The p97 ATPase and the *Drosophila* Proteasome: Protein Unfolding and Regulation

Kristian Björk Grimberg
To my supportive family
Abstract

For all living systems, there is a requirement to recycle and regulate proteins. In eukaryotic organisms, this is accomplished by the proteasome, a large multi-protein complex that recognizes and carries out proteolysis on proteins tagged with ubiquitin chains. The p97 ATPase is another highly conserved and essential complex present throughout the eukaryotic cell. Past investigations have suggested an important link between the p97 ATPase and efficient proteasomal degradation. One of the major aims of the work presented in this thesis has been to understand the specific functions that the p97 ATPase and its cofactors play in proteasome degradation of soluble substrates.

In Paper I we have utilized ubiquitin-fusion degradation fluorescent substrates and RNA interference to address the role of p97 ATPase cofactor proteins in soluble ubiquitin proteasome degradation. I have utilized ATP hydrolysis and binding defective Drosophila p97 mutants, as well as mutants defective in binding substrates, to discern the cytoplasmic function of p97 in proteasomal degradation, a potential unfoldase for the proteasome. Both p97 ATPase domains were found to be required for proteasome degradation to occur on unmodified ubiquitin-fluorescent protein substrates. We found that soluble proteasome substrates with C-terminal extended peptide regions 15 residues or longer allowed apparent efficient degradation to take place in the absence of the p97 ATPase. Through RNAi depletion experiments, the non-UBX cofactors Ufd1/Npl4 were also shown to be required for soluble degradation while cofactors of p97 containing the common UBX domain did not have an observed impact on degradation of the model substrates. These results support the concept that p97 functions upstream of the proteasome as an important unfoldase chaperone together with its Ufd1/Npl4 cofactors on cytosolic proteasome targets. Our results implicate p97 as an important factor for degradation of proteasome substrates lacking natural extended peptide regions.

In Paper II we focused on identifying transcription factors essential for production of proteasomal subunits and associated proteins in Drosophila S2 cells. We utilized an RNA library targeting 993 known or candidate transcription factors and monitored RNAi depleted Drosophila S2 cells expressing the ubiquitin fusion domain reporter UbG76GFP. We identified a range of potential candidates for regulating proteasome transcription and focused on the bZIP transcription factor Cnc-C. Disruption of Cnc-C by RNA interference leads to decreased levels of overall proteasomal activity as well as the specific decrease in mRNA and protein levels for proteasomal
subunits indicating that Cnc-C is involved in transcription of proteasomal subunits.

In **Paper III** we applied our knowledge gained from the p97 project, **Paper I**, about p97 dependent substrates and set up a high-throughput microscopy screening method to potentially find inhibitors specifically targeting the p97 ATPase proteasomal sub-pathway. Utilizing UFD substrates with and without C-terminal peptide tails we determined if compounds inhibited the core proteasomal machinery or if the compounds targeted the p97 pathway specifically. We identified several new small molecule compounds with inhibitory activity against the ubiquitin-proteasome pathway though none from our initial screening had specificity for the p97 ATPase. A range of analogs of the most promising hit were screened in a second round and the pattern of inhibition to structure allowed us to propose a potential mechanism for inhibitory activity that can be further tested and developed in the future.
List of Papers

This thesis is based on the following papers which are referred to in the text by their roman numerals:


II. Grimberg KB, Beskow A, Lundin D, Davies M, Young P. Basic-leucine zipper protein is a substrate and transcriptional regulator of the Drosophila 26S proteasome Revised manuscript re-submitted to Molecular and Cellular Biology

III. Grimberg KB, Beskow A, Lundbäck T, Jenmalm Jensen A, Young, P Dantuma N, Salomons F. A high-throughput microscopy method to find novel p97 inhibitors Manuscript
Table of Contents

Introduction .......................................................................................................................... 13
Preface ................................................................................................................................. 13
The proteasome .................................................................................................................. 15
  20S core particle .............................................................................................................. 15
  19S regulatory particle ................................................................................................. 17
Ubiquitin ........................................................................................................................... 19
  Ubiquitylation .............................................................................................................. 20
  Degradation signals .................................................................................................. 22
Other cellular signals – ubiquitin like proteins .............................................................. 22
p97 .................................................................................................................................... 24
  Structure/domains ..................................................................................................... 25
  Cofactors and domains ............................................................................................... 26
    The UBX and UBX-like domain .............................................................................. 26
    The PUB domain .................................................................................................... 28
    The SHP box ........................................................................................................... 29
    Other p97 binding motifs and domains .................................................................. 30
Functions for the p97 ATPase and its cofactors ............................................................ 30
  The role of p97 in transcriptional control and activation ........................................... 30
  The role of p97 in membrane fusion ......................................................................... 31
  The role of p97 in cell cycle control ......................................................................... 31
  The role of p97 in protein degradation ..................................................................... 32
The p97 in human disease ............................................................................................... 32
Transcriptional regulation of proteasome subunits ......................................................... 33
  bZIP proteins ............................................................................................................ 34
    The CNC-family ...................................................................................................... 35
    ARE motifs .............................................................................................................. 38

Methods ............................................................................................................................. 39
  The model system Drosophila melanogaster ................................................................ 39
  RNA interference ...................................................................................................... 39
  Expression of exogenous proteins .......................................................................... 40
  Living cell UPS degradation assay .......................................................................... 40

Summary of publications ................................................................................................. 42
  Paper I ......................................................................................................................... 42
    A conserved unfoldase activity for the p97 AAA-ATPase in proteasomal degradation
  Paper II ....................................................................................................................... 44
    Basic-leucine zipper protein Cnc-C is a substrate and transcriptional regulator of the
      Drosophila 26S proteasome
  Paper III ....................................................................................................................... 48
    A high-throughput microscopy method to find novel p97 inhibitors

Concluding remarks ......................................................................................................... 51

Acknowledgments ............................................................................................................ 52

References ......................................................................................................................... 53

IX
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>ATPases associated with various cellular activities</td>
</tr>
<tr>
<td>AMP</td>
<td>adenosine monophosphate</td>
</tr>
<tr>
<td>APC</td>
<td>anaphase promoting complex</td>
</tr>
<tr>
<td>ARE</td>
<td>antioxidant-responsive element</td>
</tr>
<tr>
<td>Arg</td>
<td>arginine</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bZIP</td>
<td>basic-leucine zipper</td>
</tr>
<tr>
<td>Cnc</td>
<td>cap'n'collar</td>
</tr>
<tr>
<td>dsRNA</td>
<td>double stranded RNA</td>
</tr>
<tr>
<td>DUB(s)</td>
<td>deubiquitinating enzyme(s)</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ERAD</td>
<td>endoplasmic reticulum associated degradation</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>HECT</td>
<td>homologous to E6-AP carboxy terminus</td>
</tr>
<tr>
<td>Leu</td>
<td>leucine</td>
</tr>
<tr>
<td>Lys</td>
<td>lysine</td>
</tr>
<tr>
<td>MARE</td>
<td>maf recognition element</td>
</tr>
<tr>
<td>Met</td>
<td>methionine</td>
</tr>
<tr>
<td>Nrf</td>
<td>nuclear factor-erythroid 2 p45 subunit-related factor</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>Phe</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>PP&lt;sub&gt;i&lt;/sub&gt;</td>
<td>inorganic pyrophosphate</td>
</tr>
<tr>
<td>RFP</td>
<td>red fluorescent protein</td>
</tr>
<tr>
<td>RING</td>
<td>really interesting new gene</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>SCF</td>
<td>Skp1 – cullin – F-box</td>
</tr>
<tr>
<td>Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>Ub</td>
<td>Ubiquitin</td>
</tr>
<tr>
<td>UBX</td>
<td>ubiquitin regulatory X</td>
</tr>
<tr>
<td>UFD</td>
<td>ubiquitin fusion domain</td>
</tr>
<tr>
<td>UPR</td>
<td>unfolded protein response</td>
</tr>
<tr>
<td>UPS</td>
<td>ubiquitin proteasome system</td>
</tr>
<tr>
<td>YFP</td>
<td>yellow fluorescent protein</td>
</tr>
</tbody>
</table>
Introduction

Preface

In the early 1950s the lysosome was proposed as the main factor responsible for degradation of intracellular proteins. This was for a long time believed to be the only way the cell could carry out large scale proteolysis, but questions arose as to why some proteins seemed to have shorter or longer half lives, and why some seemed to be much less affected by lysosomal inhibitors. Explanations were devised, such as an entry/exit model where longer-life proteins entered the lysosome but were transferred out again (Haider & Segal 1972), or selectivity in lysosomal membrane binding (Dean 1977). The notion that the lysosome was the sole factor for proteolysis was an open question until everything changed in the proteolysis field when an ATP-dependent system for proteolysis in reticulocytes, immature red blood cells lacking lysosomes, was described (Etlinger & Goldberg 1977). This, and the discovery of ”the mark of death” ubiquitin, and the ubiquitylation cascade, (Ciechanover et al. 1978; Wilkinson et al. 1980; Hershko et al. 1983) started a revolution in the way scientists regarded proteolysis.

Currently it is well known that proteins are constantly synthesized and degraded in living systems and that a group of enzymes called proteases are directly involved in regulating the cell's proteome. Proteases have enzymatic activity, cleaving peptide bonds, the linkage between amino acid residues in proteins. Cleavage of peptide bonds has different objectives, such as activation, degradation or further degradation of already partially degraded peptides. Now we know that eukaryotic cells utilize both a membrane vesicle system, the lysosomal pathway, and a multisubunit complex, the Ubiquitin Proteasome System (UPS), for protein degradation.

The interior of the lysosome is acidic and compartmentalize a variety of enzymes for various degradation. These are engaged in dephosphorylation of mononucleotides, phospholipids and degradation of RNA, DNA and peptides. Degradation by the lysosome is accomplished via mechanisms called micro- and macro-autophagy (degradation of endogenous proteins and organelles), pino-, phago- and (receptor mediated) endocytosis (degradation of exogenous proteins, exogenous particles and proteins bound to certain receptors - heterophagy). Portions of either the cytoplasm or the endoplasmic reticulum (ER) enclose particles, subsequently merging with a primary
lysosome in order to degrade the different factors, reviewed in (Klionsky et al. 2007; Mizushima et al. 2008).

