Structural and Functional Studies on the Ubiquitin-Specific Protease Family

Mark P.A. Vargas
The research described in this thesis was performed in the Division of Biochemistry at the Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital (NKI-AVL), The Netherlands.

This research was financially supported by Rubicon Network of Excellence (EU), KWF (Koningin Wilhelmina Fonds) and NWO-CW (Nederlandse organisatie voor de wetenschappelijk onderzoek - chemische wetenschappen and CBG (Centre for Biomedical Genetics).

Copyright © by M.P.A. Luna-Vargas. All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without prior permission of the author.
Structural and Functional Studies on the Ubiquitin-Specific Protease Family

Structurele en functionele studies op de ubiquitine-specifieke protease familie

Thesis

to obtain the degree of Doctor from the Erasmus University Rotterdam by command of the rector magnificus

Prof.dr. H.G. Schmidt

and in accordance with the decision of the Doctorate Board

The public defence shall be held on
Wednesday 9 November 2011 at 09.30 o’clock

By

Mark Patrick Alexander Vargas

born in Amsterdam, The Netherlands
DOCTORAL COMMITTEE

**Promotor:**
Prof. dr. T.K. Sixma

**Other members:**
Prof. dr. J. Neefjes
Prof. dr. R. Kanaar
Prof. dr. C.P. Verrijzer
Table of contents

Chapter 1  General introduction  7

Chapter 2  A Genomic and functional inventory of deubiquitinating enzymes  
(Cell, 2005)  29

Chapter 3  Enabling high-throughput ligation-independent cloning and 
protein expression for the family of ubiquitin specific-proteases  
(Journal of Structural Biology, 2011)  51

Chapter 4  The differential modulation of USP activity by internal regulatory 
domains, interactors and ubiquitin-chain types  
(submitted)  65

Chapter 5  Ubiquitin-specific protease 4 is inhibited by its ubiquitin-like 
domain  
(EMBO reports, 2011)  89

Chapter 6  General discussion  111

Addendum  Summary  121
Samenvatting  123
List of abbreviations  127
Curriculum vitae  131
PhD portfolio  133
List of publications  135
Acknowledgements  137
General Introduction
**Post-translational modification**
When RNA translation has completed and the proteins have been properly folded, most proteins are subject to a diverse range of post-translational modifications that affect their function. It is an important cellular strategy that enables the cell to react dynamically to intracellular or environmental changes caused by exposure to stress factors, growth stimuli or differentiation signals. Proteins are modified by methylation, acetylation, hydroxylation, phosphorylation or conjugated to ubiquitin (Ub) and ubiquitin-like (Ubl) proteins. The post-translational modification with Ub of proteins has emerged to play an essential role in the regulation of virtually all aspects of cell biology.

**Ubiquitination pathway**
Protein ubiquitination was discovered in the 1980s as a post-translational modification in which lysine residues of a target protein are modified with the addition of Ub. This small protein is a highly conserved 76 amino-acid polypeptide of ~8500 Da. The crystal structure of Ub revealed a distinctive fold, the β-grasp fold, which is also present in a group of proteins with distinct functions: ubiquitin-like (Ubl) proteins.

Through the sequential action of three enzymes a target protein is modified by covalent ligation to Ub (Figure 1). In the first step of the ubiquitination pathway, Ub is activated by a specific activating enzyme (E1) UBA1. UBA1 first binds MgATP and Ub and catalyzes Ub C-terminal acyl-adenylation. Secondly, the catalytic cysteine residue in the E1 attacks the ubiquitin-adenylate to form the activated ubiquitin-E1 complex via a high-energy thioester bond. In the second step of the ubiquitination pathway, the Ub loaded E1 engages one of up to tens of related E2 conjugating enzymes and then transfers the activated Ub to an active site cysteine residue of the E2. Finally the Ub is covalently linked by its C-terminus in an amide isopeptide linkage to an ε-amino group of a lysine residue of a target protein, which is catalyzed through the coordinated function of E3 Ub ligases. There are three different types of E3 ligase, RING (the Really Interesting New Gene), U-box and HECT (Homologous with E6-associated protein C-Terminus) domains and contain binding sites for both charged E2s and ubiquitinated substrates. For the largest class of E3s the RING family and the RING-related U-box family an ε-amino group of a lysine residue in the associated substrate attacks the thioester of the transiently associated charged E2 to make an isopeptide bond with Ub. The discharged E2 then dissociates from the -E3, allowing a second charged E2 to interact with the E3. This facilitates a second round of Ub transfer, either by attack of a lysine residue on Ub itself or by attack of a different lysine residue on the substrate.

In addition, through a similar cascade of reactions catalyzed by evolutionary related enzymes, the Ubl proteins (URM1, ATG12, Nedd8, SUMO, FAT10, ISG15 etc.) can also be conjugated to target proteins.

Because Ub itself contains seven acceptor lysines that can be a target of Ub conjugation, different types and lengths of Ub chains can be formed. Thus target proteins not only are ubiquitinated with a single Ub on a single lysine residue...
(mono-ubiquitination) or on multiple lysine residues (multi-mono-ubiquitination), but they are also modified with different types of poly-ubiquitin chains (Figure 2).

In addition Ub chains can be assembled in a head-to-tail or linear configuration through the α-amino group at the N-terminus. An important difference between linear and lysine-linked chains is the chemistry of the linkage (isopeptide compared with peptide bond). To add to the complexity of ubiquitination, quantitative studies have shown the existence of branched Ub chains, although the specific pathways in which these more complex poly-ubiquitin chains are involved remain poorly understood.

Deubiquitination pathway

Protein ubiquitination, similar to other regulated targeting pathways, is a reversible process and in the last decade protein deubiquitination has emerged as an important regulatory step in the ubiquitin-dependent pathways. In this deubiquitination process the isopeptide bond between Ub and a target protein or between Ub molecules in a poly-ubiquitin chain, is hydrolyzed by proteases named deubiquitinases (DUBs) (Figure 1).

Proteases

Proteases likely arose at the earliest stages of protein evolution as simple destructive enzymes necessary for protein catabolism and the generation of amino acids in primitive organisms. However, many years of studies on proteases have shown their relevance in the control of multiple biological processes in all living organisms. Thus proteases regulate the fate, localization and activity of many proteins, modulate protein-protein interactions, create new bioactive molecules and generate, transduce and amplify molecular signals. Because of their essential roles in different cellular processes, alterations in the structure and expression patterns of proteases underlie many human pathological conditions such as cancer, neurodegenerative disorders, and inflammatory and cardiovascular diseases.

Based on the mechanism of catalysis, proteases are classified into six distinct classes: aspartic, glutamic, metalloproteases, cysteine, serine and threonine proteases. The first three classes utilize an activated water molecule as a nucleophile to attack the peptide bond of the substrate.

In contrast to the remaining protease classes, the nucleophile is an amino acid residue (Cys, Ser or Thr) located in the active site. Most proteases hydrolyse the α-peptide bonds between naturally occurring amino acids, but there are some proteases that perform slightly different reactions. In particular, the DUBs are able to hydrolyze the isopeptide bonds in Ub and Ubl protein conjugates.

DUBS

The human genome encodes more than 100 putative DUBs that are predicted to be active. Even though there is a lack of knowledge about the regulation and roles of many DUBs, several generalizations have emerged in recent years.

Most DUB activity is cryptic. That is DUBs require substrate association or a scaffolding protein to achieve competent conformation. So like most other proteases, their activity is carefully controlled to prevent unnecessary cleavage of non-substrates. Other DUBs are covalently modified by phosphorylation, ubiquitination or sumoylation, which are likely to influence activity, localization and half-life.

In addition to their active-site core domains, most DUBs contain insertions within the catalytic domain and N- and C-terminal extensions, which participate in the substrate binding and recognition and direct the assembly of multi-protein complexes that localize DUBs and assist in substrate selection. These extensions contain predicted ubiquitin-binding domains (UBDs), including the zinc finger ubiquitin-specific protease domain (ZnF-UBP domain), the ubiquitin-interacting motif (UIM) and the ubiquitin-associated domain (UBA domain). The presence of one or multiple UBD domains is also widely predicted.

DUB activities fall into three major functional categories (Figure 2). First, Ub can be transcribed from several genes as a linear fusion of multiple Ub molecules or with ribosomal proteins, such that the generation of free Ub requires DUB activity (Figure 2a). Second, DUBs can remove Ub chains from post-translationally modified proteins, leading to reversal of Ub signaling or to protein stabilization by rescue from degradation (Figure 2b, c). However, once proteins are targeted for degradation, associated DUB activities can prevent degradation of Ub and maintain Ub homeostasis (recycle of Ub) (Figure 2d,e). Third, DUBs can...
be used to edit the form of Ub modification by trimmingUb chains and thereby help to exchange one type of Ub signal for another (Figure 2f)\(^{38,39}\).

The DUB family can be classified into five distinct subfamilies: ubiquitin C-terminal hydrolases (UCHs), ubiquitin-specific proteases (USPs), ovarian tumor proteases (OTUs), Josephins and Jab1/MPN/Mov34 metalloenzymes (JAMMs). The UCH, USP, OTU and Josephin subfamilies are Cys proteases, whereas the JAMM family members are zinc metalloproteases. A more comprehensive view on these five subfamilies of DUBs will be given in chapter 2: A genomic and function inventory of deubiquitinating enzymes.

Ubiquitin signaling

Types of ubiquitination

Depending on the type of Ub modification (attachment of a single Ub molecule or a Ub chain), a target protein can undergo different cellular fates (Figure 3). Multi-mono-ubiquitination has been shown to be involved in triggering the internalization of cell-surface receptors and their subsequent degradation in lysosomes or recycling to the cell surface\(^{40}\). Mono-ubiquitination is also involved in the DNA-damage response pathway, where histones or the DNA sliding clamp, PCNA (Proliferating-Cell Nuclear Antigen), are mono-ubiquitinated\(^{41,42}\).

The most well studied example of Ub signaling is the Ub chain conjugation through the lysine at position 48 of Ub which leads to the proteasomal degradation of the modified target protein\(^{7,43-45}\).

Also, proteins modified with K11-linked poly-ubiquitin chains are targeted for degradation by the proteasome. In addition proteins from diverse cellular processes were identified as being modified and regulated by K11-linked chains, suggesting that K11 is employed in many different pathways (e.g. in ERAD (Endoplasmic-Reticulum-Associated Degradation) and in the cell cycle regulation\(^{21,46,47}\)).

Another degradation signal is the K29-linked poly-ubiquitination. Proteins conjugated with K29-linked chains are degraded by lysosomal rather than proteasomal degradation pathways\(^{48,49}\). K29-linked poly-ubiquitin chains are also implicated in the UFD (Ubiquitin Fusion Degradation) pathway, in which the N-terminus of target proteins are attached to a Ub leading to the extension with a K29- or K48-linked Ub chain, resulting in efficient degradation of the fusion protein\(^{50}\). Besides
playing a role in protein degradation pathways, the K29-linked poly-ubiquitination also is involved in kinase modification, blocking their activity\textsuperscript{51}. In contrast to the degradation signals of K11-, K29- and K48-linked Ub chains, K63-linked poly-ubiquitin chains participate in endocytosis, DNA repair and in signaling kinase complexes\textsuperscript{52-59}. While several reports show K6-linked poly-ubiquitination, mediated by the heterodimeric RING E3 ligase complex, BRCA1/BARD1, might be involved in DNA repair\textsuperscript{60-62}. Although mass spectrometry proteomics did find K27- and K33-linked Ub chains, no clear cellular role has been associated with these types of Ub chains\textsuperscript{21}. A U-box E3 ligase, Ufd2, was found to catalyze the formation of K27- and K33-linked Ub chains, which suggests a role in the UFD pathway\textsuperscript{63}. The K33-linked Ub chains have also been shown to be involved in kinase modification\textsuperscript{51}.

The most recently described type of Ub polymer is the linear Ub chains, which are assembled by a specific ligase complex called the linear Ub-chain assembly complex (LU BAC) and are crucial for NF-\textsuperscript{κ}B signaling\textsuperscript{19,64-66}. Mass spectrometry proteomics has also shown that Ub chains can have more complex topologies. Doubly modified Ub peptides indicative of branched Ub chains have been detected\textsuperscript{67}. The ubiquitination of lysine residues in close proximity of each other as a product of \textit{in vitro} reactions were detected and the K27/K29 forks have been found in yeast cells\textsuperscript{68}. It is still unclear which type of signals these branched Ub chains mediate, however they show the complexity of the Ub system.

The function, structure and physiological roles of K48- and K63-linked chains in ubiquitination have been studied and published extensively. However, relatively only a few reports deal with the remaining chain types, some of which have not been studied at all. A likely reason is that only K48- and K63-linked chains have been available and protocols for generation of these Ub chains have been published\textsuperscript{69-72}. El Oualid and colleagues recently published the chemical synthesis of all types of di-ubiquitin\textsuperscript{73}, which in \textit{chapter 4} are utilized for investigating the differential modulation of the activity of a sub-family of DUBs.

**Figure 3.** Different types of ubiquitination. Image adapted from\textsuperscript{19}.
Reading out of ubiquitin signals in different cellular processes

In all the processes described above, Ub acts as a signaling component that can trigger molecular events. This is achieved by operating as a reversible and highly versatile regulatory signal for ubiquitin-binding domains (UBDs) in cellular proteins. The effects of protein ubiquitination are mediated through a class of specific UBDs, with more than twenty different families identified to date. UBDs are diverse modules in a protein that can bind, and often distinguish, different types of Ub modifications. The UBDs differ both in structure and in the type of Ub recognition that they use. Most commonly, they fold into α-helical structures that bind a hydrophobic patch in the β-sheet of Ub. Other UBDs bind Ub through two discontinuous α-helices. It is currently unknown why so many variations of helical structures have evolved to interact with Ub and regulate its downstream signaling. The structures of many UBDs in complex with mono-ubiquitin have been determined revealing interactions with multiple surfaces on Ub and not just the Ile44 hydrophobic patch region. The affinities of the mono-ubiquitin binding of these domains span a wide range (Kd = 1-100µM). As a crude generalization binding domains in enzymes with activities in or regulated by ubiquitination tend to have affinities in the low micromolar range. Adaptors that bind ubiquitinated proteins contain domains that typically bind mono-ubiquitin with affinities of about 100µM or lower.

Linkage-specific Ub recognition contributes to the diverse set of functional cellular processes associated with ubiquitination. Although some UBDs showed little discrimination between different Ub chain types, others do prefer certain ubiquitin-linkages. Several UBDs can selectively bind to K63-linked Ub chains. For example, NZF domains of TAK1-binding protein 2 (TAB2; also known as TRAF-binding protein domain) preferentially bind to K63-over K48-linked Ub chains. The sequence between the two UIMs of RAP80 promotes an appropriate protein conformation such that the UIMs are positioned for efficient binding across a single K63 linkage, thus defining selectivity. As mentioned in the previous paragraph, the linear Ub chains have been implicated in the activation of NF-κB signaling pathway. Several proteins that regulate this pathway, including NF-κB essential modulator (NEMO), A20-binding inhibitor of NF-κB proteins (ABIN) and optineurin, contain the Ub binding in ABIN and NEMO domain (UBAN domain), which specifically binds to linear Ub chains. Specific mutations of NEMO that block interactions with linear Ub chains impair the activation of IkB kinase (IKK) and NF-κB in response to tumor necrosis factor-α (TNF-α) stimulation.

UBDs in DUBs

As stated previously, also DUBs contain UBDs, which might directly regulate their activity or specificity. USP25 for example requires UIMs to efficiently hydrolyze poly-ubiquitin chains. UIMs can also play a role in Ub linkage specificity. The UIM of ATXN3 for instance is necessary for K63 selectivity. Another common UBD found in DUBs is the ZnF-UBP domain, which in the case of USP5, binds to proximal Ub in the chain and induces an allosteric conformational change leading to an increased catalytic rate for Ub chain processing. Additionally, USP5 also possess two UBA domains, which are involved in binding two different poly-ubiquitin chains, linear and K48-linked. Recently, the crystal structure of the yeast Ubp8 (USP22) in complex with the SAGA (Spt-Ada-Gcn5 acetyltransferase) module was solved showing how the ZnF-UBP domain acts as a scaffold for complex assembly by holding all subunits of the module together. New classes of UBDs are continuously being reported, revealing new information not only about their functionality but also about the mechanisms of intermolecular regulation. The accumulated knowledge of Ub-UBD interactions will also have an impact on pharmacological and medical applications, as several UBDs have now been linked to various human pathologies, including cancer and immune deficiencies, thus becoming interesting as putative drug targets.
cellular proteins is a major post-translational modification that can have profound effects on protein stability, localization or interaction pattern. These changes in the fate of individual proteins can cause alterations in cell signaling, which regulate cell cycle, proliferation or apoptosis. Tampering with the ubiquitination machinery has been observed in various cancers and neurological diseases. Multiple Ub pathway proteins, which directly control stability of key signaling molecules, have been identified as tumor suppressors or oncogenes.

The best studied example in which the deregulation of ubiquitination plays an important role in tumorigenesis and cell death is the p53 signaling pathway. It is a highly complex, multi-component signaling network that is of fundamental importance in tumor biology. p53, commonly known as 'the guardian of the genome', is a tumor suppressor that is capable of inducing cell-cycle arrest, cell senescence and apoptosis. Loss of p53 expression causes cells to lose pivotal signaling responses to DNA damage events, hypoxia and aberrant oncogene expression as seen in a large percentage of human tumors. The proliferation of oncogenic cells results in the repression of p53 expression and activation. P53 is predominantly regulated through the ubiquitin-proteasome pathway and maintained at low protein levels during normal homeostasis. The major player in controlling p53 levels is Mdm2. Mdm2 is a RING finger domain containing protein that exhibits E3 ubiquitin-protein ligase activity and is capable of regulating its own protein levels through auto-ubiquitination. Genotoxic stress can induce high levels of p53 promoting the expression of many proteins associated with apoptosis and cell-cycle arrest to counteract the stressors causing DNA damage. On the contrary, if p53 is overactive, the cell will die, so the p53-mediated increase in Mdm2 level helps to regulate p53 levels by polyubiquitination and targeting p53 for proteasomal degradation. Besides Mdm2 a few more E3 ligases have been reported to have similar activities in regulating p53 levels. COP1, Pirh2 and ARF-BP1 directly interact with p53 and target p53 for proteasome-mediated degradation. Decreased p53 levels and mutations in p53 have been linked to high incidences of cancer, thus p53 levels being elevated act as a mechanism of tumor suppression. Mdm2-dependent and independent ubiquitinations that target p53 for degradation are considered oncogenic and prevent p53 from regulating the cell cycle or inducing apoptosis.

Recent examples of the implication of ubiquitination in cancer include TRAF6-mediated K63-linked poly-ubiquitination of the kinase Akt, which is required for the membrane localization and activation of this proto-oncogene. Mutations in Akt in tumor cells, have shown to enhance its non-degradative ubiquitination and thus promote oncogenic signaling. Another example of the implication of the ubiquitination machinery in cancer is the NF-κB signaling pathway. The NF-κB transcription factor is activated through K63-linked poly-ubiquitination and its anti-apoptotic activity has been associated with tumor progression, chemotherapy resistance and metastasis. The key event in NF-κB activation is the degradation of the inhibitor of NF-κB (IκB) which releases NF-κB from its sequestration in the cytoplasm. Various pro-inflammatory stimuli, including TNF-α, IL-1, and DNA damage activate the IKK (IκB kinase) complex that phosphorylates IκB and recruits the SCF Ub ligase. In turn, this SCF Ub ligase mediates the IκB ubiquitination and proteasomal degradation. The recruitment of the IKK complex depends on the K63-linked poly-ubiquitination of adaptor proteins TRAF2 and RIP, where TRAF2 is the responsible E3 ligase. Respectively, NK-κB signaling can be terminated via either deubiquitination of the adaptor proteins or their regulated degradation. Deregulation of either mechanism can lead to tumor formation. Thus, SCF E3 ligase, which ubiquitinates IκB and down regulates NF-κB signaling, has been found mutated in a high percentage of human malignancies.

Mediation of DNA repair is another established function of non-degradative Ub signaling. And it has been shown that non-repaired Ub lesions can lead to tumorigenesis and defects in the DNA repair system are common in several cancer predisposition syndromes. BRCA1 is a tumor suppressor whose mutation leads to a high incidence of breast and ovarian cancers. And together with BARD1, BRCA1 forms a heterodimeric E3 ligase, which synthesizes poly-Ub chains of different topology. This activity of BRCA1 plays an important role in the homologous recombination (HR) DNA repair pathway. In the absence of BARD1, BRCA1 is destabilized and degraded by the proteasome. UBE2T, the E2 conjugating enzyme, and HERC2, the E3 ligase enzyme, are the newly identified factors that mediate pro-
general introduction

Teasomal degradation of BRCA1. HERC2 is also the E3 ligase that is necessary for the Ub signaling at the DNA damage sites. It forms a complex with other Ub ligases, RNF8 and RNF168 (mutated in the RIDDLE syndrome) thereby facilitating the recruitment of the E2 conjugating enzyme Ubc13 to the sites of DNA damage. After detection of DNA damage, the Ubc13/RNF8 complex initiates poly-ubiquitination of histone H2A, thereby promoting the recruitment of Ub-binding proteins, including RAP80 and RNF168. RNF168, amplifies histone ubiquitination increasing the local concentration of K63-linked Ub chains and ensuring the recruitment of other DNA repair factors, such as BRCA1 and 53BP1 to the sites of DNA damage.

More recently, the Ub machinery has been implicated in the other major degradation system: autophagy. Autophagy is an evolutionary conserved lysosomal degradation pathway that targets bulky cargos, such as protein aggregates and mitochondria. Tumors resort to the autophagy to survive hypoxia and lack of nutrition. Yet, in normal mammalian cells, autophagy is more important for cell cleansing of aggregated proteins and damaged organelles. Inhibition of autophagy in healthy tissues leads to accumulation of ubiquitinated protein aggregates and damaged mitochondria. This process has been linked to neoplastic transformation, which has shown to be associated with enhanced mutagenesis and deregulated cell signaling. The inhibition of autophagy in cancer cells, besides limiting their survival in the face of hypoxia and starvation, could lead to accumulation of damaged mitochondria and protein aggregates due to the inhibition of the selective, Ub-regulated forms of autophagy. This could sensitize tumor cells to cancer therapies that aim at inducing apoptosis.

It has become clear that the deregulation of the Ub-proteasome system (UPS) plays a vital role in the most important age-related neurodegenerative diseases, Alzheimer’s disease. In Alzheimer’s disease, hyper-phosphorylation of the microtubule-associated protein tau results in the accumulation of paired helical tau filaments involved in the formation of neurofibrillary tangles and neuritic plaques. Relative dysfunction or inhibitory overloading of the UPS may contribute to the abnormal accumulation of phosphorylated and ubiquitinated tau. Interestingly, a frameshift mutant of Ub, Ub+1, found in some sporadic and hereditary Alzheimer disease patients, inhibits the UPS and enhances the toxic protein aggregation in a yeast model.

Implication of DUBs

There is a growing recognition of DUBs that are mutated in human cancers playing an important role as oncogenes and tumor suppressors. As mentioned previously, stabilization of p53 activates downstream targets to initiate cell-cycle control, DNA repair mechanisms and apoptosis. This p53 stabilization can be achieved by inhibiting Mdm2-mediated ubiquitination. Alternatively stabilization can also be achieved by deubiquitination catalyzed by DUBs to reduce ubiquitination of p53. Several USPs, including USP7 (HAUSP), USP2 and USP10, have been reported to regulate p53 and/or Mdm2.

HAUSP was the first USP identified to deubiquitinate p53, with over-expression of HAUSP resulting in p53 stabilization leading to the induction of p53-dependent cell growth repression and apoptosis. However, depletion of HAUSP in cells doesn’t decrease p53 levels as predicted, but rather increases p53 levels, apparently because of HAUSPs ability to bind and deubiquitinate Mdm2. The relationship of HAUSP and cancer was complicated by the demonstration of HAUSP mediated regulation of Akt antagonist PTEN. Reduced nuclear localization of PTEN is a common feature in aggressive cancers, including advance stage prostate cancer. PTEN nucleo-cytoplasmic shuttling is regulated by PTEN ubiquitination. HAUSP, which is overexpressed in prostate cancer, was shown to deubiquitinate PTEN leading to reduced nuclear localization. These data suggest that HAUSP may in some instances act as a tumor suppressor by stabilizing p53 or as an oncogene by stabilizing Mdm2 and re-distributing PTEN.

In a yeast-two hybrid screen, USP2 was found to complex with Mdm2. USP2 deubiquitates Mdm2 leading to its stabilization. Oncomine data suggest that many cancers have reduced expression of USP2 including cancers of the colon, pancreas and head/neck. Interestingly, USP2 was linked to prostate cancer through a study to isolate androgen sensitive DUBs from a prostate cancer cell line. USP2 was found to bind and stabilize fatty acid synthase, a protein that is often found overexpressed in aggressive prostate cancers.

USP10 is overexpressed in breast cancer tissue compared to adjacent normal tissue and...
in glioblastoma samples\textsuperscript{127,128}. Recently, USP10 was identified as a DUB for p53. In unstressed cells, USP10 mainly localizes in the cytoplasm and regulates p53 homeostasis. After DNA damage, USP10 translocates to the nucleus and contributes to p53 activation. Increased USP10 expression in mutant p53 background increases p53 levels and promotes cancer cell proliferation, while downregulation of USP10 inhibits cancer cell growth. Increased expression of USP10 could be another mechanism responsible for increased mutant p53 expression in human cancers\textsuperscript{121}.

Another example of the implication of DUBs in cancer is the familial cylindromatosis tumor suppressor gene, CYLD. Patients with familial cylindromatosis have the predisposition for developing multiple skin tumors of the head and neck. Germline mutations in the CYLD gene were identified in patients and most likely abolish the catalytic activity of CYLD\textsuperscript{129,130}. CYLD negatively regulates NF-\kappa B signaling through its deubiquitinating activity. As noted above, several intermediates of the NF-\kappa B pathway become K63-linked poly-ubiquitinated. By removing these K63-linked Ub chains, CYLD dampens the NF-\kappa B signaling\textsuperscript{131-134}.

