# Mechanism and Structure of the Cdc48 ATPase Complex

A dissertation presented

by

Nicholas Bodnar

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

March 2018

© 2018 Nicholas Bodnar
All rights reserved.

Dissertation Advisor: Professor Tom Rapoport Nicholas Bodnar

Mechanism and Structure of the Cdc48 ATPase Complex

Abstract

The yeast ATPase Cdc48 and its mammalian homolog p97 are essential proteins that use the energy of ATP hydrolysis to perform work on polypeptide substrates. Cdc48 is a critical component of the ubiquitin-proteasome system and serves a general function as a "segregase," separating polyubiquitinated substrates from membranes or macromolecular complexes prior to their degradation by the proteasome. Cdc48 is involved in diverse protein quality control pathways operating at the endoplasmic reticulum, mitochondrion, ribosome, and chromatin, and a wide variety of substrates have been shown to require Cdc48 for proper handling by the proteasome. Despite this broad set of cellular roles, Cdc48's basic molecular mechanism has long been unclear. In this work, we use purified proteins to reconstitute Cdc48 substrate processing *in vitro* and interrogate its mechanism. We show that Cdc48, in cooperation with its essential cofactors Ufd1 and Npl4, recognizes and unfolds substrates that bear polyubiquitin chains of appropriate length and linkage. This unfolding reaction is mediated by substrate translocation through the central pore of the ATPase. Substrate release requires trimming of the polyubiquitin chain by a Cdc48-bound deubiquitinating enzyme, with concurrent passage of a portion of the chain through the central pore. Using cryo-electron microscopy and X-ray crystallography, we also provide the first high-resolution structures of the Cdc48/Ufd1/Npl4 complex. Unlike all previously characterized Cdc48 cofactors, Npl4 contains a zinc finger domain that anchors it to the top of Cdc48's ATPase rings and positions the cofactor directly over the central pore in an arrangement that is likely to facilitate substrate insertion. Together,

our results establish the essential structural and functional features of the Cdc48/Ufd1/Npl4 complex. These features are likely applicable to the handling of most or all of the polyubiquitinated substrates that depend on this ATPase in various cellular contexts.

**TABLE OF CONTENTS**

**Acknowledgements**

Five years under Tom's tutelage have had what I am certain will be a lifelong impact on me. Tom sports a unique brand of skepticism, a relentless curiosity, and a rigorous and methodical logic, and spending enough time around him has, I hope, rubbed some of those qualities off on me. I'm not sure how many hours we've spent sitting in my corner of the lab debating, looking through data, or sometimes just staring into space contemplating new models, but I feel that my brain now works in a way it never could when I first encountered Tom in my biochemistry tutorial in the first year of medical school. I didn't know how important my presentation on hemochromatosis in that class would turn out to be for my trajectory, but I'm so glad that Tom liked it well enough to let me rotate in his lab, and very grateful for the enormous investment he has made in me in the ensuing years.

I thank Alex Stein for supervising my rotation when I first joined the lab and introducing me to most of the skills I needed essentially from scratch. I try to emulate his patient and thorough explanatory style when I am teaching new people myself. Over the years here, many people have helped me with new techniques and discussions, including Stefan Schoebel, Eunyong Park, Benedikt Bauer, Tina Liu, Julie Chen, and Songyu Wang. Three excellent scientists, Neil Blok, Ryan Baldridge, and Thomas Guettler, have all recently left the lab but were constant fixtures throughout my PhD. These three all made big impacts on my development and I am very grateful to them. My co-grad students, Marco Catipovic and Robert Powers, provided the dose of daily hilarity necessary for the long grad school journey. Carol Sawyer and Lorna Fargo keep the lab humming, and our ability to spin out of control during even their brief absences is a testament to their importance here. My collaborator Kelly Kim made the cryo-EM project possible, and I am very grateful for the endless refinements and tweaks she tried out to

make the structure as good as it is. Thanks also to Tom Walz for his central role in those experiments. My DAC members, Steve Blacklow, Dan Finley, and Randy King, were excellent sources of advice over the years.

My parents, Andy Bodnar and Amy Pruitt, have listened to more long monologues about Cdc48 than they might have expected they would when I enrolled in medical school. Next year, they'll be hearing monologues of the 3rd-year medical student variety, which they are more used to. I thank them for their love, their support, and their realization that both types of monologue have their virtues. My brother Benjamin inspired my first science project, which was an attempt to booby trap his room with a Lego robot, so I view this thesis as mostly his doing. My parents-in-law and sisters-in-law have been great sources of support throughout grad school.

Most importantly, I can't begin to imagine what this experience would have been like without my wife Simin, who has been with me every step of the way and has made my life full and happy. Our professional paths diverged when I started my PhD and she started her clinical training, but our personal paths have become more intertwined every day. She is a brilliant doctor, an endlessly loving partner, and a constant source of inspiration for me.

Many others have helped me in large and small ways over the years, in ways I can probably never fully repay, and I am grateful to all of them.

# Chapter 1: Introduction

## 1.1 Overview of the ubiquitin-proteasome system

All cells must cope with the fundamental problem of waste disposal. Quality control systems that sense and eliminate a broad variety of unneeded or deleterious biomolecules are essential for viability. These waste molecules may be created as byproducts of ongoing cellular processes, generated by damaging effects of metabolic or environmental stress, or produced as consequences of the occasional mistakes that occur even in the tightly controlled processes of DNA replication, RNA transcription, and protein translation. It is now recognized that the failure of proper cellular waste disposal plays a central role in the pathogenesis of many human diseases.

In eukaryotic cells, protein quality control is managed by several systems with overlapping functions. One of the most important and best-studied such pathways is the ubiquitin-proteasome system (UPS). In its most general formulation, this system uses a set of several enzymes to covalently attach a small protein called ubiquitin to a protein targeted for disposal. Attachment of multiple ubiquitin moieties to the protein and to the originally appended ubiquitin produces a polyubiquitin signal that is recognized by a large complex called the 26S proteasome. The proteasome uses the energy of ATP hydrolysis to thread the target protein into a central chamber containing proteolytic sites, resulting in the degradation of the substrate into small peptides.

Since the discoveries of ubiquitin and the proteasome, this system has been shown to be intimately involved in essentially all facets of eukaryotic cellular life. In this section, I will describe the structure and function of the proteasome, the machinery required to conjugate ubiquitin to target proteins, and the system of proteasome accessory factors that ensure proper

functioning of the UPS. Cdc48/p97, the subject of this dissertation, belongs to the latter category and will subsequently be discussed in further detail.

*1.1.1 The 26S proteasome*

The existence of a non-lysosomal system for protein turnover was first appreciated in rabbit reticulocyte lysates through experiments that demonstrated the existence of a soluble, ATP-dependent protease that eliminates abnormal hemoglobin molecules[1]. Further work demonstrated this system's reliance on a small thermostable polypeptide, which was eventually identified as ubiquitin[2]. Since these seminal studies, the structure and mechanism of the 26S proteasome have been characterized in increasing detail.

The proteasome can be separated into two distinct assemblies, referred to as the 19S and 20S proteasomes[3]. Each of these components is itself a multisubunit complex (19 subunits in the 19S, and 28 subunits in the 20S)[4,5]. Generally, the 19S proteasome is responsible for recognition of substrates, removal of the ubiquitin chain, and ATP-dependent unfolding of substrates, which is accomplished by threading the polypeptides through a narrow central pore[6]. The 20S proteasome, on the other hand, houses the proteolytic sites[7]. The two particles are coaxially aligned such that the 19S central pore is contiguous with a central opening in the 20S and a substrate exiting from the 19S pore passes directly into the proteolytic chamber [8,9].

The 19S proteasome has several distinctive features that have been characterized in detail. First, it has multiple binding sites for ubiquitin chains. In *Saccharomyces cerevisiae*, three permanently incorporated (as opposed to transiently associated) 19S subunits - Rpn1, Rpn10, and Rpn13 - can participate in the recruitment of polyubiquitinated substrates to the 19S particle[10,11,12]. Interestingly, these sites cluster on one face of the 19S particle, suggesting that long or multiple polyubiquitin chains could be recognized cooperatively[10]. Second, the 19S particle contains a

heterohexameric set of ATPases (Rpt1-6 in yeast) that form a ring assembly with a central pore[9,13]. These proteins use ATP hydrolysis to translocate polypeptide substrates from their binding sites at the ubiquitin receptors through the central pore toward the 20S particle[14,15]. This translocation is dependent on aromatic residues on loops facing the central pore, which are presumed to interact with the substrate and move in response to ATP hydrolysis, although the mechanistic details of this process remain unclear[16]. Third, the 19S particle contains both intrinsic and transiently associated deubiquitinating enzymes (in yeast, these are Rpn11 and Ubp6, respectively) that can remove the polyubiquitin chain from arriving substrates[17-19]. Rpn11 is an essential subunit that is required for full passage of substrates through the 19S pore, as the tightly folded ubiquitin moieties cannot easily be passed through the pore and are instead recycled[20]. Fourth, the 19S ATPases harbor conserved C-terminal sequences, several of which include a hydrophobic-tyrosine-any residue (HbYX) motif at the extreme C terminus[21]. These motifs insert into grooves on the surface of the 20S core particle and mediate the interaction between the 19S and 20S proteasomes[22]. Each of these features - ubiquitin binding, ATP-mediated pore translocation, deubiquitination-dependent substrate release, and even the C-terminal HbYX motif - is recapitulated by the Cdc48 complex, as will be discussed in detail in this dissertation.

The 20S proteasome is a barrel-shaped peptidase consisting of four stacked heteroheptameric rings[23]. The outer rings are composed of alpha subunits and regulate access to the interior chamber, while the inner rings are composed of beta subunits, three of which (beta-1, beta-2, and beta-5) contain proteolytic sites[24]. These proteolytic sites display caspase-like, trypsin-like, and chymotrypsin-like specificities, respectively, resulting in an ability to cleave a diverse array of substrates[7]. Indeed, it is generally assumed that any polypeptide substrate

entering the proteolytic chamber is committed to degradation, while ATP-dependent unfolding and translocation into the chamber are the rate-limiting steps.

Together, these structural and functional features of the 19S and 20S proteasomes produce an integrated molecular machine that efficiently recognizes substrates, unfolds them via ATP hydrolysis, and degrades them. Over the last several decades, many substrates - both individual proteins and classes of degradative targets - have been identified as proteasome clients. The next section will describe the general mechanisms by which these clients are tagged with the polyubiquitin chains that signal their degradation.

*1.1.2 Ubiquitin and ubiquitin ligases*

Ubiquitin is a highly conserved and abundant protein that is 76 amino acids in length. Its stability and tightly folded nature render it an appropriate and recyclable tool for regulation of shorter-lived proteins[25]. The cascade of enzymatic events leading to the covalent attachment of ubiquitin to a substrate protein begins with an ATP-dependent activation step catalyzed by an E1 activating enzyme[26]. The E1 first adenylates the C-terminus of ubiquitin, resulting in a ubiquitin-AMP intermediate and releasing inorganic pyrophosphate. The ubiquitin C-terminus is then transferred to an active site cysteine in the E1, producing a thioester linkage and releasing AMP[27]. In the next step, ubiquitin is transferred to the active site cysteine of an E2 ubiquitin-conjugating enzyme, again forming a thioester through the carboxy terminus[28]. In the final step, an E3 ubiquitin ligase promotes the transfer of ubiquitin from the E2 to the target substrate[29]. E3 enzymes may form a third ubiquitin thioester as an intermediate in this process (as is the case for HECT domain E3 ligases) or may simply position the E2 and substrate in an appropriate manner for ubiquitin transfer (as is the case for RING domain ligases)[30]. The product of ubiquitin ligation is, in most cases, an isopeptide bond between the C-terminus of ubiquitin and a lysine side chain

of the target polypeptide, although conjugation to serine, threonine, cysteine, and the substrate amino terminus have been described[31,32]. The diversity and number of ubiquitination enzymes increases with each step in the cascade. E1 enzymes facilitate a general activation reaction for which only one such protein is necessary in the yeast genome. Yeast encode 11 E2 ubiquitin-conjugating enzymes and dozens of E3 enzymes, which are often composed of multiple subunits and recognize specific substrate sequences, biochemical characteristics, or post-translational modifications[33].

Ubiquitin's sequence includes seven lysines at positions 6, 11, 27, 29, 33, 48, and 63. These lysines are critical for the assembly of polyubiquitin chains[34]. Many E3 enzymes catalyze both the attachment of the initial ubiquitin to the substrate and the addition of further ubiquitin molecules to the growing chain. Isopeptide bond formation between the C terminus of a "distal" ubiquitin moiety and one of the lysines in the "proximal" ubiquitin moiety produces a chain defined by its linkage type. For example, K48-linked chains of at least 4 ubiquitin moieties constitute the canonical proteasome degradation signal[35]. Other linkage types participate in a variety of cellular processes such as vacuolar protein sorting or non-degradative histone modification[36,37].

### *1.1.3 Nuances of UPS substrate handling*

The canonical ubiquitination and degradation reactions described above form an accurate model for the processing of many UPS substrates. However, it is increasingly appreciated that the proteasome is not capable of degrading all substrates that are conjugated to ubiquitin. To cope with substrates that present certain types of difficulties, the UPS has evolved a set of auxiliary factors, including the AAA ATPase Cdc48.

The 19S proteasome threads its substrates processively through its central pore. To initiate this threading process, a target polypeptide must contain a loosely folded region that can insert into the pore and be engaged by the ATPases[38,39]. Many proteins that undergo regulated degradation by the UPS, such as cyclins, include such intrinsically disordered regions, suggesting that these substrates have evolved to be easily processed by the proteasome[40]. The flexible initiation segment must be positioned relative to the ubiquitination site such that polyubiquitin binding to the 19S ubiquitin receptors places the initiation segment in spatial proximity to the central pore[41]. This geometric requirement poses relatively strict constraints on the types of proteins that can be considered "natural" proteasome substrates. However, the proteasome must be able to handle a broad variety of quality control substrates, including misfolded or damaged polypeptides derived from genes that have not specifically evolved to promote proteasome processing. Such polypeptides, when ubiquitinated, may not present an appropriate initiation segment.

One set of proteins that may assist in the degradation of substrates that are not easily processed by the 26S proteasome comprises the so-called "shuttling factors." In yeast, three proteins have been identified as members of this class: Rad23, Dsk2, and Ddi1[33]. These proteins each contain a UBL (ubiquitin-like) domain and at least one UBA (ubiquitin-associated) domain. The UBL domain binds to the same sites on the 19S proteasome as polyubiquitin[42], while the UBA domain binds to ubiquitin itself[43]. Thus, these shuttling factors can form bridges between ubiquitinated substrates and the proteasome. However, as the proteasome already contains multiple intrinsic ubiquitin receptors, the shuttling factors likely play some additional role beyond simple delivery[10]. One possibility is that UBL-UBA proteins are required for substrates in which the ubiquitin attachment site and any potential initiation regions are geometrically

related such that the flexible segment does not approach the central ATPase pore when the ubiquitin chain is bound by the proteasome's intrinsic ubiquitin receptors[44]. The addition of a moderately flexible UBL-UBA protein as a bridge between the proteasome-binding site and the ubiquitin chain might, for some substrates, allow access of initiation regions to the ATPases. However, this mechanism would only apply to substrates that already include appropriate flexible segments somewhere in their sequence.

A more difficult class of substrates comprises those that do not expose any flexible segment that could be engaged by the 19S proteasome. Substrates of this type might be membrane proteins without significant cytosolic flexible regions, constituents of tightly associated complexes that are difficult to disassemble, or simply well-folded proteins lacking disordered segments. In these cases, the action of the Cdc48 ATPase is often required[45,46]. Cdc48's general function in the UPS is to preprocess substrates destined for proteasomal degradation[47]. Although several Cdc48 targets that do not proceed to degradation have been identified[48], it appears that the majority are eventual 26S proteasome clients, and that all require modification with ubiquitin. Cdc48 is commonly described as a "segregase," because it often liberates target polypeptides from their original environments, in which they are not accessible to the proteasome[49]. It has also been proposed that Cdc48 is an "unfoldase" that can generate flexible segments in substrates that would otherwise be poor proteasome targets[50]. Cdc48 is essential for viability[51], confirming that these activities are central to proper functioning of the ubiquitin-proteasome system. In the next section, I will describe several of Cdc48's specific functions and discuss general themes of its action.

## 1.2 Functions of Cdc48

### *1.2.1 Early work on Cdc48/p97*

The gene encoding Cdc48 was first isolated in *S. cerevisiae* by a screen for mutants defective in cell division cycle progression[52]. The gene was sequenced in the early 1990s, shown to be essential, and identified as a homolog of mammalian p97/VCP, which was already known to exhibit similarity to the N-ethylmaleimide-sensitive fusion protein (NSF; Sec18 in yeast) [51]. The basic structure of the ATPase was evident even in early studies, in which it was identified as an abundant species migrating slightly more slowly than the 20S proteasome in cell fractionation experiments[53]. Electron microscopy of these fractions revealed a ring-shaped hexameric assembly that was shown to exhibit ATPase activity[53]. Higher resolution structural studies in the early 2000s demonstrated that Cdc48 is a type II AAA ATPase, meaning that it contains two stacked hexameric rings of ATPase domains, referred to as D1 and D2 (Figure 1.1)[54]. In addition to the ATPase rings, Cdc48 also contains a conserved N-terminal domain (N domain) and a C-terminal tail that is not visible in structures to date[55].

In initial studies, Cdc48 was assumed to exert functions similar to those of NSF, which at the time was a better characterized ATPase and was known to be involved in vesicle fusion[51]. However, subsequent work uncovered a role for Cdc48 in the UPS, a connection which was first appreciated in the case of the ubiquitin-fusion degradation (UFD) pathway[56]. This system was discovered through screens in *S. cerevisiae* for mutants defective in the degradation of model substrates with in-frame, N-terminal fusions of ubiquitin[57]. Although Cdc48, an essential gene, was not isolated in the original screen, it was soon shown that it interacts with multiple components of the UFD pathway and is required for UFD substrate elimination[56]. From this point

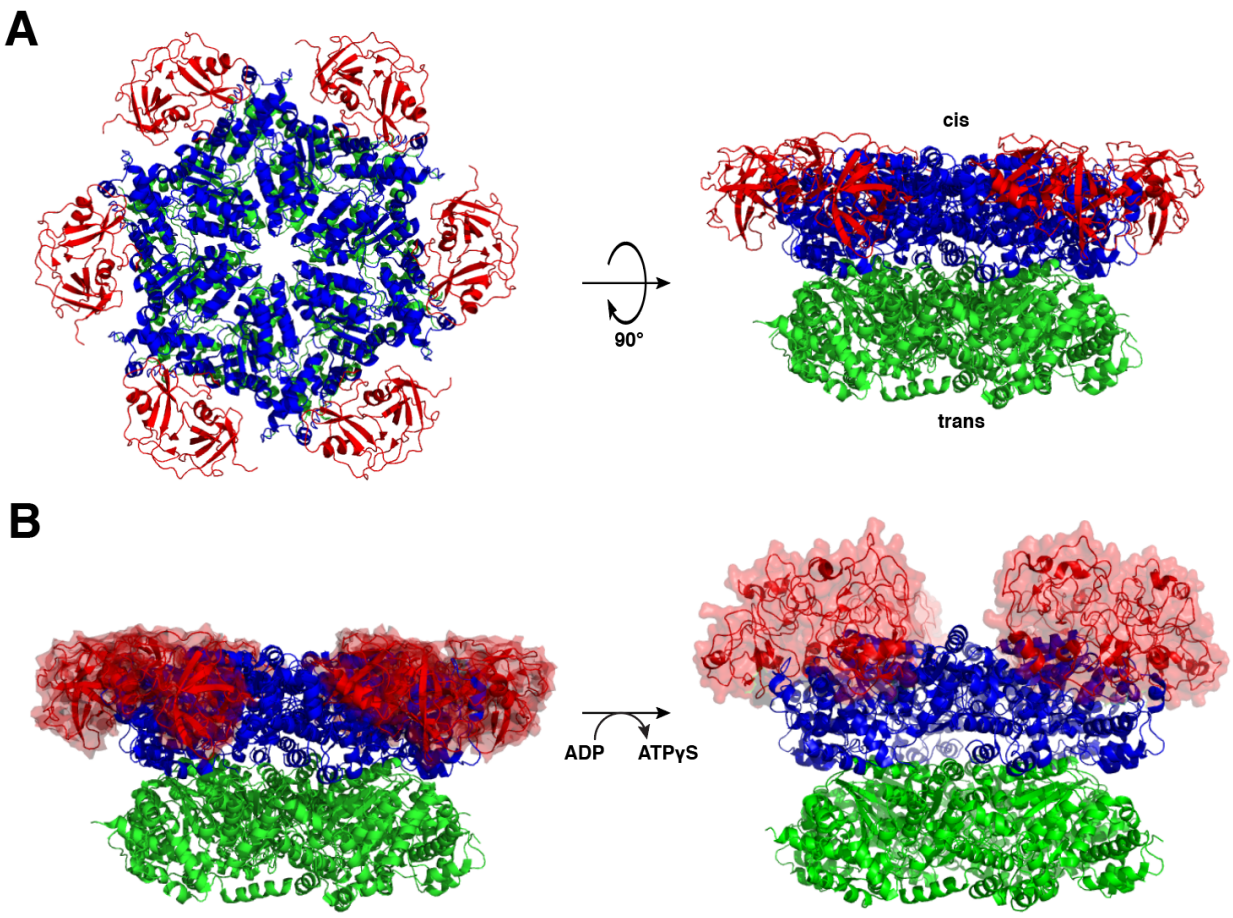



**Figure 1.1: Architecture of the Cdc48/p97 ATPase.** (A) Cdc48 is a homohexamer, and each monomer comprises an N-terminal (N) domain (red) and two AAA ATPase domains: D1 (blue) and D2 (green). The N-terminal (D1) side of the central pore is referred to as the cis side, and the C-terminal (D2) side as the trans side. (B) ATP binding produces an upward rotation of the N domains into a so-called "up conformation", in which they are positioned above the plane of the D1 ring. Left, ADP-bound state (PDB code 5FTK); right, ATPγS-bound state (PDB code 5FTN). PDB, Protein Data Bank. This figure has been previously published in: Bodnar NO and Rapoport TA. "Toward an understanding of the Cdc48/p97 ATPase." *F1000Research* 2017, 6(F1000 Faculty Rev):1318.

onward, identification of further cellular roles, binding partners, and structural aspects of Cdc48 have provided increasing insight and established this ATPase as a central node in the UPS.

*1.2.2 Cdc48 cofactors*

Before discussion of Cdc48's diverse cellular roles, it is important to note that this ATPase cooperates with a large number of cofactors that recruit substrates, regulate Cdc48's subcellular localization, modify client proteins' post-translational modifications, and in some cases - as will be described in this thesis - are essential for Cdc48's basic biochemical activity[58]. These proteins, which number around 30 for human p97, are usually classified as either substrate-recruiting cofactors, substrate-processing cofactors, or regulatory cofactors. They can also be classified according to the type of binding site they use to interact with Cdc48/p97 or according to the cellular pathways in which they operate. I will not attempt an exhaustive review here, but will instead highlight important themes of Cdc48 cofactor function.

All known Cdc48 cofactors bind either to the N domain or to the flexible C-terminal tail[58]. A hydrophobic cleft is found between the two subdomains of the N domain, and this cleft is the binding site for most cofactors that interact with this part of the ATPase[59]. A variety of motifs that interact with this site, including UBX (ubiquitin regulatory X), UBX-like, VIM (valosin-interacting motif), and VBM (valosin-binding motif) have been identified[58]. Likewise, several types of cofactor domain, including the PUB and PUL domains, can interact with the C-terminal tail, although all appear to require the HbYX motif found in the final three residues of the Cdc48 sequence[60,61]. The presence of six Cdc48 monomers per complex theoretically allows for modular assembly of a wide variety of complexes containing multiple cofactors, and several such complexes have indeed been demonstrated experimentally[62].

Perhaps the simplest role for Cdc48 cofactors is that of recruiting the ATPase to particular cellular locations, such as the endoplasmic reticulum or mitochondrion[58,63]. This function is accomplished in a variety of ways, as will be detailed below, but requires no enzymatic activity and does not directly influence Cdc48's action on its substrates. A slightly more complex function is the recruitment of Cdc48 to a cellular location by a cofactor that also binds substrates, which is likely the case for the multiple proteins that have both a UBX domain and a UBA domain[64]. The presence of both of these domains in a single cofactor is partially analogous to the presence of UBL and UBA domains in the shuttling factors discussed in section 1.1.3, as this domain structure would appear to facilitate bridging between the ATPase and its substrates.

Some cofactors have direct effects on the ATPase itself, either modulating its ATP hydrolysis activity or its structure. The first cofactor that was shown to modify the intrinsic ATPase rate of Cdc48/p97 was p47 (Shp1 in yeast), which reduces the activity of purified rat p97 by 85%[65]. A more exotic function has been ascribed to the mammalian UBX protein ASPL, which disassembles p97 hexamers and sequesters the ATPase in inactive heterotetramers of 2 p97 molecules and 2 ASPL molecules[66]. Whether a protein with a similar function exists in yeast is unclear.

Still other cofactors directly modify substrates. Cdc48/p97 associates with deubiquitinases that can trim or remove the polyubiquitin chain attached to the substrate[67]; the function of one such DUB, Otu1, will be discussed in detail in chapter 2. It appears that, as with the 19S proteasome, deubiquitination can either promote or inhibit substrate processing[68]. Other cofactors are ubiquitin ligases, such as the UFD pathway member Ufd2. This protein has been termed an "E4" ligase because it appends additional ubiquitin moieties to a pre-existing chain.

The action of Ufd2 is required for efficient processing of only a subset of Cdc48 targets, but the substrate characteristics that produce a requirement for Ufd2 are not clear[69]. Another form of substrate modification is deglycosylation, which is particularly important in the context of endoplasmic reticulum-associated degradation (see section 1.2.3)[70].

Perhaps the most important individual Cdc48 cofactor is the heterodimer of Ufd1 and Npl4 (referred to as UN in much of this dissertation). Like Cdc48, but unlike other cofactors, both Ufd1 and Npl4 are essential for viability[71]. Both proteins can be considered substrate-recruiting cofactors, as each contains binding sites for ubiquitin and for Cdc48[72]. With isolated exceptions, the two cofactors are found in complex with each other[71]. The UN complex and the p47/Shp1 cofactor bind to Cdc48 in a mutually exclusive manner[73], but many Cdc48 cofactors can associate with the ATPase in the presence of UN. Indeed, in some cases, UN binding is required for recruitment of additional cofactors in a hierarchical manner[62]. UN appears to be required in nearly all cases in which polyubiquitinated substrates must be handled by Cdc48. Chapter 2 will detail the mechanistic role of UN in substrate processing by the ATPase, while Chapter 3 will discuss the structure of the Cdc48/UN complex. These studies illuminate a general mechanism of substrate handling that is likely applied in a substantially similar manner in many of Cdc48's diverse cellular functions, several of which will now be discussed.

*1.2.3 Endoplasmic reticulum-associated degradation*

Approximately one third of the cellular proteome is directed through the secretory pathway, which handles proteins that are exported from the cell or embedded in membranes[74]. The critical initial step in biogenesis of these proteins is either co-translational or post-translational translocation into the endoplasmic reticulum (ER), a protein sorting and quality control hub that ensures appropriate folding and modification of secretory cargo and membrane

proteins and directs these clients to further steps in the pathway[75]. Many of the proteins handled by the ER are naturally fastidious in their folding process; these include proteins with disulfide bonds, multiple transmembrane segments that are prone to misincorporation, complex structures that require the cooperation of several ER-luminal chaperones for proper folding, or obligate post-translational modifications such as N-linked glycosylation or glycophosphatidylinositol anchor addition[76]. Inevitably, some such proteins will terminally fail to fold appropriately, and these waste molecules must be targeted for elimination in a process termed ER-associated degradation (ERAD).

The mechanisms by which ERAD substrates are recognized are only partially characterized. One well-studied such mechanism is the so-called "glycan timer"[77]. This pathway operates via the initial attachment of a $Glc_3Man_9GlcNAc_2$ glycan to certain asparagine residues on substrates that are entering the ER through the Sec61 protein translocation channel[78]. Removal of two glucose residues by ER-resident glucosidases trims the glycan to a species that is recognized by the luminal chaperones calnexin and calreticulin, which assist in client folding[79]. Removal of the final glucose releases substrates from the chaperones, at which point they may either have completed folding or be reglucosylated and pass through another folding attempt cycle[80]. This cycle continues until the substrate is properly folded or until mannosidases trim the glycan to a species that is no longer recognized by chaperones and becomes a signal for entry into the ERAD pathway[81]. A luminal lectin, Yos9 in yeast, is important for recognition of this terminal glycan, but it appears that direct recognition of a misfolded conformation by ERAD machinery can also be involved[82,83]. The latter step awaits further characterization. Importantly, the glycan timer mechanism does not apply to non-glycoprotein substrates, and there are

probably many other substrate characteristics, such as improper disulfide bonds or absence of binding partners, that can trigger ERAD.

By contrast with substrate recognition mechanisms, substrate disposal mechanisms in ERAD are relatively well characterized. In yeast, ERAD can be generally subdivided into three pathways termed ERAD-L, ERAD-M, and ERAD-C, according to the luminal, membrane, or cytosolic locations of their defects[84]. Typically, ERAD-C substrates are routed to an E3 ubiquitin ligase called Doa10, while ERAD-L and ERAD-M substrates are routed to an analogous E3 called Hrd1[84]. Both ligases are ER membrane proteins with cytosolic RING domains. The mechanism of Hrd1 action has been characterized in some detail. This ligase cooperates with several additional ER membrane proteins to recognize substrates and maintain its own stability in the membrane, but it is the only essential component for the basic reactions that remove substrates from the ER[85]. Hrd1 can recognize misfolded substrate segments and functions as a retrotranslocation channel, moving portions of ERAD-L and ERAD-M targets across the membrane from the lumen to the cytosol[86]. This activity requires autoubiquitination of the ligase[86]. Once substrates have emerged on the cytosolic face of the ER, they are modified with K48-linked polyubiquitin chains by Hrd1[83,86].

Polyubiquitinated substrates on the cytosolic surface of the ER are recognized by Cdc48 in a manner that requires the UN cofactor[87,88]. This recognition is aided by the concentration of a population of Cdc48 at the membrane via binding to anchors such as the UBX-containing protein Ubx2, an ER transmembrane protein[89]. Upon recognition of the target protein, Cdc48 extracts the polypeptide from the membrane using the energy of ATP hydrolysis[83]. This serves to move the substrate completely into the cytosol, where it is eventually degraded by the 26S proteasome. The steps between Cdc48-mediated extraction and 26S-mediated degradation are poorly

characterized, but can involve glycan removal by the Cdc48-associated deglycosylase Png1, recognition by UBL-UBA shuttling factors, or binding by "holdase" complexes such as Bag6, which recognize exposed hydrophobic segments and may serve to keep substrates soluble between extraction and degradation[90-92]. Although there are multiple upstream substrate recognition pathways and potential downstream fates for Cdc48's ERAD targets, the actual extraction step requires the ATPase in all examples described so far.

*1.2.4 Chromatin-associated and nuclear processes*

The presence of a significant population of Cdc48 in the nucleus indicates that at least some of its expected substrates are found in this compartment[93]. The UPS has a well-documented role in the nucleus, but the requirement for Cdc48 in upstream processing of diverse nuclear UPS targets has only been appreciated more recently[94]. Here, I will highlight several of these targets, which correspond to fundamental activities including the initiation and termination of DNA replication, the DNA damage response, and RNA transcription.

The licensing factor Cdt1 plays a critical role in the initiation of DNA replication by recruiting replisome components to origins of replication. Cdt1 degradation is important for the prevention of multiple initiation events at a single origin, and this degradation is effected by the ubiquitination and proteasomal destruction of Cdt1 after a single replisome has fired from a given origin[95]. Cdt1 can be ubiquitinated but is not removed from chromatin or degraded in the absence of functional Cdc48/p97 and UN[96]. This recognition is enhanced by a UBX domain-containing cofactor, UBXN3/FAF-1, which binds to both Cdt1 and Cdc48/p97[97]. The exact reason for the requirement for Cdc48-mediated extraction is unclear in this case, but may derive from the very strong affinity of the Cdt1 PIP box for its chromatin-associated binding partner, PCNA[98]. Selective removal of the ubiquitinated component of this high-affinity complex, Cdt1,

may be dependent on the strong segregase activity of Cdc48. Alternatively, the relatively bulky 26S proteasome may be unable to access Cdt1 efficiently when the target is embedded in a complex with other replication factors, thus requiring the smaller Cdc48/UN complex to perform an initial extraction step.

Cdc48 plays a similar role in the termination of replication in that it is essential for timely removal of the replication helicase from DNA. The observation that blocking polyubiquitination resulted in abnormal retention of helicases on chromatin was followed by identification of the crucial polyubiquitinated component, Mcm7}[99-101]. This subunit of the Mcm2-7 replicative helicase is decorated with K48-linked polyubiquitin chains by an Skp1-Cul1-Fbox (SCF) E3 ligase complex, $SCF^{Dia2}$, as replication is completed, although the exact events triggering this ubiquitination remain to be dissected[99]. Disassembly of the helicase does not require proteolytic activity of the 26S proteasome, but does require the action of Cdc48/p97. Mcm2-7 is itself a complex of AAA ATPases that in its active form surrounds a single DNA strand[102,103], and removal of this stable complex may require an unfolding activity that only Cdc48/p97 is equipped to provide.

A third critical chromatin-associated process coordinated in part by Cdc48/p97 is the DNA damage response. Double-stranded DNA breaks are recognized by the Ku70/Ku80 heterodimer, which forms a ring structure and threads the broken DNA end through a central channel[104]. Assembly of the Ku complex onto DNA forms a scaffold that recruits additional DNA repair pathway components. After DNA repair, the Ku complex, like Mcm2-7, remains topologically trapped on DNA[105]. Subsequent ubiquitination of the Ku80 subunit by an SCF E3 ligase, $SCF^{Fbxl12}$, triggers Cdc48/p97- and UN-dependent extraction of Ku80 from DNA and disassembly of the Ku heterodimer, allowing normal replication, transcription, and association of

other DNA-binding proteins to occur at the former break site[106,107]. Additional DSB repair proteins including L3MBTL1 and KAP1 are also ubiquitination targets with evidence for direct Cdc48/p97-mediated extraction[108,109]. Overall, the emerging picture is that double-stranded breaks result in a highly coordinated series of recruitment and disassembly events involving numerous repair factors, and that a significant subset of these are ubiquitinated and removed from chromatin through the action of Cdc48[110].

Surveillance of the RNA transcription process is yet another critical nuclear function for the Cdc48 ATPase. During the transcription process, RNA polymerase II (Pol II) may stall as it encounters a variety of obstacles, including other chromatin-associated proteins such as nucleosomes, inhibitory secondary DNA structures, or sites of DNA damage[111]. The latter, in the form of ultraviolet-induced lesions, is often used as a model for Pol II stalling[111]. Stalling produces ubiquitination of the polymerase itself, which is subsequently extracted from chromatin and degraded in a manner that requires the proteasome, Cdc48, UN, and two UBX-containing proteins, Ubx4 and Ubx5 in yeast[112]. The chromatin remodeling protein INO80 has also been implicated in Pol II extraction, likely because it is required to shift nucleosomes that are sterically hindering access to the stalled polymerase[113].

Apart from the extraction of chromatin-bound proteins discussed above, a general role for Cdc48 in the elimination of defective nuclear proteins is also emerging. The E3 ligase San1 is a general nuclear quality control factor that, like the Hrd1 ERAD ligase, directly recognizes a variety of misfolded substrates, binding to exposed disordered regions[114,115]. Although Cdc48 appears to be required for all Hrd1 substrates, potentially because the presence of the ER membrane impedes proteasomal access to targets, this is not the case for San1 substrates. Instead, substrates' dependence on Cdc48 correlates with their insolubility, with model

substrates that are more prone to aggregation exhibiting greater reliance on Cdc48[116]. Cdc48 may be required to disassemble nuclear inclusions of these model substrates, although these experimental conditions of substrate overexpression do not necessarily mimic scenarios likely to be encountered by Cdc48 in normally growing cells. However, certain disease states, most prominently a subset of neurodegenerative disorders, are associated with the accumulation of nuclear inclusions[117], and under these circumstances the putative aggregate disassembly activity of Cdc48/p97 might be important for limiting nuclear damage.

The diverse clients of Cdc48 in the nucleus thus include both spontaneous quality control substrates, i.e. damaged or misfolded polypeptides in the nuclear compartment, and substrates that undergo a defined program of ubiquitination and extraction as part of their normal activities. The unifying characteristic of all of these targets appears to be that the 26S proteasome is not competent to unfold or degrade them on its own; rather, the specialized activity of Cdc48 is required to remove them from inaccessible locations or to unfold them to an extent that renders them appropriate 26S substrates.

*1.2.5 Mitochondrion-associated degradation*

A critical function of ubiquitination in mitochondrial quality control is the induction of mitophagy, the autophagic engulfment and destruction of damaged mitochondria. This process is orchestrated by the Parkinson's disease-associated proteins PINK1, a serine/threonine kinase, and Parkin, an E3 ubiquitin ligase with diverse functions[118]. PINK1, which senses defects in mitochondrial import and maintenance of membrane potential, accumulates on stressed or damaged mitochondria and phosphorylates both Parkin and ubiquitin itself, both at residue S65[119]. These events lead to ubiquitination of a variety of mitochondrial surface proteins, eventually triggering recognition by autophagy receptors.

In addition to full engulfment of entire mitochondria in a ubiquitin-dependent manner, recent studies have revealed a more precise degradation system anchored by Cdc48/p97 that selectively eliminates individual polyubiquitinated mitochondrial proteins. One such degradation reaction participates directly in mitophagy: the removal of two GTPases, Mfn1 and Mfn2, which promote mitochondrial fusion, is required for appropriate separation of damaged mitochondria from the remainder of the mitochondrial network[120,121]. After Parkin-mediated ubiquitination of these GTPases, Cdc48/p97 extracts them for subsequent proteasomal degradation, thereby preventing them from re-fusing damaged mitochondria.

Ubiquitin-dependent quality control operates on surface proteins even on mitochondria that are not terminally damaged and en route to mitophagy. Several outer mitochondrial membrane (OMM) proteins are naturally short-lived and proceed to degradation only after Cdc48/p97-dependent extraction[122]. This appears to be a component of the normal functioning of these proteins, as the wild-type proteins exhibit this behavior even in the absence of mitochondrial stress[123]. Cdc48/p97 also plays a role in the mitochondrial stress response distinct from its role in mitophagy via the cofactor Vms1. Uniquely, Vms1 seems to replace Ufd1 as a binding partner for Npl4, such that the relevant ATPase complex for the "mitochondrion-associated degradation" pathway is Cdc48-Vms1-Npl4[124]. This complex is recruited to the OMM via Vms1 association with a unique lipid, ergosterol peroxide, which is formed upon mitochondrial stress[125]. One possibility is that ubiquitination of OMM proteins produces a Cdc48-dependent quality control response that, in conjunction with systems operating at the inner membrane and matrix, can rescue mitochondria from full autophagic destruction if they are not damaged beyond repair. This intermediate rescue response would be analogous to the various levels of the unfolded protein response operating at the ER, which first attempts to relieve

protein-folding stress through a variety of mechanisms including the induction of ERAD, and only after failing to relieve severe stress proceeds to initiation of apoptosis[126].

*1.2.6 Ribosomal quality control*

The organelle-specific functions of Cdc48 discussed above are complemented by activities operating at the level of individual proteins and protein complexes in the cytosol. One such recently characterized pathway is the ribosomal quality control, or RQC, system. This system complements surveillance mechanisms that ensure appropriate ribosome assembly[127] and accurate mRNA sequences[128] by handling the nascent protein products of translation. Translating ribosomes may stall due to the absence of a stop codon or the presence of a premature stop codon, inhibitory RNA secondary structure, or low levels of a particular aminoacylated tRNA, among other potential causes[129]. Stalled ribosomes are subject to a quality control system that splits them into 60S and 40S subunits, with any nascent chain that has been translated remaining in the ribosomal exit tunnel in the 60S subunit[130]. The "degron" in such cases is the exposed peptidyl-tRNA, which is normally not accessible to the cytosol in the context of the fully assembled ribosome[131]. Recognition of this abnormal feature is accomplished by the Rqc2 and Ltn1 proteins in yeast[132]. Ltn1 is an E3 ubiquitin ligase that wraps from the underside of the 60S subunit, where the tRNA is exposed, around to the exit tunnel, where the nascent chain to be ubiquitinated is exposed[133,134]. Ubiquitination of this chain results in recruitment of the Cdc48/UN complex[135,136]. The pathway-specific Cdc48 cofactor Rqc1, which binds both to ribosomes and to Cdc48, aids in this recruitment event[136]. A unique feature of this pathway is that the Rqc2 protein specifically recruits alanyl and threonyl tRNAs to the 60S ribosome and appends C-terminal alanine and threonine (CAT) tails to the nascent chain[134]. This step may serve several purposes, including extension of the nascent chain from the exit tunnel to expose additional lysines for

ubiquitination and, more generally, induction of the heat shock response to augment cytosolic protein quality control[134].

The requirement for Cdc48 in this process may derive from several features of the substrate. One possibility is that some nascent chains are not long enough to provide the ~20-residue initiation sequence preferred by the proteasome. Another is that ubiquitin chains that are too close to the exit tunnel of the ribosome are sterically inaccessible to the proteasome, as may also be the case with certain chromatin-associated substrates. A third is that interactions between the nascent chain and the exit tunnel are sufficient to counteract the potential extraction force provided by the 19S proteasome, necessitating the action of Cdc48, which presumably can exert more force. Finally, the presence of the tRNA at the end of the nascent chain may inhibit full translocation of the nascent chain through the exit tunnel. Even with the action of Cdc48, hydrolysis of the tRNA-nascent chain bond may be required for completion of extraction, and this step is not yet well characterized[129].

*1.2.7 Other functions*

Cdc48/p97 has been implicated in a large number of additional processes, many of which are not as well characterized as the above described functions. Here, I will highlight several additional activities that may be relevant to the construction of a general model for Cdc48/p97 substrate handling.

Stress granules are collections of tightly associated mRNAs and RNA binding proteins that form under conditions of stalled or limited translation initiation[137]. These granules contain non-translating RNAs and some initiation factors, and their formation is reversible, suggesting that they represent storage intermediates for RNAs that are intended to re-enter the active pool[137]. Stress granules can be induced to form and disassemble by heat shock and subsequent cooling,

respectively, and the disassembly process has been shown to require the action of Cdc48[138]. Although a specific role for ubiquitination in this process has not yet been described, one possibility is that the action of Cdc48 is needed to break apart intermolecular interactions between ribonucleoproteins, which are mediated by "aggregation domains" and can be extensive[139].