Proteasomes on the other hand are mobile complexes found both in the nucleus and the cytoplasm. These complexes have a large barrel-like multi-protein structure (20S) with active cleavage sites inside its catalytic core. The top and bottom of the barrel formed by the core subunits is usually associated with regulatory complexes, typically the 19S, which recognize certain motifs or proteins “marked” with polyubiquitin chains. Ubiquitin is a small highly conserved peptide expressed in all eukaryotic organisms. Proteins targeted for degradation by the proteasome complex are usually marked by ubiquitin through a well explored ubiquitylation cascade. First the E1 activating enzyme activates ubiquitin by adenylation. Adenylated ubiquitin can then be transferred onto the E2 conjugating enzyme, which binds an E3 ligating enzyme. E3 enzymes associate with substrates to which ubiquitin can be transferred, either directly from the E2 or through the E3 onto the substrate (Wilkinson et al. 1980; Haas et al. 1982; Hershko et al. 1983). The first ubiquitin is built upon to form polyubiquitin chains recognized by the ubiquitin interacting subunits Rpn10 and Rpn13 in the 19S regulatory complex (Deveraux et al. 1994; Elsasser et al. 2004; Husnjak et al. 2008; Schreiner et al. 2008). Substrates are transferred via the 19S into the core of the 20S complex where they are processed into short peptides by three different catalytic activities cleaving peptides after acidic, basic and hydrophobic amino acids (Dick et al. 1998) while ubiquitin chains are cleaved and recycled. In Paper I we show that there is a requirement for an additional complex, the p97 AAA ATPase, to target certain substrates for degradation.

Extensive work has been invested in understanding how the ubiquitin proteasome system is regulated. It is known that polyubiquitin chain initiation, expansion and possible reduction are highly regulated events. Regulation of proteasome subunit expression has on the other hand only recently been further clarified. Earlier studies discovered a proteasomal transcriptional regulator in the yeast Saccharomyces cerevisiae and closely related species (Mannahaupt et al. 1999; Xie & Varshavsky 2001). It has recently been proposed that transcriptional regulation of proteasome subunits in humans are regulated by Nrf1 (Radhakrishnan et al. 2010). This result in the human system was supported by the shared communication between our groups of our results indicating that the Drosophila melanogaster homologues of Nrf proteins, Cnc-C, is the flies transcription factor for the proteasome (Paper II).
The proteasome

20S core particle

Eukaryotic 20S core particles consist of four stacked rings, two inner β-rings and two outer α-rings. This means that two β-rings with seven subunits each are sandwiched between two α-rings, also with seven subunits each, creating a closed chamber, the catalytic core (Figure 2, Löwe et al. 1995; Groll et al. 1997). Directed into the chamber are active groups containing peptide cleaving activities. In its free form, 20S is generally believed to be inactive. However, it has been proposed that the 20S have a ubiquitin independent proteolytic role in degradation of oxidized proteins (reviewed by Davies 2001; Shringarpure et al. 2001). Proteasome activity assays typically use low levels of SDS to open up and “activate” the capacity for cleavage of peptides by 20S.

Assembly of the 20S core particle is done by POMP/Ump1 (in human and yeast, respectively; reviewed by Kruger et al. 2001). The α- and β-rings are assembled independently of each other whereupon one α-ring and one β-ring are joined together as a half 20S. Two of these are then assembled by POMP/Ump1 into a complete 20S trapping its creator inside the catalytic core, subsequently degrading it (Kruger et al. 2001). The active sites responsible for peptide degradation exist as inactive precursors with N-terminal peptides. These N-terminal peptides are cleaved off upon 20S assembly, activating the peptide cleaving activities of the subunits β1, β2 and β3. β1 has a PGPH activity (post glutamyl peptide hydrolysing activity) cleaving after acidic amino acids, β2 has trypsin like activity cleaving after basic amino acids and β5 has chymotrypsin like activity cleaving after hydrophobic amino acids (Dick et al. 1998). Assembly of the complete 26S proteasome involves a second class of chaperones that have recently been discovered (Roelofs et al. 2009). The chaperones associate with Rpt subunits of the 19S complex and are released as the 19S attaches to the 20S.

Peptides are normally not able to enter into the catalytic core of the 20S because of the gate that is formed by the α-subunits through their N-termini, the α3 subunit in particular. The subunits block the central pore (Groll et al. 1997) while maintaining contact with every other subunit in the heptameric α-ring (Groll et al. 2000). In order to open the gate to the pore a regulatory particle or complex is needed, like the 19S, 11S (PA28) or PA200. Biochemical studies done using the PAN complex, an Archaea homolog of
Figure 1. The 20S proteasome. Schematic representation of the closed 20S core particle, consisting of two β-rings, seven subunits each, sandwiched between two α-rings, also seven subunits each (top left). Upon 20S activation by a 19S regulatory particle, proteins are unfolded and threaded through the catalytic core, where peptide fragments with an average size of 9 residues are generated (top right). Subunit distribution in the eukaryotic 20S proteasome, white circles representing the catalytically active β subunits (bottom).

the 19S (Zwickl et al. 1999), show that association to either end of the 20S by regulatory particles may function as a key in a lock, loosening the constraints of the α-subunits in order to open the gate into the 20S core (Smith et al. 2007). Opening the gate into the pore is not sufficient for the majority of substrates to be degraded though, since their stable folded structures give a steric hindrance causing the substrates to be too big for
passing through the gate. Wenzel and Baumeister (Wenzel & Baumeister 1995) have demonstrated this by doing model experiments with \(\alpha\)-lactalbumin confirming the need for unfolding of proteins for proteasomal degradation. This demonstrates that unfolding of proteasome substrates is essential for overall proteasome activity. Previously it has been assumed that the unfolding activity is carried out by the ATPases present directly above the gate of the 20S proteasome.

The cleaved peptides from proteins degraded by the proteasome range in size from 3 to 23 amino acids, but are generally in the range of 7 to 9 amino acid residues (Kisselev et al. 1998; Kisselev et al. 1999). After processing by the proteasome, peptides are further degraded to single amino acid residues by a number of proteases including TPPII (tripeptidyl peptidase II), a protease that can counterbalance low proteasome levels (Glas et al. 1998).

In order to suppress and inhibit the catalytic peptide cleaving activity possessed by the 20S, proteasome inhibitors may be utilized. These inhibitors are being used for both research purposes and medical therapy. These compounds binds to one or several of the three active sites in the 20S, effectively blocking any cleavage of peptides. Most widely used in the laboratory are epoxomicin and MG132 (Z-Leu-Leu-Leu-CHO). Epoxomicin inhibits all three catalytic activities irreversibly, but inhibition of the chymotrypsin like activity with MG132 can be reversed (Elofsson et al. 1999; Myung et al. 2001). So far only one proteasome inhibitor, bortezomib (Velcade) has actually passed clinical trials. Three phases of clinical trials have been passed and clinical use of bortezomib has been approved for both initial and relapsed treatment of multiple myeloma as well as treatment of relapsed mantle cell lymphoma (LeBlanc et al. 2002; Orlowski et al. 2002; Richardson et al. 2003; Adams & Kauffman 2004; Richardson et al. 2005; Alinari et al. 2009).

19S regulatory particle

Proteasomal degradation does not only require the core particle with its catalytic subunits, but also an activator (as mentioned earlier). The 19S regulatory particle, or the 19S regulatory cap, is one of these activators (Hough et al. 1986). It has informally been nicknamed the “dragon head” because of its shape observed by electron tomography imaging (Walz et al. 1998). It consists of several different subunits, whereof six of them are ATPases collected in a certain order in a circular base (Tomko et al. 2010). Several other subunits are constructed into a lid sub-complex sharing
considerable sequence homology with the COP9 signalosome, a nuclear enriched complex required for removal of ubiquitin-like NEDD8 proteins from an E3 ubiquitin ligase of the SCF family (Lyapina et al. 2001; Schwechheimer & Deng 2001; Glickman et al. 1998). Components of the 19S RP are in contrast to those of the 20S CP still being added.

The 19S regulatory particle is a multi-subunit complex known to activate the 20S proteasome when associated with it by loosening the conformational constraints present in the ring of α-subunits, thereby opening the “gate” to the 20S central catalytic core (Groll et al. 2000). The 19S crystal structure has not yet been solved, in contrast to the 20S core proteasome. However, a wide range of biochemical and genetic studies on the 19S have given insight into the composition and subunit interactions of this important complex, for example with yeast two-hybrid assays (Chen et al. 2008).

The 19S regulatory particle is generally divided into two parts: the lid and the base. The base consists of six ATPase subunits and four non-ATPase subunits, Rpn1, Rpn2, Rpn10 (S5a) and Rpn13; on top of the base is the non-ATPase lid typically described having the subunits Rpn3, Rpn5-Rpn9, Rpn11, Rpn12 and Rpn15 (Voges et al. 1999). Lid and base are connected by the subunit Rpn10 and when the lid is dissociated from the base, peptides may still be degraded by the 19S RP base sitting on top of the 20S CP, but there is no recognition of ubiquitin-tagged substrates (Glickman et al. 1998; Fu et al. 2001). This implicates that ubiquitin-interacting proteins are located in the RP lid, catching and transferring substrates via the ATPases in the RP base to the 20S CP. Ubiquitin interaction is achieved through the ubiquitin receptors Rpn10 and Rpn13. Rpn10 binds polyubiquitin chains through its two ubiquitin interacting motifs UIMs (Young et al. 1998) while Rpn13 has a pleckstrin-like receptor through which ubiquitin binding is achieved (Schreiner et al. 2008; Husnjak et al. 2008). Recently, Rpn1 and Rpn2 was discovered to physically link the sites of substrate recruitment and proteolysis (Rosenzweig et al. 2008). The 20S has been found in all three branches of life, and has high conservation between Archaea and eukaryotes (Baumeister et al. 1998). In mammals, the 26S proteasome is localized to both the cytoplasm and nucleus (Brooks et al. 2000). However, the 19S is not found in Archaea or Eubacteria. As mentioned previously in this thesis, in Archaea there is a simpler related ATPase, termed PAN, which has similarities to the 19S base and p97 (Zwickl et al. 1999; Medalia et al. 2009). PAN is a homohexamer like p97 but directly binds the 20S like the 19S base.
Ubiquitin

In 1978 a small protein was found and termed "active principle of fraction 1" (or APF-1) (Ciechanover et al. 1978). Soon, APF-1 was identified as ubiquitin (Wilkinson et al. 1980), followed by the finding of the ubiquitylation cascade (Hershko et al. 1983), an effort that in 2004 was awarded the Nobel Prize in chemistry.