The importance of the DUB activity in tumor progression is also illustrated by the recent findings that USPs are often overexpressed in tumors to stabilize oncogenic potential. USP9X for example was found overexpressed in various hematological malignancies. Initially USP9X in mammals was found to have a variety of functions including important intersections with cancer pathways: the Wnt and TGF\beta pathway\textsuperscript{135-138}. Whether USP9X mediated deubiquitination plays a role in tumorigenesis remained unclear, and one could predict that overexpressed USP9X would enhance tumorigenesis by enhancing the Wnt and TGF\beta signaling. Recent findings by Schwickart and co-workers support this idea and found USP9X overexpressed in various hematological malignancies\textsuperscript{139}. USP9X overexpression leads to the promotion of cell survival by removing K48-linked poly-ubiquitin chains that would otherwise target the pro-survival BCL-2 member, MCL1, for proteasomal degradation. This contributes to the disease progression and chemoresistance\textsuperscript{139}.

Other USPs have been identified to be involved in tumorigenesis, however how these USPs are implicated remains unclear. The increase of USP6 transcription has been shown to lead to the development of bone tumors\textsuperscript{140}. USP1 has been identified as a negative regulator of FANCD2 (an important factor in Fanconi Anemia) mono-ubiquitination and DNA repair and was also found to deubiquitinates PCNA, an important component of the trans-lesion synthesis (TLS) repair pathway\textsuperscript{141-143}. The Oncomine database reveals the overexpression of USP28 in primary colon and breast cancer samples\textsuperscript{124}. USP28 has been identified to deubiquitinates Myc, a central player in many forms of cancer, thereby salvaging it from proteasomal degradation\textsuperscript{144,145}.

It is clear that the ubiquitination and deubiquitination pathway is tightly regulated and the deregulation in both processes has been shown to play a direct or indirect role in tumorigenesis and several inherited neurological diseases.

**USP family**

The ubiquitin-specific protease subclass represents the largest of the DUB family encoded by the human genome\textsuperscript{146}. In the last decade the crystal structures of the catalytic domain of several USPs, with or without Ub bound, have been solved (Figure 4) (USP8CD [PDB code: 2GFO], USP7CD [1NB8 & 1NBF], USP14CD [2AYN & 2AYO], USP2CD [2HD5], USP21CD [3I3T], Ubp8 [3M99 & 3MHH] and CYLD [2VHF])\textsuperscript{88,89,147-150}. The catalytic domain structures revealed several common structural features between the different USPs. First, despite the low sequence similarity the overall structure of the catalytic domain is highly conserved. Three well-defined sub-domains: the ‘Fingers’, the ‘Thumb’ and the ‘Palm’ domain form a structure that resembles a right-hand (Figure 4). Secondly, the catalytic centre resides at the interface between the Thumb and the Palm regions. The Thumb is predominantly \alpha-helical and contains the Cys Box, a motif that includes the active site cysteine. The Palm is composed of \beta-strands supported by \alpha-helices and contains the remaining active site residues that form the catalytic triad: a His and an Asp or Asn residue. The junction between the Thumb and the Palm domains forms a cleft that accommodates the C-terminal tail of Ub and the active site residues involved in the catalysis. Third, the Finger domain is composed of four \beta-strands and in USP8, USP2 and USP21 it contains a CXXCxCXXC motif that chelates one zinc ion and form the so-called...
Figure 4. Crystal structures of the catalytic domains of USPs.
Although not all of the solved structures contain this motif, the overall fold of the Finger domain is maintained. The CXX-CXnCXxC motif is lacking in nine of the 54 putative human USPs, suggesting that the zinc binding ability is dispensable for the integrity of the Finger domain fold. The role of the Finger domain is to serve as a scaffold that contacts the globular body of Ub and the Zinc Finger ribbon does not appear to be involved in catalysis. Interestingly, the Finger domain of CYLD is significantly smaller due to the shortening of the \( \beta \)-strands. CYLD also differs from the other USPs by the insertion of a Zn binding domain that closely resembles a B-box. The B-box is not required for deubiquitinating activity, but instead appears to be important for the cytoplasmic localization of CYLD.

Furthermore multiple studies have demonstrated that the catalytic activity of USPs is regulated by substrate- or scaffold-induced conformational changes. For example USP7 in the unliganded form has a misaligned catalytic triad configuration. The active site cysteine is approximately 10Å away from the active site histidine, too far for catalysis to occur. Upon binding to Ub a major conformational change occurs in the catalytic core domain causing the active site cysteine and histidine to be positioned within hydrogen bond distance from one another, rendering the enzyme catalytically competent.

The structure of USP14 in the presence and absence of Ub reveals a different type of activation mechanism. Unlike USP7, the catalytic triad of the free USP14 is productively aligned, indicating that the active site is catalytically competent. However, the binding groove that accommodates the C-terminal tail of Ub is blocked by two surface loops that undergo significant conformational changes upon Ub binding.

A third conformational change is thought to occur in USP8. In the structure of the unliganded form, the tip of the Zinc finger ribbon of USP8 is positioned inward toward the Palm resulting in a closed conformation that leaves insufficient room for Ub. Unlike the other USPs, USP8 and USP2 have a unique \( \alpha \)-helix juxtaposed to the finger and this helix may be involved in stabilizing the closed conformation. Although there is no structure available of USP8 in complex with Ub, the closely related catalytic domain of USP2 has been solved in the presence of Ub. In this structure the Finger domain is displaced outward to adopt the conformation observed in the other USPs. It is possible that upon target protein binding, the Finger domain of USP8 moves to the position observed in the other USPs, allowing the activation of the protease and subsequent binding to Ub.

Recently, the crystal structure of the yeast Ubp8 (USP22) in complex with the SAGA module was solved both in presence and absence of Ub. The structures showed that the Finger domain that binds Ub in the complex remains in an open conformation in the apo form. The Finger domain doesn’t collapse and occlude the Ub-binding pocket as seen in USP8. Furthermore, the active site residues are in their catalytically competent orientation in both the presence and absence of Ub. Interestingly, several residues located adjacent to the Ub C-terminus are disordered in the absence of Ub and a small loop containing the active site cysteine (Cys-loop) move inward toward the Ub tail-binding groove. Both the Cys-loop and the disordered region contain residues that contact Ub directly. Ubp8 is activated by binding of the Sgf11 subunit of the SAGA module to the catalytic domain of Ubp8 near the active site, supporting the construction of a competent catalytic center of the enzyme.

The USP family members have modular domain architecture and besides having the conserved catalytic domain they also feature additional protein-protein and localization domains. A number of domains outside the catalytic domain have been identified and characterized, for example the TRAF-like domain of USP7, the DUSP (domain in USPs) domain of USP15 and as described previously several UBDs found in USPs. Interestingly, Ubl domains were recently predicted in a subset of USPs.

Ubl domain

In a sequence analysis integrated ubiquitin-like (Ubl) domain were identified in a large number of USPs. The integrated Ubl domain is a member of a subfamily of the Ub superfamily. Despite the poor sequence conservation for all members of the Ub superfamily, they all share the conserved \( \beta \)-grasp fold. The integrated Ubl folds are stretches of 45-80 amino acids and are found in a wide array of eukaryotic proteins. Only a small number of these integrated Ubl domains have been studied and characterized functionally. These integrated
Ubl domains are present in proteasomal shuttle factors like Rad23 and Dsk2 and are believed to play a role in recruitment of ubiquitinated proteins to the proteasome. Integrated Ubl domains in other proteins such as Parkin and USP14 can also be recruited to the proteasome. However, the integrated Ubl folds not only function in recruitment but also play a role in enzymatic activities of certain immune-response inducible kinases, such as IKKβ\(^{155,156}\).

The Ubl domains found in the USP family are located at the N-terminus, within or at the C-terminus of the catalytic domain. Till recently only the Ubl domain at the N-terminus of USP14 has been characterized. It has been shown to be involved in proteasome binding, which promotes the DUB activity of USP14\(^{149}\). In chapter 5 a Ubl domain located within the catalytic domain of a USP is described and its effect on DUB activity.

**USP4**

Like USP14, USP4 has a Ubl domain N-terminal of its catalytic domain, but more interestingly it has a second Ubl domain within the catalytic domain. Furthermore, the DUSP domain is located N-terminal of the first Ubl domain. Not much is known about the structure and function of USP4, previously known as UNP for ubiquitous nuclear protein. It has been identified as a proto-oncogene related to *tre-2/tre-17* (USP6) by its ability to transform NIH3T3 cells and lead to increased tumorigenesis in nude mice\(^{157,158}\). In a study of primary human lung tumor tissue, USP4 was observed to have a consistently elevated gene expression levels in small cell tumors and adenocarcinomas of the lung, suggesting a possible causative role for USP4 in neoplasia\(^{159}\).

Besides the described domains of USP4, it also possesses functional nuclear import/export signals, the basis of its ability to shuttle between nucleus and cytoplasm\(^{160}\). Because of this ability, USP4 seems to have different cellular functions depending on its localization. When located in the cytoplasm, it has been reported that knockdown of USP4 activates the β-catenin-associated transcription\(^{161}\). By interacting with two known Wnt signaling components Nemo-like kinase and T-cell factor 4, USP4 may play a role in the Wnt signaling pathway in a variety of physiological conditions.

Interestingly, Song and co-workers show that USP4 is recruited to the spliceosome by forming a complex with Sart3\(^{162}\). The N-terminal DUSP domain and the adjacent DUF1055 domain (amino acids 27-216) of USP4 is primarily responsible for this interaction with Sart3. The Sart3 protein is a recycling factor of the U4/U6 splicing snRNP, which promotes the re-annealing of U4 and U6 snRNPs. The crystal structure of the N-terminus of USP4 (PDB: 3JYU) shows that the DUF1055 domain is actually a Ubl domain which interacts extensively with the DUSP domain. It is likely that binding to the Sart3 protein might abolish this interaction and hence recruit USP4 to the spliceosome complex where it preferentially deubiquitinates K63-linked chains on the U4 component Prp3. The deubiquitination of Prp3 most likely facilitates the release of Prp3 from the spliceosome during maturation of its active site\(^{162}\).

In addition to its cytoplasmic roles, recent evidence shows that USP4 also plays important nuclear roles. Fan and co-workers show that the tumor necrosis factor-α (TNFα), a pro-inflammatory cytokine, induces association of USP4 with the transforming growth factor-β-activated kinase 1 (TAK1), which leads to the deubiquitination of Lys63-linked polyubiquitinated TAK1 and thereby down-regulating the TAK1-mediated NF-κB activation\(^{163}\).

Furthermore, Zhang and co-workers recently identified a new important role for USP4. They show that USP4 interacts directly with and deubiquitinates ARF-BP1, leading to the stabilization of ARF-BP1 and subsequent reduction of p53 levels\(^{164}\). ARF-binding protein 1 (ARF-BP1; also known as HUWE1) was recently identified as another critical E3 Ub ligase in regulating p53 levels\(^{99}\). ARF-BP1 is a HECT domain-containing E3 Ub ligase, which interacts directly and ubiquitinates the p53 protein. This p53 ubiquitination is strongly repressed by the binding of ARF to ARF-BP1. The inactivation of ARF-BP1 stabilizes p53 and induces apoptosis.

It seems that USP4 plays several important cellular roles depending on its location in the cell. And it is therefore interesting to see what future research will unveil on how USP4 function is being regulated. Chapter 5 shows how the catalytic activity of USP4 is being regulated.

**Structural genomics**

In the 21st century with the rise of new efficient
genome sequencing techniques, microarray experiments and new proteomics technologies, biology has seen huge increase in DNA sequence, gene expression and proteomics data. The Human Genome and other genome sequencing projects continue to deliver new protein sequences in a rapid pace. New protein families are being discovered in newly sequenced genomes, for which a large fraction of these new protein families we virtually don’t have any functional or structural data.

To counteract this trend a number of structural genomics (SG) programs were created with the aim to determine rapidly a large number of novel structures in order to expand structural and functional knowledge for proteins found in genomes. The largest SG centers are funded by Japan (RIKEN), the United States [Protein Structure Initiative] (PSI and PSI2) and by an international consortium of governments, charitable foundations and industry [the Structural Genomics Consortium] (SGC), while in Europe The Structural Proteomics In Europe (SPINE and SPINE2) was set up. These SG centers are using different approaches to structurally characterize the protein world165-168. RIKEN and the SGC focuses on the human proteome165, while PSI and PSI2 target the representative members of the largest protein families, proteins from human parasite and Mycobacterium tuberculosis168,169. SPINE and SPINE2 introduced high-throughput proteomics in a wide collaborative platform in which SPINE initially was focused on targets of medical relevance. SPINE2 continued and extended these methodologies to study macromolecular complexes170.

Additionally, most SG centers have dedicated a significant part of their efforts in the development of high-throughput methods, which may now also be used for fast and more accurate determination of structures by both X-ray crystallography and NMR techniques in laboratories not involved in SG efforts. The work on thousands of target proteins has led to the development of efficient protocols for each of the steps of the structure determination process167,171. New experimental protocols that were developed through SG efforts have shifted over time the so-called ‘bottlenecks’ in the pipeline and it seems that at present the analysis of 3D structures in the context of all biological and bioinformatics information is the slowest step of the whole process.

As part of the SPINE network, we did our share of work in developing new tools and efficient protocols to aid the initial steps of structure determination and are described in chapter 3.

Outline of this thesis
Ubiquitination is a reversible protein modification that plays a central role in many important cellular functions. The enzymes involved in this ubiquitination process have been the main focus of many studies for several decades. However, deubiquitinas (DUBs), responsible for the removal and processing of ubiquitin, have been shown to be equally important. Over the last decade a lot of progress has been made in the characterization of this large family of isopeptidases. Because of their roles in key regulatory processes, these DUBs are actively pursued as new drug targets. The aim of this thesis was to study the structure and function of members of the ubiquitin-specific protease family.

In Chapter 2 we give a comprehensive overview of putative DUBs encoded by the human genome. Furthermore, the function, specificity and the regulation of DUB activity is discussed.

In order to study a large number of USPs in parallel, high-throughput tools and methods needed to be developed. In Chapter 3 a description of a set of protein expression vectors for ligation-independent cloning is given and their use on a large number of the USP family members.

In Chapter 4 a set of twelve USPs in presence and absence of modulators have been analyzed for their enzyme activity using synthetic reagents. Among these reagents are all seven wild-type lysine-linked di-ubiquitins and the first comprehensive analysis comparing Ub chain preference is given.

In Chapter 5 we present the crystal structure of the catalytic domain of USP4 and we show a new function of its Ubl domain located within the catalytic core.

A summary of the findings, concluding remarks and implications for future research are presented in Chapter 6.
References


51. Al-Hakim, A.K. et al. Control of AMPK-related kinases by USP9X and atypical Lys(29)/Lys(33)-linked polyubiquitin...


102. Wu, Y. et al. Stabilization of snail by NF-kappaB is required for inflammation-induced cell migration and inva-


141. Grunda, J.M. et al. Increased expression of thymidylate synthetase (TS), ubiquitin specific protease 10 (USP10)


153. Zapata, J.M. et al. A diverse family of proteins containing tumor necrosis fac-


A Genomic and Functional Inventory of Deubiquitinating Enzymes

Sebastian M.B. Nijman¹,* Mark P.A. Luna-Vargas¹, Arno Velds¹, Thijn R. Brummelkamp², Annette M.G. Dirac¹, Titia K. Sixma¹, and RenéBernards¹*

*Contact: s.nijman@nki.nl (S.M.B.N.), t.bernards@nki.nl (R.B.)

¹Division of Molecular Carcinogenesis and Center for Biomedical Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
²Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, USA

Cell 2005;123(5):773-86 Review
Abstract
Posttranslational modification of proteins by the small molecule ubiquitin is a key regulatory event, and the enzymes catalyzing these modifications have been the focus of many studies. Deubiquitinating enzymes, which mediate the removal and processing of ubiquitin, may be functionally as important but are less well understood. Here, we present an inventory of the deubiquitinating enzymes encoded in the human genome. In addition, we review the literature concerning these enzymes, with particular emphasis on their function, specificity, and the regulation of their activity.

Introduction
Over the last few decades, protein modification by ubiquitin (Ub) and ubiquitin-like (Ubl) molecules has emerged as a critical regulatory process in virtually all aspects of cell biology. Indeed, the 2004 Nobel Prize in Physiology or Medicine was awarded for the discovery of Ub-mediated proteolysis.

More than a dozen different Ub and Ubl modifications have been described, and up to 20% of yeast proteins are conjugated to Ub under standard culture conditions\(^1,2\). In yeast, potentially all seven conserved lysines of Ub itself (K6, 11, 27, 29, 33, 48, and 63) are used as branching sites for the generation of Ub polymers.

The topic of ubiquitination, the proteins involved, and their functions in various pathways and signaling networks has been well reviewed\(^3,4\). Here, we discuss the enzymes that remove Ub from polypeptides. These deubiquitinating enzymes (DUBs) play key regulatory roles in a multitude of processes from hereditary cancer to neurodegeneration. Despite the importance of DUBs, our knowledge of their mode of regulation and substrate specificity is surprisingly scant. A detailed annotation of individual family members of this enzyme group is an important step toward elucidating the molecular functions of DUBs in health and disease.

To this end, we provide a comprehensive overview of putative DUBs encoded in the human genome. In addition, we discuss the lacunae in our understanding of these enzymes by drawing on examples from yeast and higher eukaryotes.

In our attempt to classify these enzymes, we have made some arbitrary decisions as to which genes to include or exclude as potential DUBs. Therefore, we present three caveats to this list. First, we cannot exclude the fact that proteins or protein families not included in this overview can remove Ub from polypeptides. For instance, a recent in silico effort to predict new Ub signaling components suggested a previously undetected family of Ub peptidases\(^5\). Second, protein domain prediction based on gene transcripts depends on consensus sequences. Thus, divergent but true family members can be missed due to low homology scores. Finally, we wish to emphasize that it is unlikely that all predicted DUBs are truly specific for Ub: some will display additional activity or exclusive activity toward Ubl molecules.

DUBs Are Proteases
DUBs belong to the superfamily of proteases, of which an estimated 561 members are present in the human genome\(^6\). Based on the mechanism of catalysis, proteases are divided into five classes—aspartic, metallo, serine, threonine, and cysteine proteases—and further subdivided based on phylogeny.

Two classes of proteases (cysteine and metallo) contain DUBs, although most DUBs are cysteine proteases. By definition, the enzymatic activity of cysteine proteases relies on the thiol group of a cysteine in the active site. Deprotonation of this cysteine is assisted by an adjacent histidine, which is polarized by an aspartate residue. These three residues make up the catalytic triad. During catalysis, the cysteine performs a nucleophilic attack on the carbonyl of the scissile peptide bond, which, in the case of DUBs, is between the target and Ub. The intermediate, which contains an oxanion, is stabilized in the so-called oxanion hole. This oxanion hole is generally provided by a glutamine, glutamate, or asparagine residue and the main chain of the catalytic cysteine. The result of the reaction is release of the target protein and formation of a covalent intermediate with the Ub moiety. Reaction of this intermediate with a water molecule results in the release of the free enzyme and Ub.

In contrast to cysteine proteases, metalloproteases generally use a Zn\(^{2+}\) bound po-
The Human DUB Genes

The cysteine protease DUBs can be further organized into four subclasses based on their Ub-protease domains: ubiquitin-specific protease (USP), ubiquitin C-terminal hydrolase (UCH), Otubain protease (OTU), and Machado-Joseph disease protease (MJD). All DUBs that are metalloproteases have a Ub protease domain called JAMM (JAB1/MPN/Mov34 metalloenzyme). The structures of the catalytic domains of the different subclasses of DUBs reveal an impressive diversity in secondary structure (Figure 1).

We used the ENSEMBL human genome database (v32, July 2005) to retrieve all putative DUBs from the human genome by selecting genes whose transcripts encode one of the five Ub-protease domains. Our search identified all known DUBs except two DUBs with OTU domains (Otubain-1 and Otubain-2). This analysis indicated that the human genome encodes approximately 95 putative DUBs, including many that have not been previously reported. These can be broken down into 58 USP, 4 UCH, 5 MJD, 14 OTU, and 14 JAMM domain-containing genes, many of which are associated with multiple transcripts (see Table S1 in the Supplemental Data available with this article online). For six unnamed genes, we have submitted gene names to the HUGO gene nomenclature committee (HGNC) (Table S1). To determine whether the putative DUB genes are expressed, we searched NCBI human-expressed sequence tag (EST) databases for transcripts corresponding to the predicted protein sequence. We obtained further evidence for expression of a number of genes with relatively low numbers of ESTs from additional sources (such as SAGE and UniGene). For five predicted DUBs, we could not find any convincing data supporting transcription.

Next, we generated sequence alignments to ensure conservation of the catalytic residues and made an inventory of DUBs reported to display Ub protease activity. This indicated that of the 90 putative DUBs that are expressed, 11 are unlikely to display Ub-protease activity. Together, these data indicate that humans express approximately 79 putative DUBs that are functional (Table S1).

To investigate sequence homology between the various putative DUBs, we used two strategies. We used CLANS (Cluster analysis of sequences) software to visualize pairwise all-against-all sequence BLAST matches. As expected, very few positive BLAST results were found between the five subclasses, whereas the members within the subclasses clustered together (Figure 2A). This analysis revealed that within the subclasses some relatively divergent members are present (for example, Otubain-1 and Otubain-2, USP55, and CYLD). Similar results were obtained using the CLUSTAL alignment algorithm (Figure 2B).
The Five DUB Subclasses

Three-dimensional structures of DUB catalytic domains from all subclasses, some of them in complex with Ub derivatives, have been solved (Figure 1). These studies reveal intriguing similarities and differences between the four cysteine protease subclasses. This topic, including the structure of a JAMM domain, has recently been extensively reviewed by Amerik and Hochstrasser9. Here, we comment on the new structural features of the DUB subclasses.

**Ubiquitin C-Terminal Hydrolases (UCHs)**

The human UCH subclass of DUBs consists of four proteins that share close homology in their catalytic domains. Structural and biochemical studies have indicated that the UCH subclass of DUBs prefers to cleave relatively small protein substrates (up to 20–30 amino acids) from Ub9. This size limit is thought to be imposed by a loop that partially occludes the active site of these enzymes. However, recent biochemical and structural studies show that certain large substrates can nevertheless be accommodated10.

Although UCHs were the first described DUBs, their specific functions remain poorly understood. UCHs are thought to mainly act in the recycling of Ub when Ub is inappropriately conjugated to intracellular nucleophiles (for example, glutathione, polyamines). They also may be involved in the processing of newly synthesized Ub, which is translated either as a polyubiquitin precursor or fused to ribosomal protein precursors. However, other DUBs also display *in vitro* activity toward linear Ub fusions, suggesting that processing of newly synthesized Ub is performed by multiple DUBs.

Some studies suggest a role for UCHs in specific Ub-regulated processes. Mutations in UCH-L1 (a UCH

![Figure 2. Phylogenetic Map of Human DUBs](https://example.com/figure2.png)

**Figure 2. Phylogenetic Map of Human DUBs**

(A) Graphic two-dimensional representation of sequence similarities between all Ub protease domains of DUBs using CLANS software. CLANS performs all-against-all BLAST searches and uses the significant high-scoring segment pairs (HSPs) to draw a three-dimensional graph represented here in two dimensions. Each node represents a Ub protease domain and each edge (line) represents a significant HSP (edges are shaded according to p value). DUB subclasses are highlighted in the graph. The start and end positions of the DUB Ub-protease domains, as defined by Interpro, were used to generate the protein sequences. Proteins with a partial or short and misaligning DUB domain were excluded from the analysis. (B) Unrooted dendrogram of the DUBs using Clustal software. Clustal generates a multiple sequence alignment file based on pairwise alignments. From this information a phylogenetic tree can be constructed. The robustness of the phylogenetic relations can be assessed by “bootstrapping,” a mathematical technique that introduces noise in the alignment and measures how often the phylogenetic relationships reproduce. An asterisk indicates a bootstrapping percentage <10% (lowest branch only). The numbers correspond to the genes in Table S1.
specifically expressed in neurons) that reduce its DUB activity have been described in two siblings with Parkinson’s disease (PD), and a polymorphism in this gene has been linked to reduced PD risk\textsuperscript{11,12}. However, not all studies have found a strict relationship between UCH-L1 activity and PD. Furthermore, although mice that have a mutation in \textit{Uch-L1} exhibit neurodegeneration, they do not display PD-like symptoms\textsuperscript{13}.

\textbf{Ubiquitin-Specific Proteases (USPs)}

The USP subclass represents the bulk of the DUBs encoded by the human genome. As the number of Ub E3 ligases (the third factor in the ubiquitination cascade that determines target specificity) increased during evolution, so did the number of USPs, suggesting an intimate relationship between the two resulting in their co-evolution\textsuperscript{14}.

The catalytic domain of USPs contains two short and well-conserved motifs, called Cys and His boxes, which include the residues critical for catalysis. However, the size of the complete domain varies from approximately 300 to 800 amino acids due to the large unrelated sequences that are interspersed between the two motifs, which may serve a regulatory function (Figure 3).

Upon closer examination of the catalytic domains of USPs, we noted that a subset (USP16, USP30, USP39, USP45, and USP52) lack catalytic residues previously thought to be critical for protease activity (Figure S1). USP30 and USP16 lack only the aspartate in the catalytic triad but retain enzymatic activity against a model substrate (Table S1 and M.P.A.L.-V. and T.K.S., unpublished data). This indicates that, as is the case for the Otubain-2 protein, USP30 and USP16 may use a different residue to stabilize the active site histidine. Additional structural information about USPs may shed light on this issue.

USP39 (also known as SAD1) does not contain the conserved catalytic cysteine or histidine and does not cleave a model substrate \textit{in vitro}, indicating that USP39 is not a bona fide DUB (M.P.A.L.-V. and T.K.S., unpublished data). However, USP39 plays a critical role in spliceosome maturation in both yeast and human cells, and many of the other residues within the catalytic domain are conserved\textsuperscript{15,16}. Therefore, it is tempting to speculate that USP39 can still interact with Ub. An analogous situation exists in Ub conjugation. Here, a Ub interaction motif known as UEV (ubiquitin-conjugating enzyme variant) strongly resembles the catalytic domain of E2s (the second enzyme in the ubiquitination cascade) but lacks activity\textsuperscript{17}. In keeping with this nomenclature, these catalytically inactive USP domains are hereby referred to as USPV (ubiquitin-specific protease variant). The functions of these variants with respect to Ub await further investigation.

