Stony Brook University

Biochemical and Structural Characterization of Interactions Mediated by the N- and C-terminal Domains of Mouse PNGase and p97

A Dissertation Presented

by

Xiaoke Zhou

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

August 2007

Stony Brook University

The Graduate School

Xiaoke Zhou

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree,
hereby recommend acceptance of this dissertation.

Hermann Schindelin, Ph.D.
Advisor
Professor, Rudolf-Virchow-Center, University of Würzberg, Germany
Formerly Associate Professor, Department of Biochemistry, Stony Brook University

William J. Lennarz, Ph.D.
Co-advisor
Distinguished Professor, Department of Biochemistry, Stony Brook University

Erwin London, Ph.D.
Chairperson of Defense
Professor, Department of Biochemistry, Stony Brook University

Nicolas Nassar, Ph.D.
Assistant Professor, Department of Physiology and Biophysics,
Stony Brook University

Nisson Schechter, Ph.D.
Outside Examiner
Professor, Department of Psychiatry and Behavioral Sciences, Stony Brook University

This dissertation is accepted by the Graduate School.

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Biochemical and Structural Characterization of Interactions Mediated by the N and C-terminal Domains of Mouse PNGase and p97

by

Xiaoke Zhou

Doctor of Philosophy

in

Biochemistry and Structural Biology

Stony Brook University

2007

The inability of N-linked glycoproteins to adopt their native conformations in the endoplasmic reticulum (ER) leads to their retrotranslocation into the cytosol and subsequent degradation by the proteasome. This pathway is termed the ER associated degradation (ERAD) pathway. During this process peptide:N-glycanase (PNGase) hydrolytically removes the N-linked glycan chains from denatured glycoproteins in the cytosol. PNGase is highly conserved in eukaryotes and plays an important role during ERAD. In higher eukaryotes, PNGase has an N-terminal and a C-terminal extension, in addition to the central catalytic domain, which is structurally and functionally related to transglutaminases. The function of the C-terminal domain has not previously been characterized, whereas the N-terminal domain of PNGase was proposed to be involved in

protein-protein interactions. Recently, the N-terminal domain of p97 has been found to directly interact with Derlin1, a multifunctional AAA ATPase. P97 has been suggested to extract the unfolded ERAD substrates from the ER through the protein dislocating channel located in the ER membrane. P97 associates via its N-terminal domain with various cofactors, which recruit and process ubiquitinated substrates.

In this dissertation, the high resolution crystal structures of the mouse PNGase N-terminal and C-terminal domains were determined by X-ray crystallography. The interaction between p97 and the N-terminal domain of PNGase was further characterized by a crystallographic study of this domain in complex with a peptide derived from p97. For the mouse PNGase C-terminal domain, it could be demonstrated that it binds to an oligo-mannose unit that is an integral part of the N-linked oligosaccharide chains. Biochemical studies revealed that the C-terminal domain of mPNGase enhances the deglycosylation activity of its core domain, presumably through a tighter binding of mPNGase to the glycan chains of misfolded glycoproteins. Finally, the structure of the mPNGase C-terminal domain in complex with mannopentaose defined the interactions in atomic detail and confirmed the biochemical studies.

Recently, Derlin1, an ER membrane protein, has also been shown to directly interact with p97. Derlin1 has been suggested to participate in the retrotranslocation of misfolded glycoproteins during ERAD. In the context of this thesis, the regions mediating the interaction between p97 and Derlin1 were mapped by biochemical techniques, and the corresponding complex was crystallized. The multiple interactions between PNGase, p97, Derlin1 and other p97 cofactors indicate that the process of ERAD is highly cooperative.

Table of Contents

List of Abbreviations

Å	Ångstrom (10^{-10} m)
AAA	ATPase associated with various activities
AMFR	Autocrine motility factor receptor
Cα	Alpha carbon
CD	Circular dichroism
CHO-IAc	Iodoacetamidyl-modified oligosaccharides
CPY	Carboxypeptidase
cv	Column volume
Da	Dalton
DTT	Dithiothreitol
E1	Ubiquitin activating enzyme
E2	Ubiquitin conjugating enzyme
E3	Ubiquitin ligating enzyme
EMTS	Sodium ethylmercurithiosalicylate
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated protein degradation
FOM	Figure of merit
Glc	Glucose
GlcNAc	N-acetylglucosamine
Δgpαf	A nonglycosylated derivative of the yeast pro-α factor mating pheromone
GST	Glutathione-S-transferase
H	Enthalpy

HEK	Human embryonic kidney
IPTG	Isoproyl-β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_d	Dissociate constant
LC-ESI-MS	Liquid chromatography electrospay ionization mass spectrometry
M	mouse/murine
Man	Mannose
MHC	Major histocompatiblity complex
NMR	Nuclear magnetic resonance
NSLS	National Synchrotron Light Source
MME	Monomethyl ether
MWCO	Molecular weight cut off
OT	Oligosaccharyltransferase
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PNGase	Peptide:N-glycanase
PUB	Peptide:N-glycanase/UBA or UBX-containing protein domain
rms	Root mean square
RNase	Ribonulease
RNAi	RNA interference
S	Entropy
SBD	Sugar binding domain

SCF	Skp1, Cul1, Roc1/Rbx1 and F-box proteins complex
SDS	Sodium dodecyl sulfate
SIRAS	Single isomorphous replacement and anomalous scattering
SRP	Signal recognition particle
TCRα	T-cell receptor α chain
Tris	Tris (Hydroxylmethyl) aminomethane
UBA	Ubiquitin associated
UBL	Ubiquitin like
Ufd	Ubiquitin fusion degradation
VIMP	VCP-interacting membrane protein
WH	Winged Helix
XPCB	xeroderma pigmentosum group C complementing protein binding
y	yeast
Z-VAD-fmk	Carbobenzyloxy-Val-Ala-Asp-α-fluoromethylketone

List of Figures

List of Tables

Acknowledgements

First, I would like to express my sincere acknowledgements to my dissertation advisors Drs. Hermann Schindelin and Caroline Kisker, whose wonderful supervision and encouragement helped me all the time during my research. And I would also like to thank my co-advisor Dr. William J. Lennarz for his great help and support with my project. I thank my committee members Drs. Erwin London, Nicolas Nassar, Nisson Schechter for their suggestions and guidance. I would recognize the assistance of Dr. Robert Haltiwanger and Nadia Rana with the Mass spectrometry experiments, Dr. Jae-Hyun Cho for the help with the CD spectroscopy. I am grateful for all the members of the Kisker and Schindelin group, especially Liqun Wang, Drs. Gang Zhao, James Truglio and Song Xiang for their help and during my research. I also thank my collaborators: the Lennarz group members, especially the help from Drs. Gangtao Li and Geng Tian. Finally, I dedicate this thesis to my parents, my brother and my husband, for their love, patient and support.

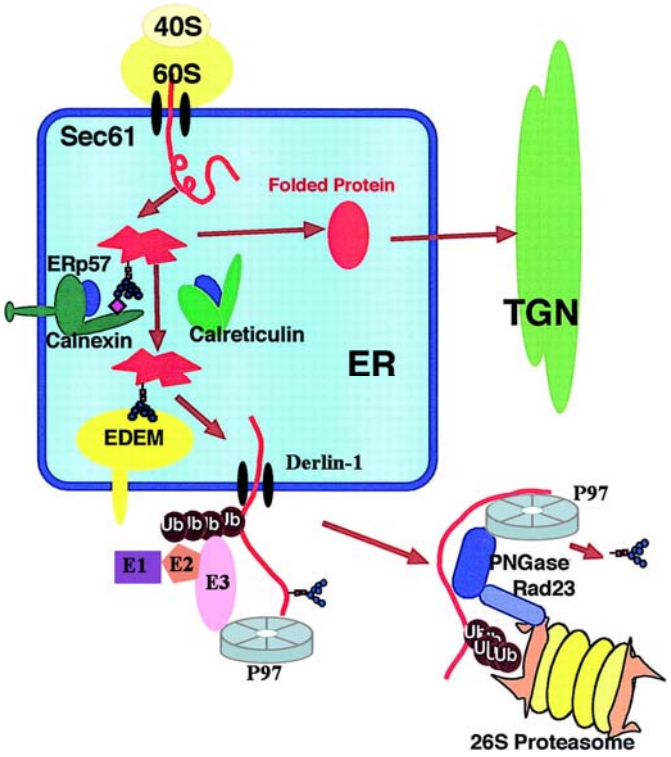
Chapter 1

Main Introduction

A. Endoplasmic reticulum-associated protein degradation

The endoplasmic reticulum-associated protein degradation (ERAD) pathway is a major component of the quality control system in the secretory pathway. It is responsible for the removal of misfolded proteins and unassembled protein subunits (Figure 1.1*A*). It is predicted that a third of all mammalian genes encode proteins passing through the secretory pathway [1]. Secretory protein synthesis is initiated in the cytosol and the signal recognition particle (SRP) guides the ribosome to associate with the ER translocation machinery. The SRP recognizes and binds to the ER signal sequence on the nascent polypeptide and recruits the complex to the SRP receptor located in the ER membrane [2]. Once this interaction is formed, it allows the SRP-ribosome complex to dock to the translocator. SRP and its receptor will then be released. After docking, the growing polypeptide chain is translocated into the lumen of the ER, or inserted into the ER membrane with the aid of the Sec61 complex (Sec61α, Sec61β and Sec61γ in mammals, Sec61p, Sbh1p, Sss1p in yeast) [3]. The homolog of Sec61 in *E. coli* is referred as SecY and the recent crystal structure of a bacterial SecY complex suggested that nascent proteins are transported through its central cavity [4]. Following protein synthesis, many secretory proteins are glycosylated either at the side chains of Asn residues (N-glycosylation) in the ER or at Ser/Thr residues (O-glycosylation) in the Golgi [5]. N-glycosylation requires the consensus sequence Asn-X-Ser/Thr and is carried out by the oligosaccharyltransferase (OT) complex located in the ER membrane [6]. OT catalyzes the transfer of a glycan chain from dolichol-linked pyrophosphate to the conserved Asn residue [6]. Following translation and glycosylation the newly synthesized proteins will

A



B

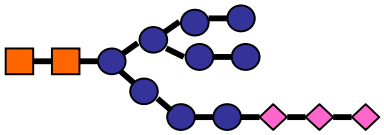


Figure 1.1 **ER-associated degradation pathway of N-glycoproteins.** *(A)*After the N-glycan is added, the nascent peptide associates with some chaperones to help its folding. Correctly folded proteins will exit the ER and be transported to the Golgi for further modification. Aberrantly folded proteins will be dislocated to the cytosol through a putative retrotranslocon. In the cytosol, the unfolded proteins is then ubiquitnated, de-N-glycosylated and degraded by the proteasome. This figure is modified from Yoshida [7]. *(B)* The structure of $(GlcNAc)_2(Man)_9(Glc)_3$. N-acetylglucosamine residues are colored in orange, Mannose in blue and the Glucose in magenta.

be released into the ER lumen, or in the case of membrane proteins into the ER membrane.

However, completion of protein synthesis and N-glycosylation does not guarantee protein maturation. Proteins have to fold into their native conformation in order to carry out their biological functions. Protein folding in the ER is assisted by an elaborate folding machinery including the molecular chaperones, calnexin and calreticulin [7], which utilize N-linked oligosaccharides attached to newly synthesized proteins as tags to detect their folding status [8, 9]. The oligosaccharide chains initially consist of a tetradecamer with the composition $(GlcNAc)_2(Man)_9(Glc)_3$ (Figure 1.1*B*). Immediately after attachment of the tetradecamer, two terminal glucose residues will be cleaved off by glucosidase I and glucosidase II [10]. The resulting protein substrate is then recognized by calnexin and calreticulin and dynamic processing of the innermost glucose residue is essential for proper folding. The misfolded protein will undergo de-/reglucosylation cycles until it is correctly folded or considered to be terminally misfolded [10]. The reglucosylation is catalyzed by UDP-Glc:glycoprotein glucosyltransferase, which senses the conformation of the protein substrate and only acts on misfolded proteins. After successful completion of one or more de-/re-glucosylation cycles, correctly folded proteins are transported to the Golgi for further carbohydrate modification, whereas aberrantly folded proteins are retro-translocated to the cytosol for degradation. It is estimated that about 30% of the nascent proteins are degraded immediately after synthesis [1].

The degradation process takes place in the cytosol, which requires the potential substrate to be transported first from the ER lumen or the ER membrane to the cytosol.

Therefore a retrotranslocon is required for this step, but its identity has not yet been unambiguously established. Recent studies [4, 11-13] suggested that either of two membrane proteins may have this function: the ER translocon Sec61 or Derlin1 [11, 14], a newly discovered membrane protein that has four transmembrane regions (see section D in this chapter for more details).

Once in the cytosol, aberrant proteins will be ubiquitinated, deglycosylated and digested by the proteasome [7]. The ubiquitination system associated with ERAD consists of the canonical cascade usually associated with this process which involves three different types of enzymes, E1 (ubiquitin-activating enzyme), E2 (ubiquitin-conjugating enzyme) and E3 (ubiquitin-ligating enzyme) [7]. While there are only two E1 enzymes [15], there are close to 20 E2 enzymes (in mammals) and hundreds of E3 enzymes [16], which are ultimately responsible for recognizing the substrates. In the context of ERAD the following E3 enzymes are of importance: AMFR (autocrine motility factor receptor) and $SCF^{Fbs1,2}$. AMFR, also known as gp78, is an ER membrane protein that has a 350 amino acid C-terminal tail, which contains a ring finger and a Cue domain [17]. The ring finger domain is essential for its E3 activity and, like other members of this family, is characterized by a consensus sequence containing eight conserved cysteines and histidines that coordinate two Zn ions. Unlike AMFR, $SCF^{Fbs1,2}$ are soluble protein complexes that specifically recognize unfolded N-glycosylated substrates through the interaction between its F-box protein and the $(GlcNAc)_2$-dimer [7], which is also referred to as chitobiose of the attached carbohydrate chains.

Ubiquitin is a highly conserved eukaryotic protein containing 76 residues, which is covalently linked to Lys residues either in the target protein or another ubiquitin

already associated with the target protein. The attachment of several ubiqutin molecules in the form of an oligo-ubiquitin chain functions as a tag for protein degradation. Ubiquitinated N-linked glycoprotein substrates are usually de-N-glycosylated before they are passed on to the proteasome for degradation. The cytosolic enzyme peptide:N-glycanase (PNGase) is the only enzyme that catalyzes this reaction. It functions on high mannose type unfolded proteins and it has been suggested that this process facilitates efficient substrate degradation by proteasome (see section B for details) [18].

The function of the 26 S proteasome is to degrade unneeded or misfolded proteins. The proteasome is composed of two 19 S regulatory subunits and one 20 S proteolytic core subunit [19]. The 19 S subunits recognize and bind to the polyubiquitin chains and unfold the substrate before it enters into the central chamber of the 20 S subunit, the site of proteolytic degradation. The 20 S subunit is composed of 28 subunits arranged in four stacked heptameric rings, which result in a cylindrically shaped particle with a central channel along the cylinder axis, where the proteolytic reactions occur [20]. The entrance to the central channel is very narrow so that only unfolded proteins can enter. Substrates are degraded into short peptides (~8 amino acids), which then exit the proteasome and are further degraded.

Until recently these three processes, ubiquitination, deglycosylation and proteolytic degradation, were thought to be separate events. However, more and more reports indicate that they are tightly regulated. Recently, our collaborators, Drs. Guangtao Li and William J. Lennarz (Stony Brook University), discovered the presence of a murine (m) multi-protein complex which contains the AAA ATPase p97 (which is described in more detail below), mAMFR, mY33K (a protein of unknown activity), mPNGase and

mHR23B [21]. Moreover, since mHR23 also binds to the 19 S subunit of the proteasome [22], interactions between these components establish a tight connection between the ubiquitination, deglycosylation and degradation machineries.

B. PNGase and ERAD

PNGase catalyzes the de-glycosylation of many misfolded N-linked glycoproteins by cleaving the β-aspartyl-glucosamine bond between the first GlcNAc (N-acetylglucosamine) of the attached carbohydrate chain and the amide side chain of the asparagines to which it is attached [23] (Figure 1.2*A*). This reaction produces a free glycan chain with a terminal hydroxyl group and an ammonia molecule, together with the deglycosylated protein in which the glycosylation site asparagine has been converted into an aspartate residue. The activity of PNGase has been suggested to improve ERAD efficiency by removing the bulky glycan chain before the proteins are degraded by the proteasome [18]. PNGase is highly conserved in eukaryotes and possesses a catalytic Cys, His and Asp triad typical of the cysteine protease superfamily. Within this superfamily it is most closely related to the transglutaminases, which catalyze the bond formation between the free amino group of a lysine residue and the γ-carboxamide group of Gln. A well-known member of this family is factor XIII, which catalyzes the formation of fibrin crosslinks as the last step during the blood coagulation cascade. A comparison of the reactions catalyzed by PNGase and transglutaminase reveals that PNGase carries out a reaction, which in essence is a reversal of the reaction catalyzed by transglutaminase. Recent studies revealed that a PNGase homolog in *A. thaliana* (AtPng1p) possesses

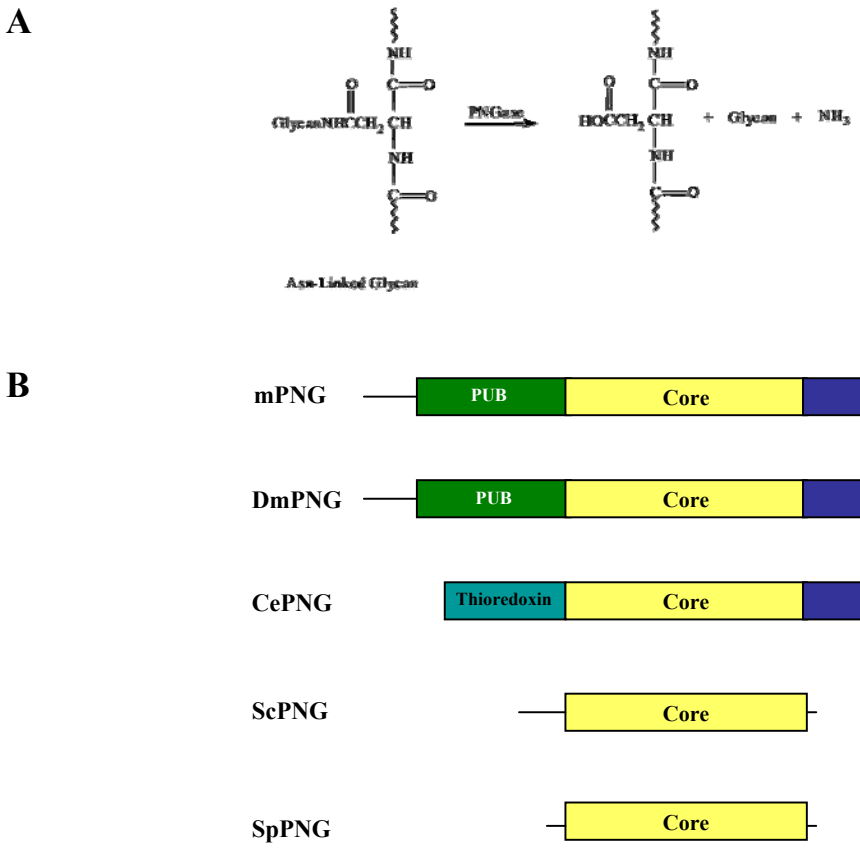


Figure 1.2 **PNGase reaction and domain distribution.** *(A)* Schematic representation of the reaction catalyzed by PNGase. This figure is taken from Katiyar et al. [24]. *(B)* Domain distribution of PNGase in different species. mPNG: mouse PNGase; Dm PNG: *Drosophila melanogaster* PNGase, CePNG: *Caenorhabditis elegans* PNGase; ScPNG: *Saccharomyces cerevisiae* PNGase; SpPNG: *Schizosaccharomyces pombe* PNGase. The conserved catalytic domain is colored in yellow and labels as core. The PUB domain is colored in green, Thiroredoxin domain in cyan and C-terminal domain of PNGase in blue. This figure was modified from Susuki et al. [23].

transglutaminase instead of PNGase activity [25]. Not surprisingly, the core domain of mPNGase and the yPNGase structure were found to adopt the transglutaminase fold [26, 27].

Mutations of the residues in the catalytic triad result in inactivation of yeast PNGase [24]. Initially PNGase was shown to only function on small synthetic glycopeptide substrates [28], however, more recent studies demonstrated that yeast (y) and mouse (m) PNGase recognize denatured, high-mannose type N-glycosylated proteins [18, 29, 30]. PNGase's *in vitro* substrates include full-length TCRα (T-cell receptor α chain), a mutant form of CPY (carboxypeptidase) referred to as CPY*, RNase B (ribonulease B) and class I MHC (major histocompatiblity complex). Very recently, Kim et al. identified the A-chain of ricin (a plant protein toxin) as the first *in vivo* substrate of yPNGase, which indicated that PNGase is very important in the degradation of at least selected substrates [31].

Both mouse and yeast PNGase have been reported to interact with HR23B and Rad23, respectively [32, 33]. HR23/Rad23 was initially discovered as being involved in DNA damage recognition in the nucleotide excision repair pathway [34], and in this context HR23/Rad23 forms a tight complex with the XPC/Rad4 protein, which is involved in DNA damage recognition. HR23 has a modular architecture with four domains, an N-terminal UBL domain, an internal UBA domain, a XPCB (xeroderma pigmentosum group C complementing protein binding) domain and a C-terminal UBA domain. The NMR structure of HR23A [35], a close homolog of HR23B, showed that these four domains are connected by long and intrinsically unstructured loops. The XPC binding domain of HR23/Rad23 is responsible for binding to PNGase and XPC/Rad4.

Recently, the structures of yeast and mouse PNGase in complex with Rad23/HR23B have been solved [26, 27]. The yeast protein corresponds to the central region of mPNGase, and in both proteins this structural entity is composed of three units, the catalytic transglutaminase-like domain, an accessory Zn-binding domain and an XPC-binding motif, which, however, is different between the yeast and mouse proteins. The catalytic domain forms a binding cleft, which is located at the interface between the catalytic and Zn-binding domains and is surrounded by highly conserved residues. The XPC-like motifs of yPNGase and mPNGase recognize the XPC binding domain of Rad23 and HR23B, respectively. Careful sequence comparisons between mPNGase and the XPC protein indicate that the core region of its catalytic domain has a statistically significant level of sequence identity with XPC, suggesting that the XPC/HR23 complex adopts a similar structure as its PNGase/HR23 counterpart.

The activity of PNGase can be inhibited by several compounds including the modified tripeptide, carbobenzyloxy-Val-Ala-Asp-α-fluoromethylketone (Z-VAD-fmk) [36], a known caspase inhibitor, and iodoacetamidyl-modified oligosaccharides (CHO-IAC) [37]. Although the latter two inhibitors are quite different in their chemical structures, they have one feature in common: they both covalently attach to the active site Cys residue, which otherwise carries out a nucleophilic attack on the β-N-glycosidic bond linking the Asn side chain and the first GlcNAc residue, thereby cleaving the glycan chain from the protein. However, Z-VAD-fmk mimics the peptide moiety of the substrate, whereas CHO-IAc mimics the carbohydrate component of the substrate.

Whereas yPNGase only contains the catalytic domain, in higher eukaryotes PNGase has both N- and C-terminal extensions (Figure 1.2*B*). The C-terminal extension

is conserved from *C. elegans* to human and is evolutionarily unique. The function of this domain was unknown until I solved the crystal structure of the mPNGase C-terminal domain and biochemically characterized this fragment. In *C. elegans*, the N-terminal domain of PNGase contains a thioredoxin domain, which function as an oxidoreductase [38, 39], whereas in higher eukaryotes, the N-terminal extension contains a conserved region, which has been termed the PUB domain (peptide:N-glycanase/UBA or UBX-containing protein domain). The PUB domain was suggested to be involved in protein-protein interactions [40], and in agreement with this hypothesis, our collaborators (Drs. Gangtao Li and William Lennarz) recently discovered that the PUB domain of mouse PNGase interacts directly with p97, the aforementioned AAA ATPase which plays a critical role in ERAD. Because of this interaction PNGase forms a super-complex with a variety of ERAD related proteins, including AMFR, Y33K and Derlin1.

C. p97 and ERAD:

p97, also known as CDC48 in yeast, belongs to the AAA ATPase family of proteins. It contributes about 1% of the cytosolic protein mass and is also referred to as valosin-containing protein (VCP). The majority of p97 is located in the cytosol and the rest either associates with various organelle membranes, or is present in the nucleus [41]. P97 is highly conserved in eukaryotes [42]. Structural studies [43] revealed that p97 consists of an N-terminal domain (N), two AAA ATPase domains (D1 and D2) and a short disordered C-terminal region (C) (Figure 1.3*A* and *B*) and forms very stable hexamers in solution.

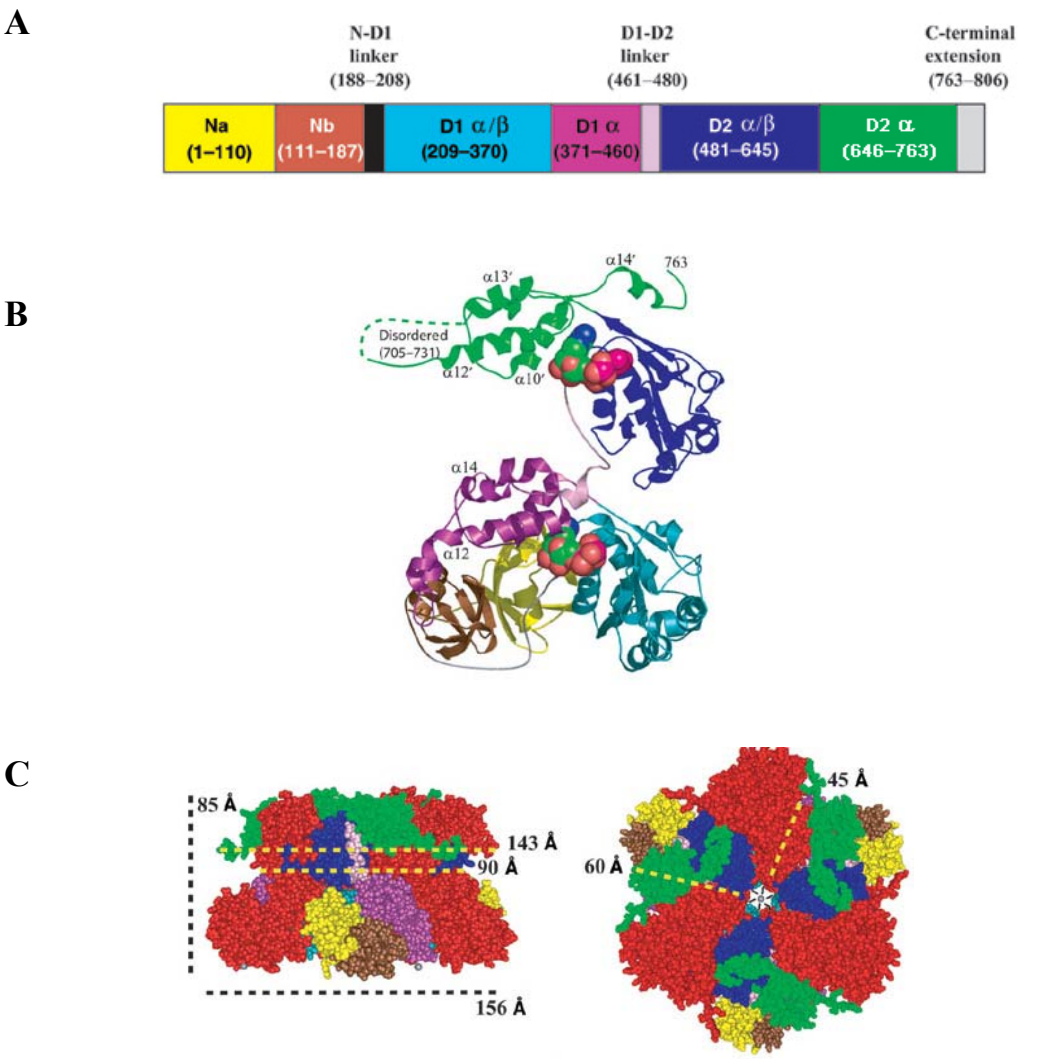


Figure 1.3 **Domain architecture and the three-dimensional structure of p97**. *(A)* Domain diagram of p97. *(B)* Structure of the p97 monomer. The color code is as in *(A)*. *(C)* Two orthogonal views of the p97 hexamer in surface representation, with subdomains of alternating monomers colored red or as in *(A)*. This figure was taken from DelaBarree and Brunger [43].

The hexameric p97 displays a short stemmed-mushroom shaped structure with a channel at the center (Figure 1.3*C*). It has been proposed that ERAD substrates are threaded mechanically through the retrotranslocon by p97 [44]. RNAi and mutagenesis experiments have confirmed the essential role of p97 in the proteasome-mediated degradation of misfolded proteins [45]. In addition, p97 has been shown to associate with E3 ligases, such as AMFR and $SCF^{Fbs1,2}$. These properties indicate that p97 plays an essential role in the ERAD pathway. However, studies also showed that the retrotranslocation of Δgpαf (a nonglycosylated derivative of the yeast pro-α factor mating pheromone) is independent of both p97 and ubiquitin; instead it only requires ATP and the 19 S regulatory particle of the 26S proteasome at the cytoplasmic site [12]. Since Δgpαf is neither glycosylated nor ubiquitinated, it is expected to behave differently from other ERAD substrates. Nevertheless, the function of p97 in the ERAD may not be universally required for all ERAD substrates.