Ubiquitin is a 76-residue polypeptide of 8 kDa and has distinct roles in a plethora of different mechanisms. One of the main cellular roles for ubiquitin is to work as a degradation signal in proteasomal degradation. Ubiquitin can reside as an individual gene or as a multicistronic gene, where several ubiquitins are translated after each other in a long band of up to nine ubiquitins (Ozkaynak et al. 1984; Wiborg et al. 1985) that are then processed to separate molecules by C-terminal hydrolases. Mature, monomeric ubiquitin are subsequently ready to attach covalently via its C-terminal glycine 76 to internal lysines of other substrates. The ubiquitin field was initially focused on protein degradation but today ubiquitylation is known to be a general post-translational modification. Ubiquitin is not only needed for degradation of proteins, but also for features such as transcriptional regulation, DNA replication and repair, membrane trafficking and histone modifications. This is achieved through the use of poly ubiquitylation or single or multiple mono ubiquitylation. Basically, single mono ubiquitylation is the conjugation of a single ubiquitin molecule to a protein, while multiple mono ubiquitylation is the conjugation of several single ubiquitin molecules to a protein. Poly ubiquitylation is the conjugation of first one ubiquitin molecule to a protein that is further built upon to create a chain of ubiquitin molecules, linked through one or several of the seven different lysine moieties present in ubiquitin (Figure 2). Conjugation of ubiquitin chains is hypothesized to be achieved through any of the seven internal lysines (Lys) of ubiquitin (Figure 2). Among the most studied of the internal lysines are Lys48, held as the main signal for targeted proteolysis, needing a chain of at least four ubiquitins for efficient protein degradation (Thrower et al. 2000), and Lys63, which have been reported in control of DNA repair, translation and kinase activation in the NFB signaling network. Lys29 has been shown to recruit chain-elongation factors that polymerize lys48-chains, targeting the substrates for degradation (Haglund & Dikic 2005). Not much is known about Lys6-chains, although it has been suggested that it has a role in DNA repair (Morris & Solomon 2004). Apart from the chain-linked ubiquitylation stands monoubiquitination, having an impact on subcellular localization, conformation and protein interactions. It can be seen to serve similar functions as phosphorylation events, possibly regulating binding of self-monoubiquitinated proteins to other proteins (Haglund & Dikic 2005).
Ubiquitylation

Conjugation of ubiquitin molecules to different proteins is the basis for all proteasomal degradation. This is achieved through a pathway utilizing activation, conjugation and ligation of ubiquitin, the ubiquitylation cascade. This cascade is carried out by a set of enzymes, ubiquitin-activating enzyme (E1), ubiquitin conjugating enzyme (E2), ubiquitin ligase (E3) and, possibly in some cases, chain elongating factors (E4). Through hydrolysis of ATP, ubiquitin is adenylated at its C-terminus and then transferred to the active site cysteine on E1, thereby releasing AMP and inorganic pyrophosphate (PP) (Ciechanover et al. 1981; Haas et al. 1982). E1 transfers the ubiquitin to the active site cysteine on E2, and once conjugated the E2 binds to an E3 that has bound a substrate. Ubiquitin is finally transferred from E2 to the substrate and the ubiquitin is elongated into a chain through its internal lysines (Figure 3; Hershko et al. 1983; Hershko et al. 1986). Organisms usually have only one or two E1 (one or two isoforms), about forty or fifty E2's and more than a thousand E3's. In human there are now two known E1 ubiquitin activating enzymes, UBA1 and UBA6 (Jin et al. 2007; Pelzer et al. 2007; Chiu et al. 2007; Handley et al. 1991).

The E3 ligases are generally divided into two families: HECT (homologous to E6-AP carboxy terminus) and RING (really interesting new gene) finger (Ardley & Robinson 2005). These two families differ in how the ubiquitin is transferred from the E2 to the substrate. HECT domain E3's have an active site cysteine to where ubiquitin is transferred from the E2, and then from the E3 to the substrate. RING finger E3's have a conserved zinc-binding site and function more as a connection receptor between E2 and substrate, leaving the transfer of ubiquitin to the E2. The RING finger proteins function as E3's in complex with other proteins, like the SCF (Skp1-cullin-F-box) or the APC (anaphase promoting complex). In SCF complexes the RING finger Skp1 is connected via cullin to a F-box protein recognizing specific phosphorylated substrates and APC complexes consist of at least thirteen subunits, its activity coordinated by the cell cycle (Ho et al. 2006; Nakayama & Nakayama 2006; Thornton & Toczyski 2006). Usually the E3 alone is sufficient for chain elongation, but in some cases a separate chain elongation factor (E4) might be needed (Koegl et al. 1999; Hoppe 2005). The first E4 was identified as Ufd2 in yeast, a known cofactor for the AAA ATPase p97, connecting the ATPase in one of many ways to ubiquitylation and substrate degradation (Richly et al. 2005).
**Figure 2. Ubiquitin and ubiquitylation.** Structure of the small polypeptide ubiquitin (PDB code 1G6J), similar to structures of other ubiquitin like proteins (top). Different possibilities for conjugation of ubiquitin molecules (middle). Single mono ubiquitylation (left), multiple mono ubiquitylation (middle) and poly ubiquitylation (right). Chain elongation of ubiquitin is mediated through its internal lysines (bottom). Depending on what kind of lysine chain is synthesized, different pathways are utilized, such as for example protein degradation, DNA repair or kinase activation.
Degradation signals

Ubiquitin ligases recognize specific degradation signals, also called degrons, in substrates. One of these, the N-end rule, states that the half-life of a protein depends on the N-terminal amino acid of a protein. That means, if a protein has its original Met removed, depending on the new amino acid residue, proteins have either destabilization signals (Arg, Lys, Phe, Leu, Trp), giving a half-life of just a few minutes, or stabilization signals, which most proteins have naturally (Bachmair et al. 1986; Varshavsky 1992).

Another signal for degradation is the ubiquitin fusion domain pathway (UFD). Typically, ubiquitin fused to the N-termini of a peptide is cleaved off by deubiquitinating enzymes (DUBs); mutating the C-terminal residue of ubiquitin from a glycine to valine renders the ubiquitin-peptide fusion resistant to cleavage by DUBs. Instead, the ubiquitin becomes elongated into a polyubiquitin causing a rapid turnover of the fusion peptide (Johnson et al. 1995). So far, the only UFD substrate found naturally is UBB, a ubiquitin with an additional C-termini of 19 residues (Lindsten et al. 2002).

Other examples of degradation signals are the destruction box and PEST sequences. The destruction box motif is recognized by the APC complex and present in cyclins and other cell cycle factors in a stretch of nine amino acid residues (RxxLxxIxN) followed by a lysine rich domain (Pagano 1997). PEST sequences are rich in proline, glutamic acid, serine and threonine residues and have been proven to reduce degradation of IκB when removed (Rechsteiner & Rogers 1996).

Other cellular signals – ubiquitin like proteins

An ever-growing number of polypeptides similar to ubiquitin have been found, not always sharing any high degree of sequence conservation, but strong structural conservation. In the same manner as ubiquitin, these polypeptides can attach to other proteins and function as signals for different cellular pathways, although not usually degradation. They also share enzymatic activating and conjugating machinery similar to that of ubiquitin, but typically have their own enzymes to carry out these enzymatic steps. Now, there are around 10 ubiquitin like proteins, but more will likely be discovered over time. Most studied among ubiquitin like proteins are NEDD8 and SUMO. Both molecules have important regulatory roles with ubiquitin degradation.

NEDD8 can attach to all of the cullin family members except APC2, and regulate activity of the SCF ligases by attachment to CUL1; it has also been
**Figure 3. The ubiquitylation cascade.** Ubiquitin (Ub) is bound to a cysteine on the ubiquitin-activating enzyme (E1) and ATP is hydrolyzed into AMP and inorganic pyrophosphate (PPI, step 1). E1 transfers ubiquitin to the ubiquitin conjugating enzyme (E2) and E1 is free to bind a new ubiquitin (step 2). E2 binds to a ubiquitin ligase (E3, step 3), which has attached to a specific substrate (step 4). Ubiquitin is transferred from E2 to the substrate, a ubiquitin chain is generated on the substrate (step 5), which is then recognized by specific factors in the proteasome, targeting the substrate for degradation (step 6) while ubiquitins are cleaved off by de-ubiquitinating enzymes to be used once again.
implicated as a regulator of p53, where lack of neddylation increases transcriptional activity (Xirodimas et al. 2004).

SUMO has three isoforms in humans (SUMO-1, -2, -3), but differ surprisingly in sequence similarity. SUMO-2 and -3 contain a motif at lys11 (hKxE, where h = hydrophobic residues) allowing chain formation, which SUMO-1 lacks. SUMO has been implicated in processes such as transcription, DNA repair and chromatin structure maintenance (Tatham et al. 2001; Bylebyl et al. 2003; Kim et al. 2005).

There are quite a few other types of degradation signals, many which have a ubiquitin-like domain as the responsible factor and are internal domains withing proteins (Buchberger 2002). Ubiquitin domains are not conjugated to other proteins and are not cleaved by DUBs. Two important classes of ubiquitin domains are the ubiquitin-like (UBL) and the ubiquitin regulatory X (UBX) domains. UBX domains play a critical role in the interaction between the p97 AAA ATPase, see below, and its cofactor proteins (Buchberger et al. 2001; Decottignies et al. 2004; Schuberth & Buchberger 2005).

p97

p97 is a type II protein of the AAA+ ATPase super-family that hexamerizes with itself, forming a structure with two rings on top of each other (Figure 4) (Wang et al. 2003), which have been reported to bind ubiquitin chains in vitro (Dai & Li 2001). As one of the most abundant proteins in the eukaryotic cell (Peters et al. 1990), this member of the AAA+ ATPases is involved as a molecular chaperone in a wide variety of pathways such as membrane fusion, cell cycle regulation, transcriptional activation, ERAD (endoplasmic reticulum associated degradation, a quality control system that recognize and remove misfolded proteins from the ER) and a multitude of other processes. Not surprisingly, since p97 is abundant and involved in so many processes, removal of p97 is embryonically lethal, demonstrated in p97−/− mice (Muller et al. 2007). Its presence spans a wide field of organisms and has been studied in such diverse model systems as Drosophila (Pinter et al. 1998), yeast (Moir et al. 1982; Frohlich et al. 1991), plants (Feiler et al. 1995) and protozoa (Roggy & Bangs 1999). In mammals p97 was first believed to be a precursor molecule to a biologically active peptide, valosin, isolated from porcine upper gut (Schmidt et al. 1985), but the valosin peptide was later shown to be only a purification artifact (Koller & Brownstein 1987).
Valosin-containing protein, cdc48, TER94, p97, all these names refer to the same homologous protein, a type II member of the AAA+ ATPase family involved in a plethora of pathways. For all the model systems p97 have been studied in, the protein exists as a homohexamer with two highly conserved ATPases for each monomer in the complex. These ATPases allow the p97 complex to function as a molecular unfolding motor for the cell.

Structure/domains

There are three major domains in p97 that have been found to be essential for activity and regulation: the N-terminal, D1 and D2 (Figure 4). The N-terminal is the interaction site of most cofactors and D1 and D2 are the two ATPase domains, both containing the Walker A and B motifs. The Walker motifs have highly conserved amino acid residues between species; Walker A contains a P-loop (conserved residues GxxxxGK[S/T]) where ATP is bound, and Walker B has a DExx-box motif (hnhhDExx, h = hydrophobic residues) where ATP is hydrolyzed (Snider & Houry 2008). From previous studies, it is proposed that D1 is responsible for keeping the hexamer together via conserved arginine residues (Wang et al. 2005) and D2 is the domain largely responsible for ATP-hydrolysis and enzymatic activity (DeLaBarre & Brunger 2003; Song et al. 2003). Both D1 and D2 have been reported to bind ATP to similar extent, though D1 exhibits a higher binding affinity to ADP indicating why D2 exhibit higher ATPase activity than D1 (Briggs et al. 2008). Both domains are important, because studies have shown that ATPase activity in D1 requires nucleotide binding and hydrolysis in D2, and that activity in D2 requires nucleotide binding in D1 (Ye et al. 2003).