\textbf{Machado-Joseph Disease Protein Domain Proteases (MJDs)}

A bioinformatics search for other classes of Ub proteases identified Ataxin-3 and a number of Ataxin-3-like proteins\textsuperscript{18}. Experiments \textit{in vitro} confirmed that wild-type Ataxin-3, but not a mutant with the active site cysteine mutated, could deubiquitinate a model substrate\textsuperscript{19}. Sequence similarity between the catalytic domain of Ataxin-3 and other DUBs is low (Figure 2A), but recent NMR structures show that the overall arrangement of the catalytic triad is conserved (Figure 1)\textsuperscript{20,21}.

Instability of a CAG nucleotide repeat in the \textit{Ataxin-3} gene leads to a hereditary neurological condition known as spinocerebellar ataxia type-3 or Machado-Joseph disease (OMIM 607047). Like other polyglutamine neurodegenerative disease-associated genes, expansion of the CAG repeat in \textit{Ataxin-3} leads to protein misfolding, resulting in aggregation and cellular toxicity. Some experimental evidence indicates that the normal function of Ataxin-3 involves transcriptional regulation, but whether its DUB activity plays a role in this process remains unclear\textsuperscript{22}. In evolutionary terms, MJDs likely represent a relatively late addition to the Ub system, as no homologs have been identified in yeast. However, protease activity of the other family members has not yet been demonstrated, and their biological functions remain unknown.

\textbf{Ovarian Tumor Proteases (OTUs)}

A bioinformatics approach also led to the identification of the Ovarian Tumor (OTU) subclass of Ub proteases\textsuperscript{23}. The \textit{otu} gene is involved in the development of the \textit{Drosophila melanogaster} ovary where it may regulate the localization and translation of certain RNA transcripts\textsuperscript{24,25}. Using the \textit{Drosophila otu} gene and its homologs as a starting point, Makarova and colleagues found sequence similarity between these genes and those encoding viral cysteine
Figure 3. Comparison of the Domain Structures of Putative DUBs. For each primary DUB transcript (the transcript associated with the HUGO or RefSeq ID), we retrieved information concerning domain architecture and signal motifs using ENSEMBL SMART (simple modular architecture research tool; http://smart.embl-heidelberg.de/), Pfam (protein families database; http://www.sanger.ac.uk/Software/Pfam/), and PROSITE databases. The USPs without additional domains are indicated as “generic USP.” Only JAMM and MJD domain proteins with predicted catalytic activity are shown. An asterisk indicates that the ENSEMBL-predicted translational start site is uncertain. Proteins and domains are plotted on an approximate scale. Select abbreviations: ZnF, zinc finger; NLS_BP, bipartite nuclear localization signal; MATH, meprin and TRAF homology; DUSP, domain in ubiquitin-specific proteases. For additional information concerning the indicated domains visit http://www.ebi.ac.uk/interpro.
proteases. A recently solved OTU structure shows that, unlike other cysteine protease DUBs, the catalytic triad is incomplete and is stabilized by a new method involving a hydrogen bonding network.

Otubain-1 and Otubain-2 were the first two OTU proteins found to display in vitro DUB activity. Shortly thereafter, Cezanne, another OTU-domain containing protein, was found to interact with poly-Ub in a yeast two-hybrid assay and to contain DUB activity in vitro, suggesting that this is a general OTU feature. However, for most OTU proteases, their physiological role in vivo, including their putative role as DUBs, remains to be investigated.

**JAMM Motif Proteases**

The JAMM domain is found in all three major kingdoms of life (bacteria, archaea, and eukarya). However, bacteria do not contain Ub protease activity, and an analogous Ub-like conjugation system has not yet been identified in prokaryotes. This suggests that JAMM domains have adopted new protease functions during evolution and indicates that at least some of the human JAMM proteases may be involved in more than Ub (or Ubl) processing. Indeed, recent work has identified the protein product of the *Mycobacterium tuberculosis* gene *mec* as a JAMM domain peptidase involved in cysteine biosynthesis by cleaving cysteine from a peptide intermediate.

Sequence alignment of the JAMM-domain proteins revealed that seven of the 14 members have at least one amino acid change in the conserved Zn$^{2+}$ ion-stabilizing residues, indicating that they may not be functional proteases. Three family members (POH1, CSN5, and AMSH) will be briefly discussed in the section concerning DUB function.

**DUB Specificity**

Accumulating evidence indicates that most DUBs regulate a limited number of proteins and pathways, suggesting that they target specific substrates (Table 1). In the case of DUBs, specificity can refer to either the Ub or Ubl moiety itself (substrate specificity) or the target protein to which the moiety is conjugated (target specificity). In reality, it may not be possible to separate these types of specificities. It is likely that in many cases a combinatorial mechanism relying on recognition of both the target and the attached moiety determines overall DUB specificity. Additional mechanisms, such as protein localization and interactions with binding partners, may further contribute to in vivo specificity. In the following sections, we will discuss current insights into DUB specificity.

**Substrate Specificity: Ubiquitin Polymers**

Protein ubiquitination comes in many different flavors that serve distinct functions. Whereas poly-Ub chains linked through the lysine residues of Ub at position 48 (K48) target proteins for proteasomal degradation, the attachment of a single Ub moiety (monoubiquitination) appears to regulate subcellular localization and recruitment of Ub binding proteins. Besides mono- and K48-linked polyubiquitination, other poly-Ub branches using alternative lysines on Ub have been described. The relevance and function of most of these different types of described for a number of mammalian proteins, including RIP, NEMO and TRAFs. These signaling molecules are involved in activation of NF-$\kappa$B signaling, a pathway involved in inflammation, apoptosis, and tumorigenesis. As in the case of mono-Ub, K63 polyubiquitination is required for the activation of downstream molecules, like kinases, or recruitment of other proteins. K63-linked Ub molecules differ remarkably from K48 chains in their three-dimensional structure, which probably accounts for their distinct functions. Furthermore, this suggests that certain DUBs may act on specific Ub branches. Indeed, the yeast DUB Ubp2 prefers K63 over K48-linked Ub chains as a substrate. Conversely, examples of DUBs cleaving K48 but not K63-linked Ub polymers include USP8 and USP14. Another DUB, UCH-L5 can cleave various types of branches but does not display activity toward linear Ub dimers. Other examples further support the notion that DUBs cleave poly-Ub variants with varying efficiency, at least in vivo. One such example is the protein product of the Cylindromatosis tumor-suppressor gene (CYLD), a DUB involved in inhibiting NF-$\kappa$B signaling. CYLD cleaves linear Ub fusions in vitro; yet in vivo it appears to be specific for non-K48-linked Ub chains. The basis for the specificity observed in vivo remains unclear, but phylogenetic analysis indicates that the catalytic domain of CYLD is relatively divergent from other DUBs, as indicated by its unique protease family identifier, C67 (Table S1 and Figure 2). This information leads us

<table>
<thead>
<tr>
<th>Name</th>
<th>Substrate(s)</th>
<th>Process</th>
<th>Remarks</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCH-L1</td>
<td>Unknown</td>
<td>Parkinson's disease</td>
<td>Homodimer has E3 activity; mutant mice display ataxia</td>
<td>Levy et al. (1998); Liu et al. (2002); Salgido et al. (1999)</td>
</tr>
<tr>
<td>BAP1</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Binds to BRCA1</td>
<td>Jensen et al. (1998)</td>
</tr>
<tr>
<td>UCH-L5</td>
<td>Unknown</td>
<td>Ubiquitin editing, TGF-β signaling</td>
<td>Binds to proteasome</td>
<td>Lam et al. (1997); Wicks et al. (2005)</td>
</tr>
<tr>
<td>USPs</td>
<td>CYLD</td>
<td>Traf2/6, NEMO</td>
<td>NF-κB and JNK signaling</td>
<td>Familial tumor suppressor (cyclindromatosis); Brannenkamp et al. (2003); Kovalenko et al. (2003); Trumpp et al. (2003); Reiley et al. (2004)</td>
</tr>
<tr>
<td>USP1</td>
<td>FNCD2</td>
<td>DNA repair</td>
<td></td>
<td>Nijman et al. (2005)</td>
</tr>
<tr>
<td>USP2</td>
<td>Fatty acid synthase</td>
<td>Androgen signaling</td>
<td>Androgen-regulated</td>
<td>Graeter et al. (2004); Onishi et al. (2003)</td>
</tr>
<tr>
<td>USP4</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Transforming activity</td>
<td>Gupta et al. (1994)</td>
</tr>
<tr>
<td>USP5</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Binds K29 chains, binds ISG15 and Ub</td>
<td>Horelia et al. (2004); Russell and Wilkinson (2004)</td>
</tr>
<tr>
<td>USP6</td>
<td>Unknown</td>
<td>Patau's orogeny, actin remodeling</td>
<td>Transforming activity, rearrangements and fusions found in cancer</td>
<td>Masuda-Roberts et al. (2003); Oliveira et al. (2005); Puhl et al. (2003)</td>
</tr>
<tr>
<td>USP7</td>
<td>HDM2, p53, H2B</td>
<td>p53 signaling, Polycomb silencing</td>
<td></td>
<td>Cummins et al. (2004); Everett et al. (1997); Li et al. (2004); van der Knaap et al. (2005)</td>
</tr>
<tr>
<td>USP8</td>
<td>NRDPI</td>
<td>Endocytosis</td>
<td>Oncogenic fusion with p85 PI3K</td>
<td>Jainsson et al. (1998); Kato et al. (2006); Wu et al. (2004)</td>
</tr>
<tr>
<td>USP9X</td>
<td>β-catenin, epsins, AF-6</td>
<td>Wnt-, Notch signaling, endocytosis</td>
<td></td>
<td>Murray et al. (2004); Overstreet et al. (2004)</td>
</tr>
<tr>
<td>USP9Y</td>
<td>Unknown</td>
<td>Spermatogenesis</td>
<td>Mutants associated with azoospermia</td>
<td>Sun et al. (1999)</td>
</tr>
<tr>
<td>USP11</td>
<td>BRCA2</td>
<td>DNA repair</td>
<td>Interacts with RanBP1</td>
<td>Ideguchi et al. (2002); Schoenfeld et al. (2004)</td>
</tr>
<tr>
<td>USP14</td>
<td>Unknown</td>
<td>Synapse function</td>
<td>Mutant mice develop axonal</td>
<td>Bokovits et al. (2001); Wilson et al. (2002)</td>
</tr>
<tr>
<td>USP15</td>
<td>RBX1</td>
<td>COP9 signalosome</td>
<td></td>
<td>Herfeld et al. (2005)</td>
</tr>
<tr>
<td>USP16</td>
<td>H2A?</td>
<td>Chromosome condensation</td>
<td></td>
<td>Mimmack et al. (2001)</td>
</tr>
<tr>
<td>USP18</td>
<td>Unknown</td>
<td>Jak-STAT signaling, immunity, brain function</td>
<td>ISG15-specific mRNA is induced by JNK and IFN</td>
<td>Malakhova et al. (2003); Ritchie et al. (2004); Ritchie et al. (2002)</td>
</tr>
<tr>
<td>USP20</td>
<td>DIO2?</td>
<td>Thyroid hormone metabolism, hypoxia signaling</td>
<td>Interacts with pVHL</td>
<td>Curtis-Morelli et al. (2003); Li et al. (2002b)</td>
</tr>
<tr>
<td>USP21</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Cleaves Ub and NEDD8 but not SMO</td>
<td>Gong et al. (2000)</td>
</tr>
<tr>
<td>USP26</td>
<td>Unknown</td>
<td>Spermatogenesis</td>
<td>mUsp26 is testis specific</td>
<td>Padlock et al. (2005); Strohmeier et al. (2005); Wang et al. (2005)</td>
</tr>
<tr>
<td>USP33</td>
<td>HIF1α-DIO2?</td>
<td>Hypoxia signaling</td>
<td>Interacts with pVHL</td>
<td>Curtis-Morelli et al. (2003); Li et al. (2002b); Li et al. (2005)</td>
</tr>
</tbody>
</table>

**MJDs**

<table>
<thead>
<tr>
<th>Name</th>
<th>Substrate(s)</th>
<th>Process</th>
<th>Remarks</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxin-3</td>
<td>Unknown</td>
<td>MJD disease</td>
<td>Sequence has CAG repeats</td>
<td>Burnett et al. (2003); Schel et al. (2003)</td>
</tr>
</tbody>
</table>

**OTUs**

<table>
<thead>
<tr>
<th>Name</th>
<th>Substrate(s)</th>
<th>Process</th>
<th>Remarks</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A20</td>
<td>RIP</td>
<td>NF-κB signaling</td>
<td>Also E3 ligase</td>
<td>Wertz et al. (2004)</td>
</tr>
<tr>
<td>UCS1</td>
<td>Unknown</td>
<td>Golgi disassembly</td>
<td></td>
<td>Wertz et al. (2004)</td>
</tr>
</tbody>
</table>

**JAMM6**

<table>
<thead>
<tr>
<th>Name</th>
<th>Substrate(s)</th>
<th>Process</th>
<th>Remarks</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNH1</td>
<td>Unknown</td>
<td>Proteasome</td>
<td></td>
<td>Vermala et al. (2002); Yao and Cohen (2002)</td>
</tr>
<tr>
<td>AMSH</td>
<td>FGFR2</td>
<td>Endocytosis</td>
<td></td>
<td>McCallaugh et al. (2004)</td>
</tr>
<tr>
<td>CSN5</td>
<td>Collins</td>
<td>CSN function</td>
<td>Mainly NEDD8 as substrate</td>
<td>Cope et al. (2002); Grossman et al. (2003)</td>
</tr>
<tr>
<td>BRCC36</td>
<td>Unknown</td>
<td>G2/M checkpoint signaling</td>
<td>Enhances BRCA1/BARD1 E3 ligase activity</td>
<td>Dong et al. (2003)</td>
</tr>
</tbody>
</table>

Listed are DUBs that have been linked to specific pathways, processes, and substrates (based on published studies).
to speculate that the architecture of the enzymatic cleft contributes to its specificity. The OTU-type DUB A20 is another potent inhibitor of NF-κB signaling. In vitro, A20 can cleave both K48- and K63-linked Ub polymers with similar efficiency. Yet in vivo, A20 deubiquitinates K63 but not K48 polyubiquitinated RIP (the protein we mentioned previously that is involved in NF-κB activation). However, given that A20 also contains (K48) E3 ligase activity toward RIP, it is not clear if the apparent K63 DUB activity in vivo is due to true substrate preference or simply due to its ability to catalyze the addition of K48-linked Ub polymers.

In at least some cases, domains outside the catalytic domain may contribute to Ub chain specificity. Splice variants of USP2 and a mutant containing only the core catalytic domain of this DUB cleave both linear Ub fusions and K48-linked Ub polymers. However, their relative efficiency varies considerably. The core domain prefers linear fusions, but full-length USP2b was most efficient in cleaving K48-linked Ub. Although USP2 does not contain an additional, known Ub-interaction motif, it is conceivable that sequences outside the catalytic domain contribute to selection and positioning of specific Ub chains. In fact, Ub-interaction motifs found in some E2 ligases have been implicated in determining linkage specificity. Indeed, it was recently described that a previously unnoticed Ub-interacting Zinc finger domain in USP15 is needed for disassembling branched Ub polymers but not for cleavage of a linear Ub-GFP fusion. Another recent study showed that addition of a UBA domain that recognizes K48-Ub chains to USP5 skewed its substrate preference toward this type of Ub polymer. Similarly, other Ub binding domains frequently encountered in DUBs, like UIM and ZnF-UBP (also called PAZ) may also

Figure 4. DUB Specificity and Regulation

(A) DUB/E3 interactions. DUBs and E3 are often found in a complex together. These interactions, which occur between USP7 and HDM2, for example, serve to reverse E3-mediated autoubiquitination (left panel) or allow the E3 to regulate the target and its DUB simultaneously as in the case of USP20 and pVHL (middle panel). Alternatively, DUB/E3 interactions confer specificity to the DUB, as in the case of Ubp2 and Rsp5 (right panel). E3 Ub-ligase and Ub-protease activity is indicated with black arrows and red arrows, respectively. Ub conjugated to E2 is not shown for clarity. (B) DUB activity is regulated at various levels, including transcription (left panel), degradation, and binding to stimulatory or inhibitory cofactors (right panel). The exact mechanism whereby these cofactors regulate DUB activity is unknown but may occur at multiple levels (for example, phosphorylation, subcellular localization), stimulating conformational changes or conferring specificity.
contribute to Ub chain selection (Figure 3).

Substrate Specificity: Ubiquitin and Ubiquitin-like-Molecules
The vast majority of putative DUBs tested so far display Ub protease activity in vitro (Table S1). Nonetheless, some predicted DUBs may be active toward Ubl moieties. Therefore, when considering DUB specificity, we wish to extend our discussion to both Ub and Ubl moieties. USP21 and UCH-L3 cleave both Ub and the Ubl molecule NEDD8. USP18 has been proposed to specifically cleave another Ubl, ISG15. Furthermore, although some circumstantial evidence indicates that CSN5 may contain Ub protease activity, its main proteolytic target is thought to be the Ubl NEDD8. For other Ubl molecules, distinct proteases have been identified. For instance, newly synthesized Ubl Atg8 is processed by a distinct protease (Apg4) of which five family members are found in the genome. Likewise, protease activity toward SUMO (small ubiquitin-like modifier) has thus far been restricted to the SENP family of cysteine proteases, of which seven genes are present in the human genome. However, within this family proteolytic activity is not limited to SUMO; SENP8 (also known as DEN1) is a NEDD8-specific protease.

Our current knowledge of the motifs or residues in these proteases that are responsible for distinguishing Ub from Ubl moieties is limited. Clearly, more in vitro and in vivo analysis of DUBs and Ubl proteases, including structural information, is required.

Target Specificity and DUB/E3 Interactions
The recognition of targets by DUBs may be directed by sequences and motifs outside the conserved catalytic core. For instance, one of the Cap-Gly domains of CYLD mediates its interaction with NEMO, a potential CYLD substrate (Figure 3). However, like most enzyme/substrate interactions, DUB/target interactions are expected to be weak and transient in nature, making the identification of in vivo targets frustrating. A more stable complex between a DUB and its target may occur in the case when proteins are inappropriately K48 polyubiquitinated (Figure 4A, left panel). These proteins need to be continuously deubiquitinated to protect them from unwanted degradation. Autoubiquitination by ring fingertype E3 ligases is a frequently observed phenomenon resulting from nonspecific ubiquitination of proximal lysines (careless gunplay). For instance, the E3 ligase NRDP1 stimulates its own turnover as well as a number of cellular targets. The DUB USP8 associates with NRDP1 resulting in its deubiquitination and stabilization, suggesting that interaction with a DUB may simply serve to antagonize this self-inflicted degradation. Similarly, the interaction of USP7 with HDI2 and USP15 with Rbx1 results in the stabilization of these E3 ligases. Interestingly, USP7 was also found to stabilize the herpes virus E3 ligase ICP0, indicating that viruses can hijack cellular DUBs to stabilize viral proteins. The importance for controlling Ub dynamics in the herpes virus life cycle is further underscored by the recent finding that the herpesviridae contains a distinct class of cysteine protease DUBs without known mammalian homologs.

Not all DUB/E3 interactions strictly serve to regulate E3 ligase stability. The E3 tumor suppressor protein pVHL regulates the stability of HIF1 transcription factors that are important regulators of angiogenesis. USP33 interacts with this E3 and appears to regulate HIF1 stability by deubiquitination. This suggests that, in this case, interaction of the DUB with the E3 allows the E3 to differentially regulate the primary proteasomal target (HIF1) as well as its deconjugating enzyme (Figure 4A, middle panel). Kee and colleagues recently suggested a third type of DUB/E3 interaction. They postulated that in some cases the DUB may hitch along with the E3 ligase. They showed that the target specificity of the yeast DUB Ubp2 is strictly dependent on the E3 ligase Rsp5, which is responsible for recognition of the substrate (Figure 4A, right panel). A remarkable variation on this theme is the previously mentioned protein A20. Here, E3 ligase and DUB activity reside in the same polypeptide.

DUB Function
Gene deletion studies in yeast have indicated that none of the USPs are required for cell growth or viability. Nonetheless, USPs and other DUBs in lower and higher eukaryotes including mammals have been implicated in regulating various critical cellular processes in a nonredundant manner. Human DUBs (or their murine homologs) of particular interest that have been linked to defined cellular processes or substrates are listed in Table 1.

The functions of DUBs at the proteasome lid, in endocytosis and regulation of
DUBs and Proteosome Function

Proteins that must undergo fast and dramatic changes in abundance are often regulated by proteolysis. These proteins are targeted to the proteasome by K48-linked polyubiquitination, where they are degraded. The 26S proteasome consists of two 19S regulatory particles and a 20S cylinder-shaped multiprotein complex possessing the proteolytic activity. The 19S subunit restricts access to the interior of the 13 Å cylinder of the proteasome, which is where the catalytic residues for proteolysis are located. Deubiquitination of proteins arriving at the proteasome allows recycling of Ub and is required for protein degradation. In fact, deubiquitination, protein unfolding, translocation into the proteasome and degradation are intimately linked processes. A number of DUBs from various subclasses have been found in complex with the 19S proteasome regulatory component, including the JAMM protease POH1 (Rpn11 in yeast), UCH-L5, and USP14 (Ubp6 in yeast). Interestingly, residents of a paralogous multiprotein structure known as the COP9 signalosome are the JAMM protein CSN5 and USP15. Like the proteasome, the COP9 signalosome has been implicated in a diverse array of biological processes. At least some of these functions can be explained by its ability to inhibit the activity of the cullin family of ubiquitin E3 ligases63,64.

The main DUB activity at the proteasome appears to be generated by POH1, since deletion of the gene that encodes this enzyme results in defective proteasomal degradation and is lethal in yeast. The functions of the other DUBs may be partially redundant with POH1, only playing a role in the deubiquitination of specific substrates, or in “Ub editing.” The Ub-editing concept was postulated as a mechanism to rescue proteins that have been mistakenly ubiquitinated (as recognized by having short Ub chains) from destruction. The suggested Ub-editing mechanism would remove Ub polymers, starting at the distal end, independently of the substrate moiety. Although UCH-L5 indeed cleaves Ub chains from the distal end, compelling evidence for an Ub-editing function for UCH-L5 has not yet been provided, and no ortholog of UCH-L5 has been found in Saccharomyces cerevisiae.

DUBs and Chromatin Structure

An increasing body of evidence implicates dynamic histone ubiquitination in the regulation of transcription and silencing, and even double-strand-break formation during meiosis. Although most histone proteins can be ubiquitinated, the dynamics of H2B monoubiquitination are best understood. In yeast, deubiquitination of H2B by the DUB Ubp10 is required for telomeric silencing. In contrast to Ubp10, deubiquitination of H2B by Ubp8 correlates with transcriptional activation. At least at some sites of active transcription, Ub-H2B levels are high during activation and subsequently decrease in an Ubp8-dependent manner. Importantly, both the ubiquitination and deubiquitination of H2B are necessary for optimal transcription, indicating a requirement for dynamic H2B modification by Ub.

Similar to yeast Ubp10, Drosophila USP7 interacts preferentially with silenced genomic regions, including telomeric domains, where it has been suggested to deubiquitinate H2B and thereby contribute to Polycomb-mediated silencing. In mammals, USP7 associates with HDM2, an E3 ligase critical for regulating p53 turnover, and thereby inhibits degradation of both HDM2 and p53. Indeed, different levels of USP7 can have opposite outcomes with respect to p53 stability. Intermediate inhibition of USP7 results in increased p53 degradation, whereas complete inhibition of USP7 enhances p53 stability. Interestingly, a recent report has suggested that HDM2 can mediate H2B ubiquitination. Together, these data suggest an attractive model in which HDM2/p53/USP7 complexes mediate transcriptional repression by regulating H2B ubiquitination.

DUBs and Endocytosis

Monoubiquitination and, at least in yeast, the attachment of a K63-linked Ub dimer, play an important role in endocytosis of receptors and sorting of proteins. After binding to ligands, receptor tyrosine kinases (RTKs) and adaptor proteins are monoubiquitinated at multiple sites, which triggers their internalization. The RTKs are subse-
quenty either recycled or transported to lysosomes for destruction. The E3 ligase for many RTKs is the proto-oncogene Cbl, which can also induce proteasome-dependent degradation by stimulating K48 polyubiquitination. Repeated addition of Ub or reduced deubiquitination may be the trigger for targeting to the lysosomal compartment, though exactly how ubiquitination determines this decision is unclear. Ub-interacting proteins like Hrs subsequently bind to the monoubiquitinated receptor and recruit protein complexes involved in budding of the endocytic vesicle.

DUBs are implicated in the endocytic pathway at multiple levels and also play important roles in other types of intracellular traffic. In yeast, the DUB Doa4 acts to recycle Ub at the late endosome to rescue Ub from destruction. Inactivation of Doa4 interferes with many Ub-related processes since it results in depletion of free Ub and many of the defects observed on Doa4 mutant cells are restored upon expression of additional Ub. The closest human relative of Doa4 is USP8, which binds the Hrs binding partner (Hbp) and inhibits EGF receptor (EGFR) endocytosis, suggesting that USP8 may act to regulate endocytic traffic. Remarkably, a second Hrs interacting protein is AMSH, a JAMM domain DUB. Inhibition of AMSH results in the accumulation of endosomal Ub and promotes EGFR endocytosis thereby accelerating EGFR downregulation.

Yet more DUBs are implicated in controlling endocytosis. In Drosophila, Fat facets (Faf; the homolog of human USP9X), deubiquitinates Liquid facets (Lqf), resulting in enhanced Lqf activity. Lqf and Faf play a role in Drosophila eye development by enhancing the internalization of a receptor implicated in cell patterning, called Delta. In humans, Lqf homologs are known as epsins, adaptor molecules involved in the initial steps of endocytosis.

