning – from others – either new methods or how to organize work and life. Oana Mihalache and Günter Pfeiffer have taught me ways of working, without which the rest of the work could not be. My colleagues in this department have generated a working atmosphere based on helping one another and friendship, rather than competition.

I would like to thank in particular the former group members of Kay Grünewald for adopting me as one of their own. Again, this work would have turned out much less valuable or not valuable at all, if Juha Huiskonen, Lars-Anders Carlson, Ulrike Maurer and Iosune Ibiricu Urriza would not have given great examples of scientific work and life motivation in their very special way. Elizabeth Villa, Marion Jasnin and Alexander Rigort helped to judge with their experience thoughts on work and beyond. All FSCs have been discussed in detail with

Mathias Eibauer. Not forgotten are the great times with Thomas Löwe, Marco Gartmann and Petr Chlanda, who I hope to meet ever again in the future.

Some friends have shared the route of life with me for an immeasurable amount of time – their ideas and personalities have changed me to who I am today. These will not be forgotten and it makes me happy, thinking that we will share a long way still.

Katrin Klaus gave up more for this, than one ever would like to ask for.