Since p97 is similar to other AAA+ ATPases that thread substrates through an axial pore, such as Clp-proteases, (Yu & Houry 2007) one might think that the same could be expected for p97, but structural studies have shown that a Zn atom obstructs threading of substrates through the pore (DeLaBarre & Brunger 2003). Because there is no evidence for substrate threading, DeLaBarre and Brunger have instead proposed that substrates enter a large cavity or pore at the bottom of the p97 hexamer. Two central and adjacent pore residues (tryptophan and phenylalanine) that sit at the top surface of the internal pore region mediate substrate interaction (DeLaBarre et al. 2006). Surprisingly, it has proven difficult to study the expected unfolding activity of purified p97 in vitro. Unfolding is only observed after specific site directed mutagenesis. Interestingly, p97 has a very distinct motion and flexibility in different stages of ATP binding and hydrolysis,
suggestions a probable substrate-processing mechanism (DeLaBarre & Brunger 2005).

Cofactors and domains

Since the p97 ATPase has been identified to be essential for many distinct processes, how does it function in these different roles? The p97 associates with many different cofactor proteins to form unique complexes that carry out specific functions for the cell. Increasing numbers of cofactors of p97 are still being discovered and divided into different classes with regards to their different structural motifs. Cofactor proteins may compete for the same binding region on p97, while others have their own specific binding regions on the p97 ATPase hexamer. Cofactors for p97 can be divided into several groups based on domains or motifs; among these are UBX domain proteins (see above), PUB domain and PUL domain proteins, SHP box proteins and VIM (VCP interacting motif) and VBM (VCP binding motif) domain proteins.

The UBX and UBX-like domain

The protein SAKS1 was the first protein suggested to have a UBX domain. The UBX domain was discovered as a sequence of about 80 amino acid residues in an analysis of UBA (ubiquitin associated) domain proteins (Hofmann & Bucher 1996). Presence of a UBA domain suggested a connection to the UPS, while the function of the UBX itself remained unclear. Further study of the UBX domain revealed a structure highly similar to that of ubiquitin. But while ubiquitin is conjugated to other proteins, the absence of a carboxyl-terminal binding site with a di-glycine motif and properly positioned lysine residues suggested that UBX domain proteins are not conjugated to other proteins (Buchberger et al. 2001). A wide-range of cofactors for p97 have UBX or UBX-like domains, such as p47, p37, Ubx2 and VCIP135 (VCP–p47 complex-interacting protein p135) (Dreveny et al. 2004). Structure alignments of UBX domain containing proteins revealed that proteins can be grouped into at least five evolutionarily conserved subfamilies; FAF1, p47, SAKS1 UBXD1 and Rep8 proteins (Buchberger et al. 2001; Schuberth & Buchberger 2008).
Figure 4. Structure of the p97 AAA ATPase hexamer. The hexamer is viewed from above (top) and a cross section of the hexamer (a trimer) is viewed from the side (below). Different colors represent the main domains (N-terminal (N), D1 AAA ATPase (D1) and D2 AAA ATPase (D2)) as well as site specific mutants generated in the D1 and D2 AAA ATPase domains in Paper I (Walker A (WA), Walker B (WB) and pore interaction site (pore)). Illustration based on experimental data by Weis et al. (2008), PDB code 3CF3.
Best characterized so far of UBX domain containing proteins and most studied of p97 cofactors is p47, identified in p97/p47-mediated membrane fusion and named after its relative molecular mass of 47 kDa (Kondo et al. 1997). The p47 binds to the N-terminal part of p97 via its UBX domain. For each p97 hexamer, three p47 molecules bind (Pye et al. 2007) and interestingly, p47 can only bind ubiquitin when in complex with p97, and preferably mono- rather than polyubiquitin (Meyer et al. 2002). Both p47 and the Ufd1/Npl4 (Ufd stands for ubiquitin fusion domain and Npl is for nuclear protein localization) dimer compete for the same binding site on p97 (Meyer et al. 2000). This is due in part to a UBD (ubiquitin D) motif in Npl4 which have affinity to the same site on p97 as p47, and which also share similarity in structure with p47 (Meyer et al. 2000; Isaacson et al. 2007).

Among the less studied UBX domain cofactors are the Cui1 (Ubx4), Cui2 (Ubx6) and Cui3 (Ubx7) proteins (Cui for cdc48 UBX-containing interactor). In yeast, these proteins are needed for a fully functional proteolysis and sporulation of diploid cells. They interact both with p97 and the non-UBX adaptor Ufd3 that is involved in ubiquitin fusion proteolysis (Decottignies et al. 2004). In human, p97 have been demonstrated to bind all known UBX domain containing proteins, shown by p97-Myc immunoprecipitations done by Deshaies and coworkers (Alexandru et al. 2008).

**The PUB domain**

The first PUB (peptide:N-glycanase/UBA or UBX-containing proteins) domain were initially found in the N-terminal part of PNGase (peptide:N-glycanase) and database searches using the PNGase sequence yielded four more proteins containing sequences of high similarity to the query (Suzuki, Park, Till et al. 2001). Cytoplasmic PNGase is a deglycosylating enzyme and interestingly both the yeast and mouse homologue have been shown to bind to Rad23, the mouse also displaying association with the proteasome subunit Rpt2 (Suzuki, Park, Kwofie et al. 2001; Park et al. 2001). A later study revealed that PUB domain binding to p97 could be attained with association to the C-terminal part of p97 without competing with p47 for binding (Allen et al. 2006). Binding partners of PNGase are numerous, though. Not only associating with Rad23 and Rpt2, but also with gp78, an ERAD E3 enzyme, the UBX domain containing SAKS and the ERAD retrotranslocation protein Derlin-1 (McNeill et al. 2004; Li et al. 2006; Katiyar et al. 2005). Other PUB domain containing proteins, in addition to PNGase, is RNF31 (RING finger protein 31), an E3 ligase that also contains a UBA domain, two RING domains and Zn-finger domains (Thompson et al. 2004), and Ubxd1, which
also harbors a UBX domain. While RNF31 have not been reported to interact with p97, Ubxd1 have both been shown to interact with the p97 N-terminal via its first 150 amino acid residues and strongly to the C-terminal part of p97 via the PUB domain (Madsen et al. 2008; Kern et al. 2009). In fact, Ubxd1 due to these two interaction sites has been suggested to “reset” cofactor interaction with p97 when arrested in a state of unproductivity (Kern et al. 2009). Binding to p97 via the PUB domain might be abolished by phosphorylation by the next-to-last amino acid residue of p97, a tyrosine at position 805 (Li et al. 2008).

The SHP box

Two of the most studied interaction partners of p97, Ufd1 and p47, have SHP box motifs. This was first discovered in a yeast Derlin-1 homologue, Dfm1. Derlin-1 is a substrate recruiting factor functioning as a channel in membranes to transport misfolded peptides back out to the cytoplasm from the ER (Ye et al. 2004). Dfm1 has been shown to physically associate with p97 via two cytosolic SHP box motifs composed of eight residues (FxGxGQRn, where x is any amino acid and n is a nonpolar residue) which is generally situated downstream of the UBX domain in proteins containing both a SHP box motif and UBX domain (Yeung et al. 2008; Sato & Hampton 2006). Both yeast Derlin-1 homologues, Der1 and Dfm1 are included in separate complexes together with Cdc48 (p97), Dfm1 being associated with Ubx1 and Ubx7 and Der1 being associated with Ufd1, thereby also having Npl4 in complex since Ufd1 and Npl4, as earlier mentioned, prefer to form heterodimers (Meyer et al. 2000; Goder et al. 2008). Interestingly, Ufd1 itself also contains a SHP box motif (Hitt & Wolf 2004).

Ufd1 was first found in a genetic screen with four other Ufd proteins with importance to the UFD (Ubiquitin Fusion Domain) pathway (Johnson et al. 1995) and Npl4 was originally identified as a nuclear transport factor in yeast (DeHoratius & Silver 1996). Both parts of the dimer can bind ubiquitin in its free form and in cooperation with p97, though neither of the two can bind p97 on their own. Interestingly, association of the Ufd1/Npl4 dimer to p97 enhances ubiquitin binding compared to the free form of Ufd1/Npl4 (Park et al. 2005). p97/Ufd1/Npl4 have been extensively studied in the retrotranslocation of (ubiquitin-tagged) misfolded peptides in ERAD, upstream of the proteasome (Meyer et al. 2000; Ye et al. 2001; Jarosch et al. 2002).

Binding of the Ufd1 SHP box motif to p97 does not hinder binding of UBX or UBD domains to p97. In fact, it is suggested that for Ufd1/Npl4
dimer binding, both the Ufd1 SHP box motif and the Npl4 UBD can bind p97. The solved crystal structure of p97 (N-terminal and D1 domains) with the p47 UBX domain bound to it as well as data regarding the non-UBX-competitive binding of the SHP box motif suggest that both UBX and SHP box motifs can bind to p97 without interfering with each other (Bruderer et al. 2004; Isaacson et al. 2007).

Other p97 binding motifs and domains

**PUL** (after PLAP (human homologue of Doa1/Ufd3), Ufd3 and Lub1) domain proteins and PUB domain proteins are similar in that both domains can bind to the last C-terminal part of p97 which can be eliminated by phosphorylation of the tyrosine residue at position 805 in p97 (Qiu et al. 2010; Mullally et al. 2006; Iyer et al. 2004). The PUL domain was first discovered in members of a family of permuted papain fold peptidases of double stranded RNA viruses and eukaryotes, proteins that may function as de-ubiquitinating enzymes (Iyer et al. 2004). Ufd3 and Doa1 have been reported to be important when sorting ubiquitinated membrane proteins into lysosomes via multivesicular bodies (Ren et al. 2008).

**VIM** (VCP interacting motif) is a module prevalent in ubiquitin E3 ligases Hrd1 and gp78. These E3 ligases are tethered to membranes with a cytosolic C-terminal part that interact with the N and D1 domains of p97 (Zhong et al. 2004; Ye et al. 2005). In SVIP (small VCP interacting protein) a VIM motif have been identified in the N-terminal of the protein and the protein, as is apparent from the name, interacts with p97 (Nagahama et al. 2003). A recent study suggested that SVIP in a regulatory manner directs p97 away from the gp78 (Ballar et al. 2007).

**VBM** (VCP binding motif) was discovered in the de-ubiquitinating enzyme ataxin-3. This protein contains a poly-glutamine region that may be expanded, thereby causing Machado Joseph disease (MJD) (Boeddrich et al. 2006). Upstream of the poly-glutamine region is the p97 binding motif. A VBM might also be present in the cytosolic p97 interacting part of E3 ligase Hrd1 (Ye et al. 2005).