Ubiquitination and deubiquitination appears to be a common theme in vesicle dynamics; monoubiquitination plays a critical role in budding of some viruses. Additionally, VCIP135, a OTU, has been implicated in Golgi assembly after mitosis.

Other Potential Roles for DUBs
Indications for the function of DUBs may come from various sources, including genetic screens in model organisms, interactome data, and domain-and signaling-motif predictions. Genetic screens in model organisms like the worm Caenorhabditis elegans and the fruit fly Drosophila are pointing at new roles for DUBs in various pathways. For instance, screens in C. elegans for modulators of RNA interference or longevity suggest an involvement of a USP and a UCH. In addition, a protein with both a USP and OTU domain (Duo-2) has recently been implicated in synapse function. These studies further solidify the broad involvement of the Ub conjugation/deconjugation system in biological processes and will certainly spark research into the functions of the human DUB orthologs.

Data derived from large-scale human protein-protein interaction experiments has implicated a number of USPs in several signaling cascades such as the TGF-β and NF-κB pathways. For instance, USP45 binds specifically to the phosphorylated TGF-β receptor in a mammalian two-hybrid. Similarly, USP11 and USP9 interact with the NF-κB transcription factors RelB and p100, respectively. Although the significance of these interactions remains to be determined, these findings suggest that DUBs may play a regulatory role in these pathways.

Hints to the function of DUBs may be obtained from additional domains and signal motifs present in the primary amino acid sequence of these enzymes (Figure 3). For instance, USP7 might be involved in specific signaling pathways as it contains a MATH (meprin and TRAF homology) domain. These domains are found in members of the TRAF family of ring finger E3 ligases which mediates signaling via TNF receptors. Similarly, the JAMM2 protein might play a role in transcription and chromatin remodeling, as it contains domains (SWIRM and Myb DNA binding motif) implicated in these processes.

Regulation of DUBs
In contrast to many proteases, such as caspases that are translated as inactive precursors, DUBs are generally produced as active enzymes. Structural analysis has pointed out that the catalytic triad of UCHs and USPs only assume the active confirmation when bound to Ub, thereby preventing spurious protease activity against other substrates. In addition, various studies show that a diverse array of mechanisms regulates DUB activity and additional ones are likely to be discovered.
In the case of at least two JAMM domain proteins (POH1 and CSN5), it appears that incorporation into higher-order protein structures (the 19S proteasome and COP9 signalosome, respectively) is required for peptidase activity\(^{46,63}\). Similarly, accessibility of the enzymatic cleft of USP14 appears to be regulated by activity of the 26S proteasome, its resident complex\(^{68}\).

Bre5, a cofactor for the yeast DUB Ubp3, is largely responsible for its \textit{in vivo} activity toward Sec23, a protein involved in anterograde transport between the endoplasmic reticulum and the Golgi compartment\(^{84}\). Bre5 does not bind directly to Sec23, suggesting that the interaction between Bre5 and Ubp3 regulates Ubp3 activity. Surprisingly, the human homolog of Bre5, G3BP1, inhibits the activity of USP10, at least \textit{in vitro}, indicating that cofactors can either restrict or enhance protease activity (Figure 4B, right panel)\(^{85}\). A USP7 cofactor called GMPS that strongly augments USP7 activity was recently identified in \textit{Drosophila}\(^{69}\). In addition, USP7 is regulated during apoptosis by cleavage by caspases. This cleavage presumably inactivates USP7\(^{70}\).

DUBs have frequently been found to be degraded by the proteasome, indicating that their abundance is an important regulatory mechanism (Figure 4B, right panel). Moreover, some DUBs have been reported to be transcriptionally regulated (Figure 4B, left panel), sometimes in a cell-cycle-regulated manner (for example, USP1) or as part of a negative feedback loop (such as CYLD)\(^{87,88}\).

In another case, inhibitory phosphorylation of CYLD after TNF-\(\alpha\) stimulation is required for the accumulation of one of its proposed substrates, K63-ubiquitinated TRAF2. Interestingly, this event does not appear to modulate the affinity of CYLD for TRAF2, suggesting that phosphorylation may directly regulate CYLD activity by an unknown mechanism\(^{89}\).

**Concluding Remarks**

A large number of studies over the last decade have uncovered an unanticipated diversity of protein regulation by Ub and Ubl molecules. Nature has utilized the versatility of Ub in almost any conceivable way. Strikingly, the ubiquitin conjugation/deconjugation system out competes the protein phosphorylation system in terms of diversity and complexity. Although the reversal of ubiquitination by DUBs has been firmly established as a critical regulatory mechanism, we are only beginning to uncover the different mechanisms that control the activity of these enzymes.

Remarkably, the Ub E3 ligases greatly outnumber the DUBs encoded in the human genome. This is in contrast to tyrosine kinases and phosphatases, which are roughly equal in number. One possible explanation is that we have not yet identified all DUBs or their associated cofactors that may determine specificity. Indeed, Serine/Threonine kinases outnumber Serine/Threonine phosphatases, but a large variety of cofactors provide additional specificity to these phosphatases. It is also possible that many DUBs have poor substrate specificity and regulate on average up to 10 times more substrates than the average E3 ligase. However, most DUBs studied thus far appear to regulate a small number of targets. Another more likely explanation for the excess of E3 ligases could be that only a fraction of the targets that are ubiquitinated are regulated by specific deubiquitination. For destruction mediated by K48 Ub polymers, we would predict that many proteins are not deubiquitinated prior to arrival at the proteasome. Unless of course when you have made a mistake, why recycle a protein that you have decided to throw away? Possibly, only proteins that require extremely tight regulation, such as p53, require additional regulation by deubiquitination. Indeed, other types of Ub-based modifications, like K63-linked polymers or monoubiquitin require DUBs to “reset” the protein to its unmodified state and are thus more likely to be critically regulated by DUBs. Undoubtedly, future studies aided by detailed genomic annotation, structural information, and other new tools and methods to characterize this intriguing protein family will result in the demystification of these proteases.

**Supplemental Data**

Supplemental Data include one figure and one table and can be found with this article online at http://www.cell.com/cgi/content/full/123/5/773/DC1/.

**Acknowledgments**

We would like to thank Dr. Tony Huang, Dr. Huib Ovaa, and Dr. Helen Pickers-
gill for critical reading and helpful discussions. This work was supported by grants from The Netherlands Organization for Scientific Research (NWO) to T.K.S. and R.B.
References


Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases

Mark P.A. Luna-Vargas, Evangelos Christodoulou¹, Andrea Alfieri², Willem J. van Dijk, Magda Stadnik, Richard G. Hibbert, Danny D. Sahtoe, Marcello Clerici, Valeria De Marco¹, Dene Littler, Patrick H.N. Celie³, Titia K. Sixma⁴, Anastassis Perrakis⁵

*Corresponding author: a.perrakis@nki.nl (A. Perrakis).
¹Present address: Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.
²Present address: Department of Experimental Oncology, European Institute of Oncology, via Adamello 16, I-20139 Milan, Italy.
³Correspondence and requests for all vectors presented in this manuscript should be addressed to P.H.N.C.
⁴Correspondence concerning the USP family and related reagents should be addressed to T.K.S.
Abstract
High-throughput methods to produce a large number of soluble recombinant protein variants are particularly important in the process of determining the three-dimensional structure of proteins and their complexes. Here, we describe a collection of protein expression vectors for ligation-independent cloning, which allow co-expression strategies by implementing different affinity tags and antibiotic resistances. Since the same PCR product can be inserted in all but one of the vectors, this allows efficiency in versatility while screening for optimal expression strategies. We first demonstrate the use of these vectors for protein expression in Escherichia coli, on a set of proteins belonging to the ubiquitin specific protease (USP) Family. We have selected 35 USPs, created 145 different expression constructs into the pETNKIHis-3C-LIC-kan vector, and obtained 38 soluble recombinant proteins for 21 different USPs. Finally, we exemplify the use of our vectors for bacterial co-expression and for expression in insect cells, with USP4 and USP7 respectively. We conclude that our ligation-independent cloning strategy allows for high-throughput screening for the expression of soluble proteins in a variety of vectors in E. coli and in insect cells. In addition, the same vectors can be used for co-expression studies, at least for simple binary complexes. Application in the family of ubiquitin specific proteases led to a number of soluble USPs that are used for functional and crystallization studies.

Introduction
The sequences of viral, bacterial and eukaryotic genomes have been unravelled over the last decade, providing a large protein database in demand for structure–function analysis. To meet this challenge, an international effort to understand protein function was initiated, and pioneered by efforts such as the US-based “Protein Structure Initiative” (PSI), the Japan-based RIKEN Structural Genomics Initiative, and private–public open access consortia pioneered by the Structural Genomics Center (SGC) and many others. In Europe, the SPINE project introduced high throughput proteomics in a wide collaborative platform, and SPINE-2-complexes aimed to extend many of these methodologies to the study of macromolecular complexes. A direct consequence of such efforts was the rapid development of new high-throughput technologies for molecular cloning, expression and purification of recombinant proteins, that are crucial in the effort of determining the three-dimensional protein structures on a genomic scale\textsuperscript{1,2}.

Any structure determination project is dependent on the ability to obtain sufficient amounts of soluble recombinant protein. A difference of three amino acids between constructs can already improve either the protein solubility or significantly alter its propensity to crystallize\textsuperscript{3,4}. It is therefore common practice to test many protein constructs, typically combinations of N- and C-terminal truncations and assess them for the amount of protein expressed and solubility. The use of high-throughput methods enables screening of multiple constructs in parallel, to quickly determine the construct, which produces soluble recombinant protein and is more likely to crystallize.

One technique for molecular cloning that was described two decades ago\textsuperscript{5} but has been popularized in the Structural Genomics field, is Ligation Independent Cloning (LIC). LIC is a procedure for the directional cloning of PCR products that is independent of restriction endonucleases, DNA ligase or alkaline phosphatases. The LIC method takes advantage of the $3^\prime\rightarrow 5^\prime$ exonuclease activity of T4 DNA Polymerase to create very specific single-stranded overhangs in both the vector and in PCR products of the protein of interest. When vector and insert are transformed into \textit{Escherichia coli} cells, covalent bond formation at the vector-insert junctions occurs within the cell to yield circular plasmid and the resulting construct is ready for expression either in bacteria or in insect cells. Here, we first describe the procedures to design and produce a series of LIC applicable vectors for recombinant protein expression. All but one of the vectors accommodate the same PCR product: this allows to create in parallel a simple hexahistidine tag construct, or a GST tag construct to facilitate solubility, or a Strep-tag II, or a His-tag fused to the Trigger Factor in a cold inducible system that has shown very promising results\textsuperscript{6}. At the same time the target vectors have different antibiotic resistance combinations and/or origins of replica-
tion (Fig. 1) making them possible for protein co-expression experiments. The same PCR product can even be used for insertion in vector that can be used for protein production in insect cells. The only vector requiring a different PCR reaction is the his-SUMO2 vector: this peculiarity is necessary to take advantage of the SENP-2 protease property to leave no residual amino-acids after tag cleavage. To enable the high throughput design of PCR primers for these LIC reactions, the web server ‘Protein Crystallisation Construct Designer’ (ProteinCCD; http://xtal.nki.nl/ccd) has been designed.

To exemplify the use of our vector suite, we show our work in the study of the largest subfamily of Deubiquitinating Enzymes (DUBs), the Ubiquitin Specific Protease (USP) family.

DUBs are proteases that are capable of removing ubiquitin (Ub) from Ub conjugated proteins by specifically hydrolyzing ester, thiol ester and amide bonds to the carboxyl group of Glycine-76 of Ubiquitin. The large USP family comprises approximately sixty proteins, which are papain-like proteases that vary greatly in size. These USPs contain three conserved regions surrounding the cysteine, histidine and aspartate/asparagine residues that form the catalytic triad. Only a small number of substrates have been identified and little is known about the substrate specificity and selectivity of the USP family. The large size of the USP family and its associated specificity opens up perspectives for its use as drug target.

We present the expression testing for a large collection of USPs, using one of our vectors. Then, we show the use of a selection of our different bacterial expression vectors for single protein expression and co-expression studies using the minimal catalytic domain of one of the members of the USP family, USP4, which is formed by two domains, USP4-D1 and -D2. Finally we illustrate the use of an insect cell expression LIC vector for the production of another member of the USP family, USP7.

**Material and methods**

**The design and construction of the pETNKI-LIC vector suite**

An overview of all the vectors used in this study is shown in Fig. 1.

The pETNKI-His-3C-LIC-kan vector is engineered to express the target protein immediately downstream of an N-terminal hexahistidine tag, followed by the human rhinovirus 3C protease (3C protease) cleavage site, enabling removal of all vector-encoded fusion sequences. This vector was based on the popular pET-28a expression vector (Novagen): the sequence between the NcoI and XhoI restriction sites of the original plasmid was replaced with the 5’CCATGGCACATCACCAC-CACCATCACTCCGGGCTTTTGAGGTGCTTTTCAGGGACCCGGGTAC-CAAGAAACAAAACATGTCAGAACAGAGATCTGATTGATACCCGCCGTTCTCCTCGAG 3’ sequence. This sequence contains two KpnI restriction sites (bold) and the LIC sequences that are used to create the overhanging ends for the annealing reaction (italics). The 5’ sequence upstream the first KpnI site encodes the MAHHHHHHSAALEVLFQ↓GPG sequence (underlined; ↓ indicates the cleavage site for 3C protease).

To create the pCDFNKI-StrepII-3C-LIC-strep vector, pET52b (Novagen) was cut with XbaI and KpnI to obtain a 109 bp fragment containing the StrepII-tag and 3C protease recognition site. This fragment was ligated into pETNKI-His-3C-LIC-kan that was cut with XbaI and KpnI, to yield pETNKI-StrepII-3C-LIC-kan. The resulting vector was cut with NcoI and XhoI to obtain a 84 bp fragment containing the LIC cassette. This fragment was cloned into the pCDF-Ib expression vector (Novagen) cut with the same enzymes.

To create the pGEXNKI-GST-3C-LIC-amp vector, a PCR reaction was performed using the pETNKI-StrepII-3C-LIC-kan vector as template to obtain a fragment containing the LIC cassette that could be cloned into the pGEX-2T expression vector (GE Healthcare). The forward primer (50-CATAGGGATCCTCCGCGGCTTTGAGGTG-30) contained a BamHI restriction site and the reverse primer (50-ATCCGAATTCTAAGTTATTGCTCAGCGG-30) contained an EcoRI site. The 250 bp PCR fragment was digested with BamHI and EcoRI and ligated into the pGEX-2T vector cut with the same enzymes.

To construct the pColdNKI-His-TF-3C-LIC-amp vector containing a hexa-histidine-tagged Trigger Factor (TF) as a solubility tag, pETNKI-His-3C-LIC-kan was digested with SacII and XhoI to obtain a 90 bp SacII-XhoI fragment containing
the LIC cassette. This fragment was ligated into the pCOLD™ TF vector (Takara Bio Inc, Otsu, Japan), that was cleaved with the same enzymes.

To create the pET-NKI-His-SUMO2-LIC-kan vector, a Quikchange reaction (Stratagene, CA) was performed with forward primer CCAGCAGCAGACGGGAGGGTACC-GGGCTCCGCCGCCAAGCTTGCGGCCG-CACTCG and the corresponding reverse/complement oligonucleotide according to manufacturers instructions, using the SUMOPRO pSUMO2 vector (LifeSensors, PA), which contains the SMT3A gene in frame with a his-tag, as a template to introduce a LIC cassette. This is the only vector from this collection where different overhangs are required for the PCR fragment.

To create the pFastBac-NKI-His-3C-LIC-amp vector, the pETNKI-His-3C-LIC-kan cassette was PCR amplified and inserted in the baculovirus expression vector pFastBac-HTb between the Rsfl and HindIII restriction sites using the following primers: forward 5’-GAGA GTCCGTCGACAGGATCC-GGCTCCGCCGCAAGCTTGCGGCCG-CACTCG and the corresponding reverse/complement oligonucleotide according to manufacturers instructions, using the SUMOPRO pSUMO2 vector (LifeSensors, PA), which contains the SMT3A gene in frame with a his-tag, as a template to introduce a LIC cassette. This is the only vector from this collection where different overhangs are required for the PCR fragment.

Preparation of Linear Plasmid DNA for LIC

Any of these plasmids is transformed into an E. coli host strain like DH5a, amplified and retrieved via a mini or maxi plasmid prep kit (Qiagen). The vector DNA (10 µg) is then digested with KpnI for 2–3 h at 37°C in a 100 µL reaction volume and purified with a QIAquick spin column (Qiagen) according to the manufacturer’s protocol and eluted in 100µL TE buffer to achieve DNA concentration of ~100 ng/µL. The linearized vector DNA is then treated with T4 DNA Polymerase to create the single-strand overhangs as follows: In a sterile 1.5-ml micro-centrifuge tube 10 µl of linearized vector, which contains the SMT3A gene in frame with a his-tag, as a template to introduce a LIC cassette. This is the only vector from this collection where different overhangs are required for the PCR fragment.

To create the pFastBac-NKI-His-3C-LIC-amp vector the pETNKI-His-3C-LIC-kan cassette was PCR amplified and inserted in the baculovirus expression vector pFastBac-HTb between the Rsfl and HindIII restriction sites using the following primers: forward 5’-GAGA CTCCG TCGACAGGATCC-GGCTCCGCCGCAAGCTTGCGGCCG- CACTCG and reverse 5’-ACTTAAGCTTCTCAGAGGAGAAGCCC-GGGTA-3’. The start ATG codon was maintained in the same position as in pFastBac-HTb.
Production of gene products for LIC cloning

To amplify the target gene construct by PCR to create the LIC insert, we typically use the high-fidelity polymerase Pfu from Stratagene. A main requirement here is that the polymerase used needs to be one with minimal activity for the addition of non-templated 3’ nucleotides. For all vectors (except pETNKI-His-SUMO2-LIC-kan) the 5’-end of the primers must contain the CAGGGACCCGGT sequence upstream of the forward PCR primer and CGAGGAGAAGCCCGTTA sequence upstream of the reverse primer (which includes a TAA stop codon). For the pETNKI-His-SUMO2-LIC-kan vector the 5’-end of the forward PCR primer must contain the CCAGCAGCAGACGGGAGGT sequence upstream of the primer followed by the sequence of the gene of interest; the GGCGGCGGAGCCCGTTA sequence is needed upstream at the 5’-end of the reverse primer. Importantly, the SUMO2 tag can be removed from the target protein after treatment with SENP2 protease, leaving no extra amino acid residues at the N-terminus of the sequence of the target protein.

To prepare DNA fragments for the T4 DNA Polymerase treatment, the dNTPs from the PCR product are removed with a spin column kit (QIAquick PCR purification kit) and the purified PCR products are eluted in TE buffer (10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0). In a sterile 1.5-ml microcentrifuge tube (or microwell plate to enable high throughput production) 0.2 pmol of purified PCR DNA (measured in a Nanodrop1000 by Thermoscientific), 2 µl 10X T4 DNA Polymerase Buffer (NEB), 2 µl 25 mM dATP and 1 µl T4 DNA Polymerase (NEB) are added to a final volume of 20µL. The reaction is started by adding the enzyme and incubated at RT for 30 min. The T4 DNA polymerase is inactivated by incubating at 75°C for 20 min.

Annealing the vector and the LIC insert

For each construct, 1 µl vector (50 ng/µl) and 2 µl insert (0.02 pmol) prepared as above are assembled in a sterile reaction plate. The reactions are incubated at RT for 5 min, and stopped by adding 1 µl of 25 mM EDTA. Typically, half of the annealing reaction (2 µl) is transformed into Novablu competent cells that are plated onto LB agar plates containing the appropriate antibiotic. After overnight incubation at 37°C, two single colonies for each construct are picked and used for plasmid amplification, isolation and sequencing, using standard protocols. Alternatively, colonies containing the insert can be identified using colony PCR and sequence confirmation can be reserved for only constructs that produce soluble protein. It should also be noted that annealing reactions can be transformed directly into the expression strain; or each half of the reaction can go to an expression strain and a DNA amplification strain respectively.

Testing different pETNKI-LIC vectors in medium-throughput co-expression

To identify constructs likely to yield sufficient amounts of soluble protein for scale-up and protein purification, we first performed a small-

Figure 2. A schematic overview of the LIC cloning procedure. 1) Cleavage of the different expression vectors with Kpn1 (except for the pETNKI-His-SUMO2-LIC). 2) Treatment of the cleaved vector with T4 DNA polymerase in presence of dTTP. 3) Generation of PCR fragment of gene of interest flanked with specific LIC sequences. 4) Treatment of the LIC PCR fragment with T4 DNA polymerase in presence of dATP. 5) Annealing and ligation of insert into expression vector. The sequence of Kpn1 cleavage site is indicated in pink, the LIC sequence is indicated in blue for the vector and brown for the PCR insert, and the gene of interest is indicated in green.
scale protein expression and solubility screening. 50 µl E.coli Rosetta2(DE3) or BL21(DE3) cells are transformed with 50 ng of plasmid in a Thermowell 96-well PCR microplate (Corning, Inc.) and heat shocked for 40 s at 42°C in a water bath. After incubation on ice for 2 min, 100 µl SO medium is added to each well and the 96-well microplate is sealed with a silicone rubber mat (Costar) and incubated at 37°C for 1 h. The transformation reactions are plated out on 6-well LB agar plates (Falcon multiwell, Beckton Dickinson) containing the appropriate antibiotics. After overnight growth at 37°C, single colonies of each construct are picked and grown in 200 µl auto-induction medium (ZYP5052, made in-house)11 (Studier, 2005) with appropriate antibiotics, in a 2 ml 96-deepwell block (Nunc) and sealed with Airpore tape sheets (Qiagen). The 96-deepwell block is incubated in a shaker at 300 rpm for 4 h at 37°C. When the cells reach an OD600 of about 2–3, typically after 4 h, the temperature is lowered to 15°C for overnight induction. For higher volume production (in our case for all vector comparison tests) 0.5 ml of overnight pre-culture was used to inoculate 50 ml LB medium (LB Broth (Miller) from Molecular Dimensions Limited) with corresponding antibiotics in 250 ml baffled flasks and grown at 37°C until OD600 of 0.8 units is reached. Since USP4 contains two zinc finger motifs, prior to induction we supplement the media with 12.5 µl 1M ZnCl₂ (250 mM), a practice we know to be beneficial based on prior experience. Finally, we added 25 µl 1M IPTG (500 mM) to induce protein production and the cells were grown overnight at 15°C.

For high throughput testing, cells are collected by centrifuging the 96-deepwell block at 4000 rpm for 15 min. All liquid handling steps are done by hand with a multi-channel pipette (Matrix Technologies). The pellet in each well is resuspended with 200 µl lysis buffer (50 mM Tris–HCl pH8.0, 200 mM NaCl, 5 mM imidazole, 5 mM β-mercaptoethanol, 200 U Lysozyme (Novagen)). The 96-deepwell block containing the cell suspension is shaken at 800 rpm for 20 min at room temperature using the thermomixer (Eppendorf). When the cell extract is clear, 10 µl of MagneHis Ni²⁺-beads (Promega), are added to each well. After 5 min the Magnetight HT96 stand (Novagen) is used to pull down the MagneHis beads to remove unbound proteins. The MagneHis beads are washed 3 times with 50 µl wash buffer (50 mM Tris–HCl pH 8.0, 200 mM NaCl, 5 mM imidazole & 2 mM β-mercaptoethanol). Protein elution from the MagneHis-beads is done by adding 20 µl elution buffer (wash buffer containing 500 mM imidazole) to each well and incubating the 96-deepwell block for 10 min at room temperature. The MagneHis-beads are finally pulled down with the magnetic stand and the eluted proteins removed and analyzed on SDS–page gels.

For higher volume testing (in our case for all vector comparison experiments and co-expression tests), cells are harvested by centrifugation and resuspended with 3 ml lysis buffer (25 mM Hepes pH7.5, 150 mM NaCl, 10 mM Imidazole, 5 mM β-mercaptoethanol and protease inhibitor tablet). The cells are broken by subjecting the cell suspension to a 10 s pulse with a pause of 20 s after each pulse for a total of 2 min using the Misonics sonicator S-4000 at 40% maximum setting. The cell lysate is centrifuged at 13,200 rpm for 25 min at 4°C. The resulting supernatant is incubated with 100 µl of corresponding resin (Talon resin (Clontech) for His-tag, Glutathione Sepharose resin (GE Healthcare) for GSTtag and Strep Tactin resin (Novagen) for StrepII-tag) for 15 min at 4°C. The resin is washed with 2 times 5 mL lysis buffer and the protein is eluted from the beads with elution buffer (lysis buffer containing 300 mM Imidazole or 25 mM glutathione or 2.5 mM desthiobiotin). Samples of total cells, supernatant cell lysate and eluate are analysed on an SDS–PAGE gel.

Scale-up of expression and protein purification of USP8
As in the small scale expression trials, the E. coli host Rosetta2(DE3) was used for the large scale protein expression. 5 ml of overnight pre-culture was used to inoculate 500 ml autoinduction media in 3L baffled flasks and grown at 37°C until OD600 of 2–3 units was reached. The temperature was lowered to 15°C for overnight induction. Cells were harvested by centrifugation and resuspended in 200 ml lysis buffer (50 mM Tris–HCl pH8.0, 150 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol and 1 mM PMSF). The cells were broken by subjecting the cell suspension to a 10 s pulse with a pause of 30 s after each pulse for a total of 5 min using the Misonics sonicator S-4000 at 80% maximum setting. The broken cell lysis was centrifuged...
at 48,000 g (J-26XP Avanti Centrifuge, Beckman Coulter) for 30 min at 4°C to remove cellular debris and unbroken cells. The resulting supernatant was incubated with washed 2 ml Talon metal affinity resin (Clontech, Inc., Palo Alto, CA) for 20 min at 4°C and the beads were then washed with 200 ml lysis buffer. The beads were eluted with 20 ml lysis buffer containing 400 mM imidazole. The eluted sample was diluted 10–15 times with 50 mM Bis-Tris pH6.5. The diluted sample was applied to a 5 ml Poros S column equilibrated with buffer A (20 mM BisTris pH6.5, 10 mM NaCl and 5 mM β-mercaptoethanol) and bound protein was eluted with buffer A containing 1 M NaCl using a 60% gradient in 20 column volumes. Peak fractions were pooled and concentrated by ultrafiltration using an Amicon Ultra centrifugal unit (Millipore) and applied to a Superdex 75 16/60 gel filtration column (GE Healthcare) equilibrated with buffer containing 25 mM Tris–HCl pH8.0, 150 mM NaCl and 5 mM β-mercaptoethanol. Peak fractions from the gel filtration column were pooled and concentrated in an Amicon Ultra centrifugation unit to a concentration of 10 mg/ml. The concentrated protein was flash frozen in liquid nitrogen and stored at -80°C.