Cdc48 has many individual protein substrates that are not embedded in organelles or aggregates; these include important signaling and metabolic proteins such as the NFκB regulator IκBα[140], the transcription factor HIF1α[64], the enzyme glutamine synthetase[141], and the cell cycle commitment regulator Far1[142]. Many of these are constituents of multi-protein complexes, suggesting that efficient proteasomal degradation of this type of substrate often depends on extraction of the individual target polypeptide, although the 26S proteasome can handle some substrates of this class alone[143]. The accessibility of an initiation site or the thermodynamic stability of the substrate protein may determine which multi-protein complexes require the additional action of Cdc48. Consistent with this hypothesis, degradation of antibody-bound viral capsids mediated by the intracellular IgG receptor and ubiquitin ligase TRIM21 requires p97, while degradation of free intracellular IgG Fc does not, suggesting that the large and stable capsid assembly produces a requirement for p97's disassembly activity[144].

Systematic studies of the characteristics that render substrates Cdc48/p97-dependent will be necessary to distinguish among the various possibilities discussed in the previous sections. Such studies will be less straightforward in the context of membrane proteins or components of protein complexes, but may be more feasible for free polypeptide substrates. Indeed, *in vivo*, the addition of a putative initiator sequence to a UFD substrate allows it to bypass Cdc48 in a manner that depends on the length of the unstructured region[50]. Further studies along these lines,

incorporating *in vitro* work and substrates of varying folding stability, may shed light on this question.

*1.2.8 General principles*

The various Cdc48 functions discussed above give rise to a set of principles of Cdc48 action that form a basis for further investigation of its mechanism. These are:

i) *Cdc48 is selective for K48-linked polyubiquitin chains*. In each case in which linkage type has been evaluated, the extraction function of Cdc48 requires the same type of chains that form the canonical proteasome degradation signal.

ii) *Cdc48 exerts force on its substrates*. The segregation activities of Cdc48 strongly imply the existence of a force-generating mechanism. This force generation might also involve substrate unfolding, but the *in vivo* studies described above do not directly demonstrate such unfolding.

iii) *The UN complex is required for most Cdc48 activity*. With the exception of Vms1-dependent mitochondrial functions, Cdc48 requires UN when it is acting on polyubiquitinated substrates. Other cofactors appear to have more specialized roles in one or a few pathways.

iv) *Substrates are directed to downstream components of the UPS after Cdc48 action*. The fact that most substrates are eventually degraded by the 26S proteasome indicates that a transfer mechanism or mechanisms from Cdc48 to the proteasome must exist; the first step would be release of the substrate from Cdc48.

These principles suggest that successful mechanistic investigation of Cdc48 substrate processing by *in vitro* reconstitution requires a model substrate with a K48-linked polyubiquitin chain or chains, the presence of the UN cofactor, and assays to monitor substrate unfolding and release. The development of this reconstitution is discussed in Chapter 2.

## 1.3 Structure of Cdc48 and its cofactors

### *1.3.1 The AAA ATPase family*

The AAA+ (ATPases associated with a variety of cellular activities) family is a large group of ATP-hydrolyzing enzymes that form a subgroup of the larger P-loop NTPase family[145]. This family is very diverse, but contains a core fold of four or five beta strands (β1-4 or 1-5) that are "sandwiched" between two sets of two alpha-helices (α1-4)[146] (Figure 1.2). The eponymous P-loop is a phosphate-binding motif anchored by a conserved lysine[147]. Within the P-loop NTPase group, the AAA+ proteins are part of the ASCE (additional strand, catalytic E) group[145]. This group is named for an additional beta strand that is inserted in the core set of strands in the P-loop NTPase fold, and for the conserved glutamate that is involved in nucleotide hydrolysis.

Several essential features characterize the AAA+ ATPase family. In most cases, AAA+ proteins associate to form hexameric ring structures with either a single ring (Type I) or two stacked rings (Type II). Each individual AAA domain contains two subdomains. The N-terminal subdomain contains the core P-loop NTPase fold, while the C-terminal subdomain is formed from a group of four helices and is unique to the AAA+ family (Figure 1.2) [145]. In the oligomeric assembly, the N-terminal (large) subdomain of one monomer is in contact with the C-terminal (small) subdomain of the neighboring monomer[148,149]. Each unit of one large and one small subdomain is thought to move as a rigid body during the course of the ATP hydrolysis cycle[148].

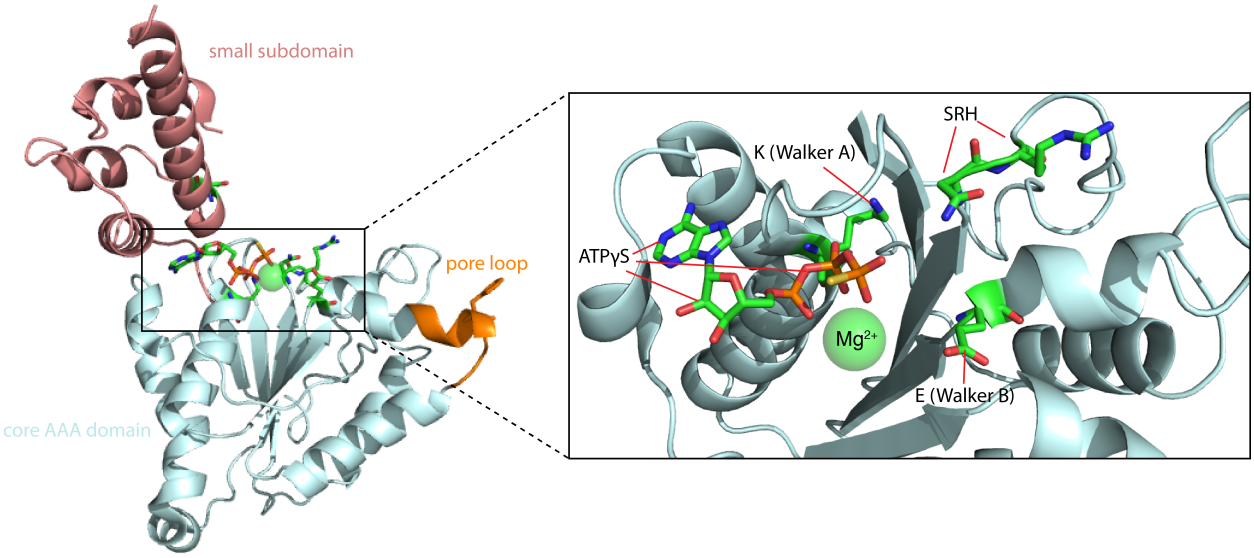



**Figure 1.2: The AAA domain.** Left, overall structure of the AAA domain, including the core AAA fold and the C-terminal small subdomain. The boxed region is the nucleotide binding site, and residues discussed in the text are highlighted in the closeup view at right. The AAA domain shown is the human p97 D2 domain bound to ATP**γ**S (PDB: 5FTN).

Each nucleotide-binding site in a AAA ring is found between two adjacent monomers, and several individual binding and hydrolysis motifs have been investigated in detail. In the N-terminal subdomain, two motifs termed Walker A and Walker B are located on either side of the "additional strand" that is inserted in the P-loop NTPase core fold (Figure 1.2)[150]. The Walker A motif is located in the P loop, immediately after beta strand 1, and consists of the motif GxxGxGK[S/T], where x is any amino acid. This motif recognizes the beta and gamma phosphates of the bound ATP nucleotide, and mutation of the invariant lysine abolishes nucleotide binding[151]. The Walker B motif is found at the apex of β3 and is thus separated in sequence from the Walker A motif by α1, β2, and α2, but is nearby in three-dimensional space.

This motif has the consensus sequence hhhhDE, where h is a hydrophobic residue[145]. It is responsible for coordinating a magnesium ion that assists in nucleotide binding (via the aspartate) and for activating a water molecule for nucleophilic attack on the gamma phosphate (via the glutamate, which is the "catalytic E")[149]. Mutation of the conserved E to Q or, more drastically, to A, typically abolishes nucleotide hydrolysis, but not binding. Walker A and Walker B mutations are thus often used to mimic the nucleotide-free and ATP-bound states, respectively[145].

An additional conserved motif in many AAA proteins is the second region of homology, or SRH, which spans the C-terminal portion of β4, α4, and the loop between α4 and β5 (Figure 1.2)[149]. At the N terminus of the SRH is a polar residue termed Sensor-1 (asparagine in both domains of Cdc48) that has an additional role in contacting the gamma phosphate and promoting hydrolysis. The SRH also contains a so-called arginine finger, which interacts with the gamma phosphate of the nucleotide in a neighboring AAA subunit and may help to coordinate nucleotide hydrolysis between subunits[152]. Although the Sensor-1 and arginine finger residues are well conserved, their exact effects on ATPase function are not as obvious as those of the Walker motifs[153]; in Chapter 2, only Walker mutations are used to evaluate the roles of the D1 and D2 ATPase domains of Cdc48 in *in vitro* experiments.

The assembly of AAA proteins into oligomeric assemblies produces a central pore. Each AAA domain projects a "pore loop," formed by the linker between β2 and α2 (Figure 1.2), into this central channel. In most cases, the pore loop includes at least one aromatic residue. In cases for which functional assays have been established, the aromatic residue has usually been found to be required for substrate processing[154,155]. The exact mechanism by which these residues might move in response to ATP hydrolysis has not yet been determined, although their interaction with

substrate proteins has been confirmed by structures of AAA ATPases actively translocating polypeptides[156,157]. Interestingly, the D2 pore loop of Cdc48 has two aromatic residues, including a bulky tryptophan that is rare among AAA proteins but conserved among Cdc48 orthologs, while the D1 pore loop contains no aromatic residues. The implications of this contrast for substrate processing will be further discussed in Chapter 2.

Cdc48 is a member of the "classical" clade of AAA ATPases, which also includes the bacterial unfoldase and protease FtsH, the ATPases of the 19S proteasome, the peroxisomal biogenesis complex Pex1/Pex6, and NSF[145]. The AAA domains of these proteins cluster together in evolutionary analysis by virtue of two distinguishing features: the replacement of a conserved arginine with an aspartate at the so-called Sensor 2 position in the small subdomain, and an extra helical insert between β2 and α2, immediately before the pore loop[145,158]. Whether this helical insert has any unique role in substrate processing has not been explored. As will be detailed below, Cdc48's additional N-terminal domain and binding partners differentiate it from its relatives and produce its specific activity.

*1.3.2 X-ray, cryo-EM, and NMR analyses of Cdc48 and selected cofactors*

The ring structure observed in initial electron micrographs of the cellular fractions containing Cdc48 was confirmed at higher resolution with the first crystal structure of p97 in 2000[54]. This structure, which was solved with the protein in the ADP-bound state, contained only the N and D1 domains, but revealed several important features of the ATPase. Notably, the N domains protrude from the D1 ATPase ring, forming an additional outer ring coplanar with the ATPases, although individual N domains do not contact each other[54]. Two subdomains are present in the N domain, with the N-terminal subdomain forming a double psi barrel fold and the C-terminal subdomain forming a four-stranded beta barrel[54]. This subdomain architecture is also

present in the N domain of NSF[159]. In the crystallized form of p97, the N domains make contacts with the alpha-helical C-terminal subdomains of the D1 AAA domains through an interface that is fairly extensive and well-conserved, indicating that this interaction is biologically important.

Within a few years, crystal structures of the full-length human and mouse p97 proteins were available, including structures in the presence of the ATP analog AMP-PNP, the transition state mimic ADP-AlFx, and ADP[160-162]. The relatively low resolution of many of these structures (between 4.7 and 3.5 Å) limited their ability to explain how nucleotide binding, hydrolysis, and release might be coupled to relative movements of the ATPase domains and the N domain to generate force. Several of these structures could be re-refined and improved using a higher resolution (3.0 Å) model of the D2 ATPase[163]. The isolated D2 AAA domain crystallized as a heptameric (rather than hexameric) ring, highlighting the plasticity of inter-domain contacts[163]. This flexibility might allow for relative motions of the domains during the hydrolysis cycle while maintaining the overall hexameric arrangement.

Notably, regardless of the nucleotide included during crystallization, only ADP was observed in the D1 ring in these structures[163]. In addition, only minor rearrangements of inter-domain contacts and angles were observed, possibly because of the restrictions imposed by crystal packing[163]. A comparison of the AMP-PNP and ADP states revealed a rearrangement of the aromatic residues of the D2 pore loop, which flip upward in the ATP analog state and downward in the ADP states, suggesting motion of these residues as a potential mechanism for transmission of force to the substrate[163]. However, a zinc ion occluding the central pore in the D1 region led to the hypothesis that substrates do not pass through the pore and are instead remodeled or unfolded through interactions with the exterior of p97[153,163].

An advance in understanding nucleotide-induced rearrangements in p97 came with the crystallization of disease mutant proteins. These mutants corresponded to several of those observed in the human disease IBMPFD (inclusion body myopathy with Paget's disease of bone and frontotemporal dementia). This autosomal dominant multisystem degenerative syndrome manifests with cytosolic protein aggregates in cells of the affected organs and is caused by mutations in p97 that cluster at the N-D1 interface described above[164,165]. Mutation of several of these residues resulted in the ability to crystallize p97 with the ATP analog ATP**γ**S bound to both the D1 and D2 domains[166]. In this nucleotide state, the N domains undergo a significant rearrangement, adopting a so-called "up conformation" above the plane of the D1 ring, as opposed to the coplanar "down conformation" observed with D1 in the ADP state[166]. Presumably, weakening of the N-D1 interface allows the N domains to transition to the up conformation with a lower energetic cost. Although the significance of this conformational change for disease pathogenesis was not immediately obvious, it did demonstrate that large-scale domain rearrangements can occur in p97, and led to speculation that relative motions of the N domains might be responsible for substrate manipulation.

Recently, high-resolution cryo-EM reconstructions of p97 in ADP (2.4-2.3 Å) and ATP**γ**S (3.3-3.2 Å) states have become available[55]. These structures avoid the caveats imposed by crystal packing restrictions, and largely confirm the older ADP-bound structures, albeit with notably improved resolution. 3D classification of the ATP**γ**S particles allowed for separation into three classes, which were differentiated according to nucleotide state: the first had ADP bound in both domains, the second had ATP**γ**S bound to D2 and ADP to D1, and the third had ATP**γ**S bound in both domains[55]. These structures allowed more accurate delineation of the effects of ATP binding than the previous crystal structures, showing that ATP binding to D2 causes a 10 to

15 degree relative rotation of the D1 and D2 rings, while ATP binding to D1 causes the same upward rotation of the N domains that had previously been observed in the disease mutants[55]. Thus, N domain motions are a bona fide feature of the wild type ATPase. Nevertheless, because these cryo-EM structures were determined with six-fold symmetry imposed, it was not possible to determine whether the true ATP hydrolysis cycle involves similar motions of all of the monomers at once, or instead operates via a rotational mechanism or via stochastic firing of subunits. Indeed, the latter appears to be the case for the related bacterial ATPase ClpX, in which subunits are partially coordinated but do not execute the same hydrolysis phases simultaneously[167]. Furthermore, despite the new high-resolution information, the fundamental mechanism of substrate processing remained elusive.

The mode of Cdc48 binding by several cofactors has been investigated in detail. Particularly relevant to the proteins discussed in this dissertation are the binding mechanisms of Ufd1, Npl4, and the DUB Otu1. Each of these proteins interacts with the N domain of Cdc48; Ufd1 binds via a pair of SHP motifs, and Npl4 and Otu1 via UBX-like domains[73,168]. However, because each Cdc48 complex contains six N domains, the relative arrangement of binding sites in a multi-cofactor complex is unknown.

The UBX-like (UBXL) domains of Npl4 and Otu1 share a common beta-grasp fold with the larger UBX family and with ubiquitin itself, although the UBXL and UBX primary sequences do not overlap extensively[168,169]. The beta-grasp fold found in UBX proteins inserts a conserved Phe-Pro-Arg (FPR) motif with a cis-proline into the hydrophobic cleft found between the two subdomains of the Cdc48 N domain[170]. Although the FPR motif is not present in UBXL proteins, the inter-subdomain interaction site is conserved, as both the Npl4 and Otu1 UBXL domains insert loops into this cleft[168,169]. Because the UBXL domain is roughly the size of a single

subdomain of the N domain, the complex of the full N domain with a UBXL produces a tri-lobed arrangement. The interdomain cleft faces outward from the D1 ring in both the down conformation and the up conformation of the N domains, such that a UBXL domain could conceivably bind to Cdc48 in both states.

The SHP motif, or SHP box, is a short sequence found both in the Cdc48 cofactor Shp1 (p47 in higher organisms) and in the unstructured C-terminal region of Ufd1, the UT6 domain[58]. This motif has the consensus sequence $FxGxGx_2H$, where x is any residue and h is hydrophobic. It has recently been shown that this sequence, unlike other Cdc48 binding motifs, interacts with the side of the C-terminal subdomain of the N domain[171,172]. This binding site would position the SHP box close to the body of the D1 ATPase in the down conformation of the N domains, and at the tip of the N domain closest to the central pore axis in the up conformation. The UT6 domain of Ufd1 has two SHP boxes separated by a relatively lengthy linker (54 residues in the *S. cerevisiae* protein). Deletion mapping has shown that this linker contains the region of Ufd1 required for heterodimerization with Npl4[73].

**1.4 Summary of this work**

The work described here addresses the biochemical mechanism and molecular structure of the Cdc48/Ufd1/Npl4 complex. In Chapter 2, I describe our work with purified components to elucidate the mechanism of polyubiquitinated substrate processing by Cdc48. Several newly developed biochemical assays show that Cdc48 is an unfoldase that functions by passing its substrates through its central pore. In Chapter 3, I describe a set of structural biology experiments that have yielded the first high-resolution views of the Cdc48/UN complex. Together, these data provide new insights into the fundamental mechanism of a critical protein quality control factor.

## Chapter 2: Molecular Mechanism of Substrate Processing by the Cdc48 ATPase Complex

This chapter has been published previously as:

Bodnar NO, Rapoport TA. "Molecular mechanism of substrate processing by the Cdc48 ATPase complex." *Cell* 2017 May 4;169(4):722-735.e9.

### 2.1 Abstract

The Cdc48 ATPase and its cofactors Ufd1/Npl4 (UN) extract polyubiquitinated proteins from membranes or macromolecular complexes, but how they perform these functions is unclear. Cdc48 consists of an N-terminal domain that binds UN and two stacked hexameric ATPase rings (D1 and D2) surrounding a central pore. Here, we use purified components to elucidate how the Cdc48 complex processes substrates. After interaction of the polyubiquitin chain with UN, ATP hydrolysis by the D2 ring moves the polypeptide completely through the double ring, generating a pulling force on the substrate and causing its unfolding. ATP hydrolysis by the D1 ring is important for subsequent substrate release from the Cdc48 complex. This release requires cooperation of Cdc48 with a deubiquitinase, which trims polyubiquitin to an oligoubiquitin chain that is then also translocated through the pore. Together, these results lead to a new paradigm for the function of Cdc48 and its mammalian ortholog p97/VCP.

### 2.2 Introduction

The yeast Cdc48 ATPase and its metazoan ortholog p97 (or VCP) are critical components of many ubiquitin-dependent cellular pathways that require the segregation of individual proteins from binding partners or membranes (for review, see[173,174,175]). This ATPase is present in all eukaryotic cells and is essential for their viability. The function of Cdc48/p97 is best understood in endoplasmic reticulum (ER)-associated protein degradation (ERAD), in which it extracts polyubiquitinated, misfolded proteins from the ER and transfers them to the proteasome for degradation[176]. Cdc48/p97 also functions in mitochondrion-associated protein

degradation[177], ribosomal quality control[135], and the extraction of chromatin-bound proteins[178,179]. Consistent with this central role in protein quality control, mutations in human p97 cause several neurodegenerative proteopathies[180,181]. Despite its biological significance, the mechanism of Cdc48/p97 action is poorly understood.

Cdc48 belongs to the AAA+ family of ATPases (ATPases associated with various cellular activities), whose members use ATP hydrolysis to exert force on macromolecules[182,183]. Many of the AAA proteins form hexamers with either a single or double ring of ATPase domains (type I and II ATPases, respectively). Cdc48 is a type II ATPase consisting of an N-terminal (N) domain, two tandem AAA domains (D1 and D2) separated by a short linker, and a flexible C-terminal tail[160]. The AAA domains form two stacked rings surrounding a central pore. Both ATPase rings hydrolyze ATP[184,72], but their roles in substrate processing are unknown. ATP hydrolysis in the D1 ring seems to move the N domains from a so-called "up-conformation" in the ATP state to a position co-planar with the D1 ring in the ADP-bound state ("down-conformation")[55]. The function of this N domain movement is unclear.

Cdc48 cooperates with a large number of protein cofactors that provide pathway selectivity and fine-tune substrate processing. One of the most important cofactors is the Ufd1/Npl4 heterodimer (UN), an essential complex that participates in many Cdc48-dependent processes, including ERAD[87]. UN binds to the N domain of Cdc48 and recruits polyubiquitinated substrates to the ATPase[72]. In a reconstituted *in vitro* system of partial ERAD reactions, a polyubiquitinated, misfolded protein could be extracted from the membrane in a UN- and Cdc48- dependent manner[83]. However, most experiments with the Cdc48/p97 complex have been performed with intact cells, the complexity of which makes it impossible to obtain mechanistic insight. Dissection of Cdc48 function has also been hampered by a lack of suitable

*in vitro* substrates. The only reported experiments employed non-ubiquitinated proteins[185,153], which are not the principal substrates *in vivo*. Cdc48's physiological substrates are generally polypeptides modified with K48-linked polyubiquitin chains, which also serve as a major targeting signal for the proteasome[35]. *In vitro* experiments with the proteasome have employed polypeptides decorated with K63-linked polyubiquitin chains[186], but these substrates are not appropriate for the more specific Cdc48 ATPase complex[72].

One of the most important questions is how the Cdc48 ATPase can "pull" on a substrate, thereby releasing it from a protein complex or membrane. Some AAA ATPases, including the Clp ATPases in bacteria, the archaeal Cdc48 homologs, and the 19S proteasome, use a translocation mechanism (for review, see[187]). In these cases, central loops with conserved aromatic residues are thought to contact a polypeptide chain and move in response to ATP hydrolysis, thereby dragging the substrate through the pore. Archaeal Cdc48 and some mammalian p97 mutants can also translocate polypeptides into associated 20S proteasomes under certain conditions[188,189], implying movement through the central pore. However, archaeal Cdc48 undergoes conformational changes that have never been observed with eukaryotic homologs[190], and wild type yeast Cdc48 and mammalian p97 are thought not to use a translocation process[55,161]. A mechanism that does not involve translocation is also employed by Cdc48's closest relative, the NEM-sensitive fusion protein (NSF)[191]. NSF is thought to bind the SNARE complex and the cofactor αSNAP with its N domains, and to use ATP hydrolysis to unwind the supercoiled SNARE proteins without polypeptide passage through the central pore. A major argument against a translocation mechanism for Cdc48/p97 is that its central pore is very narrow or occluded in crystal or cryo-EM structures. These observations have led to the proposal of several alternative mechanisms. In one model, a substrate transiently enters the D2

ring or moves through the D2 ring, exiting between the D1 and D2 rings[161]. Other models postulate that the polypeptide chain inserts only shallowly into the D1 ring (for review, see[175]). Finally, it has been suggested that the large, nucleotide-dependent conformational changes of the N domains are sufficient to move a polypeptide chain[192]. In each of these cases, it is unclear how a continuous pulling force could be generated, although such a force would likely be required to extract a subunit from a tight multimeric complex or a multi-spanning protein from the ER membrane. Clearly, experiments are needed to test whether a translocation mechanism or any of the alternative models can explain the function of Cdc48.

Another important question concerns the mechanism by which a polyubiquitinated substrate is released from the Cdc48 complex and passed on to downstream components, such as the proteasome. There are at least three proposed models. In the first, a ubiquitinated substrate is transferred by ubiquitin-binding shuttling factors to the proteasome[193]. In the second, the substrate is completely deubiquitinated, processed by Cdc48, and then re-ubiquitinated to enable shuttling factor and proteasome binding (for review, see[67]). In the third, substrate is only partially deubiquitinated, leaving sufficient ubiquitin moieties for interaction with downstream components. A candidate for a deubiquitinase involved in Cdc48-dependent reactions is Otu1 (called YOD1 in mammals). Otu1 binds via its UBX-like domain to the N domain of Cdc48, an interaction that stimulates its deubiquitination activity[83,194]. Expression of an enzymatically inactive Otu1 mutant blocks Cdc48 function *in vivo* and leads to the accumulation of polyubiquitinated proteins on the Cdc48/p97 complex. However, the precise role of Otu1 has yet to be clarified. Specifically, it is unclear whether it deubiquitinates substrates before they are processed by Cdc48 and thus acts as a negative regulator[68], or synergistically cooperates with Cdc48.

Here, we have elucidated the mechanism of Cdc48 from *S. cerevisiae* with purified components that include Cdc48 itself, its UN cofactor, and model substrates bearing K48-linked polyubiquitin chains. We demonstrate that polyubiquitinated substrate first binds to UN. Following ATP binding to the D1 ring, Cdc48 unfolds the substrate by passing it through the central pore, starting from the D1 side and moving it all the way through the D2 ring. This translocation process relies on ATP hydrolysis by the D2 domain. We show that subsequent substrate release from Cdc48 requires hydrolysis of ATP in D1 and partial trimming of the ubiquitin chain by the deubiquitinase Otu1. The remaining ubiquitin molecules follow the substrate through the central pore and are released on the D2 side of the pore. Our results establish a model for protein extraction and segregation that is broadly applicable to the various Cdc48-dependent quality control systems.

## 2.3 Results

### *Generation of a purified Cdc48 substrate*

Cdc48's substrates in ERAD and other pathways commonly carry K48-linked polyubiquitin chains. To obtain a well-defined substrate modified with a single ubiquitin chain of this linkage, we employed a system based on the N-end rule pathway. This pathway degrades proteins with certain N-terminal amino acids, such as arginines[195]. The substrate used for our studies is produced as a fusion protein (Figure 2.1A), consisting of an N-terminal SUMO tag, a 43 amino acid degron sequence that has been shown to be polyubiquitinated at a lysine residue[195], and superfolder GFP (sfGFP). The fusion protein was expressed in *E. coli* and purified on the basis of an N-terminal His-tag. The SUMO tag was then cleaved off with the SUMO-specific protease Ulp1.

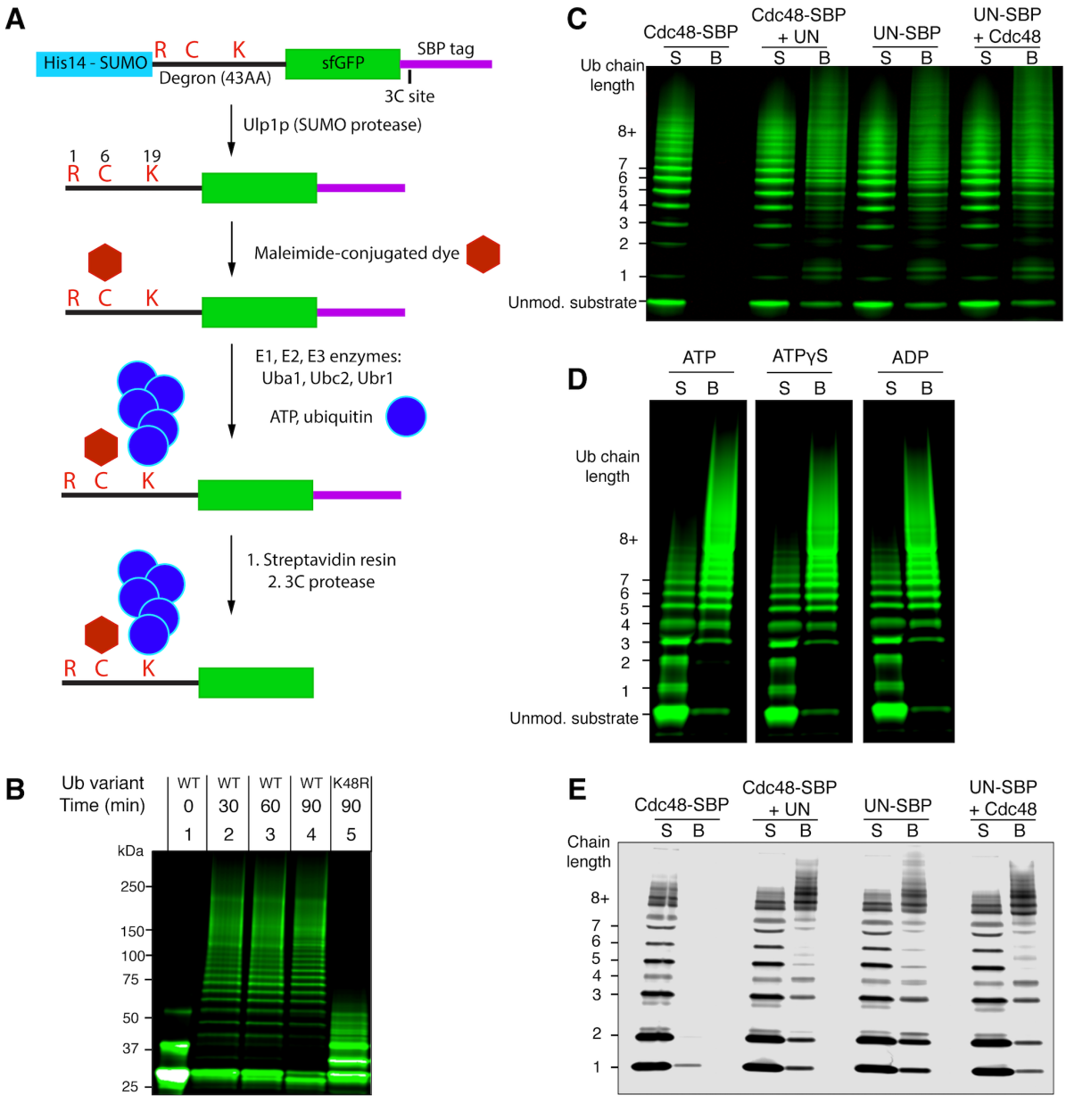



**Figure 2.1: Cdc48/UN complex interacts with a polyubiquitinated substrate**

(A) Scheme for the synthesis and purification of polyubiquitinated superfolder GFP (sfGFP) a Cdc48 substrate containing an N-end degron; see text for details.

(B) Dye-labeled sfGFP bearing the N-end degron was incubated with ubiquitination machinery, ATP, and the indicated ubiquitin variant. Samples were analyzed by SDS-

**Figure 2.1 (Continued)**

PAGE and fluorescence scanning. The unmodified substrate appears as two bands due to incomplete denaturation in SDS.

(C) The indicated proteins were immobilized through a SBP tag on streptavidin beads and incubated with ubiquitinated sfGFP. Supernatants (S) and bound material (B) were analyzed by SDS-PAGE and fluorescence scanning.

(D) As in (C), with Cdc48 and SBP-tagged UN used in each condition. The indicated nucleotide was included in all steps following purification. Irrelevant lanes of the gel have been removed for display (white spaces).

(E) As in (C), but with free fluorescently labeled K48-linked polyubiquitin chains instead of polyubiquitinated sfGFP.

The cleavage product contains an N-terminal arginine, which renders it a substrate for Ubr1, an E3 ubiquitin ligase of the N-end rule pathway[196]. To facilitate detection of the substrate, a fluorescent dye was attached to a cysteine close to the N-terminus. The labeled protein was then incubated with a mixture of purified ubiquitin-activating enzyme (Uba1), ubiquitin-conjugating enzyme (Ubc2), ubiquitin ligase (Ubr1; for purity of the proteins, see Figure 2.S1A), ubiquitin, and ATP. The polyubiquitinated substrate was purified on streptavidin beads via a C-terminal streptavidin binding peptide (SBP) tag and eluted from the resin by cleavage with 3C protease.

The substrate contained ubiquitin chains of different lengths (Figure 2.1B; lanes 2-4). Essentially all chains are composed of K48 linkages, as most high-molecular weight chains failed to form when wild type ubiquitin was replaced with the ubiquitin mutant K48R (Figure

2.1B; lane 5). Mass spectrometry suggested that the remaining modification is monoubiquitination at several sites in the sfGFP moiety (Figure 2.S1B). Analysis of tryptic peptides derived from the polyubiquitinated substrate also suggested that K63-linked chains occurred in only low abundance (three orders of magnitude lower than K48 linkages) and that polyubiquitin chains were attached almost exclusively at a single lysine in the degron segment preceding sfGFP (Figure 2.S1B).

*Substrate binding to the Cdc48 complex*

To test substrate binding to the Cdc48 complex, polyubiquitinated sfGFP substrate (Ub(n)-GFP) was incubated with SBP-tagged *S. cerevisiae* Cdc48 ATPase, which was purified after expression in *E. coli* and immobilized on streptavidin beads. The incubation was performed in the presence or absence of the UN cofactor from *S. cerevisiae*, which was also expressed in *E. coli* and purified as a heterodimer (Figure 2.S1A). Consistent with previous studies[83], substrate binding was only observed in the presence of the cofactor (Figure 2.1C). Substrate molecules with chains of five or more ubiquitin moieties bound more efficiently than those with shorter chains. The UN complex alone was capable of substrate binding, as shown by immobilizing the cofactor on streptavidin beads through an SBP tag on Ufd1 (Figure 2.1C). The extent of substrate binding was comparable in the presence of ATP, ATPγS, and ADP (Figure 2.1D). Similar results were obtained when the binding of free K48-linked ubiquitin chains was tested (Figure 2.1E). These chains were synthesized with purified Uba1, the ubiquitin-conjugating enzyme Ubc1[197], and ubiquitin that was labeled with a fluorescent dye at a cysteine appended to the N-terminus. Again, the UN complex was responsible for all interaction with the Cdc48 complex (Figure 2.1E). Taken together, these results indicate that substrate is recruited to the Cdc48 complex through an interaction between the polyubiquitin chain and the UN cofactor.

*The effect of substrate on Cdc48 ATPase activity*

To evaluate the effect of substrate binding on Cdc48 activity, we measured the rate of steady state ATP hydrolysis. UN reduced the intrinsic ATPase rate of Cdc48 by a factor of about two (Figure 2.2A). When saturating Ub(n)-sfGFP was added to the Cdc48/UN complex, the ATP hydrolysis rate was stimulated approximately 5-fold. The substrate had no stimulatory effect in the absence of the UN complex (Figure 2.2A), and no stimulation was seen with non-ubiquitinated sfGFP (Figure 2.2B). Free K48-linked polyubiquitin chains stimulated ATP hydrolysis to the same extent as polyubiquitinated sfGFP (Figure 2.2B). Stimulation was dependent on the length of the ubiquitin chains, with mono- and di-ubiquitin having negligible effects, and penta-ubiquitin producing the highest rate of hydrolysis (Figure 2.2C). K63-linked penta-ubiquitin, by contrast, had no effect on ATP hydrolysis (Figure 2.S2A).

To test how substrate affects each of Cdc48's two ATPase domains, we generated mutations designed to prevent ATP hydrolysis, but not ATP binding, in either the D1 (E315A) or D2 (E588A) domains (mutations in the Walker B motifs). A Cdc48 mutant in which D2 is the only active ATPase (E315A) eliminated ATP hydrolysis in the absence or presence of the UN

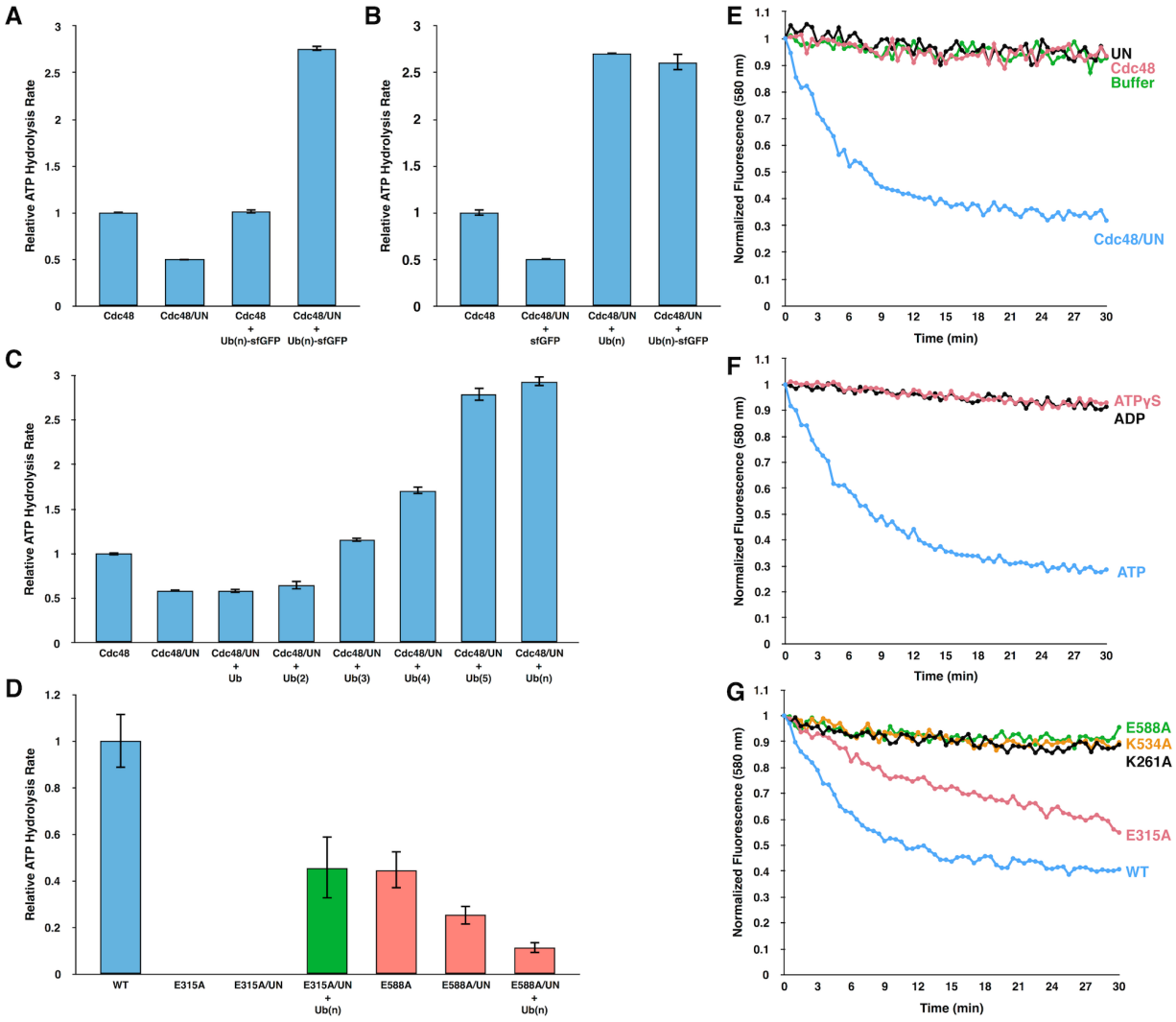



**Figure 2.2: Ubiquitin stimulates Cdc48 ATPase activity for substrate unfolding**

(A) ATP hydrolysis rates were determined with the indicated combinations of purified proteins and substrate (Cdc48; UN; Ub(n)-GFP). Substrate was included at 5-fold excess over the cofactor to maximize occupancy. The rates were normalized with respect to that of Cdc48 alone. Shown are the means and standard deviations of three experiments.

(B) As in (A), but also with free K48-linked polyubiquitin chains (Ub(n)), carrying up to ~15 ubiquitin moieties.

(C) ATPase stimulation with free K48-linked ubiquitin chains of increasing length.

**Figure 2.2 (Continued)**

(D) As in (B), but with Cdc48 Walker B mutations in D1 (E315A; green bars) or D2 (E588A; red bars). E315A and E588A can bind, but not hydrolyze, ATP in D1 and D2, respectively (Walker B mutations).

(E) Irradiated, fluorescent Eos, consisting of two polypeptide segments, was polyubiquitinated, as described for sfGFP (Figure 1A). This substrate (Ub(n)-Eos) was incubated with an ATP-regenerating system and excess of the indicated proteins. After addition of ATP, Eos fluorescence was followed over time.

(F) As in (E), but with Cdc48 and UN present in all conditions, and the nucleotide varied as indicated.

(G) As in (F), but with UN and an ATP regenerating system present in all conditions. Cdc48 or its point mutants were included as indicated. K261A and K534A are defective in ATP binding to the D1 and D2 ATPases, respectively (Walker A mutations), and E315A and E588A can bind, but not hydrolyze, ATP in D1 and D2, respectively (Walker B mutations).

complex (Figure 2.2D). However, the addition of both polyubiquitin and UN restored ATP hydrolysis to about half the level seen with wild type Cdc48 alone. Thus, substrate stimulates ATP hydrolysis in the D2 domain.

A mutant in which D1 is the only active ATPase (E588A) had reduced baseline activity, which was further suppressed by the addition of UN (Figure 2.2D). Interestingly, addition of

polyubiquitin resulted in an even more pronounced decrease in ATP hydrolysis (Figure 2.2D). Thus, substrate inhibits the ATPase activity of the D1 domain.

Similar results were obtained with more commonly used Walker B mutations, in which E315 or E588 are replaced by Gln. The E315Q mutation in the D1 domain did not entirely abolish ATP hydrolysis, but stimulation by the simultaneous presence of UN and polyubiquitin was still observed (Figure 2.S2B). Likewise, the E588Q mutation in the D2 domain resulted in a high baseline ATP hydrolysis rate, which was suppressed by the presence of cofactor and substrate. Thus, despite the fact that the conservative Gln mutations do not completely abolish ATPase activity, substrate still stimulates the D2 domain and slows the D1 domain. The effect on D2 is dominant, explaining why substrate stimulates the overall ATPase activity of the wild type protein.

*Cdc48 unfolds substrate*

The stimulatory effect of substrate on ATPase activity raised the possibility that Cdc48 actively unfolds the polypeptide chain. To test this possibility, we replaced the sfGFP moiety of the substrate (Figure 2.1A) with the monomeric fluorescent protein mEos3.2[198]. This protein undergoes a peptide backbone cleavage when irradiated with near-UV light, producing a complex of two fragments. If Cdc48 unfolds this complex, the constituent polypeptides are separated and fluorescence is lost; the two chains cannot re-associate, rendering any unfolding irreversible[199]. In contrast, the sfGFP substrate used initially has the ability to refold, which counteracts unfolding by Cdc48 and therefore allows only a small decrease in fluorescence (Figure 2.S2C).

The Eos substrate was polyubiquitinated and subjected to gel filtration to enrich for molecules carrying 5-10 ubiquitin moieties (Ub(n)-Eos). When incubated with Cdc48, UN, and an ATP regenerating system, Ub(n)-Eos was indeed unfolded, as evidenced by a decrease in fluorescence over time (Figure 2.2E). Unfolding did not occur when either Cdc48 or UN was omitted. The reaction was dependent on ATP hydrolysis, as no decrease in fluorescence was observed with ATPγS or ADP (Figure 2.2F).