Functions for the p97 ATPase and its cofactors

*The role of p97 in transcriptional control and activation*

It has been reported that p97 is involved in the activation of NFκB via
interactions with IκB (Dai et al. 1998). NFκB associates with IκBα and is trafficked in an inactive form to the cytoplasm. After IκBα is phosphorylated and ubiquitinated, p97 binds to it and detaches it from NFκB, which can now activate its target genes. Similar transcription controls have been found in yeast, for the regulation of fatty acid levels, with the NFκB relatives Mga2 and Spt23 (Hoppe et al. 2000; Rape et al. 2001). They are both made as inactive homodimeric precursors, p120, which binds to the ER membrane via their C-termini. One subunit of the two in the p120 pair is ubiquitinated and processed by the proteasome, truncating it at its C-termini. The resulting dimer has one p120 and one p90 subunit still anchored to the ER membrane. A few ubiquitins are still left on the p90, recognized by p97 together with Ufd1-Npl4, releasing p90 for regulation of genes (Hoppe et al. 2000; Rape et al. 2001).

The role of p97 in membrane fusion

In nuclear envelope growth and assembly, both p97 coupled with the non-UBX Ufd1/Npl4 dimer and the UBX p47 is required. The nuclear envelope is closed via the physical interaction of p97/Ufd1/Npl4 with the mitotic kinase Aurora B. p97/Ufd1/Npl4 associates with ubiquitylated Aurora B, removes it from the nuclear envelope and promotes reformation of the nucleus. Although Aurora B is ubiquitylated it is not degraded, p97 probably merely relocates it or regulates its activity (Ramadan et al. 2007). After that, p97/p47 provides a transfer mechanism for adding to the growing nuclear envelope (Hetzer et al. 2001). ER fusion and reformation of Golgi complexes occur at the end of mitosis through interaction of p97/p47 with a t-SNARE (target-soluble NSF attachment protein receptor), syntaxin 5. For membrane fusion to work, syntaxin 5 needs to be ”primed” by binding of p97/p47. Then, dissociation of p97/p47 from syntaxin 5 is performed by VCIP135 (Uchiyama et al. 2002), mediating dimerization of syntaxin 5 on different membranes, fusing them.

The role of p97 in cell cycle control

p97, cdc48, was first found to be involved in cell cycle control in a temperature sensitive mutant in yeast that arrest in G2/M phase (Moir et al. 1982). Further research has revealed that p97 is required for degradation of a G1 cyclin, cyclin E, needed for entry into S phase (Dai & Li 2001) and that, in extracts from Xenopus eggs, spindle disassembly has been shown to require p97/Ufd1/Npl4 (Cao et al. 2003). Recently, it was also reported that nucleus reformation at the end of mitosis require p97/Ufd1/Npl4 to extract a
The p97 in protein degradation

In the ER, misfolded peptides induce the unfolded protein response (UPR), which adjust ER-stress pathways (Travers et al. 2000). As long as misfolded peptides are dealt with in the ER, there is no need for degradation. However, if the problem cannot be solved by the UPR, ERAD (endoplasmic reticulum associated degradation) takes over, transferring the misfolded peptides out of the ER to the cytosol (called retrotranslocation), where they are ubiquitinated via different ubiquitin ligases (Morito et al. 2008; Gnann et al. 2004; Bordallo et al. 1998). Depending on what kinds of peptide domains are misfolded (ER-luminal, intra-membrane or cytosolic domains), different ubiquitin ligases are employed (Carvalho et al. 2006). After ligation of ubiquitin chains to the misfolded peptides, p97/Ufd1/Npl4 binds and transfers the peptides out into the cytoplasm (Meyer et al. 2002; Ye et al. 2003) and possibly all the way to the proteasome. A yeast two-hybrid screen revealed that Ubx2 and Ubx3 (a yeast homologue of p47) interacted with p97. p97/Ubx3 was indicated to deliver substrates to the proteasome, but Ubx2 showed no indication of involvement in proteolysis (Hartmann-Petersen et al. 2004). Later that same year it was reported that among the UBX containing proteins, all of them interacted with p97 (Schubert et al. 2004), and Ubx2 is a catalyst for proteolysis by supplying a connection for transfer of ubiquitin chain-linked ER substrates from ubiquitin ligases to p97 (Schuberth & Buchberger 2005). Until recently it has been believed that the p97 ATPases were always responsible for the extraction of ERAD substrates while the 19S ATPases had the sole job of unfolding substrates in the cytoplasm. Recent work has shown that the 19S ATPase sometimes can switch roles with the p97 and actually serve to extract ERAD substrates from the ER (Lipson et al. 2008). Our current work complements this latest finding and indicates that p97 may also serve unfolding roles for the proteasome, Paper I.

The p97 in human disease

Endogenous p97 has been implicated in many different neurological diseases, for example co-localized with polyglutamine repeat aggregates in brain of Huntington's and Machado-Joseph disease patients (Hirabayashi et
al. 2001), as well as with Lewy bodies in Parkinson's disease and dementia patients (Hirabayashi et al. 2001; Mizuno et al. 2003; Kitami et al. 2006). Also in several cases of cancer, p97 has been implicated as an upregulated factor, probably due to involvement in regulating p53 levels (Tsujimoto et al. 2004; Yamamoto et al. 2004; Yamamoto et al. 2005; Sherman et al. 2007).

When a GFP-tagged p97 was expressed in cells with induced polyglutamine aggregates, p97 stayed localized to aggregates until they were fully dissipated, though knockdown or expression of a dominant negative mutant caused formation of aggregates (Kobayashi et al. 2007). Other studies have also implicated p97 in prevention of aggregate formation, more specifically heat denatured luciferase aggregate formation (Thoms 2002; Song et al. 2007). On the other hand, expression of loss of function Drosophila p97 mutants prevents aggregate formation in a polyglutamine disease model system (Higashiyama et al. 2002). The p97 has been seen localizing to inclusion bodies in patients with IBMPFD (inclusion body myopathy associated with paget disease of bone and frontotemporal dementia), a late onset disease, which led to identification of a group of single amino acid residue p97 mutants as the responsible factors for the disease. So far nine different p97 mutations have been found (Watts et al. 2004; Haubenberger et al. 2005; Forman et al. 2006). Currently it is still unclear whether p97 induced aggregates part of the normal process to remove proteins or signs of cellular stress and compromised degradation; are p97 induced aggregates the cells own treatment of the disease, or are they the cause of the disease? All evidence points to p97 ”regulating” aggregates, but there is still much work to be done until a clear understanding of the function for p97 and aggregation is obtained.

Transcriptional regulation of proteasome subunits

A wide range of studies have been initiated in studying how the proteasome machinery participates in transcriptional regulation, but barely none touch upon the transcriptional regulation of proteasome genes themselves. What little data have been collected so far is that in Saccharomyces cerevisiae, proteasome levels are closely regulated by Rpn4p, an extremely short-lived protein (Xie & Varshavsky 2001; Mannhaupt et al. 1999). This factor is degraded by proteasomes and if the amount of proteasomes decrease, the amount of Rpn4 increase, thereby stimulating the transcription of proteasome subunits.
Rpn4 is known to associate to a cis-regulatory element, PACE, upstream of proteasome subunits, thereby inducing the transcription. The problem is that Rpn4 is only present in *Saccharomyces cerevisiae* and close relatives thereof, but not in several other eukaryotes such as *Saccharomyces pombe*, *Arabidopsis thaliana*, or humans (Mannhaupt & Feldmann 2007).

Recently, a study was carried out that showed the human transcription factor Nrf1 (nuclear factor erythroid-derived 2-related factor 1) specifically exhibited proteasome regulatory activity (Radhakrishnan et al. 2010), which was originally initiated by our early findings that are presented in our second paper/manuscript, showing that *Drosophila* may be utilized as a model system for studying proteasome regulation more closely.

Both research and drug development could benefit greatly from increased understanding of the regulation of mammalian proteasome transcription factors, in such that proteasomes could be regulated on a whole different level without having to handle compounds inhibiting proteolytic activity, but instead inhibiting the transcription of proteasome subunits in a negative or positive manner.

bZIP proteins

Over two decades ago Landschulz and coworkers discovered one of the first mammalian transcription factors, CCAAT/enhancer binding protein α (C/EBP α). Through these studies was established that the bZIP DNA-binding domain (DBD) consists of a region rich in basic residues linked to a “zipper” formed by a heptad repeat of hydrophobic residues, most often leucines (Landschulz et al. 1988; Landschulz et al. 1989). DNA binding is achieved through specific motifs, such as the MARE (Maf recognition element), the TRE (12-0-tetradecanphylphorbol-13-acetate (TPA)-responsive element), and the ARE (Xie et al. 1995; Jaiswal 1994; Kataoka et al. 1994), and is mediated by the basic domain through hydrogen bonds and hydrophobic interactions between amino acid side chains and bases in the major groove of DNA. Hetero- or homodimerization of bZIP factors are achieved through the leucine zipper, which through a coiled-coil structure with two parallel α-helices wrapped around each other yielding a hydrophobic surface through which dimerization can be achieved (Figure 5; Ellenberger et al. 1992; Glover & Harrison 1995; Miller et al. 2003; Rauscher et al. 1988).

In the bZIP class of molecules are included a large variety of different proteins such as the NF-E2 (nuclear factor-erythroid 2), v-Maf, p45 and the CNC-family. Because of the similarity in nature between bZIP domains of
**Figure 5. Structure of the bZIP protein.** C/EBPα in complex with DNA, found by Landschulz and coworkers (1988). Shown in red is the Leucine zipper, in orange are the heptad repeat leucines forming the zipper, in yellow is the basic region and in blue is the DNA. Illustration based on experimental data by Miller et al. (2003), PDB code 1NWQ.

Different proteins different kinds of heterodimers can form, essentially causing an interaction between bZIPs. Utilizing this, a platform is provided where a small number of molecules can regulate many processes (Newman & Keating 2003).

**The CNC-family**

In 1989 a mammalian factor with a binding site resembling NF-E1 (nuclear factor erythroid 1) was found and named NF-E2 due to the similarities between the two in DNA binding. NF-E2 was found to be comprised of a
factor limited to certain tissues, the CNC (Cap'n'Collar) domain containing protein p45 and a smaller, ubiquitously expressed p18 subunit similar to the v-Maf oncogene thereby belonging to the Maf protein family (Andrews et al. 1993; Mignotte et al. 1989). Following the discovery of NF-E2 and p45 the CNC gene was identified as a family member belonging to the bZIP pedigree of transcription factors (Mohler et al. 1991). Because of the presence of a bZIP domain and based on similarity between the fos transcription factor and CNC leucine zipper repeats, CNC was expected to form DNA associating heterodimers with other bZIP proteins (Mohler et al. 1991; O'Shea et al. 1989). The name CNC was assumed because of the expression in cap (head or anteriormost laberal) and collar (mandibular) segments of Drosophila melanogaster embryos and was from here on the collective name for this specific family of heterodimer bZIPS (Mohler et al. 1991).