### Small-scale test expression in insect cells

The production and isolation of the recombinant bacmid and the infection of the P0 culture is performed as described in the Invitrogen Bac-to-Bac® Baculovirus Expression System manual. 25 ml of SF-9 insect cells at 106 cells/ml concentration is infected with 1 ml of P0 culture medium and let grow for 72 h.

Cells are harvested by centrifugation and resuspended with 10 ml lysis buffer (20 mM Hepes pH 8.0, 150 mM NaCl and protease inhibitor tablet). The cells are lysed by sonication in an ice bath for 30 s (5 s pulse with a pause of 40 s) using the Misonics sonicator S-4000 (at 50% maximum setting) and the lysate centrifuged at 48,000g (J-26XP Avanti Centrifuge, Beckman Coulter) for 25 min at 4°C. The supernatant is purified on column using 200 µl of Chelating Sepharose Fast Flow resin loaded with Ni²⁺. Samples of lysate, supernatant and elution are analysed on a SDS–PAGE gel.

### Table 1

Summary of the high-throughput cloning and protein expression screen.

<table>
<thead>
<tr>
<th>Target name</th>
<th>Targets designed</th>
<th>Constructs cloned</th>
<th>Soluble proteins (%)</th>
<th>Soluble targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human USPs</td>
<td>35</td>
<td>176</td>
<td>145</td>
<td>37 (26%)</td>
</tr>
</tbody>
</table>

### Table 2

Medium-throughput testing of different NKI-LIC vectors in single and co-expressions of USP4-D1 and USP4-D2 (c: cloned, E: expression, S: soluble).

#### Single expression

<table>
<thead>
<tr>
<th>Constructs</th>
<th>USP4-D1</th>
<th>USP4-D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>pETNKi-His-SUMO2-LIC-kan</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>pCDFNKi-StrepII-3C-LIC-strep</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pCOLDNKi-His-TF-3C-LIC-amp</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>pGEXNKi-GST-3C-LIC-amp</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

#### Co-expression

<table>
<thead>
<tr>
<th>Constructs</th>
<th>USP4-D1 &amp; D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>pETNKi-His-SUMO2-LIC-kan (D1)</td>
<td>✓</td>
</tr>
<tr>
<td>pCDFNKi-StrepII-3C-LIC-strep (D2)</td>
<td>✓</td>
</tr>
<tr>
<td>pCDFNKi-StrepII-3C-LIC-strep (D1)</td>
<td>✓</td>
</tr>
<tr>
<td>pCDFNKi-StrepII-3C-LIC-strep (D2)</td>
<td>✓</td>
</tr>
<tr>
<td>pCOLDNKi-His-TF-3C-LIC-amp (D1)</td>
<td>✓</td>
</tr>
<tr>
<td>pCDFNKi-StrepII-3C-LIC-strep (D2)</td>
<td>✓</td>
</tr>
</tbody>
</table>
Results and discussion

Evaluation of the pET-NKI vectors suite

The pETNKI-His-3C-LIC-kan vector was tested using a set of proteins that we have successfully produced previously in our lab. First, we used our method to clone two different proteins and analyzed 96 colonies from each. The number of clones without an insert was less than 6%, an acceptable rate of false positives to allow high throughput approaches. Expression levels were comparable to other vectors previously used in our lab. The pETNKI-His-3C-LIC-kan vector has thereafter been used successfully to produce well over one thousand expression constructs of different proteins in our lab. In the next section we describe its specific application to the study of the USP family.

Based on the success of the pETNKI-His-3C-LIC-kan system in our lab as well as in other labs, we have created a complementary suite of vectors (Fig. 1) which enable expression of different fusion proteins and which also allow production of protein complexes using co-expression strategies. To evaluate these vectors, we used as a test system the catalytic domain of the deubiquitinating enzyme USP4 (USP4CD) that is made by two domains, D1 and D2, which assemble together to make USP4CD. Neither D1 nor D2 could be obtained as a soluble protein when expressed individually as hexahistidine-tagged fusion proteins. However, upon co-expression of the two domains, a soluble protein complex could be isolated. For our vector evaluation experiment, D1 was fused to different N-terminal affinity and solubility tags (His-SUMO2-, GST- and His-TF- tag), whereas D2 was fused only to a StrepII affinity tag. We expressed each protein individually or in different D1-D2 co-expression combinations (Fig. 3, Table 2). For the comparison of expression experiments we did not apply any other normalization or optimization, other than using identical culture conditions and volumes in all cases. For His-SUMO2- and GST- D1 fusion constructs, only partially soluble protein could be obtained within the individual expression tests, whereas a fair amount of soluble D1 was obtained when fused to the His-TF-tag. For StrepII-tagged D2 protein alone, only a minor portion of the expressed protein appeared in the soluble fraction. However, when D2 was co-expressed together with the D1 constructs, a significant amount of soluble D2 could be co-eluted when D1 was purified via affinity purification (either Talon beads for the his-tag or glutathione Sepharose beads for the GST-tag), demonstrating the presence of a soluble D1-D2 complex. Co-expression improved the solubility of the D1 fragments in all cases. Using the pETNKI-His-SUMO2-LIC-kan and pCDFNKI-StrepII-3C-LICstrep vectors, D1 and D2 were also expressed in similar amounts, representing a stoichiometric ratio of 1:1 as physiologically expected for D1 and D2 being the two parts that constitute USP4CD. These results suggest that the NKI-LIC vector suite can be used to produce soluble proteins and protein complexes when used in co-expression experiments.

The utility of affinity tags has been previously shown to be of great use for purifying recombinant proteins. However, most tags come with their own advantages and disadvantages. For example, the use of immobilized metal-ion affinity chromatography (IMAC) sometimes yields insufficiently pure proteins and does not enhance the protein solubility, while the affinity media are very cheap. In contrast, the StrepII-tag displays increased specificity and re-

Figure 3. SDS–PAGE assessment of co-expression trials for USP4-D1 and USP4-D2 domains in different NKLIC E.coli expression vectors. T: total cells; S: supernatant; E: elution. * indicates the D1 fusion protein and • indicates the D2 fusion protein.
results to more pure protein, but the affinity resin is very expensive, while it does not improve protein solubility\textsuperscript{14}. The GST-, SUMO2- and TF-tags have been developed to aid in protein solubility and folding\textsuperscript{6,15,16}. Furthermore, the SUMO2-tag has been shown to also increase the level of protein expression and solubility. However, also for these tags there are disadvantages: the GST-tagged recombinant protein tends to form dimers, SUMO2 is mostly constrained to \textit{E.coli}, since SUMO proteases that are present in eukaryotes may cleave the fusion protein during expression, and all these tags often result in finally insoluble protein after they are cleaved. Providing a vector suite where the utility of each tag can be tested experimentally easy and with minimal cost (no additional PCR products needed, with the exception of the SUMO2 tag vector) makes the choice of this vector suite particularly appealing. Combining different tags with different origins of replication and antibiotic resistance enables co-expression experiments with the same vectors.

**Application to the USP family**

The cDNA of 35 different USPs were collected and 176 constructs were designed (Table 1 and Supplementary Table 1). Initial constructs design was enabled by studying existing crystal structures of USP catalytic domains (USP7, 1nb8; USP2, 2hd5; USP8, 2gfo; USP14, 2ayo)\textsuperscript{17-20} and also based on sequence alignment\textsuperscript{21}, secondary structure prediction\textsuperscript{22} and protein disorder prediction\textsuperscript{23}. Additional constructs were designed if a link to cancer was found in literature for a specific USP or when a positive but insufficient hit for protein expression was identified. These constructs included the full length protein as well the additional domains found outside the catalytic domain predicted by SMART\textsuperscript{8,24} with variations at the C- and N-terminal ends of the predicted domains. Although PCR problems precluded the amplification of a significant number of constructs (30 out of 176), once PCR products were made, all but one of them have been successfully cloned into the pETNKI-His-3C-LIC-kan vector. The high throughput protein expression and solubility screening of 145 different constructs resulted in 26% of the constructs giving soluble protein for 21 different USPs. Large scale protein production is ongoing to verify whether these constructs can produce soluble protein in amounts suitable for structural studies. In Fig. 4 we exemplify the large scale protein expression and purification of the catalytic domain of USP8.

From all constructs tried, 98 of them (68%) expressed recombinant His-tagged protein for 26 different USPs. Absence of protein expression in the remaining constructs was caused either by failure of transformation reactions or by low cell culture density during protein induction. Of the 98 expressing constructs, 37 constructs produced soluble recombinant protein for 21 different USPs. In some cases only particular constructs were able to give soluble protein. For example, only one construct of USP1 (Supplementary Table 1) expressed soluble protein. This construct lacks the first 29 residues at the N-terminus of the protein, which were predicted to belong to a flexible loop\textsuperscript{22}. The removal of this flexible region seemed to be necessary for the soluble expression of USP1. Also for example, in USP21 the addition and removal of several residues at the N-terminal and C-terminal ends respectively, greatly enhanced expression and solubility compared to the \textit{in silico} (SMART) predicted domain (Supplementary Table 1). USP7 was a particular USP family target that was not expressed in \textit{E.coli} despite significant effort. For convenience we created a vector (pFastBacNKI-His-3C-LIC-amp) to facilitate recombinant virus production for protein expression in insect cells, which could accommodate the exact same PCR product as the \textit{E.coli} vectors we use. We show that this vector can be used straightforwardly for LIC

![Figure 4. Large scale protein expression and purification profile of the catalytic domain of USP8CD on SDS–PAGE gel.](image-url)
cloning and will yield viruses that successfully mediated USP7 production in insect cells (Fig. 5).

Initially, all constructs were designed based on the sequence alignment and comparison with the previously determined X-ray crystal structures of several catalytic domains of USPs. However, it is clear that designing and screening several different constructs for a particular USP should also be based on domain prediction, disorder prediction and secondary structure prediction, which can greatly improve the protein expression and solubility. A tool to enable such studies has been developed partially based on this experience and is available from http://xtal.nki.nl/ccd7.

Potential for automation
All described steps were done with a manual multipipette. However all these steps can be automated with the help of robotics25, in which the different steps such as cell lysis, affinity binding and elution of recombinant protein are interlinked. Moreover, the analysis of protein samples by running conventional SDS–PAGE gels is time-consuming and laborious (e.g. pouring gels, pipetting samples, staining and destaining gels). Using other methods for rapid analysis of protein samples such as 96-well gels or microfluidics, might increase the throughput in the last step of identifying soluble recombinant proteins in the high throughput protein expression and solubility screening26,27.

Conclusions
We presented the procedures to create and use a collection of vectors for protein expression, the ‘NKI-LIC suite’ which can all (but one) accommodate the same PCR product. Combining different affinity and solubility tags with different antibiotic resistance markers and origins of replication, this enables straightforward and convenient expression strategies. The only vector not compatible with the exact same PCR product contains the small and soluble SUMO2 protein fused to a His-tag for purification purposes. This vector has the advantage that upon cleavage with the SENP protease, no residual residues are left at the N-terminus of the protein of interest. Finally, we show one vector that is suitable to initiate insect cell expression. We showcase the applicability of this strategy to a large collection of targets within the Ubiquitin Specific Protease Family, one of the SPINE-2-complexes priority areas, which enabled the production of many soluble recombinant proteins, suitable for functional and structural studies.

Potential for automation
All described steps were done with a manual multipipette. However all these steps can be automated with the help of robotics25, in which the different steps such as cell lysis, affinity binding and elution of recombinant protein are interlinked. Moreover, the analysis of protein samples by running conventional SDS–PAGE gels is time-consuming and laborious (e.g. pouring gels, pipetting samples, staining and destaining gels). Using other methods for rapid analysis of protein samples such as 96-well gels or microfluidics, might increase the throughput in the last step of identifying soluble recombinant proteins in the high throughput protein expression and solubility screening26,27.

Acknowledgments
Author contributions: MLV identified USP domains, prepared the target list, produced expression clones, performed small scale protein expression and purification, and wrote the manuscript together with AP. EC designed, made and tested the pETNKI-His-3CLIC-kan vector, and initiated the use of the LIC system at the NKI. AA worked on additional USP domains. PvD performed large-scale protein production and purification. MS did the comparative (co-)expression testing of USP4 constructs in all vectors. RGH designed and produced the pETNKI-His-SUMO2-LIC-kan vector. DS and MC designed and tested the pFastBac-NKI-His-3C-LIC-amp vector; VDM and DL participated in testing and establishing the pETNKI-His-3C-LIC-kan for routine use at the NKI. PC designed and produced the pGEXNKI-GST-3C-LIC-amp, pCOLDNKI-His-TF-3C-LICamp and pCDFNKI-StrepII-3C-LIC-strep vectors with various antibiotic resistance markers. TKS initiated and supervised the Ubiquitin Specific Protease project. AP initiated and supervised the high throughput technologies project and wrote the paper with MLV. Funding for this project has been made available through the EC integrated networks SPINE-2 and Rubicon, and through a grant from the KWF (Netherlands).

Appendix A. Supplementary data
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2011.03.017.
CHAPTER 3

References


The differential modulation of USP activity by internal regulatory domains, interactors and seven Ubiquitin chain types

Alex C. Faesen¹,#, Mark P.A. Luna-Vargas¹,#, Paul P. Geurink², Marcello Clerici¹, Remco Merkx², Willem J. van Dijk¹, Dharjath S. Hameed², Farid El Oualid², Huib Ovaa², and Titia K. Sixma¹,*

Submitted

¹Division of Biochemistry and Center for Biomedical Genetics, ²Division of Cell Biology, the Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands.
*Correspondence: t.sixma@nki.nl (T.K.S.)

# These authors contributed equally to this study.
Abstract
Ubiquitin-specific proteases (USPs) are papain-like isopeptidases with variable inter- and intra-molecular regulatory domains. To understand the effect of these domains on USP activity, we have analyzed enzyme kinetics of a set of twelve USPs in presence and absence of modulators using synthetic reagents. We synthesized all seven wild-type lysine-linked di-ubiquitins and provide the first comprehensive analysis comparing ubiquitin (Ub) chain preference. Our data reveal large variations in both the catalytic turnover and Ub binding between USPs and modest preferences for di-Ub topoisomers. Interestingly, our data show that the preference of USP7 for di-Ub topoisomers can be attributed to the binding affinity ($K_m$) for the substrate, while the intermolecular activators UAF1 and GMPS mainly increase the catalytic turnover ($k_{cat}$). Together, this comprehensive kinetic analysis highlights the variability within the USP family.

Introduction
Since the 1980s, the post-translational modification of proteins by Ub has been the focus of many studies due to their important roles in many cellular processes. However, the processing and removal of Ub and thus reversal of the modification of target proteins is equally important and is carried out by De-ubiquitinating enzymes (DUBs).

The human genome encodes nearly 100 putative DUBs, belonging to at least five subfamilies of isopeptidases. The Ubiquitin-Specific proteases (USP) family is the largest class of DUBs, with more than sixty members. USPs are cysteine proteases that use a papain-like mechanism to hydrolyze the isopeptide bond between the carboxy terminus of Ub and the ε-amine of the target lysine.

USPs are variable both in size and their modular domain architecture, which can include substrate binding domains, ubiquitin-like (UBL) domains and other protein-protein interaction domains. They share a common papain-like fold, but the catalytic domains can have large insertions, possibly directly affecting activity. Ub binding or localization as seen in USP4, USP5, USP14 and CYLD. Additionally, some USPs need structural rearrangements to bind their substrate and catalyze hydrolysis.

USPs are often found in large protein complexes and many interaction partners of USPs have been identified. Although the function of most interaction partners is still unclear, some play a role in the modulation of USP activity. For example, GMP synthetase (GMPS) interacts and activates USP7, USP5, USP14 and CYLD. Additionally, some USPs need structural rearrangements to bind their substrate and catalyze hydrolysis.

Results
Protein cloning, expression and purification
After protein expression trials, we identified constructs suitable for large-scale protein expression of twelve USPs in either E.coli or in Sf9 insect cells. In this study we could therefore include sixteen constructs containing either the (almost) full-length constructs (USP1ΔN, USP7FL, USP11FL, USP16FL, USP25FL and USP46ΔN, with ΔN and ΔC denoting N- and C-terminal truncations respectively), the catalytic domain (USP4CD, USP7CD, USP8CD, USP16CD, USP21CD, USP25CD and USP46ΔN).
Figure 1. Overview of the characterized USPs. A) Domain architecture of the USPs used in this study. The constructs used in this manuscript are highlighted with corresponding residue numbers and expression system. B) Final purification product of the USP constructs shown on SDS-PAGE gel. An asterisk indicates the expressed USP7FL has an N-terminal GST tag.

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Catalytic residues</th>
<th>Cysteine residues</th>
<th>DUSP domain</th>
<th>Ubiquitin-like domain</th>
<th>TRAF/MATH domain</th>
<th>Zinc-finger</th>
<th>Rhodanese domain</th>
<th>Mitochondrion &amp; Transport domain</th>
<th>SUMO Interacting</th>
<th>Ubiquitin Interacting</th>
<th>Localization/protein protein interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>USP1FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP1__N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP4FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP4CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP4-D1D2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP7FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP7CD-HUBL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP7CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP8FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP8CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP11FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP12FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP21FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP21CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP25FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP30FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP30CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP39FL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USP39CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Expression Systems:
- Sf9: +
- E.coli: +

Residue Numbers:
- USP1FL: 1 - 785
- USP4FL: 1 - 963
- USP4CD: 296 - 954
- USP4-D1D2: 296 - 990
- USP7FL: 1 - 1102
- USP7CD-HUBL: 208 - 1102
- USP7CD: 208 - 560
- USP8FL: 1 - 1118
- USP8CD: 776 - 1110
- USP11FL: + 1 - 920
- USP12FL: + 1 - 370
- USP16FL: + 1 - 823
- USP16CD: + 193 - 823
- USP21FL: 1 - 565
- USP21CD: + 211 - 565
- USP25FL: + 1 - 1055
- USP30FL: 1 - 517
- USP30CD: + 65 - 500
- USP39FL: 1 - 565
- USP39CD: + 222 - 565
- USP46FL: + 8 - 366
USP30CD and USP39CD) (Figure 1A and B). Additionally, we expressed and purified two known USP activity modulators: UAF1\textsuperscript{23} and GMPS\textsuperscript{19}. Cloning, expression and purification protocols are provided in the materials and methods section.

**Large variations in both catalytic turnover and Ub binding**

Although USP family members share a homologous catalytic domain, many contain insertions within their catalytic domain or have additional domains that could influence their activity\textsuperscript{6,7} (Figure 1A). To study these effects, we determined the kinetic parameters of all the USPs. To this end, we produced a minimal synthetic Ub substrate with fused at its C-terminus the small molecule 7-amino-4-methylcoumarin (UbAMC)\textsuperscript{22,24}. The UbAMC substrate is a widely used reagent to assay DUB activity. Upon hydrolysis by the USP, the free AMC reporter molecule produces a fluorescent signal, which allows for a direct read-out of activity. Since this universal DUB substrate contains an AMC moiety instead of the endogenous USP target, it is suitable for comparing the relative activity among the USP family members. With this substrate, we observed variations of several orders of magnitude in both $K_M$ and $k_{cat}$ between the USP constructs (Figure 2 and Supplemental Figure S1)). Our data are in agreement with earlier reports for USP1, USP4, USP7, USP8 and USP39\textsuperscript{7,11,21,25,26}. Based on their $K_M$ and $k_{cat}$ values, the USPs could be classified in three groups (Figure 2D). Group 1 represents the USPs, whose activity is very limited due to a low $k_{cat}$ (USP1\textsubscript{ΔN}, USP4CD, USP5CD-D1D2). Group 2 contains USPs with intermediate activity and group 3 contains the USPs with the highest activity. Dashed lines link the catalytic domains with the corresponding full length USPs. Solid lines show the effect of intra-molecular activating and inhibiting domains.

**Figure 2. Kinetic parameters using UbAMC.** A and B) The Michaelis-Menten curves for the different USPs obtained by determining the initial rates ($V_0$) at different UbAMC concentration. B) shows the USPs with intra-molecular modulation. C) Overview of the kinetic parameters ($k_{cat}$, $K_M$ and $k_{cat}/K_M$) for the different USPs. D) Activity classification of USPs, based on kinetic parameters, where group 1 represents the USPs with the lowest activity; group 2 contains USPs with intermediate activity and group 3 contains the USPs with the highest activity. Dashed lines link the catalytic domains with the corresponding full length USPs. Solid lines show the effect of intra-molecular activating and inhibiting domains.
USP12, USP30CD, USP39CD and USP46ΔN). The “intermediate” group 2 contains the USPs that show moderate activity (USP4-D1D2, USP11FL, USP16CD, USP16FL, USP21CD and USP25FL), and group 3 contains very active USPs (USP7FL, USP7CD-HUBL and USP8CD).

As expected, group 1 contains USP39CD. It shows no activity, since it lacks the catalytic cysteine and histidine residues. Group 1 also contains USP1ΔN, USP12FL and USP46ΔN, all three known to have low activity, which is enhanced by the external modulator UAF1. Interestingly, also USP30CD shows very little activity. However, to date there is no known activator for USP30CD, although several interaction partners have been identified.

In contrast, group 3 represents the most active USPs, and contains both USP8CD and the USP7 constructs with activating C-terminal HAUSP UBL (HUBL) domain. Interestingly, USP8CD has an unusual high $K_m$ which is possibly due to an inserted α-helix in the catalytic domain, which is suggested to stabilize the observed closed conformation. This is compensated by a very high catalytic turnover, rendering it a very active USP overall.

**Intra-molecular modulation of USP activity**

Not only do we observe differences in enzymatic behavior between the USPs, but we also observe differential effects of intra-molecular domains on the activity of the (minimal) catalytic domains in USP4, USP7 and USP16 (Figure 2B).

We recently showed that USP4 contains a UBL domain inserted in its catalytic domain (USP4CD; Figure 1A), which inhibits the activity of USP4CD (group 1; Figure 2B,D). The presence of this UBL domain in USP4CD increases the $K_m$ and is therefore less active than the minimal catalytic domain USP4-D1D2 (group 2; Figure 2D). In contrast, both $k_{cat}$ and $K_m$ are affected in USP7, where the minimal catalytic domain (group 1) shows far less activity than the full-length enzyme (group 3). Here, the activity of USP7 is modulated by its HUBL domain which is essential for both activity and Ub binding in vitro and in vivo. The activity of USP16CD is modulated by the zinc-finger Ub specific protease (ZnF-UBP). Surprisingly, the activity is enhanced by increasing catalytic turnover, rather than the $K_m$ (Figure 2B,D). Since it is a Ub binding domain, the effect of the zinc-finger could be more prominent in poly-Ub processing, which might add up to a bigger difference than observed here. USP39CD also contains a ZnF-UBP domain, but it is unlikely that this will lead to enzymatic activation since USP39CD does not have the catalytic residues.

Overall, this shows that several intra-molecular domains are able modulate USPs. The modulation can affect $K_m$ (USP4), $k_{cat}$ (USP16), or both (USP7), and both inhibitory and activating domains are found in USPs. Together, this creates an additional layer of regulation on the catalytic activity of USPs.

Some USPs show small preference for di-Ub topoisomers

Most studies of DUB specificity have focused on processing K48- and K63-linked poly-Ub. However, since the additional linkages serve equally important cellular functions, we synthesized all seven lysine-linked di-Ub topoisomers and used them in a qualitative assay to assess all linkage preferences of the panel of USPs (Figure 3 and Supplemental Figure S2). In agreement with the kinetic parameters from the UbAMC assay, the USPs from group 1 showed very little activity; USP8CD from group 3 is the most active USP, and most USPs from group 2 show an intermediate activity. However, there were two clear changes. Where USP7 was amongst the most active USPs in the UbAMC assay, now it shows an intermediate activity. In contrast, USP21CD showed intermediate activity in the UbAMC assay, but is very active in the di-Ub assay and displays activities almost matching the most active USP; USP8CD.

The USP family seems to be rather promiscuous compared to other DUB families. For example, Cezanne (K11), OTUB1 (K48) and TRABID (K29) from the OTU family display strong linkage preferences for di-Ub topoisomers, and we used them in a qualitative assay to assess all linkage preferences of the panel of USPs (Figure 3 and Supplemental Figure S2). In agreement with the kinetic parameters from the UbAMC assay, the USPs from group 1 showed very little activity; USP8CD from group 3 is the most active USP, and most USPs from group 2 show an intermediate activity. However, there were two clear changes. Where USP7 was amongst the most active USPs in the UbAMC assay, now it shows an intermediate activity. In contrast, USP21CD showed intermediate activity in the UbAMC assay, but is very active in the di-Ub assay and displays activities almost matching the most active USP; USP8CD.

The USP family seems to be rather promiscuous compared to other DUB families. For example, Cezanne (K11), OTUB1 (K48) and TRABID (K29) from the OTU family display strong linkage preferences for di-Ub topoisomers. However, figure 3 shows that the differential activity of the USPs is smaller. All the active USPs from this study hydrolyze all di-Ub topoisomers. Nevertheless, there are clear differences in efficiency. For instance, most USPs have difficulties in hydrolyzing K27- and, to a lesser extent, K29-linked di-Ub. For example, USP7 has limited activity towards hydrolyzing K27- and K29-linked di-Ub. In contrast, the K6, K11, K48 and K63 Ub topoisomers are hydro-
lyzed relatively efficiently. Another clear example is USP4, for which K63-linked di-Ub is a better substrate than K48-linked di-Ub. Apparently, some USPs seem to prefer specific di-Ub isoforms.