In addition to its role in ERAD, p97 has been found to participate in other cellular processes, including DNA repair, cell cycle control and membrane trafficking [42, 44]. Structural and biochemical studies have suggested that p97 fulfills its various functions by interacting with a wide spectrum of cofactors [46-49]. Most of its cofactors interact with its N-terminal domain, with the exceptions of Ufd2, Ufd3 and peptide:N-glycanase (PNGase), which interact with the C-terminal region of p97 [21, 46, 49]. The cofactors can be divided into substrate-recruiting cofactors, which regulate the substrate specificity of p97, and substrate-processing cofactors, which determine the fate of substrates [49]. The substrate-recruiting cofactors include the Ufd1-Npl4 heterodimer and proteins containing the ubiquitin regulatory X (UBX) domain. Ufd1-Npl4 functions in ERAD

[47], nuclear envelope formation [50] and spindle disassembly after mitosis [51]. One of the best-characterized UBX-containing proteins is p47. The complex formed between p47 and p97 mediates the reassembly of the Golgi membrane and nuclear envelope following cell division [52, 53]. Compared to the extensively studied substrate-recruiting cofactors, the functions of the substrate-processing cofactors are less well defined. Recently, Rumpf and Jentsch [49] showed that, depending on whether p97 interacts with Ufd2 or with Ufd3, the p97 substrates are either polyubiquitinated and degraded, or deubiquitinated and stabilized.

D. Derlin1 and ERAD:

Mammalian Derlin1 (Derl-like protein 1) is an ER membrane protein that was first identified [11, 13] as a homolog of yeast Derl, a protein involved in the degradation of misfolded proteins [54]. It is widely expressed throughout different organs with the exception of heart and brain. Mammalian Derlin1 has 251 residues in total and is predicted to have four trans-membrane segments with both, the N- and C-terminal ends on the cytoplasmic side of the ER membrane [55]. The C-terminal cytosolic region of Derlin1 directly interacts with p97 [56]. Mammalian Derlin is a small family consisting of three homologs, including Derlin1, Derlin2 and Derlin3, whereas in yeast the Derlin homologs include Derl and Dfm1 (Figure 1.4). Unlike Derl, Dfm1 is not involved in ERAD [57]. Although Derlin1 is named after Derl, Dfm1 actually has a slightly higher homology with mammalian Derlin1 [57]. The most significant difference between Derlin1 and Dfm1 is in their C-terminus, where Dfm1 contains two copies of an eight residue long motif termed the SHP box within the context of a longer tail than that of

Derlin1. This element has been suggested to be required for the co-precipitation of Dfm1 by CDC48, the yeast homolog of p97 [57].

In order to be successfully degraded, misfolded proteins need to exit the ER lumen and enter the cytosol. This requires the proteins to travel across the ER membrane through a conduit, which is termed the "retrotranslocon". Studies on the nature of the retrotranslocon are rather controversial and, so far, three proteins have been proposed to form the retrotranslocon [11, 14]. The first is the Sec61 complex, which constitutes the translocon that associates with the ribosome and directs the nascent synthesized polypeptide to the ER lumen, or the ER membrane. In US2 gene expressing cells, deglycosylated MHC heavy chains were found complexed with the Sec 61 channel [14]. US2 together with US11, represent two glycoproteins that are encoded by the human cytomegalovirus. Either of these proteins will cause the dislocation and degradation of MHC heavy chains [58]. However, other studies showed that degradation of some of the aberrant protein substrates are Sec61 independent [59]. The second putative protein forming the protein conducting channel is the Derlin1 protein. The Derlin1/VIMP (VCP-interacting membrane protein) complex has been shown to recruit MHC heavy chains in US11 expressing cells [11]. A very recent study revealed that Derlin1, and not Sec61, photo-crosslinks with Δgpαf [12]. In agreement with this idea, Derlin forms both homodimers and heterodimers with the two other Derlin proteins, which can be observed in the ER membrane [56]. Very recently, HRD1, the yeast homolog of AMFR, was proposed as a third candidate for the retrotranslocon (Dr. Tom Rapoport, personal communication).

To add to the controversy, besides all three retrotranslocon candidates, Hidde Ploegh very recently proposed a new “lipid based model” [60], claiming that the retrotranslocating step may involve a lipid rearrangement rather than a dedicated membrane protein channel, since certain viruses escape the ER via lipid droplets. At the present time this lipid based model is purely hypothetical and more experimental data are needed to clarify this issue in the future.

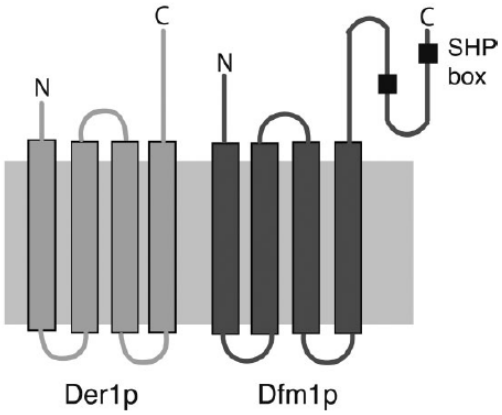


Figure 1.4 **Models of yeast Der1p and Dfm1p.** The Der1 and Dfm proteins contain four transmembrane spanning segments with both of their termini located in cytosol. The SHP boxes in Dfm1p are shown as black boxes. This figure is taken from Sato [57].

Chapter 2

Structural and Biochemical Studies of the C-terminal Domain of Mouse Peptide:N-Glycanase Identify it as a Mannose-binding Module

The major findings presented in this chapter have been presented in the publication entitled "Structural and biochemical studies of the C-terminal domain of mouse peptide-N-glycanase identify it as a mannose-binding module" by X. Zhou et al., Proc. Natl. Acad. Sci USA Vol. 103, pp. 17214-9 (2006). I have carried out all of the experiments described in this chapter except, as indicated, for the cloning of some of the constructs.

I. Introduction

PNGase has by now been recognized as one of the components of the ERAD machinery. It removes N-glycan chains from high mannose-type ERAD substrates after they have been transported from the ER lumen to the cytosol. PNGase is present in all eukaryotes and features a conserved transglutaminase-like catalytic domain. In yeast, PNGase primarily contains this catalytic domain with very short extensions at both ends. In higher eukaryotes including the fruit fly and mouse, PNGase features two extra domains besides the catalytic domain. The domain at the C-terminal end is present in *C. elegans* and higher organisms, suggesting it acquired an important biological function during the evolution of multicellular organisms. However, since no other homologs of this domain were found in the entire Gene Data Bank using blast searches against a non-redundant (nr) set of protein sequences, the function of this domain could not be established prior to the biochemical and structural studies of this domain presented in this thesis.

An early study [61] revealed that mPNGase binds to free glycan chains derived from its glycoprotein substrates, and that this binding inhibits the activity of PNGase, thus suggesting that mPNGase has a carbohydrate-binding activity. Moreover, recent studies revealed that PNGase specifically acts on the unfolded form of high mannose type N-glycoproteins [18, 29, 30]. However, how PNGase binds to glycan chains and how it recognizes the high mannose-type substrates was not understood.

In this chapter, I present the structure of the C-terminal domain of mPNGase, which is present in higher eukaryotes ranging from *C. elegans* to human with 30% sequence identity between these two species. Biochemical, biophysical and

crystallographic studies reveal that it contains a mannose-binding domain, which presumably contributes to the oligosaccharide-binding specificity of mPNGase. This suggests that the C-terminal domain increases the binding affinity between mPNGase and its substrates.

Since yPNGase, in contrast to mPNGase, only contains the catalytic domain and neither the C-terminal domain nor the PUB domain at its N-terminus, it was expected to have a similar activity as the mPNGase core domain. However, our activity assays presented in this chapter revealed that yPNGase is much more active than the mPNGase core domain and has an activity that is comparable to the mPNGase full-length protein. In addition to the conserved region of the catalytic domain there are about 30 residues at its N-terminal end, which participate in Rad23 binding and another 36 residues at its C-terminus, These latter residues are disordered in all crystal structures of yPNGase, but additional activity assays demonstrate that these C-terminal amino acids of yPNGase are very important for efficient substrate turn-over.

II. Materials and Methods

A. Constructs

The C-terminal domain of mPNGase (residues 451-651) was amplified from the gene encoding full-length mPNGase by PCR and was cloned into either the NdeI/XhoI sites of the pET28a (His_6-mPNGc) (Novagen, San Diego, CA) or the NdeI/SapI sites of the pTXB1 (Int-mPNGc) vectors (New England Biolabs, Beverly, MA). The oligonucleotide primers utilized during the various cloning procedures and site directed mutagenesis studies are summarized in Table 2.1. Another His_6-tagged construct, His_6-mPNGc2, encompassing residues 471-651 was specifically made to facilitate cocrystallization with mannopentaose. All mutations were introduced with the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) using His_6-mPNGc as the template. The construct encoding full-length yPNGase (pET28a, NcoI/BamHI, His_6-tag) was a gift from the group of Dr. Lennarz. The truncated yPNGase construct encoding residues 8-341 (pET21b, NdeI/XhoI, no tag) and yRad23 XPCB domain construct (residues 253-309, pTXB1, intein-tag) were kindly provided by a postdoctoral fellow in the laboratory, Dr. Gang Zhao. Both yPNGase and the truncated protein were cloned into the pET-28a vector and expressed under the same conditions as His_6-mPNGc. Full-length mPNGase (residues 1-651) and C-terminally truncated mPNGase (residues 1-450) were cloned into the NdeI/XhoI site of the pET28a vector, and the mPNGase core domain (residues 165-450) was cloned into the NdeI/XhoI sites of the pET21b vector (Novagen, San Diego, CA). A construct encoding a fusion protein consisting of the full length yPNGase (residues 1-363) and the mPNGase C-terminal domain (residues 451-650) was amplified using a two-step PCR. In the first step two separate PCR products of

Table 2.1 **Primers sequences**

Primers		Sequence
mPNGase451-651 /PTXB1	5’ 3’	GGAATTCCATATGGAACTTGGTGGAAGGGTATC CTTGCTCTTCCGCAGAGGTCATTGAACGT
mPNGase451-651 /pET28a	5’ 3’	GGAATTCCATATGGAACTTGGTGGAAGGGTATC CCGCTCGAGTCAGAGGTCATTGAACGT
mPNGase471-651 /pET21b	5’ 3’	CCGAGCATATGCTAGAGAGA AAGGAG CCGCTCGAGTCAGAGGTCATTGAACGT
mPNGase1-450 /pET28a	5’ 3’	GGAATTCCATATGGCGTCGGCCACACTGGGC CCGCTCGAGTCATCCAGGTCTTGGGGTTTTAG
mPNGase165-450 /pET21b	5’ 3’	GGGCCCTTCATATGGATTCAACCATCTTA CCGCTCGAGTCCAGGTCTTGGGGTTTTAG
mPNGase1-651 /pET28a	5’ 3’	GGAATTCCATATGGCGTCGGCCACACTGGGC CCGCTCGAGTCAGAGGTCATTGAACGT
yPNGase+mPNGc /pET28a	5’ 3’	GCACAGCGTGGGGAGGATGGTAAAGAACTTGGTGGAAGG GTATCC GGATACCCTTCCACCAAGTTCTTTACCATCCTCCCCACGCT GTGC
mPNGc F525A /pET28a	5’ 3’	GGAAAATGGAATCCATAGCCAGAAAAGTCGAGAAAG CTTTCTCGACTTTTCTGGCTATGGATTCCATTTTCC
mPNGcK527A /pET28a	5’ 3’	GAATCCATATTCAGAGCAGTCGAGAAAGACTGG CCAGTCTTTCTCGACTGCTCTGAATATGGATTC
mPNGc E529A /pET28a	5’ 3’	CATATTCAGAAAAGTCGCGAAAGACTGGAACATG CATGTTCCAGTCTTTCGCGACTTTTCTGAATATG
mPNGcK530A /pET28a	5’ 3’	TTCAGAAAAGTCGAGGCAGACTGGAACATGGTT AACCATGTTCCAGTCTGCCTCGACTTTTCTGAA
mPNGcW532A /pET28a	5’ 3’	GAAAAGTCGAGAAAGACGCGAACATGGTTTATTTG CAAATAAACCATGTTCGCGTCTTTCTCGACTTTTC
mPNGcY536A /pET28a	5’ 3’	GACTGGAACATGGTTGCTTTGGCCCGAAAAGAAG CTTCTTTTCGGGCCAAAGCAACCATGTTCCAGTC
mPNGcK540A /pET28a	5’ 3’	GTTTATTTGGCCCGAGCAGAAGGATCATCTTTTG CAAAAGATGATCCTTCTGCTCGGGCCAAATAAAC
mPNGc E541A /pET28a	5’ 3’	TATTTGGCCCGAAAAGCAGGATCATCTTTTGCT AGCAAAAGATGATCCTGCTTTTCGGGCCAAATA
mPNGcQ625A /pET28a	5’ 3’	GAGATGGAGACGTTGCTTGGGCACATACCCAACTGTTTAGA TCTAAACAGTTGGGTATGTGCCCAAGCAACGTCTCCATCTC
mPNGcW624A /pET28a	5’ 3’	GATGGAGACGTTGCTGCGCAACATACCCAACTG CAGTTGGGTATGTTGCGCAGCAACGTCTCCATC
mPNGc F630A /pET28a	5’ 3’	CTTGGCAACATACCCAACTGGCTAGACAAAGCTTAAATGAC GTCATTTAAGCTTTGTCTAGCCAGTTGGGTATGTTGCCAAG
mPNGcQ628A /pET28a	5’ 3’	GCTTGGCAACATACCGCACTGTTTAGACAAAGC GCTTTGTCTAAACAGTGCGGTATGTTGCCAAGC
mPNGcR631A /pET28a	5’ 3’	GGCAACATACCCAACTGTTGCACAAAGCTTAAATGAC GTCATTTAAGCTTTGTGCAAACAGTTGGGTATGTTGCC

yPNGase (5' primer: T7 promoter, 3'primer: listed in Table 2.1) and mPNGc (5'primer: listed in Table 2.1, 3'primer: T7 reverse) were generated. These two PCR products were purified using a PCR product purification kit (Roche). In the second step, these two PCR products were used as templates, with the T7 promoter and T7 reverse as primers, and PfuTurbo as DNA polymerase (Stratagene). The product was generated using a standard PCR procedure. The resulting product was gel purified and cloned into the NcoI/XhoI sites of the pET28a vector. This construct does not contain any tag.

B. Protein expression and purification

The transformation and expression steps were very similar in each case. The plasmid of interest was transformed by heat shock [62] into BL21DE3 Codon Plus RIL cells selected on LB plates containing 100 µg/ml Ampicillin (50 µg/ml Kanamycin in case of the pET28a vector), and 34 µg/ml Chloramphenicol, except for Int-mPNGc, which was transformed into BL21DE3 cells selected on LB plates containing 100 µg/m Ampicillin. A single colony grown was picked and used to inoculate 10-80 ml LB culture containing the same concentration of antibiotics for overnight growth at 37 ºC. 1% of the overnight culture was usually used for large scale (1-8 L) cell growth. His_6-mPNGc wild-type and variants were expressed in BL21DE3 Codon Plus RIL cells at 30 ºC for 4 hours after induction with 1 mM IPTG at an OD_{600} of 0.6 (2 L total volume). Int-mPNGc was expressed in BL21DE3 cells overnight at 15 ºC after induction at the same A_{600} as described for His_6-mPNGc (8 L total volume). All other proteins were expressed in BL21DE3 Codon Plus RIL cells overnight at 15 ºC after induction at the same A_{600} (1 L total volume each, except XPCB with 4 L).

After growth, the cells were harvested by centrifugation at 9000 g for 15 minutes at 4 ºC and lysed in 20 mM Tris/HCl, pH 8.5, 150 mM NaCl, a protease inhitor cocktail (Complete mini, EDTA free, Roche) by passing twice through a French pressure cell at 1,500 psi. The cell lysate was then centrifuged at 50000 g for 30 minutes. The cleared cell lysate (supernatant) was collected and incubated with the corresponding affinity-matrix coated beads. Intein-tagged mPNGc and yeast XPCB were incubated with 5-10 ml of chitin beads (New England Biolabs, Ipswich, MA) already packed in a 100 ml glass column by end over shaking for 30 minutes at 4ºC and washed with 20 column volumes (cv) of buffer containing 20 mM Tris/HCl, pH 8.5, 1 M NaCl, followed by equilibration with cleavage buffer (wash buffer plus 50 mM DTT). Beads were incubated for overnight at room temperature prior to elution with 20 mM Tris/HCl, pH 8.5, 150 mM NaCl.

The XPCB-bound protein complexes were prepared by incubating the cell lysates of either yPNGase, yPNGase8-341 and the yPNGase/mPNGc fusion with XPCB immobilized on chitin beads for 30 minutes prior to the addition of cleavage buffer to elute the XPCB-protein complexes. The cleared cell lysate of the His_6-tagged proteins (mPNGase fragments, yPNGase and mPNGc mutants) was incubated with a Ni-NTA affinity column (Qiagen, Valencia, CA) for 30 minutes in binding buffer consisting of 20 mM Tris/HCl, 150 mM NaCl, 10 mM imidazole and washed with 10 cv of binding buffer. Bound proteins were eluted with 20 mM Tris/HCl, 150 mM NaCl, 250 mM imidazole. All protein purifications followed the same two steps including MonoQ anion exchange (5/5) and Superdex 200 (10/30 for mPNGc mutants, 26/60 for all others) size exclusion chromatography (Amersham, Piscataway, NJ). Prior to the MonoQ column all proteins were diluted to lower the salt concentration to values below 50 mM NaCl in the

respective protein buffer (20 mM Tris-HCl pH 8.5, 40-50 mM NaCl, 1-5 mM DTT from the cleavage buffer) in order to allow binding to the column with the aid of a 50 ml superloop at a flow rate of 0.4-0.7 ml/min. The proteins of interest were eluted with a linear NaCl gradient (0-1 M) at a flow rate of 0.4-0.7 ml/min and examined by SDS-PAGE (12 or 15% gels) and stained with Coomassie Brilliant Blue R-250 (Fisher). Protein fractions of highest purity were pooled and concentrated to less than 5 ml by centrifuging 3220 g using centricon Plus-20 (MWCO of 30,000 for full-length mPNGase and yPNGase/mPNGc-XPCB complex, all others using MWCO of 10,000. After the concentration step they were loaded onto the size exclusion chromatography column, which was equilibrated with 20 mM Tris-HCl, pH8.5, 150 mM NaCl and 5 mM DTT. Size exclusion chromatography was carried out at a flow rate of 1 ml/min, fractions were analyzed by SDS page and pooled according to purity. Finally the proteins were concentrated to 5-15 mg/ml using calculated molar extinction coefficient of 46230 $M^{-1}cm^{-1}$ for mPNGase 451-651, 40450 $M^{-1}cm^{-1}$ for mPNGase 471-651, 66250 $M^{-1}cm^{-1}$ for mPNGase ΔN truncation, 63570 $M^{-1}cm^{-1}$ for mPNGase 165-450, 112480 $M^{-1}cm^{-1}$ for mPNGase, 70205 $M^{-1}cm^{-1}$ for yPNGase, 71695 $M^{-1}cm^{-1}$ for yPNGase/XPCB, 64705 $M^{-1}cm^{-1}$ for yPNGase8-341/XPCB, 117645 $M^{-1}cm^{-1}$ for yPNGase-mPNGc/XPCB complex before they were stored in -80° C freezer.

C. Circular dichroism analysis

The secondary structure of the mPNGase C-terminal domain and possible conformational changes introduced by the various mutants were analyzed with an Aviv Model 62A DS circular dichroism (CD) spectrometer in the wavelength range from 200-

260 nm. These measurements were performed at 18 ºC. All samples were prepared by dialysis against a buffer containing 4 mM NaH_2PO_4, pH 8.0 and 60 mM NaCl overnight, at protein concentrations ranging from 13-20 mg/ml.

D. Crystallization and structure determination

Intein-mPNGc crystals were grown at 18° C using purified protein at a concentration of 15 mg/ml using the hanging drop vapor diffusion method in 24 well Limbro plates against a reservoir solution containing 14-16% PEG 8000, 0.1 M Tris-HCl, pH 8.5, 0.12 M $MgCl_2$ and 10 mM DTT. The hanging drop is prepared by adding 1 µl of protein complex and 1 µl of reservoir solution. Larger and better diffracting crystals were obtained using macroseeding techniques. Plate-like crystals from the original well were looped out of the drop and washed twice with freshly prepared reservoir solution. Macroseeds were obtained by breaking the initial crystals into several still visible pieces to generate new growth edges and surfaces. Each of the resulting small crystals was transferred into one hanging drop that was pre-equilibrated for 3 days with 11-13% PEG 8000, 0.1 M Tris-HCl, pH 8.5, 0.12 M $MgCl_2$ and 10 mM DTT, resulting in the formation of larger crystals from these macroseeds. The heavy atom derivative was prepared by soaking crystals with mother liquor supplemented with 10 mM sodium ethylmercurithiosalicylate (EMTS) for 10 minutes at room temperature.

His_6-mPNGc crystals were obtained by handing drop vapor diffusion at 18° C against a reservoir solution containing 1.6 M Li_2SO_4 and 0.1 M Hepes, pH 7. The mPNGc-mannopentaose complex was prepared by incubating a 15 mg/ml (0.7 mM) solution of mPNGc (471-651) with 2 mM α3,α6-mannopentaose (Sigma) for 30 minutes

on ice. The crystals of the complex were grown with 19% PEG 4000, 11.5% isopropanol, 0.1 M Tris-HCl, pH 7.5 and 0.2 M Ca-acetate. The crystals were flash frozen in liquid nitrogen. Diffraction data of the apo-structures were collected on beam line X26C (wavelength: 1.1 Å for intein-tagged mPNGc, 1.0 Å for His_6-tagged mPNGc) and that of the complex structure on beam line X12C (wavelength: 0.9795 Å) of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory at 100º K. Diffraction data were indexed, integrated and scaled with HKL2000 (26).

The structure of Int-mPNGc was determined by the single isomorphous replacement and anomalous scattering (SIRAS) method. One major Hg-site was identified by SHELXD [63] and a minor site by difference Fourier methods. Phase refinement was carried out with Sharp [64] to 3.3 Å, followed by solvent flattening with Solomon [65] at the same resolution. A preliminary model consisting of ~80% of the residues without side chains was built with the aid of the program O [66], and starting from this model ARP/Warp [67] was able to build 180 out of 201 residues starting with the solvent-flattened phases and amplitudes extending to 1.9 Å resolution. The protein model was completed manually in O and was refined with Refmac5 [68]. Water molecules were added automatically with ARP/Warp and additional solvent molecules manually. The His_6-mPNGc and complex structures were solved by molecular replacement using Molrep [69] and refined with Refmac5. Solvent molecules were added as described above.

E. ITC experiments

ITC measurements were carried out using a VP-ITC microcalorimeter (Northampton, MA). Prior to the titration the proteins were dialyzed overnight at 4 ºC against a buffer containing 10 mM Tris-HCl, pH 8.5, and 0.15 M NaCl, followed by degassing for 30 minutes. The ligands were dissolved in the same buffer to minimize the heat of buffer dilution. Proteins at concentrations ranging from 15 to 30 μM were titrated with 0.6-1.2 mM of either mannopentaose, mannotriose or chitobiose at a time with a total number of 30 injections each containing 10 μl solution at 18 ºC. A 300 second waiting period was set between each titration to allow the baseline to stabilize. The binding parameters were calculated using Origin (version 7.0) by fitting the data to a single site binding model.

F. Activity assay

The reaction mixture consisting of either yPNGase or mPNGase fragments and RNase B as substrate was prepared at room temperature in a buffer containing 20 mM Tris, pH 8.5, 250 mM NaCl and 5 mm DTT. All proteins utilized in each assay were purified at the same time and following the same protocol. The RNase B substrate (Sigma, St. Louis, MO), at a concentration of 5μg/μl, was denatured by incubation at 95 ºC for 15 minutes prior to the start of the assay. 10 μg of RNase B were used in each reaction. The enzyme to substrate molar ratio was 1:200 except that in the activity assay of the yPNGase and mPNGase core domains a ratio of 1:40 was used. The reaction was stopped by the addition of SDS sample buffer and heating at 95 ºC for 10 minutes. The resulting Coomassie-stained gels were quantitated by densitometry with the ImageJ program and a HP scanjet 5590 scanner.

G. Coordinates

The atomic coordinates and structure factors (accession codes: 2G9F for intein-tagged mPNGc, 2G9G for His-tagged mPNGc and 2I74 for the complex) have been deposited in the Protein Data Bank, Research Collaboratory for Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

III. Results

A. Protein purification

All of the His_6-tagged mPNGase and yPNGase proteins display very good expression levels under the conditions described in section II. The yield is about 5-10 mg per liter culture and they were purified to more than 90% purity based on SDS-PAGE analysis. Intein-tagged proteins usually show lower expression levels than His_6-tagged proteins, about 2 mg per liter culture. The purity is also about 90 % except for two protein complexes: yPNGase/mPNGc-XPCB and yPNGase (8-341)-XPCB, with a purity of only about 60% in both cases. The proteins used for crystallization were usually concentrated to 10-20 mg/ml prior to flash freezing small drops (15 μl volume) in liquid nitrogen and storage at -80 °C.

B. Overall structure of mPNGc

Due to difficulties in obtaining large single crystals, mPNGc (residues 451-651) was expressed as an intein and a His-tagged fusion protein. Both purified proteins yielded crystals that diffracted to about 2 Å, however, the space groups differed ($P3_221$ and C2, respectively). The crystals of both proteins have plate-like shapes with the longest dimension about 400-500 μm, but a thickness of only 10-20 μm. mPNGc crystals were obtained after about two weeks of growth, whereas that of the His-tagged protein were obtained after about three months of growth. In the latter case, production of high-quality crystals could not be repeated. A packing analysis indicates that there is one molecule in the asymmetric unit of each crystal with Matthew's coefficient of 2.02 $Å^3$/Da and 1.94 $Å^3$/Da, corresponding to relative low solvent contents of 39 % and 36 %, respectively.

The structure of intein-tagged mPNGc was solved by single isomorphous replacement and anomalous scattering with the aid of a Hg-derivative (Table 2.2) and was refined at 1.9 Å resolution (Table 2.3) to an R-factor of 0.167 (R_{free}=0.217) with TLS refinement (ten bodies selected by the TLSMD server [70]) after the addition of 98 water molecules, one Cl^- ion and two glycerol molecules. The N-terminal residues 451-471 are disordered in this structure. Subsequently, the structure of His-tagged mPNGc was refined at 2 Å resolution (Table 2.3) to an R-factor of 0.15 (R_{free}=0.207) including TLS refinement (one body) after the addition of 120 solvent molecules. Both of the refined models display very good stereochemistries with 100 % for intein-tagged mPNGc and 99.5 % for His-tagged mPNGc in the allowed regions of Ramachandran statistics.

Both structures are very similar as reflected in a root mean square (rms) deviation in the Cα positions of 0.24 Å (Figure 2.1). However, in the His-tagged protein model residues 454-463 could be visualized, thus leaving only residues 451-453 and 464 to 472 as unassigned in the electron density maps. Due to the presence of these additional residues, the discussion in this chapter focuses on the His-tagged model, which is shown in Figure 2.2.

mPNGc is a slightly elongated molecule and displays a β-sandwich architecture, which is composed of two layers containing nine and eight antiparallel β-strands, respectively, and three additional short helices (Figure 2.2). One of the layers, which will be referred to as the "front" layer based on the orientation depicted in Fig. 2.2, deviates strongly from a standard β-sheet and consists of an antiparallel six-stranded β-sheet (β-strands 4-6, 11, 14 and 17) including a very long strand (β11) which is also involved in the formation of a second four-stranded antiparallel β-sheet (β-strands 8, 9, 11 and 16).

Table 2.2 **Data collection and phasing statistics**

	Native-His	**Native-Intein**	**Hg**	**Complex**
Space group	C2	$P3_221$	$P3_221$	$P2_1$
Cell dimensions (Å and °)	71.96, 40.80, 66.42 90, 95.34, 90	40.76, 40.76, 193.66 90, 90, 120	40.57, 40.57, 193.51 90, 90, 120	45.71, 41,65, 94.90 90, 94.89, 90
Resolution limits (Å)	50 – 2.0	50 – 1.9	50 – 3.3	50 – 1.75
R_{sym}	0.108 (0.485)	0.091 (0.622)	0.187 (0.424)	0.087 (0.441)
Mean Redundancy	4.6	9.4	8.7	5.5
Completeness	0.999 (0.994)	0.999 (0.986)	0.999 (0.995)	0.999 (1.0)
<I/sigI>	17.8 (3.9)	36.4 (3.5)	10.1 (2.9)	20.0 (3.3)
Heavy atom sites	-	-	2	-
Phasing Power	-	-	1.38	-
Cullis R Factor	-	-	0.695	-
FOM	-	-	0.435	-

$R_{sym} = \Sigma_{hkl}\Sigma_i |I_i - \langle I \rangle| / \Sigma_i \langle I \rangle$ where I_i is the i^{th} measurement and <I> is the weighted mean of all measurements of I. <I/sigI> indicates the average of the intensity divided by its standard deviation. Numbers in parentheses refer to the respective highest resolution data shell in each dataset. SIRAS phasing was performed to 3.3 Å resolution. Phasing power is the mean value of the heavy atom structure factor amplitude divided by the lack of closure for isomorphous / anomalous differences. R_{Cullis} is the lack of closure divided by the absolute of the difference between F_{PH} and F_P for isomorphous differences of centric data. FOM is the figure of merit given for acentric / centric data.