Following the discovery of NF-E2/p45 and CNC, the search for other NF-E2 (or NF-E2-type MARE) binding factors continued. Probing for binding of proteins to a cDNA with the NF-E2 DNA binding site led to the identification of mammalian transcription factors Nrf1 and Nrf2 (NF-E2-related factor 1 and 2, respectively) (Chan et al. 1993; Moi et al. 1994). These transcription factors are both bZIP proteins with a CNC domain and were observed to use small Maf proteins as obligatory heterodimeric partner molecules for binding to MAREs (Toki et al. 1997). Nrf3 was finally identified because it was observed that the genes nrf1, nrf2 and p45 are situated near hoxB, hoxD and hoxC genes, respectively. By using this information and the hypothesis that CNC family genes evolved from a common ancestor, as hox genes did, Nrf3 was found near the hox gene cluster hoxA and was confirmed to heterodimerize with MafK (Kobayashi et al. 1999).

Drosophila melanogaster CNC transcription factors are produced from one single gene containing different start sites, and although there are seven possible gene products only three functional proteins have been verified (McGinnis et al. 1998). In contrast to the Drosophila melanogaster CNC transcription factors are the mammalian homologues. These come from separate genes, suggesting duplications from a common ancestor. Even though Drosophila melanogaster CNC have been suggested to be a homologue of Nrf2 specifically (Sykiotis & Bohmann 2008), evidence of conserved regions suggest a more general conservation where all mammalian CNC transcription factors originate from one single transcription factor such as the Drosophila melanogaster homologue (Manuscript II of this thesis).

Despite the fact that p45 was the CNC transcription factor first found in
mammals, today the Nrf transcription factors are best understood and well described. All three of these activate ARE driven genes and have been shown to associate with small MaF's. Nrf1 and Nrf2 are expressed in basically all tissue examined while Nrf3 have a more restricted expression pattern (Moi et al. 1994; Chan et al. 1993; Kobayashi et al. 1999). Nrf transcription factors have been studied in relation to their importance for the whole organism through knock out mice studies. Nrf1 knock out mice rarely survived as embryos, with the few doing so dying only a few hours after birth (Chan et al. 1998). Overexpression of Nrf2 cannot rescue the loss of Nrf1 (Leung et al. 2003). Nrf2 and Nrf3 mice showed no early development lethality phenotype. Nrf2 deficient mice appear to handle different types of stress poorly while Nrf3 deficient mice did not exhibit any phenotype at all (Chan et al. 1996; Iida et al. 2004; Ramos-Gomez et al. 2001; Zhang et al. 2009).

The factor for regulating Nrf2 driven gene transcription have been found to be the protein Keap1 which is bound to actin. Keap1 binds Nrf2, effectively keeping it out of the nucleus in concert with an E3 ubiquitin ligase, through which Nrf2 is ubiquitylated and degraded (Itoh et al. 1999; Zhang et al. 2006). Binding of Keap1 to Nrf1 has also been studied, but it was concluded that a weaker interaction between Keap1 and Nrf1 than with Nrf2 prevented down-regulation. Finally, Nrf3 does not have the motif needed to associate with Keap1 (called ETGE; Kobayashi et al. 2002), though it does apparently share a similar N-terminal conserved region with Nrf1.

Important for our second study (Paper II) is that Nrf1 and Nrf3 have been shown to be tethered to membranes (outside of the ER membrane and the cell membrane) and need part of the N-terminal proteolytically cleaved before transport into the nucleus activating ARE driven genes (Zhang et al. 2006; Zhang et al. 2009). As mentioned above, Nrf2, lacking this N-terminal part is instead regulated by Keap1 to activate ARE driven genes. We have observed possible degradation products from a construct of the cnc-c specific N-terminal region (amino acids 1-576) fused to a V5/His tag, suggesting a possible membrane tethering also here.

As with regard to the proteasome, genes have been shown to be upregulated in liver cells of mice both treated with antioxidants and with Nrf2 overexpressed, though importantly disruption of Nrf2 did not alter gene expression in any way (Kwik et al. 2003). In another study done with colon cancer patients high Nrf2 activity was observed. While interesting in itself it was also observed that proteasome levels were higher in the tumor tissue (Arlt et al. 2009). We have observed in our second study that Drosophila melanogaster proteasome genes have potential ARE motifs upstream of their
19S base and 20S core proteasome genes and that these *Drosophila melanogaster* proteasome genes are CnC regulated. This is supported by the observation of proteasomal upregulation when Nrf2 is overexpressed as well as the fact that DeShai es and coworkers (Radhakrishnan et al. 2010) recently reported that Nrf1 mediates recovery of proteasome subunits after proteasome inhibition. This altogether support the *Drosophila melanogaster* as a useful model system to understand metazoan-proteasome regulation.

**ARE motifs**

In 1991, a DNA sequence promoter element responding to different antioxidative factors (the ARE – antioxidative-responsive element) such as hydrogen peroxide and phenolic antioxidants was found. Using mutational and deletional analysis in the 5' flanking region of the rat glutathione S-transferase Ya subunit gene and the NAD(P)H:quinone reductase gene, Picket and co-workers could identify part of the pathway to sense and respond to oxidative stress (Rushmore et al. 1991). Since then ARE consensus sequences have been identified in the 5' flanking region of numerous cytoprotective and antioxidant genes. The consensus binding sequence of the ARE motif (5'-TGA[C/T]nnnGC-3') show relation to DNA recognition sequences of NF-E2 and v-Maf (5'-TGAsTCAGC-3' and 5'-TGCTGACTCAGCA-3', respectively) (Hayes & McMahon 2001).

ARE sequences have actually been observed present in promoter regions of proteasome subunits, suggesting a transcriptional regulation acting via these elements (Kwak et al. 2003).
Methods

To study how ubiquitylated substrates are processed prior to their degradation by the 26S proteasome, I have used a variety of techniques; RNA interference, a Cu-inducible expression system and subsequently methods such as immuno blotting, flow cytometry and a live cell proteasome assay with ubiquitin-fluorescent protein constructs.

The model system *Drosophila melanogaster*

Working with molecular tools in *Drosophila melanogaster* was made much easier when the first draft of its genome was sequenced (Itoh et al. 1999; Zhang et al. 2006). With this, and a very large body of past work, it could be determined that many human proteins share a high sequence-similarity and function with *Drosophila melanogaster*. Using a *Drosophila* Schneider 2 (S2) cell line (Schneider 1972), we have an elegant and easy way to work with this model system for studying human-related molecular mechanisms. *Drosophila* S2 cells can be grown attached to surfaces or in suspension, and are easy to maintain. Stable cell lines can be established in just three to four weeks without the need to isolate and screen colonies, as is done with mammalian cell lines. Knockdown of genes is also easier to achieve with S2 cells; instead of using transfection reagents and vectors, dsRNA only needs to be spread on cells in serum free media (SFM). The dsRNA treatment can be done with long sequences that are processed into many different siRNAs, which is the preferred method for RNA interference.

RNA interference

In 1998, Fire and coworkers saw that double stranded RNA injected into *Caenorhabditis elegans* gave rise to silencing of genes (Fire et al. 1998), which was the birth of a new technique and field of research, RNA interference (RNAi).

RNA interference can be both naturally occurring and introduced from an external source. So called micro RNAs (miRNAs) are expressed endogenously and processed in the nucleus into hairpin shapes by the microprocessor complex, composed of DROSHA and PASHA. Both the endogenous miRNAs and the exogenously introduced long double stranded
RNA (dsRNA) are processed further by DICER into small interfering RNAs (siRNAs), around 20 nucleotides (nt) in length. The siRNAs are then recognized by and introduced into RISC (RNA-inducing silencing complex), which unwinds the RNA. When unwound, the anti-sense RNA guides RISC to messenger RNA (mRNA) having the complementary sequence, resulting in cleavage of the target mRNA, and thus silencing of the gene (Bernstein et al. 2001; Gregory et al. 2006).

In our studies, we use a dsRNA of around 600 nucleotides synthesized in vitro, with T7 RNA-polymerase, from a dsDNA template designed from exonic sequences in the gene of our protein of interest. For efficient knockdown in S2 cells, the cells are pre-incubated and washed with serum free media and the 500-700 nt dsRNA is added. After four to seven days a transient knockdown is achieved and protein levels can be analyzed with a western blot.

Expression of exogenous proteins

For functional studies of proteins in cells and organisms, not only RNA interference can be used, but also expression of a protein of interest. In *Drosophila melanogaster* S2 cells, this is performed by cloning a gene into a vector containing a metallothionein promoter, the pMT vector. After introduction of the vector into cells, either through transient or stable transfections, the promoter can be differently induced by adding different amounts of copper or cadmium directly into the cell medium (Bunch et al. 1988).

Since p97 hexamerizes, the mutant forms can oligomerize with wild-type monomers. For mutants that have a dominant functional phenotype we can analyze directly. For recessive effects, we are required to deplete the wild-type gene products with RNA interference.

Living cell UPS degradation assay

In order to study the ubiquitin-proteasome system in living cells, we have utilized a range of fluorescent protein reporter systems. These constructs are rapidly degraded by a functional ubiquitin proteasome pathway and contain either ubiquitin or a degradation signal conjugated to EGFP (enhanced green fluorescent protein), RFP (red fluorescent protein) or YFP (yellow fluorescent protein). Fluorescent reporter proteins fused to ubiquitin is
comprised of either ubiquitin with a G76V mutation at its C-terminal part (Ub-G76V) or a methionine to arginine substitution of the the most N-terminal amino acid of the fluorescing protein (Ub-R). Ub-G76V is a ubiquitin fusion domain, UFD, substrate meaning that ubiquitin cannot be cleaved from the fluorescing substrate resulting in a rapidly degraded protein (Dantuma et al. 2000). Ub-R is an N-end rule substrate (Bachmair et al. 1986; Dantuma et al. 2000). In contrast to Ub-G76V, Ub-R substrates have a cleavable ubiquitin. When cleaved, the substrate is recognized as as a damaged protein resulting in its subsequent removal and degradation. CL1 reporter substrates have a degradation signal, a C-terminal hydrophobic polypeptide of 16 amino acids (CL1), fused to the C-terminal part of substrates (Menéndez-Benito et al. 2005). Lastly, CD3e substrates contain a misfolded ERAD substrate T-cell receptor fused to the N-terminal part of a fluorescing protein (Menéndez-Benito et al. 2005).

Utilizing these fluorescing substrates, we can monitor the success of our experiments in live cells with manual and automated fluorescence microscopy, and also with flow cytometry. Flow cytometry can be used to count and sort cells in a fluid stream and is superior to single imaging from a microscope, since it can count and register fluorescence intensity for thousands of cells in a matter of just a few minutes.
Summary of publications

Paper I

**A conserved unfoldase activity for the p97 AAA-ATPase in proteasomal degradation**

The p97 ATPase is a complex that has been implicated in many different cellular functions. In regard to protein degradation and p97, the most studied mechanism must be the ERAD, Endoplasmic Reticulum Associated Degradation. If the UPR, Unfolded Protein Response, can't successfully process misfolded proteins, the proteins get retro-transported out of the ER, where p97 together with cofactors target the misfolded proteins to the proteasome. In this study we go further into the function of p97 in protein degradation, but with a focus on cytoplasmic degradation instead of ERAD.