We wondered whether the intramolecular modulating domains in USP4, USP7 and USP16 change the linkage preferences. However, this does not seem to be the case. The different USPs respond differently to modulation by internal domains, analogous to what was observed with UbAMC (Figure 3 and Supplemental Figure S2 B,C). However, no change in linkage preference was seen between catalytic domain and longer constructs, showing that the modulation effects are substrate independent mechanisms.

Overall, this shows that USPs can hydrolyze all Ub lysine-linked di-Ub topoiso- mers, but with differences in efficiency. Moreover, these differences are preserved in the presence of the intramolecular activity modulators.

In the case of USPs, isopeptide-linked Ub is not representative for di-Ub

To explain the Ub linkage preference, we might not need full-length di-Ub. To test this in an activity assay, we designed and synthesized a panel of fluorescence polarization-based (FP) di-Ub mimics. In
these reagents, TAMRA-labeled Ub peptides were linked via an isopeptide linkage to the carboxy-terminus of wild-type full-length mono-Ub\(^3\) (Figure 4A and Supplemental Figure S3)). Therefore, in contrast to the peptide linkage in UbAMC, these FP-reagents use the natural isopeptide linkage. The proximal Ub is represented by 14-mer peptides, each representing one of the seven lysines of Ub (Figure 3A and 4A, Table). In addition, the di-peptide (KG) was prepared to serve as a minimal substrate. Mass spectrometry and SDS-PAGE analysis of these new Ub substrates showed that the synthesis was successful for all eight different TAMRA labeled isopeptide-linked Ub FP-reagents (Supplemental Figure S3F).

As a proof of principle, we used the minimal ‘KG’ FP-reagent to determine the kinetic parameters of USP4-D1D2 (Figure 4B and Supplemental Figure S3H). With this reagent we determined \(K_M = 293 \text{ nM} \pm 53\) and \(k_{cat} = 0.07 \text{ s}^{-1} \pm 0.007\) values similar to the kinetic parameters obtained using UbAMC. Only the \(k_{cat}\) value is higher, possibly due to the difference in the chemical nature of the linkage, since the FP-reagents contain a natural isopeptide linkage in contrast to the UbAMC reagent. However, since the \(K_M\) values are similar, both represent comparable Ub reagents.

In the di-Ub time course assay, we observed linkage preferences of USP4-D1D2 and USP7; e.g. USP7 prefers the hydrolysis of K6 over K27-linked di-Ub, and USP4-D1D2 prefers K63 over K48-linked di-Ub (Figure 3B). Although difficult to fit for USP7, with our FP-reagents we observed no difference in activity for either USP4-D1D2 or USP7, and therefore could not recapitulate the preferences observed in the di-Ub assay (Figure 3B and Supplemental Figure S3G). This shows that these FP reagents do not contain the required information to mimic di-Ub for USPs (Figure 4C and Supplemental Figure S3).

### The proximal Ub does not contribute, but rather hinders binding to USP7

Since the FP-reagents were not sufficient to reproduce the observed linkage preference, we used full-length di-Ubs to determine the kinetic parameters directly. We determined \(K_M\) and \(k_{cat}\) of the hydrolysis of all seven lysine linked di-Ubs by USP7, using initial rate experiments that monitored the appearance of mono-Ub (Figure 5). These experiments showed that the linkages that are efficiently hydrolyzed by USP7 (K6, K11, K33, K48 and K63) have similar kinetic behavior (Figure 5B). Interestingly, the \(K_M\) and \(k_{cat}\) values are similar to the minimal substrate UbAMC, which contains only a single Ub moiety. This suggests that there is no induced binding or catalysis effect by the proximal Ub moiety.

In the initial di-Ub assay, two linkages (K27 and K29) showed a clear delay in hydrolysis by USP7 (Figure 3). This was nicely reproduced in this kinetic di-Ub assay. Interestingly, there was hardly any change in \(k_{cat}\), but rather the \(K_M\) increased far above the concentrations used in our assays. This suggests that the preference for the di-Ub topoisomers arises from steric hindrance, rather than an

<table>
<thead>
<tr>
<th>Linkage</th>
<th>Res. Nr.</th>
<th>Peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1-14</td>
<td>MQIFVTLTGKTTIT</td>
</tr>
<tr>
<td>K6</td>
<td>1-14</td>
<td>MQIFVTLTGKTTIT</td>
</tr>
<tr>
<td>K11</td>
<td>4-17</td>
<td>SVTIFVAKIQK--</td>
</tr>
<tr>
<td>K27</td>
<td>20-33</td>
<td>SVTIFVAKIQK--</td>
</tr>
<tr>
<td>K29</td>
<td>22-35</td>
<td>SYTVAKIQDGIPFD--</td>
</tr>
<tr>
<td>K33</td>
<td>26-39</td>
<td>SYTVAKIQDGIPFD--</td>
</tr>
<tr>
<td>K48</td>
<td>41-54</td>
<td>OQLIFAGQLEDGR--</td>
</tr>
<tr>
<td>K63</td>
<td>56-69</td>
<td>LEDYNQKESTLHL--</td>
</tr>
</tbody>
</table>

Figure 4. Kinetics of di-Ub hydrolysis. A) Schematic view of N-terminal TAMRA labeled ubiquitin peptide (K6) conjugated with ubiquitin. Table shows the peptide sequences used with the corresponding residue numbers for the different types of ubiquitin linkages. The conjugated lysine is highlighted in red. B) Michaelis-Menten curves for USP4-D1D2 (top) and USP7FL (bottom) were obtained using the TAMRA labeled ubiquitin peptides in a FP hydrolysis assay. The curves for USP7 could not be fitted.
additional binding site for the proximal Ub moiety. Therefore, the binding of some linkages to the catalytic domain is impaired, resulting in lower activity.

**Intermolecular activation of USPs by UAF1 and GMPS only affects** $k_{\text{cat}}$

Besides their intrinsic activity, some USPs are activated by intermolecular modulation. For example, USP1, USP12 and USP46 are activated by the WD40-repeat containing UAF1, and USP7 is activated by GMPS$^{17,19-21}$. Here, we used the UbAMC assay to quantify this activation (Figure 6A,B and Supplemental Figure S4). In agreement with previous data, we observe mainly a $k_{\text{cat}}$ increase (7-fold) of USP1ΔN activity in the presence of UAF1. The USP1 used in this work has a mutation in the self-cleavage site (Gly671,672Ala)$^{21}$. UAF1 also activates USP12FL and USP46ΔN, where the $k_{\text{cat}}$ is increased by 66- and 70-fold, respectively. Also in the case of USP7 we observed a $k_{\text{cat}}$ increase (5.5-fold) in the presence of its modulator GMPS. Interestingly, in contrast to variable modulation invoked by internal domains (Figure 2D), intermolecular modulation is achieved mainly by an increase in the catalytic turnover rather than in substrate binding (Figure 6B).

To investigate whether this activation also induces new linkage preferences of these USPs, we repeated the di-Ub assay in the presence of UAF1 or GMPS (Figure 6C). As expected from the UbAMC kinetics, USP1ΔN shows limited activity in the absence of UAF1, while USP12FL and USP46ΔN show no activity. However, in the presence of UAF1, the activity of all three USPs is increased, albeit not to the same level. In complex with their activators, USP1ΔN and USP7CD-HUBL show most activity, but no change in chain type preference by UAF1 or GMPS. This agrees well with an activation mechanism that only increases $k_{\text{cat}}$, but does not induce binding, which should translate in changing $K_M$ values.

**Discussion**

In this study, we used chemical reagents to determine the kinetic parameters of substrate inde-
Figure 6. Intermolecular USP activity modulation is achieved by increasing $k_{cat}$.

A) Kinetic parameters ($k_{cat}$, $K_m$, and $k_{cat}/K_m$) using UbAMC as substrate for USP1ΔN, USP12FL and USP46ΔN in presence of UAF1 and USP7CD-HUBL in presence of GMPS.

B) Alternative representation of the kinetic parameters comparing the USP activity between the USPs and in the presence of their modulator.

C) Activity modulation by UAF1 and GMPS towards all seven di-Ub topoisomers. Samples from each time-point (0, 5, 10, 30, 60, 180 min) were analyzed on coomassie stained SDS-PAGE gels.

pendent activity of 12 USPs, their di-Ub linkage preference and characteristics of both intra- and intermolecular activity modulation. We observe large variations in both the catalytic turnover ($k_{cat}$) and Ub binding ($K_m$) between USPs, which can of the DUBs so far has been tested for all Ub linkages, some DUBs show remarkable specificity. For example the OTU protease DUBA11 is K63-specific, OTUB130,31 is K48-specific, while AMSH42 a small preference for individual di-Ub topoi-

GMPS can activate USPs by increasing their catalytic turnover ($k_{cat}$). Additionally, the USPs have a small preference for individual di-Ub topoisomers. We show that in USP7 there is no additional Ub binding site, but rather that the differences in hydrolysis of the topoisomers are achieved by hindering binding ($K_m$) sterically. The combined data provide insights in the variation in the biochemical behavior of the USP enzyme family.

Based on their specific catalytic efficiency (Figure 2), the USPs can be classified into three groups: (1) a group of USPs showing very low DUB activity, (2) an `intermediate´ group and (3) USPs that exhibit high activity. This variability in activity can be explained in several ways. First, the activity can be affected by structural rearrangements in both Ub binding sites and active sites, as shown by structural studies11,12. Secondly, intra-molecular domains of USPs can modulate the DUB activity, as seen here for USP4, USP7 and USP16. External modulator proteins can further regulate the activity of the USP by enhancing its activity, as seen for USP1, USP7, USP12 and USP46 (Figure 6A,B).

Here we report a few cases where intra-molecular modulators regulate the USP catalytic efficiency: either insertions within or additional domains outside the catalytic domain. For both USP7 and USP16 the enzymatic behavior is regulated by intra-molecular domains (the HUBL and ZnF-UBP domain, respectively) outside the catalytic domain, resulting in the increase of the activity. Additionally, variations in kinetics can be induced by (large) insertions in the catalytic domains themselves, as demonstrated for USP4, where a UBL containing insert is inhibiting the catalytic efficiency7. These variations and intra-molecular modulations result in the unique activity of each USP.

For the last decade, the focus on DUB specificity has been on K48- and K63-linked poly-Ub chains. However, different Ub linkage topoiso-
mers can result in different cellular fates, some of which are very specific36-38 and others requiring a minimal chain length to invoke its function39,40.

Our study presents the first complete and comprehensive study on di-Ub preference of all seven linkages for USP family members. Although none of the DUBs so far has been tested for all Ub linkages, some DUBs show remarkable specificity. For example the OTU protease DUBA41 is K63-specific, OTUB130,31 is K48-specific, while AMSH42 and BRCC3 43 both from the JAMM/MPN+ family are K63-specific. Next to CYLD10, the USPs do not have strict chain-type specificity, but rather have preferences. Kinetic studies on USP7 showed us that there is no proximal S1' Ub binding site to induce Ub topoisomers preference, but rather the proximal Ub moiety induces steric constraints for binding to the USP in the case of K27 or K29 linkages. However, it is possible that linkage specificity is increased when using longer Ub chains.

Overall, the hydrolysis efficiency of the USPs towards K6-, K11-, K48- and K63-linked Ub was higher than for K27- and, to a lesser extent, K29- and K33-linked di-Ub. These residues localize in distinct regions on Ub (Figure 3A). The lysine residues involved in the easiest hydrolyzed linkages (K6, K11, K48 and K63) are in the β-sheet or loops. In contrast, the lysine residues of the more difficult linkages (K27, K29 and K33) are positioned on the other side of the Ub molecule, and are all in the α-helix. Additionally, K27 is barely accessible, which possibly induces a steric constraint, resulting in the lower activity. This interesting bi-polar behavior needs future investigation.

Previous studies suggested that Ub-peptide reagents were sufficient to mimic di-Ub and discriminate between topoisomers in binding34. However, in our activity assays with the FP Ub-peptide reagents, we observed no difference between Ub linkages. This suggests that the peptides do not contain enough information to mimic the proximal Ub for the USPs. Nevertheless, they may be sufficient for DUBs from families with more pronounced Ub specificity and be useful tools in those cases. In addition, the ‘KG’ FP-reagent might prove a good alternative for UbAMC, as the kinetic parameters are similar, while it contains the natural isopeptide linkage, which is not present in UbAMC.

This study confirmed that two known intermolecular USP activity modulators UAF1 and GMPS activate USP1, USP12, USP46, and USP7 respectively. This activation is mainly by increasing the k\_cat. However, the biological roles of the UAF1 and GMPS activation are distinct. UAF1 activation is almost essential for USP activity of USP1, USP12 and USP46. This resembles the Ubp8 activation by Sgf1114,15. Surprisingly, USP12 in complex with UAF1 is still not very active, possibly requiring additional partners, like WDR2044. In a different manner, GMPS hyper-activates USP7, by allosterically stabilizing the active state of the enzyme induced by the HUBL domain17.

Besides a general activation, the GMPS activity modulation most likely has additional substrate specific roles, as it induces H2B de-ubiquitination.

Although the function of an increasing number of USPs is elucidated, they still represent a relatively uncharacterized enzyme family. To aid in the biochemical understanding of these enzymes, we here report the large variations in kinetics and intra-molecular modulation (k\_cat and K\_m), the modest but surprising differential activity towards the seven di-Ub topoisomers (K\_m), and a characterization of the activation by intermolecular interactions (k\_cat).

**Significance**

Ubiquitination is a dynamic process, which is involved in numerous key cellular processes. The removal of the Ub molecules is an integral part of this process, and is carried out by Deubiquitinating enzymes (DUBs). These are increasingly recognized as interesting drug targets. However, to date we lack the markers to predict the biochemical behavior based on sequence alignments and therefore a need exists for comprehensive kinetic studies. This is where chemical tools that allow fast and accurate read-outs will contribute to answer these biological questions. In this study, we designed and produced several of such chemical reagents to determine the kinetics and di-Ub linkage preferences of twelve USPs. Despite the homologous catalytic domain, the kinetic data underline the large variability within the USP family, and the intra- and intermolecular activity modulators create an additional layer of regulation.

In addition this study for the first time reports the linkage preference of twelve USPs against all seven-lysine linked di-Ubs. Kinetic analysis of the hydrolysis of the di-Ub topoisomers, suggest
that within the USP family the preferences are induced by steric hindrance, rather than the induced binding, as seen in other DUB families.

Together, this data provides insight in the biochemical behaviour in the USP family, and validates the chemical tools that now also can be applied in characterizing other DUB families.

**Material and Methods**

**General**

General reagents were obtained from Sigma Aldrich, Fluka and Acros and used as received. Solvents were purchased from Biosolve or Aldrich. Peptide synthesis reagents were purchased from Novabiochem. USP25 cDNA was provided by Erik Meulmeester and Frauke Melchior.

**General plasmids and proteins**

Di-Ub moieties were produced as previously described. USP4CD (aa 296-954), USP4-D1D2 (aa 296-490/766-932), USP8CD (aa 776-1110), USP11FL (aa 1-920), USP16FL (aa 1-823), USP16CD (aa 193-823), USP21CD (aa 211-565), USP30CD (aa 65-500), USP39CD (aa 222-565) and USP46ΔN (aa 8-366) are cloned into the pETNKI-LIC vector for expression in bacteria as described. USP1ΔN (aa 21-785 self-cleavage site glycine 671 and 672 are mutated to alanine), USP12FL (aa 1-355) and STREP-TEV-UAF1 (6-677) are cloned into the pFastBac vector for expression in insect cells. Both USP1ΔN-HUBL (aa 208-1102) and USP1ΔN (aa 208-560) are cloned into the pGEX vector and USP25FL is cloned in the pET11a vector.

**Protein expression and purification**

As specified in figure 1, the USPs were expressed in both E.coli and insect cells and purified as described. GMPS was expressed and purified as before. USP constructs and GMPS cloned both in bacterial and baculovirus expression vector are expressed and purified as described. Depending on the type of vector, the tag was removed with either TEV or the HRV 3C protease. Bacmids were prepared following the manufacturer’s guidelines. USP1, USP12 and UAF1 were produced using Sf9 and Sf21 insect cell expression. Infection was done using a low-MOI infection protocol. The cells were harvested 72 hours after a baculovirus induced growth arrest was observed. USP46 was produced in E.coli. USP1, 12, 46 and UAF1 were purified using Ni²⁺ sepharose (GE Healthcare) in 20 mM Hepes (pH 7.5), 150 mM NaCl, 0.1 mM PMSF, and 0.1 mM DTT followed by elution using imidazole. His-tag was removed by overnight cleavage with TEV protease whilst dialyzing to remove imidazole. Uncleaved product was removed with Ni²⁺ sepharose. Size exclusion chromatography was performed using a Superdex 200 or 75 column (GE Healthcare), equilibrated against buffer containing 10 mM Hepes [pH 7.5], 100 mM NaCl and 1 mM DTT. All proteins were concentrated to ~10 mg/ml and stored at -80°C.

**UbAMC assay**

Kinetics were determined as described before. UAF1 and GMPS were added in a 1:1 stoichiometry. USP concentration varied between 1 and 100 nM, depending on relative activity. In order to calculate the kinetic parameters for the hydrolysis of UbAMC, curves obtained by plotting the measured enzyme initial rates (v) versus the corresponding substrate concentrations ([S]). These were subjected to nonlinear regression fit using the Michaelis–Menten equation \( V = \frac{V_{max} \cdot [S]}{[S] + K_M} \) (eqn 1), where \( V_{max} \) is the maximal velocity at saturating substrate concentrations and \( K_M \) the Michaelis constant. The \( k_{cat} \) value was derived from the equation \( k_{cat} = \frac{V_{max}}{[E_o]} \) (eqn 2) where \([E_o]\) is the total enzyme concentration. Experimental data was processed using Prism 5.01 (GraphPad Software, Inc.).

**Di-Ub assay**

Di-Ub hydrolysis reactions were performed at 37°C in 50 mM Hepes buffer at pH 7.5, with 100 mM NaCl, 1 mM EDTA, 5 mM dithiothreitol and 0.05% (w/v) Tween-20 with constant enzyme concentration (75 nM). When indicated UAF1 was added in a 2-fold excess (150 nM) and GMPS in a 1:1 stoichiometry. Reactions were stopped by addition of SDS loading buffer and followed by SDS-PAGE analysis. For the kinetic analysis, the reaction mixture was pre-heated to 37 °C degrees before...
adding USP7. Samples were run on a 12% Bis-Tris NuPage gel (duplicates on one gel), and western blots were performed with anti-Ub antibody (Santa Cruz; P4D1). The ChemiDoc system (Biorad) was used to read the chemiluminescence signal and subsequent quantification of mono-Ub was done using the quantification tools of ImageLab (Biorad) using the non-saturated di-Ub signal (corrected for conversion to mono-Ub). Experimental data was processed using Prism 5.01 (GraphPad Software, Inc.).

**Solid Phase Peptide Synthesis (SPPS) of the TAMRA thiolysine peptides**

SPPS was performed on a Syro II MultiSyntech Automated Peptide synthesizer using standard 9-fluorenylmethoxycarbonyl (Fmoc) based solid phase peptide chemistry at 25 µmol scale, using fourfold excess of amino acids relative to pre-loaded Fmoc amino acid Wang type resin (0.2 mmol/g; Applied Biosystems). The following protected amino acids were used during Ub peptide synthesis: Fmoc-L-Ala-OH, Fmoc-L-Arg-(Pbf)-OH, Fmoc-L-Asn (Trt)-OH, Fmoc-L-Asp (OBU)-OH, Fmoc-L-Gln (Trt)-OH, Fmoc-L-Glu (OBU)-OH, Fmoc-Gly-OH, Fmoc-L-His (Trt)-OH, Fmoc-L-Ile-OH, Fmoc-L-Leu-OH, Fmoc-L-Lys (Boc)-OH, Fmoc-L-Met-OH; Fmoc-L-Phe-OH; Fmoc-L-Pro-OH; Fmoc-L-Ser (tBu)-OH; Fmoc-L-Thr (tBu)-OH, Fmoc-L-Tyr (tBu)-OH, Fmoc-L-Val-OH. Fmoc-S-(methyldisulfanyl)-(L)-Lys (Boc)-OH was synthesized as described previously.

The coupling procedure starts off with single couplings in N-methylpyrrolidon (NMP) for 45 min using PyBOP (4 equiv) and DiPEA (12 equiv) in a total volume of 750 µL. Followed by the removal of Fmoc with 20% piperidine in NMP for 2×2 and 1×5 min. Finally the procedure ends with NMP wash steps after each coupling (3×) and deprotection (5×).

The resin was washed with diethyl ether and dried under high vacuum. Next, the polypeptide sequence was detached from the resin and deprotected by treatment with TFA/H₂O/Phenol/iPr₃SiH 90.5/5/2.5/2 v/v/v/v for 2.5 h. After washing the resin with 3×1 mL TFA, the crude protein was precipitated with cold Et₂O/n-pentane 3:1 v/v. The precipitated protein was washed 3× with diethyl ether, the pellet was dissolved in a mixture of H₂O/CH₃CN/HOAc (65/25/10 v/v/v) and finally lyophilized. All peptides were analyzed by LC-MS and purified by RP-HPLC when necessary.

**LC-MS**

LC-MS measurements were performed on a Waters 2795 Separation Module (Alliance HT), equipped with a Waters 2996 Photodiode Array Detector (190-750 nm), Phenomenex Kinetex C18 column (2.1×50, 2.6 µm) and LCTTM Orthogonal Acceleration Time of Flight Mass Spectrometer. Samples were run using two mobile phases: A = 0.1% formic acid in water and B = 0.1% formic acid in acetonitrile. Flow rate = 0.8 mL/min, runtime = 6 min, column T = 40 °C. Gradient: 0 – 0.5 min: 5% B; 0.5 – 4 min: 5% to 95% B; 4 – 5.5 min: 95% B. Data processing was performed using Waters MassLynx Mass Spectrometry Soft-ware 4.1 (deconvolution with Maxent1 function).

**Ligation of Ub to the peptides followed by desulphurization**

Schematic overview of reaction scheme and final yields can be found in Supplemental Figure S3. A mixture of 4-mercaptophenylacetic acid (MPAA, 100 mM) and TCEP (50 mM) in 6M Guanidinium -HCl (1 mL, pH 7) was added to Ub-MesNa thioester (5 mg, prepared according to the procedure described previously). To this the TAMRA thiolysine peptide (100 µL of a 20 mM stock solution in DMSO) was added and the whole mixture was incubated at 37 °C. After overnight incubation, all low-molecular weight material was removed using a 3 kDa cutoff spin-column (Amicon Ultra) in four centrifuge cycles. The crude material was taken up in 6M Guanidinium-HCl, 0.1M sodium phosphate (4 mL, pH 6.5) and to this was added TCEP (187 mg) and glutathione (30 mg), after which the pH of the mixture was adjusted to pH 6.5 by addition of 1M NaOH. Next, the mixture was degassed with argon, after which radical initiator VA-044 was added. The mixture was incubated at 37 °C overnight. All constructs were purified by RP-HPLC and analyzed by LC-MS and gel electrophoresis and were obtained as purple solids.

**C18 RP-HPLC**

Purification by RP-HPLC was performed on a Shimadzu system equipped with a LC-20AT liquid chromatography pump, CTO-20A column oven (T = 40 °C), SPD-20A UV/VIS detector (detection simultaneously at 230 nm and 254 nm), RF-
10AXL fluorescence detector (ex/em = 540/600 nm) and an Atlantis Prep T3 column (10×150 mm, 5 μm). Samples were run using two mobile phases: A = 0.05% trifluoroacetic acid in water and B = 0.05% trifluoroacetic acid in acetonitrile. Flow rate = 7.5 mL/min, runtime = 30 min. Gradient: 0 – 6 min: 5% to 10% B; 6.5 – 26 min: 25% to 47% B; 26.5 – 29.5 min: 95% B. Pure fractions were pooled and lyophilized.

Isopentide linked Ub FP hydrolysis assay
FP assays were performed on a PerkinElmer Wallac EnVision 2010 Multilabel Reader with a 531 nm excitation filter and two 579 nm emission filters. The confocal optics were adjusted with TAMRA-KG (synthesized by SPPS as described above) and the G factor was determined using a polarization value for TAMRA-KG (25 nM) of 50 mP. The assays were performed in “non binding surface flat bottom low flange” black 384-well plates (Corning) at room temperature in a buffer containing 20 mM Tris-HCl, pH 7.5, 5 mM DTT, 100 mM NaCl, 1 mg/mL 3-[ (3-cholamidopropyl) dimethylammonio] propanesulfonic acid (CHAPS) and 0.5 mg/mL bovine gamma globulin (BGG). Each well had a volume of 20 μL. Buffer and enzyme were predisposed and the reaction was started by the addition of substrate. Kinetic data was collected in intervals of 2.5 or 3 min. From the obtained polarization values ($P$) the amount of processed substrate ($P_t$) was calculated with to the following equation: $S = S_0 - S_0 \cdot \left( \frac{P_t}{P_{max}} \right) \left( \frac{P_{max}}{P_{min}} \right)$, where $P_t$ is the polarization measured (in mP); $P_{max}$ is the polarization of 100% unprocessed substrate (determined for every reagent at all used substrate concentrations); $P_{min}$ is the polarization of 100% processed substrate (determined for every linkage at all used substrate concentrations); $S_0$ is the amount of substrate added to the reaction. From the obtained $P_t$ values the values for initial velocities were calculated, which were used to determine the Michaelis-Menten constants. All experimental data was processed using Ms Excel and Prism 4.03 (GraphPad Software, Inc.).

Acknowledgements
We thank Martin A. Cohn and Alan D’Andrea for USP1, USP12, USP46 and UAF1 ccDNA, Annette Dirac and Rene Bernards for USP4 and USP8 cDNA, Elisabetta Citterio for USP21 and USP39 cDNA, Carlos Lopez-Otin for USP30 cDNA and Erik Meulmeester and Frauke Melchior for USP25 cDNA. We thank Ovaa and Sixma group members for discussion, sharing reagents and critical reading of the manuscript. This work has been supported by grants from the Dutch Cancer Society, The Netherlands Organization for Scientific Research VIDI grant, and EU-Rubicon and NWO-CW ECHO 700.59.009 and KWF-2008-4014.