Unfolding occurred with a Cdc48 molecule in which only the D2 ATPase is active (Walker B mutation E315A), although the rate of unfolding was slower than with wild type Cdc48 (Figure 2.2G). On the other hand, Cdc48 mutants in which only the D1 ATPase is active (E588A mutation) or which cannot bind ATP in either the D1 or D2 domain (Walker A mutations K261A or K534A) did not show any unfolding activity. The Walker B mutants E315Q and E588Q both exhibited residual unfoldase activity (Figure 2.S2D), consistent with their higher ATPase rates. Taken together, these data indicate that Cdc48 uses the energy of ATP hydrolysis to unfold substrates. Unfolding requires the UN cofactor, as well as ATP binding, but not hydrolysis, in the D1 domain, and both ATP binding and hydrolysis in the D2 domain.

*Substrate passes through the central pore of Cdc48*

One possible mechanism for the unfolding activity of Cdc48 is the translocation of the substrate through the central pore; the ubiquitin chain on the substrate would be bound to the N-domain associated UN cofactor, and the GFP moiety would be pulled through the D1 and D2 rings, from the "cis" to the "trans" side of the double ring. To test this idea, we used site-specific photo-crosslinking. The photo-reactive amino acid p-benzoylphenylalanine (Bpa) was introduced at several positions in the ATPase using amber codon suppression[200] (Figure 2.3A). Upon UV

irradiation, Bpa forms covalent crosslinks to targets located within a distance of ~3Å from the selected positions[201].

We first tested a position in the D1 pore loop, which is located at the cis side of the double ring, at the entrance of the central pore (M288; Figure 2.3B). A SBP-tagged version of Cdc48 that carries Bpa at this position was incubated with fluorescently labeled Ub(n)-sfGFP in the presence or absence of the UN complex and various nucleotides. After irradiation, Cdc48-SBP was recovered with streptavidin beads, which were then washed with high-salt buffer. The remaining bound material was subjected to SDS-PAGE and analyzed with a fluorescence scanner. Products of crosslinking between substrate and Cdc48 were only seen when the UN complex was present and the samples were irradiated (Figure 2.3B, lanes 4-6 versus lanes 1-3, 7). The major product corresponds to one Ub(n)-sfGFP molecule crosslinked to one Cdc48 molecule. Higher molecular weight bands likely contain additional crosslinked Cdc48 molecules. The crosslinks were most prominent in ATPγS and ADP (lanes 5, 6), but were also observed in the presence of ATP (lane 4). These results indicate that, in the presence of UN, substrate binds to the entrance of the D1 ring even without ATP hydrolysis. Crosslinking might be weaker in the presence of ATP because, under these conditions, the D1 subunits are not all in the same nucleotide state, which may result in only some pore loops being close to substrate, or because substrate can enter the pore and might move the pore loops into an unfavorable position for interaction.

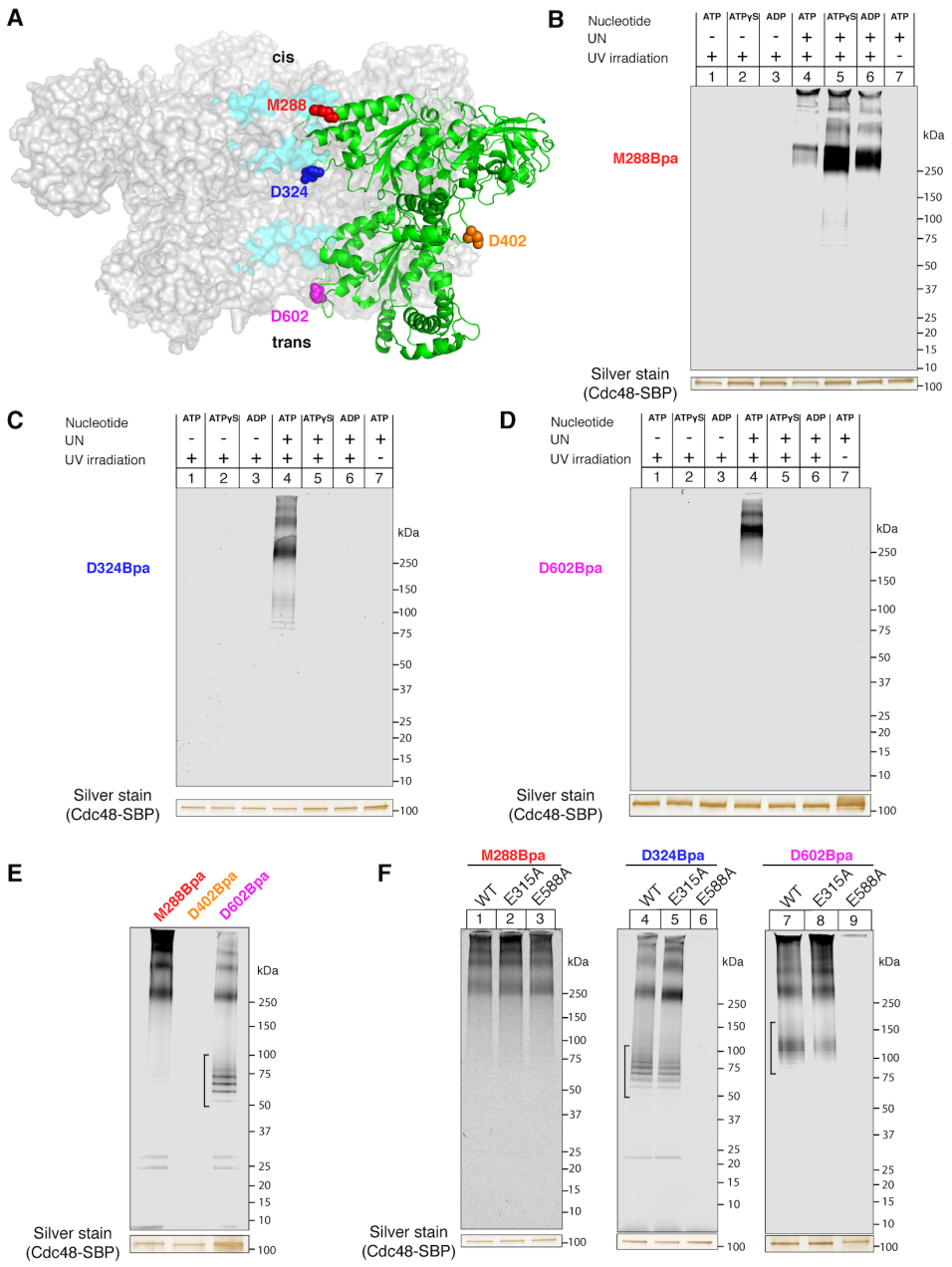



**Figure 2.3: Substrate passes through the central pore of Cdc48**

(A) Cdc48 positions used for introduction of the Bpa crosslinker. Sections of the central pore are highlighted in cyan, and a single Cdc48 monomer in the hexamer is shown in green. The Cdc48 model was generated based on PDB ID 3CF1.

**Figure 2.3 (Continued)**

(B) SBP-tagged Cdc48 (Cdc48-SBP) with Bpa at the M288 position (D1 pore loop) was incubated with dye-labeled, polyubiquitinated sfGFP and the indicated combinations of UN and nucleotide. After irradiation and pulldown with streptavidin beads, crosslinked species were detected by SDS-PAGE and fluorescence scanning. The silver-stained band of Cdc48-SBP serves as a loading control.

(C) As in (B), but with Bpa at the D324 position in the interior of the central pore.

(D) As in (B), but with Bpa at the D602 position in the D2 pore-2 loop.

(E) As in (B), but with Bpa at an external position (D402). For comparison, crosslinking was also performed with probes in the pore. UN and ATP were included in all conditions. It should be noted that some non-crosslinked substrate remained associated with Cdc48. We generally observed that in the presence of ATP, non-crosslinked material had a tendency to stick to Cdc48 (brackets; see also Figure 3F), likely because it was present in an unfolded conformation inside the pore.

(F) As in (B), but with Cdc48 mutants that contain the crosslinker and Walker mutations at the indicated positions. E315A, Walker B mutation in D1; E588A, Walker B mutation in D2. UN and ATP were included in all conditions.

Next, we incorporated Bpa at a position in the interior of the central pore (D324), immediately below the D1 constriction that has been proposed to block substrate access to the pore (Figures 2.3A, C). Crosslinks between substrate and Cdc48 were formed in the presence of ATP, but not ATPγS or ADP (Figure 2.3C, lane 4 versus lanes 5, 6). No crosslinks were observed in the absence of the UN complex or without irradiation (lanes 1-3, 7). These results

show that Cdc48 uses the energy of ATP hydrolysis to move substrate molecules past the D1 opening of the pore.

Finally, we placed the photoreactive probe at the exit of the D2 ring (position D602), in the pore-2 loop located at the trans side of the double ring (Figures 2.3A, D). The pore-2 loop is positioned below the pore-1 loop, which contains the conserved aromatic residues responsible for substrate movement in other AAA proteins. Crosslinks were again formed in the presence of ATP, but not ATPγS or ADP, or when the UN complex or irradiation were omitted (Figure 3D). As expected, no crosslinks were observed when Bpa was placed on the exterior of the hexamer (position D402; Figures 2.3A, E). Crosslinking efficiency was higher for the M288 position (~20%) than for the D324 and D602 positions (~5%) (Figure 2.S3), likely because only a fraction of the substrate bound to the cis side was translocated through the pore. Taken together, these data suggest that ATP hydrolysis allows the Cdc48 ATPase to pull the substrate all the way from the entrance of the D1 ring (cis side), through the central pore, to the trans side of the double-ring.

Consistent with its capacity to unfold Ub(n)-Eos, a Cdc48 mutant in which only the D2 ATPase is active (Walker B mutation E315A in the D1 ring) maintained its ability to form substrate crosslinks to the cis side (M288), interior (D324), and trans side (D602) of the pore (Figure 2.3F). In contrast, a mutant in which only the D1 ATPase is active (Walker B mutation E588A in the D2 ring) eliminated crosslinking to the interior and trans side of the pore, while maintaining crosslinking to the cis side (Figure 2.3F). Thus, ATP hydrolysis in D2 is required for pore entry and passage, but not for the initial interaction with the D1 pore entrance. These results confirm that the translocation and unfolding of substrates require only ATP binding, but not

hydrolysis, in the D1 domain, whereas ATP hydrolysis in the D2 domain is obligatory for both activities.

*Substrate moves entirely through the double-ring of Cdc48*

To test whether a substrate molecule moves all the way through the double-ring of the Cdc48 ATPase, we attached a ring-shaped protease to the D2 ring; if a polypeptide chain emerges from the D2 ring, it should be degraded. We took advantage of the fact that the bacterial AAA protease FtsH contains tandem ATPase and protease domains, which form stacked hexameric rings[202]. The FtsH protease domain from *Aquifex aeolicus*, which forms a hexamer on its own[203], was fused to the C-terminus of Cdc48, removing the last 10 residues of the flexible native C-terminal tail but otherwise leaving the ATPase unaltered (Figure 2.4A). As with FtsH, this hybrid construct (Cdc48-FtsH) is expected to degrade proteins that emerge from the trans side of the ATPase ring. For these experiments, we used polyubiquitinated mEos3.2 that had not been photo-converted, preserving an intact polypeptide. A single exposed cysteine near the C-terminus of Eos (C195 of the native sequence) was used to label the protein with a maleimide-conjugated dye. When this substrate was incubated with either UN or Cdc48-FtsH alone in the presence of ATP, no degradation was seen over a period of 20 min (Figure 2.4B). Addition of a large excess of the deubiquitinating enzyme Otu1 at the end of the incubation period generated unmodified substrate, with a minority of the substrate retaining 1-3 ubiquitin moieties. Importantly, when substrate was incubated with both UN and Cdc48-FtsH, it was degraded in a time-dependent manner. A stable proteolytic fragment of ~13 kDa appeared over time (Figure 2.4B, arrow head), which contains Cys238, to which the fluorescent dye is attached. The fragment likely corresponds to the C-terminus of the substrate, as shown by the identification of a peptide by mass spectrometry (Figure 2.S4A), but the abundance of the fragment was too low

to unambiguously define its boundaries. A comparison of the intensity of bands after Otu1 treatment indicated that this stable cleavage product accounts for approximately 60% of the total

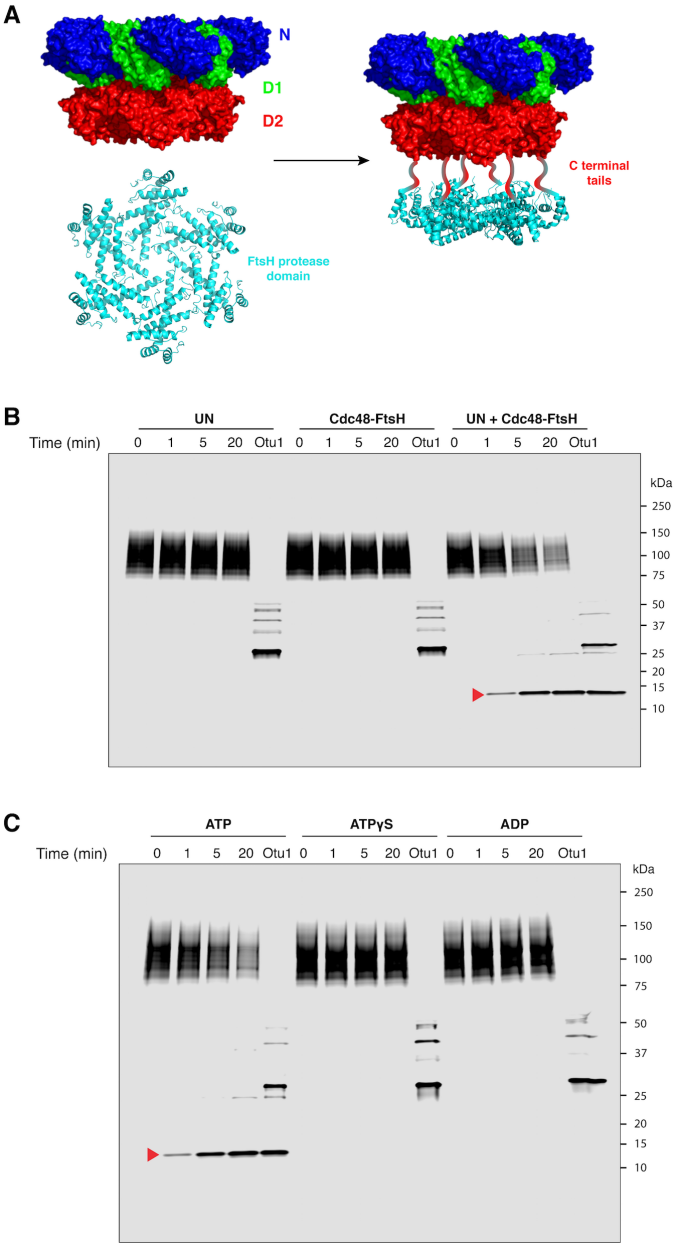



**Figure 2.4: Substrate exits from the D2 side of the double ring ATPase**

(A) The proteolytic domain of FtsH (PDB ID 2DI4) was fused to the C-terminal tail of the Cdc48 ATPase to produce a hybrid AAA protease (Cdc48-FtsH).

**Figure 2.4 (Continued)**

(B) Dye-labeled, polyubiquitinated Eos was incubated with an ATP regenerating system and the indicated proteins. Aliquots were removed at the indicated time points and analyzed by SDS-PAGE and fluorescence scanning. After the time course, a 50-fold excess of Otu1 was added to remove ubiquitin chains. The red arrow head indicates a stable fragment containing the fluorescent dye, which was generated when both Cdc48-FtsH and UN are present. The dye is attached to C238, near the C-terminus of the 275-residue substrate.

(C) As in (B), with Cdc48-FtsH and UN in all conditions, and the nucleotide varied as indicated.

substrate signal, consistent with the efficiency of the unfolding reaction (Figure 2.2E). Mass spectrometry analysis of the full reaction mixture without further proteolytic processing identified peptides representative of the full substrate sequence (Figure 2.S4A). Thus, the entire substrate polypeptide is moved through the D2 ring of Cdc48 into the FtsH ring. Consistent with this conclusion, no substrate degradation was observed when ATP was replaced with ATPγS or ADP (Figure 2.4C). Both substrate unfolding and positioning of the proteolytic sites at the D2 pore exit are required for this reaction, as no degradation was observed when wild type Cdc48 was incubated with the isolated FtsH protease domain (Figure 2.S4B).

*Substrate release requires ubiquitin chain trimming*

If substrate is pulled through the central pore from the cis to the trans side, but the associated ubiquitin chain remains bound to the UN cofactor on the cis side, how is substrate released from the Cdc48 complex? Indeed, we found that substrate does not spontaneously

dissociate, as labeled substrate remained bound over a 20-minute time course even in the presence of excess unlabeled substrate (Figure 2.S5A, compare lanes 1-4 and 5-8). These experiments were performed in the presence of ATP, i.e. under conditions in which substrate translocation and unfolding occur. Thus, an additional mechanism is required for substrate release. Previous experiments have implicated the deubiquitinating (DUB) enzyme Otu1 in Cdc48 function[83]; indeed, addition of Otu1 rapidly reversed the sfGFP fluorescence loss mediated by Cdc48 and UN, suggesting that deubiquitination liberates the substrate and allows it to refold (Figure 2.S2C).

To further test the role of deubiquitination in substrate release, we immobilized preformed complexes of Ub(n)-sfGFP, Cdc48, and SBP-tagged UN on streptavidin beads. Addition of Otu1 resulted in the efficient release of substrate from the beads (Figure 2.5A, lanes 5-8 versus 1-4; Figure S5A, lanes 9-12 versus 1-4). Release was much slower with an Otu1 mutant that lacked the Cdc48-interacting UBX-like domain (Figure 2.5A, lanes 9-12), and no substrate dissociation was observed with a catalytically inactive Otu1 mutant (C120S) (Figure 2.5A, lanes 13-16).

Analysis of the supernatants from the Otu1-mediated release reaction showed that only a small percentage of the dissociating substrate molecules were completely deubiquitinated; the majority contained short ubiquitin chains with up to about 10 ubiquitin moieties (Figure 2.5B, lanes 8-10). Some portion of these released species probably rebind to the Cdc48 complex, as

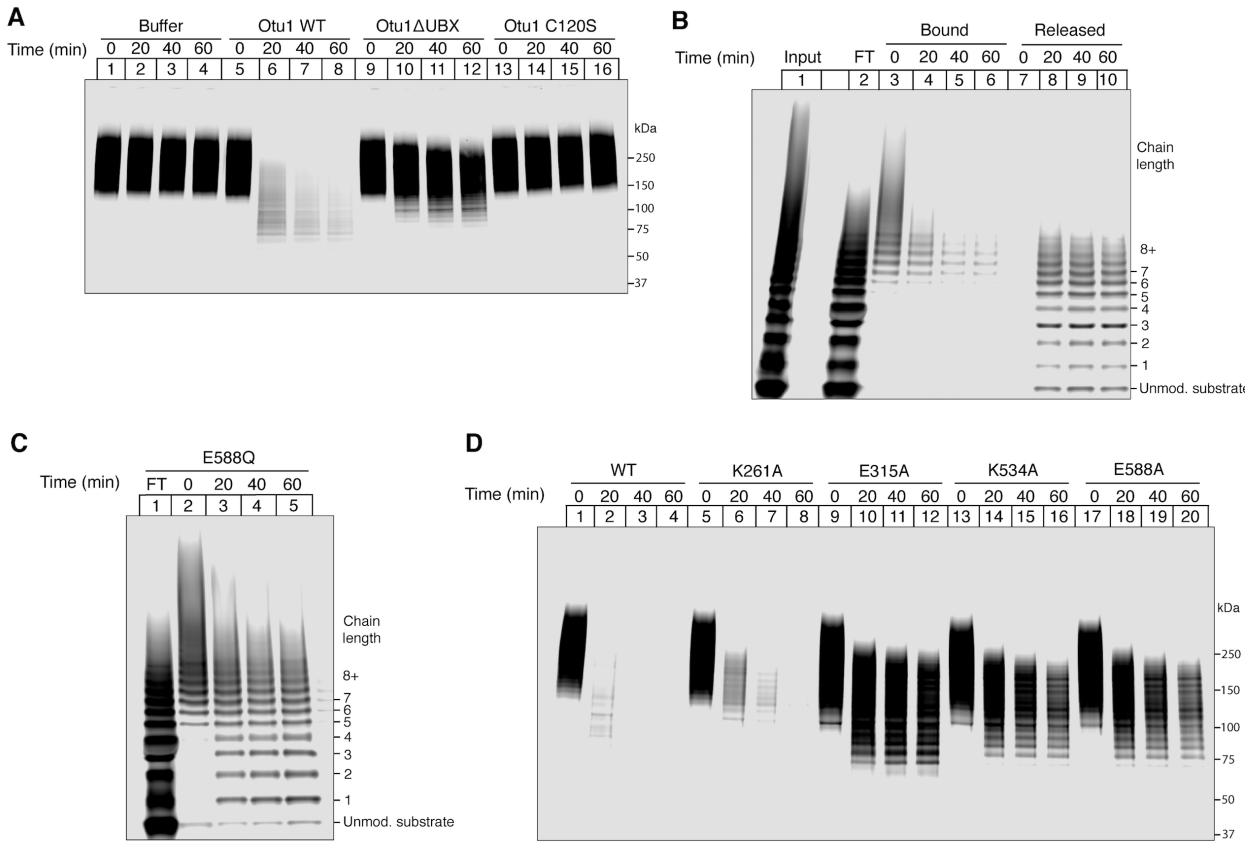



**Figure 2.5: Substrate release from the Cdc48 complex requires deubiquitination**

(A) Complexes of Cdc48, SBP-tagged UN, and dye-labeled, polyubiquitinated sfGFP were immobilized on streptavidin beads in the presence of ATP. After washing, the indicated Otu1 variants were added and samples of the bound material were analyzed at the indicated time points by SDS-PAGE and fluorescence scanning. WT, wild type; Otu1ΔUBX, Otu1 lacking the UBX domain; Otu1 C120S, catalytically inactive Otu1.

(B) As in (A), but with both bound and released material analyzed over time. Lane 1 shows the input material before incubation with streptavidin beads, and lane 2 the fraction that did not bind (flow-through; FT).

(C) As in (A), but with wild type Otu1 and the Walker B mutant E588Q in D2, which slowly hydrolyzes ATP.

**Figure 2.5 (Continued)**

(D) As in (C), but with other Walker mutations in D1 or D2. K261A, Walker A mutation in D1; E315A, Walker B mutation in D1; K534A, Walker A mutation in D2; E588A, Walker B mutation in D2.

they bear chains long enough to interact with UN (lanes 5 and 6). Interestingly, a similar pattern of oligoubiquitinated substrate molecules was retained by the Cdc48 complex when Cdc48 carried the Walker B mutation E588Q (Figure 2.5C), in contrast to the situation with wild type Cdc48, where Otu1 incubation resulted in essentially complete release of all substrate molecules (Figure 2.5B, lanes 3-6). Most of the ubiquitin chains retained by the mutant Cdc48 complex are too short to mediate the initial interaction with the complex (Figure 2.5C; compare lanes 1 and 3-5), suggesting that they represent pre-release intermediates. Thus, the slow ATPase rate in the D2 ring of the E588Q mutant must have delayed the release of Otu1-processed substrate molecules from the Cdc48 complex.

We next tested whether Otu1-mediated substrate release depends on the nucleotide bound to the D1 and D2 domains. A mutant in the Walker A motif of the D1 domain (K261A) allowed substrate release with similar kinetics as wild type Cdc48 (Figure 2.5D, lanes 5-8 versus 1-4). In contrast, a mutant in the Walker B motif of the D1 domain (E315A) had markedly slowed release kinetics (Figure 2.5D, lanes 9-12). This altered deubiquitination rate is not attributable to impaired assembly of the Otu1-Cdc48 complex, as Otu1 bound equivalently to wild-type Cdc48 and all Walker mutants (Figure 2.S5B). Instead, given that the D1 domain is in the ATP bound state in the Walker B mutant and the N domains therefore adopt the "up-conformation", Otu1 may not have full access to the ubiquitin chain. In the Walker A mutant, the N domain is likely in the “down- conformation”, allowing efficient deubiquitination by Otu1.

Both ATP binding and hydrolysis in the D2 domain are required for efficient substrate release by Otu1, as mutations in the Walker A (K524A) or B (E588A) motifs resulted in slow release kinetics (Figure 2.5D, lanes 13-16 and 17-20). These results suggest that a full cycle of ATP hydrolysis has to occur before D1 can hydrolyze ATP and return the N domains to the "down conformation", in which the ubiquitin chains are accessible to Otu1.

*Ubiquitin molecules can pass through the central pore*

The retention of one or more ubiquitin moieties on released substrate molecules raised the possibility that not only an unmodified substrate segment, but also the oligoubiquitin chain, can be translocated through the central pore. We therefore tested the path of free fluorescent polyubiquitin chains through the Cdc48 ATPase with photo-crosslinking experiments.

As with Ub(n)-sfGFP, free polyubiquitin chains crosslinked to the cis side of the D1 ring (position M288) in all nucleotide states (Figure 2.6A). Again, the interaction was dependent on the presence of the UN complex. Crosslinking to the interior and trans positions (D324 and D602, respectively) was observed with ATP, but not with ATPγS or ADP (Figures 2.6B and 2.6C). These results are comparable to those obtained with Ub(n)-sfGFP (Figure 2.3) and indicate that Cdc48 can translocate at least one ubiquitin molecule, even in the absence of an associated substrate or flexible peptide region. Interestingly, penta-ubiquitin crosslinked only weakly to the position at the entrance of the D1 ring and not at all to positions inside the central pore (Figure 2.6D), although it can maximally stimulate ATPase activity and must therefore be

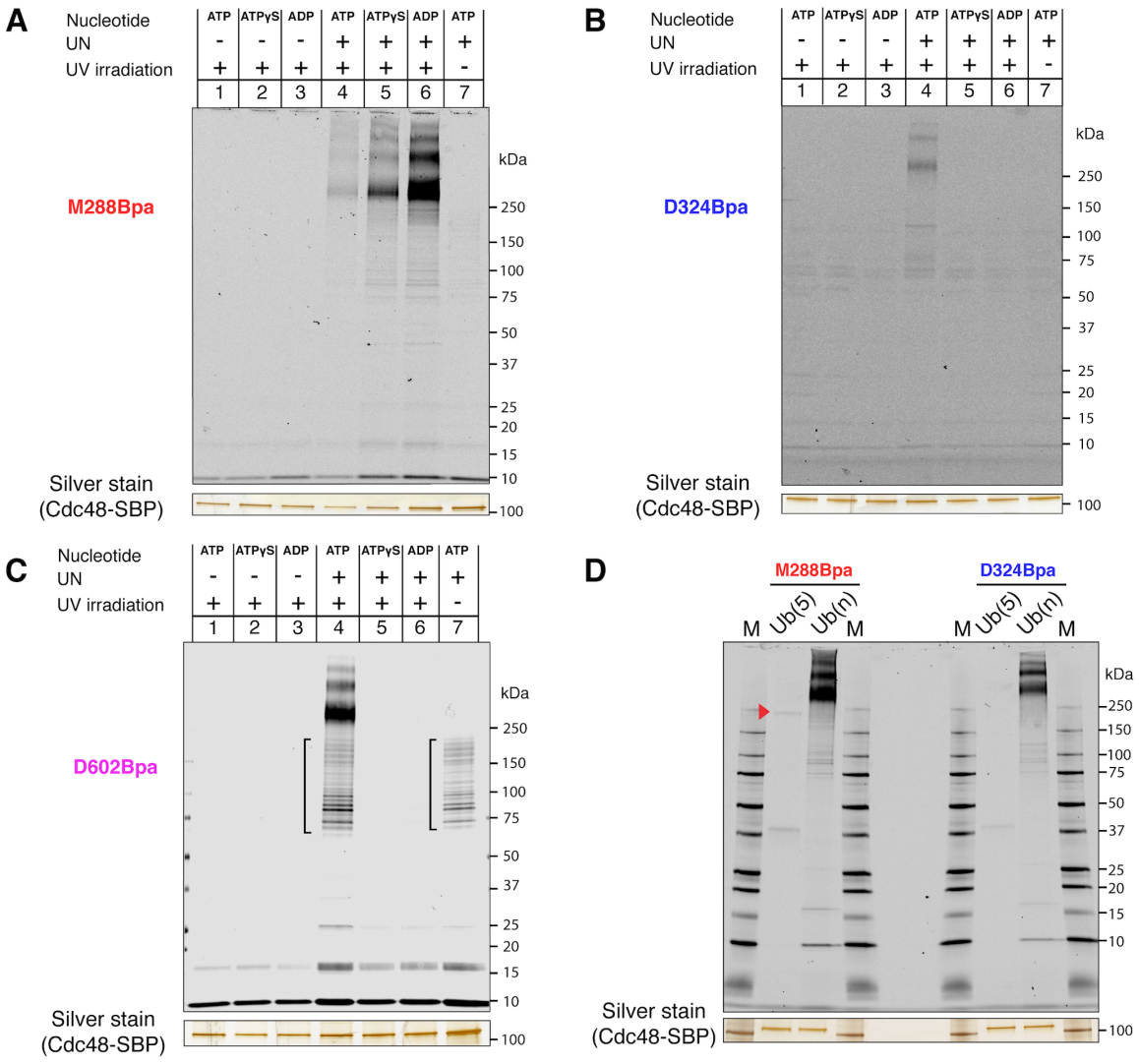



**Figure 2.6: Ubiquitin passes through the central pore of Cdc48**

(A) SBP-tagged Cdc48 (Cdc48-SBP) with Bpa at the M288 position (D1 pore loop) was incubated with dye-labeled free, K48-linked polyubiquitin chains and the indicated combinations of UN and nucleotide. After irradiation and streptavidin pulldown, crosslinked species were detected by SDS-PAGE and fluorescence scanning. The silver-stained band of Cdc48-SBP serves as a loading control.

(B) As in (A), but with Bpa at the D324 position in the interior of the central pore.

**Figure 2.6 (Continued)**

(C) As in (A), but with Bpa at the D602 position in the D2 pore-2 loop. Brackets, non-crosslinked material.

(D) As in (A), but with Bpa at positions M288 or D324 and either penta-ubiquitin ((Ub(5)) or polyubiquitin (Ub(n)). The arrowhead indicates a weak crosslink of Ub(5) to position 288. M, molecular weight markers.

able to bind to the Cdc48 complex (Figure 2.2C). Additional ubiquitin molecules in the chain seem to be required to allow efficient insertion of a polypeptide into the central pore of Cdc48.

## 2.4 Discussion

Our results clarify major aspects of the molecular mechanism of the Cdc48 ATPase. They lead to a model that can explain how Cdc48 and its mammalian ortholog p97/VCP function together with the conserved UN cofactor to disassemble protein complexes and extract proteins from membranes.

In the model, Cdc48 starts out with the D1 ATPases in the ADP-bound state (Figure 2.7, stage 1). The N domains are in the "down-conformation", co-planar with the D1 ring. When the D1 domain binds ATP, the N domains move upwards (stage 2). The UN complex can bind to either conformation, inhibiting the overall ATPase activity in the absence of substrate. Substrate is initially bound to the Cdc48 complex exclusively through an interaction of the attached K48-linked polyubiquitin chain with the UN cofactor (stage 3). Most of the UN cofactor in the cell is probably bound to Cdc48, which is present at high concentrations, so free UN would not compete for substrate. Substrate binding to the Cdc48 complex reduces ATPase activity in the D1 domain, biasing it toward its ATP-bound state with the N domain in its "up-conformation".

The UN cofactor and D1 ring ATPases might form a composite binding surface that can locally denature substrate without energy input, generating an unfolded polypeptide loop that can reach into the central pore. Substrate binding also stimulates ATP hydrolysis in the D2 domain. This activity allows pore loops in the D2 ring to move and drag the substrate polypeptide through the central pore (stage 4). The pulling force exerted by the D2 ATPases results in the unfolding of the substrate (stage 5). During translocation, most of the polyubiquitin moiety remains on the cis side, bound to the UN cofactor. However, a portion of the ubiquitin chain can enter the central pore along with the substrate (stage 5). The final step is substrate release. Once D1 has hydrolyzed ATP, the N domains convert back to the "down-conformation", allowing access of

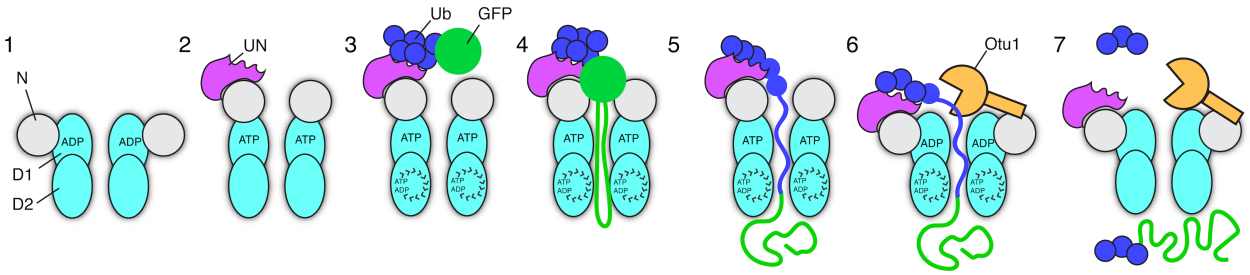



**Figure 2.7: Stages of substrate processing by the Cdc48 complex.** See text for details.

the polyubiquitin chain to the deubiquitinating enzyme Otu1 (stage 6). When the ubiquitin chain has been shortened sufficiently, its affinity for the UN complex is reduced or lost, and the remaining ubiquitin moieties are unfolded and pulled through the central pore (stage 7). The ubiquitin probably refolds rapidly after translocation[204], although this needs to be tested by future experimentation.

Our results argue against several alternative models proposed for Cdc48 function. We show that Cdc48 is an unfoldase that pulls the polypeptide substrate through the central pore, indicating that models of substrate segregation without unfolding are incorrect. Likewise, models in which the substrate inserts shallowly into only the D1 or D2 side of the pore are inconsistent

with the observed interactions along the full length of the pore. The crosslinking data also invalidate models in which relative motions of the N domains displace substrates or substrates interact with the exterior of the ATPase.

Our data indicate that the D1 domain only needs to hydrolyze ATP a few times, or perhaps even just once, during the processing of a given substrate molecule, whereas the D2 domain hydrolyzes ATP many times. Consistent with this model, only D2 contains the canonical aromatic pore loop residues implicated in substrate binding and translocation. However, the two ATPase rings likely communicate with each other. ATP hydrolysis in D1 not only controls the position of the N domains, but also affects D2 activity, as substrate binding at the cis side stimulates the ATPase rate of D2. Inter-ring communication may also occur in the reverse direction, as our substrate release data indicate that the nucleotide state of D2 influences the activity of Otu1, which is bound to the N domain. It is also likely that the ATPase subunits influence one another within each ring when the subunits are in different nucleotide states. Because all members of the ring are forced to be in the same state in the presence of ADP or ATPγS, or when Walker motifs are mutated, our model is likely a simplification; in the presence of ATP, only some of the six subunits of the D1 or D2 rings may undergo the described nucleotide binding and hydrolysis events in a synchronous manner.

Several findings support the proposition that not only substrate but also some ubiquitin moieties are translocated by Cdc48 and released from the D2 side of the central pore. First, ubiquitin itself can serve as a translocation substrate, forming crosslinks to positions along the full length of the pore. Thus, at least one ubiquitin molecule completely traverses the double-ring ATPase. Second, a slowly hydrolyzing Cdc48 mutant retained oligoubiquitin chains, indicating that these are slowly translocated through the ATPase and released with delay. Third, fully

released substrates are primarily oligoubiquitinated, rather than fully deubiquitinated. The last two points suggest that not only a single ubiquitin molecule, but also a short ubiquitin chain can be translocated through the central pore. We attempted to confirm ubiquitin translocation by incubating Ubn-Eos with Cdc48-FtsH but were unable to reproducibly detect ATP-dependent peptides by mass spectrometry, perhaps because FtsH does not efficiently cleave ubiquitin. It should be noted that the proteasome can also translocate ubiquitin under a variety of circumstances[205].

A translocation mechanism for Cdc48/p97 had previously been dismissed because the ATPase has a narrow central pore relative to other double-ring AAA ATPases of known structure, and its D1 pore is even occluded by a $Zn^{2+}$ ion in a crystal structure[161]. However, our data now suggest that at least two polypeptide strands can be accommodated, as the ubiquitin attachment site on the substrate is likely moved into the pore. Our results thus imply that the pore diameter widens during translocation, although this has yet to be confirmed experimentally. A hairpin structure inside the central pore has been demonstrated for other AAA ATPases, and there is even evidence that three strands can be present[206]. This mechanism differs from that of NSF, where the SNARE substrate is not translocated through the central pore, although the D1 ring has canonical pore residues[191], and a single ATPase cycle disassembles the SNARE complex[207].

In our model, Cdc48 activity is coupled with substrate deubiquitination. Indeed, free Otu1 has much lower deubiquitination activity than Otu1 bound to the Cdc48 complex[83]. This stimulation is in part due to the UN cofactor, which may present bound K48-linked ubiquitin molecules to Otu1 in an appropriate conformation. In addition, our data suggest that deubiquitination is delayed until the substrate is at least partially translocated. The slow

deubiquitination and substrate release exhibited by the D1 Walker B mutant indicate that Otu1 may not have full access to the UN-associated ubiquitin chain while the D1 domains are in the ATP-bound state and the N domains in the "up-conformation". Movement of the N domain to the "down-conformation" by ATP hydrolysis in D1 would relieve this inhibition and permit full deubiquitination. Efficient deubiquitination also requires ATP binding and hydrolysis in D2, ensuring that substrate translocation commences before Otu1 can act. Furthermore, the repression of D1 ATP hydrolysis by substrate gives the D2 ring a chance to translocate a substantial portion of the polypeptide before deubiquitination occurs. Taken together, these mechanisms allow substrate binding, translocation, and release to occur in a defined order. Our results show that Otu1 cooperates with Cdc48 and argue against an earlier model in which Otu1 antagonizes Cdc48 function by preventing substrate recognition[68].

The overall mechanism of the Cdc48 complex resembles that of the 19S regulatory subunit of the proteasome, which also uses receptor proteins to bind polyubiquitin chains attached to a substrate and employs a translocation mechanism (for review, see[208]). As with Cdc48, full substrate movement through the central pore requires deubiquitination, a reaction performed by Rpn11 of the 19S subunit. In contrast to Otu1, which cleaves between ubiquitin moieties and does not fully remove a substrate-associated ubiquitin chain, Rpn11 cleaves off the entire chain. However, in both cases, deubiquitination cooperates with, rather than antagonizes, substrate processing by the ATPase complex. This is distinct from the function of other DUBs, such as the Ubp6 component of the 19S subunit, which serve as negative regulators.

In many cellular functions, such as ERAD, Cdc48 acts upstream of the proteasome. Cdc48 may be required in cases where a substrate does not have a flexible segment[50], which is needed to initiate translocation into the 19S proteasomal subunit[39]. Indeed, in our experiments,

the Cdc48 complex can translocate a polyubiquitin chain, which does not expose extended flexible segments. The exact reason why Cdc48, but not the proteasome, can deal with folded substrates remains to be clarified. The mechanism of substrate transfer from Cdc48 to the proteasome also needs further investigation. Most of the ubiquitin chains released by Cdc48 would be long enough to bind directly, or through the shuttling factors Rad23 or Dsk2, to the proteasome [91,209], and chains that are too short could first be extended by the Cdc48-associated "E4 ligase" Ufd2[193]. Alternatively, some substrates may be transferred from Cdc48 into the 20S proteasome without the involvement of the 19S subunit[188,189]. Regardless of the downstream events, our results show that Cdc48 pulls on substrate polypeptides and unfolds them, explaining how the ATPase complex disassembles protein complexes and extracts proteins from membranes.

**2.5 Methods**

*Experimental models*

Recombinant proteins, with the exception of Uba1 and Ubr1, were expressed in *Escherichia coli* BL21 (DE3) or BL21 (DE3) RIPL cells grown in Terrific Broth. Uba1 was expressed in *Saccharomyces cerevisiae* INVSc1, and Ubr1 was expressed in *Saccharomyces cerevisiae* BY4741; both strains were grown in YPD medium.

*Purification of proteins*

Uba1 was purified as described[83]. Ubr1 was expressed in *S. cerevisiae* from the plasmid pFlagUBR1SBX, a gift from Alexander Varshavsky. The protein was purified by FLAG affinity chromatography from the yeast strain BY4741 according to a previously established protocol [210].

Cdc48, Otu1, and Ufd1/Npl4 were purified as described [83]. Point mutants, truncated constructs, SBP-tagged variants, and the Cdc48-FtsH chimera were purified by the procedures

for their parent constructs, with the exception of constructs incorporating Bpa (see below).

Ubc1, Ubc2, sfGFP, mEos3.2, the FtsH proteolytic domain (residues 406-634 of the *Aquifex aeolicus* sequence) and Cys-ubiquitin (*S. cerevisiae* ubiquitin with a cysteine introduced at position 1) were expressed in *E. coli* BL21 DE3 RIPL cells with N-terminal His14-SUMO fusion tags [211]. Cells were grown in Terrific Broth to an OD600 of 1. Protein expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 hrs at 18ºC. Cells were harvested by centrifugation at 4,000 x g for 10 min and resuspended in lysis buffer (50 mM Tris pH 8, 200 mM NaCl, 30 mM imidazole). Phenylmethylsulfonyl fluoride (PMSF; 1 mM), a protease inhibitor cocktail, and DNase I (5 μg/mL) were added, and cells were lysed by sonication. Lysates were cleared by ultracentrifugation in a Ti45 rotor (Beckman) at 40,000 rpm for 30 min at 4ºC. Supernatants were incubated with Ni-NTA resin for 90 min at 4ºC. The resin was washed three times with 20 column volumes of buffer (50 mM Tris pH 8, 200 mM NaCl, 30 mM imidazole). Proteins were eluted twice with 10 mL elution buffer (50 mM Tris pH 8, 100 mM NaCl, 400 mM imidazole), and the eluates were combined and supplemented with 1 mM tris(2-carboxyethyl)phosphine (TCEP). His14-SUMO tags were removed by incubation with 2 μM Ulp1 (SUMO protease) for 2 hrs at 4ºC. Ubc1, Ubc2, sfGFP, and the FtsH protease were dialyzed against 50 mM Tris pH 8, 100 mM NaCl, and mEos3.2 was dialyzed against 50 mM HEPES pH 7, 50 mM NaCl. Proteins were loaded onto ion exchange columns (MonoQ 10/100 GL for Ubc1, Ubc2, sfGFP, and the FtsH protease; MonoS 10/100 GL for mEos3.2) equilibrated in the respective dialysis buffers, and eluted by linear gradients to 500 mM NaCl over 10 column volumes. Peak fractions were pooled, concentrated to 2 mg/mL or greater, snap frozen in liquid nitrogen, and stored at -80ºC.