A study done earlier within our group (Lundgren et al. 2005) implicated a co-regulation of p97 and proteasome subunits that led us to study in closer detail the degradation of cytoplasmic substrates, specifically the UFD substrates Ub$^{G76V}$GFP and Ub$^{G76V}$RFP.

Knockdown of *Drosophila melanogaster* p97 causes a massive accumulation of ubiquitin conjugates, as well as stabilization of the proteasome reporter substrates. Since p97 might need cofactors to assist the proteasome in degradation of cytoplasmic substrates we performed a set of RNAi screenings to identify essential cofactors such as Npl4 and Ufd1 as well as other potential cofactor proteins that contained conserved UBX domains. We observed the importance of two p97 cofactors, Ufd1 and Npl4. Other tested cofactors of the p97 ATPase were not observed to stabilize the soluble proteasome substrates in our screening.

To address if the ATPase activity of p97 is required for degradation of the soluble substrates p97 mutants were constructed in the proposed active sites on inducible plasmids, the Walker A and B sites in the D1 and D2 ATPase domains, as well as in the proposed substrate interacting site, the pore. Mutants in the ATPase domains have been described earlier in dislocation of proteins from the ER membrane (Ye et al. 2003). A site directed mutation in the Walker A site of the ATPase domain in p97 complexes prevents ATP binding, while mutations in the Walker B site still allows binding but
prevents ATP hydrolysis. Reports suggest that the D2 domain is responsible for the majority of p97's ATPase activity and the D1 domain have importance in the formation of the p97 hexamer (Song et al. 2003; Wang et al. 2003). Abolishing the pore activity has been shown to implicate ERAD negatively (DeLaBarre et al. 2006).

Stable Drosophila S2 cell lines were generated from the mutants and analyzed using the RNAi+c technique, where the endogenous protein is knocked down and the exogenous protein is induced. With this we could show that all sites were important for full degradation of Ub⁸⁶V-GFP, but that the D2 domain was of particular importance. Also the pore activity was essential in degradation of our model substrate.

It is clear that the p97 and its ATPase domains are important in degradation of a cytoplasmic substrate, but what is the actual effect or change that the p97 makes on the proteasome substrates? It was hinted to us through the previous finding (Takeuchi et al. 2007) that GFP with a proteasome association element could not be directly degraded by the proteasome even though a proteasome association element docked the GFP near the 26S proteasome ATPase base. However, with an added unstructured peptide extension the GFP could be rapidly degraded (Higashiyama et al. 2002). By using this past observation and designing a Ub⁸⁶V-GFP with a V5/His C-terminal peptide extension we could remove the requirement for the p97 dependent degradation of the reporter proteasome substrate. We proposed the hypothesis that within the cell the p97 ATPase could generate partial unfolded ends that allowed the proteasome to degrade substrates that lacked proper extensions. Mapping the length of the extension, we found that 20 amino acid residues or longer was needed to circumvent substrate stabilization when the cell was depleted of the p97 ATPase, but 15 amino acids or shorter behaved as with the original proteasome reporters. We also examined the importance of different amino acid composition for the extended peptide sequence. All glutamine or all glycine extensions could not escape the p97 dependence for degradation. Glutamine and glycine extensions induced aggregation of the reporter when stabilized with proteasome inhibitor treatment, therefore an aggregate prone all alanine extension was also examined. This extension could still bypass the need for p97 dependent degradation, indicating that aggregation was not the reason why the glutamine and glycine extensions could not escape the p97 independence.

Finally, we confirmed our findings in Drosophila melanogaster had importance in human cells by looking at HeLa cells stably transfected with UFD reporters with and without a C-terminal unfolded peptide extension. As
in Drosophila, C-terminal peptide extensions in human cells bypass the p97 dependent degradation. Also amino acid length of said extension behave as in Drosophila, with 20 amino acids or more abolishing the need for p97 to degrade the UFD substrates.

These data lead us to propose a model where p97, via interaction with its pore region transfer substrates to the proteasome, possibly exhibiting an unfolding activity via the ATPase domains to extract free peptide sequences that can be fed to the proteasome. Recent additional studies in the yeast support our model where the proteasome requires the presence or the generation of unfolded extensions on its substrates for degradation to occur (Heinen 2010). The 26S proteasome interacts with a number of additional proteins that shuttle substrates to the proteasome. These shuttling factors escape degradation from the 26S proteasome and increase the complexity and function of the proteasome in our cells. If the proteasome ATPases could unfold any protein that was in close proximity of the 26S proteasome would likely also target shuttling factors for degradation and limit the complexity and flexibility of the 26S proteasome.

Paper II

Basic-leucine zipper protein Cnc-C is a substrate and transcriptional regulator of the Drosophila 26S proteasome

A wide range of studies have addressed the roles that proteasomes have on transcription, yet barely any have touched upon the subject of functional regulation of proteasomes themselves. Researchers have found the factor responsible for transcription of proteasome subunits in the yeast Saccharomyces cerevisiae and the closely related Hemiascomycetes, but not in other species such as Schizosaccharomyces pombe or metazoans. The factor in question is named Rpn4 and associates with PACE elements upstream of Saccharomyces proteasome genes.

Since neither PACE elements nor any apparent homologue exist in Drosophila melanogaster, we believed our best option to understand proteasome transcriptional regulation was to do a large genomic RNAi screen depleting transcription factors in a comprehensive approach. From the DRSC RNAi screening center at Harvard Medical School we obtained different double stranded RNA sequences which targeted 993 genes in 384-well plates, aimed specifically towards known and predicted Drosophila
*melanogaster* transcriptional elements. Most genes were targeted twice with different dsRNAs at different exonic regions, except for the exceptionally short *Drosophila* genes. By plating stable S2 cells expressing the substrate Ub$^{G76V}$GFP into wells containing dsRNA directed towards the different transcriptional elements we could screen for factors modifying proteasome expression. The strategy was that when mRNA levels dropped proteasomal component (subunit) levels correspondingly drop and Ub$^{G76V}$GFP accumulates suggesting that the transcriptional element that was depleted in the specific well contributes to proteasomal expression.

Nine 384-well plates from the DRSC RNAi library were screened over a consecutive period of three days (two, three and four days after addition of cells to RNA containing plates). The library was screened at two different times, each plate of the library by two scorers. A scale of 1 to 4 was used to score the wells and the scores were averaged. Wells with a positive score were then distributed into three groups. Wells with a positive score during all three consecutive screening days were termed group I, wells with a positive score during only two of the screening days were termed group II and wells with a positive score on only one of the three days were recorded as group III. Since the top six hits showed the greatest stabilization of Ub$^{G76V}$GFP we decided to examine and re-target these as best potential candidates. New dsRNAs for each of these genes was designed and the stable S2 cells with the reporter was yet again assayed. Out of these six genes, four were re-confirmed. We focused on the top hit, Cnc-C, for which we constructed two new dsRNA. Cnc-C belongs to the Cap'N'Collar family of transcription factors which may form heterodimers with maf proteins. In our top hits we found one maf protein, maf-S, though surprisingly we could not re-confirm proteasomal inhibition by targeting of this gene.

The *Drosophila melanogaster* cnc gene can hypothetically encode seven different gene products, though only three have been confirmed, Cnc-A, Cnc-B and Cnc-C. Mammalian homologues of *Drosophila melanogaster* Cnc have on the other hand three different genes, Nrf1, Nrf2 and Nrf3. Protein alignments of Cnc-proteins from different species display shared characteristics between vital domains supposed to associate with the transcriptional inhibitor Keap1. We performed a phylogenetic analysis of metazoan cnc bZIP protein sequences that show that Cnc-C is not a direct homologue of any specific mammalian Nrf protein. Rather, it shows that a duplication of the ancestral Cnc gene occurred in Chordata after the divergence of vertebrates from the other chordates.

Earlier studies done in our group (Lundgren et al. 2005) had found that in *Drosophila melanogaster* cells an upregulation of proteasome gene
transcription as well as for the gene for the AAA ATPase p97 occurred after the RNAi knock-down of a non-essential proteasome subunit, S5a. We hypothesized that if our top hit among the transcriptional regulators, Cnc-C, would regulate proteasome genes, we should be able to knock S5a down without an increase in proteasome subunits if Cnc-C was also depleted. We did RNAi knock-down experiments with the two different Cnc-C dsRNAs and S5a dsRNA, both by themselves and combined. After inhibition we examined the protein expression levels of p97 and the proteasome subunits α7 S5a and Rpt1. When S5a was knocked down by itself we could see an up-regulation of proteasome subunits though when combined with Cnc-C knock-downs an up-regulation could not be achieved. In fact, proteasome and p97 protein levels decreased. Finally, dsRNA treatment that only targets the Cnc-C transcripts was able to show decreases in proteasome levels. From these experiments we could conclude that Cnc-C did indeed regulate the levels of proteasome subunit expression. Also, by utilizing an in-gel native gel electrophoresis proteasome activity assay and analysis of proteasome activity in whole cell lysates of cells treated with the different RNAi combinations we could conclude that both 26S and 20S proteasomes decreased when treated with any of the Cnc-C dsRNAs. Interestingly, knock-down of the Cnc-C inhibitor Keap1 did not increase proteasome levels as would be expected. Rather, a small decrease in proteasome and p97 levels could be perceived.

Utilizing the different combinations of dsRNA and real time PCR we also examined what happened at the transcriptional level of proteasomal mRNA expression in differently treated cells. As expected, knock-down of S5a induced mRNA expression of the examined 19S lid and base subunits Rpn11 and Rpt1 as well as for the 20S subunit α7 and the p97 ATPase. Expression of mRNA did on the other hand decrease when Cnc-C either by itself or together with S5a was targeted for RNAi. It is known that by inhibiting proteasome activity in mammalian and Drosophila cells through the use of proteasome inhibitors proteasome mRNA levels are increased. This increase in proteasome and p97 mRNA levels is blocked when Cnc-C is knocked down.