Table 2.3 **Refinement statistics**

	Native-His	Native-Intein	Complex
Resolution limits (Å)	20–2.0	20-1.9	20-1.75
Number of reflections	12,453	14,827	33,995
Number of protein/ solvent atoms	1,543 / 142	1,475 / 111	2,921 / 453
R (R_{free})	0.150 (0.207)	0.167 (0.213)	0.174 (208)
Deviations from ideality			
Bond distances (Å)	0.014	0.017	0.015
Bond angles (°)	1.495	1.413	1.543
Chiral volumes (Å^3)	0.096	0.094	0.125
Planar groups (Å)	0.006	0.007	0.007
Torsion angles (°)	7.12, 32.44, 15.34	6.95, 36.77, 13.56	6.59, 36.55, 12.53
Ramachandran statistics	0.951/0.995	0.961/1.0	0.966/0.997
Average B factor (Å^2)	32.8	40.1	18.0

$R_{cryst}=\Sigma||F_o|-|F_c||/\Sigma|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes. R_{free} same as R_{cryst} for 5% of the data randomly omitted from refinement. Ramachandran statistics indicate the fraction of residues in the favored and allowed regions of the Ramachandran diagram as defined by Molprobity [71].

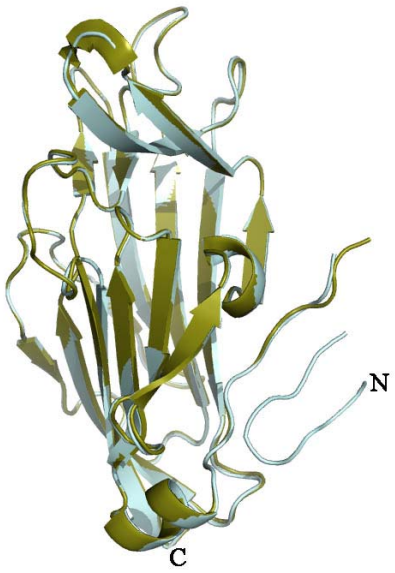


Figure 2.1 **Superposition of the two models of the mPNGase C-terminal domain.** Ribbon representation of the two mPNGc strucutures with His-tagged mPNGc model in cyan and intein-tagged mPNGc model olive.

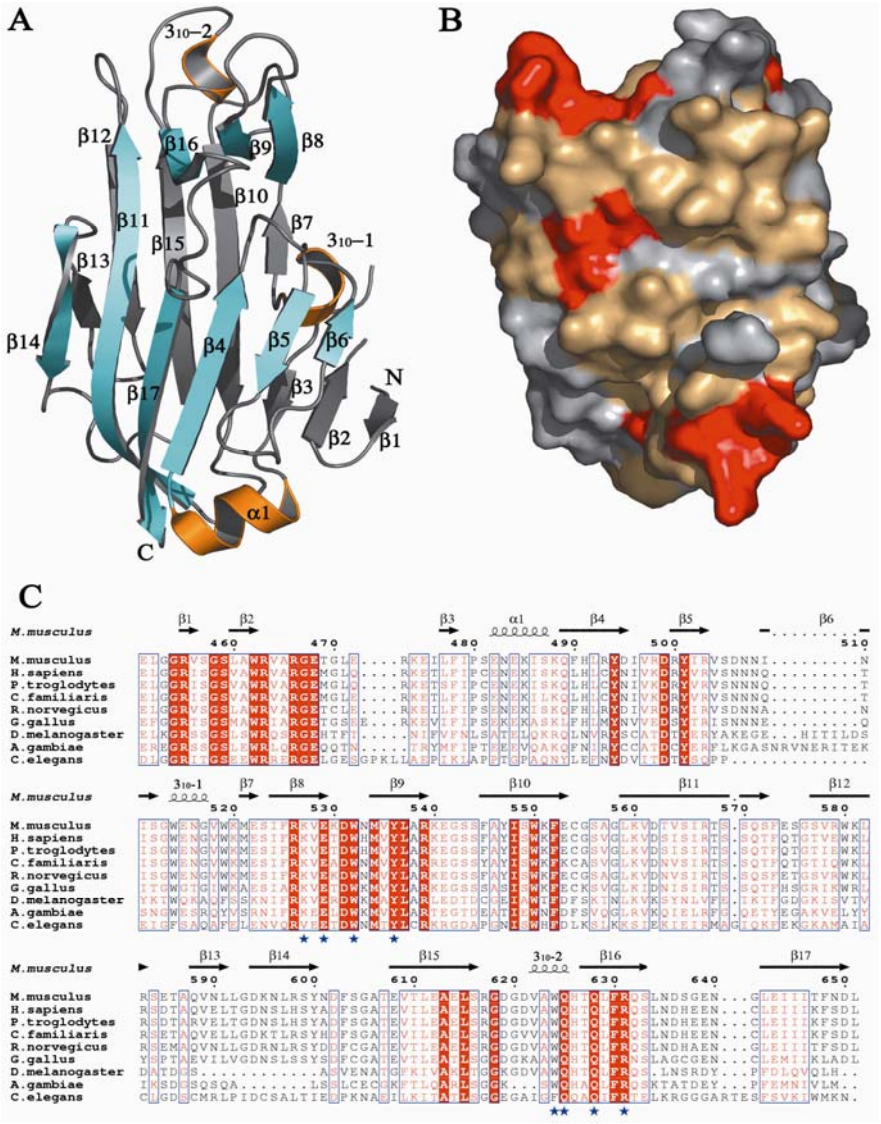
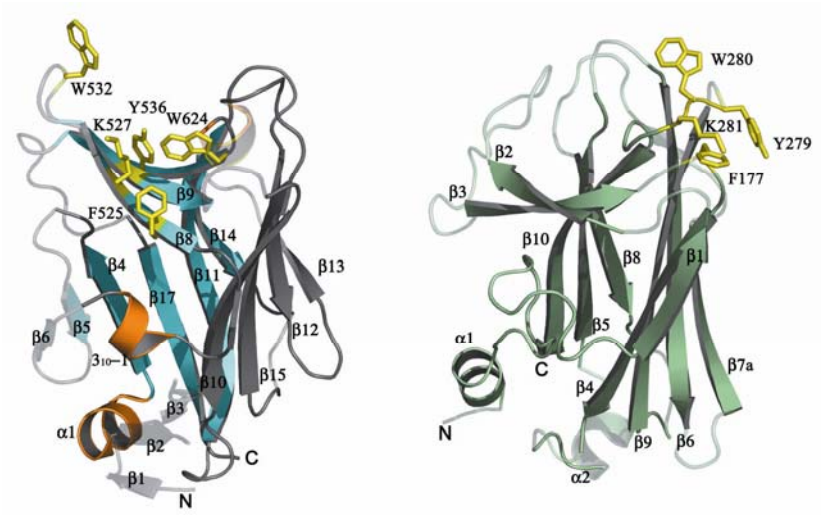


Figure 2.2 **mPNGc structure and multiple sequence alignment.** (*A*) Ribbon representation of the mPNGc structure. The front layer of the β-sandwich is colored in cyan, while the back sheet and loops are in gray with helices in orange. β-strands have been labelled. Figs. 2.1, 2.2*AB*, 2.3, 2.5*C*, 2.8 and 2.9 have been generated with Pymol (34). *(B)* Sequence conservation of mPNGc in the context of its three-dimensional structure. Strictly conserved residues of mPNGc have been mapped in red and conserved residues in light-orange onto a surface representation of the molecule. (*C*) Multiple sequence alignment of PNGase C-terminal domains (*H. sapiens*: human; *P. troglodytes*: chimpanzee; *M. musculus*: mouse; *R. norvegicus*: rat*; G. gallus*: chicken; *D. melanogaster*: fruit fly; *A. gambiae*: mosquito; *C. elegans*: nematode*)*. Strictly conserved residues are displayed in white on a red background and type-conserved residues are shown in red. Secondary structure elements and residues numbers refer to mPNGc. Blue stars: Ala substitutions of these residues abolish mannopentaose binding.

This sheet is rotated by ~90º relative to the six-stranded β-sheet and is located at one end of the molecule. β-Strands 16 and 17, which reside in the 4-stranded and 6-stranded β-sheets, respectively, are separated by a nine-residue long loop, while the loops connecting β4 and β5 in the 6-stranded sheet as well as β8 and β9 in the 4-stranded sheet completely disrupt this front layer. The "back" layer (Figure 2.2*A)* displays a more traditional architecture with a 5-stranded antiparallel β-sheet (β-strands 7, 10, 12, 13 and 15) containing a short edge strand (β7), which leaves sufficient room to allow a short 3-stranded β-sheet (β1-3) to also hydrogen bond in a parallel fashion with β10. The three helices are distributed throughout the structure with the longest helix (α1) at one end of the molecule, which will be referred to as the proximal end since it is adjacent to the N- and C-termini, and the shortest helix (3_{10}-2) at the distal end of the molecule. The first 3_{10} helix (3_{10}-1) links β-strands 6 and 7 and is located between the proximal and distal ends.

An analysis of the degree of sequence conservation in the context of the three-dimensional structure of mPNGc reveals that a depression between two loop regions and the adjacent β-strands (β8 and β9 as well as the β15 and β16 junctions) is one of the two most highly conserved regions (Figure 2.2*B,C* and Figure 2.3). The residues decorating the saddle-shaped depression at the distal end include two tryptophans, Trp532 and Trp624, which sit on opposite sides on the ridges flanking the saddle and are separated by 11 Å. Three additional residues, Tyr536, Phe525 and Lys527, are located on the concave side of the saddle.

A



B

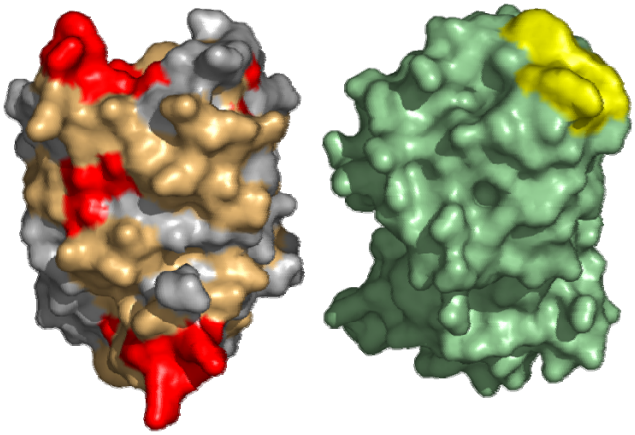


Figure 2.3 **Structural comparison of mPNGc and the SBD of Fbs1.** *(A)* mPNGc (left) is color coded as in Fig. 1*A*, and the SBD (right) of Fbs1 is in green. Conserved secondary structure elements are shown in solid colors, while non-conserved elements are rendered transparent. The side chains of highly conserved residues of mPNGc (left) that cluster around the concave surface and the chitobiose binding site of the SBD domain (right) are shown in stick representation. *(B)* Sequence conservation of mPNGc (left) in the context of its three-dimensional structure compared to the chitobiose binding site of the SBD (right). Strictly conserved residues in mPNGc are shown in red and conserved residues in light-orange. The chitobiose binding site of SBD is colored in yellow.

C. The mPNGc structure is similar to that of the sugar-binding domain of Fbs1

The mPNGc structure was compared to a non-redundant set of proteins from the Protein Data Bank using the Dali server [72]. Dali identified 34 proteins as close structural homologs with Z-scores of at least 5. Most of these structures were either carbohydrate binding proteins or hydrolases involved in carbohydrate degradation. The best match with a Z-score of 9.1 was the sugar-binding domain (SBD) of Fbs1 [73]. Fbs1 is an F-box protein, a component of one of many SCF E3 ubiquitin ligases, which are generally composed of the Skp1, Cul1, Roc1/Rbx1 and F-box proteins [73]. The F-box proteins represent the substrate-binding components of each E3 complex, and in the case of Fbs1 it was shown to recognize N-linked glycans, especially the chitobiose core via its SBD [74, 75]. The next best match with a Z-score of 7 was the CUB2 domain of the mannose-binding protein associated serine protease [76], followed by several other carbohydrate binding domain structures with similar Z-scores.

Although mPNGc shares only 10% sequence identity with the SBD of Fbs1, the two proteins are very similar in their tertiary structures and can be superimposed with an rms deviation of 3.7 Å for 128 aligned residues out of 184 present in Fbs1 (Figure 2.3). Like mPNGc, the SBD of Fbs1 features a β-sandwich architecture and its "front" sheet also displays a strong curvature. The chitobiose-binding region of the SBD has been deduced from cocrystal structures with either chitobiose [73] or the N-linked glycan of RNase B [77] and mapped to two loop regions, which connect the two β-sheets at the distal end of the elongated molecule (Figure 2.3). The surrounding loops in this area adopt dissimilar conformations between mPNGc and the SBD of Fbs1, which result in completely different surface models of the two proteins (Figure 2.3*B*).

The structural similarity with the SBD of Fbs1 suggested that mPNGc is also involved in carbohydrate binding with the saddle-shaped depression being the most likely binding site based on sequence conservation. The types of residues located in this putative binding pocket, including Trp, Tyr and Phe, are entirely consistent with those commonly observed in carbohydrate binding, which often include aromatic residues involved in van der Waals interactions in which a hexose and an aromatic ring system are arranged face-to-face [78, 79]. In addition, polar residues are found which could engage in hydrogen bonds with suitable carbohydrate acceptors and donors. Within this region of the protein, the polar residues Glu529, Tyr536, Gln625, Gln628 and Arg631 together with several water molecules form a hydrogen bonded network. Some of the conserved residues are presumably important for the structural integrity of mPNGc including the hydrophobic residues Ile524, Val535, Leu537 and Phe630, which are located in the hydrophobic core of the protein, and Gly618 in the β-turn connecting β15 and the second 3_{10} helix.

In addition to the structural similarities between the SBD and mPNGc of Fbs1, these two proteins also share common biochemical properties. Recently, Yoshida et al. [80] reported that Fbs1 interacts with N-glycoproteins, especially denatured glycoproteins, in agreement with the fact that PNGase acts on misfolded N-glycosylated proteins [29, 30, 80]. In the same study, SCF^{Fbs1} was shown to co-immunoprecipitate with p97, a protein that also interacts with mouse PNGase [81]. As discussed in the introduction, p97 is an AAA ATPase which functions as an extractor for misfolded proteins from the ER to the cytosol [82, 83]. The fact that both SCF^{Fbs1} and PNGase interact with p97 suggests that p97 may form a platform for SCF^{Fbs1} and PNGase, which

may sequentially bind to the same N-glycoprotein immediately following its extraction from the ER via the retro-translocon. In this model, the denatured glycoprotein is immediately recognized by SCF^{Fbs1} and ubiquitinated by the SCF E3 ligase, once it emerges from the ER lumen, and then forwarded to PNGase for de-glycosylation, suggesting that the ERAD pathway is highly cooperative and more processive than previously realized.

D. mPNGc binds to oligomannose carbohydrates

To investigate whether mPNGc indeed binds to carbohydrates, isothermal titration calorimetry (ITC) experiments were performed with three oligosaccharides, α3, α6-mannopentaose, α3, α6-mannotriose and N, N'-diacetyl-chitobiose (Figure 2.4), as possible ligands, all of which are part of the N-linked glycan chain (Figure 2.5). mPNGc was shown to bind to mannopentaose with a dissociation constant (K_d) of ~67 μM. Full-length mPNGase was found to have a very similar binding affinity (K_d ~72 μM) as the C-terminal fragment, whereas the mPNGase fragment without the C-terminal domain (1-450) had no detectable binding to the mannopentaose. Therefore, the mPNGc domain is exclusively responsible for the binding of mPNGase to mannopentaose.

On the other hand, neither mPNGc nor full-length mPNGase showed any affinity towards chitobiose in ITC experiments. This unexpected result indicates that mPNGase does not engage in high-affinity interactions with the first two acetyl glucosamine residues of the glycan-chain, but instead also requires the peptide part of the substrate for a high affinity interaction. For comparison, the same ITC experiments were also carried

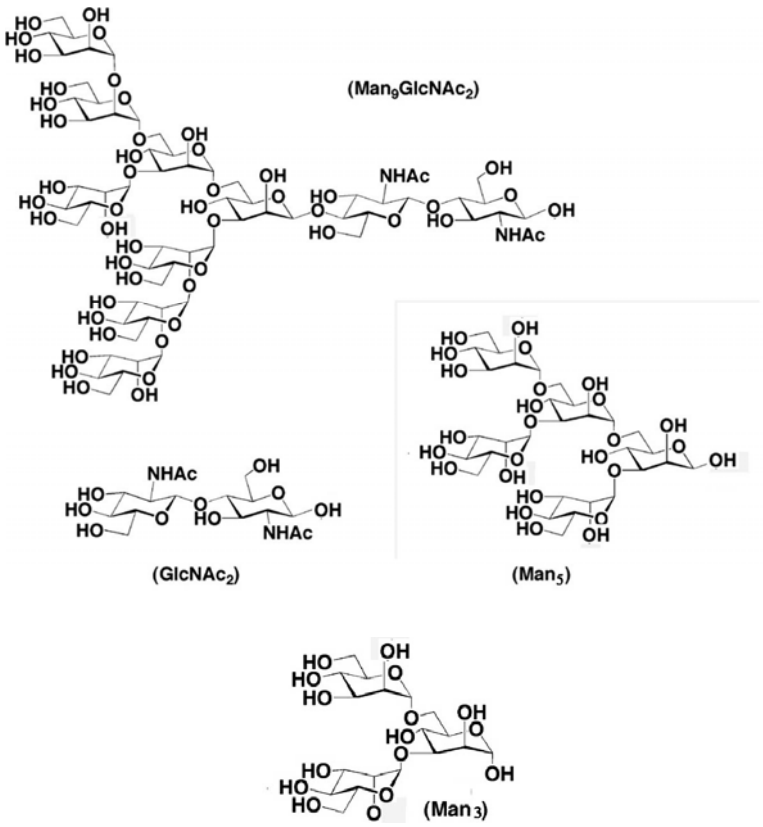


Figure 2.4 **The chemical structures of carbohydrate compounds.** The structures of $Man_9GlcNAc_2$, chitobiose ($GlcNAc_2$), mannopentaose (Man_5), mannotriose (Man_3) are shown. This figure has been modified from Suzuki, T. et al**. [37].**

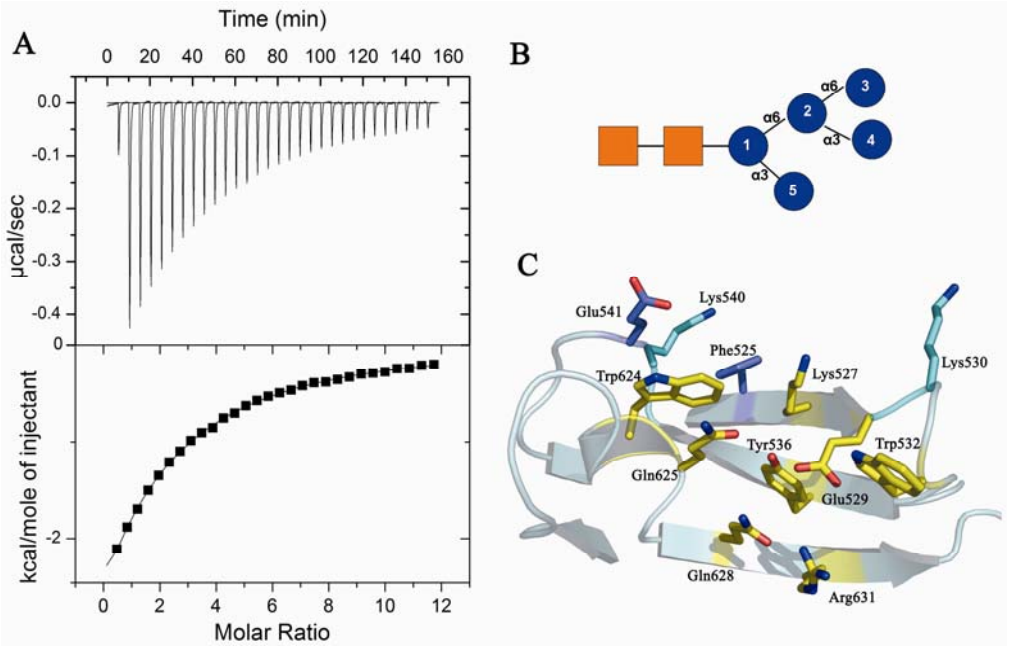


Figure 2.5 **Ligand binding to mPNGc.** *(A)* (Top) Raw ITC data showing the binding of mannopentaose to mPNGc. (Bottom) Fit of the experimental data (black squares) with a one-site binding model (thin line). *(B)* Schematic representation of the first seven residues of the N-linked oligosaccharide (orange squares: N-acetyl-glucosamine; blue circles: mannose residues with numbers). Mannoses 1 and 2 as well as 2 and 3 are linked via α-1,6-glycosidic linkages, while mannoses 2 and 4 as well as 1 and 5 are connected via α-1,3-glycosidic bonds. (*C*) Close-up view into the putative oligomannose binding pocket. Residues altered by site directed mutagenesis are shown in stick representation. Carbon atoms of residues that retain no binding affinity to mannopentaose after mutation to Ala are colored in yellow, while those of residues, which retain partial binding affinity are colored in blue. Carbon atoms of residues that have no effect after substitution with Ala are shown in cyan.

out with yPNGase (data not shown). yPNGase neither binds to mannopentaose, which is consistent with the absence of the C-terminal domain in yPNGase, nor does it interact with chitobiose as one would expect based on the close structural relationship between the core domains of mPNGase and yPNGase [27, 84].

Since α3, α6-mannotriose has been reported to inhibit the function of mPNGase [61], it was also used as substrate in the ITC binding assay. Mannotriose showed basically identical binding affinity (K_d ~ 59 μM) to mPNGc as mannopentaose (Figure 2.6), indicating that this branched structure represents the minimal binding unit required for interactions with mPNGc. PNGase apparently mainly recognizes the high mannose type oligosaccharide substrates at the second branch site, which consists of mannose residues 2, 3 and 4 (Figure 2.4 and 2.5*B*), in agreement with the complex structure described below.

E. Mannose binding site of mPNGc

Based on the conservation of solvent exposed residues in mPNGc and the structural similarity between mPNGc and the SBD of Fbs1, site directed mutagenesis was carried out to probe the role of selected residues in oligomannose binding. The following residues, Phe525, Lys527, Glu529, Lys530, Trp532, Tyr536, Lys540, Glu541, Gln625, Gln628, Trp624, Arg631(Figure 2.5*C*) were individually replaced with Ala and their mannopentaose-binding affinities were investigated by ITC (Table 2.4). These studies revealed that the K527A, E529A, W532A, Y536A, W624A Q625A, Q628A and R631A substitutions completely abolished binding of mPNGc to mannopentaose, whereas the F525A and E541A mutants resulted in slightly reduced binding affinities. Two additional

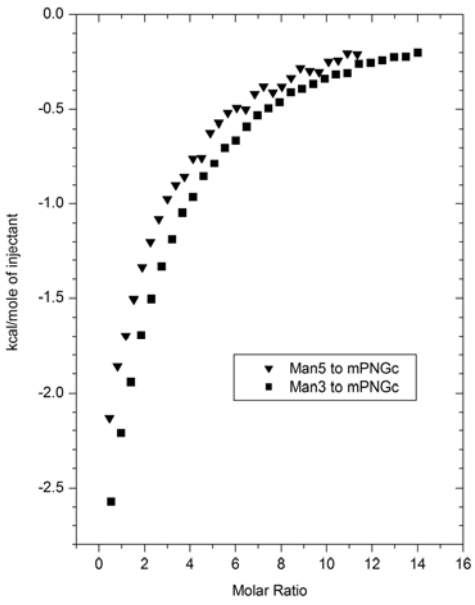


Figure 2.6 **ITC experiments with mPNGc and oligo-mannoses.** Overlaid binding isotherms of mannopentaose (solid triangles) or mannotriose (solid squares) titrated to mPNGc.

Table 2.4 **Binding affinities of mPNGc wild-type and variants to mannopentaose**

Variant	K_d (μM)
Wild-type	67
K527A	nd
E529A	nd
W532A	nd
Y536A	nd
Q625A	nd
Q628A	nd
W624A	nd
R631A	nd
F525A	95
E541A	88
K530A	65
K540A	65

nd: Binding could not be detected by ITC.

substitutions, K530A and K540A, which are not highly conserved, revealed no detectable effects on binding. To confirm that the substitutions do not affect the overall structure of mPNGc, the corresponding variants were analyzed by CD spectroscopy (Figure 2.7). Except for the Q628A mutation, the CD spectra of all other mutants are almost identical to that of the wild-type. For the Q628A variant, the minimum of the curve shifted from 214 nm to 226 nm, which may indicate that minor local conformational changes have occurred.

From an analysis of the mPNGc structure, it became clear that all of these residues are located on the concave part of the saddle, except K530 and K540, which are pointing away from the groove (Figure 2.5*C*). The oligomannose binding site therefore appears to be formed by β-strands 8, 9 and 16 which provide the concave part of the saddle and the loops between β8 and β9 as well as β15 and β16 and the second 3_{10} helix which form the ridges on either side of the saddle. Trp532 resides in the loop between β-strands 8 and 9 and Trp624 in the 3_{10}-2 helix, and these residues are on either side of the binding site.

F. Structure of mPNGc in complex with mannopentaose

To visualize the mPNGc-carbohydrate interaction the mPNGc-mannopentaose complex was crystallized. In one setup, crystals appeared as very thin leaf-like structures with five crystals clustered at one end, after 3 month of growth. One crystal was prepared from this cluster and used for structure determination. The corresponding model was refined at 1.75 Å resolution to an R-factor of 17.2% (R_{free} of 20.7%) with TLS refinement (one body for the protein/carbohydrate complex). In the complex there are two molecules

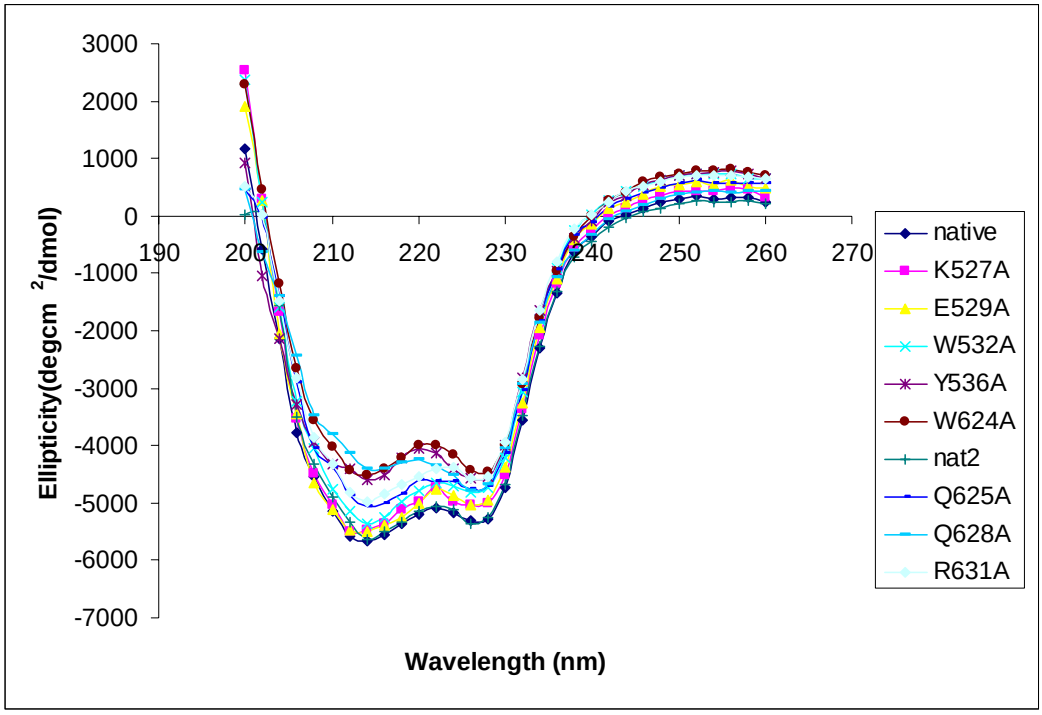


Figure 2.7 **CD spectra of the mPNGc wild-type and variants** at concentrations varying between 13 and 19 μM.