Cdc48 constructs for crosslinking experiments were expressed in *E. coli* BL21 DE3 cells

harboring the plasmid pEVOL-pBpF, a gift from Peter Schultz (Addgene plasmid #31190) [212]. Cells were grown in terrific broth without glycerol to an OD of 0.5. For induction, arabinose (0.02%), IPTG (0.1 mM), and p-benzoylphenylalanine (Bpa; 1 mM) were added and expression was carried out for 16 hrs at 18ºC. Cell harvesting, lysis, and Ni-NTA purification was performed as described above, with 5 mM $MgCl_2$ included in buffers throughout. After Ulp1 cleavage, proteins were diluted to an imidazole concentration of < 200 mM and applied to Ultra HBC streptavidin agarose resin (Gold Bio) for 2 hrs at 4ºC. The resin was washed 3 times with 20 column volumes of buffer S (50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM MgCl2). Proteins were eluted with 5 column volumes of buffer S containing 5 mM biotin, then dialyzed overnight (to remove biotin) against buffer S containing 0.5 mM TCEP. The protein solutions were supplemented with 10% glycerol, concentrated to 2 mg/mL or greater, snap frozen in liquid nitrogen, and stored at -80ºC.

*S. cerevisiae* ubiquitin and *H. sapiens* ubiquitin chains of defined lengths were purchased from Boston Biochem.

The sequence inserted at the N terminus of sfGFP or mEos3.2 to facilitate ubiquitination by the N-end rule enzymes is as follows (note that the N-terminal arginine is exposed after SUMO cleavage): RHGSG(C/S)GAWLLPVSLVKRKTTLAPNTQTASPPSYRALADSLMQ. For labeling of Eos, which already contains an exposed cysteine, Cys6 was changed to Ser to avoid dual labeling.

*Labeling with fluorescent dyes*

Prior to labeling, sfGFP, Eos, Cys-ubiquitin, and penta-ubiquitin chains were exchanged into buffer free of reducing agents (50 mM HEPES pH 7.5, 150 mM NaCl) by dialysis or PD-10 desalting columns. Proteins were incubated with a 5:1 molar excess of DyLight 800 maleimide,

DyLight 680 maleimide, or Dylight 680 NHS ester for 1 hr at room temperature. Reactions were quenched by addition of 1 mM DTT (for maleimides) or 100 mM Tris (for NHS esters). Samples were then dialyzed or treated with dye-removal columns (Thermo Scientific).

*Photoconversion of mEos3.2*

The Eos protein (2 mg/mL) was placed in a 1.5 mL Eppendorf tube in an ice bath. A long-wave UV lamp (Blak-Ray) was positioned 5 cm from the tube, and the sample was irradiated for 2 hrs, with occasional mixing. Any precipitated protein was removed by filtration.

*Ubiquitination of substrates*

Ubiquitination of sfGFP or mEos3.2 was carried out as follows. Substrates (500 nM) were incubated with *S. cerevisiae* ubiquitin (100 μM), Uba1 (100 nM), Ubc2 (12 μM), Ubr1 (300 nM), and ATP (5 mM) for 60 to 90 min at 30ºC in ubiquitination buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$). Reaction mixtures were applied to Ultra HBC streptavidin agarose beads for 30 min at room temperature. The resin was washed 3 times with 20 column volumes of ubiquitination buffer, then resuspended in 5 column volumes of ubiquitination buffer containing 1 mM TCEP. Substrates were cleaved from the resin by incubation with 2 μM 3C protease for 2 hrs at room temperature. Substrates used for binding and release experiments, ATPase stimulation assays, and crosslinking were collected and snap frozen at this point.

Substrates used for unfolding and degradation experiments were concentrated after 3C protease cleavage to < 500 μL and applied to a gel filtration column (S200 10/300 GL) in ubiquitination buffer. Fractions were collected, analyzed by SDS-PAGE, supplemented with 50% glycerol, and stored at -20ºC.

*Synthesis of free polyubiquitin chains*

Yeast ubiquitin (200 μM) was incubated with Uba1 (100 nM), Ubc1 (5 μM), and ATP (5

mM) for 2 hrs at 30ºC in ubiquitination buffer. To generate fluorescently labeled chains, DyLight 680-conjugated Cys-ubiquitin was included at 5 μM. ATP was removed by dialysis into nucleotide-free ubiquitination buffer, and chains were snap frozen and stored at -80ºC.

*Binding and release assays*

SBP-tagged Cdc48 or UN (200 nM) was applied to 10 μL magnetic streptavidin beads (Pierce) in a total of 250 μL buffer B (50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, 0.5 mg/mL protease-free BSA, 2 mM nucleotide [ATP, ATPγS, or ADP]). When applicable, untagged Cdc48, UN, or Otu1 was included at 500 nM. Binding was carried out for 30 min at room temperature, followed by three washes and resuspension in 250 μL buffer B. Polyubiquitinated substrate or polyubiquitin chains (200 nM) were added and the binding reaction continued for another 30 min. Supernatants were recovered, and beads were again washed 3 times and eluted with BSA-free buffer B containing 5 mM biotin.

For Otu1 release assays, after the second set of washes, the beads with bound Cdc48, UN, and substrate were resuspended in 250 μL buffer B with ATP. Otu1 or its variants were added at 400 nM at time 0, and 50 μL of the reaction mix, including beads, was removed for each time point. Supernatants were recovered from the 50 μL samples, and beads were washed 3 times and eluted with buffer B containing 5 mM biotin. Supernatants and eluates were subjected to SDS-PAGE and scanned on an Odyssey imager (LI-COR).

The competition experiment with unlabeled substrate was performed as described for the Otu1 release assays, except that instead of Otu1, unlabeled substrate amounting to a 5-fold excess (1 μM) over the original fluorescently labeled input was added.

*ATPase assays*

ATPase rates were measured using an absorbance-based phosphate release assay

(EnzChek, Thermo Fisher). Assays were performed at 30ºC in ATPase buffer (50 mM HEPES pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 0.5 mM TCEP, 0.1 mg/mL protease-free BSA). As appropriate, reaction components included Cdc48 or its mutants (150 nM), Ufd/Npl4 (500 nM), and substrates, ubiquitin, or polyubiquitin (2.5 μM; the concentration of mixed polyubiquitin chains was calculated on the basis of an average chain length of 12). Proteins were pre-incubated for 10 min prior to the addition of 2 mM ATP. Absorbance (360 nm) was measured at 20 sec intervals in an M5 plate reader (Spectramax). The slope of the initial, linear portion of the curve was used to calculate the rate, and rates were normalized to that of wild type Cdc48. ATPase rates are reported as the average of three replicates, and error bars show one standard deviation.

*Unfolding assays*

Unfolding experiments were performed at 30ºC in reaction buffer (50 mM HEPES pH 7.5, 130 mM KCl, 10 mM $MgCl_2$, 0.5 mM TCEP, 0.5 mg/mL protease-free BSA). As appropriate, reaction components included Cdc48 or its mutants (200 nM), Ufd/Npl4 (500 nM), and polyubiquitinated mEos3.2 or sfGFP (50 nM). Proteins were pre-incubated for 10 min prior to the addition of an ATP regeneration system (2 mM ATP, 20 mM phosphocreatine, 100 μg/mL creatine kinase), ATPγS (2 mM), or ADP (2 mM). Fluorescence (excitation: 485 nm for GFP, 540 nm for Eos; emission: 516 nm for GFP, 580 nm for Eos) was measured at 30 sec intervals in an M5 plate reader (Spectramax) for 30 min. Fluorescence values were corrected by subtracting the measured fluorescence of polyubiquitinated mEos3.2 or sfGFP denatured in 6M guanidine-HCl.

*Homology modeling*

A structural model of yeast Cdc48 was calculated based on the p97 crystal structure (PDB: 3CF1) using the program SWISS-MODEL [213]. The structure was displayed using PyMol

(Schrödinger, LLC).

*Crosslinking assays*

Crosslinking was performed in reaction buffer (50 mM HEPES pH 7.5, 130 mM KCl, 10 mM $MgCl_2$, 0.5 mM TCEP, 0.5 mg/mL protease-free BSA). As appropriate, reaction components included Cdc48 or its mutants (200 nM), Ufd1/Npl4 (500 nM), and dye-labeled polyubiquitinated sfGFP or free polyubiquitin (1 μM). An ATP regeneration system (2 mM ATP, 20 mM phosphocreatine, 100 μg/mL creatine kinase), ATPγS (2 mM), or ADP (2 mM) were added. Reactions were assembled on ice, incubated at 30ºC for 10 min, and transferred to individual wells of a black polystyrene plate. A long-wave UV lamp (Blak-Ray) was positioned 5 cm from the plate, and the samples were irradiated for 30 min. To prevent overheating, an ice-cold metal block was placed in contact with the bottom of the plate.

After irradiation, samples were diluted 10-fold in dissociation buffer (50 mM Tris pH 8, 800 mM KCl, 1% (v/v) Triton X-100, 1 mM EDTA, 0.5 mM TCEP) and incubated at room temperature for 5 min. Samples were then applied to 5 μL magnetic streptavidin beads (Pierce) equilibrated in dissociation buffer for 30 min at room temperature. Beads were washed three times, then eluted in 50 mM HEPES pH 7.5, 100 mM NaCl, 5 mM biotin. The eluted material was subjected to SDS-PAGE and the gel scanned on an Odyssey imager (LI-COR). For the experiment in Figure 2.S3, the dissociation and pulldown steps were omitted, and only 200 nM substrate was used.

*Degradation assays*

Experiments were performed in reaction buffer plus zinc (50 mM HEPES pH 7.5, 130 mM KCl, 10 mM $MgCl_2$, 0.5 mM TCEP, 0.5 mg/mL protease-free BSA, 200 nM $ZnCl_2$). As appropriate, reaction components included the Cdc48-FtsH chimera (200 nM) or a mixture of

wild-type Cdc48 and the free FtsH protease domain (200 nM each), Ufd1/Npl4 (500 nM), and polyubiquitinated, dye-labeled Eos (50 nM) that had not been photoconverted. An ATP regeneration system (2 mM ATP, 20 mM phosphocreatine, 100 μg/mL creatine kinase), ATPγS (2 mM), or ADP (2 mM) were added. Reactions were assembled on ice, incubated for 10 min, then transferred to 30ºC. Aliquots were removed at appropriate time points, followed by addition of Otu1 (2.5 μM) and further incubation for 20 min, after which the final aliquot was removed. Samples were subjected to SDS-PAGE and scanned on an Odyssey imager (LI-COR). For mass spectrometry analysis, the deubiquitination step was omitted. Instead, reactions were quenched after 20 min by the addition of 50 mM EDTA and submitted for peptide identification without further processing.

*Quantification and statistical analysis*

Bar graphs shown in Figure 2.2 and Figure 2.S2 display mean ± standard deviation of three experiments.

**Acknowledgements**

We thank Alex Stein for providing reagents, the Institute for Chemistry and Chemical Biology Longwood for use of a plate reader, and Dan Finley, Alex Stein, and Ryan Baldridge for critical reading of the manuscript. This work is supported by the NIH/NIGMS Award R01GM052586 and by award #T32GM007753 from the NIGMS. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. T.A.R. is a Howard Hughes Medical Institute Investigator.

## Chapter 3: Structure of the Cdc48 ATPase with its ubiquitin-binding cofactor Ufd1/Npl4

Attributions: I was responsible for purifying the proteins used in this study, growing crystals, collecting X-ray diffraction data, solving the crystal structure, and performing biochemical and yeast experiments. Hyojin Kim, a postdoctoral fellow in the laboratory of Tom Walz, was responsible for cryo-EM data collection and processing and for producing the cryo-EM structures. Vladimir Svetlov, a scientist in the laboratory of Evgeny Nudler, collected crosslinking mass spectrometry data, and Thomas Wales, a scientist in the laboratory of John Engen, collected hydrogen-deuterium exchange mass spectrometry data. Zhejian Ji, a postdoctoral fellow in the Rapoport lab, collected data relating to the Cdc48 triphenylalanine mutants.

This chapter is in preparation for submission as:

**Structure of the Cdc48 ATPase with its ubiquitin-binding cofactor Ufd1/Npl4**

**Nicholas O. Bodnar[#], Kelly H. Kim[#], Zhejian Ji, Thomas E. Wales, Vladimir Svetlov, Evgeny Nudler, John R. Engen, Thomas Walz*, and Tom A. Rapoport***

**[#] These authors contributed equally and are listed alphabetically.**

*** Corresponding authors**

### 3.1 Abstract

Many poly-ubiquitinated proteins cannot be directly degraded by the proteasome because they are well folded or located in membranes, chromatin, or multimeric complexes. These proteins are generally extracted from such assemblies and unfolded by an ATPase, called Cdc48 in yeast and p97 or VCP in mammals, before being transferred to the proteasome[214]. Cdc48/p97 contains two stacked hexameric ATPase rings (D1 and D2) and an N-terminal (N) domain that sits next to or above the D1 ring[54]. Cdc48/p97 binds various cofactors, including a heterodimer of Ufd1 and Npl4 (UN complex), which engages the ATPase in many cellular processes[71]. Npl4 is a target of the

potential cancer drug disulfiram[215]. Here, we report cryo-electron microscopy (cryo-EM) and crystal structures that clarify the interaction of the UN cofactor with the Cdc48 ATPase. Several major contacts are mediated by Npl4, which interacts through its UBX-like domain with one of the N domains of Cdc48, and through two $Zn^{2+}$-finger domains with the top of the D1 ATPase ring. The $Zn^{2+}$-fingers anchor the MPN domain of Npl4 above the central pore of the ATPase ring, similarly to how the MPN domain of the de-ubiquitinase Rpn11 sits in the regulatory 19S subunit of the proteasome[216]. Our results indicate that Npl4 is unique among the Cdc48/p97 cofactors, and suggest a mechanism for how poly-ubiquitinated substrates are bound to and moved into the ATPase ring.

**3.2 Results and discussion**

Npl4 is an essential and conserved cofactor of Cdc48/p97[71]. It contains an N-terminal UBX-like domain and is predicted to have $Zn^{2+}$-finger (zf-Npl4) and MPN domains (Figure 1A). MPN domains are found in several $Zn^{2+}$-dependent isopeptidases, including AMSH/AMSH-LP, the COP9 signalosome subunit CSN5, and the proteasomal de-ubiquitinase (DUB) Rpn11[217-219]. In many cellular functions, including ER-associated protein degradation (ERAD), Npl4 associates with Ufd1, a protein with a ubiquitin-binding UT3 domain that shares structural homology with the N domain of Cdc48 (Figure 3.1a)[220]. Ufd1 interacts with Npl4 through its UT6 domain, a short segment predicted to be unstructured[221]. Alternatively, Cdc48 and Npl4 can interact with Vms1, which then recruits the ATPase complex to mitochondria[124].

All known cofactors of Cdc48 bind to either the N domain or the unstructured C-terminal tail[58]. Ufd1 and Npl4 bind to the N domain through two short SHP motifs and the UBX-like domain, respectively (Figure 3.1a)[73]. These interactions do not explain how a substrate would enter the central pore of the ATPase, raising the possibility that Ufd1 and Npl4 have additional

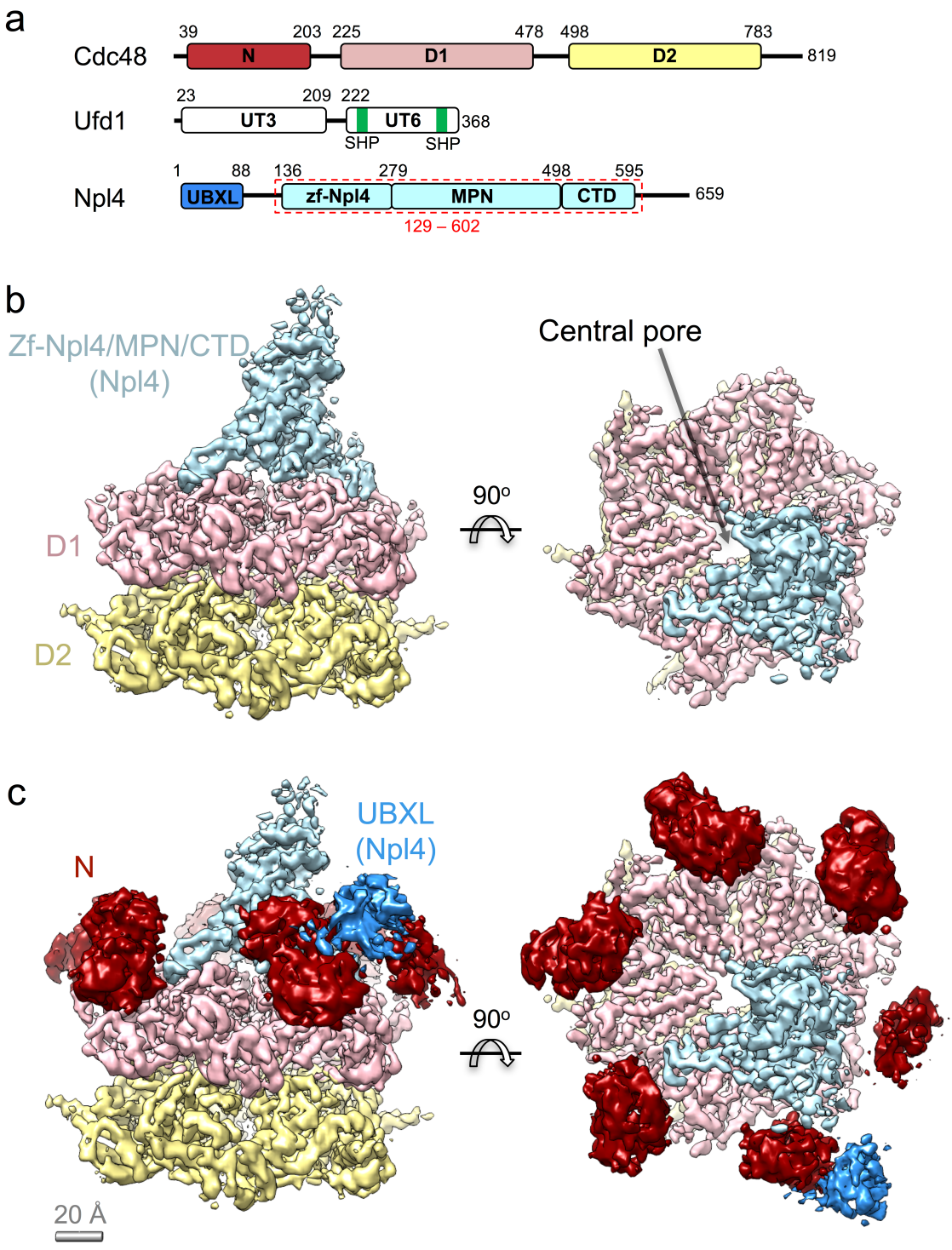



**Figure 3.1: Structure of the Cdc48/UN complex with ATPγS. a**, Domain organization of Cdc48 and its co-factors, Ufd1 and Npl4. The Npl4 region indicated with a red dashed box was crystallized. **b**, Cryo-EM density map of the Cdc48/UN complex, colored as in **a**. The N domains

**Figure 3.1 (Continued)**

were masked out in the final refinement step. **c**, The N domains of Cdc48 (dark red) from a map refined without a mask are shown relative to the map obtained with masking. The contiguous extra density next to one of the N domains (blue) is assigned to the UBX-like domain of Npl4.

binding sites on Cdc48/p97. Previous EM reconstructions of the ATPase complex showed electron density for the UN cofactor near the N domains, but the resolution was insufficient to derive molecular models[222,223].

We first determined cryo-EM structures of Cdc48 alone. To this end, Cdc48 from the thermophilic fungus *Chaetomium thermophilum* was expressed in *E. coli* and purified as a hexamer (Figure 3.S1a,b). Structures of Cdc48 were determined in the presence of ADP or ATPγS. After 3D classification and refinement, structures at 7.2 Å and 8.2 Å overall resolution were obtained (Figure 3.S2, 3.S3, 3.S6, 3.S7). As reported for mammalian p97[55], both structures showed stacked D1 and D2 ATPase rings with the N domain in the down-conformation in the ADP-bound state, and in the up-conformation in the ATPγS-bound state (Figure 3.S2, 3.S3). ATP binding to the D1 ring likely triggers this conformational switch[166]. It should be noted that in these structures all subunits of the ATPase were forced into the same nucleotide state, whereas during ATP hydrolysis, some N domains may be in the up- and others in the down-conformation.

Next, we purified a complex of Cdc48 and UN. The UN complex from *Chaetomium thermophilum* was again expressed in *E. coli* and had the expected 1:1 stoichiometry after purification (Figure 3.S1c). A complex of hexameric Cdc48 and UN (Figure 3.S1d) was

subjected to single-particle cryo-EM analysis in the presence of ADP or ATPγS (Figure 3.S4-7). The refined structures had overall resolutions of 6.7 Å and 4.3 Å, respectively. The presence of the cofactor had only a small effect on the structure of the ATPase rings (Figure 3.S4-7). However, even in the ADP-bound state, a sizable population of the Cdc48 molecules had their N domain in the up-conformation, although the percentage was lower than in ATPγS (58% versus 93%). Thus, ATP and cofactor binding act synergistically to move the N domains into the up-conformation, a state likely required for initiation of substrate processing[224].

The most obvious density contributed by the cofactor complex is a central tower that lies directly above the central pore (Figure 3.1b). In addition, some cryo-EM classes show density close to one of the N domains, which can be attributed to the UBX-like domain of Npl4; this domain is known to bind to this side of the N domain of Cdc48[169] (Figure 3.1c). Deletion of the UBX-like domain abolishes binding of Npl4 to Cdc48[73], indicating that the cofactor domains constituting the tower density interact only weakly with the ATPase. The central location of the tower excludes the binding of a second cofactor molecule, explaining why one Cdc48 hexamer binds only one UN heterodimer. Although the tower occupies a large fraction of the space immediately above the D1 ring, the central pore remains unobstructed (Figure 3.1b), thus allowing substrate to move into it.

Although the density map of the Cdc48/cofactor complex permitted the visualization of helices, the resolution was insufficient to build a molecular model for the cofactor. We therefore first identified cofactor regions that are in close proximity to the D1 ATPase ring. The Cdc48/UN complex was treated with a bifunctional lysine crosslinker, the sample was digested with protease, and crosslinked peptides were identified by mass spectrometry. The data showed that the N terminus of Ufd1 interacts promiscuously with multiple locations in the ATPase and

Npl4 (Figure 3.S8a-b), suggesting that the succeeding UT3 domain is flexible. On the other hand, several lysines in the zf-Npl4 and MPN domains of Npl4 crosslinked specifically to residues on the surface of the D1 ATPase ring (Figure 3.S8a-b). Thus, these domains were the best candidates to form the central tower.

Using limited proteolysis, we found that the zf-Npl4 and MPN domains form a stable fragment (Figure 3.S8c). A crystal structure of this construct was determined using the central tower density of the cryo-EM map as a molecular replacement model (data collection and refinement statistics in Table 3.S1). The crystal structure indeed fits well into the cryo-EM map (Figure 3.2a-b), indicating that the Npl4 domains undergo only small changes upon Cdc48 binding. The bottom of the tower is formed by the zf-Npl4 domain, the central portion by the MPN domain, and the top portion by a C-terminal domain (CTD) of five **α**-helices (Figure 3.2a,b). A small region of density remained unexplained and likely corresponds to a segment of UT6 in Ufd1 (Figure 3.2a). Indeed, hydrogen/deuterium (H/D) exchange experiments showed that several Npl4 peptides in this region were protected when Ufd1 was present (Figure 3.S8d). The Npl4-interacting region of UT6 is likely located between the two SHP motifs that anchor Ufd1 to the N domains of Cdc48[73]. The UT3 domain of Ufd1 is not visible in the density map, confirming that it is flexible. Interestingly, a domain with the same fold is also flexible in the Pex1/Pex6 ATPase[225], suggesting that these domains might only be fixed when they bind ubiquitinated substrate.

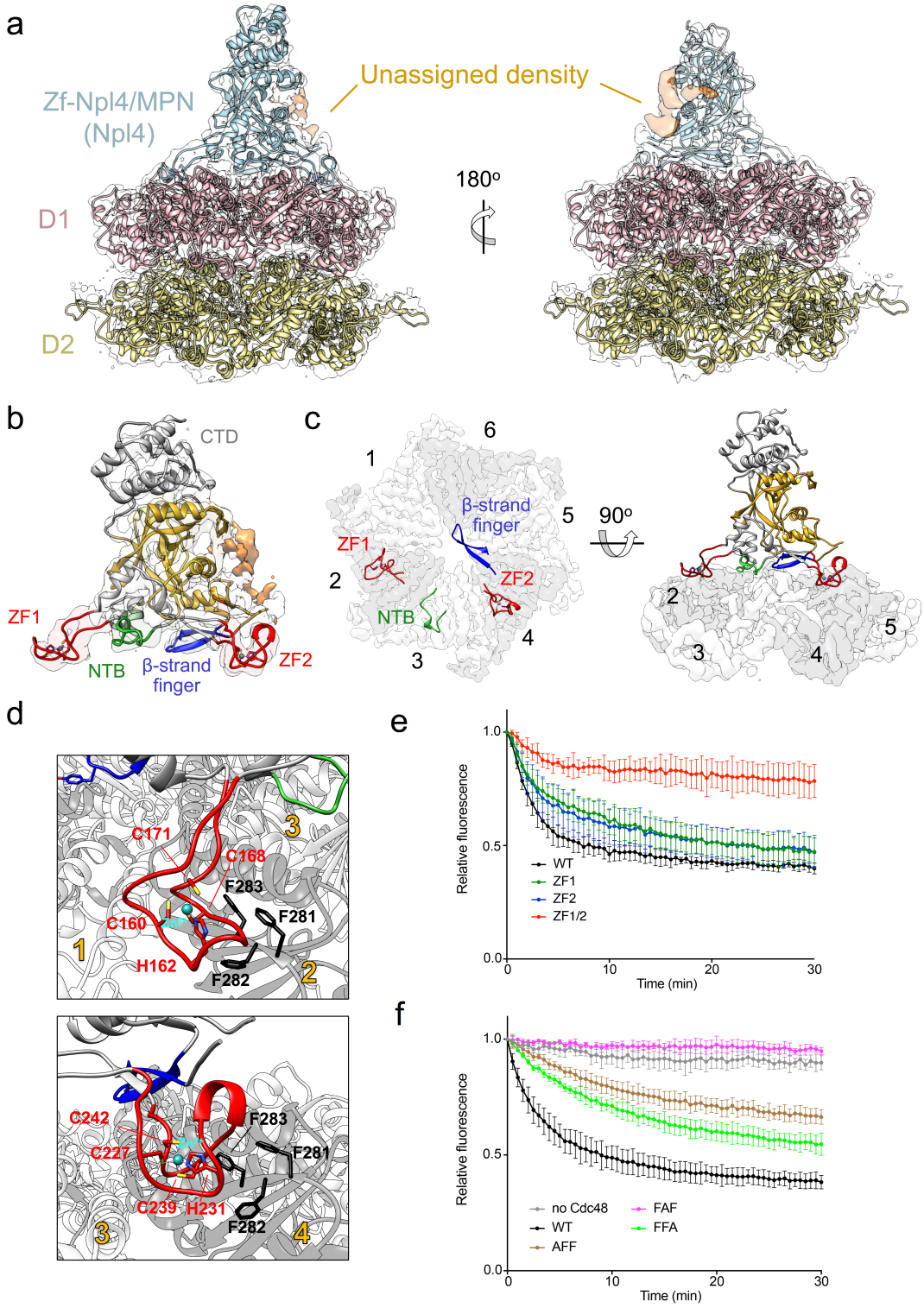



**Figure 3.2: Interactions between Cdc48 and its cofactor. a**, A crystal structure of Npl4 (light blue) and a homology model of Cdc48 (D1 in pink; D2 in yellow) were docked into the cryo-EM

**Figure 3.2 (Continued)**

map. Unassigned density (orange) likely belongs to parts of the UT6 domain of Ufd1. **b**, Close-up view of the Npl4 crystal structure docked into the cryo-EM map. The MPN domain is shown in yellow and D1-interacting regions extending from it are highlighted: the two $Zn^{2+}$-fingers (ZF; red), an N-terminal bundle (NTB; light green), and the 'β-strand finger' (blue). **c**, Top view of D1-interacting regions, colored as in **b**, with the D1 ring shown as a white/gray surface. For clarity, D2 was omitted. ATPase subunits of the Cdc48 hexamer are numbered. **d**, Close-up views of the $Zn^{2+}$-finger, with $Zn^{2+}$-coordinating residues in stick representation. The interacting tri-phenylalanine (FFF) sequence in Cdc48 is highlighted. **e**, Unfolding of poly-ubiquitinated Eos by wildtype *S. cerevisiae* Cdc48 and the indicated UN variants. ZF1: H139A/C145A. ZF2: H208A/C216A. ZF1/2: H139A/C145A/H208A/C216A. Data are shown as mean ± SD of n=3 technical replicates. **f**, As in **e**, but with FFF variants of Cdc48 and n=4 technical replicates.

The MPN domain is anchored to the top of the D1 ATPase rings via the preceding $Zn^{2+}$-finger domains, which are both of the CHCC type, i.e. use His and Cys for coordination of the $Zn^{2+}$ ion (Figure 3.2b-d). The $Zn^{2+}$-fingers form two "stalks" that project into grooves between the D1 ATPase subunits (Figure 3.2b,c). When numbered from the position of the N-terminal $Zn^{2+}$-finger, the interacting grooves are between ATPase subunits 1 and 2 and between subunits 3 and 4 (Figure 3.2c). A third "stalk" is formed by segments preceding the first $Zn^{2+}$-finger and residues located between the $Zn^{2+}$-fingers. This domain makes only a few contacts with the surface of ATPase subunit 3 and is less conserved than the $Zn^{2+}$-fingers. Finally, a fourth "stalk" is formed by two β-strands with a loop at their tip. This loop projects over the axial pore and

faces ATPase subunit 6 (Figure 3.2c), but it makes no clear contact with the D1 ring. At the tip of the "β-strand finger" is a highly conserved tyrosine residue.

To test the functional role of the $Zn^{2+}$-fingers, we used *S. cerevisiae* Cdc48, Npl4, and Ufd1 in an *in vitro* unfolding assay. A fusion between a short degron sequence and the fluorescent protein mEos3.2 was poly-ubiquitinated and incubated with the ATPase complex; the loss of fluorescence is an indication of Eos unfolding[224]. The results show that mutation of the central His and Cys residues in either of the individual $Zn^{2+}$-finger domains had little effect on unfolding, but a defect was seen when both domains were mutated together (Figure 3.2e). Similarly, mutants in individual $Zn^{2+}$-fingers could rescue the temperature-sensitive growth phenotype of an *npl4-1* yeast strain, but a mutant in both $Zn^{2+}$-fingers could not (Figure 3.S10a). We also tested mutations in the Cdc48 ATPase in the unfolding assay. Both $Zn^{2+}$-finger domains are in close proximity to a conserved tri-phenylalanine (FFF) sequence in the D1 domain. Indeed, mutation of the first or third Phe reduced the unfolding activity of Cdc48 without affecting hexamer formation (Figure 3.2f; Figure 3.S9). Mutation of the central Phe in the FFF motif abolished unfolding completely, but also reduced hexamerization, consistent with the fact that it faces a hydrophobic pocket in the ATPase domain. The importance of the $Zn^{2+}$-fingers of Npl4 is supported by a recent report in which these domains were identified as a target of the drug disulfiram [215]. The role of the "ß -strand finger" of Npl4 remains unclear, as its mutation or deletion did not alter unfoldase activity *in vitro* or affect the ability to complement a temperature-sensitive npl4 mutant (Figure 3.S10b-c). It remains possible that a subclass of Cdc48 substrates is dependent on the ß-strand finger.

Like other members of this family, the MPN domain consists of a core MPN fold with two inserts (insert-1 and -2) (Figure 3.3a). Npl4 is enzymatically inactive, as it lacks the $Zn^{2+}$-

binding motif in the core, which is essential for the de-ubiquitination activity of other MPN domains[226]. However, the position of Npl4's MPN domain above the D1 ATPase ring (Figure 3.2a-c) is similar to that of MPN domains in the 19S regulatory subunit of the proteasome and the COP9 signalosome, where an enzymatically active MPN domain (Rpn11 or CSN5) is located over a ring of six homologous subunits[216,227]. These MPN domains dimerize with enzymatically inactive MPN domains (Rpn8 or CSN6, respectively), whereas Npl4's MPN domain is a monomer, with its C-terminal helical domain occupying the dimerization site in the other MPNs (Figure 3.3b).

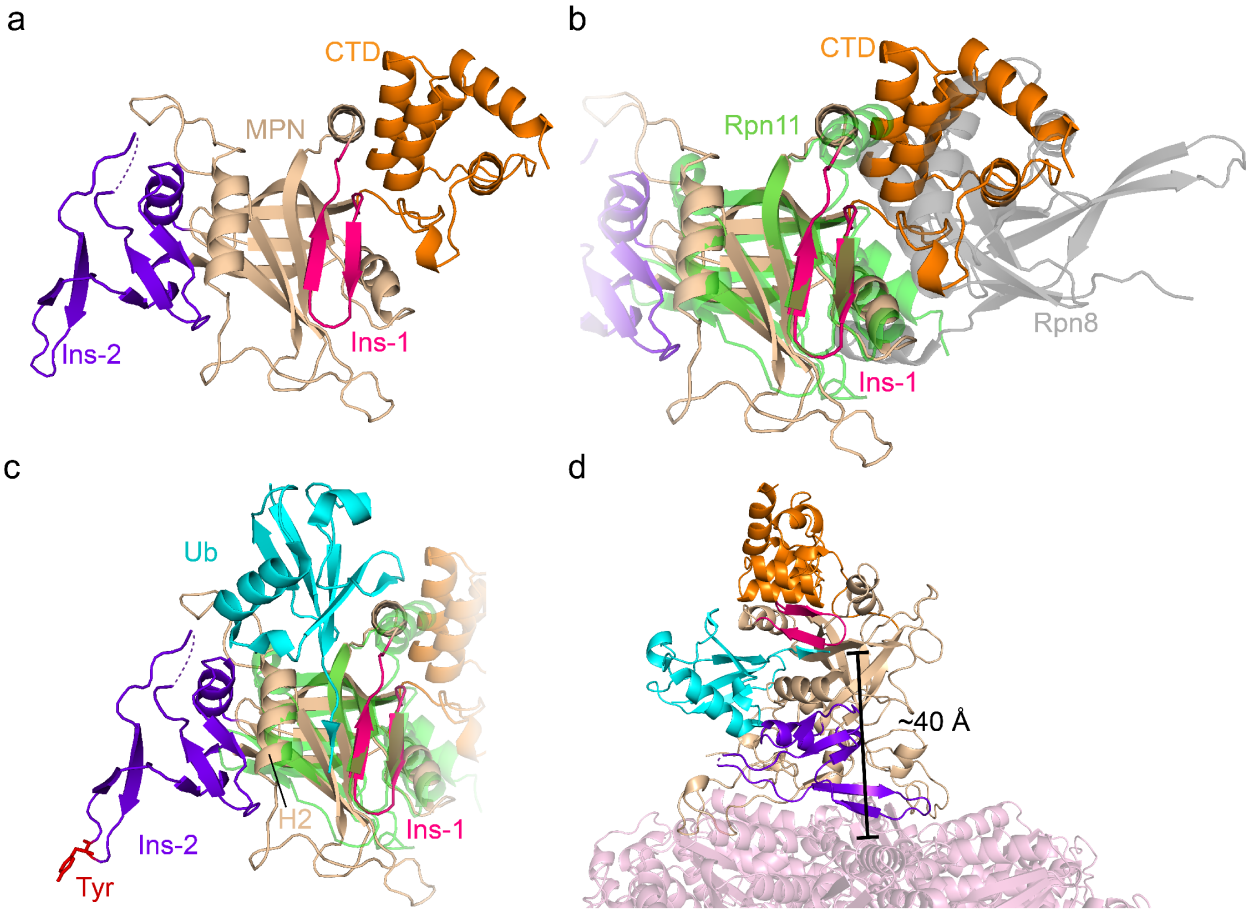



**Figure 3.3: MPN domain of Npl4. a**, Crystal structure of the MPN/CTD domains of Npl4. The core MPN region is in tan, the insert-1 (Ins-1) region in magenta, the insert-2 (Ins-2) region in

**Figure 3.3 (Continued)**

purple, and the CTD in orange. The dashed line indicates a 9-residue acidic loop unresolved in the crystal structure. **b**, As in **a**, but with the Rpn11/Rpn8 structure overlaid (PDB: 5U4P). Rpn11 is in green and Rpn8 in gray. **c**, As in **a**,, with the Rpn11 MPN and its associated ubiquitin overlaid (PDB: 5U4P). The conserved Tyr at the tip of Ins-2 is shown in stick representation in red. H2: MPN helix 2. **d**, As in **c**, but in the context of the Cdc48/Npl4 structure. The D1 ATPases are shown in pink.

The MPN domains of Rpn11 and AMSH-LP accommodate the C-terminal tail of ubiquitin in a cleft between insert-1 and helix 2, positioning ubiquitin's C-terminus next to the active site (shown for Rpn11 in Figure 3.3c)[228,229]. Insert-1 of Npl4 closely resembles the structure assumed by the corresponding region of Rpn11 in the presence of ubiquitin (Figure 3.3c), but it adopts this conformation even in the absence of ubiquitin, as observed for AMSH and several other MPN family members[230]. Given the similarity with Rpn11 and AMSH, it is likely that the cleft in Npl4's MPN domain also accommodates the C-terminal tail of a ubiquitin molecule, but in our structure the groove is covered by a segment of UT6 of Ufd1, suggesting that the UT6/MPN interaction is broken to allow ubiquitin binding.

The clefts of MPN domains seem to have generally only weak affinity for ubiquitin, as mutation of only one residue of Rpn11 affected ubiquitin binding[231], and mutation of residues near the cleft of Npl4 have no effect (Figure 3.S10). However, we found that the helical CTD following the MPN domain of Npl4 is required for interaction with ubiquitin, as a construct lacking the CTD was inactive (Figure 3.S10d,e). Thus, the poly-ubiquitin chain of a substrate is

recognized both by the UT3 domain of Ufd1 and by Npl4, perhaps at the interface between the MPN and CTD regions (Figure 3.4). As in the case of Rpn11[10], several ubiquitin receptors are required to mediate the interaction with the MPN domain.

Although the MPN domains of Rpn11 and Npl4 both position the polypeptide substrate for entry into the ATPase ring, the putative ubiquitin binding groove and central pore are at an approximately right angle in Cdc48/Npl4 (Figure 3.3d), whereas they are aligned in the proteasome[9]. Furthermore, the catalytic site of Rpn11 is located immediately above the central ATPase pore, whereas in the Cdc48/UN complex, there is a ~40 Å gap between the putative C-terminus of ubiquitin and the pore (Figure 3.3d). A gap of this size is large enough to accommodate an additional ubiquitin moiety. We therefore speculate that a ubiquitin molecule serves as a universal initiating signal for translocation into the pore, regardless of the substrate to which it is attached, which would explain why Cdc48 can deal with even folded substrates. The poly-ubiquitin chain would bind to both the UT3 domain of Ufd1 and the CTD/MPN domains of Npl4, and the polypeptide chain would then be inserted into the central pore of the Cdc48 ATPase (Figure 3.4). Although the details of pore insertion need to be clarified, our data show that Npl4 is unique among the known Cdc48 cofactors, as it binds directly to the ATPase ring and likely serves as a universal gatekeeper for all Cdc48-dependent reactions that require translocation through the central pore.

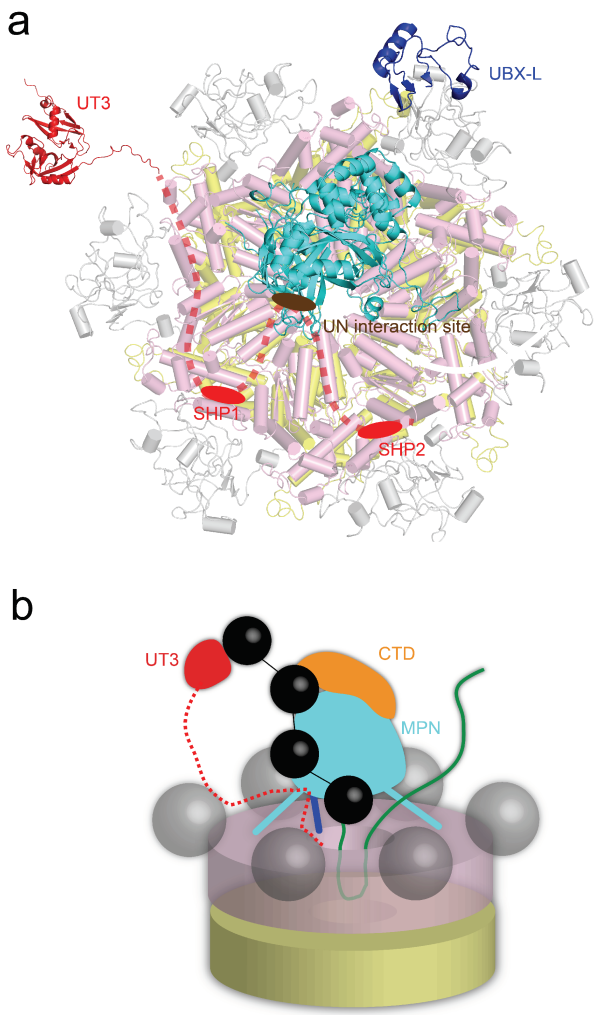



**Figure 3.4: Model for Cdc48/UN function. a**, Arrangement of the various domains of the UN complex in the Cdc48 ATPase complex. The UBX-like and $Zn^{2+}$-finger/MPN domains of Npl4 are shown in dark blue and cyan, respectively. The UT3 domain of Ufd1 (in red) is flexibly attached to the complex. The Npl4-interacting region of UT6 is shown as a brown oval, and the SHP boxes anchoring Ufd1 to the N domains of Cdc48 as red ovals. The intervening segments are shown as dashed lines. **b**, Model for the path of a poly-ubiquitinated substrate (black: ubiquitin; green: substrate protein) into the central pore of Cdc48. The three putative ubiquitin-interaction sites – UT3, CTD, and MPN – position the substrate for insertion into the central pore.