Since the yeast proteasomal transcription factor Rpn4 is itself regulated by the proteasome, we wanted to examine the proteasomal effect on the Drosophila Cnc-C. We did this by expressing the first 576 amino acids of the Cnc-C exclusive N-terminal region. This large region is unique to the C-transcript of the Drosophila cnc gene and was previously isolated in a cDNA Drosophila library. The Cnc-C construct was tagged with a V5/His tag on the C-terminal end of the large protein domain. Both proteasome inhibitor
treatment and proteasome subunit S5a depletion induced stabilization of induced stabilization of tagged Cnc-C polypeptide. The tagged form of the Cnc-C polypeptide could also be depleted by targeting it with the RNAi that were previously used. Soluble and membrane fractions of treated cells were made from cells targeted by a proteasome inhibitor and S5a RNAi. Cnc-C migrated remarkably slow on SDS-PAGE gels, which has also been observed previously with human Nrf1 and Nrf2 (Zhang et al. 2006). Interestingly, the Cnc-C N-terminal part have strikingly different size comparing between membrane and soluble fractions. This may suggest that Cnc-C has a mechanism similar to mammalian Nrf1 which is bound to the ER and membrane and proteolytically cleaved for activation and transport to the nucleus without the help of Keap1 (Zhang et al. 2006). Truncation of Cnc-C is likely achieved through a proteasome independent N-terminal cleavage of Cnc-C since the C-terminal part contains the V5/His peptide and visualization of protein levels is achieved with a V5 antibody. Even though this is an exciting initial discovery, other forms of post-translational modifications such as glycosylation can not be currently excluded as explanations for the size migration differences observed for tagged Cnc-C.

The current experimental data and the bioinformatic identification of potential ARE like motifs in critical regulatory proteins of proteasome genes suggest that Cnc-C truly plays an important role in transcriptional regulation of proteasome subunits, possibly independent of Keap1. The size difference observed between membrane and soluble fractions as well as the earlier observed Keap1 independent membrane tethering of the mammalian Nrf1 further support this theory. We propose a model where Cnc-C may activate either proteasome genes by a cleavage of the membrane bound N-terminal releasing Cnc-C for translational activation or by activating anti-oxidative genes by the oxidation of the Keap1 protein that releases Cnc-C and allows entry into the nucleus. Currently we are far from understanding the similarities and differences between proteasome regulation and anti-oxidative stress response. Our results point to both critical transcription pathways being regulated by a single transcription factor in Drosophila, Cnc-C. The Drosophila model system should be a powerful system to understand the detail mechanisms and components of these forms of regulation.
Paper III

A high-throughput microscopy method to find novel p97 inhibitors

Proteasome inhibition has proved important in treating multiple myeloma (with the drug Velcade/Bortezomib) and suggests that other drugs inhibiting parts of the ubiquitin proteasome system may be out there to find. Several screening methods to find proteasome inhibitors exist, but few work within a living cell context. These previous screens utilize purified 20S proteasomes and do not typically take into account if compounds might affect other parts of the ubiquitin proteasome system. For the current first generation 20S proteasome inhibitor Velcade/Bortezomib it has been documented that there are several potentially severe side effects associated with this drug among which peripheral neuropathy is a fairly common consequence (Cavaletti et al. 2007). One way to decrease side effects might be to target only a sub-pathway of the proteasome. In order to be able to efficiently screen for compounds inhibiting a part of the proteasomal system we have utilized fluorescent protein ubiquitin fusion domain (UFD) substrates and our knowledge about these acquired in Paper I (Beskow et al. 2009). We previously observed that addition of an unfolded C-terminal peptide tail made UFD substrates able to circumvent the need of the AAA ATPase p97 for their degradation. Using this knowledge we constructed a two-step high-throughput screening method. In the first screening stage we employed a MelJuSo cell line expressing the UFD substrate Ub\(^{G76V}YFP\). Compounds inducing accumulation of fluorescence affect the ubiquitin proteasome system. The second screening stage re-examines the small number of positive hits found in the first stage. This second screen uses a proteasome substrate that is p97 independent for its degradation. It allows the identification of inhibitors that target any area of the p97 sub-pathway by measuring for compounds that lose their inhibitory effect on substrate degradation.

The screening for proteasome inhibitors was initially set up by using a smaller chemical compound library that is representative of a larger library that was designed to omit general toxins or chemicals with non-specific reactive groups. All screenings used automatic microscopy on living cells in a 96-well format. By screening the effect of 5720 chemical compounds for UFD substrate accumulation in living cells we found several compounds
causing significant accumulation (three times more than the mean plate standard deviation). Each well containing compound treated cells were measured twice and the mean cytoplasmic fluorescence were calculated. The effect of these compounds were subsequently evaluated for cellular structural changes such as: cell rounding, vesicle formation, chromosome condensation (apoptosis) and approximate measurements of cell cycle arrests. Seventeen of the original 5720 compounds were deemed to have substrate accumulations to be classified as positive hits to function as potential ubiquitin proteasome pathway inhibitors. Out of 17 compounds 9 were evaluated further. Six of these caused significant auto-fluorescence, whereas one compound caused consistent high cytoplasmic and nuclear substrate accumulation without auto-fluorescence, compound 35-B6. Dose response experiments were done for 35-B6 and the optimum concentration for substrate stabilization was determined to be 5 µM. Subsequent experiments were performed to establish if 35-B6 inhibited the p97 dependent pathway. Transient transfections of UFD substrates with and without a C-terminal peptide tail (V5/His) was subjected to 35-B6 but both types of UFD substrates accumulated indicating that 35-B6 has an effect on a global factor in the proteasome system.

As a follow up eleven chemical analogs for 35-B6 were selected from the larger chemical library and subsequently applied to the MelJuSo cells expressing UbG76V-YFP. Comparing 35-B6 and these analogs reveal that several analogs apparently caused a higher accumulation of substrates than 35-B6 and comparable to a commercially available proteasome inhibitor, epoxomicin. These early results with analogs suggests that there may exist a more potent proteasome inhibitor than 35-B6 among the analogs but requires further testing to determine which chemical modifications makes the most potent inhibitor of this newly found proteasome inhibitor. The use of analogs will also help in narrowing the actual site of action for the 35-B6 compound, by allowing one to identify the specific groups in the compound that are essential for proteasome inhibition.

While the first attempts to identify p97 pathway specific inhibitors did not identify a candidate compound, setting up a screening method for finding novel inhibitors for proteasomal pathways has been successful. In the future additional automation of cell handling should allow for increased compound library size and increased chances to find p97 pathway specific inhibitors. With our current screening protocol we identified several potential compounds and out of these potential compounds we concentrated on our top hit for continued examination including finding analogs and their effect on proteasome substrate accumulation. With our high-throughput cellular
method we have a robust and novel way to find inhibitors for both overall proteasomal activity and specific to the p97-proteasome sub-pathway. Being able to dissect the ubiquitin proteasome pathway with unique assays should in the future allow new novel drugs to be developed and hopefully increase potency and decrease harmful side-effects.
Concluding remarks

Targeted protein degradation that allows specific removal of proteins is crucial for cells to function and thereby regulation of the proteasome levels is also essential. Both the proteasome and p97 ATPase are utilized in a wide range of functions such as mere housekeeping functions to remove misfolded and damaged proteins as well as rather more complex tasks such as degradation of cell cycle related proteins. Interest is gaining in utilizing proteasome inhibition in treatment of disease (cancer) such as multiple myeloma. The studies conducted in this thesis gives a brief overview over substrate unfolding and degradation (**Paper I**) as well as transcriptional control of proteasome subunits (**Paper II**). Methodology and technology has recently advanced in allowing for high-throughput screening of thousands of chemical compounds in a matter of a few days, demonstrated by the high-throughput microscope screening of chemical compounds to target the proteasome as we performed in **Paper III**. Knowledge of substrate specific sub-pathways of the proteasome may be utilized in finding new candidate compounds for treating proteasome and p97 ATPase influenced disease. In the future the regulatory part of the proteasome machinery should be an interesting target to develop inhibitors or activators against for both research and treatment of disease.
Acknowledgments

A big thank you to

Pat, for the opportunity to join his group in the first place. There's been ups and downs, but it seems like I managed to get through this. You've had lots to teach, not all of which I've managed to absorb.

Anne, du stormade, och stormar fortfarande, in med energi både på labbet, kontoret och i lunchrummet.

Neus, för ditt stöd och för att du fick mig lite mer strukturerad i slutet av avhandlingsarbetet.

Micke, för spelkvällar på Molbio och för introduktionen till klättringen, min nya last

Thomas och Thorsten för trevliga luncher, diskussioner, öl and all that jazz.

Johan, för hjälp med qPCR och lunchdiskussioner;

Ingegerd, Cissi och Petra, för terapisessioner och allt annat

Helene, för barndiskussionerna som tråkade ut alla icke-föräldrar på avdelningen

Daniel och Helge, det finns inga bättre personer att dela kontor med än er!

All of the people at the Molbio and GMT departments: the “old” and the “new”. Both departments have been really good to me and I'm happy to have had the opportunity to be a part of two departments at once.

Matlaget! You rule! Klättring snart, Erik och Daniel?

Alla mina syskon, Ellinor, Sandra, Ida, Edwin, Signe, Fredrik

Min mamma och Leif; Min pappa och Anna

Min förlängda familj, Gäredalingarna: Inge, Mona, Carro, Henrik (inte Gäredaling men typ ingift, eller hur?), Noel, Ylva

Och sist men inte minst min egen lilla familj: Anna och Isa, och Isas kommande lillasyster/-bror. Jag älskar er ❤️!
References


Arlt, A., Bauer, I., Schafmayer, C., Tepel, J., Müerköster, S.S., Brosch, M., Röder, C., Kalthoff, H. et al., 2009. Increased proteasome subunit protein expression and proteasome activity in colon cancer relate to an enhanced activation of nuclear factor E2-related factor 2 (Nrf2). Oncogene, 28(45), 3983-3996.


Chan, K., Lu, R., Chang, J.C. & Kan, Y.W., 1996. NRF2, a member of the NFE2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24), 13943-13948.


Dick, T.P., Nussbaum, A.K., Deeg, M., Heinemeyer, W., Groll, M., Schirle, M., Keilholz, W., Stevanovic, S. et al., 1998. Contribution of proteasomal beta-subunits to the


Glickman, M.H., Rubin, D.M., Coux, O., Wefes, I., Pfeifer, G., Cjeka, Z., Baumeister, W., Fried, V.A. et al., 1998. A Subcomplex of the Proteasome Regulatory Particle Required for Ubiquitin-Conjugate Degradation and Related to the COP9-


Li, G., Zhao, G., Schindelin, H. & Lennarz, W.J., 2008. Tyrosine phosphorylation of ATPase
p97 regulates its activity during ERAD. *Biochemical and Biophysical Research Communications*, 375(2), 247-251.


Pinter, M., Jekely, G., Szepesi, R.J., Farkas, A., Theopold, U., Meyer, H.E., Lindholm, D.,


Richly, H., Rape, M., Braun, S., Rumpf, S., Hoege, C. & Jentsch, S., 2005. A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and


Zhang, Y., Crouch, D.H., Yamamoto, M. & Hayes, J.D., 2006. Negative regulation of the Nrf1 transcription factor by its N-terminal domain is independent of Keap1: Nrf1, but not Nrf2, is targeted to the endoplasmic reticulum. The Biochemical Journal, 399(3), 373-385.

Zhang, Y., Kobayashi, A., Yamamoto, M. & Hayes, J.D., 2009. The Nrf3 transcription factor is a membrane-bound glycoprotein targeted to the endoplasmic reticulum through its N-terminal homology box 1 sequence. The Journal of Biological Chemistry,