Author contributions
USP expression and purification was performed by M.P.A.L.V. with assistance from W.J.v.D; enzyme assays were designed and analyzed by A.C.F. and executed by M.P.A.L.V and A.C.F.; USP1, 12, 46 and UAF1 were expressed, purified and analyzed by M.C.; UbAMC was designed and synthesized by R.M.; di-Ubs were designed by F.E. and H.O. and synthesized by D.S.H.; FP reagents were designed by F.E., P.P.G. and H.O.; FP reagent assays were performed by P.P.G.; H.O. supervised all synthesis efforts and FP experiments; T.K.S. designed and supervised USP project; Data analysis and manuscript writing by A.C.F. with M.P.A.L.V. and T.K.S..
References


## Table of contents Supplementary information

<table>
<thead>
<tr>
<th>Table/Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table S1</td>
<td>Comparison with published UbAMC kinetics</td>
<td>82</td>
</tr>
<tr>
<td>Figure S1</td>
<td>Exemplary raw data UbAMC assay. Related to Figure 2.</td>
<td>82</td>
</tr>
<tr>
<td>Figure S2</td>
<td>Di-Ub assay for inactive USPs and comparison USP7CD versus USP7CD-HUBL. Related to Figure 3.</td>
<td>83</td>
</tr>
<tr>
<td>Figure S3</td>
<td>Synthesis, LC-MS spectra of FP-reagents, exemplary raw data and Michaelis-Menten curves of USP4 and USP7. Related to Figure 4.</td>
<td>84</td>
</tr>
<tr>
<td>Figure S4</td>
<td>Curves of the Michaelis-Menten analysis of the UAF1 and GMPS modulation. Related to Figure 6.</td>
<td>86</td>
</tr>
</tbody>
</table>
Supplemental Table S1. Related to Figure 1. Comparison with published UbAMC kinetics.

<table>
<thead>
<tr>
<th></th>
<th>$k_{cat}$ [$s^{-1}$]</th>
<th>$K_M$ [$\mu$M]</th>
<th>$k_{cat}/K_M$ [$10^{-3} M^{-1} s^{-1}$]</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>USP1FL</td>
<td>0.014</td>
<td>1.4</td>
<td>10.2</td>
<td>Cohn et al. (2007)</td>
</tr>
<tr>
<td>USP1FL+UAF1</td>
<td>0.26</td>
<td>0.7</td>
<td>371</td>
<td>Cohn et al. (2007)</td>
</tr>
<tr>
<td>Ubp8/SAGA</td>
<td>0.17</td>
<td>1.5</td>
<td>110</td>
<td>Samara et al. (2011)</td>
</tr>
<tr>
<td>USP2CD</td>
<td>0.62</td>
<td>2.5</td>
<td>250</td>
<td>Zhang et al. (2011)</td>
</tr>
<tr>
<td>USP2CD</td>
<td>0.14</td>
<td>0.55</td>
<td>252</td>
<td>Renatus et al. (2006)</td>
</tr>
<tr>
<td>USP7FL</td>
<td>3.56</td>
<td>17.5</td>
<td>203</td>
<td>Fernandez-Montalvan et al. (2007)</td>
</tr>
<tr>
<td>USP7CD</td>
<td>0.077</td>
<td>44.2</td>
<td>1.7</td>
<td>Fernandez-Montalvan et al. (2007)</td>
</tr>
<tr>
<td>USP7CD-HUBL</td>
<td>0.805</td>
<td>22.8</td>
<td>35</td>
<td>Fernandez-Montalvan et al. (2007)</td>
</tr>
<tr>
<td>USP8CD</td>
<td>2.4</td>
<td>10.2</td>
<td>235</td>
<td>Avvakumov et al. (2006)</td>
</tr>
<tr>
<td>USP21CD</td>
<td>0.041</td>
<td>0.26</td>
<td>158</td>
<td>Ye et al. (2011)</td>
</tr>
<tr>
<td>USP25FL</td>
<td>0.12</td>
<td>5</td>
<td>24</td>
<td>Meulmeester et al. (2008)</td>
</tr>
</tbody>
</table>

Supplemental Table S1. Related to Figure 1. Comparison with published UbAMC kinetics.

Supplemental Figure S1. Related to Figure 2. Exemplary raw data of UbAMC hydrolysis by USP7CD-HUBL. Measurements were done using a five minute interval. The signal was stable for at least one hour.
Supplemental Figure S2. Related to Figure 3. Di-Ub assay for inactive USPs and comparison USP7CD versus USP7CD-HUBL. A and B) Di-Ub assay for USP12FL, USP39CD, USP46FL, USP7CD, USP16CD and USP30CD. Time-course using all di-Ub topoisomers (5 µM) (Linear, K6, K11, K27, K29, K33, K48 and K63) for USPs (75 nM). Samples from each time-point (0, 5, 10, 30, 60, 180 min) (A) and (0, 10, 30, 60 min) (B) were analyzed on coomassie stained SDS-PAGE gels. C) Di-Ub assay USP7CD versus USP7CD-HUBL. Time-course using all di-Ub topoisomers (5 µM) (Linear, K6, K11, K27, K29, K33, K48 and K63) for USPs (75 nM). Samples from each time-point (0, 10, 30, 60 min) were analyzed on coomassie stained SDS-PAGE gels.
Supplemental Figure S3. Related to Figure 4. Synthesis, LC-MS of FP reagents, exemplary raw data and Michaelis-Menten curves of USP4 and USP7. A. Ligation of Ub to the peptides. B. Peptide sequence, molecular weight and typical yield of the reaction. C. Coomassie staining and fluorescence scan of SDS electrophoresis analysis of the FP-reagent. D. Kinetic parameters of USP4-D1D2 using the FP-reagents. E. Exemplary data of hydrolysis of FP reagents.
Supplemental Figure S3. Related to Figure 4. Synthesis, LC-MS of FP reagents, exemplary raw data and Michaelis-Menten curves of USP4 and USP7. F. LC-MS spectra of the FP-reagents. G. Michaelis-Menten analysis of USP7 hydrolysis of the FP reagents. Data could not be fitted with an exponential Michaelis-Menten curve.

Figure and legend continued from previous page.

Figure and legend continue on next page.
Supplemental Figure S3. Related to Figure 4. Synthesis, LC-MS of FP reagents, exemplary raw data and Michaelis-Menten curves of USP4 and USP7. H. Michaelis-Menten analysis of USP4 hydrolysis of the FP reagents. Kinetic parameters are in (D).

Supplemental Figure S4. Related to Figure 6. Curves of the Michaelis-Menten analysis of the UAF1 and GMPS modulation. The Michaelis-Menten curves for the different USPs obtained by determining the initial rates (V0) at 2-fold serial dilutions of UbAMC.
Ubiquitin-specific protease 4 is inhibited by its ubiquitin-like domain

Mark P.A. Luna-Vargas, Alex C. Faesen, Willem J. van Dijk, Michael Rape, Alexander Fish & Titia K. Sixma


1Division of Biochemistry and Center for Biomedical Genetics, The Netherlands Cancer Institute, Amsterdam, The Netherlands, and
2Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California, USA

*Correspondence: t.sixma@nki.nl (T.K.S.)
Abstract
USP4 is a member of the ubiquitin-specific protease (USP) family of deubiquitinating enzymes that has a role in spliceosome regulation. Here, we show that the crystal structure of the minimal catalytic domain of USP4 has the conserved USP-like fold with its typical ubiquitin-binding site. A ubiquitin-like (Ubl) domain inserted into the catalytic domain has autoregulatory function. This Ubl domain can bind to the catalytic domain and compete with the ubiquitin substrate, partly inhibiting USP4 activity against different substrates. Interestingly, other USPs, such as USP39, could relieve this inhibition.

Introduction
Post-translational modification by the small, highly conserved ubiquitin (Ub) protein has an essential role in the regulation of many cellular processes in eukaryotes. In this process, the carboxy-terminus of Ub forms an isopeptide with lysines on the target proteins, or on Ub itself, to form poly-Ub chains. The activity of the conjugating enzymes E1–E2–E3 is actively balanced through hydrolysis by deubiquitinating enzymes (DUBs). Deregulation of the ubiquitination pathway can lead to cancer and neurodegenerative diseases. More than 100 putative DUBs are known so far, belonging to five subfamilies of isopeptidases. The Ub-specific protease (USP) family is the largest, with more than 60 members in the human genome. USPs share a papain-like catalytic domain and crystal structures show a conserved catalytic core that undergoes conformational changes on Ub binding. USPs are variable in size with modular domain architecture including, for example, TRAF-like, DUSP or Znf domains. Sequence analysis predicted the presence of Ub-like (Ubl) domains in 17 different USPs. Integrated Ubl domains are stretches of 45–80 amino acids that share the b-grasp fold of Ub, but often have poor sequence conservation among subfamilies. The Ubl domains in the USP family are located amino-terminally, within or C-terminally to the catalytic domain. Structural studies of the N-terminal Ubl domain of USP14 confirmed the Ubl-fold (Protein Data Bank (PDB): 1WGG) and showed involvement in proteasome binding that promotes the DUB activity of USP14. Similar to USP14, USP4 has a Ubl domain N-terminal of its catalytic domain, but it has an additional Ubl domain embedded in the catalytic domain. USP4 was previously known as ubiquitous nuclear protein (UNP). Identified as a proto-oncogene related to Tre 2/Tre 17 (USP6), USP4 shows a consistently elevated gene expression level in small cell tumours and lung adenocarcinomas, suggesting that it may have a possible causative role in neoplasia. Besides possible roles in Wnt signalling and recruitment to the A2A receptor, USP4 is recruited to the spliceosome by complex formation with Sart3. Here, it preferentially deubiquitinates K63-linked chains on the U4 component Prp3. Another component of the spliceosome complex is the catalytically inactive USP39, which controls the messenger RNA levels of Aurora B. Here, we report on the crystal structure of the catalytic domain of USP4 without the internal Ubl domain, and show how this Ubl domain acts as an autoregulatory domain that partially inhibits catalytic activity by competitive inhibition.

Results
Identification of USP4-D1D2
To gain insight into the structure and function of USP4, we expressed and purified the USP4 catalytic domain (amino acids 296–954, Fig 1A) in Escherichia coli. To improve the chances for crystallization, we used limited proteolysis. After treatment with thermolysin, two fragments—domain 1 (D1) and 2 (D2)—were obtained, which copurified on size exclusion chromatography and together retained DUB activity (supplementary Fig S1A,B online). We identified the composition of D1 and D2 using mass spectrometry and N-terminal sequencing (supplementary Fig S1C online). The protease treatment removed an insertion between Leu 481 and Leu 766 (supplementary Fig S2 online), yielding a minimal catalytic domain consisting of two fragments: USP4–D1D2.

Structure of the USP4–D1D2 catalytic domain
We crystallized and determined the USP4–D1D2 structure by molecular replacement using the...
USP8 catalytic domain (PDB: 2GFO) as the search molecule, and refined it to 2.4Å resolution with an R/R_free of 0.178/0.21 and good geometry (Fig 1B; supplementary Table S1 online). There are six molecules of USP4–D1D2 per asymmetrical unit, with a pairwise root-meansquare deviation of approximately 0.7Å over 344 residues using the PISA. Similar to crystal structures of other USPs, the catalytic domain of USP4–D1D2 resembles an extended right hand comprising three domains: Fingers, Thumb and Palm (Fig 1B; supplementary Fig S3 online). The D1 fragment contains the Thumb domain and part of the Fingers domain with the Cys box (amino acids 303–320) and QQD box (amino acids 390–403) of the active site, whereas the D2 fragment completes the active site with the His box (amino acids 864–885, 894–903, 915–922) and makes the remaining part of the Fingers and the Palm9 (supplementary Fig S2 online). Like other USP structures10,12-15, except USP711, the catalytic triad is in a catalytically competent configuration, wherein His 711-ND1 is 3.2Å away from Cys 311-SG and His 711-ND2 is hydro- gen bonding with Asp 898-OD1 (2.7Å; Fig 1B).

The zinc-finger ribbon observed in USP2 and USP8 is present in USP4 (Fig 1B,C). The Zn²⁺ ion brings together the D1 and D2 domains, tetrahedrally coordinated by cyste- ines on anti-parallel β-strands β1 and β2 in D1,
Figure 2. Insert inhibits the DUB activity of USP4CD. (A–C) The full-length USP4 catalytic domain (A) is much less active than (B) USP4–D1D2 or (C) USP4 fusion in deubiquitinating K63 di-Ub (Coomassie-stained SDS–PAGE gels). D) Quantification of mono-Ub in K63 di-Ub cleavage assays. The intensity of the mono-Ub band is plotted against time. E) The inhibitory effect of the insert is observed in a Ub-AMC assay. On comparing $K_m$ between USP4CD and D1D2, we observed a 90 times lower enzyme efficiency for the insert containing USP4CD. F) Inhibition of USP4–D1D2 in trans in Ub-AMC assays at different Ubl-insert concentrations (5, 15, 45 and 90 µM) can be jointly fit as a competitive inhibitor. SDS–PAGE, SDS–polyacrylamide gel electrophoresis; Ub, ubiquitin; Ub-AMC, ubiquitin-7-amido-4-methylcoumarin; Ubl, ubiquitin-like; USP, ubiquitin-specific protease.

and b4 in D2. This zinc-finger ribbon in the Fingers domain seems to be in the contracted ‘closed-hand’ configuration seen in USP8 that blocks access of Ub to its binding site. A similar role was assigned to the two Ub-binding surface loops (BL1 and BL2) in USP14 that block the active site, but relocate on Ub binding. In USP4, both loops (Fig 1D)—as well as a third blocking loop (BL3) that hinders access of the C-terminal tail of Ub to the binding pocket—are observed. Superposition of the six non-crystallographic symmetry-related molecules of USP4–D1D2 shows that both the zinc-finger ribbon and the three blocking loops show flexibility (maximal $\Delta \alpha$ displacement 4Å; Fig 1C,D), which is in agreement with their role in activation.

The insert inhibits deubiquitinating activity

We compared the catalytic activity of the USP4 catalytic domain with and without the large insert, by using in vitro deubiquitinating assays. In these assays we followed the hydrolysis of K63- and K48-linked di-Ub into mono-Ub (Fig 2A,B; supplementary Fig S4A,B online). We observed that K63 di-Ub is more efficiently degraded than K48, in agreement with the role of USP4 in splicing. Interestingly, quantification (Fig 2D; supplementary Fig S4D online) shows that USP4–D1D2 without insert is more efficient at degrading both di-Ub than the...
complete catalytic domain. When D1 and D2 are fused through a short linker, as found in USP7 (supplementary Fig S2 online), their activity is similar to that of USP4–D1D2, showing that the cause of the activation is the lack of insert and not the chain break (Fig 2C; supplementary Fig S4C online).

In Ub-7-amido-4-methylcoumarin (Ub-AMC) assays the intact USP4 catalytic domain is also less active than USP4–D1D2 or the fusion protein. As only AMC is cleaved off, the inhibition is not dependent on the protein target. When analysed by Michaelis–Menten kinetic analysis (Fig 2E) the \( V_{\text{max}} \) values were similar, but the \( K_{\text{m}} \) for the intact catalytic domain (13.5 µM) was weaker than that for USP4–D1D2 (0.20 µM), leading to approximately 90-fold lower catalytic efficiency overall (\( k_{\text{cat}}/K_{\text{m}} \)) for USP4CD than for USP4–D1D2.

As the insert seems to inhibit the DUB activity of USP4, we tested whether it could do so in trans. We expressed and purified the insert (amino acids 483–765) and added it in increasing amounts to USP4–D1D2 in the Ub-AMC assay (supplementary Fig S5A online). We observed that the insert slows deubiquitination by USP4–D1D2. To investigate whether this reduction in DUB activity is due to molecular crowding, we repeated

![Figure 3. Ubiquitin competes with the insert or Ubl-domain for binding to USP4–D1D2.](image-url)

(A–C) Interaction of Ub and the insert fragments with USP4–D1D2 was studied by SPR experiments. Top: (A) GST-tagged Ub, (B) GST-insert and (C) GST-Ubl domain were immobilized on anti-GST antibodies coupled to a CM5 Biacore chip and USP4–D1D2 was flowed over the chip at different concentrations. Bottom: Langmuir binding curves. D) Competition experiment with immobilized GST insert on USP4–D1D2 with varying concentrations of Ub. A one-site competition binding model was fitted (\( K_i = 1.4 \) µM). E) The interaction of Ub with USP4–D1D2 (left) and with full-length USP4CD (right) were studied by ITC analyses. Thermodynamic values for USP4–D1D2 (\( \Delta H = -14.3 \) kcal/mol and \( \Delta S = -16.9 \) cal/mol/deg), for USP4CD (\( \Delta H = -11.4 \) kcal/mol and \( \Delta S = -10.0 \) cal/mol/deg). GST, glutathione S-transferase; ITC, isothermal titration calorimetry; Ub, ubiquitin; Ubl, ubiquitin-like; USP, ubiquitin-specific protease.
the *in trans* inhibition assay with USP4–D1D2 in the presence of either SUMO or BSA (supplementary Fig S6 online). Neither of these reduced DUB activity, confirming that the insert is intrinsically able to inhibit the catalytic activity of USP4.

**Competitive inhibition of the USP4 insert**

We tested whether USP4–D1D2 would directly interact with the insert. In a surface plasmon resonance (SPR; Fig 3B) experiment, we observed binding of USP4–D1D2 to the insert, with a $K_d$ of 1.32 µM after equilibrium fitting. This affinity closely resembled the affinity of USP4–D1D2 for Ub itself ($K_d$ of 1.39 µM; Fig 3A).

Therefore, we tested whether the insert could compete with Ub for binding to USP4–D1D2, and would therefore bind to the same binding site. In an SPR competition experiment we flowed USP4–D1D2 over a glutathione S-transferase (GST)-tagged insert in the presence of increasing amounts of Ub (Fig 3D). We observed decreasing binding of USP4–D1D2 to the GST-insert as the Ub concentration increased. The data could be fitted with a one-site competition binding model with a $K_i$ of 1.4 µM, showing that the USP4 insert competes with Ub for binding to USP4–D1D2. Interestingly, the $K_d$ of intact USP4CD for Ub is only fourfold less, compared to USP4–D1D2 in an isothermal titration calorimetry (ITC) experiment (Fig 3E). Although the exact $K_s$ are slightly tighter in the ITC experiment, qualitative analysis of SPR experiments agrees with this assessment. Non-specific binding at high concentrations precluded detailed fitting of these data (supplementary Fig S7 online), but the curves show that binding of Ub to USP4CD has a slower off-rate than that of Ub to USP4–D1D2, and together with the $K_s$ value also suggests that it has a slower on-rate. As the $K_m$ is dependent on $K_s$ and the binding rate, the combination of slow kinetics and slightly lower affinity explains the differences in $K_m$ values. Apparently, the insert prevents rapid binding as well as rapid release of the Ub substrate, allowing competitive binding.

Finally we analysed whether the enzymatic activity is competitively inhibited by the addition of the insert in *trans*. We tested the enzymatic activity with varying inhibitor concentrations against a range of substrate concentrations (Fig 2F), and fitted the data against different inhibition models. We found that the data were best explained by competitive inhibition with $K_i = 47$ µM. Although this value is lower than expected on the basis of the binding data alone, it explains why the USP4CD is not completely inhibited in the continuous presence of the insert. It seems that...
additional conformational changes take place. One possibility is that the enzyme reaches a state after turnover that has lower affinity for the insert, and is therefore not as effectively inhibited.

**The Ubl domain is sufficient for inhibition**

The presence of a Ubl domain within the insert was predicted (supplementary Fig S2 online)\(^1\). To test whether the Ubl domain is sufficient for binding to the USP4 catalytic domain, we performed the SPR experiment with the purified Ubl domain (amino acids 483–571, Fig 1A) and found a \(K_d\) of \(1.36\) \(\mu\)M towards USP4–D1D2, which is similar to that for the complete insert (Fig 3C). This suggests that the Ubl domain is the functional part of the insert.

To test whether the Ubl domain can inhibit the DUB activity of USP4, we repeated the *in trans* inhibition assay with USP4–D1D2 in the presence of increasing amounts of the Ubl domain (supplementary Fig S6 online) and found that it provides inhibition equal to the insert. We therefore conclude that the Ubl domain is sufficient to inhibit the DUB activity of USP4, through competitive inhibition of Ub binding.

**Regulation by other USP enzymes**

As the Ubl domain seems to bind in the substrate Ub-binding site of USP4, we wondered whether other USP enzymes could also bind to the Ubl domain. We tested whether our Ubl domain containing insert could bind to the catalytic domain of USP39 and USP8, and found similarly high affinities as for USP4CD (Fig 4B,C).

Then, we analysed whether these DUBs could modulate USP4CD activity. We repeated the *in trans* Ub-AMC assay with USP4CD in the presence of the intrinsically inactive USP39CD or an inactive variant of the USP8 catalytic domain, USP8CD-mut (Fig 4A). For both USPs we observe a modest activation of USP4CD that was dependent on the presence of the Ubl-containing insert, as it does not increase the DUB activity of USP4–D1D2 in this manner.

![Figure 5. Model for Ubl domain inhibition on USP4. A) Structural model in which interaction of the Ubl domain with USP4CD inhibits the binding of Ub. B) Schematic model of the auto-inhibitory role of the Ubl domain in USP4. C) Other USP enzymes, such as USP39, may relieve the inhibition by binding to the Ubl domain. Ub, ubiquitin; Ubl, ubiquitin-like; USP, ubiquitin-specific protease.](/image)
Apparently, other USP enzymes can regulate USP4 activity by competing for binding to the Ubl domain. This effect could be larger when the USPs have further interactions. As USP39 forms a stable complex with USP4 in cells\textsuperscript{23,24}, it is a prime candidate for an activating role \textit{in vivo}.

Discussion

We show that the predicted Ubl domain within a large insert embedded in the USP4 catalytic domain partially inhibits DUB activity by competing with Ub for binding. Superposition of the crystal structure of USP4–D1D2 and any Ubl domain on USP7 in complex with Ub-aldehyde (PDB: 1NBF), respectively, shows that the Ubl domain would fit like a Ub molecule into the hand of USP4–D1D2 (Fig 5A), only requiring movements in the blocking loops and the zinc-finger ribbon. Hence, we propose a model in which the Ubl domain partly inhibits DUB activity through competitive inhibition by binding into the hand of USP4 and thus preventing Ub substrate binding (Fig 5B).

This function of an integrated Ubl domain is relatively new. The Ubl domains in proteasomal shuttle factors Rad23 and Dsk2, as well as in Parkin and USP14, function in recruitment of ubiquitinated proteins to the proteasome\textsuperscript{12,27}. Other Ubl domains regulate the enzymatic activities of immune-response inducible kinases such as IKKβ (a subunit of IκB kinase complex)\textsuperscript{28}, or as PB1 (Phox and Bem1) domains, have a role in the regulation of signal transduction in proteins such as P62, MEK5 and protein kinase C\textsuperscript{29,30}. However, all these Ubl-domain families have low sequence similarities, indicating that their functions are probably distinct between subfamilies.

The activity of USPs is regulated through an inactive conformation of the catalytic triad, as in USP7, or through a series of blocking loops or a blocking zinc-finger ribbon. USP4 seems to combine the blocking loops and zinc-finger ribbon with a further regulation through the Ubl domain.

Whether Ubl domains provide a common regulation mechanism for the DUB activity of USPs is an interesting question for future research. A second Ubl domain is found within USP4, at its N-terminus. A recent crystal structure (PDB: 3JYU, amino acids 139–226) shows that this Ubl domain interacts extensively with the adjacent DUSP (domain in USP) domain (amino acids 27–125). This region of the protein is primarily important for interaction with Sart3\textsuperscript{23} and hence might not have this function.

However, USP4 is not the only DUB with a Ubl fold within its catalytic domain. Sequence analysis by Zhu and co-workers identified an integrated Ubl fold within the catalytic domain of USPs 6, 11, 15, 19, 31, 32 and 43, embedded in a larger insert, like in USP4\textsuperscript{16}. In particular, USP11 and USP15 are closely related to USP4. This subgroup of USPs probably also regulates DUB activity through its Ubl domain.

The way in which Ubl-domain inhibition itself is regulated is an exciting question. One could imagine that further posttranslational modification by, for example, phosphorylation or acetylation would enable the release of the full activity of the DUB enzyme. In addition, we have shown that binding partners such as USP39, can activate USP4 function by binding to the Ubl domain (Fig 5C). Although the activation is modest (Fig 4A), this could be increased by further interactions, as observed in the spliceosome complex.

Whatever the mechanisms that are identified to regulate USP4 activation, it is clear that this type of internal regulation by a Ubl domain allows the creation of an extremely fast response element to external signals.

Material and Methods

Plasmids and cloning
cDNA for human USP4 and USP8 was a gift from Hidde Ploegh and cDNA for USP39 was a gift from R. Medema. USP4CD (aa 296-954) and any Ubl domain on USP7 in complex with Ub-aldehyde (PDB: 1NBF), respectively, shows that the Ubl domain would fit like a Ub molecule into the hand of USP4–D1D2 (Fig 5A), only requiring movements in the blocking loops and the zinc-finger ribbon. USP4CD (aa 296-490) of human USP4, USP39CD (aa 219-565) and USP8CD (aa 771-1118) were cloned using ligation independent cloning into pET-46 Ek/LIC vector (Novagen). The D2 fragment (aa 766-932) of USP4 was cloned into the pET-NKI b/3C (Luna-Vargas, in preparation). The fused USP4-D1D2 was created by inserting aa 353-359 of USP7 (SIKGKNN) between residues Leu479 and Leu777. The USP4 insert (aa 483-765) was generated by site-directed mutagenesis of the catalytic cysteine (C786A).
Protein preparation
Purification of E2-25K\(^{31}\), Ubc13/Mms2\(^{32}\) was as described. GST-tagged proteins were overexpressed in *Escherichia coli* strain Rosetta2(DE3)-T1R using IPTG (200µM) induction overnight at 15°C. Cells were lysed by microfluidizer into buffer A (50mM Hepes pH7.5, 150mM NaCl, 5mM β-mercaptoethanol, 1mM PMSF). The fusion protein was purified using glutathione sepharose resin, eluted, followed by removal of the GST-tag with 3C protease and size-exclusion using HiLoad 16/60 Superdex 200 (GE Healthcare). Peak fractions were concentrated to 10mg/ml in 25mM Hepes (pH7.5), 200mM NaCl and 5mM β-mercaptoethanol.