## 3.3 Methods

*Protein expression and purification*

Full-length *Chaetomium thermophilum* and *Saccharomyces cerevisiae* Cdc48, Ufd1, and Npl4, as well as the Npl4 $Zn^{2+}$-finger /MPN fragment (residues 129-602) and its tagged versions and mutants, were overexpressed in E. coli Bl21 DE3 RIPL cells as follows. Cdc48, Ufd1, and the $Zn^{2+}$-finger/MPN fragment were expressed with N-terminal His14-SUMO tags from the K27SUMO vector [211]. Full-length Npl4 was expressed untagged from the pET21b vector. Buffers used for Cdc48 purifications contained 5 mM $MgCl_2$ throughout. Cells were grown to an $OD_{600}$ of 0.6 to 0.8 in Terrific Broth and induced with 0.25 mM isopropyl β-thiogalactopyranoside (IPTG) for 16 hrs at 18°C. Cells were pelleted at 4000xg and resuspended in lysis buffer (50 mM Tris pH 8, 300 mM NaCl, 30 mM imidazole, 5 mM $MgCl_2$). For the UN complex, Ufd1 and Npl4 pellets were mixed at this step. The suspension was supplemented with protease inhibitors and DNAse I, and cells were lysed by sonication. Lysates were cleared by centrifugation in a Ti45 rotor (Beckman) for 30 min at 40,000 rpm and applied to Ni-NTA agarose for 1hr at 4°C. Beads were washed with lysis buffer, and His-tagged proteins were eluted in 10-20 mL elution buffer (50 mM Tris pH 8, 100 mM NaCl, 400 mM imidazole). The eluates were supplemented with 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP) and 100-200 nM SUMO protease (Ulp1p), then incubated at 4°C for 1 hr to remove His14-SUMO tags. Protein solutions were diluted to <200 mM imidazole with 50 mM Tris pH 8 and loaded onto a MonoQ 10/100 GL column (GE Healthcare) equilibrated in 50 mM Tris pH 8, then eluted by linear gradients to 500 mM NaCl over 8 column volumes. Peak fractions were pooled and concentrated. UN variants used for binding experiments and unfolding assays were stored at this point. Other proteins were further purified by size exclusion chromatography as follows. For

Cdc48 and Cdc48/UN complexes, proteins were incubated with 1 mM nucleotide (ADP or ATPγS) for 45 min prior to gel filtration, and the gel filtration buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP) additionally contained 100 uM nucleotide. For Cdc48/UN complexes, UN was added at a 3-5 fold molar excess of heterodimer to hexamer prior to gel filtration. For the $Zn^{2+}$-finger /MPN crystallization construct, the buffer was 25 mM HEPES pH 7.5, 100 mM NaCl. All proteins were separated on S200 size exclusion columns (GE Healthcare). Samples for cryo-EM analysis were used immediately after gel filtration, while crystallization samples were supplemented with 5% vol/vol 1,2-propanediol and flash frozen in liquid nitrogen.

*EM Specimen Preparation and Data Collection*

Sample homogeneity was first examined by negative-stain EM with 0.7% (wt/vol) uranyl formate, as previously described[232] . The images were recorded using a 1K x 1K CCD camera (Gatan) on a Philips CM10 electron microscope (FEI) operated at an acceleration voltage of 100 kV and a nominal magnification of 52,000x.

Prior to preparing grids for cryo-EM, all samples were concentrated to 2-3 mg/mL and centrifuged at 13,000 x g for 10 min to remove protein aggregates. NP-40 was added to the samples to a final concentration of 0.05% immediately before vitrification to lower the propensity of the particles to adopt preferred orientations in the ice layer.

The specimens for cryo-EM were frozen using a Cryoplunge 3 (Gatan). A 3.5-µL aliquot of the sample was applied to a glow-discharged Quantifoil Cu 1.2/1.3 grid (Quantifoil). The grid was blotted for 2.5-3.5 s and then plunge frozen in liquid ethane, which was maintained at a temperature of –172°C.

For analysis of Cdc48 alone in the presence of ADP or ATPγS, cryo-EM data collection was carried out at Harvard Medical School (Boston, MA), using a Polara electron microscope (FEI) operated at 300 kV and equipped with a K2 Summit direct electron detector (Gatan). All images were recorded in super-resolution counting mode using the semi-automated data collection software UCSFImage4 [233]at a nominal magnification of 31,000x (corresponding to a calibrated super-resolution pixel size of 0.62 Å). The defocus was set to range from –1.6 μm to –3.0 μm. The total exposure time of 6 s was dose-fractionated into 30 frames (200 ms per frame), with a dose rate of 8 electrons per pixel per second. A total of 1,628 and 569 image stacks were collected for Cdc48(ADP) and Cdc48(ATPγS), respectively.

For Cdc48 in complex with the UN cofactor in the presence of ADP or ATPγS, all data collection was performed in the Cryo-EM Resource Center at the Rockefeller University using the automated data collection software SerialEM[234]. Using a Titan Krios electron microscope (FEI) operated at 300 kV, 3,279 and 9,299 image stacks were collected for Cdc48-cofactor(ADP) and Cdc48-cofactor(ATPγS), respectively. The image stacks were recorded at a nominal magnification of 22,500x (yielding a calibrated super-resolution pixel size of 0.65 Å) with a K2 Summit camera in super-resolution counting mode. The defocus was set to range from –1.2 μm to –2.8 μm. The total exposure time of 15 s was dose-fractionated into 50 frames (300 ms per frame), with a dose rate of 10 electrons per pixel per second.

*Image Processing*

For the datasets of Cdc48 alone with ADP or ATPγS, the image stacks were motion-corrected and binned over 2 × 2 pixels with MotionCor[235], yielding a pixel size of 1.23 Å. The defocus values were estimated by CTFFIND4[236]. Each image was then manually inspected and rejected if considered of inadequate quality for further image processing (e.g., ice contamination, blurriness,

bad CTF fit, etc.). Particles were manually picked from the remaining images using the e2boxer.py command in EMAN2 [237].

For the Cdc48(ADP) dataset, 52,348 particles were manually picked from 1,440 images. The particles were extracted into 256 x 256-pixel boxes in RELION[238]. A subset of these particles was used to generate 2D class averages using the iterative stable alignment and clustering (ISAC) algorithm[239]. These were then used to calculate an initial 3D map using the validation of individual parameter reproducibility (VIPER) algorithm, both implemented in SPARX [240]. Using the map obtained with VIPER as a reference, all the particles were subjected to RELION 3D classification into 5 classes. One class containing 9,751 particles (~19% of the dataset) showed the most detailed structural features with apparent six-fold symmetry. This class was subjected to 3D refinement and yielded a density map at 8.9 Å resolution. When the same class was refined with C6 symmetry imposed, an improved density map at 7.2 Å resolution was obtained.

For the Cdc48(ATPγS) dataset, 29,313 particles were manually picked from 354 images and extracted into 200 x 200-pixel boxes in RELION. The particles were used to calculate 2D class averages with ISAC, and the resulting averages were used to calculate an initial 3D map with VIPER. Using this map as reference, the particles were subjected to RELION 3D classification into 10 classes. One class containing 3021 particles (~10% of the dataset) showed the most well-defined structural features. This class was subjected to 3D refinement and yielded a density map at 10.3 Å resolution. Since the structure also exhibited apparent six-fold symmetry, the same class was also refined with C6 symmetry imposed, which resulted in a map with an improved resolution of 8.2 Å.

For the Cdc48-cofactor complex in the presence of ADP or ATPγS, the image stacks were motion-corrected, dose-weighted, and binned over 2 × 2 pixels (yielding a pixel size of 1.3 Å)

using MotionCor2[241]. The defocus values were estimated with CTFFIND4. Each image was manually inspected and imperfect images were excluded from further processing. The remaining images were subjected to template-free particle auto-picking using Gautomatch (www.mrclmb.cam.ac.uk/kzhang/Gautomatch/).

For the Cdc48-cofactor(ADP) complex, 145,947 auto-picked particles were extracted from 2,551 images into 256 x 256-pixel boxes in RELION. These particles were subjected to reference-free 2D classification and classes showing poor averages were removed. To obtain an initial 3D model, the remaining 141,422 particles were aligned to Cdc48(ADP) cryo-EM map (obtained as described above and filtered to 40 Å) using RELION 3D classification with the number of classes set to one. The resulting map showed density not present in the reference map. Using this map as reference, the particles were sorted into 6 classes by 3D classification. The classes that showed strong cofactor density (class 2, 5 and 6; 82,249 particles; ~58% of the dataset) were combined and subjected to 3D classification into 10 classes. The resulting classes showed that the Cdc48-cofactor(ADP) complex also exhibits conformational variability, most notably in the N domain and the cofactor region. Classes 2, 8, 9 and 10 showing the strongest cofactor density and relatively well-ordered N domains were combined, and the resulting 52,178 particles were subjected to 3D refinement in RELION, yielding a map at 6.7 Å resolution.

For the Cdc48-cofactor(ATPγS) complex, the 808,059 auto-picked particles were binned over 4 x 4 pixels, resulting in a pixel size of 5.2 Å, extracted into 64 x 64-pixel boxes in RELION-2[242], and subjected to reference-free 2D classification in RELION-2. After removing classes giving poor averages, the remaining 616,772 particles were used as input for Cryosparc[243] to calculate an initial 3D map. Using this map as reference, the particles were sorted into 8 classes by RELION-2 3D classification. Only one of the classes (containing 91,883

particles; ~15% of the dataset) showed strong density for the cofactor bound to Cdc48 and detailed structural features. Refinement of this class yielded a density map at 10.4 Å resolution. The refined particles were then re-extracted from the original micrographs as re-centered and unbinned particles into 256 x 256-pixel boxes (pixel size of 1.3 Å). 3D refinement of the newly extracted particles was performed using the orientation parameters determined from the dataset of 4x binned particles as the starting point for further optimization. The final density map had a resolution of 4.6 Å, according to the gold-standard Fourier shell correlation (FSC) curve and using the FSC = 0.143 cutoff. Masking out the flexible N domains of Cdc48 and continuing the refinement in RELION-2 improved the resolution of the remaining map region to 4.3 Å.

*Model Building and Refinement for the Cdc48-cofactor(ATPγS) Structure*

The structure of human p97 bound to ATPγS (PDB: 5FTN) was used to build a homology model of *C. thermophilum* Cdc48 using SWISS-MODEL[244]. The homology model of Cdc48 and the crystal structure of the zinc finger and MPN domains of Npl4 were docked into the cryo-EM density map using UCSF Chimera[245]. The atomic model was optimized by cycles of real-space refinement using phenix.real_space_refine [246] against half-map 1 from RELION-2 and manual re-building in Coot[247]. FSC curves were calculated between the refined models and half-map 1 (work), half-map 2 (free), and the combined map.

*Protease-protection experiments*

Purified *C. thermophilum* UN was treated with increasing concentrations of trypsin (0, 2.4, 7.3, 22, and 66 ug/mL), incubated at room temperature for 30 min, and subjected to SDS-PAGE and Coomassie staining. Bands were subjected to mass spectrometry.

*Crystallization*

The $Zn^{2+}$-finger /MPN fragment (57 mg/mL) was thawed immediately before crystallization setup and centrifuged for 10 min at 20,000xg. Crystals were grown using the hanging drop method at 4°C by mixing 2 uL of protein solution with 2 uL of well solution (0.2 M Na/K phosphate pH 7, 9.2% wt/vol poly-γ-glutamic acid [Molecular Dimensions], 0.2 M potassium thiocyanate, 5% propylene glycol). Crystals appeared in 1-2 days and were harvested, incubated briefly in cryoprotection solution (0.2 M Na/K phosphate pH 7, 10% wt/vol poly-γ-glutamic acid, 0.2 M potassium thiocyanate, 30% propylene glycol), and flash frozen in liquid nitrogen.

*X-ray data collection and structure determination*

Crystals were screened at NE-CAT beamline 24-ID-C at the Advanced Photon Source (Argonne National Laboratory). An X-ray absorption scan showing a peak at 9665.7 eV confirmed the presence of zinc, and data were accordingly collected at this energy. The dataset used for structure determination (doi:10.15785/SBGRID/565) was gathered from a single crystal at 100K. Data were processed with XDS[248] and analyzed with Aimless[249]. The crystal belonged to the $P12_11$ space group, diffracted to 2.58Å (outer shell, 2.675-2.582 Å: I/sigma = 0.49, CC1/2 = 0.322) and contained two copies per asymmetric unit. To solve the structure, the cryo-EM density corresponding to the cofactor central tower was used as a molecular replacement (MR) model according to the protocol of Jackson et al.[250] , using the programs Chimera, Phaser, RESOLVE, and the CCP4i and Phenix crystallographic suites[245,246,251,252,253]. The MR solution was passed to the MR-SAD module in Phenix. Initial phases were of sufficient quality for the majority of the structure in the asymmetric unit to be assembled using the Autobuild function in Phenix[254]. The structure was completed by iterative rounds of manual adjustment in Coot[247] and

refinement in Phenix, with TLS parameter refinement enabled and a riding hydrogen model. Zinc fingers were refined with geometry restraints as suggested in[255]. Figures were generated using UCSF Chimera[245] and PyMOL (Schrödinger, LLC). Crystallographic software was maintained by SBGrid[256].

*LC-MS/MS and cross-link mapping*

Buffer components for cross-linking were BioUltra grade (Sigma Aldrich). LC-MS/MS was carried out with Thermo Scientific LC-MS grade reagents and solvents. The cross-linkers bis[sulfosuccinimidyl] suberate (BS3) and disuccinimidyl sebacate (DSSeb) were purchased from Thermo Scientific and ProteoChem, respectively.

Purified protein complexes (0.5 mg/mL in 50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, 0.1 mM ATPγS) were cross-linked with 100, 200, or 400 μM of either BS3 (3.5 mM stock in water) or DSSeb (5 mM stock in dimethyl sulfoxide) for 30 min at room temperature. The reactions were quenched by addition of Tris-HCl (pH 7.5) to a final concentration of 10 mM. Samples were reduced with 50 mM TCEP at 60°C for 10 min and alkylated with 50 mM iodoacetamide in the dark for 60 min at room temperature. Digestion was carried out at 37°C overnight with 0.5 μg sequencing grade modified trypsin (Promega) in 100 mM ammonium bicarbonate. The resulting peptides were passed though C18 Spin Tips (Thermo Scientific) before elution with 40 μL of 80% acetonitrile (ACN) in 0.1% trifluoroacetic acid. Eluted peptides were dehydrated in vacuum and resuspended in 20 μL 0.1% formic acid for MS analysis.

Peptides were analyzed in the Orbitrap Fusion Lumos mass spectrometer[257] (Thermo Scientific) coupled to an EASY-nLC (Thermo Scientific) liquid chromatography system, with a 2 μm, 500 mm EASY-Spray column. The peptides were eluted over a 120-min linear gradient

from 96% buffer A (water) to 40% buffer B (ACN), then continued to 98% buffer B over 20 min with a flow rate of 250 nL/min. Each full MS scan (R = 60,000) was followed by 20 data-dependent MS2 (R = 15,000) with high-energy collisional dissociation (HCD) and an isolation window of 2.0 m/z. Normalized collision energy was set to 35. Precursors of charge state 4-6 were collected for MS2 scans; monoisotopic precursor selection was enabled and a dynamic exclusion window was set to 30.0 s.

Raw LC-MS/MS data files were converted into mgf format using Proteome Discoverer (Thermo Scientific) and searched using pLink[258,259] with default FDR<5%, maximum e-value set at =0.001, trypsin digest with up to 3 missed cleavages, constant modification at 1=carbamidomethyl[C], variable modification at 1=oxidation[M]. Cross-linker was set to BS3 ([K [K 138.068 138.068 156.079 156.079) or DSSeb ([K [K 166.099 166.099 181.11 181.11). Mass tolerances for fragments and precursors were left unaltered. mgf files were searched against a database comprising Fasta sequences of Cdc48, Ufd1, and Np14.

*Substrate unfolding assays*

mEos3.2 was purified and polyubiquitinated as described[224]. Briefly, the protein was expressed as an N-terminal His14-SUMO fusion. After SUMO cleavage, an N-terminal arginine is exposed, facilitating ubiquitination by the purified *S. cerevisiae* enzymes Uba1, Ubc2, and Ubr1. Eos was separated from ubiquitination machinery on the basis of a C-terminal streptavidin-binding peptide tag, which was then removed by 3C protease. Finally, substrates bearing ubiquitin chains of 5-10 moieties were isolated by gel filtration.

Substrate unfolding was monitored as described[224]. Briefly, the substrate (200 nM) was mixed with 300 nM UN variants and 400 nM Cdc48 variants in 50 mM HEPES pH 7.2, 100 mM KCl, 10 mM $MgCl_2$, 0.25 mg/mL protease-free bovine serum albumin. After 10 min incubation

at 30°C, ATP (2 mM) was added, and fluorescence (excitation: 540 nm, emission; 580 nm) was monitored in a Spectramax M5 plate reader for 30 min.

*Substrate binding experiments*

The Npl4 $Zn^{2+}$-finger/MPN fragment (residues 129-602) with an N-terminal streptavidin-binding peptide (SBP) tag was incubated with magnetic streptavidin agarose beads (Pierce) in binding buffer (50 mM Tris pH 8.0, 150 mM NaCl) for 15 min at room temperature. The beads were washed three times with binding buffer to remove excess bait protein. Next, DyLight 800-labeled, polyubiquitinated GFP generated as described for the Eos substrate above was incubated with the beads for 15 min at a concentration of 30 nM in 250 uL binding buffer. The beads were again washed three times. Bound material was eluted with binding buffer plus 5 mM biotin and subjected to SDS-PAGE and fluorescence scanning on an Odyssey CLx infrared scanner (Licor).

*Yeast experiments*

The *npl4-1* strain (Mata npl4-1 ura3-52 leu2Δ1 trp1Δ63) was transformed with plasmids derived from pPS402 (gift of Pedro Carvalho, originally generated by the lab of Pamela Silver [260]). The original plasmid encodes wild-type Npl4 under its endogenous promoter and includes a Ura3 cassette. Initial cultures were grown at room temperature, as the *npl4-1* strain grows poorly at 30°C. Yeast were spotted in 10-fold serial dilution on SD-Ura plates and incubated at room temperature, 30°C, or 37°C for 2-3 days.

**Acknowledgements**

We thank Xudong Wu and Long Li for assistance with crystallography, the SBGrid consortium at Harvard Medical School, and the ICCB Longwood for use of equipment. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the NIH/NIGMS (P41 GM103403). The Pilatus 6M detector on 24-ID-C beam line is

funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This research was supported in part by a Helmsley Postdoctoral Fellowship at The Rockefeller University (to K.H.K.), funding from the Blavatnik Family Foundation (to E.N.), a research collaboration with the Waters Corporation (J.R.E.), and NIGMS grants R01GM052586 and T32GM007753. E.N. and T.A.R. are Howard Hughes Medical Institute investigators.

## Chapter 4: Conclusion

### 4.1 Overview

The studies described in this dissertation resolve several longstanding questions regarding the Cdc48 ATPase, including its basic mechanism of substrate handling and its overall structure in complex with its essential cofactors Ufd1 and Npl4. In this chapter, I will place these insights in the broader context of Cdc48's cellular roles. The chapter will discuss the implications of the substrate processing model elucidated in chapter 2 for Cdc48's function in *in vivo* pathways, particularly ERAD. Outstanding questions regarding Cdc48's integration into these pathways and cooperation with downstream factors including the proteasome will be highlighted. The chapter will also discuss Cdc48's structural and functional similarities to the proteasome and speculate on the nature of cooperation between these two critical machines. Future avenues of experimentation will be outlined throughout. Finally, I will provide a short discussion of Cdc48's potential as a therapeutic target.

### 4.2 Substrate processing in context

This section is partly adapted from a previously published review article:

Bodnar NO and Rapoport TA. "Toward an understanding of the Cdc48/p97 ATPase." *F1000Research* 2017, 6(F1000 Faculty Rev):1318

The mechanistic studies described in Chapter 2 resulted in a model for a generic Cdc48 substrate processing event, for which the minimal components are a polypeptide conjugated to a K48-linked polyubiquitin chain, the cofactors Ufd1/Npl4, and the ATPase itself. However, Cdc48 acts on a broad range of substrates with various biochemical characteristics and subcellular locations. It also cooperates with a large number of accessory cofactors beyond Ufd1/Npl4 and is integrated into a broader system of chaperones that funnel substrates to the proteasome. An understanding of how the essential recognition, unfolding,

and release mechanisms discussed in Chapter 2 might be differentially deployed or fine-tuned in these various contexts will be an important target of future work.

Based on our work and other mechanistic studies of the ERAD process, we propose one example of the integration of Cdc48's unfoldase activity into a specific quality control pathway. In ERAD, the Cdc48 complex extracts polyubiquitinated, misfolded proteins from the ER membrane. This process can be reproduced with proteoliposomes and purified proteins [83], demonstrating that no other component is required. Our model for ERAD activity is as follows. After retrotranslocation of a substrate segment across the ER membrane, a Lys48-linked polyubiquitin chain is appended to the substrate by an ER-resident ubiquitin ligase (for example, Hrd1) (step 1). Next, the polyubiquitin chain binds to the Cdc48 complex (step 2), which is recruited to the ER membrane via an interaction of Cdc48's N domain with the membrane protein Ubx2 [89]. Cdc48 processively extracts the substrate from the ER membrane, passing the entire polypeptide through its central pore (step 3). Once the substrate is fully removed from the ER, Cdc48 dissociates from the Ubx2 anchor, diffusing away from the membrane with its associated substrate (step 4). This dissociation step may be triggered or simply stochastic but would be required to free an N domain binding site for a DUB, such as Otu1 (step 5), as these proteins bind to the same site[83]. Hydrolysis of ATP by the D1 domain exposes the ubiquitin chain to the DUB, resulting in chain trimming and translocation of the remainder of the oligoubiquitinated substrate through the Cdc48/p97 pore (step 6). Finally, the unfolded substrate is retrieved by one of the accessory factors discussed above and passed on to the proteasome (step 7). In the case of glycosylated ERAD substrates, the N glycans probably are moved through the central pore and removed after translocation, as the cytosolic N-glycanase Png1 binds to the C-terminus of Cdc48[261,262].

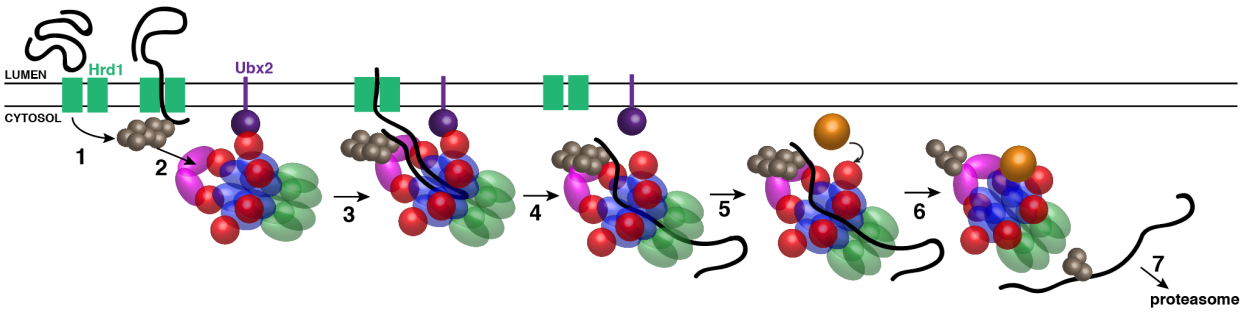



**Figure 4.1**: **Model for Cdc48 function in endoplasmic reticulum–associated protein degradation.**

In step 1, a segment of a misfolded luminal substrate is moved into the cytosol and polyubiquitinated by Hrd1. In step 2, the Cdc48 complex is bound to the cytosolic face of the endoplasmic reticulum via an interaction of its N domain with Ubx2. The complex recognizes the polyubiquitin chain via Ufd1/Npl4 (UN). In step 3, Cdc48 uses ATP hydrolysis in the D2 domains to extract the substrate from the membrane, translocating the polypeptide through its central pore. In step 4, Cdc48 completes translocation and eventually diffuses away from Ubx2, together with the bound substrate. In step 5, a deubiquitinase (DUB) such as Otu1 (orange) binds to the newly vacant N domain. In step 6, the D1 domains hydrolyze ATP, moving the N domains to the "down conformation" and allowing trimming of the ubiquitin chain. The oligoubiquitinated substrate then is released from the trans side of the pore. In step 7, the substrate is transferred to downstream factors, eventually arriving at the proteasome.

Similar models might be proposed for Cdc48's activities in analogous pathways such as mitochondrion-associated degradation, ribosomal quality control, and chromatin-associated degradation. In each of these cases, dedicated E3 ligases detect anomalous substrates and append ubiquitin chains to them. In many of these pathways, accessory Cdc48 cofactors that serve to

recruit the ATPase to particular locations have been identified. However, it is unclear whether cofactors of this class are always required for appropriate substrate processing *in vivo*. For example, the effect on ERAD of Ubx2 deletion is less pronounced than the effect of loss of function of more central components such as the E3 ligases, Cdc48, or UN [263]. However, Ubx2 is not the only ER membrane anchor that can serve to recruit Cdc48 to this organelle, as other proteins such as VIMP and Dfm1 have also been shown to perform this function[264 265]. Although additional recruiting cofactors are not required for reconstituted proteoliposome extraction and substrate unfolding events *in vitro*, this may not hold true in the *in vivo* case. Future work will need to carefully compare the relative efficiencies of Cdc48 substrate handling in the presence and absence of localization cofactors. Precise definition of the roles of these proteins might allow for detailed experimental manipulation of individual Cdc48-dependent pathways, rather than the more common and deleterious suppression of most or all Cdc48 function through genetic or pharmacologic inhibition.

An interesting caveat to the general concept of Ufd1/Npl4 as an obligate heterodimer and essential component of the basic Cdc48 reaction comes from the Vms1 protein, which appears to replace Ufd1 in the context of mitochondrion-associated degradation[124]. The existence of one such exception should prompt a search for additional proteins of this class. One possibility supported by our structural studies is that Npl4 is a unique cofactor that binds astride the central ATPase pore and is critical for substrate positioning and pore insertion, while cofactors such as Ufd1, Vms1, and possibly others bind interchangeably to Npl4. These Npl4 heterodimerization factors might serve to stabilize Npl4, recruit ubiquitin, or modulate the function of the MPN domain, which appears to be the binding site for Ufd1. Future experiments probing the ability of

Npl4 to support *in vitro* unfolding in the presence of Vms1 instead of Ufd1, or possibly even in the absence of both, would be a first step toward characterizing the roles of these proteins.

**4.3 Details of substrate recognition and pore insertion**

Around the time of our work described in Chapter 2, the group of Ray Deshaies used substantially similar polyubiquitinated fluorescent proteins to show that human p97 exhibits an unfoldase activity similar to that of Cdc48, including requirements for Ufd1/Npl4, K48-linked polyubiquitin chains, and predominant hydrolytic activity in the D2 ATPase ring[266]. This group also explored the role of branched polyubiquitin chains, finding that purified substrates with more elaborate ubiquitin conjugates were better targets for p97 *in vitro*. A mechanistic explanation for this observation is not immediately obvious. Multiple ubiquitin chains might serve to increase substrate affinity for the UN complex, but this would be less likely to affect processing efficiency in the setting of purified components at relatively high concentrations. Assuming that single chain and branched-chain substrates are recruited to UN with similar efficiency *in vitro,* another possible explanation is that the presence of multiple chains or branchpoints somehow allows more efficient initiation of substrate unfolding. This possibility raises the question of how substrate translocation is initiated. Specifically, which residues of any given substrate polypeptide are the first to enter the central pore and interact with pore residues of Cdc48?

Notably, p97 and Cdc48 lack conserved aromatic residues in the D1 ATPase pore loop. These residues are present in the archaeal Cdc48 homolog VAT, which translocates substrates through its central pore in the absence of additional cofactors[267]. Indeed, these residues have been visualized in contact with polypeptide chains moving through the central channel of this protein in cryo-EM structures[157]. The absence of similar residues in eukaryotic Cdc48 and p97

indicates that central pore insertion in these ATPases may proceed by a different mechanism. Because Cdc48 does not require a flexible initiation region similar to that needed by the proteasome, but does require polyubiquitination, one possibility is that substrate translocation is actually initiated within the ubiquitin chain. The most likely mechanism in this scenario would be that the Npl4 zinc finger/MPN domain positions the ubiquitin chain in a manner that allows pore entry, and possibly also alters the D1 pore conformation to promote ubiquitin insertion. This model would be consistent with Cdc48's ability to translocate free ubiquitin chains and to release substrates with oligoubiquitin chains still attached, but would require structural validation.

This structural validation would be best approached through cryo-EM, which in recent years has yielded several structures of AAA ATPases with polypeptide substrates visible in the central pore[9,156,157]. In the case of eukaryotic Cdc48, sample preparation will be a particular challenge, as the UN cofactor complex will need to be included and substrates will need to bear ubiquitin chains of appropriate length and linkage. Because ubiquitin is relatively hydrophobic and prone to aggregation, these preparations may be difficult to handle as EM samples, which we have indeed found to be the case in preliminary attempts. Furthermore, active ATP hydrolysis would be required to obtain translocating complexes, which would further increase conformational heterogeneity and hence the difficulty of particle classification.

We initially hypothesized that pore insertion might be accomplished by the conserved aromatic residue at the tip of the Npl4 beta-strand finger (see Chapter 3). This aromatic residue might function similarly to the one found in the ATPase SecA, which uses a two-helix finger to push substrates through the SecY protein-conducting channel[268]. In the absence of canonical D1 ATPase pore loops, this mechanism would have provided a tidy explanation for how substrates are initially pushed into the central pore. However, we have not yet found a role for the beta-

strand finger, as alteration or deletion of this portion of the cofactor had no obvious *in vitro* or *in vivo* phenotype. Because this motif is highly conserved, it is nevertheless likely that it plays some role in substrate processing. By analogy to the proteasome, one possibility is that it is essential in cases for which no flexible initiation region is present. Our *in vitro* substrate contains such a region, and it is possible that substrates responsible for Cdc48's essentiality *in vivo* also contain such regions. However, the beta-strand finger may be necessary for processing of a particular class of substrates for which the initiation of translocation is particularly problematic. Future experiments with *in vitro* substrates that do not contain any flexible loops or tails will be the first steps toward evaluating this hypothesis. Interestingly, mutation of the D1 pore loop residues in *S. cerevisiae* Cdc48 to conform to the aromatic residues found in archaeal variants is toxic[269], suggesting that pore insertion mechanisms have diverged in the eukaryotic and archaeal lineages.

**4.4 Cdc48 cooperation with the proteasome**

This section is partly adapted from a previously published review article:

Bodnar NO and Rapoport TA. "Toward an understanding of the Cdc48/p97 ATPase." *F1000Research* 2017, 6(F1000 Faculty Rev):1318

Cdc48 is generally thought to function upstream of the proteasome, handling substrates that ultimately are destined for proteasomal degradation. Our *in vitro* experiments show that many of the substrate molecules released from Cdc48 contain four or more ubiquitin molecules, which would be a suitable targeting signal for the proteasome[35], provided that ubiquitin molecules emerging from the D2 ring rapidly refold. At least in the case of ERAD, substrates do not directly bind to the proteasome but rather first interact with Rad23 or Dsk2[91,209]. These proteins seem to serve as "shuttling factors" by harboring both ubiquitin- and proteasome-

binding domains. However, the *in vitro* experiments show that some substrates released from Cdc48 bear ubiquitin chains that are too short to directly bind to the proteasome or shuttling factors, and in these cases the Cdc48-associated "E4" ubiquitin ligase Ufd2 might need to extend the oligoubiquitin chains before handoff to the downstream components[193]. Released substrates that contain exposed hydrophobic segments may require the additional activity of a "holdase" complex anchored by Bag6, which reduces aggregation of substrates destined for degradation[92]. Finally, some substrates may be transferred directly from Cdc48 into the 20S proteasome without involvement of the 19S subunit[188,189,270], although evidence for such an interaction in eukaryotes is lacking[271]. Interestingly, in the absence of Ufd3, a cofactor that binds to the C-terminal tail of Cdc48, ubiquitin is degraded abnormally quickly by the proteasome in yeast cells[57,272]. One possibility is that Ufd3 antagonizes the Cdc48-20S interaction, preventing the degradation of unfolded ubiquitin molecules emerging from the D2 ring of Cdc48. Alternatively, Ufd3 may somehow facilitate the refolding of ubiquitin molecules that emerge from the D2 ring.

The direct reconstitution of a Cdc48-dependent proteasomal degradation reaction would be a major step toward elucidating the mode or modes of cooperation between the two ATPases. The substrates developed in this work have some characteristics that would make them useful for such experiments. The presence of a fluorescent tag allows reproducible and relatively fast binding and release experiments, which would need to be adapted to query substrate handling by multiple downstream components such as the shuttling factors. The ability to follow the length of the ubiquitin chain through deubiquitination and reubiquitination steps would allow dissection of DUB and Ufd2 activity in a potential E4-mediated handoff reaction. Finally, the GFP or Eos moieties can be used to follow unfolding, which is hypothesized to be the essential requirement for Cdc48 in proteasomal reactions that cannot proceed through the 19S ATPases alone. The

essential first step for any of these experiments, regardless of the specific downstream factors used, will be to establish a purified substrate that cannot be degraded by purified 26S proteasomes, but does exhibit unfolding when incubated with Cdc48 and UN. Most likely, such a substrate will have minimal flexible regions. After this substrate is validated, experiments incorporating purified Rad23, Dsk2, Ufd2, Ufd3, Bag6, or other potential intermediary components can be initiated. The initial goal would be the identification of a set of components that enables 26S substrate degradation when added in conjunction with Cdc48 and UN, followed by mechanistic dissection of the resulting reaction.

Our biochemical and structural experiments reveal a striking set of similarities between the Cdc48/UN complex and the 19S proteasome. Both complexes are AAA ATPases that unfold polypeptide substrates by passage through the central pore[25]. Both require initial recognition of a K48-linked polyubiquitinated substrate by intimately associated ubiquitin receptors, a binding event that causes the stimulation of ATP hydrolysis[25]. Both feature a ubiquitin-interacting MPN domain poised over the central pore[228], and both require the function of deubiquitinating enzymes to complete a substrate processing event. Despite these similarities, clear-cut differences include the presence of two ATPase rings in Cdc48 (as opposed to one in the 19S), the lack of catalytic activity of the Npl4 MPN domain (as opposed to the essential DUB activity of Rpn11), the variant positioning of the two MPN domains relative to the pore opening (see Chapter 3), the trimming function executed by Otu1 (as opposed to the en bloc removal of full ubiquitin chains by Rpn11), and the apparent ability of Cdc48 to handle substrates that do not expose enough of a flexible segment or loop for efficient processing by the 19S ATPases. The fact that archaeal Cdc48 can cooperate with the 20S proteasome[188,189,270] suggests that the

ancestral functions of Cdc48 and the 19S were very similar, but that over time Cdc48's role has evolved to encompass pre-processing of particularly difficult or inaccessible substrates.

**4.5 Connections to medicine (with apologies to Tom Rapoport)**

The general role of Cdc48 in handling difficult polyubiquitinated substrates immediately suggests several therapeutic avenues. The first is the use of Cdc48/p97 inhibitors to block substrate processing in cells that are highly dependent on protein quality control. A similar approach has been used successfully for many years in the context of multiple myeloma, a tumor of plasma cells that, due to their high production of antibodies, are dependent on an intact proteasome to degrade misfolded proteins[273]. Proteasome inhibitors such as bortezomib and carfilzomib are now a mainstay of treatment in this tumor type[274]. The specific type of proteasomal degradation on which these cells are most dependent is likely to be ERAD, which oversees quality control of antibodies proceeding through the secretory pathway[275]. This provides a compelling rationale for trials of p97 inhibitors in myeloma, as these compounds would presumably more directly target the particular substrates rendering these cells susceptible, rather than all proteasomal substrates[276]. However, as discussed above, the breadth of p97 functions indicates that these drugs will still have a wide variety of influences on pathways as diverse as the DNA damage response, aggregate handling, amino acid metabolism, and the UPR, making their effects and side effects in cancer difficult to predict. A more demanding challenge is to find ways to target one or a few p97-dependent pathways without compromising all of them at the same time.

Recently, the candidate chemotherapeutic drug disulfiram has been shown to form a copper compound metabolite that binds and inactivates Npl4 via the zinc finger domain[215]. Our structural study provides a rationale for the essential nature of this zinc finger domain and

explains why inactivating it would compromise function of the entire Cdc48/UN complex. Disulfiram, by analogy to p97 inhibitors, may thus be a good candidate for pharmacologic targeting of tumors that are highly dependent on protein quality control. However, our data also suggest that Npl4 is a unique p97 cofactor that, when compromised, might have effects nearly as broad as direct inhibition of the ATPase. Nevertheless, resistance to p97 inhibitors is already a known phenomenon[277], and disulfiram, in targeting Npl4, might constitute a potential rescue or adjunct therapy for resistant tumors.

Mutations in p97 are directly responsible for a set of neurodegenerative syndromes on the frontotemporal dementia – amyotrophic lateral sclerosis (FTD-ALS) spectrum. Patients present with a variety of phenotypes ranging from inclusion body myopathy to familial ALS to Paget's disease of bone[278]. Ubiquitin-positive cytosolic inclusions reminiscent of those found in other neurodegenerative diseases are commonly present[279]. Interestingly, recent *in vitro* work has indicated that p97 harboring disease mutations is hyperactive in its activity, unfolding model substrates more quickly than the wild type protein[266]. Consistent with this, p97 inhibitors were found to rescue, rather than exacerbate, mitochondrial defects found in patient-derived fibroblasts and fly disease models[280]. Whether p97 inhibition will be a viable strategy for the treatment of diseases caused by these mutations remains to be seen.

It is clear that neurodegeneration can be directly caused by p97 mutation, but several lines of evidence suggest that even the wild-type protein is involved in the complex pathogenesis of these diseases. For example, p97 can be found in the hallmark aggregates associated with various neurodegenerative syndromes, including dystrophic neurites (Alzheimer's disease), Lewy bodies (Parkinson's disease), and huntingtin inclusions (Huntington's disease)[281-283]. In several cases, it has been shown that overexpression of p97 and UN can ameliorate various

phenotypes associated with accumulation of pathogenic proteins in cellular models[284,285]. It is currently poorly understood why cytosolic accumulation of misfolded proteins appears to correlate with ER stress, activation of the UPR, and mitochondrial dysfunction, all of which are observed in neurodegeneration[286]. One hypothesis for the pleiotropic defects observed in many of these diseases is that sufficient inactivation or sequestration of p97 by pathologic oligomers or aggregates simultaneously produces dysfunction in the many quality control pathways that rely on this ATPase. This model could be tested by producing animal models of neurodegeneration that overexpress p97 or its cofactors and analyzing the course of disease progression. If p97 is partially responsible for clearance of pathogenic aggregates, then enhancing this function might constitute a disease-modifying therapy for multiple neurodegenerative diseases. Gene therapies or small molecules that cause upregulation of this critical quality control factor might allow neurons to survive longer even without modifying baseline production of pathogenic proteins, thus altering the course of these diseases.

## Appendix A: Ddi1, a ubiquitin-dependent protease

This appendix describes preliminary experiments investigating the mechanism of Ddi1. I would like to thank Fred Goldberg, Zhe Sha, and Galen Collins for bringing this protein to our attention and for many helpful discussions.


### Abstract

The yeast protein Ddi1 is part of a group of proteins known as proteasomal shuttling factors. These proteins bind to polyubiquitin chains via a ubiquitin-associated (UBA) domain and to the proteasome via a ubiquitin-like (UBL) domain, thus serving to recruit substrates to the proteasome. As compared with the related UBA-UBL proteins Rad23 and Dsk2, Ddi1 has several unique features, including a central retroviral protease (RVP) domain, which mediates homodimerization, and an atypical UBL that binds ubiquitin. Here, we show that Ddi1 is an active aspartyl protease that targets substrates bearing long polyubiquitin chains. Under physiologic buffer conditions, purified Ddi1 cleaves a model fluorescent protein substrate in a manner that requires substrate polyubiquitination. Efficient proteolysis is dependent on the UBL domain and on the presence of very long ubiquitin chains of > 8 moieties. Together, our data suggest that Ddi1 is a unique accessory factor in the ubiquitin-proteasome system that may specialize in handling substrates with long polyubiquitin chains that are otherwise difficult to process.


### Introduction

The proteasomal shuttling factors Rad23, Dsk2, and Ddi1 exhibit a domain architecture that qualifies them to serve as bridges between polyubiquitinated substrates and the 26S proteasome[33]. Specifically, each of these proteins contains an N-terminal ubiquitin-like (UBL) domain and a C-terminal ubiquitin-associated (UBA) domain. UBL domains have been shown to

interact with the 19S proteasomal ubiquitin receptors, indicating that a proteasome-bound shuttling factor would be positioned in the same location as a standard polyubiquitinated substrate. UBA domains all share the property of binding to ubiquitin, although purified UBA domains vary in their relative affinities for monoubiquitin and polyubiquitin with different lysine linkages[287]. Thus, proteins containing both of these domains can provide an accessory pathway of substrate recruitment to the proteasome. This pathway may be important in cases where the standard mechanism of direct interaction between the 19S ubiquitin receptors and the substrate does not result in efficient degradation, perhaps due to geometric constraints[288].

Rad23, Dsk2, and Ddi1 can be distinguished on the basis of their varying *in vivo* phenotypes, which indicate that the three proteins are not functionally redundant[289]. Furthermore, the region between the UBA and UBL domains varies considerably among the three shuttling factors. In Rad23, a second, central UBA domain is present, which might contribute to selectivity for particular chain linkage types. In Dsk2, the central region consists of a methionine-rich set of Sti1-like repeats (STI), which may recognize misfolded polypeptide segments or transmembrane domains in the cytosol[290]. In Ddi1, the central region consists of a unique "helical domain of Ddi1" (HDD) and a retroviral protease (RVP) domain with homology to HIV protease[291]. The RVP domain mediates Ddi1 homodimerization[291].

Ddi1 has been shown to affect the stability of several individual substrates. These include the Ho endonuclease, which is polyubiquitinated but not degraded in the absence of Ddi1[292], and the F-box protein Ufo1[293]. Interestingly, Ufo1 is the F-box component of the Skp1-cullin-F-box (SCF) E3 ubiquitin ligase that targets Ho for degradation, indicating that Ddi1 regulates at least two components of this system. Ddi1 mutant and deletion strains also exhibit the unusual

phenotype of secreting high amounts of protein into the medium[294,295], an observation that may be related to Ddi1's interaction with certain SNARE proteins[296,297].