(A and B) in the asymmetric unit, which are very similar to each other (rms deviation of 0.2 Å for all Cα atoms), except in the flexible loop between β16 and β17. This indicates that the structure of mPNGc does not undergo major conformational changes upon substrate-binding. Overall, molecule B is slightly better defined in the electron density maps and is shown in Figure 2.8. Three of the five mannoses are well-defined in the electron density maps (Figure 2.8A), while Man1 is rather flexible and Man5 is apparently completely disordered. Man1 represents the reducing end of the mannopentaose and is connected to the chitobiose core. However, it does not form hydrogen bonds with mPNGc (Figure 2.8*B* and Table 2.5). Man2, 3 and 4 lie within the binding grove and form fifteen hydrogen bonds with the protein. Man2 is located at the center of the binding grove and hydrogen bonds to the side chain of Asp531, Trp532 and Gln625. Man3 has the best defined density and H-bonds to Glu529, Gln625, Gln628 and Arg631. Man4 only interacts with Lys527 via a direct H-bond. 396 $Å^2$ of surface area is buried upon complex formation for the protein and the ligand. In addition to these direct protein-substrate interactions, there are water-mediated interactions (Figure 2.8*B* and Table 2.5), which differ in number (3 in molecule A and 5 in molecule B) between the two molecules in the asymmetric unit. Besides Tyr536, all of other residues, including Lys527, Glu529, Trp532, Gln625, Gln628, Trp624 and Arg631, which are important for the interaction with mannopentaose according to our mutagenesis studies, have been found to hydrogen bond to mannopentaose either directly or through the water molecules. The importance of Tyr536 is thought to maintain several side chain conformations, since it hydrogen bonds to other residues including Glu529, Gln625 and Gln628. Compared to

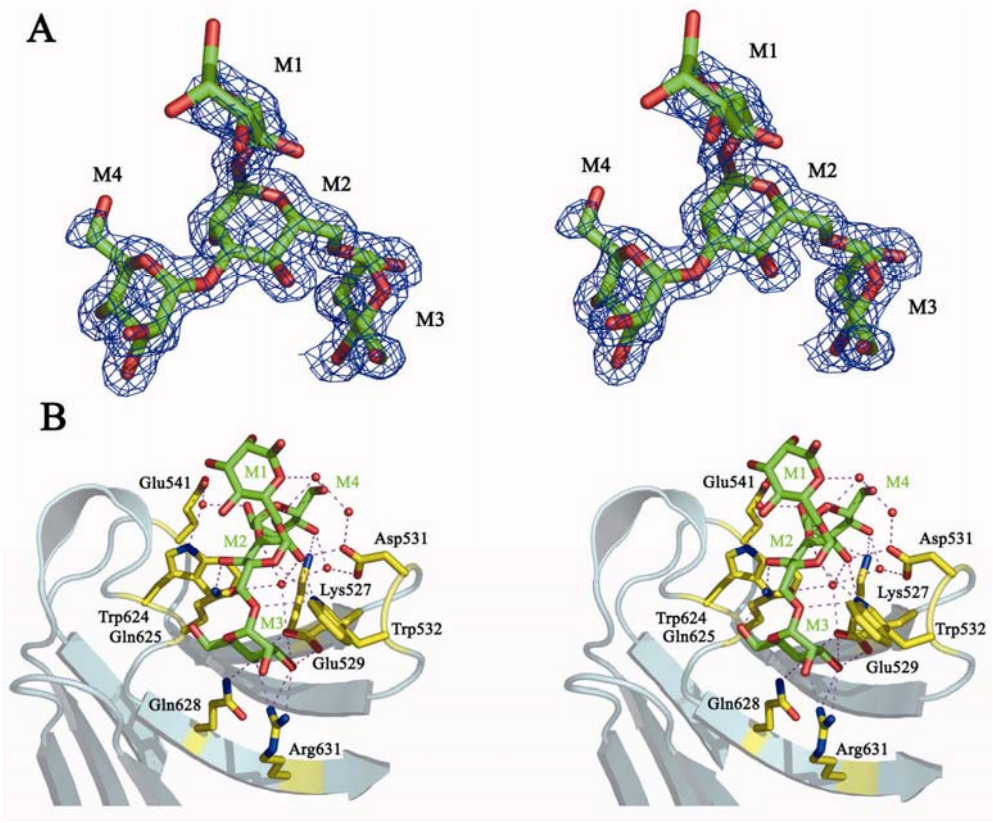


Figure 2.8 **Oligomannose binding to mPNGc.** *(A)* Stereo view of a $2F_o$-F_c omit map contoured at one times the rms deviation with the mannose residues omitted from molecule B. The carbon atoms of the four visible mannose residues are colored in green and their numbering is the same as in Fig. 2.5*B*. *(B)* Stereo view of the hydrogen bonded interactions between mPNGc and the mannotetraose as observed in molecule B. Carbon atoms of interacting residues are shown in yellow, potential hydrogen bonds are indicated as dashed lines in magenta, and water molecules are shown as red spheres.

Table 2.5 **H-bonds between mannopentaose and the C-terminal domain of mPNGase**

Direct mannose-protein interactions				
Mannopentaose	**Atom**	**mPNGc**	**Atom**	**Distance (Å)**
Man2	O2	Asp531 Trp532	Oδ2 Nε1	2.5 3.1
Man2	O3	Arg631 Glu529	NH2 Oε2	2.9 3.0
Man2	O4	Gln625	Nε2	2.7
Man2	O5	Trp532	Nε1	2.8
Man2	O6	Trp532	Nε1	3.2
Man3	O3	Arg631 Glu529	NH2 Oε1	2.7 3.0
Man3	O4	Glu529 Gln628 Arg631	Oε2 Nε2 NH1	2.6 3.0 3.0
Man3	O6	Gln625	Nε2	2.8
Water mediated interactions				
H_2O		**Residues**	**Atoms**	**Distance (Å)**
#1		Man1 Man1 Man4 #2 H_2O	O5 O6 O6	2.9 3.3 3.1 2.6
#2		Asp531	Oδ2	2.6
#3		Glu529 Gln625 Man2 Man2 #4 H_2O	Oε2 Oε1 O2 O3	2.8 3.0 3.0 2.9 3.0
#4		Asp531 Man4	Oδ1 O4	2.7 2.7
#5		Glu541 Trp624 Man4	Oε2 Nε2 O2	2.8 3.0 3.0

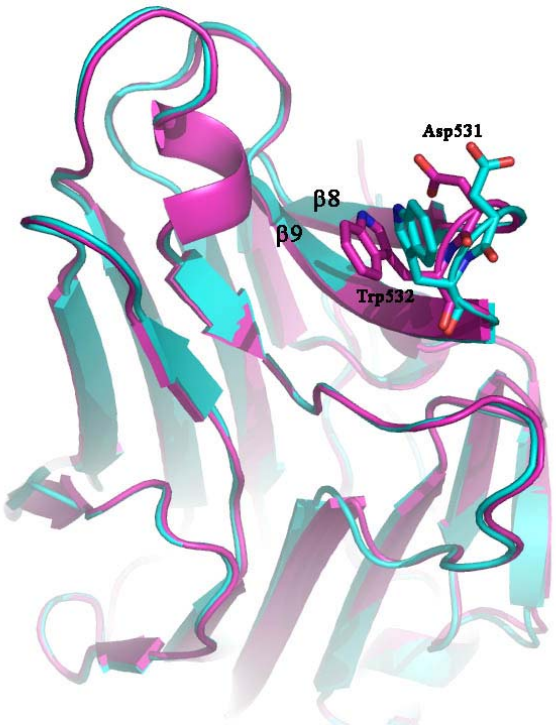


Figure 2.9 **Superposition of mPNGc in the apo-structure and in complex with mannopentaose.** The apo-structure is shown in cyan and the structure of the complex in magenta. Asp531 and Trp532 are shown in stick representation.

the apo-structure, the loop between β8 and β9 moved about 1.4 Å closer towards the binding pocket resulting in the formation of additional H-bonds with the mannopentaose (Figure 2.9). At the same time, rotations of the side chains of Asp531 and Trp532 in this loop allow interactions with the substrate. Overall there is an excellent agreement between the cocrystal structure and the binding studies involving altered residues.

G. mPNGc enhances the catalytic activity of mPNGase

To investigate whether the C-terminal domain contributes to the de-glycosylation activity of mPNGase (Figure 2.10), we compared the enzymatic activity of full-length mPNGase, mPNGase ΔC (residues 1-450) and the mPNGase K527A mutant, a mutant in which the binding of mPNGase to mannopentaose is abolished according to our ITC studies (Table 2.4). RNase B, a high mannose type N-glycosylated protein with one N-glycosylation site and a well-characterized PNGase substrate, was used as substrate [29]. The assay showed that the half-life ($t_{1/2}$) of glycosylated RNase B in the presence of full-length mPNGase is about 2.5 minutes, whereas it is about 80 minutes in the presence of mPNGase ΔC. This dramatic difference between the full-length protein and the mPNGase ΔC truncation confirmed that the C-terminal domain is very important for mPNGase activity. The mPNGase K527A mutant displayed an intermediate activity in this assay with a $t_{1/2}$ of about 5 minutes. This indicates that although no binding to mannopentoase of the corresponding mPNGc variant could be detected in the ITC experiments, a residual affinity remains, which increases the catalytic activity of the core domain. In conclusion, although the catalytic core domain of mPNGase exhibits de-N-glycosylation activity in the absence of the C-terminal domain, the presence of this

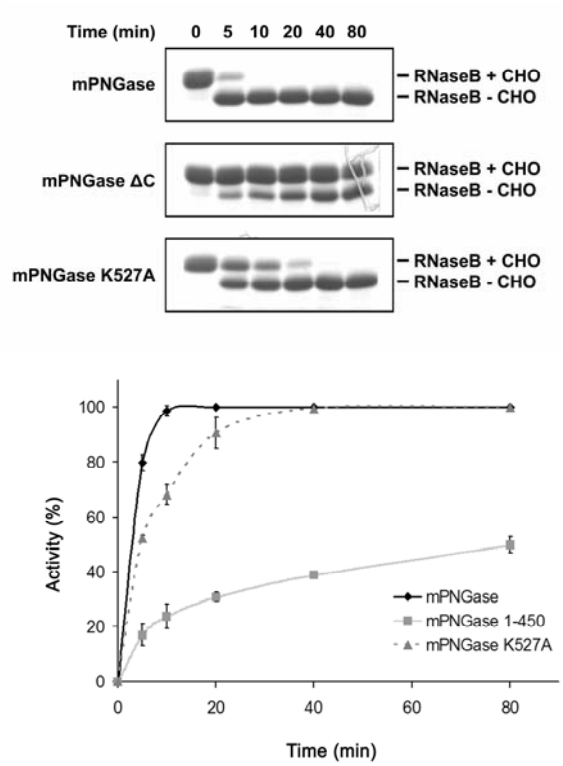


Figure 2.10 **Ribonuclease B digestion by PNGase.** Time course of the reactions between full-length mPNGase, mPNGase ΔC (residues 1-450), and the K527A variant of full-length mPNGase with RNase B as the substrate. Samples were analyzed at the indicated time points by SDS PAGE. Glycosylated (+CHO) and deglycosylated (-CHO) forms of RNase B migrate as indicated. The curves were obtained by laser densitometric analysis of SDS PAGE gels and represent the average of two experiments. The error bars indicate the resulting standard deviations.

domain greatly accelerates substrate turn-over.

H. The last 36 residues of yPNGase are important for efficient turn-over

Sequence comparisons between mPNGase and yPNGase showed that the yeast enzyme basically only contains the catalytic domain, which is present as the central domain in mPNGase. Not surprisingly the structures of yPNGase [27] and the mPNGase core domain [26] displayed high structural homology as reflected in a rms deviation of 1.5 Å for 295 out of 320 residues. However, there are two additional N-terminal α-helices in yPNGase which have no structural counterparts in mPNGase. As described elsewhere [26] this difference is a key factor for the different interaction between yPNGase and the XPCB domain of Rad23 compared to how mPNGase interacts with the XPCB domain of HR23B [26, 85]. In addition, yPNGase contains residues 328 to 363 at the C-terminus which were absent from all crystal structures. These residues also have no counterparts in the mPNGase sequence.

Given the high sequence and structural similarity one would predict that the activities of yPNGase and the mPNGase are comparable. To test this hypothesis, an activity assay was performed using RNase B as the substrate. Surprisingly, yPNGase displays a much higher activity than the mPNGase core domain (Figure 2.11*A*), however, comparable activities were obtained when full-length mPNGase was used instead of the core domain (Figure 2.11*B*). In order to understand this result, additional activity assays were designed. Since the C-terminal domain of mPNGase (451-651) helps to accelerate the enzyme activity of mPNGase, it was attached directly to the C-terminus of full-length yPNGase to see whether this domain could improve yPNGase activity as well. At the

same time, a truncated form of yPNGase (residues 8-341), which was used in the structural studies of yPNGase, was assayed to compare its activity with that of the full-length protein. In contrast to the previous activity assay, these proteins were purified in complex with XPCB binding domain of yeast Rad23 to increase the stability of the enzymes. The result unexpectedly revealed that yPNGase with the fused mPNGase C-terminal domain did not display an increase in activity (Figure 2.11*C*), instead both proteins have very similar activities. However, the truncated yPNGase 8-341 construct has a lower activity compared to full-length yPNGase. Since there are only 7 residues missing at its N-terminal end, which are at least 65 Å away from the substrate-binding cleft, but instead are close to the XPCB domain, these missing residues at the N-terminus are not likely to be the reason for the loss in activity. Since there are 22 residues missing from the C-terminus in this construct, it is very likely that the last 36 residues (328-363) of yPNGase are responsible for the difference in catalytic activity, since this region has no counterpart in mPNGase as mentioned above (Figure 2.12). When compared with other fungal (*S. pombe*, *K. lactis* and *M. grisea*) PNGase proteins, this very C-terminal region has somewhat low sequence conservation (19% identity and 47% similarity with PNGases), suggesting that this region might be unique to *S. cerevisiae* enzyme. It is tempting to speculate that these residues are involved in binding the carbohydrate moiety of the substrate and hence play a similar role to the C-terminal domain of mPNGase.

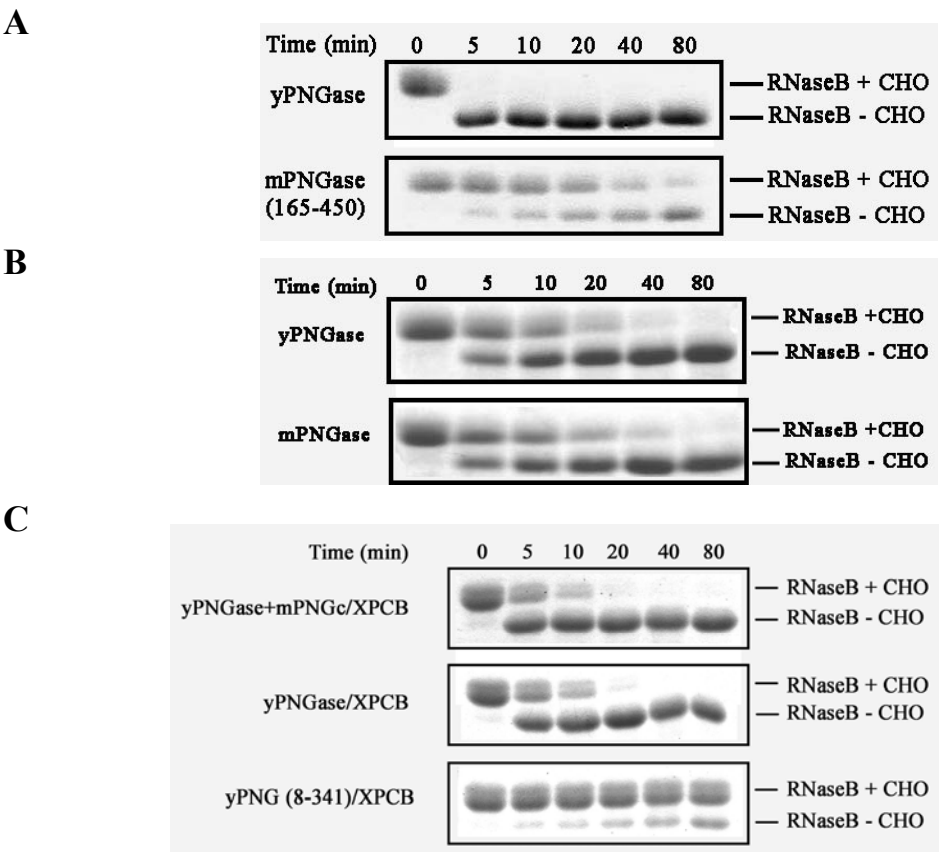


Figure 2.11 **PNGase activity assay using RNase B as a substrate.** *(A)* Activity comparison between yPNGase and the core domain of mPNGase in the absence of respective XPCB domain. The molar ratio between enzyme and substrate is 1: 40. *(B)* Activity comparison between yPNGase and mPNGase in the absence of the respective XPCB domain. The molar ratio between enzyme and substrate is 1:200. *(C)* Activity comparison between the fusion protein containing yPNGase and the C-terminal domain of mPNGase (yPNGase+mPNGc), full-length yPNGase and C-terminally truncated yPNGase (residues 8-341). In this assay, all three proteins were purified in complex with the XPCB domain of yeast Rad23. The molar ratio between enzyme and substrate is 1:200.

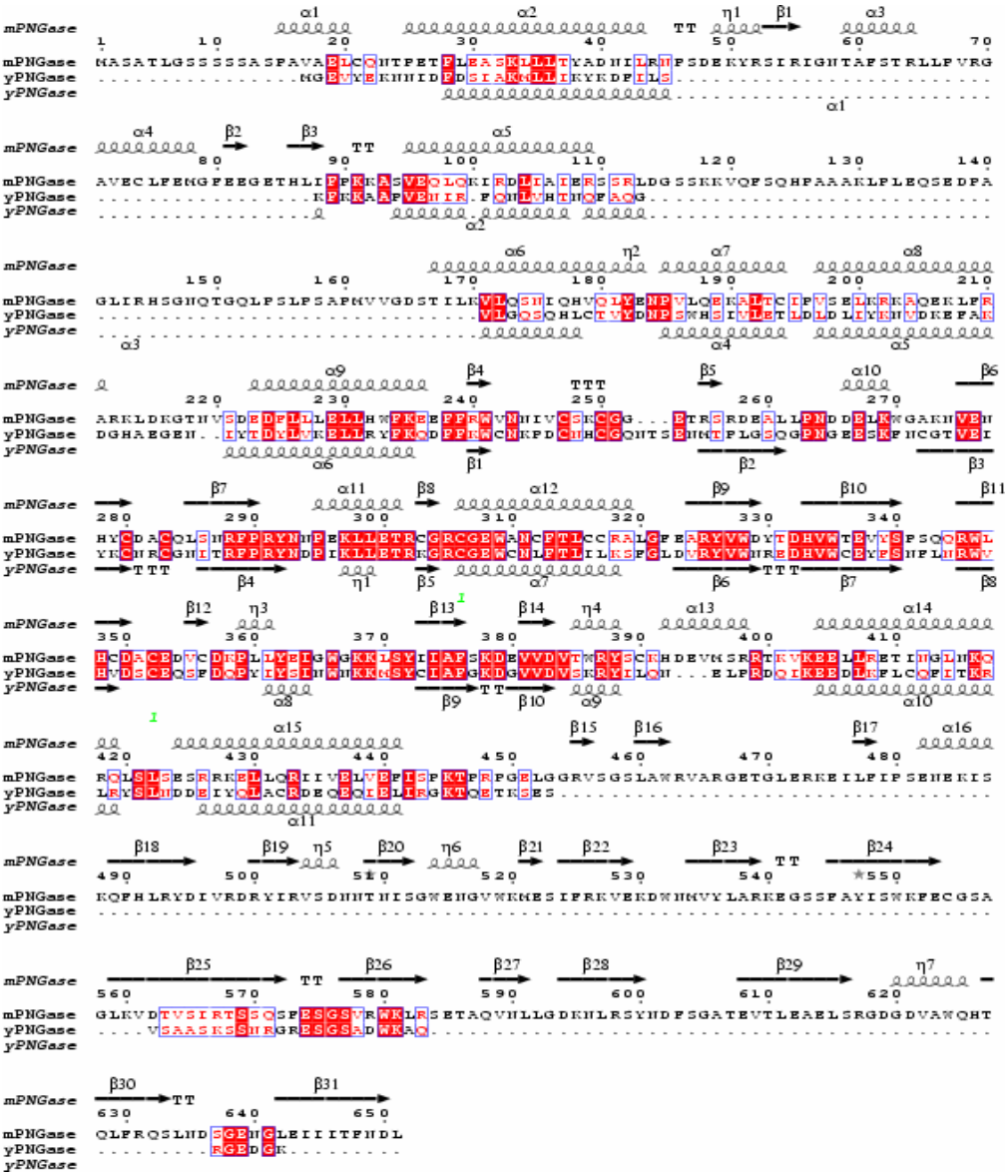


Figure 2.12 **Sequence alignment of mPNGase and yPNGase.** The secondary structure of mPNGase is labeled on top of the alignment and that of yPNGase on the bottom.

IV. Discussion

The importance of PNGase in the ERAD relies on the activity of this enzyme, primarily due to its de-N-glycosylation activity, but also its ability to bind to other proteins involved in ERAD. In lower eukaryotes, PNGase consists mainly of its catalytic domain. In higher eukaryotes, however, PNGase has evolved two extra domains at either terminus. The C-terminal domain is highly conserved in this subclass of the PNGase family, suggesting that it may be involved in an important function of this enzyme. In this study, the function of the mPNGase C-terminal domain was identified as an oligomannose-binding module. The carbohydrate binding property of this domain was first suggested through the comparison of the mPNGc crystal structure with that a non-redundant set of proteins derived from the Protein Data Bank. Subsequently this activity was demonstrated by site directed mutagenesis studies and ITC experiments. In addition, the crystal structure of the mPNGc-mannopentaose complex revealed the detailed atomic interactions between this domain and the glycan substrate.

Activity assays with RNase B as substrate showed that deletion of this C-terminal domain results in a significantly reduced activity. Therefore, the presence of this domain accelerates substrate turn-over. Unlike mPNGase, yPNGase apparently does not have a corresponding carbohydrate recognition domain. However, its overall activity is nearly identical to that of mPNGase. This was an unexpected result given the absence of a C-terminal carbohydrate-binding domain in yPNGase. At the same time a significant drop in activity was observed when the mPNGase C-terminal domain was removed from mPNGase. An attempt to further increase the activity of yPNGase by fusing the mPNGase C-terminal to it did not create a hyperactive enzyme, at the same time no

reduction was observed either. Additional activity assays revealed that the last 36 residues of yPNGase play an important role for the efficient cleavage of glycan chains. However, how the C-terminal 36 residues affect the overall activity is unclear at present since these residues are absent in all crystal structures of the yeast enzyme and they have not yet been studied by site directed mutagenesis. Our current working hypothesis postulates that these residues are involved in carbohydrate recognition and therefore have a similar function as the mPNGase C-terminal domain, which represents the oligomannose binding entity.

Although in ITC studies, neither mannopentaose nor chitobiose showed any binding to yPNGase, we have not yet tested oligsaccharides that contain both the chitobiose and mannopentaose part. Therefore, it is still possible that yPNGase could have a low affinity for the substrate once the peptide part of the substrate is bound. In fact it has been demonstrated that iodo-acetate labeled chitobiose binds to yPNGase [37]. Although this involves a covalent modification of the active site cysteine there must be some specificity provided by the chitobiose since the active site cysteine is the only cysteine reacting with this compound out of fourteen cysteines present in yPNGase.

Chapter 3

The Structure of the PNGase-P97 Complex Suggests that P97 Phosphorylation Modulates Endoplasmic Reticulum-Associated Degradation

The major findings presented in this chapter have been presented in the publication entitled "The structure of the PNGase-P97 complex suggests that P97 phosphorylation modulates endoplasmic reticulum-associated degradation" by G. Zhao et al., Proc. Natl. Acad. Sci USA Vol. 104, pp. 8785-90 (2007). I was a joint first author of this manuscript and have carried out the following experiments: Purification of the PUB domain proteins and mutants, crystallization of the PUB domain protein, data collection, model building and ITC measurements.

I. Introduction

The presence of the PUB domain was first detected by analyzing and comparing PNGase sequences from vertebrates and drosophila [40]. Additional sequence searches with Blast [40] revealed that four PUB domain homologs are also found in *Arabidopsis thaliana*, however, these are proteins which are distinct from PNGase and besides this domain also possess either a UBA (ubiquitin associated) or UBX (Ubiquitin regulatory X) domain. Subsequent yeast two-hybrid experiments [32] using full-length mPNGase as bait suggested the PUB domain to be a protein-protein interacting domain. Recently, biochemical experiments demonstrated that p97 binds to mPNGase via its PUB domain [21]. Since p97 interacts with other proteins involved in ERAD [46], it provides a structural and possible functional link between E3 ligases, such as mAMFR and SCF ligases containing the sugar-binding F-box protein Fbs1 and the deglycosylation activity of mPNGase.

In this chapter, I demonstrate that the PUB domain binds to a previously undetected protein-protein interaction motif of p97, contained within its last ten residues (p97-C10); this module also interacts with Ufd3 and possibly other cofactors of p97. Phosphorylation of the penultimate tyrosine residue of p97 (Tyr805 in mouse p97) abolishes p97's interaction with both PNGase and Ufd3. The crystal structure of the PUB domain in complex with p97-C10 reveals that the C-terminus of p97 fits into a conserved and positively charged groove on the surface of the PUB domain. These results open up new possibilities to elucidate the roles of p97 in various cellular processes, and to study proteins that possess the PUB domain, an evolutionarily new domain.

II. Materials and Methods

A. Construct design

The gene encoding the N-terminal domain of mouse PNGase (residues 12-111) was amplified from full-length mPNGase by PCR and was cloned into the NdeI/SapI sites of the pTYB1 vector (pTYB1-PUB) (New England Biolabs, Beverly, MA) by Dr. Gang Zhao. Another construct of the N-terminal domain of mouse PNGase (1-111) was cloned into the NdeI/XhoI sites of pET28a vector (His_6-PUB). All PUB domain mutations were introduced with the QuikChange site-directed mutagenesis kit using pTYB1-PUB as template. The full-length, N (1-199), ND1 (1-458), D2C (459-806) constructs of mouse p97 were amplified from pQE9-p97 [42] and ligated into the NdeI/BamHI sites of pET28a vector by Dr. Gang Zhao. The p97 mutants were introduced with the same kit mentioned above using full-length p97 as template. GST-mPNGase(1-130) was constructed by cloning into the EcoRI/BamHI sites of the pGEX-5X-1 vector (GE Healthcare Life Sciences) by Dr. Guangtao Li (Lennarz laboratory). The primers used in this study are listed in Table 3.1.

B. Protein expression and purification

The transformation step was similar to that described in Chapter 2, except that full-length p97 with a His_6-tag was transformed into *E. coli* M15 cells and selected on LB plates containing 100 μg/μl of Ampicillin and 50 μg/μl Kanamycin, and GST-mPNGase (1-130) was transformed into BL21DE3 cells and selected on LB plates containing 100 μg/μl of Ampicillin. All other proteins were transformed into BL21DE3 Codon Plus RIL cells (Ampicillin and Chloramphenicol resistances). GST-mPNGase (1-130) was

Table 3.1 **Primer sequences:**

Primers	**Sequences**
mPNGase(1-111) 5’	GGAATTCCATATGGCGTCGGCCACACTGGGC
/pET28a 3’	CCGCTCGAGTCAGCTGCTTCTCTCTATGGCAATC
PUB K33E 5’	CTTTCTGGAGGCCTCCGAGCTGCTGCTCACCTAC
/pTYB1 3’	GTAGGTGAGCAGCAGCTCGGAGGCCTCCAGAAAG
PUB Y38F 5’	CAAGCTGCTGCTCACCTTCGCCGACAAC
/pTYB1 3’	CAGGATGTTGTCGGCGAAGGTGAGCAGC
PUB K50E 5’	GAAACCCCAGTGATGAAGAATACAGATCCATCCGTATTG
/pTYB1 3’	CAATACGGATGGATCTGTATTCTTCATCACTGGGGTTTC
PUB R52E 5’	CAGTGATGAAAAATACGAATCCATCCGTATTGGG
/pTYB1 3’	CCCAATACGGATGGATTCGTATTTTTCATCACTG
PUB R55E 5’	GAAAAATACAGATCCATCGAGATTGGGAACACAGCG
/pTYB1 3’	CGCTGTGTTCCCAATCTCGATGGATCTGTATTTTTC
PUB N58D 5’	GATCCATCCGTATTGGGGACACAGCGTT
/pTYB1 3’	GTAGAAAACGCTGTGTCCCCAATACGGA
PUB R64E 5’	CACAGCGTTTTCTACTGAACTCTTGCCTGTCAGAG
/pTYB1 3’	CTCTGACAGGCAAGAGTTCAGTAGAAAACGCTGTG
PUB H86E 5’	GAAGAGGGAGAAACGAATCTTATCTTTCCTAAAAAAGC
/pTYB1 3’	GCTTTTTTAGGAAAGATAAGATTCGTTTCTCCCTCTTC
PUB K91E 5’	CATCTTATCTTTCCTGAAAAAGCTTCAGTGGAAC
/pTYB1 3’	GTTCCACTGAAGCTTTTTCAGGAAAGATAAGATG

expressed at 37° C after induction with 0.1 mM IPTG at an A_{600} of 0.6–0.8.overnight. All other proteins were expressed at 15° C after induction with 0.3 mM IPTG at an A_{600} of 0.6–0.8. The purifications of the p97 fragments were performed by Dr. Gang Zhao. After growth, pTYB1-PUB and its mutant cells were harvested by centrifuging at 9000 g for 15 minutes at 4 ºC and lysed in 20 mM Tris/HCl, pH 8.5, 150 mM NaCl, a protease inhitor cocktail (Complete mini, EDTA free, Roche) by passing twice through a French pressure cell at 1,500 psi. The cell lysate was then centrifuged at 50000 g for 30 minutes. The cleared cell lysate supernatant was collected and incubated with chitin beads (New England Biolabs, Ipswich, MA) for 30 minutes at 4º C and washed with 20 cv of buffer containing 20 mM Tris/HCl, pH 8.5, 1 M NaCl, followed by equilibration with cleavage buffer (wash buffer plus 50 mM DTT). Beads were incubated overnight at room temperature before elution with 20 mM Tris/HCl, pH 8.5, 150 mM NaCl.