Recently, the human and worm orthologs of Ddi1 have been implicated in the processing of the transcription factor Nrf1, which is a critical component of the cellular response to proteasomal stress in higher eukaryotes[298,299,300]. Nrf1 is an ER-resident single-pass transmembrane protein that is constitutively retrotranslocated from the ER by p97 and degraded by the 26S proteasome[299]. However, when proteasome activity is pharmacologically compromised, Nrf1 is retrotranslocated but not fully degraded. Instead, it is deglycosylated and processed to a shorter form that translocates to the nucleus and directs the transcription of proteasome subunits, p97, Ufd1, and Npl4[300-302]. The partial proteolysis of Nrf1 requires the activity of Ddi2 (the human Ddi1 ortholog), and the most parsimonious explanation for this is that Ddi2 is itself the protease that cleaves Nrf1[300,303]. The overall logic of this system serves to upregulate critical ubiquitin-proteasome system (UPS) components when flux through the proteasome pathway overwhelms its capacity at any given time.

The RVP domain of a Ddi1 ortholog from *Leishmania major* was found to have proteolytic activity *in vitro*, but this activity was dependent on atypical buffer conditions (i.e. pH 4). Purified yeast and human Ddi1 orthologs have not been shown to have *in vitro* protease activity, even in screens against peptide libraries derived from full proteomes of the respective organisms[291,304]. These results raise the possibility that the Ddi1 RVP domain is somehow degenerate, although structures of the yeast and human protease domains reveal no obvious evidence for this[291,304]. Alternatively, Ddi1 might exhibit substrate specificity that is not captured by screens done to date.

Here, we use purified yeast Ddi1 and a model substrate with K48-linked polyubiquitin chains to show that Ddi1 is indeed a competent aspartyl protease. To our knowledge, Ddi1 is the first identified ubiquitin-dependent endoprotease other than the proteasome. Ddi1's preference for substrates with long ubiquitin chains may position it to serve a unique function in the UPS under conditions of proteasome compromise.

**Results**

Yeast Ddi1 was purified after overexpression in *E. coli*. Constructs with the native N terminus were poor SUMO protease substrates and thus retain an N-terminal His-SUMO tag. The model substrate used for these experiments was polyubiquitinated superfolder GFP (sfGFP), which was generated as described in chapter 2. Ddi1 was incubated with the fluorescently labeled substrate, and samples of the reaction were taken at various time points and analyzed by SDS-PAGE and fluorescence scanning (Figure A.1A). We observed disappearance of the high molecular weight fluorescent species and corresponding accumulation of a low molecular weight fluorescent peptide. Because the fluorescent dye is attached at Cys6 of the substrate and the ubiquitin chain is attached at K19, this result is consistent with proteolysis between these two residues, which would liberate a short, dye-labeled N-terminal peptide.

To confirm that proteolysis was occurring in the C6-K19 region, we analyzed the sequence of the substrate and compared it to the known cleavage site of Nrf1, which is between W103 and L104 of the human protein[299]. The substrate contained a similar WL sequence in the region between the dye and the ubiquitin chain (Figure A.1B), which was mutated to GG. This mutation abolished the appearance of the low molecular weight fluorescent peptide upon incubation with Ddi1, confirming that the WL sequence is important for this cleavage event (Figure A.1C). Ddi1's exact tolerance for amino acid substitutions in the cleavage site is not

known, but could readily be investigated with this system. Interestingly, an apparent decrease in substrate molecular weight was still observed even with the diglycine cleavage site mutant (Figure A.1C), suggesting that residual proteolysis might occur at another site in the substrate.

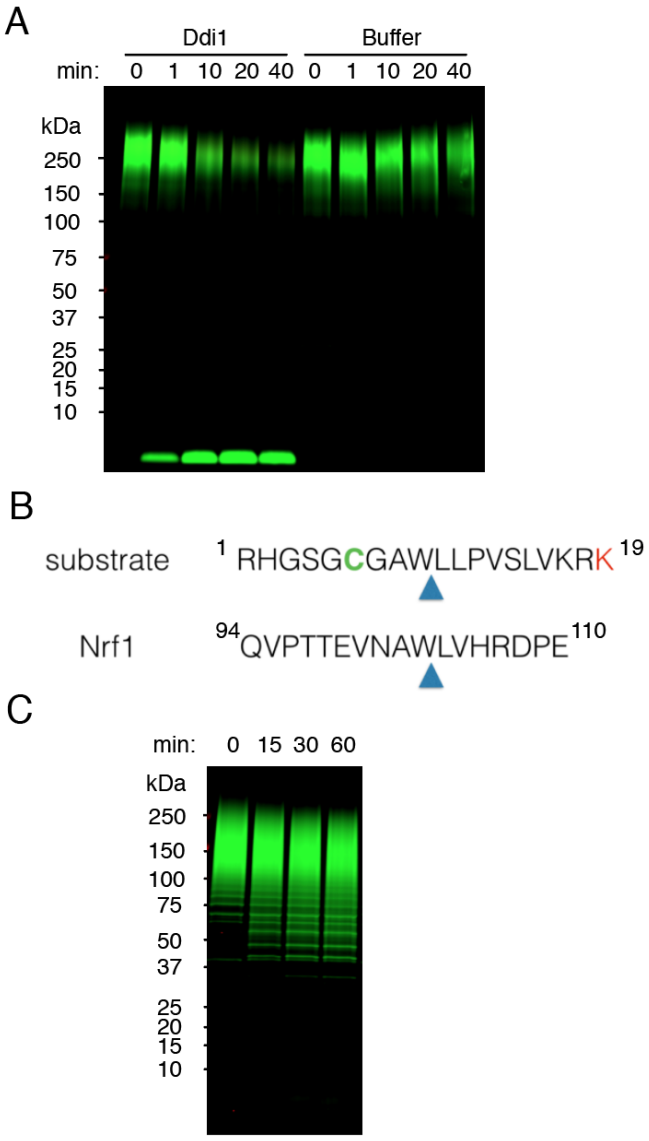



**Figure A.1:** Ddi1 cleaves a polyubiquitinated model substrate at a WL motif. (A) Purified yeast Ddi1 (1 uM dimer) or buffer was added to 1 uM polyubiquitinated sfGFP labeled with an infrared fluorescent dye. Samples were taken at the indicated time points and subjected to SDS-PAGE and fluorescence scanning. (B) Comparison of selected sequences from the model

**Figure A.1 (Continued)**

substrate and human Nrf1. The arrows indicate the putative substrate cleavage site and confirmed Nrf1 cleavage site. For the substrate, the cysteine bearing the fluorescent dye is marked in green, and the lysine bearing the polyubiquitin chain is marked in red. (C) Ddi1 was incubated with the substrate and samples analyzed as in (A), except that the WL motif in the substrate was mutated to GG.

Substrate cleavage by Ddi1 was dependent on the presence of the polyubiquitin chain, as unmodified substrate was not proteolyzed (Figure A.2A). To further explore this dependence, we used substrates with ubiquitin chains of varying lengths separated by size exclusion chromatography. Ddi1 efficiently cleaved substrates with very long ubiquitin chains of 15-25 moieties, but was inactive against substrates with chains of 8 ubiquitin moieties or fewer (Figure A.2B). This length requirement contrasts with that of the proteasome, for which the canonical minimal degradation signal is only 4 ubiquitin moieties[35].

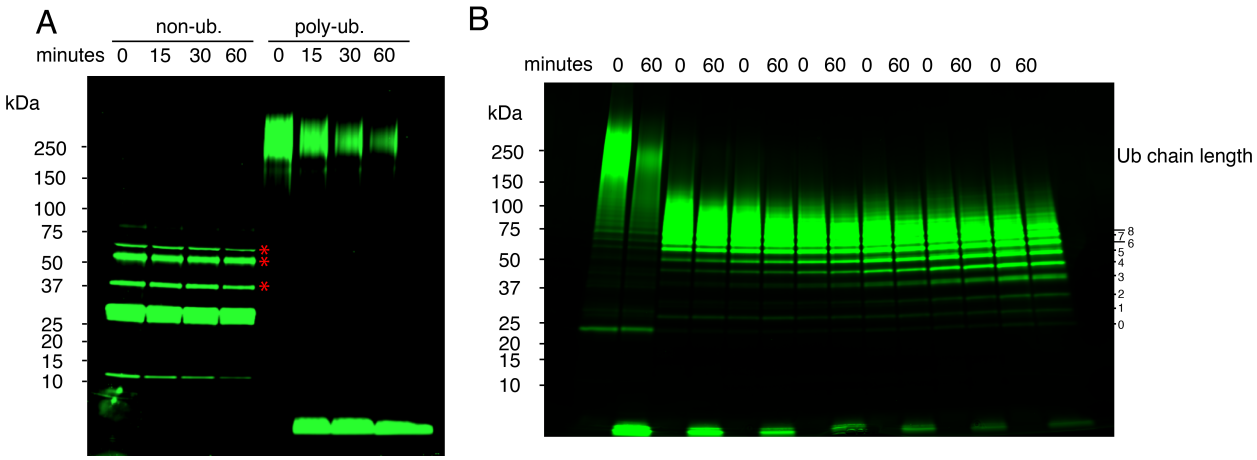



**Figure A.2:** Dependence of Ddi1 cleavage on chain length. (A) Ddi1 was incubated with either non-ubiquitinated substrate (left) or polyubiquitinated substrate (right). Samples were taken at the indicated time points and subjected to SDS-PAGE and fluorescence scanning. The red asterisks indicate contaminants in the substrate sample that are removed by gel filtration of the

**Figure A.2 (Continued)**

polyubiquitinated proteins. (B) As in (A), with various fractions of ubiquitinated substrate separated by gel filtration. Note that cleavage efficiency decreases with decreasing chain length.

To determine the contributions of individual domains of Ddi1 to substrate cleavage, we generated constructs with deletions of (1) the UBL domain (ΔUBL), (2) the UBA domain (ΔUBA), (3) both the UBL and UBA domains (HDDRVP), and (4) the UBL, UBA, and HDD domains (RVP). Each of these proteins migrated as a dimer on gel filtration. We incubated the polyubiquitinated substrate with varying concentrations of each of the deletion mutants and assayed degradation. In comparison to wild-type Ddi1 (Figure A.3A), a deletion of the UBL domain had a more pronounced deleterious effect on substrate degradation than a deletion of the UBA domain (Figures A.3B-C), consistent with *in vivo* results indicating that UBL deletion compromises Ddi1 function while UBA deletion does not[294]. This result is also consistent with data showing that the Ddi1 UBL domain, unlike other UBLs, binds not only the proteasome but also ubiquitin[305]. The HDDRVP construct, despite lacking both ubiquitin-binding domains, was still a competent protease, although higher concentrations were required to see activity (Figure A.3D). In contrast, the RVP domain alone was not active even at a 10:1 enzyme:substrate ratio (Figure A.3E), indicating that the HDD domain is required for catalytic activity. Interestingly, all of the proteolytically competent constructs retained the requirement for long ubiquitin chains (Figure A.3F). This result suggests that the HDD domain may mediate recognition of long chains and license the RVP domain to cleave the substrate.

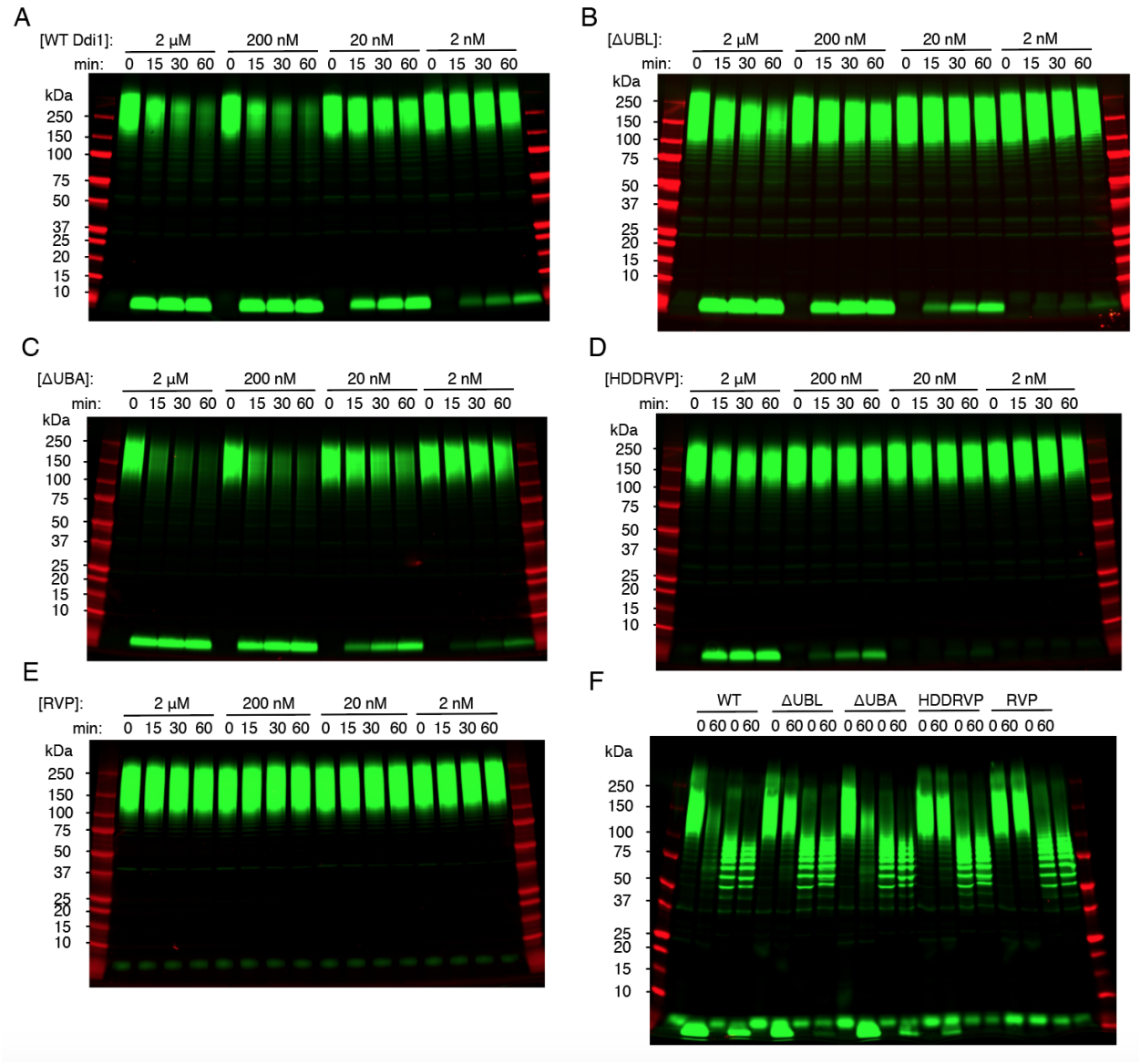



**Figure A.3:** Contributions of Ddi1 domains to activity. (A-E) Ddi1 variants at the indicated concentrations were incubated with polyubiquitinated, dye-labeled sfGFP (200 nM). Samples were taken at the indicated time points and subjected to SDS-PAGE and fluorescence scanning. M, molecular weight marker. The HDD-RVP construct is a deletion of both the UBL and UBA domains, while the RVP construct is deletion of the UBL, UBA, and HDD domains. (F) As in (A), except that substrates with both long and short chains were assayed (time points at 0 and 60 minutes for each type of chain). Substrates and Ddi1 variants were all at 100 nM.

**Discussion and future directions**

The demonstration that Ddi1 is an active protease resolves previous uncertainty regarding the role of the RVP domain and demonstrates that this protein likely has a role beyond simple, canonical proteasome shuttling. Although direct reconstitution of Nrf1 cleavage by Ddi1 would be the best way to demonstrate that Ddi1, rather than an unidentified intermediary protease, is responsible for Nrf1 processing *in vivo*, our data indicating that Ddi1 cleaves a polyubiquitinated substrate with a sequence close to that of Nrf1 constitute strong evidence that a similar process occurs in cells.

The unexpected dependence of Ddi1 on very long polyubiquitin chains raises the question of when such chains might be encountered *in vivo*. One possibility is that inhibition or dysfunction of the proteasome results in increased half-lives of some or all proteasome substrates, giving these polypeptides time to accumulate additional ubiquitin. Whereas a substrate might normally be immediately degraded upon achieving a ubiquitin chain length of 4 or more, conditions that lead to compromised processing of polyubiquitinated proteins might stabilize substrates and eventually produce a sizeable population with chains of 8 or more ubiquitin moieties. The chain lengths accumulating on Nrf1 in the presence of proteasome inhibitors have not been characterized, and such experiments might shed light on this issue.

Although our experiments show that multiple ubiquitin moieties are required for Ddi1 substrate recognition, they do not indicate which arrangements and combinations of ubiquitin chains are permissible. There are at least four ubiquitin-binding modules (two UBAs and two UBLs) per Ddi1 dimer, and possibly up to six if the HDD domain also recognizes ubiquitin. These might cooperatively recognize certain types of chains, but not others. In the most permissive case, rather than recognizing a certain chain length or linkage, Ddi1 might instead

recognize the presence of a certain number of ubiquitin molecules, regardless of their geometric arrangement. Indeed, multiple mono-ubiquitination has been shown to be sufficient for proteasome targeting for some substrates[306]. Alternatively, Ddi1 might require a specific chain linkage or subset of linkages (e.g., K48 and K11 linked chains), but discriminate against others (e.g. K63 linked chains). Experiments with substrates conjugated to other chain types, e.g. by the K63-selective E3 ligase Rsp5, might begin to address these questions.

Another source of substrate selectivity is the sequence surrounding the peptide bond to be hydrolyzed. Mutation of the WL motif to GG in our substrate abolished Ddi1 cleavage, but the data suggest that another, unidentified Ddi1 cleavage site might be present elsewhere in the substrate sequence. Systematic mutation of the WL motif to various residues could define Ddi1's stringency in selecting particular sequences for proteolysis, and positioning of the motif in different locations could define the required relative geometry of the ubiquitinated lysine and the cleavage site.

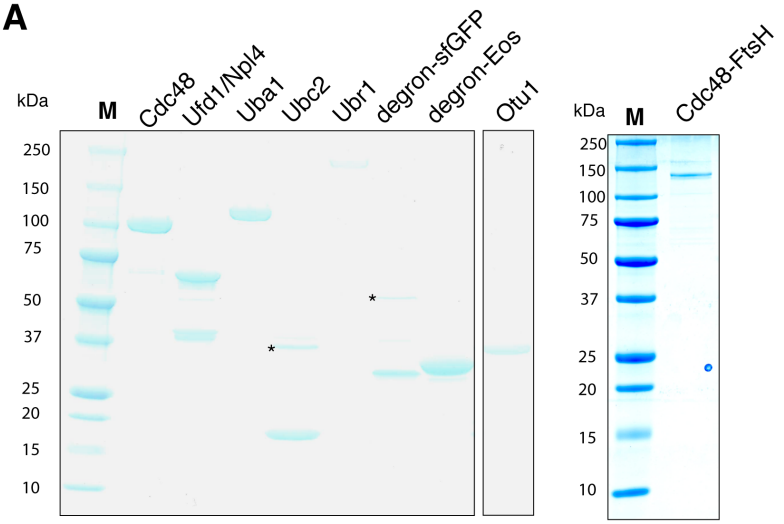
Whether Ddi1 cleaves only a few substrates with appropriate sequences or acts relatively promiscuously on any substrate that bears the correct type of polyubiquitin chain remains to be determined. To identify natural substrates of Ddi1, a screening approach might be most appropriate. One possibility would be to purify polyubiquitinated material from whole cell lysates, bind it to ubiquitin affinity resin, and cleave it with the active recombinant Ddi1 described here. Proteins appearing in the supernatant in this experiment would be candidate Ddi1 substrates. Alternatively, a quantitative mass spectrometry approach could be used to compare the relative abundance of proteins in wild-type and Ddi1 knockout yeast, with proteins enriched in the latter forming another set of candidate substrates. The *in vitro* system described here would yield a simple biochemical validation method for any identified candidates.

**Methods**

Ddi1 and its variants were expressed from the K27 SUMO vector and purified by Ni-NTA and ion exchange chromatography as described in Chapter 2 for other similarly purified proteins. Polyubiquitinated substrates were generated as described in Chapter 2. Cleavage assays were carried out in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM TCEP at 30ºC. Reactions were quenched at various time points by addition of SDS sample buffer, then subjected to SDS-PAGE and fluorescence scanning on an Odyssey CLx imager (Li-COR).

## Appendix B: Supplementary figures for Chapters 2 and 3

**A**




**B**

| Lysine ID | Peptide | Peak intensity |
|---|---|---|
| Substrate K19 | R.K#TTLAPNTQTASPPSYR.A | 2.3E+07 |
| Substrate K19 | K.RK#TTLAPNTQTASPPSYR.A | 1.2E+06 |
| Substrate K199 | K.LEYNFNSHNVYITADK#QK.N | 5.0E+04 |
| Substrate K17 | R.HGSGCGAWLLPVSLVK#R.K | 4.0E+04 |
| Ubiquitin K48 | R.LIFAGK#QLEDGR.T | 1.6E+08 |
| Ubiquitin K63 | R.TLSDYNIQK#ESTLHLVLR.L | 2.5E+05 |
| Ubiquitin K48 | R.LIFAGK#QLEDGRTLSDYNIQK.E | 1.3E+05 |
| Ubiquitin K6 | -.M*QIFVK#TLTGK.T | 1.0E+05 |

**Figure 2.S1: Purity of proteins used in this study**

(A) Purified proteins used in this study were subjected to SDS-PAGE and staining with Coomassie blue. Black stars indicate contaminating proteins. Degron-sfGFP and degron-Eos are substrates before being subjected to fluorescent labeling and ubiquitination. Cdc48-FtsH was run on a different gel; an irrelevant lane was removed from the gel on the left (white space). M, molecular weight markers.

(B) Tryptic peptides of the polyubiquitinated sfGFP substrate were analyzed by mass spectrometry, with a search for ubiquitin conjugation sites. Intensities of the peptides corresponding to ubiquitin chain linkage and ubiquitin conjugation sites are shown. Note that the predominant linkage is K48 and the predominant substrate ubiquitination site is K19. The # sign indicates the position of the isopeptide bond.

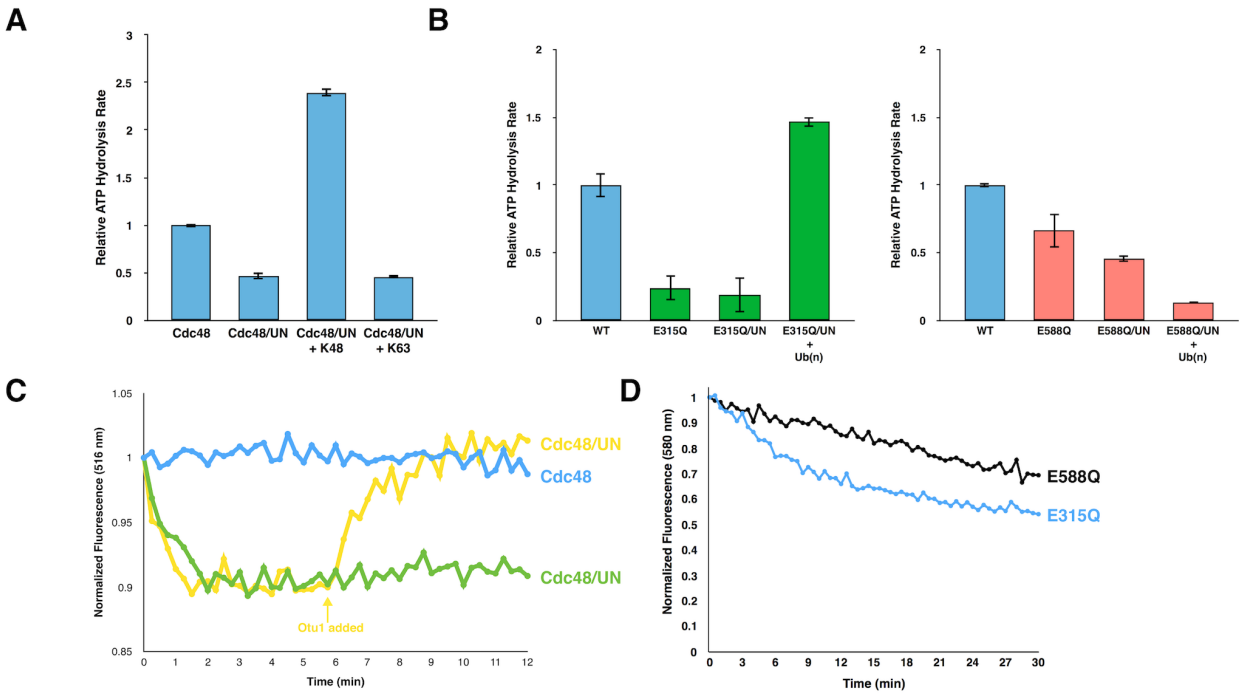



**Figure 2.S2: ATPase activity and substrate unfolding**

(A) ATP hydrolysis rates were determined with the indicated combinations of purified proteins and substrate. K48 and K63 indicate penta-ubiquitin with different linkages. The rates were normalized with respect to that of Cdc48 alone. Shown are the means and standard deviations of three experiments.

(B) As in (A), but with either wild type (WT) Cdc48 or Walker B glutamine mutants (Walker B mutation in D1 (E315Q; green bars) or D2 (E588Q; red bars)).

(C) Polyubiquitinated sfGFP (Ub(n)-sfGFP), generated as in Figure 1A, was incubated with an ATP regenerating system and excess of either Cdc48 alone or Cdc48 plus UN. After addition of ATP, GFP fluorescence was followed over time. One sample (yellow curve) received an equimolar amount of Otu1 at the indicated time point (arrow).

**Figure 2.S2 (Continued)**

(D) Substrate unfolding assays with Ub(n)-Eos and Walker B glutamine mutants. Assays were performed as in Figure 2F.

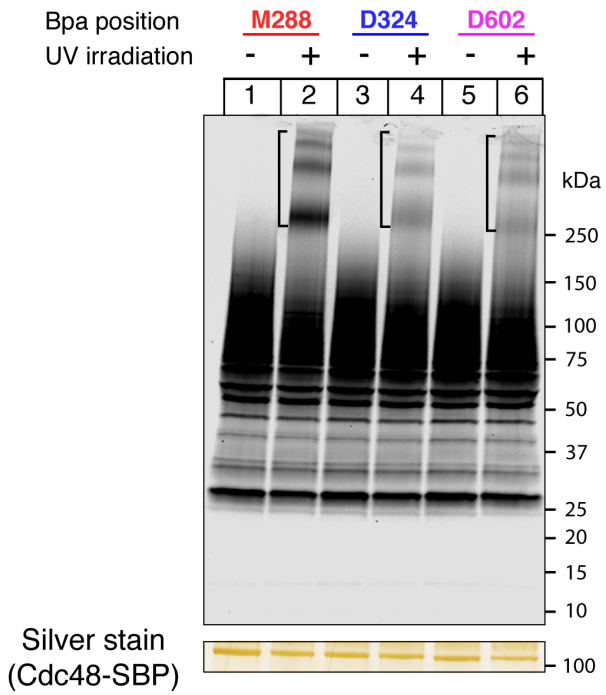



**Figure 2.S3: Efficiency of substrate crosslinking**

SBP-tagged Cdc48 (Cdc48-SBP) with Bpa at the indicated positions was incubated with dye-labeled, polyubiquitinated sfGFP (Ub(n)-sfGFP), UN, and an ATP regenerating system. Where indicated, reactions were irradiated. The samples were then subjected to SDS-PAGE and fluorescence scanning. The silver-stained band of Cdc48-SBP serves as a loading control. The brackets indicate the positions of crosslinked products. Crosslinking efficiencies were estimated by comparing the fluorescence intensities of non-crosslinked and crosslinked species (~20% for M288, and ~5% for D324 and D602).

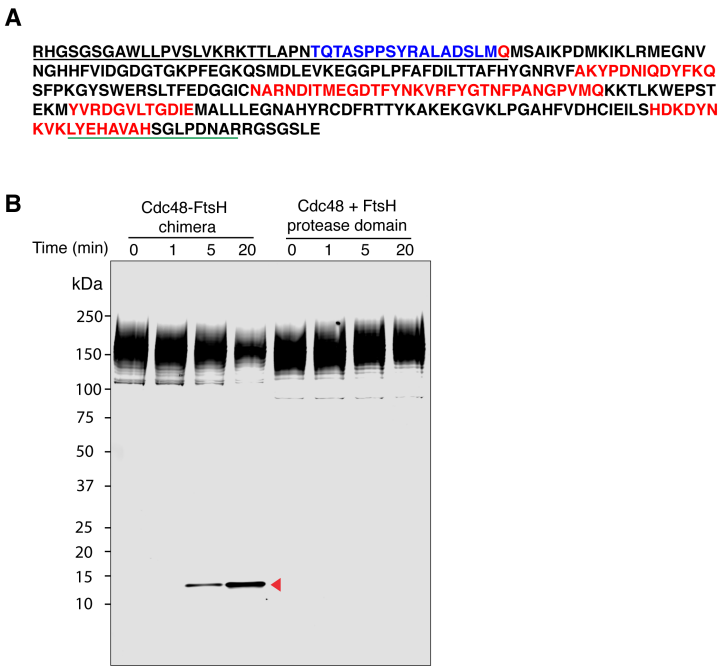



**Figure 2.S4: Substrate cleavage by the Cdc48-FtsH chimera**

(A) Dye-labeled, polyubiquitinated Eos (Ub(n)-Eos) was incubated with UN, the fusion of Cdc48 and FtsH (Cdc48-FtsH), and ATP, ATPγS, or ADP. One sample was directly analyzed by mass spectrometry. Sequences detected only in ATP are shown in red (15 unique peptides). Two unique peptides (shown in blue) were detected in all nucleotide states and correspond to part of the degron preceding Eos (underlined in black), which is likely flexible. Another sample was subjected to SDS-PAGE and the stable proteolytic fragment of ~13 kDa was trypsinized and analyzed by mass spectrometry. The identified peptide is underlined in green.

(B) Dye-labeled Ub(n)-Eos was incubated with an ATP regenerating system, UN, and either Cdc48-FtsH or a mixture of wild type Cdc48 and the isolated FtsH domain. Aliquots were removed at the indicated time points and analyzed by SDS-PAGE and fluorescence scanning. The arrow head indicates a stable fragment containing the fluorescent dye, which was generated with the Cdc48-FtsH fusion.

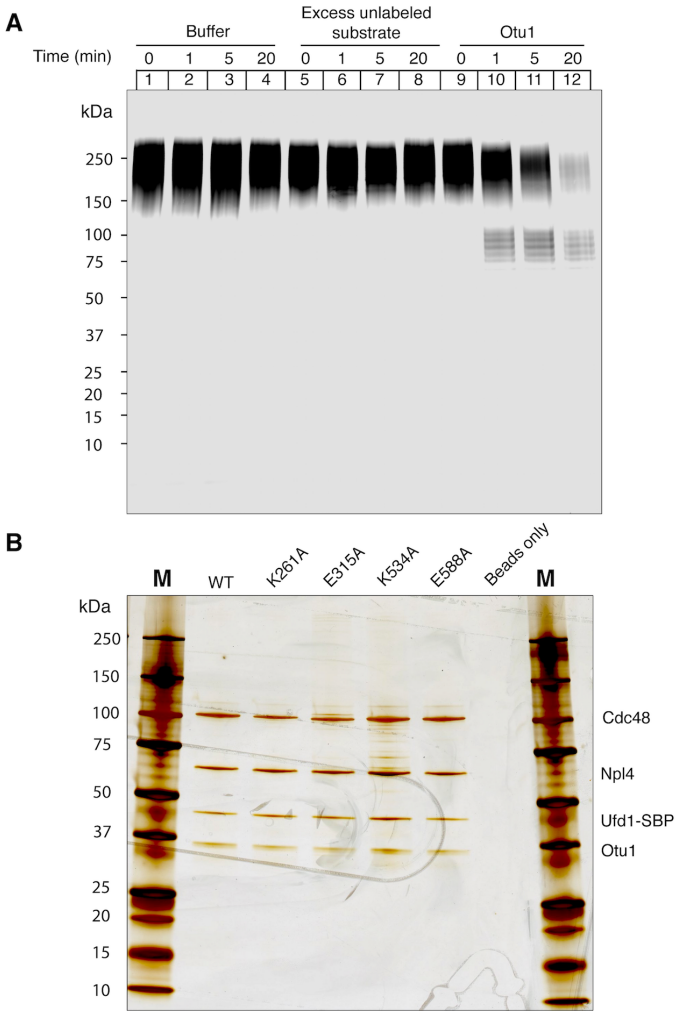



**Figure 2.S5: Otu1 binding to the Cdc48 complex**

(A) Complexes of Cdc48, SBP-tagged UN, and dye-labeled, polyubiquitinated sfGFP were immobilized on streptavidin beads in the presence of ATP. After washing, the indicated proteins were added and samples of the bound material were analyzed at the indicated time points by SDS-PAGE and fluorescence scanning. Where indicated, unlabeled, polyubiquitinated sfGFP was added in 5-fold excess over the original fluorescently labeled input.

**Figure 2.S5 (Continued)**

(B) SBP-tagged UN and the indicated Cdc48 mutants were immobilized on streptavidin beads, and wild-type Otu1 was added. After washing, the bound material was recovered and subjected to SDS-PAGE and silver staining.

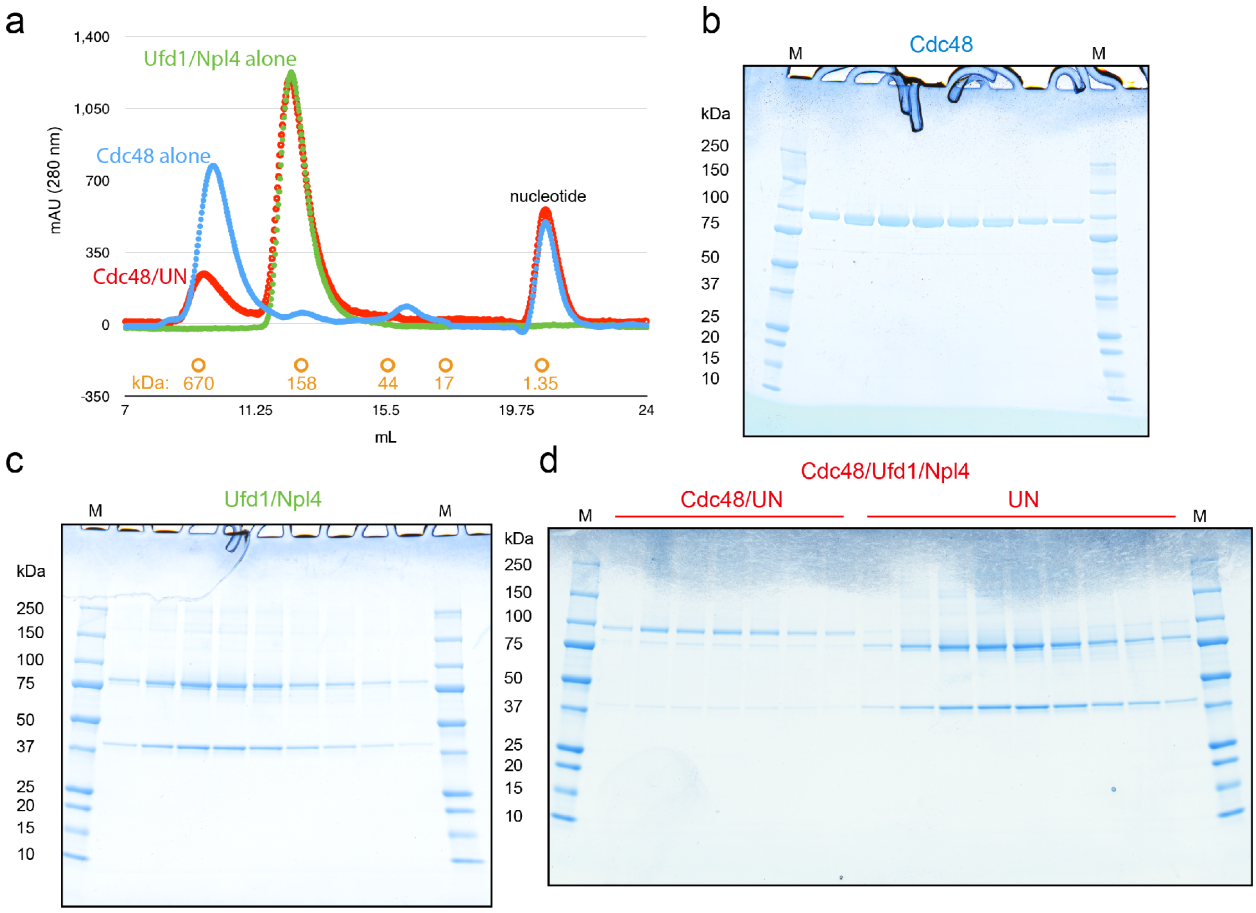



**Figure 3.S1: Purification of proteins. a**, Gel filtration chromatograms for *C. thermophilum* Cdc48 (blue), Ufd/Npl4 (UN; green), and Cdc48/Ufd1/Npl4 (Cdc48/UN; red). The elution positions of molecular weight standards are shown in orange. **b-d**, Samples from the major peaks of the gel filtration runs in **a** were subjected to SDS-PAGE and Coomassie blue staining.

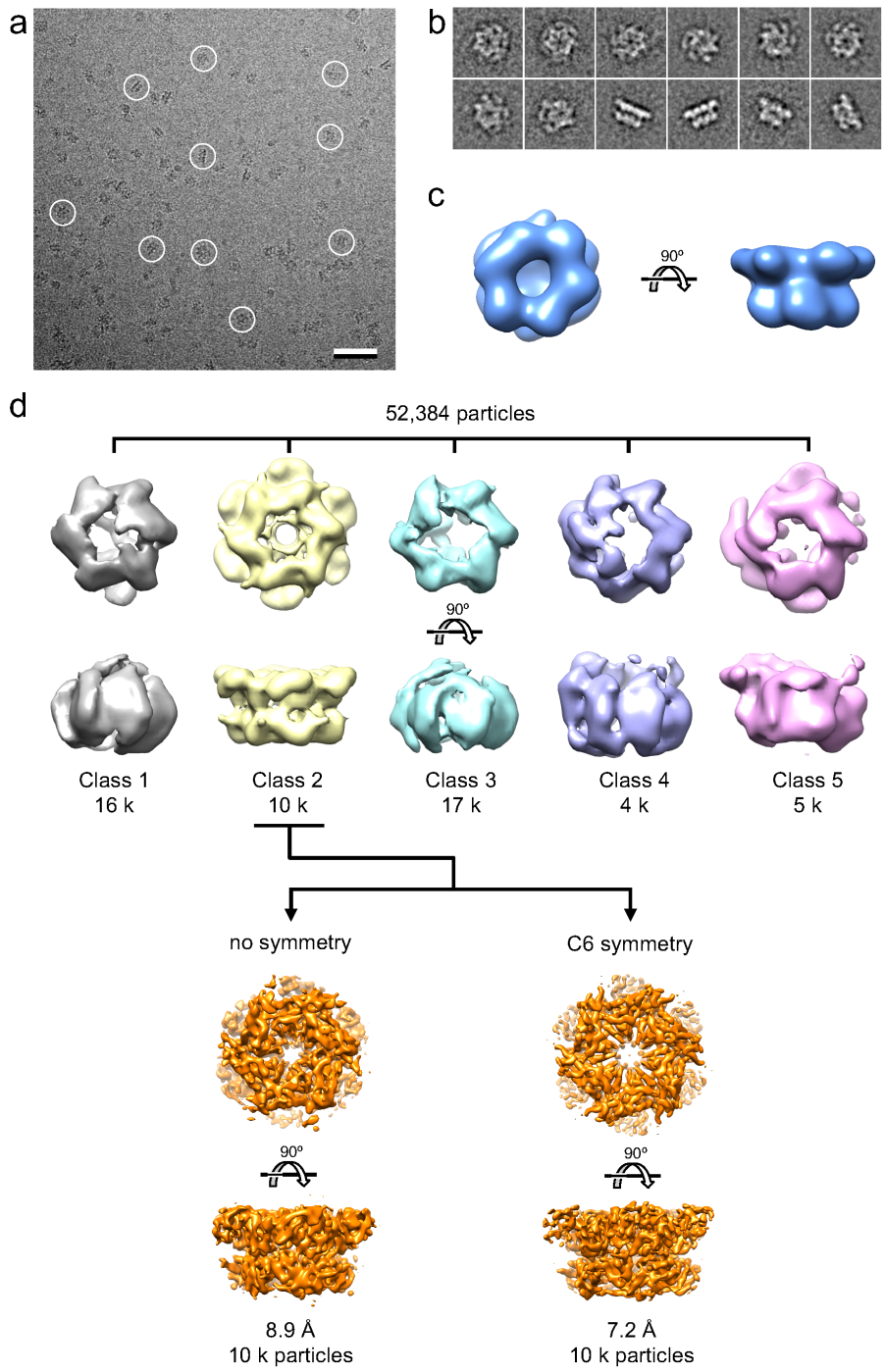



**Figure 3.S2: Image processing of Cdc48 in the presence of ADP. a**, An area of a cryo-EM image of a vitrified sample is shown with some particles circled. Scale bar: 50 nm. **b**, Selected 2D class averages obtained with ISAC. Side length of individual averages: 32 nm. **c**, Initial 3D

**Figure 3.S2 (Continued)**

map obtained with VIPER. **d**, Image-processing workflow for 3D classification and refinement in RELION-1.4 that yielded density maps at resolutions of 8.9 Å (with no symmetry imposed) and 7.2 Å (with C6 symmetry imposed). See Methods section for details.