The His_6-PUB cells were collected and lysed using the same conditions as the pTYB1-PUB cells. The cleared cell lysate was incubated with a Ni-NTA affinity column for 30 minutes in binding buffer consisting of 20 mM Tris/HCl, 150 mM NaCl, 10 mM imidazole and washed with 10 cv of binding buffer. Bound proteins were eluted with 20 mM Tris/HCl, 150 mM NaCl, 250 mM imidazole. All protein purifications included the next two steps consisting of MonoQ anion exchange 5/5 and Superdex 200 (10/30 for PUB domain mutants, 26/60 for all others) size exclusion chromatography. The detailed protocols are the same as those described in Chapter 2. The proteins were concentrated with Centricon plus-20 5,000 MWCO (Millipore) to a concentration of 10-20 mg/ml using calculated molar extinction coefficient of 3105 $M^{-1}cm^{-1}$ for PUB domain. The mPNGase-p97 complex was prepared by mixing the two purified proteins together on ice

for 30 minutes followed by size exclusion chromatography on a Superdex 200 (26/60) column. This experiment was carried out by Dr. Gang Zhao.

C. Protein-protein interaction studies

ITC measurements were carried out using a VP-ITC microcalorimeter (Northampton, MA). Three synthesized peptides (Anaspec, San Jose, CA) used in this study, C40 (GFGSFRFPSGNQGGAGPSQGSGGGTGGSVYTEDNDDDLYG), C10 (TEDNDDDLYG), and C13 (SVYTEDNDDDLYG) were derived from the C-terminus of p97. They were chemically synthesized and purified by HPLC using reverse C18 column by the company resulting in a purity of more than 95%. Prior to each experiment the proteins and peptides were dialyzed overnight at 4° C against a buffer containing 20 mM Tris-HCl pH 8.5, 150 mM NaCl, 2% glycerol and 1 mM β-mercaptoethanol, followed by degassing. Proteins at concentrations of ~15 μM were titrated with 10 μl of 0.2 mM of either peptide or protein with a total number of 30 injections at 25° C. A 300 second waiting period was set between each injection to allow the baseline to stabilize. The binding parameters were calculated using Origin (version 7.0) by fitting the data to a single site binding model. The ITC experiments involving mutant PUB domains were performed by myself whereas the remaining titrations were performed by Dr. Gang Zhao.

GST pull-down experiments were performed with purified proteins by Dr. Guangtao Li. Two μg of GST-mPNGase (1-130) bound to 8 μl of GSH-agarose beads were mixed with 4 μg of either wild-type (WT) or mutated His_6-tagged p97 in 0.5 ml binding buffer (0.2 M Na-Phosphate buffer pH 7.4, 150 mM NaCl, 1% Triton-X 100, 1 mM $MgCl_2$, 1 mM ATP, 5 mM DTT, 5% glycerol, 4 mM PMSF). The binding

experiments were performed at 4°C for 1 h, after which the beads were washed three times with binding buffer. Bound proteins were eluted with SDS sample buffer and analyzed by SDS-PAGE followed by Western blot analysis with a monoclonal antibody against the His-tag.

D. Crystallization and structure determination

The PUB domain at a concentration of 13 mg/ml was crystallized by hanging drop vapor diffusion against a reservoir solution containing 18-22% PEG 5000 MME and 14-18% Tacsimate, pH 7.0 (Hampton research). The hanging drop is prepared by adding 1 µl of protein complex and 1 µl of reservoir solution. The heavy atom derivative was prepared by soaking the crystals in mother liquor containing 2.5 mM sodium ethylmercurithio-salicylate (EMTS) for 20 minutes. After mixing the PUB domain (1 mM) and p97-C10 (2 mM) in a 1:2 molar ratio, the resulting complex was crystallized under similar conditions containing 16-22% PEG 5000 MME and 14% Tacsimate, pH 7.0. The crystals were flash frozen in liquid nitrogen. The native dataset was collected with a Rigaku RU-H3R rotating anode X-ray generator operating at 50 kV and 100 mA and equipped with confocal multilayer optics and a Rigaku RAXIS-IV^{++} detector. The mercury derivative and the complex datasets were collected on beam line X26C of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory at a wavelength of 1.0074 Å. All diffraction data were collected at 100° K, indexed, integrated and scaled with HKL2000 [86].

The structure was determined by the single isomorphous replacement and anomalous scattering (SIRAS) method. Phase determination and density modification

were carried out with Solve/Resolve [87, 88] to 2.5 Å resolution including building of a partial structure. Two Hg sites were also identified by Solve/Resolve. The resulting model was input into ARP/Warp [67] which was able to build 98 out of 100 residues. The protein model was completed manually with the aid of the program O [66] and was refined with Refmac [68]. Water molecules were added automatically with ARP/Warp, additional solvent molecules were added manually. The structure of the complex was solved by molecular replacement using Refmac because both crystals are isomorphous. The peptide in the complex structure was added manually in O and refined in Refmac.

E. Coordinates

The atomic coordinates and structure factors (codes 2HPJ for the apo-structure and 2HPL for the complex) have been deposited in the Protein Data Bank, Research Collaboratory for Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

III. Results

A. Protein purification

The PUB domain constructs including the His-tagged mPNGase (1-111), intein-tagged mPNGase (12-111) and its mutants all showed very good expression levels under the condition described in the Materials and Methods. The typical yield of the His-tagged mPNGase (1-111) was about 10 mg pure protein from 1 L of cell culture and that of the intein-tagged PUB domain protein was about 2 mg per liter. After purification the proteins were about 95% pure according to SDS-PAGE analysis. The purification of p97 fragments were performed by Dr. Gang Zhao. The yield of these proteins was about 10 mg/L culture, and the purity was about 90% as judged by SDS-PAGE. The PUB domain proteins used in the crystallization set up were usually concentrated to 10-20 mg/ml prior to being flash frozen as 15 μl drops in liquid nitrogen and stored at -80 °C.

B. The C-terminus of p97 harbors a novel protein-protein interaction motif

Previously, our collaborators, Drs. Li and Lennarz and we (Drs. Zhao, Schindelin and myself) reported that the PUB domain of mouse PNGase interacts with the D2-C fragment (residues 459 to 806) of mouse p97 [21]. To further narrow down the region of p97 that interacts with the PUB domain, various truncations of p97 were prepared and their interaction with PNGase was studied by native gel electrophoresis. As shown in Figure 3.1, in the upper gel, both full-length mPNGase and the PUB domain shifted when incubated in a 1:1 molar ratio with the p97 full-length protein. This indicated that the PUB domain is responsible for mediating the binding with p97, which is consistent with our previous report. In the lower gel, various fragment of p97 including the full-length

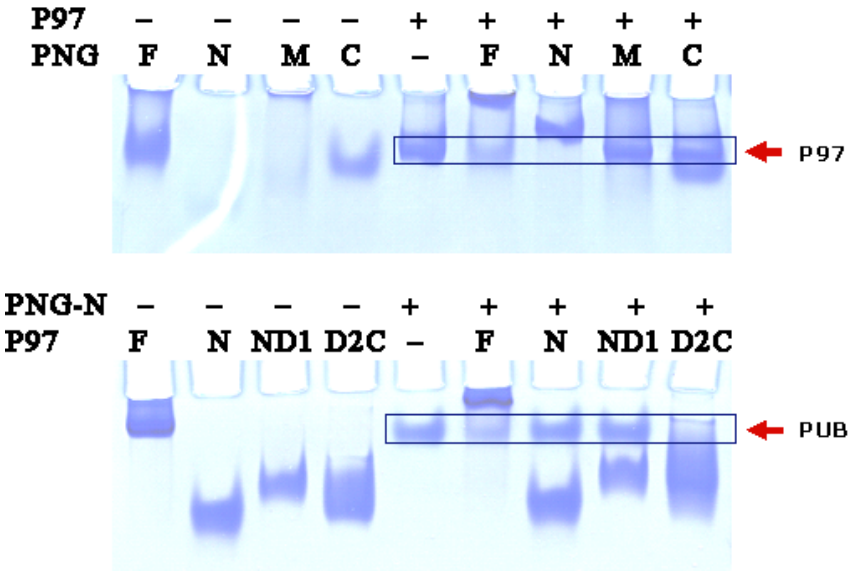


Figure 3.1 **Native gel electrophoretic analysis of the interactions between p97 and PNGase.** 4 μg of p97 (upper panel) or the PNGase PUB domain (lower panel) were incubated with various PNGase or p97 variants in a 1:1 molar ratio for 20 minutes at room temperature. The samples were resolved on a 6% native PAGE gel and stained with Coomassie Blue. Abbreviations: PNG F, N, M and C are full-length mouse PNGase and its N-terminal, core and C-terminal domains, respectively. P97 F, N, D1, D2 and C are full-length mouse p97 and its N-terminal, first and second AAA ATPase and C-terminal domains, respectively.

protein as well as the N, ND1 and D2C domains were incubated with the PUB domain of mPNGase, respectively. In the lanes containing full-length p97/PUB and the D2C domain of p97/PUB, a gel shift was observed. This indicates that the D2C domain of p97 mediates the direct interaction with the PUB domain. Considering that the D1 and D2 domains are both ATPase domains and highly homologous to each other, it seemed likely that the last 40 residues of p97 (p97-C40), which are absent from the D1 domain, are necessary for binding to PNGase.

Additional ITC experiments showed that p97-C40 indeed interacts with the PUB domain of PNGase with an apparent K_D of 11.1 μM, an affinity similar to that observed with the full-length p97 protein (Table 3.2). A notable feature of the C-terminal region of p97 is a cluster of acidic residues at the immediate C-terminus (Figure 3.2*A*). To test whether this sequence is involved in the PUB domain interaction, two peptides corresponding to either the last 13 (p97-C13) or 10 (p97-C10) residues were synthesized and their interactions with the PUB domain were characterized by ITC (Figure 3.3*B* and Table 3.2). These titrations revealed that the binding affinities of both p97-C13 (3.2 μM) and p97-C10 (3.6 μM) to the PUB domain are similar to that of p97-C40 and full-length p97 (Table 3.2). Consequently, only the C-terminal ten residues of p97 are necessary for its interaction with the PUB domain of PNGase. Further experiments demonstrated that this novel protein-protein interaction motif also mediates p97's interaction with Ufd3, albeit with a slightly lower affinity than that with the PUB domain (Table 3.2). Although the detailed biological function of Ufd3 is not known, it is found to be related to deubiquitination [49]. Since Ufd2, a E4 multiubquitination enzyme that adds additional ubiquitin molecules to existing ubiquitin conjugates, competes with Ufd3 for p97

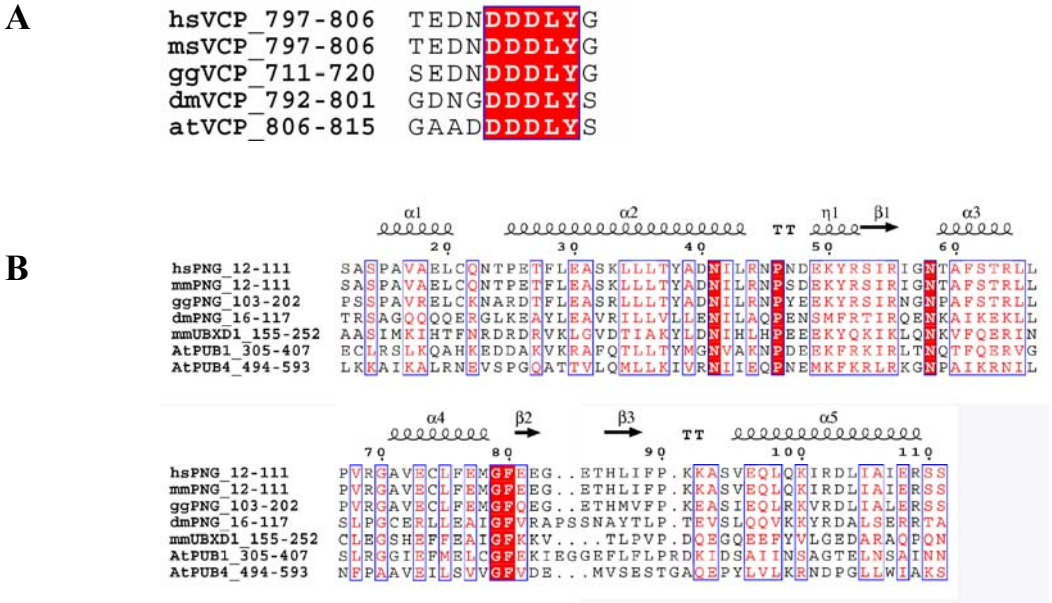


Figure 3.2 **Sequence alignment of the C-terminal residues of p97 and PUB domains.** *(A)* The last ten residues of human (NP_009057.1), mouse (NP_033529.2), chicken (XP_424984.1), *Drosophila melanogaster* (NP_477369.1) and *A. thaliana* (NP_190891.1) p97 were aligned using ClustalW [89] and the figure was prepared using ESPript [90]. Identical residues are shown in white with red background and conserved residues are shown in red with boxes. (*B*) Sequence alignment of the PUB domains of human (NP_060767.2), mouse (NP_067479.2), chicken (XP_418754.1) and *D. melanogaster* PNGase (NP_610192.1), mouse UBXD1 (NP_077752.1) and two *A. thaliana* proteins (NP_563718.1 and NP_198419.2) [40]. Residues are numbered according to the mouse PUB domain and the secondary structure elements of the mouse PUB domain are shown above the sequences.

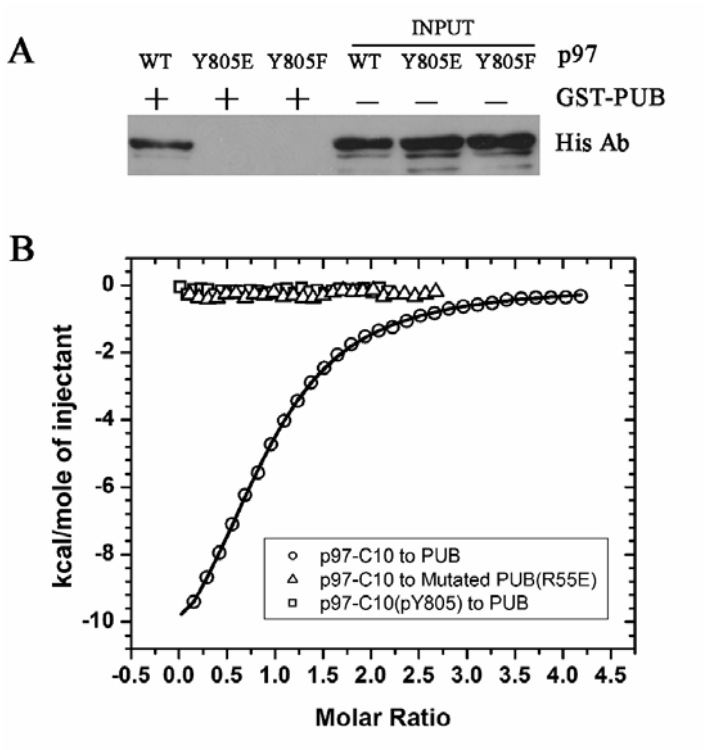


Figure 3.3 **Binding of p97 to the PUB domain of mouse PNGase.** (*A*) GST pull down assay. GST-PUB immobilized on glutathione beads was incubated with His-tagged p97 variants as indicated. Bound proteins were detected by Western blot with a monoclonal anti-His-tag antibody. (*B*) Interaction studies by ITC. Binding isotherms are shown for the last 10 residues of p97 (p97-C10) to wild-type (open circles) and the Arg55Glu variant of the PUB domain (open triangles) as well as for the Tyr805 phosphorylated form of p97-C10 to wild-type PUB domain (open squares).

Table 3.2 **ITC analyses of p97-PNGase interactions**

Injectant	**Cell**	**n**	**K_D (μM)**	**ΔH (kcal/mol)**
PNGase	p97	0.27 ± 0.07	16.7 ± 1.3	-50.5 ± 13.9
PUB	p97	0.88 ± 0.03	5.6 ± 0.3	-12.9 ± 0.6
PUB-R55E	p97	n.d.		
p97-C40	PUB	0.76 ± 0.03	11.1 ± 1.1	-8.9 ± 0.6
p97-C13	PUB	0.90 ± 0.01	3.2 ± .12	-14.2 ± 0.3
p97-C10	PUB	0.91 ± 0.01	3.6 ± 0.1	-13.5 ± 0.2
Phos-p97-C10	PUB	n.d.		
p97-C10	PUB-Y38F	1.15 ± 0.05	39.1 ± 5.3	-18.7 ± 1.5
p97-C10	PUB-K50E	0.77 ± 0.13	32.8 ± 4.1	-14.9 ± 3.2
p97-C10	PUB-R55E	n.d.		
p97-C10	PUB-N58D	n.d.		
p97-C10	PUB-R64E	0.83 ± 0.1	140.8 ± 51	-24.8 ± 1.7
p97-C10	Ufd3 330-794	0.86 ± 0.1	34.5 ± 2.5	-9.9 ± 1.4
Phos-p97-C10	Ufd3 330-794	n.d.		

n.d.: no binding detected

interaction [49], it is possible that Ufd2 also interacts with the C-terminus of p97 and all three proteins, Ufd2, Ufd3 and PNGase, may compete with each other for p97 interaction. Another scenario would predict a different binding site for Ufd2, and the Ufd2 and Ufd3 competition would simply result from spatial overlap of both proteins when bound to p97.

C. One p97 hexamer binds two PNGase molecules

During the protein interaction experiments, we observed that the stoichiometry of the interaction between full-length, hexameric p97 and full-length PNGase is significantly greater than one to one. After the PNGase-p97 complex was purified by size exclusion chromatography, resolved on SDS-PAGE and stained with Coommassie blue, the molecular quantities of PNGase and p97 in the complex were estimated by laser densitometry. Assuming that PNGase and p97 proteins stain equally well with Coommassie blue, a p97 to PNGase molar ratio of 3.2-3.5 to 1 was obtained (Figure 3.4). A similar stoichiometry was seen in an ITC experiment by titrating p97 with full-length PNGase (Table 3.2). These data suggest that p97 interacts with PNGase in a 3:1 molar ratio. Considering that p97 forms a stable hexamer, this ratio corresponds to one p97 hexamer binding two PNGase monomers. Since the isolated PUB domain interacts with full-length p97 and p97-C10 in a 1:1 molar ratio based on our biochemical and co-crystallization data (Table 3.2 and see below), the stoichiometry of the interaction of the full-length proteins is likely constrained by steric effects imposed by the core and/or C-terminal domains of mPNGase.

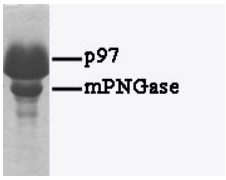


Figure 3.4 **One p97 hexamer binds to two mPNGase molecules.** The SDS-PAGE gel shows the purified p97 and mPNGase complex. The intensities of the two bands are 10968 and 2876 for p97 and mPNGase respectively. This figure has been contributed by Dr. Gang Zhao.

D. Crystal structure of the mPNGase PUB domain

Originally, the mPNGase 1-111 was used in crystallization. This protein crystallized under a condition of 1.85 M $(NH_4)_2SO_4$, 0.1 M Tris pH 7.0, 0.05 M Li_2SO_4. However, these crystals did not diffract, possibly because the N-terminal residues (1-10) are flexible. In order to overcome this problem, the N-terminal domain (residues 12 to 111) of mPNGase was purified and crystallized. The resulting crystals were rod-shaped with a longest dimension of 600-800 μm and a length of 20-25 μm perpendicular to that. They were obtained after about two weeks of growth at 18° C. The PUB domain crystals belong to space group C2 containing one molecule per asymmetric unit (Table 3.3) with Matthew's coefficients of 2.3 $Å^3$/Da, corresponding to solvent content of 46%. The structure was determined by single isomorphous replacement with anomalous scattering from a Hg-derivative and was refined at 1.7 Å resolution (Table 3.4) to an R-factor of 14.7% (R_{free} = 17.7%) with TLS refinement treating the entire domain as a single TLS body. All residues with the exception of Ser111 are well defined, and the resulting model is characterized by very good stereochemistry. The Ramachandran statistics indicate that 99% of the residues are in the most favored regions and 100% of the residues in allowed regions.

The PUB domain adopts a compact fold (Figure 3.5*A*), which mainly consists of a left-handed anti-parallel four-helical bundle (H1, H2, H4 and H5). Helices H2 and H4 are linked by an extended connection which is comprised of a 3_{10} helix, a β strand (β1) and the short α helix (H3). A twisted anti-parallel sheet is formed by β1 and two additional strands located between H4 and H5. Not surprisingly, this structure can be superimposed with the recently published structure of the human PNGase PUB domain [91] resulting in

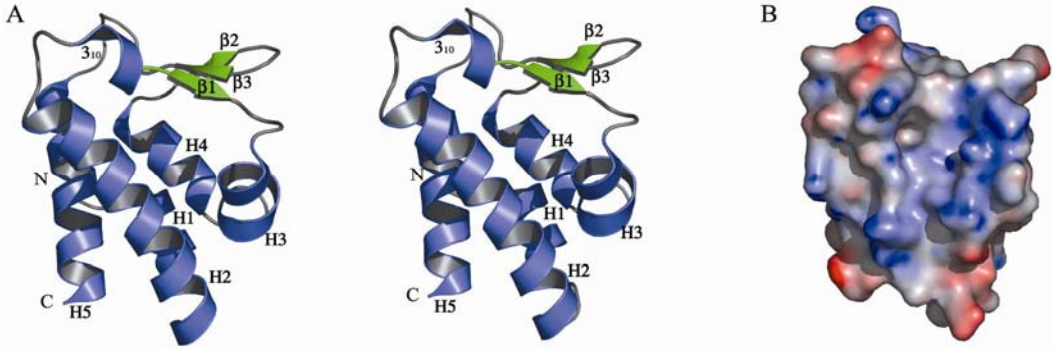


Figure 3.5 **Structure of the PUB domain.** (*A*) Ribbon diagram (stereo representation) of the mouse PNGase PUB domain. Helices are shown in blue, β-strands in green and loops in gray. Secondary structure elements and the termini are labeled. Figs. 2-4 were generated with PyMol [92]. (*B*) Surface representation of the PUB domain with electropositive patches in blue and electronegative areas in red. The electrostatic surface potential was calculated with APBS [93].

Table 3.3 **Data Collection and Structure Solution**

	PUB	**Hg**	**Complex**
Space group	C2	C2	C2
Cell dimensions (Å and °)	66.64, 52.04, 34.95, 90, 115.25, 90	66.6, 52.2, 35.2, 90, 115.5, 90	65.4, 52.9, 35.0, 90, 115.1, 90
Resolution limits (Å)	50 – 1.7	50 – 2.5	50 – 2.28
R_{sym}	0.055 (0.106)	0.114 (0.145)	0.061 (0.153)
Mean Redundancy	4.1	1.8	4.0
Completeness	0.950 (0.568)	0.970 (0.938)	0.996 (0.994)
<I/sigI>	30.5 (6.6)	8.5 (6.8)	16.3 (7.1)
Heavy atom sites	–	2	–
Phasing Power (centric / acentric)	0.90 / 1.25		–
Cullis R-Factor	0.59		–
FOM	0.49		–

$R_{sym} = \Sigma_{hkl}\Sigma_i |I_i - \langle I \rangle| / \Sigma_i \langle I \rangle$ where I_i is the i^{th} measurement and <I> is the weighted mean of all measurements of I. <I/sigI> indicates the average of the intensity divided by its standard deviation. Numbers in parentheses refer to the respective highest resolution data shell in each dataset. SIRAS phasing was performed to 2.5 Å resolution. Phasing power is the mean value of the heavy atom structure factor amplitude divided by the lack of closure for isomorphous / anomalous differences. R_{Cullis} is the lack of closure divided by the absolute of the difference between F_{PH} and F_P for isomorphous differences of centric data. FOM is the figure of merit given for all reflections.

Table 3.4 **Refinement Statistics**

	PUB	**Complex**
Resolution limits (Å)	20-1.7	20-2.28
Number of reflections	10,843	4,692
Number of protein / solvent atoms	779/155	819/76
R (R_{free})	0.148 (0.177)	0.195 (0.257)
Deviations from ideal values in		
Bond distances (Å)	0.014	0.006
Bond angles (°)	1.525	0.935
Chiral volumes (Å^3)	0.100	0.062
Planar groups (Å)	0.006	0.002
Torsion angles (°)	5.34, 36.81, 11.87	4.70, 36.97, 13.72
Ramachandran statistics	0.99/1.0	0.99/1.0
Average B factor (Å^2) (protein / (peptide) / solvent)	12.4/15.5	23.2/34.4/35.9

$R_{cryst}=\sum||F_o|-|F_c||/\sum|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes. R_{free} same as R for 5% of the data randomly omitted from refinement. Ramachandran statistics indicate the fraction of residues in the favored and allowed regions of the Ramachandran diagram as defined by MOLPROBITY [71].

a root mean square (rms) deviation in Cα positions of 1.0 Å. The major secondary structure elements of these two domains are very similar to each other except for two loop regions; one is located between H5 and β3 and the other between H2 and 3_{10} helix (Figure 3.6). In those regions rms deviations of up to 1.8 Å are observed and the resulting shifts result from differences in crystal phasing.

A surface representation of the roughly spherical molecule reveals a shallow groove with a width of 8 Å and a length of 16 Å in which a hydrophobic pocket is embedded. The top half of the groove is surrounded by highly conserved residues including Tyr38, Asp40, Asn41, Lys50, Tyr51, Ile54, Arg55, Asn58 and Phe61, which form an extensive hydrogen bond network with 12 water molecules and a glycerol molecule, which is coordinated by Arg55, Asn58 and His86. A calculation of the electrostatic potential of the PUB domain with the APBS program [93] reveals that positively charged residues including Lys33, Lys50, Arg55 and Arg64 are clustered around the groove (Figure 3.5*B*).

The PUB domain has also been found in several other proteins present in plants and mammals [40] (Figure 3.2B). Based on the crystal structure we can demonstrate that a number of the conserved residues of the PUB domain have important structural roles. For example, the side chains of Leu35, Ile42, Leu65, Leu75, Phe80 and Leu98 contribute to the hydrophobic core of the PUB domain, while Pro46, Gly79 and Pro90 are important for maintaining the conformations of loop regions. The indispensable role of Gly79 and Phe80 has been demonstrated in our previous mutagenesis study [21]. The G79A/F80A double mutant abolished the interaction between mPNGase and mp97. From our structure, both of the residues are far from the peptide-binding site. The side chain of

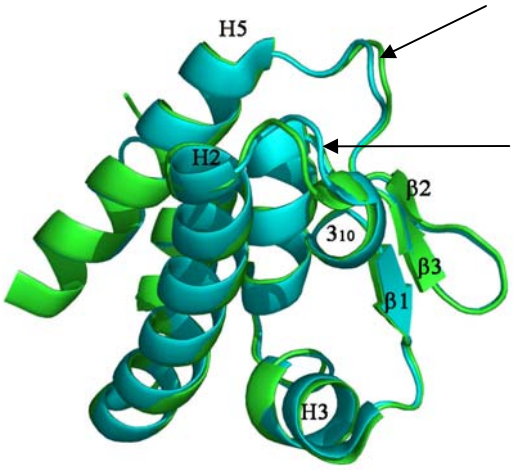


Figure 3.6 **Superposition of the PUB domains of human and mouse PNGase.** The PUB domain of mouse PNGase is shown in green and that of human PNGase is in cyan. Two loop regions (implicated by arrows) that have different conformations are between H5 and β3 as well as H2 and 3_{10}. Secondary structure elements, which are conserved between the two structures, are labeled.

Phe80 is not accessible to the solvent and mutating this residue to Ala could lead to conformational changes. In addition, several conserved hydrophilic residues, including Asn41, Arg55 and Asn58, are of functional importance and, as shown below, are directly involved in interactions with p97.

A search for structural homologs of the PUB domain with the DALI server revealed close matches with a hypothetical protein from *Vibrio cholerae* (Vc1899), to which no function has been assigned yet, and with the orange carotenoid protein with Z-scores of 5.6 and 5.3, respectively. These are followed by several members of the winged-helix (WH) family of DNA binding proteins, including Cdc6 and Orc2 with Z-scores of 5.0 and 4.0, respectively. The overall architecture of the PUB domain largely mimics the WH topology, except that H3 and H4 of the PUB domain are shorter and in a different orientation relative to the β sheet compared to typical WH domains (Figure 3.7). WH proteins are primarily involved in DNA recognition [94], and they bind to their targets via a positively charged surface. However, the location of this positively charged region and the one in the PUB domain differ, and it is therefore unclear whether the PUB domain evolved from an ancestral DNA-binding domain.

E. Crystal structure of the PUB domain in complex with a p97 derived peptide

To further investigate the interaction between p97 and PNGase, we co-crystallized the PUB domain of mouse PNGase with the p97-C10 peptide (Tables 3.3 and 3.4). The last four residues of the peptide, with the sequence Asp-Leu-Tyr-Gly (residues 803 to 806 of p97), are well-defined in the electron density (Figure 3.8). In addition weak density is present for the preceding residue (Asp802), and this residue has been

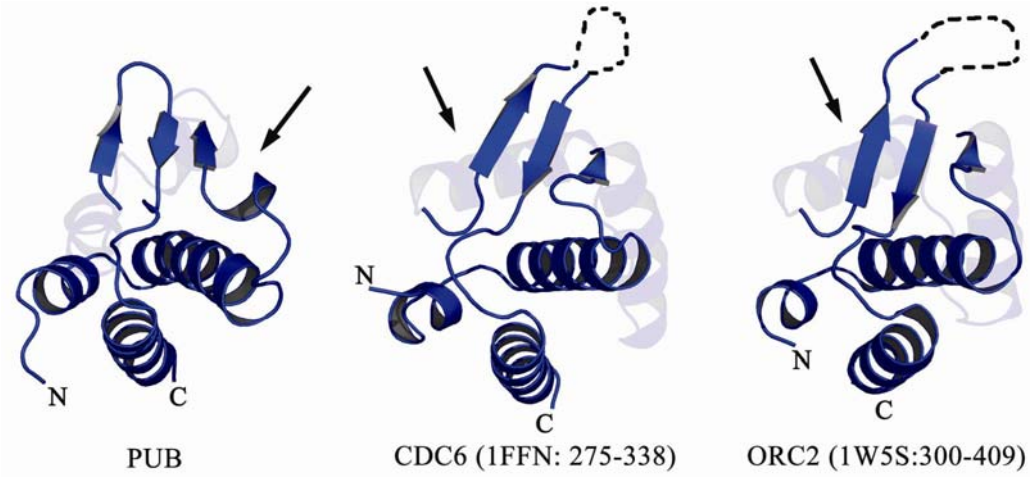


Figure 3.7 **Relationship of the PUB domain and Winged-Helix (WH) DNA-binding domains of Cdc6 (PDB entry: 1FFN, residues 275-388) and Orc2 (PDB entry: 1W5S, residues 300-409).** The homologous segments between the PUB and WH domains are shown in solid blue, whereas divergent regions are rendered transparent. The p97 binding site for the PUB domain and the DNA binding sites of Cdc6 and Orc2 are indicated by arrows.

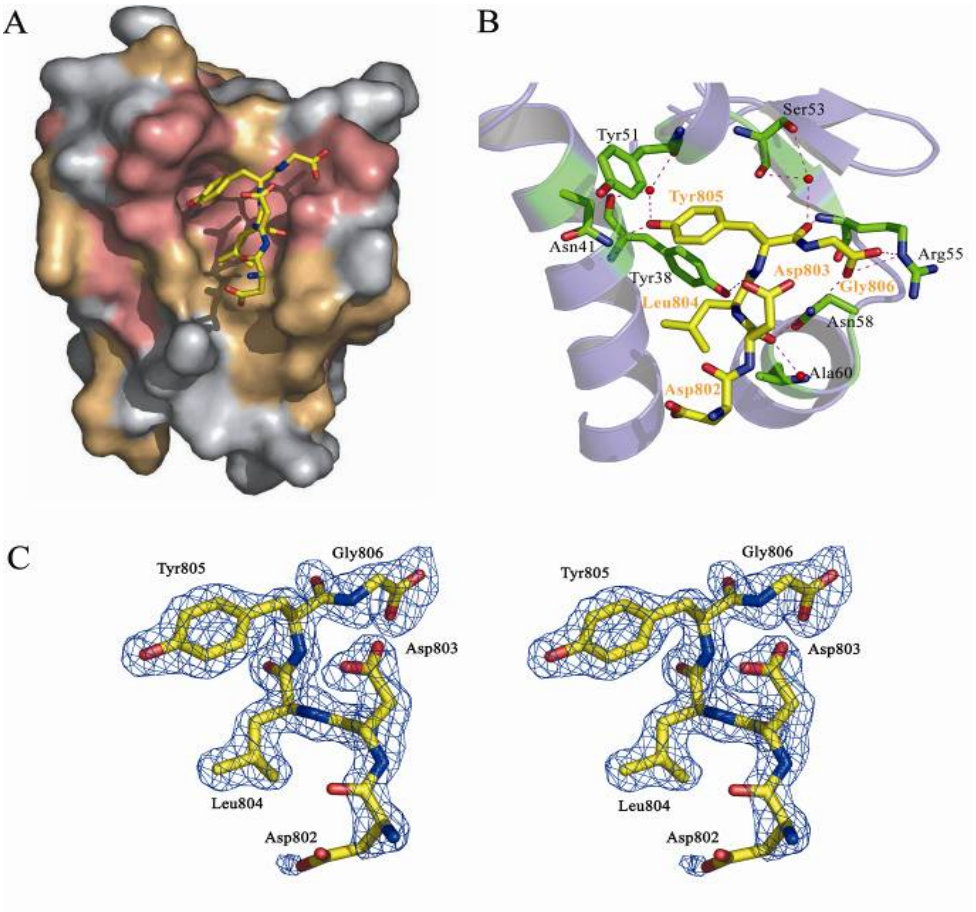


Figure 3.8 **Interaction between the PUB domain and the C-terminal residues of p97.** *(A)* Surface representation of the PUB domain with the peptide (carbon, nitrogen and oxygen atoms in yellow, blue and red, respectively) in stick representation. The surface of the PUB domain is colored according to sequence homology with highly conserved residues in pink, moderately conserved residues in orange and non-conserved residues in gray. (*B*) Detailed view of the interactions between the PUB domain and the peptide. The PUB domain is shown in a ribbon diagram with the p97 interacting residues shown in stick representation. The p97 peptide is colored as in (*A*). Hydrogen bonds are shown as red dashed lines and water molecules, which mediate interactions, as red spheres. (*C*) $2F_o$-F_c omit electron density map of the p97 C-terminal five residues contoured at a level of one times the r.m.s. deviation.

tentatively included in the model. Binding of the p97 peptide does not perturb the overall structure of the PUB domain as evidenced by the fact that the two structures can be superimposed with a root mean square (r.m.s.) deviation of 0.33 Å in Cα positions. Notable differences between the apo-PUB domain and the PUB domain in the complex are only observed for the side chains of Glu26, Lys33 and Arg64 (Figure 3.9).

As expected from the electrostatic surface potential of the PUB domain, p97-C10 binds to the positively charged groove on the surface of the PUB domain formed by H2-H4 and β1. Complex formation is mediated by both ionic and hydrophobic interactions. The C-terminal carboxylate of the peptide interacts with the side chain of Arg55, which is highly conserved and only replaced by a lysine in the PUB domain of a mouse protein (UBXD1) of unknown function [95] (Figure 3.2B). Extensive hydrogen bonded interactions are formed between residues Asp803, Leu804, Tyr805 and Gly806 of p97 and residues Tyr38, Asn41, Arg55 and Asn58 of the PUB domain (Table 3.5). To probe the importance of the ionic interactions, three basic residues, Arg55, Lys50 and Arg64, around the p97 binding site of the PUB domain were mutated to glutamate and their effects on the p97 interaction were analyzed in ITC experiments. The K50E and R64E mutants of the PUB domain showed reduced affinities, while the R55E mutant completely lost its ability to interact with p97-C10 and full-length p97 (Table 3.2). This is in agreement with the high degree of sequence conservation of these residues, especially that of Arg55 (Figure 3.2*B*).

The second and third residues at the C-terminus of p97, which are leucine and tyrosine in all eukaryotes (Figure 3.2*A*), are engaged in hydrophobic interactions with the PUB domain. Leu804 of p97 is in van der Waals contact with Tyr38 and Leu34 of the

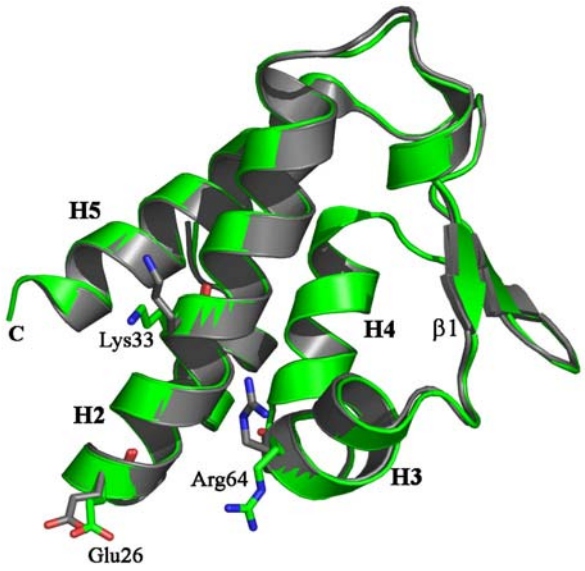


Figure 3.9 **Superposition of the apo and complex PUB domain structures.** The apo structure is shown in grey and the complex in green. The residues that have side chain conformational changes are shown in stick representation and are labeled.

Table 3.5 H-bonds between p97-C10 and PUB domain of mPNGase

Direct protein-protein interactions					
P97	**Atom**		**PUB**	**Atom**	**Distance (Å)**
Gly806	OT1 OT2 OT2		Arg55 Arg55 Asn58	Nε Nε Nδ2	2.7 3.3 2.9
Tyr805	O OH		Arg55 Asn41	N Nδ2	3.3 2.7
Leu804	O		Tyr38	OH	3.1
Asp803	O		Asn58	Nδ2	2.9
H_2O mediated interactions					
P97	**Atoms**	**H_2O**	**PUB**	**Atom**	**Distance (Å)**
Tyr805	O	#1	Ser53 Ser53	Oγ O	2.6/3.1 2.6/3.0
Tyr805	OH	#2	Tyr51 Tyr38	O O	2.7/2.6 2.7/2.7
Asp803	O	#3	Ala60	N	2.7/3.0

PUB domain, and its main chain oxygen atom hydrogen bonds with the side chain of Tyr38 (Figure 3.8). Interestingly, the side chain of Tyr805 snugly fits into the aforementioned hydrophobic pocket formed by residues Tyr38, Asn41, Lys50, Tyr51, Ser53 and Ile54 of the PUB domain (Figure 3.8*A* and 3.8*B*). A water molecule is also present in the Tyr805 binding pocket and bridges the phenolic hydroxyl group of Tyr805 to the surrounding residues.

The critical location of Tyr805 in the complex structure prompted us to further explore the possibility that modification of this residue might abolish complex formation. Tyr805 was mutated to either Phe or Glu in full-length p97, and its interaction with mPNGase was tested. The GST pull-down assay showed that only wild-type His-tagged p97, but neither the Y805E nor the Y805F mutants, were bound to GST-mPNGase (1-130) (Figure 3.3A). This result confirmed the indispensable role of the C-terminal motif of p97, especially Tyr805, in its interaction with the PUB domain.

F. Tyr805 phosphorylation abolishes the p97-PUB domain interaction

Phosphorylation of p97 has been observed during sperm capacitation [96], DNA repair [97], transitional endoplasmic reticulum assembly [98] and T-cell receptor activation [99]. In addition, phosphorylation of p97 has been shown to affect its association with ubiquitinated proteins and its nuclear localization [100, 101]. Interestingly, the primary site of phosphorylation during T-cell stimulation [102] is the highly conserved residue Tyr805. In light of the pivotal role of Tyr805 in interaction of p97 with PNGase, we compared the binding affinity of phosphorylated and non-phosphorylated p97-C10 peptides to the PUB domain. In agreement with our hypothesis,

phosphorylation of Tyr805 completely abolished p97-C10's interaction with the PUB domain (Figure 3.3 and Table 3.2). Furthermore, phosphorylation of Tyr805 also eliminated p97-C10's interaction with Ufd3 (Table 3.2). Although structural details of the p97-Ufd3 interaction are unavailable, we speculate that p97 interacts with the PUB domain and with Ufd3 via a similar mechanism, even though there is no obvious sequence similarity between these p97 binding partners.

IV: Discussion

Proper ERAD activity is essential for homeostasis and cell survival. ERAD malfunctions have been associated with various pathophysiological conditions, such as Alzheimer's and Parkinson's disease, as well as with cystic fibrosis [103, 104]. In addition, it has been shown that the ERAD pathway is "hijacked" by human cytomegalovirus (HCMV) during viral infection. The virus compromises the host's immune system by encoding two proteins, US2 and US11, which bind to the major histocompatibility complex (MHC) class 1 heavy chain and direct it to the ERAD pathway for degradation [14, 105]. The biological and potential therapeutical importance of ERAD emphasizes how important it is to have a clear understanding of its regulation.

p97 plays an essential role in the ERAD process, in which it extracts unfolded substrates through the putative retrotranslocon and recruits ERAD components to form a degradation complex [21, 48]. It has been proposed that the p97 substrates are first recognized by substrate-recruiting cofactors and then processed by substrate-processing cofactors, including Ufd2, Ufd3 and PNGase [49]. For the first time, we identified a highly conserved binding motif of p97, which mediates its interaction with at least two substrate-processing cofactors, Ufd3 and PNGase.

Furthermore, tyrosine phosphorylation of this motif abolished the interaction of p97 with both cofactors. Based on these observations, we hypothesize that tyrosine phosphorylation / dephosphorylation of p97 may dynamically modulate ERAD activity, affecting both the deglycosylation activity of PNGase and the ubiquitination functions of the Ufd proteins downstream, which lead to either an accumulation or quick degradation of ERAD substrates in the ER lumen, respectively. In this context it is interesting to note

that p97 has been identified as a substrate of protein-tyrosine phosphatases, such as PTPH1 (PTPN3) [106] and PTPN22 [107]. Because protein-tyrosine phosphatases, including PTP1B, potentiate IRE1 signaling during endoplasmic reticulum stress [108], we speculate that under these conditions p97 is de-phosphorylated so as to recruit components of the degradation complex to p97, including PNGase and the Ufd proteins.

Unassembled T cell receptor (TCR) α and CD3δ are well-characterized substrates of the ERAD pathway and have been used as established model systems in protein degradation studies [46, 109]. Although it has been observed that Tyr805 contributes to more than 90% of p97 tyrosine phosphorylation during T cell receptor activation [102], the rationale for p97 phosphorylation has been unclear so far. Based on our hypothesis it is possible that the TCR activation-induced p97 tyrosine phosphorylation down-regulates ERAD activity. As a result, more TCR subunits and CD3δ molecules could survive protein degradation and thereby assemble on the cell surface.

The novel C-terminal protein-binding motif of p97 may be utilized by other cofactors to regulate p97's function and/or direct it to various cellular processes. In fact, the PUB domain has been found in several proteins with unknown functions [40], such as UBXD1 in mammals and AtPUB1 through AtPUB4 in *Arabidopsis thaliana* (Figure 3.2B). UBXD1 is intriguing in that, in addition to the PUB domain, it also contains a UBX domain which was shown to interact with the N-terminal region of p97 [46, 110]. Possessing two p97 interacting domains, each binding to a different part of p97, UBXD1 would be expected to constrain any conformational changes of p97 (if the two domains bind to the same p97 molecule) or cross-link multiple p97 molecules into a supramolecular assembly. The fact that p97 employs a rather short and unstructured motif

to interact with proteins with no obvious sequence similarities (PUB domain, Ufd3 and possibly Ufd2) implies the versatility of this motif and raises the possibility of additional protein interaction partners.

Chapter 4

The C-terminal 15 residues of Derlin1 bind to the N-terminal domain of p97

The material presented in this chapter has not yet been published. All experiments described here have been carried out by myself.

I. Introduction

P97 interacts with a wide variety of partner proteins (see chapters 1+3 for details), including Derlin1, a recently identified integral membrane protein located in the ER membrane [11, 13]. Derlin1 contains 251 residues in total, and its architecture features four transmembrane spanning helices with both of its termini located in the cytosol (Figures 4.1 and 1.4). In mammals Derlin1 is accompanied by two homologs, named Derlin2 and Derlin3. In contrast, yeast only contains two Derlin homologs, Der1 and Dfm1. Derlin1 has been found to associate with VIMP [11], an ER membrane protein that is predicted to span the membrane once, resulting in a 25-residue N-terminal region in the ER and a larger extension (132 residues) in the cytosol. The Derlin1 and VIMP complex was proposed to participate in the retrotranslocation of ERAD substrates [11] and both proteins interact with p97 [11, 56, 111]. Recently, evidence has accumulated which suggests that the Derlin1 protein could be a component of the retrotranslocon, however, the retrotranslocation of some ERAD substrates is not dependent on Derlin1 [12]. Clearly more studies are needed to fully understand the function of Derlin1 and its paralogs Derlin2 and 3.

To identify additional interacting partners we searched for human Derlin1 (hDerlin1) interacting proteins with a pull-down assay using its C-terminal domain (residues 187-251) as bait. The only Derlin1 binding partner identified with confidence by this approach was p97, an interaction, which had already been reported [56]. Although the p97-Derlin1 interaction was already known, the detailed interacting regions in both proteins have not been mapped. To narrow down the binding sites in p97 and Derlin1, intein pull-down assays and ITC experiments were performed using various fragments of

p97 and the hDerlin1 C-terminal region. These experiments revealed that the last 15 residues of the Derlin1 and the N-terminal domain (residues 1 to 199) of p97 directly interact with each other with a dissociation constant of around 15 μM. In order to describe the atomic details of the interaction between the two proteins, the ND1 domain of p97 (residues 1 to 458) and the last 15 residue of the hDerlin1 were cocrystallized, however, the resulting crystals need to be further optimized before the structure of the complex can be determined.

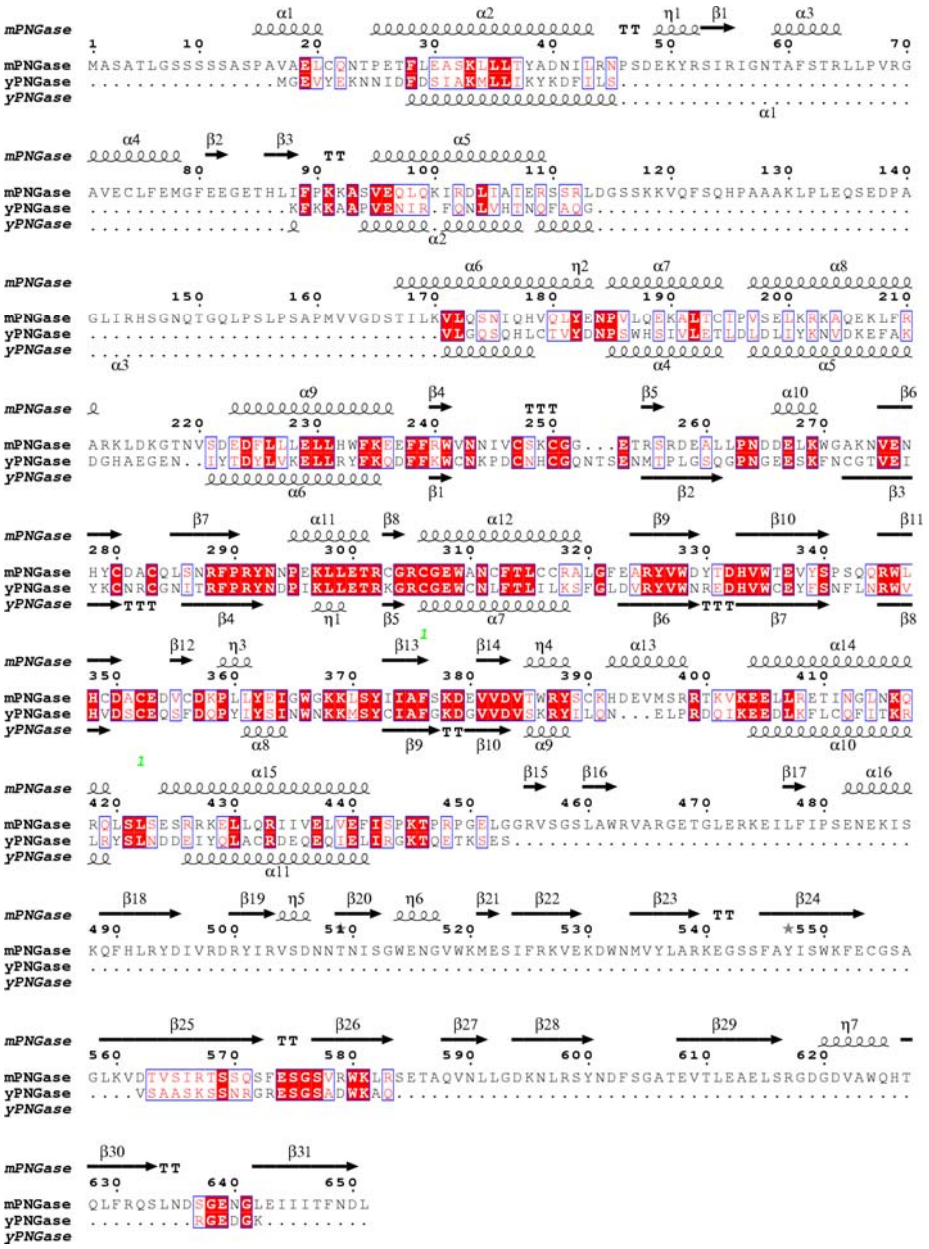


Figure 4.1 **Multiple sequence alignment of Derlin1 homologs**. *H.s. Homo sapiens*; *M. m., Mus musculus*; *C. e.,Caenorhabditis elegans*; *D. m., Drosophila melanogaster; S. c., Saccharomyces cerevisiae*. This figure was modified from Ye [11].

II. Materials and Methods

A. Constructs

The gene fragment encoding residues (187-251) of hDerlin1c was amplified from full-length Derlin1 fused to green fluorescent protein (GFP); this construct was kindly provided by Dr. Guangtao Li from the Lennarz lab. All primers used in these experiments are listed in Table 4.1. hDerlin1c was cloned into the NdeI/SapI sites of the pTXB1 vector. This particular PTXB1 vector has been modified by replacing its ampicillin resistance with that of kanamycin to improve vector stability (this construct was kindly provided by Dr. Gang Zhao). The hDerlin1c sequence was subsequently cloned into the SapI/XhoI sites of the pTYB12 vector and the NdeI/XhoI sites of the pET21b vector. In the latter construct (pET21b) the His_6-tag is located at the N-terminal end of the protein, immediately following the NdeI site, to not block the C-terminus and to somewhat mimic the primary structure of the Derlin protein. The smaller Derlin1 fragments hDerlin1-p1 (187-219) and hDerlin1-p2 (220-251) were generated with the QuikChange site-directed mutagenesis kit, using pTYB12-hDerlin1c as the template. Cloning of all p97 fragments has been performed by Dr. Gang Zhao (see Chapter 3 for details).

B. Protein Expression and Purification

All Derlin1-derived protein fragments were transformed into BL21DE3 Codon Plus RIL cells (Ampicillin and Chloramphenicol resistance). The transformation and expression protocols are the same as previously described (see Chapter 2 for details). The proteins were expressed overnight in BL21DE3 Codon Plus RIL cells at 15° C after induction with 0.3 mM IPTG at an A600 of 0.6–0.8. After growth, cells were harvested

Table 4.1 Primer sequences:

Primers		Sequence
Derlin187-251 /pTXB1	5’ 3’	TCAGACATATGTTCAGATACCCAATGGACTTG CTCGAGGCTCTTCCGCAACCCTGGTCTCCAAGTCGAAAGCC
Derlin187-251 /pTYB12	5’ 3’	TCAGACATATGTTCAGATACCCAATGGACTTG CATGCTCGAGTCACTGGTCTCCAAGTCGAAAGCC
Derlin187-251 /pET21b	5’ 3’	GGAATTCCATATGCACCACCACCACCACCACATGTTCAGATA CCCAATGGAC CCG CTCGAG TTA CTGGTCTCCAAG TCG
Derlin187-219 /pTYB12	5’ 3’	GGAGAGGAGGAGTATCAGGATAACTCGAGCCCGGGTGACTG CAGTCACCCGGGCTCGAGTTATCCTGATACTCCTCCTCTCC
Derlin220-251 /pTYB12	5’ 3’	GTTGTTGTACAGAATGCTGGTTTTGGTGTGCCCCCTGCTAGC GCTAGCAGGGGGCACACCAAAACCAGCATTCTGTACAACAA C

by centrifuging at 9000 g for 15 minutes at 4 ºC and lysed in 20 mM Tris/HCl, pH 8.5, 150 mM NaCl, a protease inhitor cocktail (Complete mini, EDTA free, Roche) by passing them twice through a French pressure cell at 1,500 psi. The cell lysate was then centrifuged at 50000 g for 30 minutes at 4° C. The cleared cell lysate of the intein-tagged proteins was incubated with 4 ml chitin beads (New England Biolabs, Ipswich, MA) already packed in a glass column by end over shaking for 30 minutes at 4º C and washed with 20 cv (column volume) of wash buffer containing 20 mM Tris/HCl, pH 8.5, 1 M NaCl. This was followed by equilibrating the bound proteins with cleavage buffer (wash buffer plus 100 mM DTT). The beads were incubated overnight at room temperature before elution with 20 mM Tris/HCl, pH 8.5, 150 mM NaCl. The eluted protein solution was examined by 15% SDS-PAGE, and the fractions containing cleaved hDerlin1c (usually corresponding to about 3 cv) were pooled together and diluted 5 times to lower the salt concentration of the protein buffer before loading it onto a MonoS 10/10 cation exchange chromatography column (Amersham, Piscataway, NJ). The Derlin C-terminal peptide was eluted using a NaCl gradient (Buffer A: 20 mM Tris/HCl, pH 8.5; Buffer B: 20 mM Tris/HCl, pH 8.5, 1 M NaCl). Fractions were analyzed by SDS PAGE and stained with Coomassie Blue. Fractions without contaminating bands were combined and the protein was concentrated to 80-130 μM (corresponding to 0.5-0.9 mg/ml) with Vivaspin-20 3,000 MWCO microconcentrators (VIVAscience).

Expression and cell harvest of the His_6-tagged Derlin fragment (residues 187 to 251) was carried out following the same protocol described for the intein-tagged Derlin fragment with the exception of the affinity chromatography step. Purification of this protein was initiated with Ni-NTA superflow (Qiagen, Valencia, CA) affinity

chromatography. The cleared cell lysate was incubated with 1 ml Ni beads pre-packed in a 20 ml plastic column for 30 minutes at 4° C in binding buffer consisting of 20 mM Tris/HCl, 150 mM NaCl and 10 mM imidazole. The column was washed with 10 cv of binding buffer. The bound proteins were eluted with a buffer containing 20 mM Tris/HCl, 150 mM NaCl, and 250 mM imidazole and analyzed by SDS PAGE. 95% pure fractions were pooled, diluted 5 fold with 20 mM Tris-HCl pH 8.5 and applied to the MonoS column described above. MonoS cation exchange chromatography followed the exact same protocol as described above. Purification of the p97 fragments has been performed by Dr. Gang Zhao (see Chapter 3 for details).

C. *In vitro* binding assay

Intein alone, intein-Derlin1c, intein-Derlin-p1 and intein-Derlin-p2 fusion protein extracts were immobilized by incubated the cell lysate resulting from a 0.5 L growth with 200 µl of a chitin bead suspension in binding buffer (20 mM Tris pH 8.5, 150 mM NaCl, 0.5% Triton X-100) at 4° C for 30 minutes. HEK (human embryonic kidney) 293 cells were kindly provided by Dr. Gangtao Li. The HEK293 cell lysate was prepared by sonicating 5 x 10^6 cells, which had been resuspended in 2 ml binding buffer, three times for 10 s with a sonicator (Sonicator 3000, Misonix) at a power setting of 1 using a microtip. About 30 µl of the respective chitin bead preparation were incubated with either the HEK293 cell lysate (overnight at 4° C) or 100 µg of purified p97 fragments (30 minutes at 4° C). Beads were washed six times with binding buffer and bound proteins were eluted with SDS-PAGE loading buffer. The protein mixtures were resolved by 12% SDS-PAGE and stained with Coomassie Blue as described before.

To quantify the interaction between p97 and Derlin1, different p97 and Derlin fragments were investigated by ITC. p97 and Derlin1 fragments were dialyzed overnight at 4 ºC against a buffer containing 20 mM Tris pH 8.5 and 150 mM NaCl. 80-150 μM of Derlin1c were titrated as the ligand into the sample cell containing 10-20 μM of p97 fragments. N18 (Derlin1 220-237) and C15 (Derlin1 237-251) (Anaspec, San Jose, CA) peptides were derived from the C-terminus of hDerlin1 and chemically synthesized. They were purified by HPLC using a C18 reverse column. The purity of both peptides is larger than 95%. About 200 μM of the N18 (sequence: FGVPPASMRRAADQNGGG) or C15 (sequence: GGRHNWGQGFRLGDQ) peptides was titrated to a 10 μM solution of the ND1 fragment of p97. 10 μl of ligand was added each time with 30 injections in total. All ITC experiments were performed using a VP-ITC instrument (MicroCal, Northampton, MA) at 25 ºC. The data were analyzed using ORIGIN assuming a single-site binding model.

D: Mass spectrometry

Fractions from the pull-down assay were resolved on a NOVEX NuPage 4-12% gradient gel (Invitrogen) and stained with colloidal coomassie blue (Pierce GelCode Blue Stain Reagent, cat.# 24590). The proteins of interest were cut out and digested with trypsin at 37° C for 4-24 hours before Ziptip cleanup using a C-18 zip tip (Millipore). The purified samples were injected onto a Zorbax 300SB-C8 HPLC column connected on-line to an Agilent XCT Ion Trap mass spectrometer. The identified peptide sequences were analyzed by the GPM searching engine (http://h319.thegpm.org/tandem/thegpm_tandem.html). The detailed procedure and

reagents were the same as described by Aleksandra Nita-Lazar and Robert Haltiwanger [112].