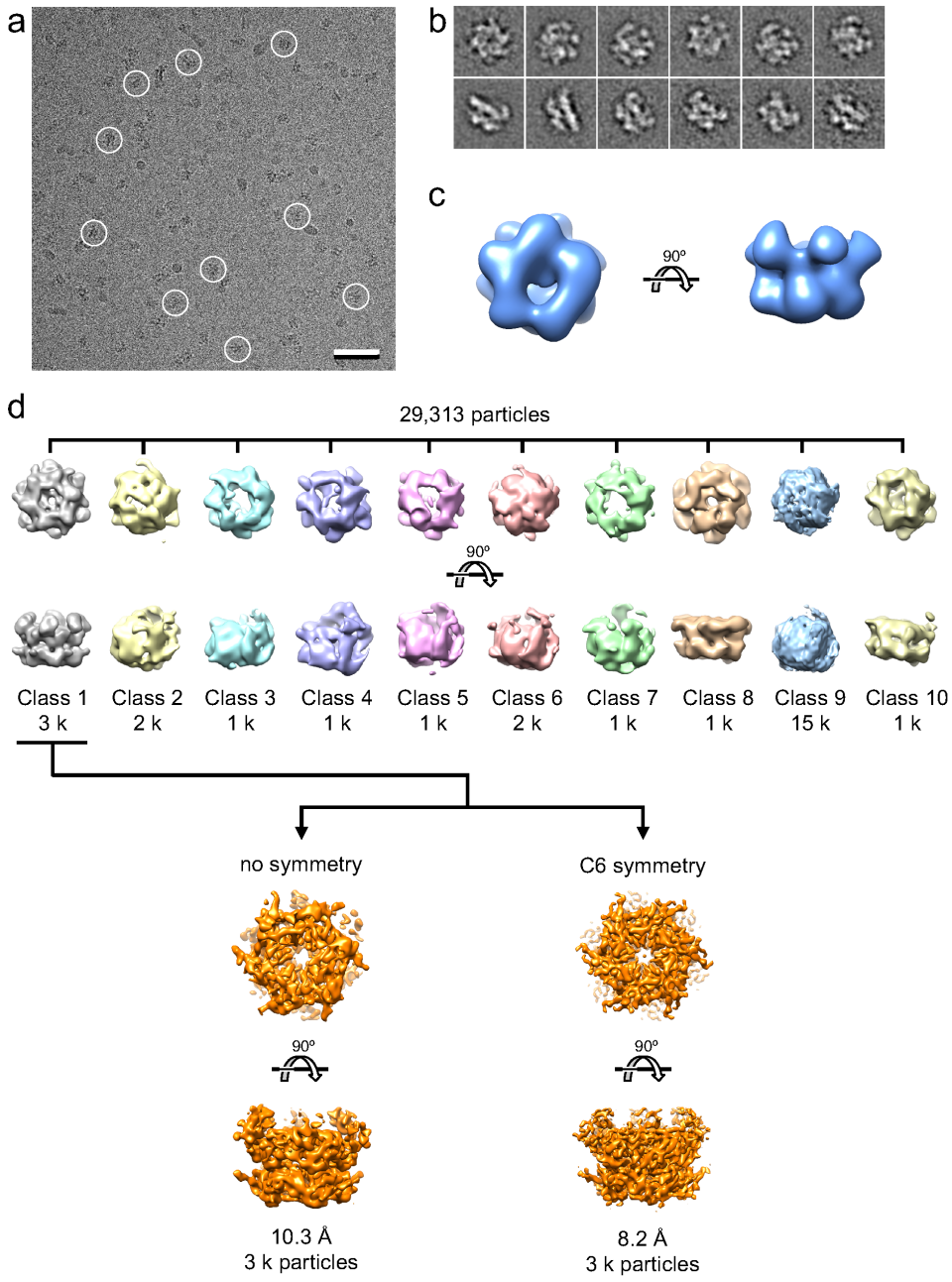



**Figure 3.S3: Image processing of Cdc48 in the presence of ATPγS. a**, An area of a cryo-EM image of a vitrified sample is shown with some particles circled. Scale bar: 50 nm. **b**, Selected 2D class averages obtained with ISAC. Side length of individual averages: 25 nm. **c**, Initial 3D map obtained with VIPER. **d**, Image-processing workflow for 3D classification and refinement

**Figure 3.S3 (Continued)**

in RELION-1.4 that yielded density maps at resolutions of 10.3 Å (with no symmetry imposed) and 8.2 Å (with C6 symmetry imposed). See Methods section for details.

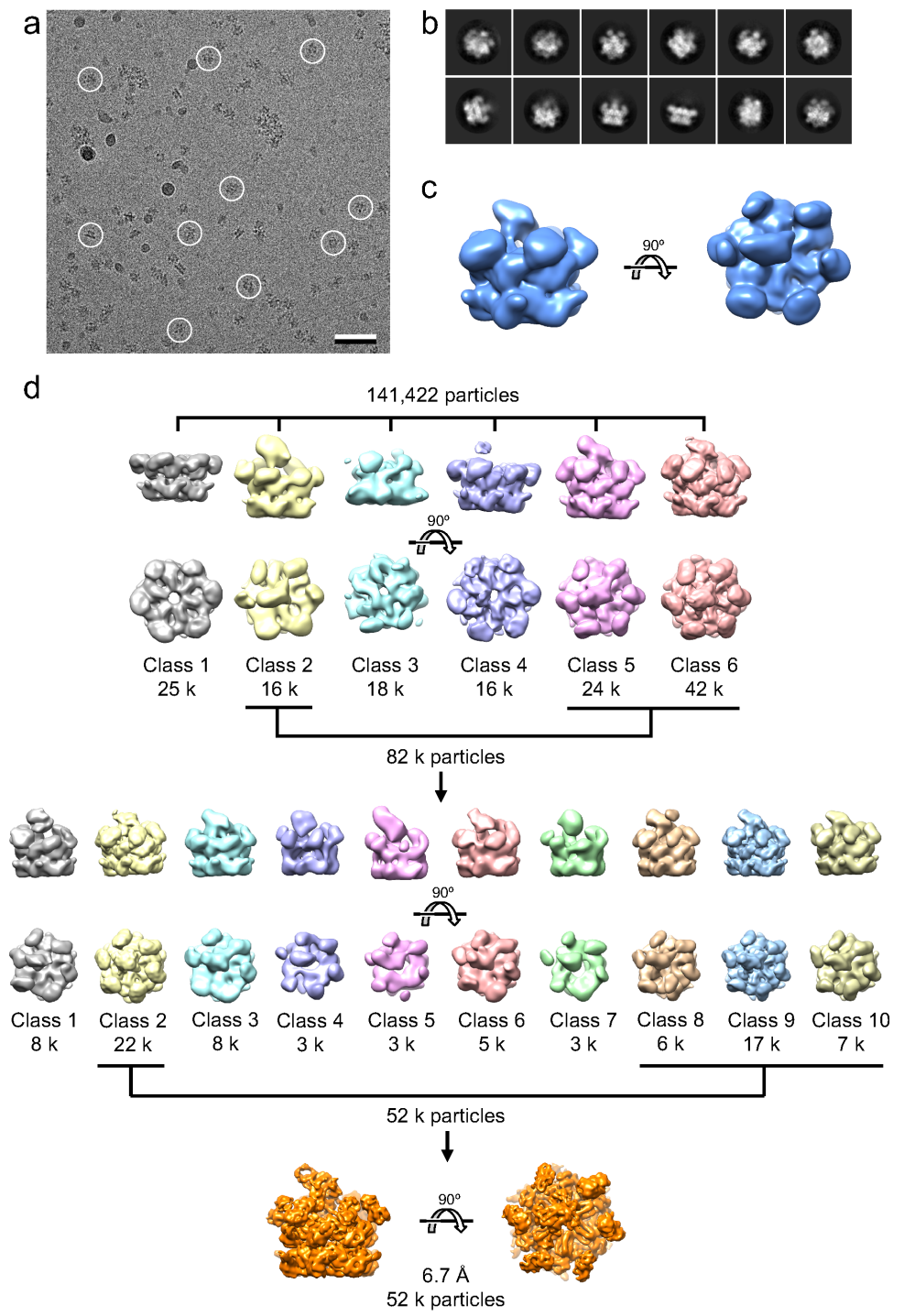



**Figure 3.S4: Image processing of the Cdc48–cofactor complex in the presence of ADP. a**, An area of a cryo-EM image of a vitrified sample is shown with some particles circled. Scale bar: 50 nm. **b**, Selected 2D class averages obtained with RELION-1.4. Side length of individual

**Figure 3.S4 (Continued)**

averages: 33 nm. **c**, Initial 3D map obtained with RELION-1.4. **d**, Image-processing workflow for 3D classification and refinement in RELION-1.4 that yielded a density map at 6.7 Å resolution. See Methods section for details.

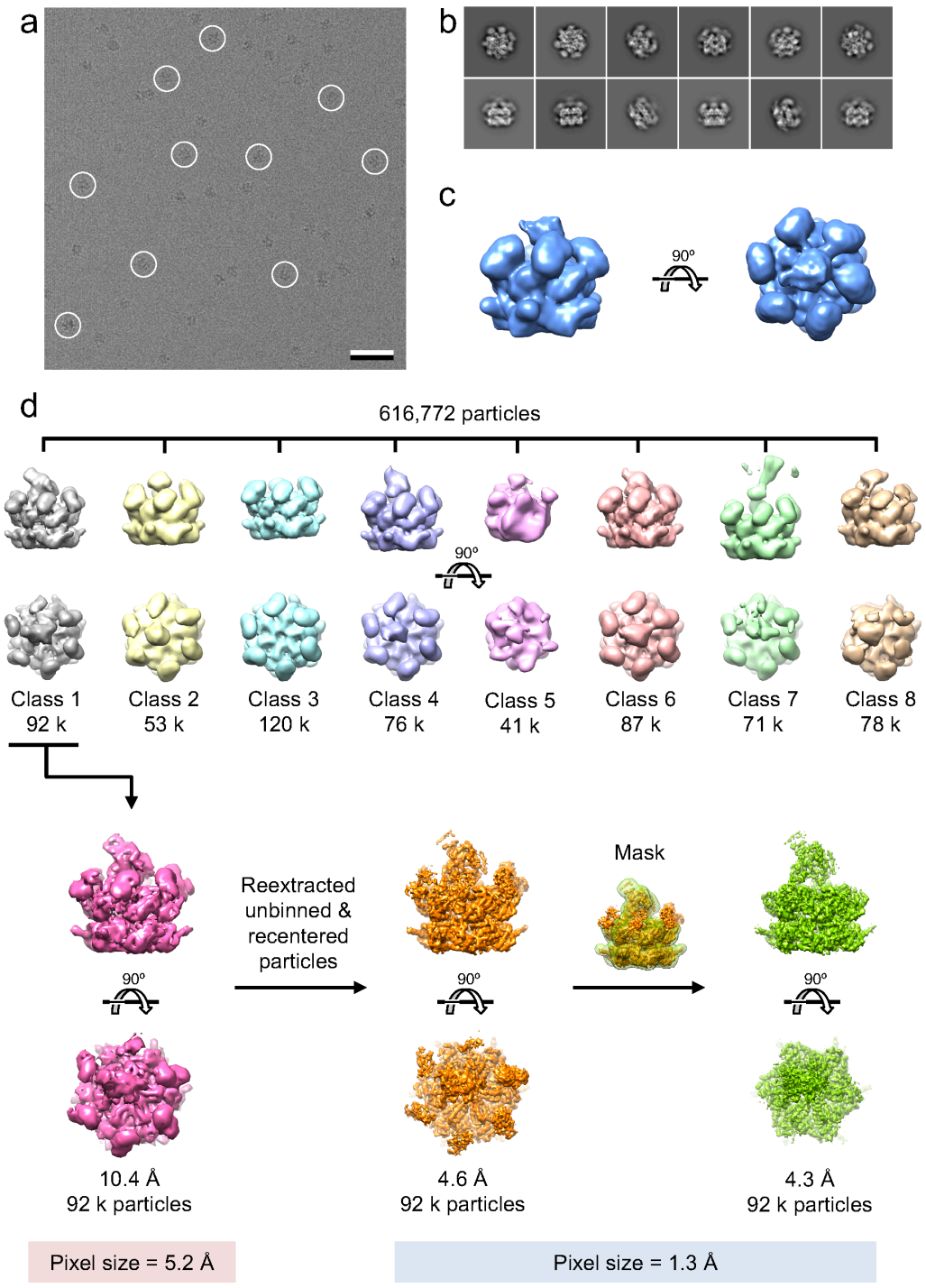



**Figure 3.S5: . Image processing of the Cdc48–cofactor complex in the presence of ATPγS.**

**a**, An area of a cryo-EM image of a vitrified sample is shown with some particles circled. Scale bar: 50 nm. **b**, Selected 2D class averages obtained with ISAC. Side length of individual averages: 33 nm. **c**, Initial 3D map obtained with cryoSPARC. **d**, Image-processing workflow for

**Figure 3.S5 (Continued)**

3D classification and refinement in RELION-2 that yielded a density map at 4.3 Å. See Methods section for details.

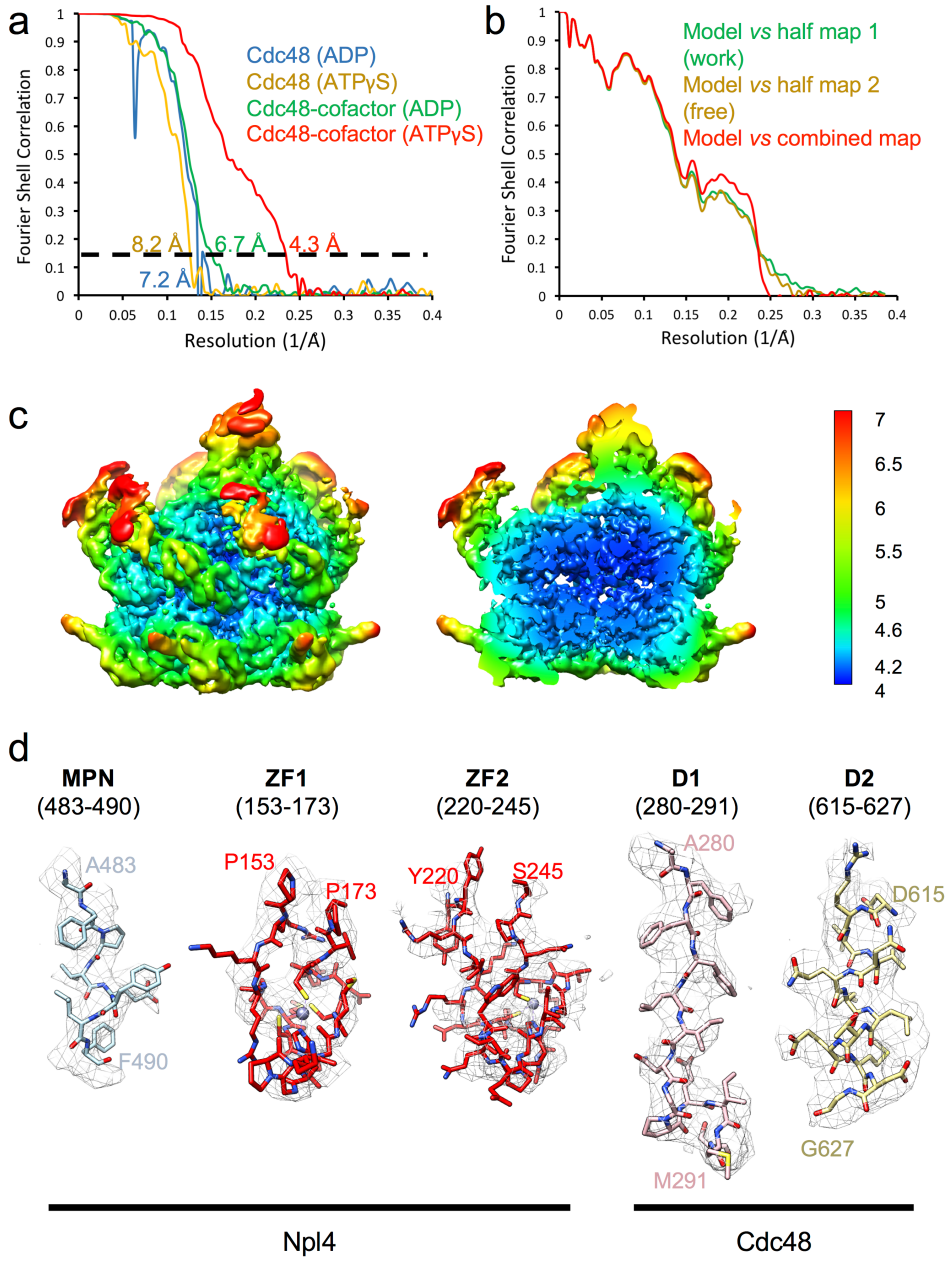



**Figure 3.S6: FSC curves and local densities. a**, FSC curves calculated between independently refined half maps for the density map of Cdc48 alone in the presence of ADP (blue) or ATPγS (yellow), and the Cdc48–cofactor complex in the presence of ADP (green) or ATPγS (red). **b**, Cross-validation FSC curves showing no significant overfitting. **c**, Local resolution map of the Cdc48–cofactor complex structure obtained in the presence of ATPγS. Views are shown for a

**Figure 3.S6 (Continued)**

complete map (left) and a map after removing the front half (right). **d**, Representative cryo-EM densities with the fit models. ZF1, ZF2, $Zn^{2+}$-fingers.

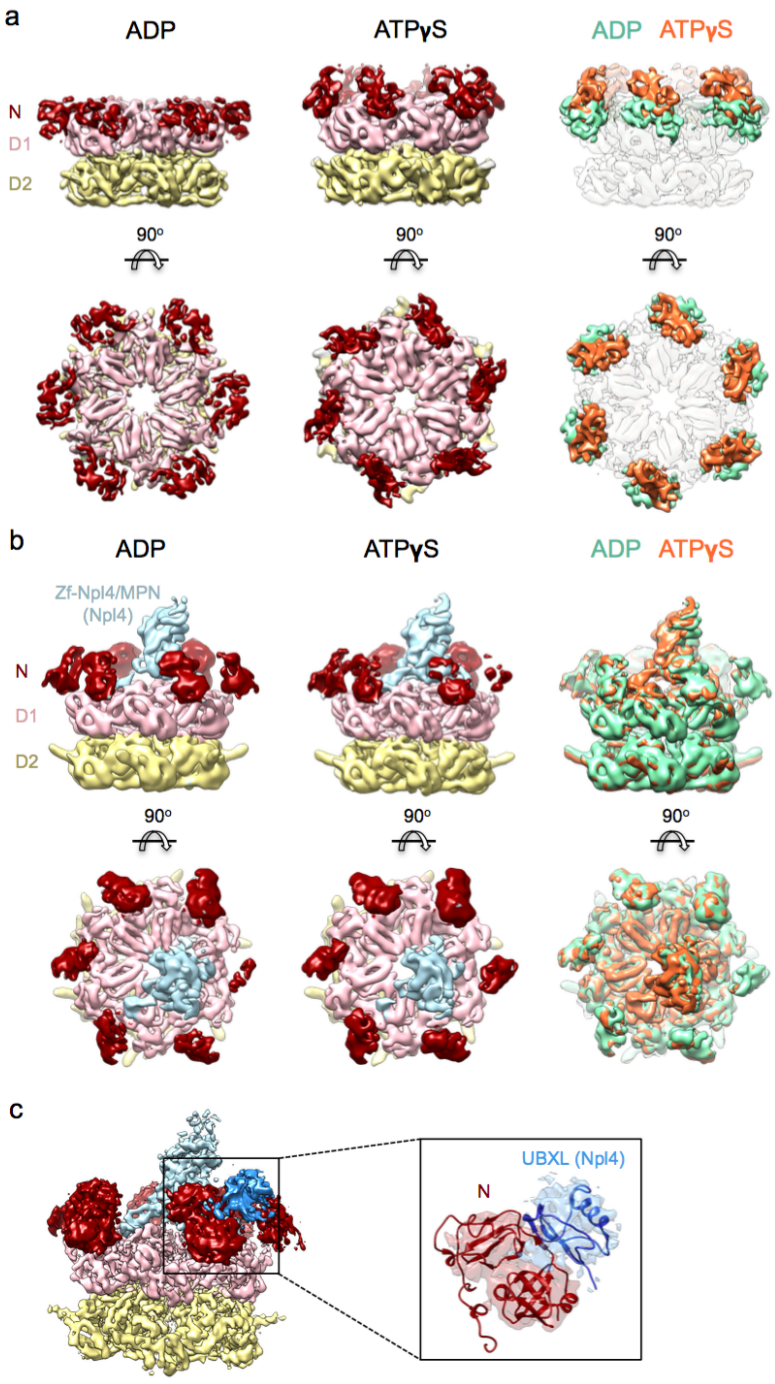



**Figure 3.S7: Comparison of the Cdc48 and Cdc48/UN structures obtained in the presence of ADP and ATPγS. a**, The cryo-EM density maps of Cdc48 obtained in the presence of ADP and ATPγS are shown separately (left and middle panels) as well as superimposed (right panel). All maps were low pass-filtered to 8 Å. **b**, As in **a**, but for the Cdc48-cofactor complex. All maps were low pass-filtered to 7 Å. **c**, The cryo-EM density map of the Cdc48–cofactor complexes

**Figure 3.S7 (Continued)**

obtained in the presence of ATPγS (left) showed density (blue) for the UBX-like domain of Npl4. An NMR structure of the complex of an N domain and the UBX-like domain (PDB: 2PJH) was docked into this cryo-EM density (inset on the right).

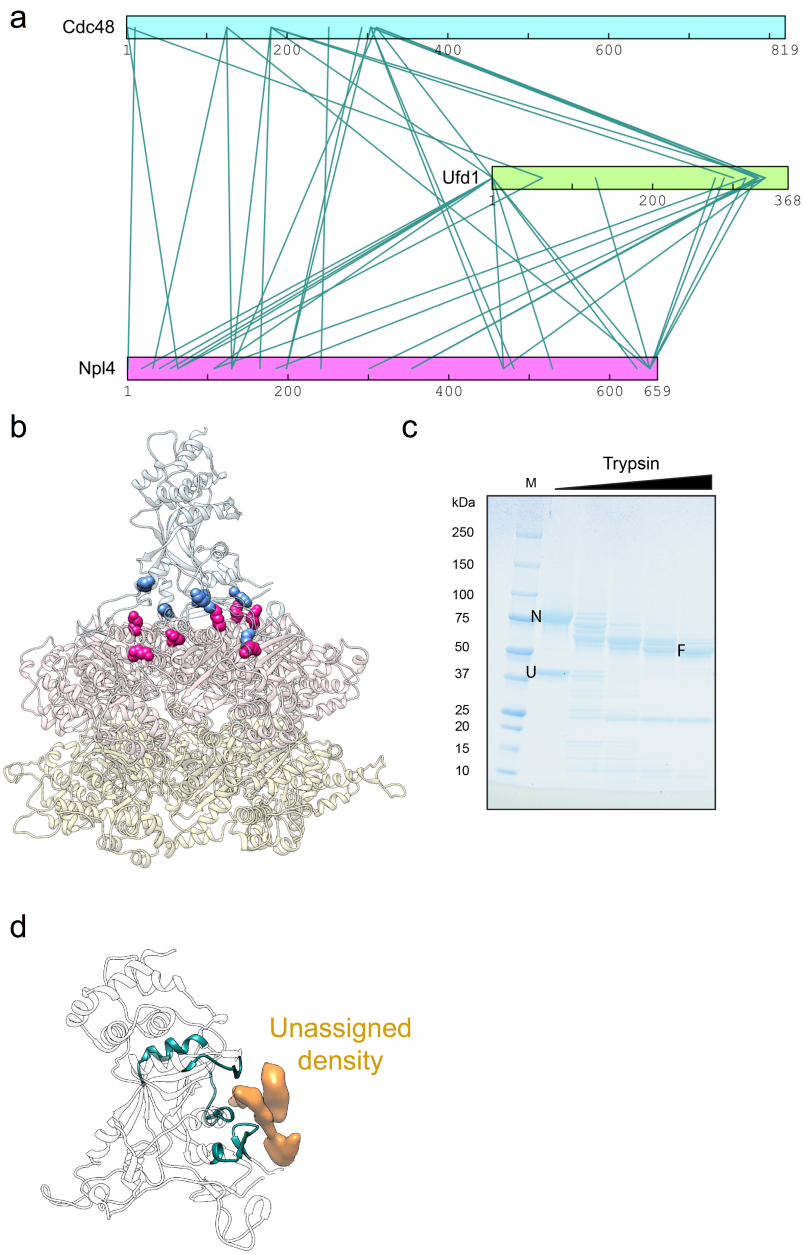



**Figure 3.S8. Analysis of interactions in the Cdc48–cofactor complex.** **a**, The *C. thermophilum* Cdc48/UN complex was treated with bifunctional crosslinkers and crosslinks were determined by mass spectrometry. **b**, Residues involved in crosslinks between Cdc48 and Npl4 are mapped onto the structure and shown as red and blue spheres, respectively. **c**, Purified *C. thermophilum* UN was treated with increasing concentrations of trypsin and subjected to SDS-PAGE and Coomassie staining. N, U: full-length Npl4 and Ufd1. F: a stable fragment, identified

**Figure 3.S8 (Continued)**

by mass spectrometry as Npl4 129-602. **d**, Hydrogen-deuterium (H/D) exchange was performed with the Npl4 construct used for crystallization in the absence or presence of Ufd1. Npl4 backbone residues protected by Ufd1 are colored in turquoise (residues 262-274, 316-327, 360-374, and 428-444) and mapped onto the structure of Npl4 in ribbon representation. The unassigned density from the cryo-EM map (in orange) is predicted to belong to Ufd1.

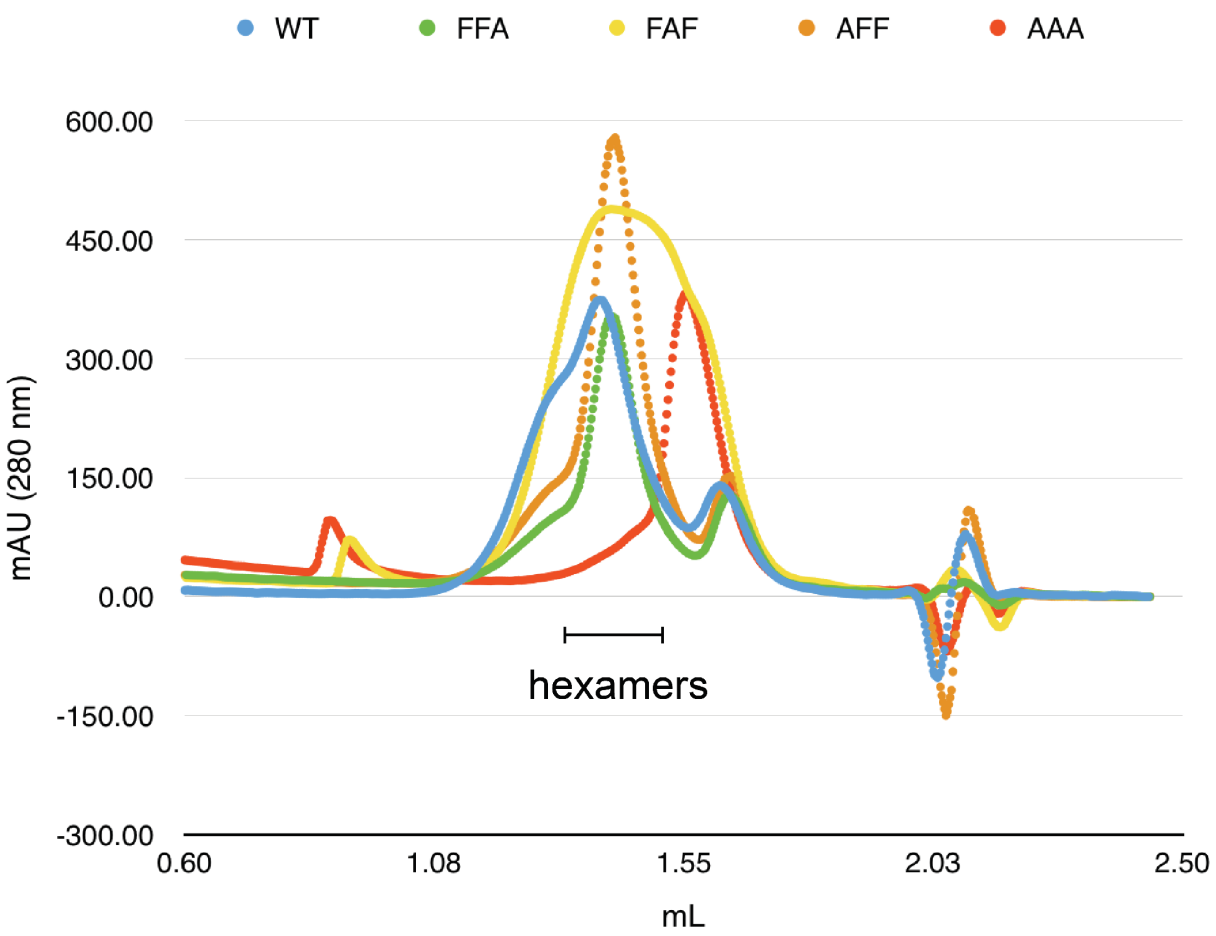



**Figure 3.S9: Hexamerization of Cdc48 FFF mutants**. The indicated Cdc48 variants were subjected to gel filtration to analyze their oligomeric states. The position of hexamers is indicated. Note that mutation of the central Phe residue (FAF and AAA mutants) affects the migration of Cdc48 in gel filtration.

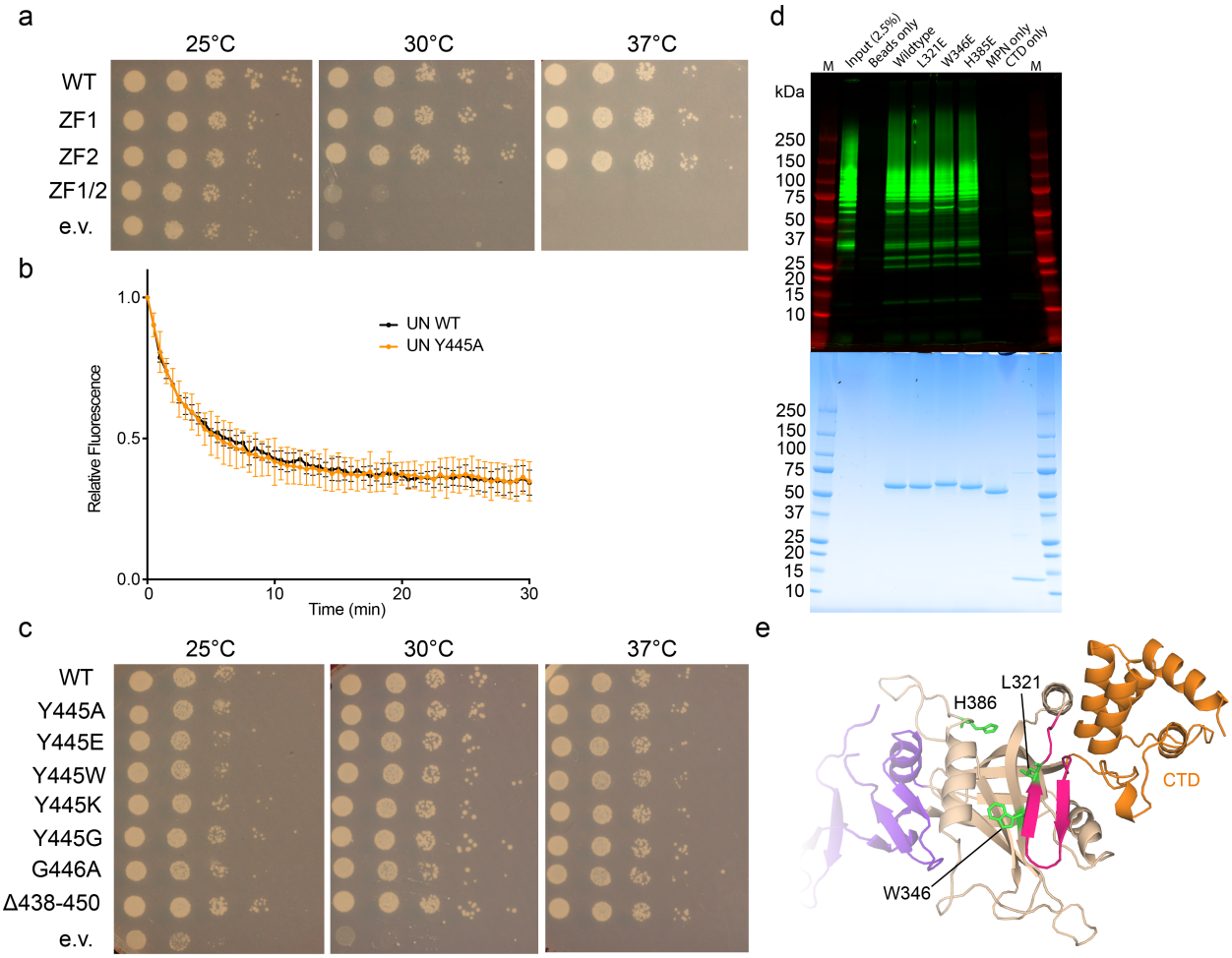



**Figure 3.S10: Functional analysis of $Zn^{2+}$-finger and MPN domain mutations**. **a,** An *npl4-1* temperature-sensitive *S. cerevisiae* strain was transformed with a plasmid encoding wild type Npl4 or the indicated $Zn^{2+}$-finger mutants, spotted in serial dilution, and incubated at the indicated temperatures for two days (30 and 37 °C) or three days (25 °C). ZF1: H139A/C145A. ZF2: H208A/C216A. ZF1/2: H139A/C145A/H208A/C216A. e.v., empty vector. **b**, Unfolding of poly-ubiquitinated Eos with Cdc48 and the indicated UN variants. Data are shown as mean ± SD of n=3 technical replicates. **c**, As in **a**, with the indicated mutations in the Npl4 insert-2 beta strand finger. **d**, Binding of polyubiquitinated substrate to SBP-tagged Npl4 ($Zn^{2+}$-finger/MPN/CTD domains, residues 129-602) and to the indicated variants (MPN only: residues 129-519, CTD only: residues 519-602). The bait proteins were bound to streptavidin beads. Dye-

**Figure 3.S10 (Continued)**

labeled, polyubiquitinated superfolder GFP was added, and bound material was assessed by SDS-PAGE, fluorescence scanning (top), and Coomassie staining (bottom). M, molecular weight marker. **e**, Locations of the mutants tested in **d**.

**Table 3.S1. Data collection and refinement statistics.**

| | Npl4 zf/MPN/CTD |
|---|---|
| **Wavelength** | 1.283 |
| **Resolution range** | 96.11 - 2.582 (2.675 - 2.582) |
| **Space group** | P 1 21 1 |
| **Unit cell** | 58.86 72.22 193.54 90 96.71 90 |
| **Total reflections** | 279737 (24345) |
| **Unique reflections** | 96089 (7754) |
| **Multiplicity** | 2.9 (2.7) |
| **Completeness (%)** | 95.15 (77.79) |
| **Mean I/sigma(I)** | 9.82 (0.49) |
| **Wilson B-factor** | 81.66 |
| **R-merge** | 0.064 (2.03) |
| **R-meas** | 0.078 (2.50) |
| **R-pim** | 0.044 (1.44) |
| **CC1/2** | 0.998 (0.32) |
| **CC*** | 1 (0.70) |
| **Reflections used in refinement** | 94503 (7751) |
| **Reflections used for R-free** | 4617 (394) |
| **R-work** | 0.19 (0.43) |
| **R-free** | 0.23 (0.46) |
| **CC(work)** | 0.95 (0.67) |
| **CC(free)** | 0.97 (0.58) |
| **Number of non-hydrogen atoms** | 7515 |
| **macromolecules** | 7283 |

**Table 3.S1 (Continued)**

| | |
|---|---|
| **ligands** | 4 |
| **solvent** | 228 |
| **Protein residues** | 915 |
| **RMS(bonds)** | 0.002 |
| **RMS(angles)** | 0.53 |
| **Ramachandran favored (%)** | 96.03 |
| **Ramachandran allowed (%)** | 3.86 |
| **Ramachandran outliers (%)** | 0.11 |
| **Rotamer outliers (%)** | 0.75 |
| **Clashscore** | 1.18 |
| **Average B-factor** | 96.95 |
| **macromolecules** | 97.44 |
| **ligands** | 120.87 |
| **solvent** | 80.87 |
| **Number of TLS groups** | 6 |

Statistics for the highest-resolution shell are shown in parentheses.

## REFERENCES


1. Etlinger, J. D. & Goldberg, A. L. A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. *Proc. Natl. Acad. Sci. U.S.A.* **74,** 54–58 (1977).

2. Wilkinson, K. D., Urban, M. K. & Haas, A. L. Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. *J. Biol. Chem.* **255,** 7529–7532 (1980).

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

4. Chu-Ping, M., Vu, J. H., Proske, R. J., Slaughter, C. A. & DeMartino, G. N. Identification, purification, and characterization of a high molecular weight, ATP-dependent activator (PA700) of the 20 S proteasome. *J. Biol. Chem.* **269,** 3539–3547 (1994).

5. Dahlmann, B. *et al.* The multicatalytic proteinase (prosome) is ubiquitous from eukaryotes to archaebacteria. *FEBS Lett.* **251,** 125–131 (1989).

6. Liu, C.-W. *et al.* Conformational remodeling of proteasomal substrates by PA700, the 19 S regulatory complex of the 26 S proteasome. *J. Biol. Chem.* **277,** 26815–26820 (2002).

7. Orlowski, M., Cardozo, C. & Michaud, C. Evidence for the presence of five distinct proteolytic components in the pituitary multicatalytic proteinase complex. Properties of two components cleaving bonds on the carboxyl side of branched chain and small neutral amino acids. *Biochemistry* **32,** 1563–1572 (1993).

8. Walz, J. *et al.* 26S proteasome structure revealed by three-dimensional electron microscopy. *J. Struct. Biol.* **121,** 19–29 (1998).

9. Matyskiela, M. E., Lander, G. C. & Martin, A. Conformational switching of the 26S proteasome enables substrate degradation. *Nat. Struct. Mol. Biol.* **20,** 781–788 (2013).

10. Shi, Y. *et al.* Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. *Science* **351,** aad9421–aad9421 (2016).

11. Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453,** 481–488 (2008).

12. van Nocker, S. *et al.* The multiubiquitin-chain-binding protein Mcb1 is a component of the 26S proteasome in Saccharomyces cerevisiae and plays a nonessential, substrate-specific role in protein turnover. *Mol. Cell. Biol.* **16,** 6020–6028 (1996).

13. Rubin, D. M. *et al.* ATPase and ubiquitin-binding proteins of the yeast proteasome. *Mol. Biol. Rep.* **24,** 17–26 (1997).

14. Rubin, D. M., Glickman, M. H., Larsen, C. N., Dhruvakumar, S. & Finley, D. Active

site mutants in the six regulatory particle ATPases reveal multiple roles for ATP in the proteasome. *EMBO J.* **17,** 4909–4919 (1998).

15. Finley, D., Chen, X. & Walters, K. J. Gates, Channels, and Switches: Elements of the Proteasome Machine. *Trends Biochem. Sci.* **41,** 77–93 (2016).

16. Beckwith, R., Estrin, E., Worden, E. J. & Martin, A. Reconstitution of the 26S proteasome reveals functional asymmetries in its AAA+ unfoldase. *Nat. Struct. Mol. Biol.* **20,** 1164–1172 (2013).

17. Verma, R. *et al.* Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298,** 611–615 (2002).

18. Guterman, A. & Glickman, M. H. Complementary roles for Rpn11 and Ubp6 in deubiquitination and proteolysis by the proteasome. *J. Biol. Chem.* **279,** 1729–1738 (2004).

19. Hanna, J. *et al.* Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell* **127,** 99–111 (2006).

20. Yao, T. & Cohen, R. E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419,** 403–407 (2002).

21. Smith, D. M. *et al.* Docking of the proteasomal ATPases‘ carboxyl termini in the 20S proteasome’s alpha ring opens the gate for substrate entry. *Mol. Cell* **27,** 731–744 (2007).

22. Rabl, J. *et al.* Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol. Cell* **30,** 360–368 (2008).

23. Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* **386,** 463–471 (1997).

24. Groll, M. *et al.* The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 10976–10983 (1999).

25. Collins, G. A. & Goldberg, A. L. The Logic of the 26S Proteasome. *Cell* **169,** 792–806 (2017).

26. Ciechanover, A., Elias, S., Heller, H. & Hershko, A. ‘Covalent affinity’ purification of ubiquitin-activating enzyme. *J. Biol. Chem.* **257,** 2537–2542 (1982).

27. Ciechanover, A., Heller, H., Katz-Etzion, R. & Hershko, A. Activation of the heat-stable polypeptide of the ATP-dependent proteolytic system. *Proc. Natl. Acad. Sci. U.S.A.* **78,** 761–765 (1981).

28. Hershko, A., Heller, H., Elias, S. & Ciechanover, A. Components of ubiquitin-protein

ligase system. Resolution, affinity purification, and role in protein breakdown. *J. Biol. Chem.* **258,** 8206–8214 (1983).

29. Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373,** 81–83 (1995).

30. Metzger, M. B., Hristova, V. A. & Weissman, A. M. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J. Cell. Sci.* **125,** 531–537 (2012).

31. Tait, S. W. G. *et al.* Apoptosis induction by Bid requires unconventional ubiquitination and degradation of its N-terminal fragment. *J. Cell Biol.* **179,** 1453–1466 (2007).

32. McDowell, G. S. & Philpott, A. Non-canonical ubiquitylation: mechanisms and consequences. *Int. J. Biochem. Cell Biol.* **45,** 1833–1842 (2013).

33. Finley, D., Ulrich, H. D., Sommer, T. & Kaiser, P. The ubiquitin-proteasome system of Saccharomyces cerevisiae. *Genetics* **192,** 319–360 (2012).

34. Hershko, A. & Heller, H. Occurrence of a polyubiquitin structure in ubiquitin-protein conjugates. *Biochem. Biophys. Res. Commun.* **128,** 1079–1086 (1985).

35. Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19,** 94–102 (2000).

36. Wu, R. S., Kohn, K. W. & Bonner, W. M. Metabolism of ubiquitinated histones. *J. Biol. Chem.* **256,** 5916–5920 (1981).

37. Bishop, N., Horman, A. & Woodman, P. Mammalian class E vps proteins recognize ubiquitin and act in the removal of endosomal protein-ubiquitin conjugates. *J. Cell Biol.* **157,** 91–101 (2002).

38. Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M. & Matouschek, A. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol. Cell* **7,** 627–637 (2001).

39. Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. & Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat. Struct. Mol. Biol.* **11,** 830–837 (2004).

40. van der Lee, R. *et al.* Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep* **8,** 1832–1844 (2014).

41. Inobe, T., Fishbain, S., Prakash, S. & Matouschek, A. Defining the geometry of the two-component proteasome degron. *Nat. Chem. Biol.* **7,** 161–167 (2011).

42. Hiyama, H. *et al.* Interaction of hHR23 with S5a. The ubiquitin-like domain of hHR23 mediates interaction with S5a subunit of 26 S proteasome. *J. Biol. Chem.* **274,** 28019–28025 (1999).

43. Mueller, T. D., Kamionka, M. & Feigon, J. Specificity of the interaction between ubiquitin-associated domains and ubiquitin. *J. Biol. Chem.* **279,** 11926–11936 (2004).

44. Fishbain, S., Prakash, S., Herrig, A., Elsasser, S. & Matouschek, A. Rad23 escapes degradation because it lacks a proteasome initiation region. *Nat Commun* **2,** 192 (2011).

45. Gödderz, D. *et al.* Cdc48-independent proteasomal degradation coincides with a reduced need for ubiquitylation. *Sci Rep* **5,** 7615 (2015).

46. Stolz, A., Hilt, W., Buchberger, A. & Wolf, D. H. Cdc48: a power machine in protein degradation. *Trends Biochem. Sci.* **36,** 515–523 (2011).

47. Meyer, H., Bug, M. & Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat. Cell Biol.* **14,** 117–123 (2012).

48. Rape, M. *et al.* Mobilization of processed, membrane-tethered SPT23 transcription factor by CDC48(UFD1/NPL4), a ubiquitin-selective chaperone. *Cell* **107,** 667–677 (2001).