To determine the actual molecular mass of the purified Derlin1c peptide after Mono S cation exchange chromatography, 0.3 µg of protein were used and cleaned by Ziptip before injecting into the mass spectrometer. The ESI-MS (Liquid chromatography electrospray ionization mass spectrometry) generates cations ions in multiple charged states, and the resulting charge state envelop was de-convoluted to derive the molecular mass of the peptide.

E: Crystallization

The complex of the p97 ND1 domain and C15 peptide was prepared by incubating separately purified p97ND1 (17 mg/ml) and C15 peptide (0.6 mg/ml) for 30 minutes on ice. Two crystallization screens: Topaz-Optimix 1 (Fluidigm) and the CSS II (MD clear strategy screen II from Molecular Dimensions) were used for initial screening with a Honeybee crystallization robot (Genomic Solutions). Two promising conditions were obtained. The optimization step of these conditions was carried out by the hanging drop vapor diffusion method in 24 well Limbro plates. The hanging drop is prepared by adding 1 µl of protein complex and 1 µl of reservoir solution. Bigger crystals were obtained after optimization.

III. Results

A. Searching for new binding partners of Derlin1

In order to identify possible soluble binding partners of the C-terminal region of hDerlin1 (hDerlin1c), a pull-down assay using immobilized intein-hDerlin1c was carried out. Cell lysates derived from HEK293 cells were incubated with chitin beads containing the conjugated hDerlin1c. The incubation was conducted at 4 ºC and overnight to minimize protein denaturation on one hand, and to increase the signal from a possible interaction on the other hand. The intein-tag alone derived from the pTXB1 vector was also expressed and used as a negative control to check for nonspecific binding in this experiment. On NOVEX NuPage 4-12% gradient gels two additional bands were (barely) visible after hDerlin1c was incubated with the HEK293 cell lysate (Figure 4.2). The first band showed an apparent mass of around 80 kDa, while the second band runs at around 120 kDa.

To identify the two proteins detected in the previous pull-down assay, gel pieces that contained the proteins of interest were excised and solubilized, the proteins were digested with trypsin and analyzed by LC-ESI-MS. By database searches, the 80 kDa band was identified as tubulin, an abundant cytoskeletal protein. Since we also detected the presence of tubulin in the control lane at the corresponding size(Figure 4.2 lane 2), tubulin might bind to hDerlin1c in a non-specific fashion. The second band was identified as p97, which was already known to bind to Derlin1.

B. Mapping the binding region in p97

Although p97 has been identified to bind to the C-terminal region of Derlin1, the

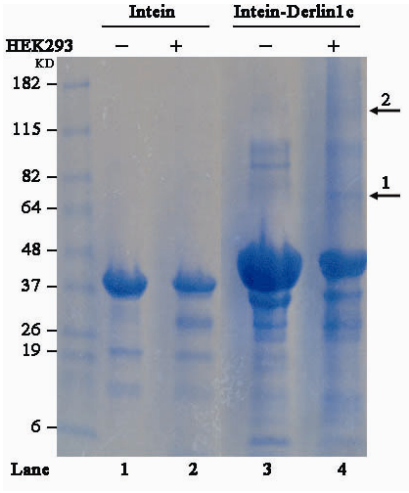
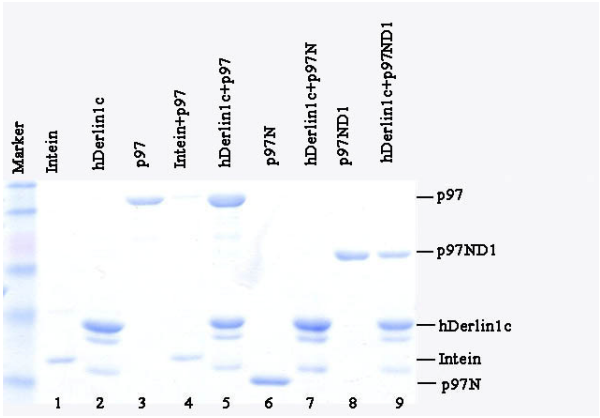


Figure 4.2 **Identification of Derlin1c interacting proteins.** Lanes 1 and 2: Intein only before and after incubation with the HEK293 cell lysate, respectively. Lanes 3 and 4: Intein-tagged Derlin1c in the presence and absence of the HEK293 cell lysate. Derlin1c binding partners are indicated by arrows. The molecular masses of the marker proteins are indicated on the left.

minimum binding region of both proteins was previously unknown. Since p97 contains four different domains and hDerlin1c is relatively small with only 64 residues, we hypothesized that hDerlin1c only binds to a small region of p97. In order to narrow down the interacting region in p97, different purified p97 fragments were used to incubate with hDerlin1c, which had been immobilized on chitin beads. Besides the full-length protein the following fragments were tested: the N-terminal domain (N) consisting of residues 1 to 199 and the N-terminal domain plus the D1 domain (ND1), which encompasses residues 1 to 458. The interaction was again probed with the same pull-down assay and the intein protein as negative control. In the gel, the full-length and the ND1 proteins clearly showed an interaction with hDerlin1c (Figure 4.3*A*, lanes 5 and 9), whereas the N-terminal domain alone did not show any binding (Figure 4.3*A*, lane 7). The intein control showed a very weak p97 band, indicating the presence of very weak, non-specific binding between the intein and p97. However, when we conducted ITC experiments to confirm the interactions, the results were different (Figure 4.3*B*). Full-length p97 (data not shown) and the N domain clearly showed binding to hDerlin1c, while a p97 construct in which the N-terminal domain was deleted (ΔN domain containing residues 200-806) does not display any binding signal. Based on the ITC results, it appears as if the N-terminal domain of p97 is the primary domain responsible for binding to hDerlin1c. At present, we do not have a satisfactory explanation for the inability of the N-terminal domain to interact with hDerlin1c in the pull-down assay. Both the pull-down assay and the ITC experiment have been repeated twice and the results were consistent.

A



B

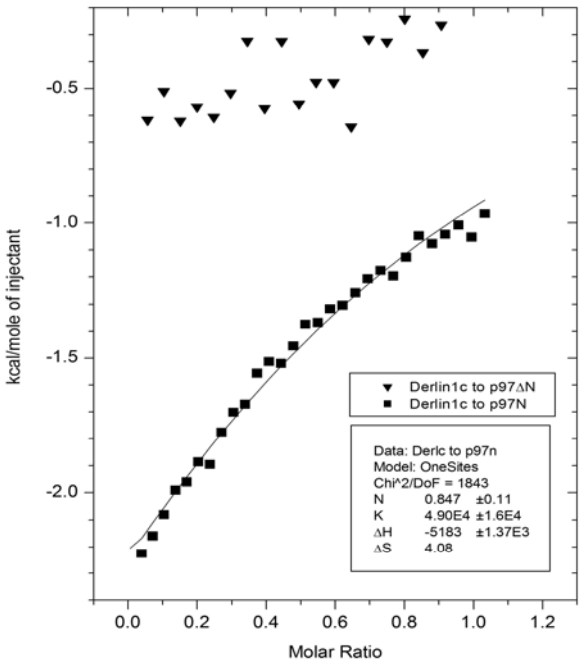


Figure 4.3 **The N-terminal domain of p97 binds to Derlin1.** *(A)* Intein pull-down assay. Different fragments of p97 were incubated for 30 minutes with intein-tagged hDerlin1 in the presence of chitin beads. After stringent washing with 20 mM Tris pH8.5, 150 mM NaCl, 0.5% Triton X-100, the proteins were eluted with SDS loading buffer and resolved by SDS-PAGE. *(B)* ITC measurements. Binding isotherms are shown for Derlin1c to a p97 construct without its N-terminal domain [p97ΔN (triangles)] and to the p97 N-terminal domain (squares). The units of the parameters are M^{-1} for K, cal/mol for ΔH, cal/mol/degree for ΔS.

C: Problems with the ITC measurements

Although we detected a clear exothermic interaction signal in ITC experiments involving various p97 fragment (full-length, ND1 and N) and hDerlin1c, we were unable to satisfactorily fit a binding isotherm to the experimental data. There are a couple of factors in this case which complicate the ITC analysis: 1) The Derlin1c protein may have several conformations due to the lack of defined tertiary structure in the absence of a binding partner. 2) hDerlin1c is not completely pure. The second complication makes it impossible to accurately calculate the concentration of hDerlin1c, whereas the first would complicate data analysis due to the fact that a competing process, the folding of hDerlin1c would overlay the binding signal. Consequently, modeling of the data becomes rather complex and the calculated binding parameters could be systematically wrong. In order to circumvent these problems, more protein and protein with a higher purity was needed. Purification of hDelin1c was attempted using various constructs including two intein-tagged proteins (in the pTXB1 and pTYB12 vectors) and one His_6-tagged hDerlin1 (in pET21b vector). However, neither of them yielded enough pure protein for ITC experiments.

D: Mapping the p97 binding region in hDerlin1c

To narrow down the binding region of p97 in hDerlin1c, two new constructs, pTYB12-hDerlin-p1 (187-219) and pTYB12-hDerlin-p2 (220-251), were designed for use in the pull-down assay. These studies revealed that the intein fused p2-peptide corresponding to the C-terminal region was still able to pull down full-length p97, while no binding was detected for the p1-peptide (Figure 4.4*A*, lane 9 versus lane 7). The

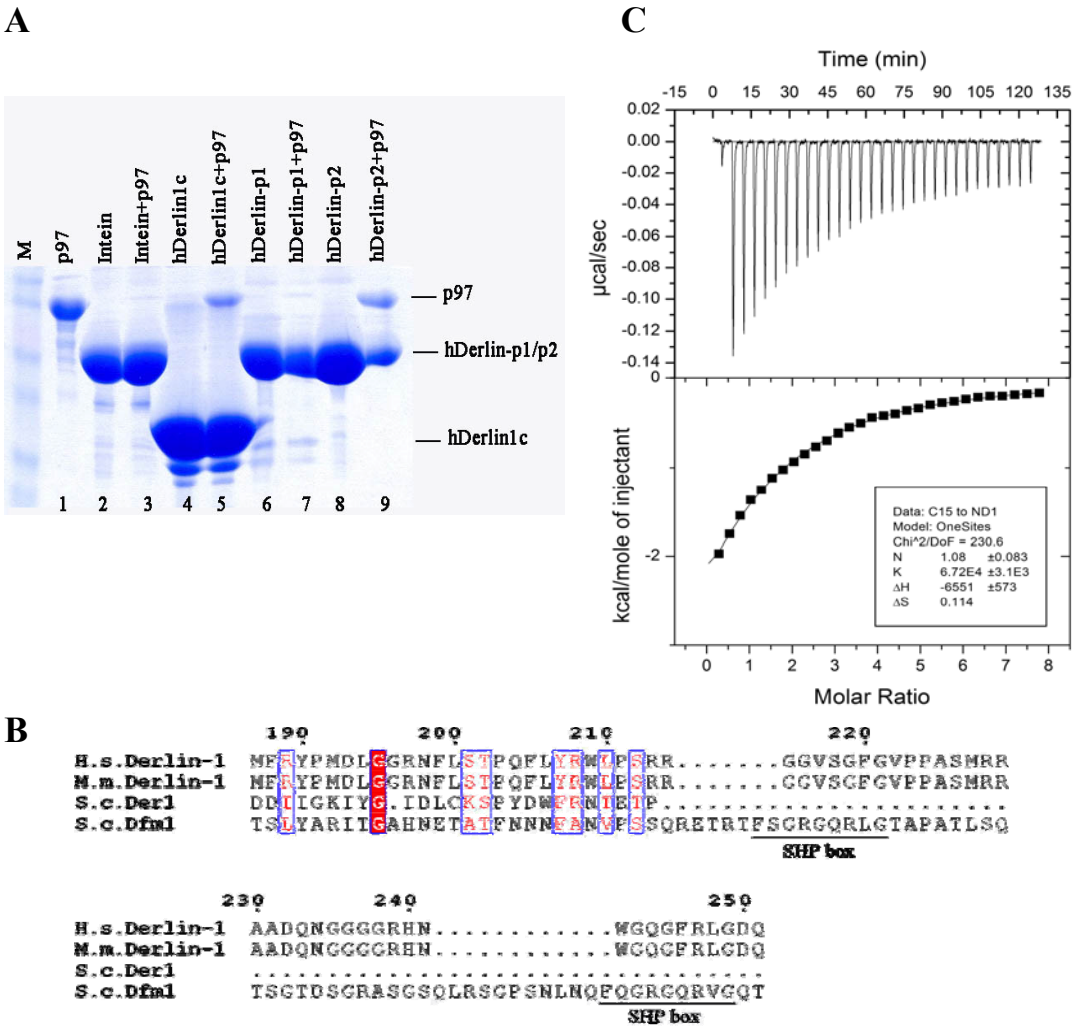


Figure 4.4 **The last 15 residue of Derlin1 are sufficient to bind to p97.** *(A)* Intein-pull down assay. P1 (187-219) and p2 (220-251) peptides were cloned into the pTYB12 vector, resulting in a fusion with a large intein domain, while the Derlin1c is fused to a smaller intein domain. Only the p2 peptide binds to p97, but not the p1 peptide. *(B)* Sequence alignment of human and mouse Derlin1 and their homologs in yeast. H.s: *H. sapiens*; M.m: *M. musculus*; S.c: *S. cerevisiae*. SHP boxes in the Dfm1 protein are labeled. Strictly conserved residues are highlighted by red boxes, whereas type-conserved residues are shown in red. *(C)* ITC measurement. The synthetic peptide containing the last 15 residue of Derlin1 was titrated to a solution containing the p97 ND1 domain. The thermodynamic parameters derived from fitting the binding isotherm are listed in the box. The units of the parameters are M^{-1} for K, cal/mol for ΔH, cal/mol/degree for ΔS.

p2-peptide consists of 32 residues with a central Gly_4 motif. Following these four glycines is the sequence "WGQGFRLG", which displays limited similarity to the SHP box. The SHP box is named after the Shp1 protein that contains this box [55]. It features the repeating pattern, FxGzGQxb (x: any amino acid, b: basic, z: acidic). Proteins containing this box are involved in processes which are regulated by either ubiquitin or ubiquitin-related protein modifiers. The sequence alignment (Figure 4.4*B*) reveals that this sequence aligns with the second SHP box with the sequence FQGRGQRVG in the Dfm1 protein in such a way that the three glycines and the one arginine are identical. In addition, the hydrophobic residue between the arginine and the last glycine is type-conserved. The Dfm1 protein is involved in ER homeostasis and both of its SHP boxes were shown to be necessary for its interaction with p97 [57]. These properties led us to hypothesize that these eight residues contain the major binding sequence. If this is true, the last 15 residues of hDerlin1c would be sufficient to mediate the interaction with p97.

In order to confirm our hypothesis, we used two synthetic peptides, which will be referred to as N18 and C15. N18 contains the sequence 220-237, while C15 contains the last 15 residues of hDerlin1 so that together they cover the entire p2 sequence. ITC experiments showed that N18 does not interact with ND1 domain of p97 (data not shown), whereas C15 showed binding to the ND1 construct with a dissociation constant (K_d) of about 15 μM (Figure 4.4*C*) using a one binding site model. Therefore, the last 15 residue of hDerlin1 are responsible for mediating the interaction with p97. In this study we used the ND1 fragment instead of the isolated N-terminal domain because the ND1 domain can be obtained with higher purity and displays higher stability presumably due

to the formation of the characteristic p97 hexamer, which is dependent on the D1 domains.

E: Crystallization trials of the ND1 domain of p97 in complex with the C15 peptide of hDerlin1

In order to investigate the detailed interaction between p97 and hDerlin1c, crystallization of the p97 ND1 domain in complex with the C15 peptide was attempted. In an initial screening attempt with a Honeybee crystallization robot utilizing the Topaz-Optimix 1 (Fluidigm) and the CSS II (MD clear strategy screen II from Molecular Dimensions) crystal screens, two promising conditions could be identified. Further optimization of these micro-crystals (Figure 4.5) resulted in the following two conditions: (1) 14% PEG 3350, 0.1 M Hepes pH 7.5 and 0.5 M ammonium citrate and (2) 10% PEG 20K, 0.1 M Bis-Tris pH 7.0 and 0.4 M Li_2SO_4. Although the crystals grown under the first condition are about 400-500 μm in diameter, they did not diffract at all after examined on beamline X29 at the NSLS. The crystals in the second condition are still small at this stage and appear as clustered needles. Further optimization for this condition is in progress.

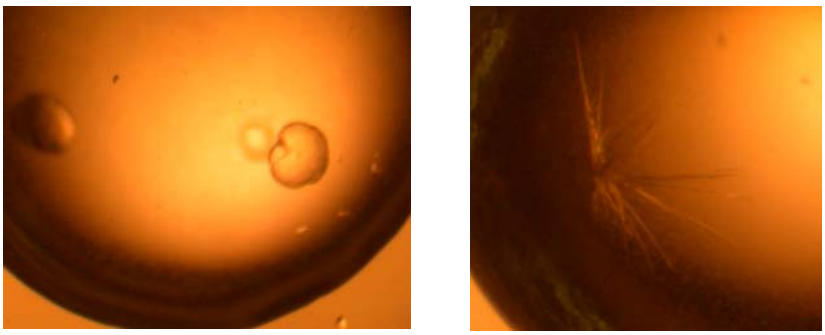


Figure 4.5 **Crystals of the ND1 and C15 peptide complex.** The reservoir solution for the experiment on the left contained 14% PEG 3350, 0.1 M Hepes pH 7.5 and 0.5 M ammonium citrate while that on the right side is composed of 10% PEG 20K, 0.1 M Bis-Tris pH 7.0 and 0.4 M Li_2SO_4.

IV. Conclusions

Derlin1 and its paralogs Derlin2 and Derlin3 are newly discovered membrane proteins, which reside in the ER membrane [13]. Delin1 has been found to associate with several proteins both in the membrane and in the cytosol [13, 56]. Recent studies provided preliminary evidence to support the idea that Derlin1 could in fact be the retro-translocating channel that researchers in the ERAD field have been searching for years [11-13, 111]. Under this assumption, Derlin1 should be able to communicate with ERAD components, which are both upstream (within the ER or in the ER membrane) and downstream (in the ER membrane and the cytosol). Previous studies [43] showed that p97 is an AAA ATPase which features two ATPase domains, which provide the driving force to extract misfolded proteins through the retro-translocating channel in the ER membrane. Therefore, p97 most likely acts downstream of Derlin1 in the ERAD, after the retrotranslocation step. Interestingly, recent biochemical data [56] demonstrated that these two proteins indeed interact with each other.

After independently confirming this interaction in a pull-down assay with HEK293 cell lysates, the interaction of these two proteins has been further characterized by mapping studies, which initially revealed that the C-terminal end of Derlin1 is critical for this interaction. Detailed domain mapping indicated that the last 15 residues of Derlin1 are essential for binding to p97. Further biochemical studies also showed that p97 interacts with Derlin1 through its N-terminal domain, the domain that is involved in many p97-protein interactions including Ufd1/Npl4 and some UBX containing proteins like p47 [46]. A thermodynamic characterization of this binding interaction revealed it to be moderately tight with a dissociation constant of 15 μM. Compared with the binding

affinity of the p97-p47 complex which has a dissociation constant of 0.5 μM and a stoichiometry of 2 to 1 [113], this affinity is 30 times weaker. However, since p97 forms a hexamer and one p97 hexamer could interact with more than one hDerlin1 protein, the affinity of p97-hDerlin1 may be within the same range as that of the p97-p47 interaction.

However, the Derlin1 C-terminal tails differs between the various Derlin homologs (Figure 4.4*B*). Der1 in yeast has the shortest tail, which is about half the size of that of mammalian Derlin1. In contrast, the yeast Dfm1 protein has the longest tail due to the presence of two insertions with a total length of 19 residues. As can be seen from the sequence alignment, the first 27 residues of the C-terminal tail are reasonably well-conserved (59% similarity) in all Derlin1 homologs, suggesting that this region carries out a similar function in all homologous proteins. The latter part of the C-terminal tail, however, differs dramatically between the four proteins listed in Fig. 4.4*B*, but is completely identical among mammalian Derlin1 proteins. The lack of conservation across eukaryotic organisms indicates differences in function between these homologs. This is in agreement with previous studies that assign different cellular functions to the Der1, Dfm1 and Derlin1 proteins [55, 57, 114]. In contrast the high conservation of the C-terminal tails in mammalian Derlin1 orthologs suggests a common function, presumably this is the interaction with p97 described here.

The biochemical and crystallization trials described in this chapter represent just a starting point for studying the function of Derlin1. In order to fully understand this protein, several issues need to be addressed. Foremost, is to understand how Derlin1 can form a retrotranlocon, if it really represents the protein export channel. Derlin1 only features four transmembrane spanning segments, which does not appear to be sufficient

to form a channel in the membrane. Although a recent study [56, 115] showed that Derlin1 dimerizes under physiological conditions, the dimer population is very small compared to the monomer. If Derlin1 cannot form the retrotranslocon by itself, are there other membrane proteins that participate in the formation of the channel? Since Derlin1 associates with VIMP, VIMP may participate in the membrane channel formation. However, VIMP only contains a single transmembrane segment and five helices might still not be enough. Another candidate for association with Derlin1 is the Sec translocon and in this case there would be more than enough helices to form a channel. The ultimate goal to aid in understanding these aspects of the Derlin1 function would be to obtain a structure of the full-length protein, possibly even in the presence of a bound substrate.

Another question is how Derlin1 binds to ERAD substrates. These proteins need to be transferred from upstream cofactors, which require Derlin1 to be in close spatial proximity, which could be mediated by a direct interaction similar to that described here for the downstream ERAD component p97. As a result of the studies presented here we now have a better understanding regarding the interaction between Derlin1 and p97. Apparently the action of Derlin1 and p97 are tightly coupled as reflected by this interaction, and the future structural characterization of this complex will shed light on the details of this interaction. In any case, this interaction represents another evidence that the retrotranslocation of misfolded glycoproteins from the ER, their subsequent ubiquitination and possible deglycosylation and ultimate degradation by the proteasome is a tightly coordinated process.

Chapter 5

Concluding Discussion

A. A cytosolic super-complex mediating all post-translocation steps during ERAD

The ERAD system can quickly detect and remove aberrant proteins to ensure that misfolded proteins in the ER do not pose a problem for cellular viability [1]. Malfunction of ERAD leads to various diseases including diabetes, inflammation, Alzheimer's disease and bipolar disorder [103]. The majority of proteins that enter the secretory pathway are N-glycosylated, and when these proteins cannot reach their native conformation in the ER, they are retrontranslocated to the cytosol and degraded there. Already in 1996, the accumulation of de-N-glycosylated Class I MHC heavy chain intermediates was observed in the presence of a proteasome inhibitor [14], indicating the involvement of PNGase activity in the ERAD prior to proteasomal degradation. PNGase is now thought to be required for cleaving off the bulky N-glycan chain(s) before the unfolded protein enters the narrow entrance leading into the 20 S proteasome core, the site of proteolytic degradation. In this manner the deglycosylation step ensures the efficient degradation of N-linked glycoproteins by the proteasome. Although a lot of attention has been paid to PNGase in recent years, the precise function of PNGase is still not clear for several reasons: 1) The degradation of MHC heavy chains is not affected even when PNGase is completely inhibited. 2) The PNGase knockout in yeast is viable. 3) *In vivo* substrates of PNGase are still mostly unknown, with the exception of the recent discovery of the ricin A chain as the first *in vivo* substrate. Therefore, further studies on ricin A degradation on one hand and the identification of additional *in vivo* substrates are needed to fully understand the cellular importance of PNGase.

The recent discovery of the quintenary complex consisting of mAMFR, mY33K, mp97, mPNGase and mHR23 by pull-down assays suggest that a complex that mediates

the ubiquitination, deglycosylation and degradation of ERAD substrate may exist and stay in close proximity of the ER membrane. This hypothesis agrees well with the fact that a population of human PNGase has been found to be associated with the ER membrane [116]. In this complex two proteins, p97 and HR23B, carry out an escort function, by bridging the interactions between the different enzymatic activities with p97 contributing its own enzymatic activity. HR23B interacts with PNGase and the proteasome through its XPCB and UBL domains, respectively. The PNGase-HR23 interaction is conserved throughout the PNGase family, although the mechanism of interaction is different between the yeast and mouse proteins. In addition to linking PNGase and the proteasome, HR23B features two UBA domains which may simultaneously interact with the polyubiquitin chains on the ERAD substrates in order to recruit them from the upstream deglycosylation complex to the proteasome.

A link between some E3 ubquitin ligases and mPNGase are mediated by p97. The PUB domain of mPNGase was identified as the primary binding region for its interaction with p97. In our study we found that one p97 hexamer binds to only two mouse PNGase molecules, presumably because a one to one complex would result in steric hindrance. Another protein involved in the protein degradation pathway is Ufd3, which is also found to interact with a similar region in p97, indicating a possible competition between Ufd3 and PNGase for interacting with p97. Ufd3 appears to participate in a de-ubiquitination process since a *ufd3* deletion leads to a severe lack of free ubiquitin [117]. Since Ufd2 has been found to compete with Ufd3 in binding with p97, Ufd2 either also interacts with the same C-terminal binding motif in p97 or there is steric overlap between the bound Ufd2 and Ufd3. However, despite the identification that Ufd3 and PNGase both bind to

the C-terminus of p97, a simultaneous binding cannot be ruled out since they could bind to different monomers given the fact that only two PNGase molecules are bound per p97. Similarily, even if Ufd2 and PNGase were to compete for the same binding site, the mutual exclusion could be overcome by a substoichiometric binding to different monomers. In the case that Ufd2 and PNGase simultaneously bind to one p97 hexamer, one would predict that the overall efficiency of ubiquitination and deglycosylation would increase even further since Ufd2 is an E4 multiubiquitination enzyme.

p97 interacts with most of its cofactors through its N-terminal domain. Among these cofactors, the interaction between p97 and the membrane E3 ligase AMFR is of special interest in the context of ERAD [21]. P97 directly binds to the C-terminal cytosolic tail of AMFR [21]. Recently, also the $SCF^{Fbs1,2}$ E3 ligase that specifically recognizes the N-glycosylated substrates was found to coimmunoprecipitate with p97 [80], suggesting that this cytosolic protein may be an alternative E3 of the ERAD machinery. Both E3 ligases utilize p97 as a platform to recruit other components of ERAD machinery such as PNGase. Considering that in mammals hundreds of E3 ligases exist which recognize different substrates, multi-subunit complexes involving either AMFR, $SCF^{Fbs1,2}$ or other E3 enzymes may be present in the vicinity of the ER membrane to facilitate the degradation of different ERAD substrates.

As described in Chapter 4, the very C-terminus of hDerlin1 has been found to interact with the N-terminal domain of p97. This interaction indicates that the retrotranslocation step may also be tightly coupled to other steps of ERAD in the cytosol, since Derlin1 has been suggested as a possible candidate for the retrotranslocon.

Taken together, a glycoprotein degradation complex consisting of Derlin1, AMFR/SCF^{Fbs1,2}, p97, PNGase, HR23B and the proteasome may exist in the cytosol. The relationship between these proteins is summarized in Figure 5.1. This complex would catalyze and possibly regulate the retrotranslocation, ubiquitination, deglycosylation and final degradation steps of ERAD substrates in a sequential and cooperative fashion, which would ensure the rapid degradation of misfolded glycoproteins.

B. PNGase in eukayotes: Domain functions and their phylogenetic relationships

The PNGase activity has been found in all eukaryotic organisms ranging from yeast to mammals. Within that group the yeast and mouse enzymes represent the two best studied members of this family, however, their architectures differ considerably. Besides the catalytic core domain, mPNGase contains two additional domains, an N-terminal domain (PUB domain) and a C-terminal domain compared to yPNGase. The PUB domain provides a link to the AAA ATPase, p97, which is involved in extracting misfolded (glyco)proteins from the ER. As described in Chapter 3, the minimal p97 binding region to the PUB domain has been identified as a novel binding motif that is located at very C-terminal end of p97. In contrast, most cofactors of p97 interact with the N-terminal domain of p97. Our structure of the PUB domain in complex with the p97 C-terminal peptide provides the first detail insights into this interaction. The binding region in p97 is confined to the last five amino acids, including Tyr805, a residue that is phosphorylated during T-cell receptor activation. Therefore, we hypothesized that during ER stress, p97 will be phosphorylated, which would prevent an interaction with PNGase and lead to a reduction in the degradation of glycoproteins.

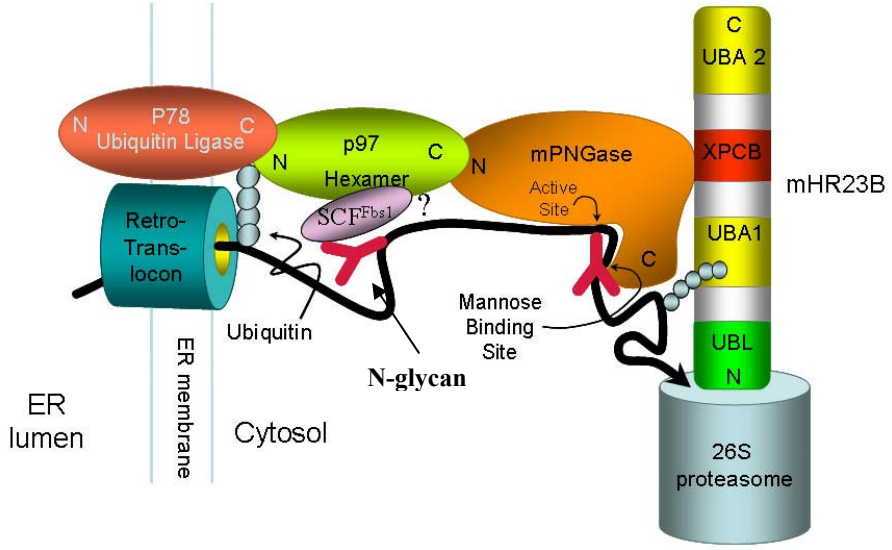


Figure 5.1 **A cytosolic super-complex mediates all post-translocation steps during ERAD.** The interactions between Derlin1, mAMFR/SCF^{Fbs1}, p97, mPNGase, Rad 23, and 26 S proteasome are as described in chapter 5. This figure has been modified a document prepared in the Lennarz lab.

Not all higher eukaryotes contain the PUB domain at their N-terminus, *C. elegans* for example, contains an N-terminal thioredoxin domain which is catalytically active [39]. The thioredoxin activity could potentially reduce disulfide bonds in the substrate and help to unfold substrates in the cytosol before the core domain of *C. elegans* PNGase cleaves off the carbohydrate chains. In those species containing the PUB domain, substrate unfolding may be ensured through the above mentioned interaction with p97, which has been suggested to accept unfolded ERAD substrates from the retrotranslocon [42]. The difference in the N-terminal domains of PNGase from multicellular organisms may result from the special requirements of the ERAD pathway in different species.

In contrast, the additional C-terminal domain of mPNGase is highly conserved in all multicellular organisms ranging from C. elegans to human. The function of this 200 residue domain was unknown until I solved the structure of this domain as described in Chapter 2. The C-terminal domain has a β-sandwich structure and contains a saddle-like feature that was identified as a mannopentaose binding motif by ITC studies coupled with site directed mutagenesis of residues located in this area. Finally the detailed interactions were visualized by a crystal structure of the protein-carbohydrate complex.

The differences between yPNGase and mPNGase discussed here could potentially result from the different substrate pools. In yPNGase, all N-glycosylated proteins are of the high mannose-type, whereas in mouse, there are also hybrid-type and complex-type N-glycosylated proteins. Although both the hybrid and complex-type N-glycan chains are generated in the Golgi [5], malfunctioning proteins can be detected by the quality control system and these proteins are transported back to the ER for further degradation. As a result, there could be some population of hybrid and complex-type substrates present in

the cytosol waiting for degradation [7]. However, whether yPNGase and mPNGase have the ability to differentiate between various types of substrates is still not clear. Since mPNGase binds to mannopentaose, it may act on those hybrid-type substrates, that have the major interacting mannose attached, but do not contain a core fucosylation. On the other hand, mPNGase will not function on complex-type N-glycosylated substrates because of a lack of the major interacting mannose. These types of substrates could be degraded via a different route. In contrast, yPNGase may not have this ability for substrate selection. It may function on any type of substrate as long as it is not fucosylated at the chitobiose site (Figure 5.2). In order to answer this question, different types of substrates need to be tested in the future using both the yeast and mouse enzymes.

C. A hypothetical model of mPNGase and its interaction with substrate

Recently, the structure of the core domain of mPNGase in complex with the XPC binding domain of HR23B has been determined in our lab [26] as well as a complex of yPNGase with an iodoacetamidyl-modified chitobiose (G. Zhao et al., unpublished data). These structures, especially the complexes with the peptide-derived inhibitor Z-VAD-fmk and the chitobiose derivative provide additional information on how PNGase recognizes its substrate. As demonstrated here, substrate-binding by mPNGase involves not only the core domain, but also the C-terminal domain. Since the relative arrangement of the two domains is important in understanding the substrate specificity of PNGase, possible domain orientations were investigated by manually positioning the structures of

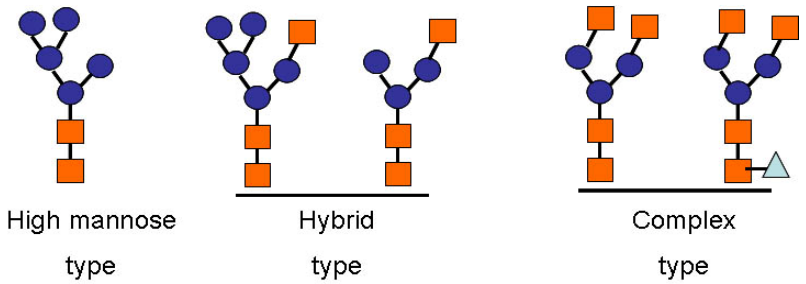


Figure 5.2 **The three types of N-glycan chains.** High mannose type, hybrid type and complex type glycan chains differ in their composition. The blue circles reprensent mannose, orange squares N-acetylglucosamine, and cyan triangles fucose.

the mPNGase core and C-terminal domains while at the same time fulfilling two conditions: (1) The C-terminus of the core domain and the N-terminus of the C-terminal domain have to be in spatial proximity since there are only four residues in between. Nevertheless, certain movements within this region appear possible as both ends are flexible. (2) Since the active site cysteine (Cys306) in the core domain has to be in close proximity of the N-glycosidic bond of the substrate, the distance between its side chain and the center of the mannose binding motif in the C-terminal domain was required to be within 25 Å, the maximum length of four pyranoses.

In the resulting model (Figure 5.3), a continuous cleft is visible which on one end accommodates the peptide represented by the Z-VAD-fmk inhibitor, followed by the chitobiose moiety in the center and finally the mannotetraose in the C-terminal domain. The peptide inhibitor and the chitobiose are adjacent to the catalytic triad in the core domain, while additional conserved residues in the core domain are involved in chitobiose and peptide binding due to their close spatial proximity. However, as stated before (see Chapter 2 for detail), we did not detect binding of chitobiose to the core domain in ITC experiments. The presence of the oligomannose binding site in the C-terminal domain extends the substrate-binding cleft of mPNGase and therefore presumably enhances the catalytic efficiency of the enzyme.

This model provides the first structural details about how mouse PNGase binds to its substrate. However, a lot of other questions still need to be answered, such as: (1) How does yPNGase bind to its substrate while lacking a corresponding C-terminal domain. (2) Are there any other *in vivo* substrates of PNGase besides ricin A? To answers the questions, further studies of PNGase function in the ERAD pathway is needed.

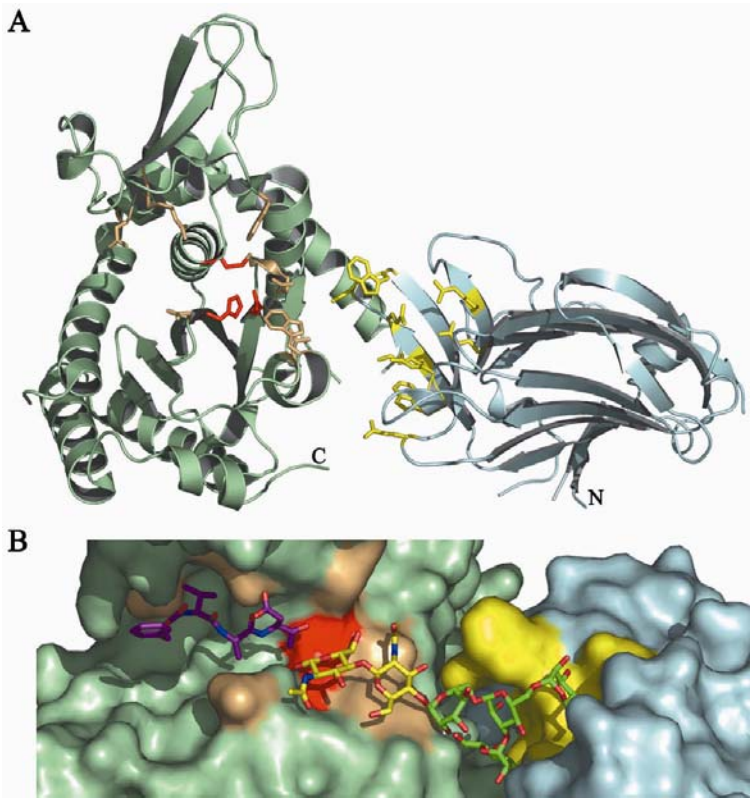


Figure 5.3 **Hypothetical model of the relative arrangements of the core and C-terminal domains of mPNGase.** (*A*) Ribbon diagram of the core domain in green and the C-terminal domain in cyan. In the core domain, residues of the catalytic triad are colored in red and other conserved residues in the binding pocket are colored in light-orange. The mannose-binding residues in mPNGc are displayed in yellow. The C-terminus of the core domain and N-terminus of the C-terminal domain are labeled with C and N, respectively. (*B*) Close-up view in the same orientation as in (*A*) with the protein in a surface representation. The carbon atoms of the chitobiose are colored in yellow and those of the mannotetraose in green. The carbon atoms of the mPNGase inhibitor Z-VAD-fmk are shown in magenta.

REFERENCES:

1. Schubert, U., et al., *Rapid degradation of a large fraction of newly synthesized proteins by proteasomes.* Nature, 2000. **404**(6779): p. 770-4.

2. Keenan, R.J., et al., *The signal recognition particle.* Annu Rev Biochem, 2001. **70**: p. 755-75.

3. Romisch, K., *Endoplasmic reticulum-associated degradation.* Annu Rev Cell Dev Biol, 2005. **21**: p. 435-56.

4. Van den Berg, B., et al., *X-ray structure of a protein-conducting channel.* Nature, 2004. **427**(6969): p. 36-44.

5. Editors, T.C.o.G., *Essentials of Glycobiology*. 1999, La Jolla: Cold spring harbor laboratory press.

6. Yan, A. and W.J. Lennarz, *Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons.* Glycobiology, 2005. **15**(12): p. 1407-15.

7. Yoshida, Y., *A novel role for N-glycans in the ERAD system.* J Biochem (Tokyo), 2003. **134**(2): p. 183-90.

8. Hebert, D.N., S.C. Garman, and M. Molinari, *The glycan code of the endoplasmic reticulum: asparagine-linked carbohydrates as protein maturation and quality-control tags.* Trends Cell Biol, 2005. **15**(7): p. 364-70.

9. Lederkremer, G.Z. and M.H. Glickman, *A window of opportunity: timing protein degradation by trimming of sugars and ubiquitins.* Trends Biochem Sci, 2005. **30**(6): p. 297-303.

10. Ruddock, L.W. and M. Molinari, *N-glycan processing in ER quality control.* J Cell Sci, 2006. **119**(Pt 21): p. 4373-80.

11. Ye, Y., et al., *A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol.* Nature, 2004. **429**(6994): p. 841-7.

12. Wahlman, J., et al., *Real-Time Fluorescence Detection of ERAD Substrate Retrotranslocation in a Mammalian In Vitro System.* Cell, 2007. **129**(5): p. 943-55.

13. Lilley, B.N. and H.L. Ploegh, *A membrane protein required for dislocation of misfolded proteins from the ER.* Nature, 2004. **429**(6994): p. 834-40.

14. Wiertz, E.J., et al., *Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction.* Nature, 1996. **384**(6608): p. 432-8.

15. Jin, J., et al., *Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging.* Nature, 2007. **447**(7148): p. 1135-8.

16. Pickart, C.M. and M.J. Eddins, *Ubiquitin: structures, functions, mechanisms.* Biochim Biophys Acta, 2004. **1695**(1-3): p. 55-72.

17. Fang, S., et al., *RING finger ubiquitin protein ligases: implications for tumorigenesis, metastasis and for molecular targets in cancer.* Semin Cancer Biol, 2003. **13**(1): p. 5-14.

18. Hirsch, C., D. Blom, and H.L. Ploegh, *A role for N-glycanase in the cytosolic turnover of glycoproteins.* Embo J, 2003. **22**(5): p. 1036-46.

19. Hartmann-Petersen, R., K. Tanaka, and K.B. Hendil, *Quaternary structure of the ATPase complex of human 26S proteasomes determined by chemical cross-linking.* Arch Biochem Biophys, 2001. **386**(1): p. 89-94.

20. Lowe, J., et al., *Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution.* Science, 1995. **268**(5210): p. 533-9.

21. Li, G., et al., *The AAA ATPase p97 links peptide N-glycanase to the endoplasmic reticulum-associated E3 ligase autocrine motility factor receptor.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8348-53.

22. Hiyama, H., et al., *Interaction of hHR23 with S5a. The ubiquitin-like domain of hHR23 mediates interaction with S5a subunit of 26 S proteasome.* J Biol Chem, 1999. **274**(39): p. 28019-25.

23. Suzuki, T., H. Park, and W.J. Lennarz, *Cytoplasmic peptide:N-glycanase (PNGase) in eukaryotic cells: occurrence, primary structure, and potential functions.* Faseb J, 2002. **16**(7): p. 635-41.

24. Katiyar, S., et al., *Site-directed mutagenesis study of yeast peptide:N-glycanase. Insight into the reaction mechanism of deglycosylation.* J Biol Chem, 2002. **277**(15): p. 12953-9.

25. Della Mea, M., et al., *AtPng1p. The first plant transglutaminase.* Plant Physiol, 2004. **135**(4): p. 2046-54.

26. Zhao, G., et al., *Structure of the mouse peptide N-glycanase-HR23 complex suggests co-evolution of the endoplasmic reticulum-associated degradation and DNA repair pathways.* J Biol Chem, 2006. **281**(19): p. 13751-61.

27. Lee, J.H., et al., *Structure of a peptide:N-glycanase-Rad23 complex: insight into the deglycosylation for denatured glycoproteins.* Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9144-9.

28. Suzuki, T., et al., *PNG1, a yeast gene encoding a highly conserved peptide:N-glycanase.* J Cell Biol, 2000. **149**(5): p. 1039-52.

29. Hirsch, C., et al., *Yeast N-glycanase distinguishes between native and non-native glycoproteins.* EMBO Rep, 2004. **5**(2): p. 201-6.

30. Joshi, S., S. Katiyar, and W.J. Lennarz, *Misfolding of glycoproteins is a prerequisite for peptide: N-glycanase mediated deglycosylation.* FEBS Lett, 2005. **579**(3): p. 823-6.

31. Kim, I., et al., *The Png1-Rad23 complex regulates glycoprotein turnover.* J Cell Biol, 2006. **172**(2): p. 211-9.

32. Park, H., T. Suzuki, and W.J. Lennarz, *Identification of proteins that interact with mammalian peptide:N-glycanase and implicate this hydrolase in the proteasome-dependent pathway for protein degradation.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11163-8.

33. Suzuki, T., et al., *Rad23 provides a link between the Png1 deglycosylating enzyme and the 26 S proteasome in yeast.* J Biol Chem, 2001. **276**(24): p. 21601-7.

34. Ng, J.M., et al., *Developmental defects and male sterility in mice lacking the ubiquitin-like DNA repair gene mHR23B.* Mol Cell Biol, 2002. **22**(4): p. 1233-45.

35. Walters, K.J., et al., *DNA-repair protein hHR23a alters its protein structure upon binding proteasomal subunit S5a.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12694-9.

36. Misaghi, S., et al., *Using a small molecule inhibitor of peptide: N-glycanase to probe its role in glycoprotein turnover.* Chem Biol, 2004. **11**(12): p. 1677-87.

37. Suzuki, T., et al., *Site-specific labeling of cytoplasmic peptide:N-glycanase by N,N'-diacetylchitobiose-related compounds.* J Biol Chem, 2006. **281**(31): p. 22152-60.

38. Kato, T., et al., *Unique peptide:N-glycanase of Caenorhabditis elegans has activity of protein disulfide reductase as well as of deglycosylation.* J Biochem (Tokyo), 2007.

39. Suzuki, T., et al., *Dual enzymatic properties of the cytoplasmic peptide:N-glycanase in C. elegans.* Biochem Biophys Res Commun, 2007. **358**(3): p. 837-841.

40. Suzuki, T., et al., *The PUB domain: a putative protein-protein interaction domain implicated in the ubiquitin-proteasome pathway.* Biochem Biophys Res Commun, 2001. **287**(5): p. 1083-7.

41. Ye, Y., *Diverse functions with a common regulator: ubiquitin takes command of an AAA ATPase.* J Struct Biol, 2006. **156**(1): p. 29-40.

42. Pye, V.E., et al., *Going through the motions: The ATPase cycle of p97.* J Struct Biol, 2006.

43. DeLaBarre, B. and A.T. Brunger, *Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains.* Nat Struct Biol, 2003. **10**(10): p. 856-63.

44. Wang, Q., C. Song, and C.C. Li, *Molecular perspectives on p97-VCP: progress in understanding its structure and diverse biological functions.* J Struct Biol, 2004. **146**(1-2): p. 44-57.

45. Wojcik, C., M. Yano, and G.N. DeMartino, *RNA interference of valosin-containing protein (VCP/p97) reveals multiple cellular roles linked to ubiquitin/proteasome-dependent proteolysis.* J Cell Sci, 2004. **117**(Pt 2): p. 281-92.

46. Dreveny, I., et al., *p97 and close encounters of every kind: a brief review.* Biochem Soc Trans, 2004. **32**(Pt 5): p. 715-20.

47. Alzayady, K.J., et al., *Involvement of the p97-Ufd1-Npl4 complex in the regulated endoplasmic reticulum-associated degradation of inositol 1,4,5-trisphosphate receptors.* J Biol Chem, 2005. **280**(41): p. 34530-7.

48. Richly, H., et al., *A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting.* Cell, 2005. **120**(1): p. 73-84.

49. Rumpf, S. and S. Jentsch, *Functional division of substrate processing cofactors of the ubiquitin-selective Cdc48 chaperone.* Mol Cell, 2006. **21**(2): p. 261-9.

50. Hetzer, M., et al., *Distinct AAA-ATPase p97 complexes function in discrete steps of nuclear assembly.* Nat Cell Biol, 2001. **3**(12): p. 1086-91.

51. Cao, K., et al., *The AAA-ATPase Cdc48/p97 regulates spindle disassembly at the end of mitosis.* Cell, 2003. **115**(3): p. 355-67.

52. Kondo, H., et al., *p47 is a cofactor for p97-mediated membrane fusion.* Nature, 1997. **388**(6637): p. 75-8.

53. Uchiyama, K. and H. Kondo, *p97/p47-Mediated biogenesis of Golgi and ER.* J Biochem (Tokyo), 2005. **137**(2): p. 115-9.

54. Knop, M., et al., *Der1, a novel protein specifically required for endoplasmic reticulum degradation in yeast.* Embo J, 1996. **15**(4): p. 753-63.

55. Hitt, R. and D.H. Wolf, *Der1p, a protein required for degradation of malfolded soluble proteins of the endoplasmic reticulum: topology and Der1-like proteins.* FEMS Yeast Res, 2004. **4**(7): p. 721-9.

56. Ye, Y., et al., *Inaugural Article: Recruitment of the p97 ATPase and ubiquitin ligases to the site of retrotranslocation at the endoplasmic reticulum membrane.* Proc Natl Acad Sci U S A, 2005. **102**(40): p. 14132-8.

57. Sato, B.K. and R.Y. Hampton, *Yeast Derlin Dfm1 interacts with Cdc48 and functions in ER homeostasis.* Yeast, 2006. **23**(14-15): p. 1053-64.

58. Story, C.M., M.H. Furman, and H.L. Ploegh, *The cytosolic tail of class I MHC heavy chain is required for its dislocation by the human cytomegalovirus US2 and US11 gene products.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8516-21.

59. Walter, J., et al., *Sec61p-independent degradation of the tail-anchored ER membrane protein Ubc6p.* Embo J, 2001. **20**(12): p. 3124-31.

60. Ploegh, H.L., *A lipid-based model for the creation of an escape hatch from the endoplasmic reticulum.* Nature, 2007. **448**(7152): p. 435-8.

61. Suzuki, T., et al., *Does an animal peptide: N-glycanase have the dual role as an enzyme and a carbohydrate-binding protein?* Glycoconj J, 1994. **11**(5): p. 469-76.

62. Sambrook, J. and D.W. Russell, *Molecular cloning : a laboratory manual*. 3rd ed. 2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

63. Sheldrick, G., and Schneider, T. Methods Enzymol. Vol. 277. 1997, New York: Academic press. 319-343.

64. Bricogne, G., et al., *Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 11): p. 2023-30.

65. Abrahams, J.P. and A.G. Leslie, *Methods used in the structure determination of bovine mitochondrial F1 ATPase.* Acta Crystallogr D Biol Crystallogr, 1996. **52**(Pt 1): p. 30-42.

66. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.

67. Perrakis, A., R. Morris, and V.S. Lamzin, *Automated protein model building combined with iterative structure refinement.* Nat Struct Biol, 1999. **6**(5): p. 458-63.

68. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.

69. Vagin, A. and A. Teplyakov, *MOLREP: an Automated Program for Molecular Replacement.* Journal of Applied Crystallography, 1997. **30**(6): p. 1022-1025.

70. Painter, J. and E.A. Merritt, *TLSMD web server for the generation of multi-group TLS models*, in *Journal of Applied Crystallography*. 2006. p. 109-111.

71. Davis, I.W., et al., *MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W615-9.

72. Holm, L. and C. Sander, *Searching protein structure databases has come of age.* Proteins, 1994. **19**(3): p. 165-73.

73. Mizushima, T., et al., *Structural basis of sugar-recognizing ubiquitin ligase.* Nat Struct Mol Biol, 2004. **11**(4): p. 365-70.

74. Yoshida, Y., et al., *Fbs2 is a new member of the E3 ubiquitin ligase family that recognizes sugar chains.* J Biol Chem, 2003. **278**(44): p. 43877-84.

75. Yoshida, Y., et al., *E3 ubiquitin ligase that recognizes sugar chains.* Nature, 2002. **418**(6896): p. 438-42.

76. Feinberg, H., et al., *Crystal structure of the CUB1-EGF-CUB2 region of mannose-binding protein associated serine protease-2.* Embo J, 2003. **22**(10): p. 2348-59.

77. Mizushima, T., et al., *Structural basis for the selection of glycosylated substrates by SCF(Fbs1) ubiquitin ligase.* Proc Natl Acad Sci U S A, 2007. **104**(14): p. 5777-81.

78. Garcia-Hernandez, E., et al., *Stereochemical metrics of lectin-carbohydrate interactions: comparison with protein-protein interfaces.* Glycobiology, 2000. **10**(10): p. 993-1000.

79. Taroni, C., S. Jones, and J.M. Thornton, *Analysis and prediction of carbohydrate binding sites.* Protein Eng, 2000. **13**(2): p. 89-98.

80. Yoshida, Y., et al., *Glycoprotein-specific ubiquitin ligases recognize N-glycans in unfolded substrates.* EMBO Rep, 2005. **6**(3): p. 239-44.

81. McNeill, H., et al., *A novel UBA and UBX domain protein that binds polyubiquitin and VCP and is a substrate for SAPKs.* Biochem J, 2004. **384**(Pt 2): p. 391-400.

82. Rabinovich, E., et al., *AAA-ATPase p97/Cdc48p, a cytosolic chaperone required for endoplasmic reticulum-associated protein degradation.* Mol Cell Biol, 2002. **22**(2): p. 626-34.

83. Jarosch, E., et al., *Protein dislocation from the ER requires polyubiquitination and the AAA-ATPase Cdc48.* Nat Cell Biol, 2002. **4**(2): p. 134-9.

84. Zhao, G., et al., *Structure of the mouse peptide N-glycanase-HR23 complex suggests co-evolution of the endoplasmic reticulum-associated degradation and DNA repair pathways.* J Biol Chem, 2006.

85. Biswas, S., et al., *The N-terminus of yeast peptide: N-glycanase interacts with the DNA repair protein Rad23.* Biochem Biophys Res Commun, 2004. **323**(1): p. 149-55.

86. Otwinowski, Z., Minor,W., *Processing of X-ray Diffraction Data Collected in Oscillation Mode* Methods in Enzymology. Vol. 276. 1997, New York: Academic press. 307-326.

87. Terwilliger, T.C., *Maximum-likelihood density modification.* Acta Crystallogr D Biol Crystallogr, 2000. **56**(Pt 8): p. 965-72.

88. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 4): p. 849-61.

89. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.

90. Gouet, P., et al., *ESPript: analysis of multiple sequence alignments in PostScript.* Bioinformatics, 1999. **15**(4): p. 305-8.

91. Mark D. Allen, A.B., and Mark Bycroft, *The PUB domain functions as a p97 binding module in human peptide N-glycanase.* J. Biol. Chem., 2006: p. doi:10.1074/jbc.M601173200.

92. DeLano, W.L., *The PyMOL User's Manual*. 2002, San Carlos, CA, USA: DeLano Scientific

93. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.

94. Gajiwala, K.S. and S.K. Burley, *Winged helix proteins.* Curr Opin Struct Biol, 2000. **10**(1): p. 110-6.

95. Carim-Todd, L., et al., *Identification and characterization of UBXD1, a novel UBX domain-containing gene on human chromosome 19p13, and its mouse ortholog.* Biochim Biophys Acta, 2001. **1517**(2): p. 298-301.

96. Ficarro, S., et al., *Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation.* J Biol Chem, 2003. **278**(13): p. 11579-89.

97. Livingstone, M., et al., *Valosin-containing protein phosphorylation at Ser784 in response to DNA damage.* Cancer Res, 2005. **65**(17): p. 7533-40.

98. Lavoie, C., et al., *Tyrosine phosphorylation of p97 regulates transitional endoplasmic reticulum assembly in vitro.* Proc Natl Acad Sci U S A, 2000. **97**(25): p. 13637-42.

99. Egerton, M., et al., *VCP, the mammalian homolog of cdc48, is tyrosine phosphorylated in response to T cell antigen receptor activation.* Embo J, 1992. **11**(10): p. 3533-40.

100. Madeo, F., et al., *Tyrosine phosphorylation regulates cell cycle-dependent nuclear localization of Cdc48p.* Mol Biol Cell, 1998. **9**(1): p. 131-41.

101. Klein, J.B., et al., *Akt-mediated valosin-containing protein 97 phosphorylation regulates its association with ubiquitinated proteins.* J Biol Chem, 2005. **280**(36): p. 31870-81.

102. Egerton, M. and L.E. Samelson, *Biochemical characterization of valosin-containing protein, a protein tyrosine kinase substrate in hematopoietic cells.* J Biol Chem, 1994. **269**(15): p. 11435-41.

103. Yoshida, H., *ER stress and diseases.* Febs J, 2007. **274**(3): p. 630-58.

104. Vij, N., S. Fang, and P.L. Zeitlin, *Selective inhibition of endoplasmic reticulum-associated degradation rescues DeltaF508-cystic fibrosis transmembrane regulator and suppresses interleukin-8 levels: therapeutic implications.* J Biol Chem, 2006. **281**(25): p. 17369-78.

105. Wiertz, E.J., et al., *The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol.* Cell, 1996. **84**(5): p. 769-79.

106. Zhang, S.H., et al., *Identification of the cell cycle regulator VCP (p97/CDC48) as a substrate of the band 4.1-related protein-tyrosine phosphatase PTPH1.* J Biol Chem, 1999. **274**(25): p. 17806-12.

107. Wu, J., et al., *Identification of substrates of human protein-tyrosine phosphatase PTPN22.* J Biol Chem, 2006. **281**(16): p. 11002-10.

108. Gu, F., et al., *Protein-tyrosine phosphatase 1B potentiates IRE1 signaling during endoplasmic reticulum stress.* J Biol Chem, 2004. **279**(48): p. 49689-93.

109. Tiwari, S. and A.M. Weissman, *Endoplasmic reticulum (ER)-associated degradation of T cell receptor subunits. Involvement of ER-associated ubiquitin-conjugating enzymes (E2s).* J Biol Chem, 2001. **276**(19): p. 16193-200.

110. Buchberger, A., et al., *The UBX domain: a widespread ubiquitin-like module.* J Mol Biol, 2001. **307**(1): p. 17-24.

111. Lilley, B.N. and H.L. Ploegh, *Multiprotein complexes that link dislocation, ubiquitination, and extraction of misfolded proteins from the endoplasmic reticulum membrane.* Proc Natl Acad Sci U S A, 2005. **102**(40): p. 14296-301.

112. Nita-Lazar, A. and R.S. Haltiwanger, *Methods for analysis of O-linked modifications on epidermal growth factor-like and thrombospondin type 1 repeats.* Methods Enzymol, 2006. **417**: p. 93-111.

113. Beuron, F., et al., *Conformational changes in the AAA ATPase p97-p47 adaptor complex.* Embo J, 2006. **25**(9): p. 1967-76.

114. Hu, D., et al., *Overexpressed Derlin-1 inhibits ER expansion in the endothelial cells derived from human hepatic cavernous hemangioma.* J Biochem Mol Biol, 2006. **39**(6): p. 677-85.

115. Crawshaw, S.G., et al., *The oligomeric state of Derlin-1 is modulated by endoplasmic reticulum stress.* Mol Membr Biol, 2007. **24**(2): p. 113-20.

116. Katiyar, S., G. Li, and W.J. Lennarz, *A complex between peptide:N-glycanase and two proteasome-linked proteins suggests a mechanism for the degradation of misfolded glycoproteins.* Proc Natl Acad Sci U S A, 2004. **101**(38): p. 13774-9.

117. Ghislain, M., et al., *Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in Saccharomyces cerevisiae.* Embo J, 1996. **15**(18): p. 4884-99.