49. Bergink, S. *et al.* Role of Cdc48/p97 as a SUMO-targeted segregase curbing Rad51-Rad52 interaction. *Nat. Cell Biol.* **15,** 526–532 (2013).

50. Beskow, A. *et al.* A conserved unfoldase activity for the p97 AAA-ATPase in proteasomal degradation. *J. Mol. Biol.* **394,** 732–746 (2009).

51. Fröhlich, K. U. *et al.* Yeast cell cycle protein CDC48p shows full-length homology to the mammalian protein VCP and is a member of a protein family involved in secretion, peroxisome formation, and gene expression. *J. Cell Biol.* **114,** 443–453 (1991).

52. Moir, D., Stewart, S. E., Osmond, B. C. & Botstein, D. Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies. *Genetics* **100,** 547–563 (1982).

53. Peters, J. M., Walsh, M. J. & Franke, W. W. An abundant and ubiquitous homo-oligomeric ring-shaped ATPase particle related to the putative vesicle fusion proteins Sec18p and NSF. *EMBO J.* **9,** 1757–1767 (1990).

54. Zhang, X. *et al.* Structure of the AAA ATPase p97. *Mol. Cell* **6,** 1473–1484 (2000).

55. Banerjee, S. *et al.* 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351,** 871–875 (2016).

56. Ghislain, M., Dohmen, R. J., Levy, F. & Varshavsky, A. Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in Saccharomyces cerevisiae. *EMBO J.* **15,** 4884–4899 (1996).

57. Johnson, E. S., Ma, P. C., Ota, I. M. & Varshavsky, A. A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.* **270,** 17442–17456 (1995).

58. Hänzelmann, P. & Schindelin, H. The Interplay of Cofactor Interactions and Post-translational Modifications in the Regulation of the AAA+ ATPase p97. *Front Mol Biosci* **4,** 21 (2017).

59. Dreveny, I. *et al.* Structural basis of the interaction between the AAA ATPase p97/VCP and its adaptor protein p47. *EMBO J.* **23,** 1030–1039 (2004).

60. Zhao, G., Li, G., Schindelin, H. & Lennarz, W. J. An Armadillo motif in Ufd3 interacts with Cdc48 and is involved in ubiquitin homeostasis and protein degradation. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 16197–16202 (2009).

61. Allen, M. D., Buchberger, A. & Bycroft, M. The PUB domain functions as a p97 binding module in human peptide N-glycanase. *J. Biol. Chem.* **281,** 25502–25508 (2006).

62. Hänzelmann, P., Buchberger, A. & Schindelin, H. Hierarchical binding of cofactors to the AAA ATPase p97. *Structure* **19,** 833–843 (2011).

63. Schuberth, C. & Buchberger, A. UBX domain proteins: major regulators of the AAA ATPase Cdc48/p97. *Cell. Mol. Life Sci.* **65,** 2360–2371 (2008).

64. Alexandru, G. *et al.* UBXD7 binds multiple ubiquitin ligases and implicates p97 in HIF1alpha turnover. *Cell* **134,** 804–816 (2008).

65. Meyer, H. H., Kondo, H. & Warren, G. The p47 co-factor regulates the ATPase activity of the membrane fusion protein, p97. *FEBS Lett.* **437,** 255–257 (1998).

66. Arumughan, A. *et al.* Quantitative interaction mapping reveals an extended UBX domain in ASPL that disrupts functional p97 hexamers. *Nat Commun* **7,** 13047 (2016).

67. Liu, Y. & Ye, Y. Roles of p97-associated deubiquitinases in protein quality control at the endoplasmic reticulum. *Curr. Protein Pept. Sci.* **13,** 436–446 (2012).

68. Rumpf, S. & Jentsch, S. Functional Division of Substrate Processing Cofactors of the Ubiquitin-Selective Cdc48 Chaperone. *Mol. Cell* **21,** 261–269 (2006).

69. Liu, C. *et al.* Ubiquitin chain elongation enzyme Ufd2 regulates a subset of Doa10 substrates. *J. Biol. Chem.* **285,** 10265–10272 (2010).

70. Hosomi, A., Fujita, M., Tomioka, A., Kaji, H. & Suzuki, T. Identification of PNGase-dependent ERAD substrates in Saccharomyces cerevisiae. *Biochem. J.* **473,** 3001–3012 (2016).

71. Meyer, H. H., Shorter, J. G., Seemann, J., Pappin, D. & Warren, G. A complex of mammalian ufd1 and npl4 links the AAA-ATPase, p97, to ubiquitin and nuclear transport pathways. *EMBO J.* **19,** 2181–2192 (2000).

72. Ye, Y., Meyer, H. H. & Rapoport, T. A. Function of the p97-Ufd1-Npl4 complex in

retrotranslocation from the ER to the cytosol: dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains. *J. Cell Biol.* **162,** 71–84 (2003).

73. Bruderer, R. M., Brasseur, C. & Meyer, H. H. The AAA ATPase p97/VCP interacts with its alternative co-factors, Ufd1-Npl4 and p47, through a common bipartite binding mechanism. *J. Biol. Chem.* **279,** 49609–49616 (2004).

74. Barlowe, C. K. & Miller, E. A. Secretory protein biogenesis and traffic in the early secretory pathway. *Genetics* **193,** 383–410 (2013).

75. Park, E. & Rapoport, T. A. Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu Rev Biophys* **41,** 21–40 (2012).

76. Araki, K. & Nagata, K. Protein folding and quality control in the ER. *Cold Spring Harb Perspect Biol* **4,** a015438–a015438 (2012).

77. Helenius, A. How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell* **5,** 253–265 (1994).

78. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73,** 1019–1049 (2004).

79. Williams, D. B. Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum. *J. Cell. Sci.* **119,** 615–623 (2006).

80. Van Leeuwen, J. E. & Kearse, K. P. Reglucosylation of N-linked glycans is critical for calnexin assembly with T cell receptor (TCR) alpha proteins but not TCRbeta proteins. *J. Biol. Chem.* **272,** 4179–4186 (1997).

81. Molinari, M., Calanca, V., Galli, C., Lucca, P. & Paganetti, P. Role of EDEM in the release of misfolded glycoproteins from the calnexin cycle. *Science* **299,** 1397–1400 (2003).

82. Gauss, R., Jarosch, E., Sommer, T. & Hirsch, C. A complex of Yos9p and the HRD ligase integrates endoplasmic reticulum quality control into the degradation machinery. *Nat. Cell Biol.* **8,** 849–854 (2006).

83. Stein, A., Ruggiano, A., Carvalho, P. & Rapoport, T. A. Key steps in ERAD of luminal ER proteins reconstituted with purified components. *Cell* **158,** 1375–1388 (2014).

84. Carvalho, P., Goder, V. & Rapoport, T. A. Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. *Cell* **126,** 361–373 (2006).

85. Carvalho, P., Stanley, A. M. & Rapoport, T. A. Retrotranslocation of a misfolded luminal ER protein by the ubiquitin-ligase Hrd1p. *Cell* **143,** 579–591 (2010).

86. Baldridge, R. D. & Rapoport, T. A. Autoubiquitination of the Hrd1 Ligase Triggers Protein Retrotranslocation in ERAD. *Cell* **166,** 394–407 (2016).

87. Ye, Y., Shibata, Y., Yun, C., Ron, D. & Rapoport, T. A. A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. *Nature* **429,** 841–847 (2004).

88. Ye, Y., Meyer, H. H. & Rapoport, T. A. The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol. *Nature* **414,** 652–656 (2001).

89. Neuber, O., Jarosch, E., Volkwein, C., Walter, J. & Sommer, T. Ubx2 links the Cdc48 complex to ER-associated protein degradation. *Nat. Cell Biol.* **7,** 993–998 (2005).

90. Stolz, A. & Wolf, D. H. Endoplasmic reticulum associated protein degradation: a chaperone assisted journey to hell. *Biochim. Biophys. Acta* **1803,** 694–705 (2010).

91. Medicherla, B., Kostova, Z., Schaefer, A. & Wolf, D. H. A genomic screen identifies Dsk2p and Rad23p as essential components of ER-associated degradation. *EMBO Rep.* **5,** 692–697 (2004).

92. Wang, Q. *et al.* A ubiquitin ligase-associated chaperone holdase maintains polypeptides in soluble states for proteasome degradation. *Mol. Cell* **42,** 758–770 (2011).

93. Madeo, F., Schlauer, J., Zischka, H., Mecke, D. & Fröhlich, K. U. Tyrosine phosphorylation regulates cell cycle-dependent nuclear localization of Cdc48p. *Mol. Biol. Cell* **9,** 131–141 (1998).

94. Vaz, B., Halder, S. & Ramadan, K. Role of p97/VCP (Cdc48) in genome stability. *Front Genet* **4,** 60 (2013).

95. Arias, E. E. & Walter, J. C. Replication-dependent destruction of Cdt1 limits DNA replication to a single round per cell cycle in Xenopus egg extracts. *Genes Dev.* **19,** 114–126 (2005).

96. Raman, M., Havens, C. G., Walter, J. C. & Harper, J. W. A genome-wide screen identifies p97 as an essential regulator of DNA damage-dependent CDT1 destruction. *Mol. Cell* **44,** 72–84 (2011).

97. Franz, A. *et al.* Chromatin-associated degradation is defined by UBXN-3/FAF1 to safeguard DNA replication fork progression. *Nat Commun* **7,** 10612 (2016).

98. Havens, C. G. & Walter, J. C. Docking of a specialized PIP Box onto chromatin-bound PCNA creates a degron for the ubiquitin ligase CRL4Cdt2. *Mol. Cell* **35,** 93–104 (2009).

99. Maric, M., Maculins, T., De Piccoli, G. & Labib, K. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science* **346,** 1253596–1253596 (2014).

100. Moreno, S. P., Bailey, R., Campion, N., Herron, S. & Gambus, A. Polyubiquitylation drives replisome disassembly at the termination of DNA replication. *Science* **346,** 477–481 (2014).

101. Maric, M., Mukherjee, P., Tatham, M. H., Hay, R. & Labib, K. Ufd1-Npl4 Recruit Cdc48 for Disassembly of Ubiquitylated CMG Helicase at the End of Chromosome Replication. *Cell Rep* **18,** 3033–3042 (2017).

102. Bochman, M. L. & Schwacha, A. The Mcm2-7 complex has in vitro helicase activity. *Mol. Cell* **31,** 287–293 (2008).

103. Koonin, E. V. A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res.* **21,** 2541–2547 (1993).

104. Spagnolo, L., Rivera-Calzada, A., Pearl, L. H. & Llorca, O. Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Mol. Cell* **22,** 511–519 (2006).

105. Postow, L. Destroying the ring: Freeing DNA from Ku with ubiquitin. *FEBS Lett.* **585,** 2876–2882 (2011).

106. Postow, L. & Funabiki, H. An SCF complex containing Fbxl12 mediates DNA damage-induced Ku80 ubiquitylation. *Cell Cycle* **12,** 587–595 (2013).

107. van den Boom, J. *et al.* VCP/p97 Extracts Sterically Trapped Ku70/80 Rings from DNA in Double-Strand Break Repair. *Mol. Cell* **64,** 189–198 (2016).

108. Meerang, M. *et al.* The ubiquitin-selective segregase VCP/p97 orchestrates the response to DNA double-strand breaks. *Nat. Cell Biol.* **13,** 1376–1382 (2011).

109. Acs, K. *et al.* The AAA-ATPase VCP/p97 promotes 53BP1 recruitment by removing L3MBTL1 from DNA double-strand breaks. *Nat. Struct. Mol. Biol.* **18,** 1345–1350 (2011).

110. Torrecilla, I., Oehler, J. & Ramadan, K. The role of ubiquitin-dependent segregase p97 (VCP or Cdc48) in chromatin dynamics after DNA double strand breaks. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **372,** 20160282 (2017).

111. Edenberg, E. R., Downey, M. & Toczyski, D. Polymerase stalling during replication, transcription and translation. *Curr. Biol.* **24,** R445–52 (2014).

112. Verma, R., Oania, R., Fang, R., Smith, G. T. & Deshaies, R. J. Cdc48/p97 mediates UV-dependent turnover of RNA Pol II. *Mol. Cell* **41,** 82–92 (2011).

113. Lafon, A. *et al.* INO80 Chromatin Remodeler Facilitates Release of RNA Polymerase II from Chromatin for Ubiquitin-Mediated Proteasomal Degradation. *Mol. Cell* **60,** 784–796 (2015).

114. Gardner, R. G., Nelson, Z. W. & Gottschling, D. E. Degradation-mediated protein quality control in the nucleus. *Cell* **120,** 803–815 (2005).

115. Rosenbaum, J. C. *et al.* Disorder targets misorder in nuclear quality control degradation: a disordered ubiquitin ligase directly recognizes its misfolded substrates. *Mol. Cell* **41,** 93–106 (2011).

116. Gallagher, P. S., Clowes Candadai, S. V. & Gardner, R. G. The requirement for Cdc48/p97 in nuclear protein quality control degradation depends on the substrate and correlates with substrate insolubility. *J. Cell. Sci.* **127,** 1980–1991 (2014).

117. Woulfe, J. Nuclear bodies in neurodegenerative disease. *Biochim. Biophys. Acta* **1783,** 2195–2206 (2008).

118. Mouton-Liger, F., Jacoupy, M., Corvol, J.-C. & Corti, O. PINK1/Parkin-Dependent Mitochondrial Surveillance: From Pleiotropy to Parkinson's Disease. *Front Mol Neurosci* **10,** 120 (2017).

119. Ordureau, A. *et al.* Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis. *Mol. Cell* **56,** 360–375 (2014).

120. Tanaka, A. *et al.* Proteasome and p97 mediate mitophagy and degradation of mitofusins induced by Parkin. *J. Cell Biol.* **191,** 1367–1380 (2010).

121. Cohen, M. M. J., Leboucher, G. P., Livnat-Levanon, N., Glickman, M. H. & Weissman, A. M. Ubiquitin-proteasome-dependent degradation of a mitofusin, a critical regulator of mitochondrial fusion. *Mol. Biol. Cell* **19,** 2457–2464 (2008).

122. Xu, S., Peng, G., Wang, Y., Fang, S. & Karbowski, M. The AAA-ATPase p97 is essential for outer mitochondrial membrane protein turnover. *Mol. Biol. Cell* **22,** 291–300 (2011).

123. Wu, X., Li, L. & Jiang, H. Doa1 targets ubiquitinated substrates for mitochondria-associated degradation. *J. Cell Biol.* **213,** 49–63 (2016).

124. Heo, J.-M. *et al.* A stress-responsive system for mitochondrial protein degradation. *Mol. Cell* **40,** 465–480 (2010).

125. Nielson, J. R. *et al.* Sterol Oxidation Mediates Stress-Responsive Vms1 Translocation to Mitochondria. *Mol. Cell* **68,** 673–685.e6 (2017).

126. Hetz, C. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat. Rev. Mol. Cell Biol.* **13,** 89–102 (2012).

127. LaRiviere, F. J., Cole, S. E., Ferullo, D. J. & Moore, M. J. A late-acting quality control process for mature eukaryotic rRNAs. *Mol. Cell* **24,** 619–626 (2006).

128. Lykke-Andersen, J. & Bennett, E. J. Protecting the proteome: Eukaryotic cotranslational quality control pathways. *J. Cell Biol.* **204,** 467–476 (2014).

129. Brandman, O. & Hegde, R. S. Ribosome-associated protein quality control. *Nat. Struct. Mol. Biol.* **23,** 7–15 (2016).

130. Shao, S., Malsburg, von der, K. & Hegde, R. S. Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation. *Mol. Cell* **50,** 637–648 (2013).

131. Shao, S. & Hegde, R. S. Reconstitution of a minimal ribosome-associated ubiquitination pathway with purified factors. *Mol. Cell* **55,** 880–890 (2014).

132. Brandman, O. *et al.* A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* **151,** 1042–1054 (2012).

133. Shao, S., Brown, A., Santhanam, B. & Hegde, R. S. Structure and assembly pathway of the ribosome quality control complex. *Mol. Cell* **57,** 433–444 (2015).

134. Shen, P. S. *et al.* Protein synthesis. Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. *Science* **347,** 75–78 (2015).

135. Verma, R., Oania, R. S., Kolawa, N. J. & Deshaies, R. J. Cdc48/p97 promotes degradation of aberrant nascent polypeptides bound to the ribosome. *Elife* **2,** e00308 (2013).

136. Defenouillère, Q. *et al.* Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 5046–5051 (2013).

137. Buchan, J. R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* **36,** 932–941 (2009).

138. Buchan, J. R., Kolaitis, R.-M., Taylor, J. P. & Parker, R. Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function. *Cell* **153,** 1461–1474 (2013).

139. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149,** 753–767 (2012).

140. Li, J.-M., Wu, H., Zhang, W., Blackburn, M. R. & Jin, J. The p97-UFD1L-NPL4 protein complex mediates cytokine-induced IκBα proteolysis. *Mol. Cell. Biol.* **34,** 335–347 (2014).

141. Nguyen, T. V. *et al.* p97/VCP promotes degradation of CRBN substrate glutamine synthetase and neosubstrates. *Proc. Natl. Acad. Sci. U.S.A.* **114,** 3565–3571 (2017).

142. Fu, X., Ng, C., Feng, D. & Liang, C. Cdc48p is required for the cell cycle commitment point at Start via degradation of the G1-CDK inhibitor Far1p. *J. Cell Biol.* **163,** 21–26 (2003).

143. Verma, R., McDonald, H., Yates, J. R. & Deshaies, R. J. Selective degradation of ubiquitinated Sic1 by purified 26S proteasome yields active S phase cyclin-Cdk. *Mol.*

*Cell* **8,** 439–448 (2001).

144. Hauler, F., Mallery, D. L., McEwan, W. A., Bidgood, S. R. & James, L. C. AAA ATPase p97/VCP is essential for TRIM21-mediated virus neutralization. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 19733–19738 (2012).

145. Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146,** 11–31 (2004).

146. Milner-White, E. J., Coggins, J. R. & Anton, I. A. Evidence for an ancestral core structure in nucleotide-binding proteins with the type A motif. *J. Mol. Biol.* **221,** 751–754 (1991).

147. Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15,** 430–434 (1990).

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

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

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

151. Story, R. M., Weber, I. T. & Steitz, T. A. The structure of the E. coli recA protein monomer and polymer. *Nature* **355,** 318–325 (1992).

152. Ogura, T., Whiteheart, S. W. & Wilkinson, A. J. Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J. Struct. Biol.* **146,** 106–112 (2004).

153. DeLaBarre, B., Christianson, J. C., Kopito, R. R. & Brunger, A. T. Central pore residues mediate the p97/VCP activity required for ERAD. *Mol. Cell* **22,** 451–462 (2006).

154. Song, H. K. *et al.* Mutational studies on HslU and its docking mode with HslV. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 14103–14108 (2000).

155. Hinnerwisch, J., Fenton, W. A., Furtak, K. J., Farr, G. W. & Horwich, A. L. Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell* **121,** 1029–1041 (2005).

156. Puchades, C. *et al.* Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science* **358,** eaao0464 (2017).

157. Ripstein, Z. A., Huang, R., Augustyniak, R., Kay, L. E. & Rubinstein, J. L. Structure of

a AAA+ unfoldase in the process of unfolding substrate. *Elife* **6,** 43 (2017).

158. Kienle, N., Kloepper, T. H. & Fasshauer, D. Shedding light on the expansion and diversification of the Cdc48 protein family during the rise of the eukaryotic cell. *BMC Evol. Biol.* **16,** 215 (2016).

159. May, A. P., Misura, K. M., Whiteheart, S. W. & Weis, W. I. Crystal structure of the amino-terminal domain of N-ethylmaleimide-sensitive fusion protein. *Nat. Cell Biol.* **1,** 175–182 (1999).

160. DeLaBarre, B. & Brunger, A. T. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nat. Struct. Biol.* **10,** 856–863 (2003).

161. DeLaBarre, B. & Brunger, A. T. Nucleotide dependent motion and mechanism of action of p97/VCP. *J. Mol. Biol.* **347,** 437–452 (2005).

162. Huyton, T. *et al.* The crystal structure of murine p97/VCP at 3.6A. *J. Struct. Biol.* **144,** 337–348 (2003).

163. Davies, J. M., Brunger, A. T. & Weis, W. I. Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure* **16,** 715–726 (2008).

164. Kimonis, V. E. *et al.* Clinical and molecular studies in a unique family with autosomal dominant limb-girdle muscular dystrophy and Paget disease of bone. *Genet. Med.* **2,** 232–241 (2000).

165. Mehta, S. G. *et al.* Genotype-phenotype studies of VCP-associated inclusion body myopathy with Paget disease of bone and/or frontotemporal dementia. *Clin. Genet.* **83,** 422–431 (2013).

166. Tang, W. K. *et al.* A novel ATP-dependent conformation in p97 N-D1 fragment revealed by crystal structures of disease-related mutants. *EMBO J.* **29,** 2217–2229 (2010).

167. Cordova, J. C. *et al.* Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158,** 647–658 (2014).

168. Kim, S. J. *et al.* Structural basis for ovarian tumor domain-containing protein 1 (OTU1) binding to p97/valosin-containing protein (VCP). *J. Biol. Chem.* **289,** 12264–12274 (2014).

169. Isaacson, R. L. *et al.* Detailed structural insights into the p97-Npl4-Ufd1 interface. *J. Biol. Chem.* **282,** 21361–21369 (2007).

170. Kloppsteck, P., Ewens, C. A., Förster, A., Zhang, X. & Freemont, P. S. Regulation of p97 in the ubiquitin-proteasome system by the UBX protein-family. *Biochim. Biophys.*

*Acta* **1823,** 125–129 (2012).

171. Le, L. T. M. *et al.* Structural Details of Ufd1 Binding to p97 and Their Functional Implications in ER-Associated Degradation. *PLoS ONE* **11,** e0163394 (2016).

172. Hänzelmann, P. & Schindelin, H. Characterization of an Additional Binding Surface on the p97 N-Terminal Domain Involved in Bipartite Cofactor Interactions. *Structure* **24,** 140–147 (2016).

173. Buchberger, A. Roles of Cdc48 in regulated protein degradation in yeast. *Subcell. Biochem.* **66,** 195–222 (2013).

174. Meyer, H. & Weihl, C. C. The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis. *J. Cell. Sci.* **127,** 3877–3883 (2014).

175. Xia, D., Tang, W. K. & Ye, Y. Structure and function of the AAA+ ATPase p97/Cdc48p. *Gene* **583,** 64–77 (2016).

176. Christianson, J. C. & Ye, Y. Cleaning up in the endoplasmic reticulum: ubiquitin in charge. *Nat. Struct. Mol. Biol.* **21,** 325–335 (2014).

177. Taylor, E. B. & Rutter, J. Mitochondrial quality control by the ubiquitin-proteasome system. *Biochem. Soc. Trans.* **39,** 1509–1513 (2011).

178. Ramadan, K. *et al.* Cdc48/p97 promotes reformation of the nucleus by extracting the kinase Aurora B from chromatin. *Nature* **450,** 1258–1262 (2007).

179. Franz, A., Ackermann, L. & Hoppe, T. Ring of Change: CDC48/p97 Drives Protein Dynamics at Chromatin. *Front Genet* **7,** 73 (2016).

180. Chapman, E., Fry, A. N. & Kang, M. The complexities of p97 function in health and disease. *Mol Biosyst* **7,** 700–710 (2011).

181. Kimonis, V. E., Fulchiero, E., Vesa, J. & Watts, G. VCP disease associated with myopathy, Paget disease of bone and frontotemporal dementia: review of a unique disorder. *Biochim. Biophys. Acta* **1782,** 744–748 (2008).

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

183. Sauer, R. T. & Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80,** 587–612 (2011).

184. Chou, T.-F. *et al.* Specific inhibition of p97/VCP ATPase and kinetic analysis demonstrate interaction between D1 and D2 ATPase domains. *J. Mol. Biol.* **426,** 2886–2899 (2014).

185. Baek, G. H., Kim, I. & Rao, H. The Cdc48 ATPase modulates the interaction between

two proteolytic factors Ufd2 and Rad23. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 13558–13563 (2011).

186. Nathan, J. A., Kim, H. T., Ting, L., Gygi, S. P. & Goldberg, A. L. Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes? *EMBO J.* **32,** 552–565 (2013).

187. White, S. R. & Lauring, B. AAA+ ATPases: achieving diversity of function with conserved machinery. *Traffic* **8,** 1657–1667 (2007).

188. Barthelme, D. & Sauer, R. T. Identification of the Cdc48•20S proteasome as an ancient AAA+ proteolytic machine. *Science* **337,** 843–846 (2012).

189. Barthelme, D. & Sauer, R. T. Bipartite determinants mediate an evolutionarily conserved interaction between Cdc48 and the 20S peptidase. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 3327–3332 (2013).

190. Huang, R. *et al.* Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E4190–9 (2016).

191. Zhao, M. *et al.* Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518,** 61–67 (2015).

192. Schuller, J. M., Beck, F., Lössl, P., Heck, A. J. R. & Förster, F. Nucleotide-dependent conformational changes of the AAA+ ATPase p97 revisited. *FEBS Lett.* **590,** 595–604 (2016).

193. Richly, H. *et al.* A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. *Cell* **120,** 73–84 (2005).

194. Ernst, R., Mueller, B., Ploegh, H. L. & Schlieker, C. The otubain YOD1 is a deubiquitinating enzyme that associates with p97 to facilitate protein dislocation from the ER. *Mol. Cell* **36,** 28–38 (2009).

195. Chau, V. *et al.* A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243,** 1576–1583 (1989).

196. Bartel, B., Wünning, I. & Varshavsky, A. The recognition component of the N-end rule pathway. *EMBO J.* **9,** 3179–3189 (1990).

197. Rodrigo-Brenni, M. C., Foster, S. A. & Morgan, D. O. Catalysis of lysine 48-specific ubiquitin chain assembly by residues in E2 and ubiquitin. *Mol. Cell* **39,** 548–559 (2010).

198. Zhang, M. *et al.* Rational design of true monomeric and bright photoactivatable fluorescent proteins. *Nat. Methods* **9,** 727–729 (2012).

199. Glynn, S. E., Nager, A. R., Baker, T. A. & Sauer, R. T. Dynamic and static components power unfolding in topologically closed rings of a AAA+ proteolytic machine. *Nat.*

*Struct. Mol. Biol.* **19,** 616–622 (2012).

200. Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science* **301,** 964–967 (2003).

201. Dormán, G. & Prestwich, G. D. Benzophenone photophores in biochemistry. *Biochemistry* **33,** 5661–5673 (1994).

202. Ito, K. & Akiyama, Y. Cellular functions, mechanism of action, and regulation of FtsH protease. *Annu. Rev. Microbiol.* **59,** 211–231 (2005).

203. Suno, R. *et al.* Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Mol. Cell* **22,** 575–585 (2006).

204. Sivaraman, T., Arrington, C. B. & Robertson, A. D. Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin. *Nat. Struct. Biol.* **8,** 331–333 (2001).

205. Shabek, N. & Ciechanover, A. Degradation of ubiquitin: the fate of the cellular reaper. *Cell Cycle* **9,** 523–530 (2010).

206. Burton, R. E., Siddiqui, S. M., Kim, Y. I., Baker, T. A. & Sauer, R. T. Effects of protein stability and structure on substrate processing by the ClpXP unfolding and degradation machine. *EMBO J.* **20,** 3092–3100 (2001).

207. Ryu, J.-K. *et al.* Spring-loaded unraveling of a single SNARE complex by NSF in one round of ATP turnover. *Science* **347,** 1485–1489 (2015).

208. Kish-Trier, E. & Hill, C. P. Structural biology of the proteasome. *Annu Rev Biophys* **42,** 29–49 (2013).

209. Kim, I., Mi, K. & Rao, H. Multiple interactions of rad23 suggest a mechanism for ubiquitylated substrate delivery important in proteolysis. *Mol. Biol. Cell* **15,** 3357–3365 (2004).

210. Du, F., Navarro-Garcia, F., Xia, Z., Tasaki, T. & Varshavsky, A. Pairs of dipeptides synergistically activate the binding of substrate by ubiquitin ligase through dissociation of its autoinhibitory domain. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 14110–14115 (2002).

211. Frey, S. & Görlich, D. A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *J Chromatogr A* **1337,** 95–105 (2014).

212. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 11020–11024 (2002).

213. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42,** W252–8 (2014).

214. Stach, L. & Freemont, P. S. The AAA+ ATPase p97, a cellular multitool. *Biochem. J.*

**474,** 2953–2976 (2017).

215. Skrott, Z. *et al.* Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4. *Nature* **552,** 194–199 (2017).

216. Lander, G. C. *et al.* Complete subunit architecture of the proteasome regulatory particle. *Nature* **482,** 186–191 (2012).

217. McCullough, J., Clague, M. J. & Urbé, S. AMSH is an endosome-associated ubiquitin isopeptidase. *J. Cell Biol.* **166,** 487–492 (2004).

218. Cope, G. A. *et al.* Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of Nedd8 from Cul1. *Science* **298,** 608–611 (2002).

219. Maytal-Kivity, V., Reis, N., Hofmann, K. & Glickman, M. H. MPN+, a putative catalytic motif found in a subset of MPN domain proteins from eukaryotes and prokaryotes, is critical for Rpn11 function. *BMC Biochem.* **3,** 28 (2002).

220. Park, S., Isaacson, R., Kim, H. T., Silver, P. A. & Wagner, G. Ufd1 exhibits the AAA-ATPase fold with two distinct ubiquitin interaction sites. *Structure* **13,** 995–1005 (2005).

221. Hetzer, M. *et al.* Distinct AAA-ATPase p97 complexes function in discrete steps of nuclear assembly. *Nat. Cell Biol.* **3,** 1086–1091 (2001).

222. Pye, V. E. *et al.* Structural insights into the p97-Ufd1-Npl4 complex. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 467–472 (2007).

223. Bebeacua, C. *et al.* Distinct conformations of the protein complex p97-Ufd1-Npl4 revealed by electron cryomicroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 1098–1103 (2012).

224. Bodnar, N. O. & Rapoport, T. A. Molecular Mechanism of Substrate Processing by the Cdc48 ATPase Complex. *Cell* **169,** 722–735.e9 (2017).

225. Blok, N. B. *et al.* Unique double-ring structure of the peroxisomal Pex1/Pex6 ATPase complex revealed by cryo-electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **112,** E4017–25 (2015).

226. Ambroggio, X. I., Rees, D. C. & Deshaies, R. J. JAMM: a metalloprotease-like zinc site in the proteasome and signalosome. *PLoS Biol.* **2,** E2 (2004).

227. Lingaraju, G. M. *et al.* Crystal structure of the human COP9 signalosome. *Nature* **512,** 161–165 (2014).

228. Worden, E. J., Dong, K. C. & Martin, A. An AAA Motor-Driven Mechanical Switch in Rpn11 Controls Deubiquitination at the 26S Proteasome. *Mol. Cell* **67,** 799–811.e8 (2017).

229. Sato, Y. *et al.* Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains. *Nature* **455,** 358–362 (2008).

230. Davies, C. W., Paul, L. N., Kim, M.-I. & Das, C. Structural and thermodynamic comparison of the catalytic domain of AMSH and AMSH-LP: nearly identical fold but different stability. *J. Mol. Biol.* **413,** 416–429 (2011).

231. Worden, E. J., Padovani, C. & Martin, A. Structure of the Rpn11-Rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. *Nat. Struct. Mol. Biol.* **21,** 220–227 (2014).

232. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online* **6,** 23–34 (2004).

233. Li, X., Zheng, S., Agard, D. A. & Cheng, Y. Asynchronous data acquisition and on-the-fly analysis of dose fractionated cryoEM images by UCSFImage. *J. Struct. Biol.* **192,** 174–178 (2015).

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

235. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).

236. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192,** 216–221 (2015).

237. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).

238. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).

239. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20,** 237–247 (2012).

240. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157,** 47–55 (2007).

241. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14,** 331–332 (2017).

242. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5,** 19 (2016).

243. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms

for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14,** 290–296 (2017).

244. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18,** 2714–2723 (1997).

245. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

246. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).

247. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

248. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).

249. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69,** 1204–1214 (2013).

250. Jackson, R. N., McCoy, A. J., Terwilliger, T. C., Read, R. J. & Wiedenheft, B. X-ray structure determination using low-resolution electron microscopy maps for molecular replacement. *Nat Protoc* **10,** 1275–1284 (2015).

251. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (2007).

252. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56,** 965–972 (2000).

253. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 (2011).

254. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64,** 61–69 (2008).

255. Touw, W. G., van Beusekom, B., Evers, J. M. G., Vriend, G. & Joosten, R. P. Validation and correction of Zn-CysxHisy complexes. *Acta Crystallogr D Struct Biol* **72,** 1110–1118 (2016).

256. Morin, A. *et al.* Collaboration gets the most out of software. *Elife* **2,** e01456 (2013).

257. Zubarev, R. A. & Makarov, A. Orbitrap mass spectrometry. *Anal. Chem.* **85,** 5288–5296 (2013).

258. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat.*

*Methods* **9,** 904–906 (2012).

259. Fan, S.-B. *et al.* Using pLink to Analyze Cross-Linked Peptides. *Curr Protoc Bioinformatics* **49,** 8.21.1–19 (2015).

260. DeHoratius, C. & Silver, P. A. Nuclear transport defects and nuclear envelope alterations are associated with mutation of the Saccharomyces cerevisiae NPL4 gene. *Mol. Biol. Cell* **7,** 1835–1855 (1996).

261. Suzuki, T., Park, H., Hollingsworth, N. M., Sternglanz, R. & Lennarz, W. J. PNG1, a yeast gene encoding a highly conserved peptide:N-glycanase. *J. Cell Biol.* **149,** 1039–1052 (2000).

262. Li, G., Zhao, G., Zhou, X., Schindelin, H. & Lennarz, W. J. The AAA ATPase p97 links peptide N-glycanase to the endoplasmic reticulum-associated E3 ligase autocrine motility factor receptor. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 8348–8353 (2006).

263. Schuberth, C. & Buchberger, A. Membrane-bound Ubx2 recruits Cdc48 to ubiquitin ligases and their substrates to ensure efficient ER-associated protein degradation. *Nat. Cell Biol.* **7,** 999–1006 (2005).

264. Tang, W. K., Zhang, T., Ye, Y. & Xia, D. Structural basis for nucleotide-modulated p97 association with the ER membrane. *Cell Discov* **3,** 17045 (2017).

265. Neal, S. *et al.* The Dfm1 Derlin Is Required for ERAD Retrotranslocation of Integral Membrane Proteins. *Mol. Cell* **69,** 306–320.e4 (2018).

266. Blythe, E. E., Olson, K. C., Chau, V. & Deshaies, R. J. Ubiquitin- and ATP-dependent unfoldase activity of P97/VCP•NPLOC4•UFD1L is enhanced by a mutation that causes multisystem proteinopathy. *Proc. Natl. Acad. Sci. U.S.A.* **114,** E4380–E4388 (2017).

267. Gerega, A. *et al.* VAT, the thermoplasma homolog of mammalian p97/VCP, is an N domain-regulated protein unfoldase. *J. Biol. Chem.* **280,** 42856–42862 (2005).

268. Erlandson, K. J. *et al.* A role for the two-helix finger of the SecA ATPase in protein translocation. *Nature* **455,** 984–987 (2008).

269. Esaki, M., Islam, M. T., Tani, N. & Ogura, T. Deviation of the typical AAA substrate-threading pore prevents fatal protein degradation in yeast Cdc48. *Sci Rep* **7,** 5475 (2017).

270. Forouzan, D. *et al.* The archaeal proteasome is regulated by a network of AAA ATPases. *J. Biol. Chem.* **287,** 39254–39262 (2012).

271. Asano, S. *et al.* Proteasomes. A molecular census of 26S proteasomes in intact neurons. *Science* **347,** 439–442 (2015).

272. Ogiso, Y. *et al.* Lub1 participates in ubiquitin homeostasis and stress response via maintenance of cellular ubiquitin contents in fission yeast. *Mol. Cell. Biol.* **24,** 2324–

2331 (2004).

273. Richardson, P. G. *et al.* Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. *N. Engl. J. Med.* **352,** 2487–2498 (2005).

274. Kouroukis, T. C. *et al.* Bortezomib in multiple myeloma: systematic review and clinical considerations. *Curr Oncol* **21,** e573–603 (2014).

275. Obeng, E. A. *et al.* Proteasome inhibitors induce a terminal unfolded protein response in multiple myeloma cells. *Blood* **107,** 4907–4916 (2006).

276. Anderson, D. J. *et al.* Targeting the AAA ATPase p97 as an Approach to Treat Cancer through Disruption of Protein Homeostasis. *Cancer Cell* **28,** 653–665 (2015).

277. Bastola, P. *et al.* Specific mutations in the D1-D2 linker region of VCP/p97 enhance ATPase activity and confer resistance to VCP inhibitors. *Cell Death Discov* **3,** 17065 (2017).

278. Weihl, C. C., Pestronk, A. & Kimonis, V. E. Valosin-containing protein disease: inclusion body myopathy with Paget's disease of the bone and fronto-temporal dementia. *Neuromuscul. Disord.* **19,** 308–315 (2009).

279. Forman, M. S. *et al.* Novel ubiquitin neuropathology in frontotemporal dementia with valosin-containing protein gene mutations. *J. Neuropathol. Exp. Neurol.* **65,** 571–581 (2006).

280. Zhang, T., Mishra, P., Hay, B. A., Chan, D. & Guo, M. Valosin-containing protein (VCP/p97) inhibitors relieve Mitofusin-dependent mitochondrial defects due to VCP disease mutants. *Elife* **6,** 2231.e1 (2017).

281. Leitman, J., Ulrich Hartl, F. & Lederkremer, G. Z. Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress. *Nat Commun* **4,** 2753 (2013).

282. Mizuno, Y., Hori, S., Kakizuka, A. & Okamoto, K. Vacuole-creating protein in neurodegenerative diseases in humans. *Neurosci. Lett.* **343,** 77–80 (2003).

283. Hirabayashi, M. *et al.* VCP/p97 in abnormal protein aggregates, cytoplasmic vacuoles, and cell death, phenotypes relevant to neurodegeneration. *Cell Death Differ.* **8,** 977–984 (2001).

284. Fujita, K. *et al.* A functional deficiency of TERA/VCP/p97 contributes to impaired DNA repair in multiple polyglutamine diseases. *Nat Commun* **4,** 1816 (2013).

285. Duennwald, M. L. & Lindquist, S. Impaired ERAD and ER stress are early and specific events in polyglutamine toxicity. *Genes Dev.* **22,** 3308–3319 (2008).

286. Scheper, W. & Hoozemans, J. J. M. The unfolded protein response in neurodegenerative

diseases: a neuropathological perspective. *Acta Neuropathol.* **130,** 315–331 (2015).

287. Su, V. & Lau, A. F. Ubiquitin-like and ubiquitin-associated domain proteins: significance in proteasomal degradation. *Cell. Mol. Life Sci.* **66,** 2819–2833 (2009).

288. Yu, H., Kago, G., Yellman, C. M. & Matouschek, A. Ubiquitin-like domains can target to the proteasome but proteolysis requires a disordered region. *EMBO J.* **35,** 1522–1536 (2016).

289. Díaz-Martínez, L. A., Kang, Y., Walters, K. J. & Clarke, D. J. Yeast UBL-UBA proteins have partially redundant functions in cell cycle control. *Cell Div* **1,** 28 (2006).

290. Itakura, E. *et al.* Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. *Mol. Cell* **63,** 21–33 (2016).

291. Sivá, M. *et al.* Human DNA-Damage-Inducible 2 Protein Is Structurally and Functionally Distinct from Its Yeast Ortholog. *Sci Rep* **6,** 30443 (2016).

292. Kaplun, L. *et al.* The DNA damage-inducible UbL-UbA protein Ddi1 participates in Mec1-mediated degradation of Ho endonuclease. *Mol. Cell. Biol.* **25,** 5355–5362 (2005).

293. Ivantsiv, Y., Kaplun, L., Tzirkin-Goldin, R., Shabek, N. & Raveh, D. Unique role for the UbL-UbA protein Ddi1 in turnover of SCFUfo1 complexes. *Mol. Cell. Biol.* **26,** 1579–1588 (2006).

294. White, R. E., Dickinson, J. R., Semple, C. A. M., Powell, D. J. & Berry, C. The retroviral proteinase active site and the N-terminus of Ddi1 are required for repression of protein secretion. *FEBS Lett.* **585,** 139–142 (2011).

295. Lustgarten, V. & Gerst, J. E. Yeast VSM1 encodes a v-SNARE binding protein that may act as a negative regulator of constitutive exocytosis. *Mol. Cell. Biol.* **19,** 4480–4494 (1999).

296. Gabriely, G., Kama, R., Gelin-Licht, R. & Gerst, J. E. Different domains of the UBL-UBA ubiquitin receptor, Ddi1/Vsm1, are involved in its multiple cellular roles. *Mol. Biol. Cell* **19,** 3625–3637 (2008).

297. Marash, M. & Gerst, J. E. Phosphorylation of the autoinhibitory domain of the Sso t-SNAREs promotes binding of the Vsm1 SNARE regulator in yeast. *Mol. Biol. Cell* **14,** 3114–3125 (2003).

298. Radhakrishnan, S. K. *et al.* Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. *Mol. Cell* **38,** 17–28 (2010).

299. Radhakrishnan, S. K., Besten, den, W. & Deshaies, R. J. p97-dependent retrotranslocation and proteolytic processing govern formation of active Nrf1 upon proteasome inhibition. *Elife* **3,** e01856 (2014).

300. Lehrbach, N. J. & Ruvkun, G. Proteasome dysfunction triggers activation of SKN-1A/Nrf1 by the aspartic protease DDI-1. *Elife* **5,** e08153 (2016).

301. Sha, Z. & Goldberg, A. L. Proteasome-mediated processing of Nrf1 is essential for coordinate induction of all proteasome subunits and p97. *Curr. Biol.* **24,** 1573–1583 (2014).

302. Tomlin, F. M. *et al.* Inhibition of NGLY1 Inactivates the Transcription Factor Nrf1 and Potentiates Proteasome Inhibitor Cytotoxicity. *ACS Cent Sci* **3,** 1143–1155 (2017).

303. Koizumi, S. *et al.* The aspartyl protease DDI2 activates Nrf1 to compensate for proteasome dysfunction. *Elife* **5,** 16929 (2016).

304. Trempe, J.-F. *et al.* Structural studies of the yeast DNA damage-inducible protein Ddi1 reveal domain architecture of this eukaryotic protein family. *Sci Rep* **6,** 33671 (2016).

305. Nowicka, U. *et al.* DNA-damage-inducible 1 protein (Ddi1) contains an uncharacteristic ubiquitin-like domain that binds ubiquitin. *Structure* **23,** 542–557 (2015).

306. Dimova, N. V. *et al.* APC/C-mediated multiple monoubiquitylation provides an alternative degradation signal for cyclin B1. *Nat. Cell Biol.* **14,** 168–176 (2012).