D1 and D2 co-expression, other USP4 variants, USP39CD and USP8CD-mut were overexpressed as above, with 200µM ZnCl\(_2\) during induction and lysed in buffer A supplemented with 1mM ZnCl\(_2\) and 10mM Imidazole. These His-tagged proteins were purified by a Co\(^{2+}\)-affinity (Talon resin) step. Upon Imidazole elution the His-tag was removed by TEV cleavage at 4°C overnight during dialysis in buffer B (25mM Hepes pH7.5, 150mM NaCl, 5mM β-mercaptoethanol). This was followed by POROS Q affinity chromatography and size-exclusion using HiLoad 16/60 Superdex 200 (GE Healthcare), where the protein eluted as a monomer. The peak fractions were concentrated to 5mg/ml in buffer B.

Limited Proteolysis and protein identification
Purified USP4CD (9mg/ml) was incubated with Thermolysin (0.8units) for 1.5hr at room temperature and subjected to size exclusion chromatography using Superdex75 16/60. Fractions containing USP4-D1 and –D2 were subject to LC-MS analysis. LC-MS measurements were performed on a system equipped with a Waters 2795 Separation Module (Alliance HT), Waters 2996 Photodiode Array Detector (190-750nm), Waters Alltime C18 (2.1x100mm, 3µm), Waters Symmetry300 T3, C4 (2.1x100mm, 3.5µm) and LCT\(_{TM}\) Orthogonal Acceleration Time of Flight Mass Spectrometer. Data processing was performed using Waters MassLynx Mass Spectrometry Software 4.1 (deconvolution with Maxent 1 function). N-terminal sequencing of USP4-D1 and –D2 were performed by AltaBioscience in Birmingham, England.

Crystallization and structure determination of the USP4-D1D2
Crystals were grown overnight in sitting-drops mixing 200nl USP4-D1D2 (~3.5mg/ml) with 200 nl 100mM Bis-Tris propane [pH8.5], 25mM Na\(_2\)SO\(_4\) and 18% PEG3350 (w/v) at 19°C. Crystals were cryoprotected in mother liquor with 25% ethylene glycol. The crystals belong to the space group P2\(_1\)2\(_1\)2\(_1\) with six molecules per asymmetric unit (supplementary Table S1). Diffraction data were collected at the ESRF (Grenoble, France) beamline ID14-2 and processed with MOSFLM\(^{33}\) and SCALA\(^{34}\). The structure was solved by molecular replacement with PHASER\(^{35}\) using USP8CD (PDB:2GFO) as search model. Iterative rebuilding and refinement were done with Coot\(^{36}\) and PHENIX\(^{37}\) and BUSTER\(^{38}\). The structure was validated with MOLPROBITY\(^{39}\) and WHAT-CHECK\(^{40}\) and structure figures were generated using PYMOL\(^{41}\). Cysteine residue 311 in all chains have been chemically modified by β-mercaptoethanol.

Ub-AMC assay
UbAMC assays were done in 50mM Hepes [pH7.5], 100mM NaCl, 5mM DTT, 0.05% Tween-20 and 1mM EDTA and reaction progress was monitored with a Fluostar Optima plate-reader (BMG Tech) by the increase in fluorescence emission at 460nm (λ\(_{ex}\) = 355nm) generated by Ub-AMC cleavage. Quantitative activity (triplicate) and *in trans* inhibition (duplicate) assays or USP modulation assays (triplicate) were performed using Ub-AMC with 10nM enzyme in 30 µl reaction volume in 384-well plates and preincubated for 15 min at 21°C, for inhibition assays with Ubl insert and with other DUBs for modulation assays. Initial velocities against Ub-AMC concentration were computed to derive steady-state kinetic parameters using GraphPad Prism5 (GraphPad Software Inc.). Non-linear fitting of four inhibition models was compared in GraphPad.

Di-ubiquitin assay
Di-Ub assays were performed in similar buffer as in UbAMC assays at 37°C in 75µl reaction volume. Aliquots (5µl) were stopped by addition of 4x SDS-sample loading buffer and subjected to SDS-PAGE analysis on a 4-12% coomassie stained gel (Invitrogen). K48 and K63 di-Ub substrates were produced and purified as described\(^{42}\). 75nM enzyme was
incubated with 3 µM di-Ub, subjected to SDS–polyacrylamide gel electrophoresis and image analysis, and quantitation was performed in duplicate with TINA 2.09 (Raytest Co.).

**Surface plasmon resonance**

SPR was performed on a Biacore T-100, with GST-Ub, GST-insert and GST-Ubl domain immobilized on anti-GST antibodies coupled to a CM5 chip. Data (duplicate) were processed using BiaEvaluation (GE Healthcare) and GraphPad Prism5. Quantitative binding analysis was done in duplicate at 25°C on a Biacore T-100 instrument (GE Healthcare). GST fused Ub, insert and Ubl domain were immobilized on α-GST antibodies lysine-coupled to a CM5 chip. USPs were injected in varying concentrations over the sensor chip at 30µl/min with a 120s association phase followed by a 10min dissociation phase. For the binding inhibition assay Ub was added in varying concentrations to USP4-D1D2. Standard double referencing data subtraction methods were used before and equilibrium curve fitting with BiaEvaluation (GE Healthcare) and GraphPad software (GraphPad Software Inc).

**Isothermal Titration Calorimetry**

ITC experiments were performed with the VP-ITC Micro Calorimeter (MicroCal, Inc.) at 25°C. Stock solutions of USP4CD, USP4-D1D2 and Ub were prepared by dialysis of the purified proteins against a buffer containing 25mM Hepes pH8.0, 150mM NaCl and 5mM β-mercaptoethanol at 4°C and were degassed before use. The sample cell (1.8ml) contained USP4-D1D2 (10µM) or USP4CD (20µM) which was titrated with 100µM Ub or 200µM Ub respectively using 16 injections. The injections after saturation were used to determine the background signal. Corrected data were analyzed using software supplied by the ITC manufacturer to calculate the dissociation constant K_d and fitted with a one to one binding model.

**Accession number**

Atomic coordinates and structure factors have been deposited to the Protein Data Bank with accession number: **2Y6E**

Supplementary information is available at EMBO reports online (http://www.emboreports.org).

**Acknowledgements**

We thank J. Lebbink for mass spectrometry, European Synchrotron Radiation Facility beamline scientists for assistance during X-ray data collection, A. Petrakis and P. Rucktooa for help with crystallographic analysis, R.G. Hibbert for help with SPR data analysis, and lab members for discussion and sharing of reagents. This study was supported by the European Union Network of Excellence project RUBICON and Dutch Cancer Society Project KWF 2008-4014.

**Conflict of interest**

The authors declare that they have no conflict of interest.
References


25. van Leuken, R.J., Luna-Vargas, M.P., Sixma, T.K., Wolhuis, R.M. & Medema, R.H. Usp39 is essential


41. Delano, W.L. The Pymol molecular graphics system. (Delano Scientific, San Carlos, CA, USA, 2002).

### Table of contents Supplementary information

| Table S1 | Data collection and refinement statistics. | 103 |
| Figure S1 | Identification of USP4-D1D2. | 104 |
| Figure S2 | Structure based multiple sequence alignment. | 104 |
| Figure S3 | Overview and superposition of USP catalytic domain structures with USP4-D1D2. | 106 |
| Figure S4 | Deubiquitinating assay with K48 di-Ub as substrate. | 107 |
| Figure S5 | In trans inhibition of USP4-D1D2 DUB activity. | 107 |
| Figure S6 | The molecular crowding of high concentrations of SUMO or BSA (100µM) does not have an effect on USP4-D1D2 DUB activity. | 108 |
| Figure S7 | Kinetic comparison of USP4-D1D2 and USP4CD binding to Ub, Ubl and insert on SPR. | 108 |
Table SI  Data collection and refinement statistics.

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.993</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Unit cell (Å)</td>
<td>110.5, 151.0, 178.7</td>
</tr>
<tr>
<td>Molecules per asymmetric unit</td>
<td>6</td>
</tr>
<tr>
<td>Resolution (Å)(^a)</td>
<td>47.0 - 2.4 (2.53 - 2.4)</td>
</tr>
<tr>
<td>R(_{merge}) (%)</td>
<td>8.8 (66.1)</td>
</tr>
<tr>
<td>(&lt;I/\sigma(I)&gt;)</td>
<td>11.0 (1.1)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.9 (74.0)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.4 (2.4)</td>
</tr>
</tbody>
</table>

Reefinement

<table>
<thead>
<tr>
<th>R(<em>{work}/R</em>{free}) (%)</th>
<th>17.5/21.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reflections</td>
<td>111078</td>
</tr>
<tr>
<td>Number of protein atoms</td>
<td>15775</td>
</tr>
<tr>
<td>Number of zinc ions</td>
<td>6</td>
</tr>
<tr>
<td>Number of waters</td>
<td>867</td>
</tr>
</tbody>
</table>

RMSD from ideal geometry

| Bond lengths (Å)           | 0.009     |
| Bond angles (°)            | 1.01      |
| Ramachandran statistic\(^b\) | 1849 / 66 / 2 |

\(^a\) Numbers in parentheses are for the highest-resolution shell

\(^b\) Calculated using Molprobity
Supplemental Figures

A) Limited proteolysis analysis of thermolysin cleavage on USP4CD at 37°C. Samples at different time-points were taken and analyzed on a SDS-PAGE gel.

B) Proteolytic sample of USP4CD was subjected to a size exclusion chromatography and fraction samples were analyzed on a SDS-PAGE gel.

C) Mass Spectrometry analysis and N-terminal sequencing determined the identity of the two fragments, D1 and D2.

Supplemental Figure S1

Identification of USP4-D1D2. A) Limited proteolysis analysis of thermolysin cleavage on USP4CD at 37°C. Samples at different time-points were taken and analyzed on a SDS-PAGE gel. B) Proteolytic sample of USP4CD was subjected to a size exclusion chromatography and fraction samples were analyzed on a SDS-PAGE gel. C) Mass Spectrometry analysis and N-terminal sequencing determined the identity of the two fragments, D1 and D2.

Calc. mass:
D1) 19687.46 Da  
D2) 22230.30 Da

Theor. mass:
D1) 19685.54 Da  
D2) 22227.46 Da

N-terminal sequencing:
D1) Gly - Met - His - Ile - Gln
D2) Leu - Gln - Pro - Gln - Lys
Supplemental Figure S2
Structure based multiple sequence alignment. Secondary structure elements are colored and labeled according to structure of USP4-D1D2 in Figure 1. The internal Ubl is depicted as a yellow bar and the two black arrows indicate where the protease thermolysin cleaved in the catalytic domain of USP4. The catalytic triad residues are indicated with an asterisk. The four black triangles indicate the positions of the Cys residues coordinating the zinc ion.
Supplemental Figure S3
Overview and superposition of USP catalytic domain structures with USP4-D1D2. Comparison between catalytic domains depicted in cartoon representation of USP4-D1D2 (red-cyan), Ubp8 (brown, PDB: 3MM9), CYLD (orange, PDB: 2VHF), USP21 (marine blue, PDB: 3I3T), USP2 (yellow, PDB: 2HD5), USP8 (purple, PDB:2GFO), USP14 (green, PDB: 2AYO) and USP7 (light pink, PDB:1NB8). The structures depicted in ribbon representation were superposed in Coot (RMSD of 2.1Å over 323 residues).
Supplemental Figure S4
Deubiquitinating assay with K48 di-Ub as substrate. A-C) The full-length USP4 catalytic domain (A) is much less active than USP4-D1D2 (B) or USP4-fusion (C) in a deubiquitinating assay using K48 di-Ub as substrate on coomassie-stained SDS-PAGE gels. D) Quantification of mono-Ub in K48 di-Ub cleavage assays. The intensity of the mono Ub band is plotted against time.

Supplemental Figure S5
In trans inhibition of USP4-D1D2 DUB activity. A) The inhibitory effect of the insert is observed in a Ub-AMC assay with increasing amounts of insert (5, 10, 25, 50 and 75µM). B) The Ubl domain (5, 10, 25, 50 and 100µM) is sufficient to show this in trans inhibition.
Supplemental Figure S6
The molecular crowding of high concentrations of SUMO or BSA (100µM) does not have an effect on USP4-D1D2 DUB activity.

Supplemental Figure S7
Kinetic comparison of USP4-D1D2 and USP4CD binding to Ub, Ubl and insert on SPR. Binding curve of 0.8 µM of USP4-D1D2 and of 1 µM of USP4CD were normalized for maximum binding in order to compare off-rates.
General Discussion
General Discussion

The USP family: small and multifunctional

More and more functional roles of USPs in important biological cellular processes are being discovered. Also the role they play in tumorigenesis is recognized in the ubiquitin field. Therefore members of the large USP family are being actively pursued as drug targets.

In chapter 2 a genomic and functional overview is given on the large family of deubiquitinating enzymes encoded in the human genome. The analysis of the ENSEMBL human genome database revealed 95 putative DUBs of which 79 are expressed in cells and display Ub/Ubl protease activity. Compared to the number of E3 ubiquitin ligases, the DUBs are remarkably outnumbered. One possible explanation is that not all DUBs have been identified yet or their associated cofactors that may determine specificity. Another explanation is that a DUB may have different targets. For instance USP7 seems to be able to deubiquitinate several different targets such as p53, Mdm2, PTEN, FOXO4 and H2B. As already mentioned in chapter 1, USP4 also has different targets e.g. the U4 spliceosome component Prp3, TAK1 and more recently the E3 Ub ligase ARF-BP1. More USPs seem to have more than one target, like USP1 (targets FANCD2 and PCNA)9,10, USP2 (fatty acid synthase and Mdm2)11,12, and CYLD (targets NEMO, TRAF-2 and TRAF-6)13-15. Another explanation for the excess of E3 Ub ligases could be that only a fraction of the targets that are ubiquitinated are regulated by specific USPs. It is possible that only proteins that demand extremely tight regulation, such as p53 and H2A/B, require additional regulation by deubiquitination. Indeed p53 and H2A/B seem to have more than one USP for their deubiquitination, USP2, USP7, USP10 and USP3, USP16, USP21, USP22 respectively. Undoubtedly, future studies will uncover more about the functional roles of members of the USP family and why the E3 Ub ligases outnumber the DUBs.

High-throughput equals low-output

In order to understand and explain the molecular and biochemical function of USPs one should turn to structural molecular biology. Once the three-dimensional structure of a USP is determined it could help decipher basic principles of protein structure and assembly, mechanisms of biochemical reactions and details of macromolecular interactions. As stated in chapter 1 only the crystal structures of the catalytic domain of several USP members and more recently of USP4 (chapter 5) have been solved. Despite the low sequence similarity the overall structure of the catalytic domain is highly conserved. The structures also reveal that the catalytic activity of USPs is regulated by substrate- or scaffold-induced conformational changes. It will be very interesting to determine full-length USPs and see how their additional domains interact with the catalytic domain. In the process of determining the three-dimensional protein structures it is important to produce a large number of soluble recombinant protein variants. In chapter 3 a set of protein expression vectors for ligation-independent cloning is described and their use for protein expression in E.coli on 35 different members of the USP family. Out of 145 different expression constructs, 38 soluble recombinant proteins for 21 different USPs were obtained. Looking more carefully, the level of protein solubility amongst the 21 USPs differ substantially. An explanation for this solubility difference could be the use of high-throughput methods. The protein expression and solubility screening was done in 96-well blocks and working with small volumes one can easily make a mistake in the different steps such as cell lysis, affinity binding and elution of recombinant protein. The use of robotics in automating these different steps can help substantially in identifying the correct construct that gives high protein expression and solubility.

Most of the constructs that give soluble protein for the 21 different USPs are either representing the catalytic domain only or with variations at the N- and C-terminal ends of the catalytic domain. The expression of the full-length protein of only two USPs (USP16 (823aa) and USP46 (366aa)) was possible in E.coli. Interestingly, the USP12 construct which has high sequence similarity with USP46 did not give soluble protein in the E.coli expression system. However expressing the USP12 construct in the Sf9 insect cell expression resulted in a high yield of soluble full-length protein (chapter 4). It looks like the Sf9 insect cell expression system might be a more...
suitable system for expressing the full-length USP protein. Indeed when testing the pFastBac- NKILIC vector in Sf9 insect cells, the construct for full-length USP7 gave high yield of soluble protein.

Another way of improving the expression and solubility of the USPs is the use of synthetic genes containing optimized codons. In E.coli the protein expression is maximized by using codons corresponding to tRNAs that retain amino acid charging during starvation. The expression and solubility of USP7 in E.coli for instance has been greatly improved by using synthetic genes (chapter 4).

Although the development and the use of high-throughput methods did not achieve the desired results in getting a large number of USP constructs giving soluble protein in large amounts, the use of these high-throughput methods are now well established and the ligation-independent cloning using the pET-NKI vectors is currently the standard method in our lab. Furthermore, the number of USPs that we finally could obtain was used for the characterization in chapter 4.

USPs look similar, but behave differently

Even though many studies show the vital role of a number of USPs in important cellular pathways and the implication in tumorigenesis, not much is known about the proteases themselves in terms of enzyme kinetics and preference for ubiquitin-chain type. We have tried to address these topics by characterizing a set of twelve USPs using synthetic tools including all seven di-ubiquitin topoisomers (chapter 4).

The analysis of the enzyme kinetics showed that there are large differences in DUB activity amongst the twelve USPs. Based on their $K_m$ and $k_{cat}$ values, these USPs can be grouped in three classes: the inactive, the intermediate and the very active group. Future studies will show whether these three classes still hold true after characterizing the enzyme kinetics for the rest of the USP family. It seems that the USP family doesn’t show any real preference for a certain ubiquitin-chain type, but rather show a modest differential activity towards the seven di-ubiquitin topoisomers which was variable between USPs.

Many USPs have additional domains and internal insertions within the catalytic domain and some USPs are known to have external modulators that regulate the DUB activity. As shown for USP7 and USP16, both show increased DUB activity in the presence of their additional domains (the HUBL and ZnF-UBP domain, respectively). The ZnF-UBP domain of USP16 enhances the DUB activity by increasing the $k_{cat}$. This activation of DUB activity has been previously shown for USP5 as demonstrated by Reyes-Turcu et al. They show that binding of free Ub to the ZnF-UBP domain enhances the USP5 DUB activity in vitro. Besides USP16 and USP5, several other USPs (USP3, USP44, USP45 and USP49) have a similar ZnF-UBP domain. It is possible that enhancement of USP activity following free ubiquitin binding would be a general regulatory mechanism. The activity of USP7 is modulated by its HUBL domain which affects both the $K_m$ and $k_{cat}$. Despite a clear difference in DUB activity, no change in ubiquitin-chain type preference was seen between the catalytic domain and full-length of USP7 and USP16. Interestingly, our data show that the preference of USP7 for di-ubiquitin topoisomers can be attributed to the binding affinity ($K_m$) for the substrate.

A similar enhancement of DUB activity by increasing the $k_{cat}$ is seen for USP1, USP12 and USP46 in presence with their external modulator, the WD40-repeat containing UAF1. Also for USP7 the $k_{cat}$ is increased by adding its external modulator GMPS. Similar to the HUBL and ZnF-UBP domain, both external modulators do not seem to affect the preference for ubiquitin-chain type.

Because of the important cellular roles of several USPs, their DUB activity must be tightly regulated. Several studies have revealed that kinases play an important role in USP control and a number of USPs, such as USP7, USP16 and USP44, are activated by phosphorylation. It seems that the DUB activity of USPs is further regulated by intramolecular modulating domains (HUBL and ZnF-UBP) and by intermolecular modulators (GMPS and UAF1). The identification of new intra- and intermolecular modulators for other USPs will be interesting and challenging for future studies and will unveil whether they also can affect the preference for ubiquitin-chain type.

New role for the Ubl domain

The knowledge about functions for the Ubl domains in USP family members is limited. Only one example has been described which is the Ubl
domain located at the N-terminus of the catalytic domain of USP14. It functions as an anchor for recruitment to the proteasome where USP14 is activated\(^2^3\). Very recently Faesen and co-workers show the crystal structure of five Ubl domains located C-terminal of the catalytic domain of USP7\(^2^4\). Here the first three Ubl domains seem to be required as a docking station for the intermolecular modulator GMPS, while the last two Ubl domains apparently are responsible for activating USP7. This activation is achieved by interacting with the Cys-loop and position the catalytic triad in a catalytic competent configuration. Finally in chapter 5, a new role for the Ubl domain in USPs is described. The Ubl domain of USP4 located within the catalytic domain plays a role as the internal inhibitor of the DUB activity of USP4. The Ubl domain inhibits USP4 activity by binding to the catalytic domain with similar affinity as for ubiquitin and thereby preventing ubiquitin substrate binding.

It seems that the Ubl domain in USPs can have different functions: recruiting, activating and inhibiting DUB activity. These functions might be a common feature in USPs as similar Ubl domains have been identified in a large number of USPs\(^2^5\). Interestingly, the Ubl domains which recruits, activates and inhibits are each located differently with respect to the catalytic domain. The recruiting Ubl domain of USP14, the activating Ubl domain of HAUSP and the inhibiting Ubl domain of USP4 are located respectively N-terminal, C-terminal and internal of the catalytic domain. Whether the location of the Ubl domain in respect of the catalytic domain plays an important role in Ubl function is an interesting question.

Interestingly, the inhibitory role of the Ubl domain within the USP4 catalytic domain can be blocked by USP39 through binding to the Ubl domain, thus activating USP4 function. The interaction between USPs opens up a new door for USP regulation. Looking at the domain architecture of the USP family containing ubiquitin binding domains and Ubl domains, it is very likely that other USPs can interact with each other and possibly alter USP function. Indeed recently, Maertens et al show the interaction between USP7 and USP11 and how they regulate the ubiquitination status of several components of the Polycomb repressive complex \(^1^2^6^, ^2^7\). How they interact remains unclear, but as USP11 looks very similar in domain architecture as USP4 with Ubl domains, it is possible that USP7 binds to USP11 through one of its Ubl domains. Or USP11 binds to one of the HUBL domains of USP7 and thereby altering each others function. Future research into Ubl domains should be conducted to investigate whether the Ubl domains predicted in other USPs have similar or might display new functions.

**Concluding remarks**

In the last decade the importance of DUBs as a key regulatory step in ubiquitin-dependent pathways and as an important factor in tumorigenesis, have been increasingly recognized. Although many exciting studies on DUBs are ongoing, there is still a long way to go before we fully characterize this intriguing protease family. The largest subfamily of DUBs, the USP family, has been the focus of this thesis. We have seen that USPs, even in small number compared to the E3 ubiquitin ligases, are multifunctional in that they have more than one target. We have also seen that a subset of USPs displays a large variation in DUB activity with no real preference for a ubiquitin-chain type. Furthermore, the USP activity not only is regulated on the level of phosphorylation, but it also involves inter- and intramolecular modulators. Finally, a new inhibitory role for the integrated Ubl domain in the catalytic domain of USP4 has been described. Hopefully, the experiments and data shown in this thesis will contribute to the further understanding of the USP family.
References


25. Zhu, X., Menard, R. & Sulea, T. High incidence of ubiquitin-like do-


Addendum

Summary
Samenvatting
List of abbreviations
Curriculum Vitae
PhD portfolio
List of publications
Acknowledgements
SUMMARY

Since the discovery of ubiquitin (Ub), it is increasingly apparent that Ub mediated events are critical in cell proliferation. In the last several decades much attention is placed on the ubiquitination pathway and recently the role of deubiquitinating enzymes (DUBs) in the reverse pathway is being recognized as important regulators of these processes. There is also a growing recognition of DUBs that are mutated in human cancers making them interesting drug targets. To better understand these DUBs, this thesis focuses on the structural and functional aspects of the largest subclass of DUBs, the ubiquitin-specific protease (USP) family.

Chapter 2 gives a genomic and functional overview of the DUB family and describes the five different subclasses. Describing examples of the cellular roles of USPs in different pathways with known protein substrates, it shows that eventhough the USP family is relatively small, the USP family member is multifunctional by having more than one substrate. It also shows how some important substrates such as p53 and H2A/B requires additional regulation by having more than one USP for their deubiquitination.

In order to better understand the molecular and biochemical roles of USPs, soluble protein is required for both crystallographic and biochemical studies. To obtain soluble protein, one must produce and screen a large number of DNA constructs. In chapter 3 the development and the use of the pET-NKI-LIC vectors as high-throughput methods to obtain DNA constructs, which results in high protein expression and improves protein solubility, is described. Ligation-independent cloning (LIC) using the pET-NKI-LIC vectors is currently the standard method in our lab and the implementation of LIC for a large number of USPs, resulted in a number of soluble USP proteins.

Chapter 4 shows the enzymatic characterization of the obtained soluble USPs using synthetic substrates including all seven lysine-linked di-ubiquitins. The analysis shows that the USPs behave differently in terms of DUB activity and that based on their kinetic behaviour they can be grouped into three classes: the inactive, the intermediate and the very active group. Furthermore, we provide the first comprehensive analysis comparing the Ub chain preference and show that USPs display a modest activity towards the seven di-ubiquitin topoisomers which was variable amongst the USPs and that this Ub chain-type preference in the case for USP7 can be attributed to the binding affinity ($K_m$) for its substrate. Finally, the existence of intermolecular modulating domains (HUBL and ZnF-UBP) and intramolecular modulators (GMPS and UAF1) help regulate the DUB activity of USPs by mainly increasing the catalytic turnover ($k_{cat}$).

In chapter 5 the three-dimensional crystal structure of minimal catalytic domain USP4-D1D2 is shown and a new role for the Ubl domain embedded within the catalytic domain of USP4 is described. The crystal structure of the USP4 catalytic domain has the conserved USP-like fold with its typical Ub binding site. Our findings show that the integrated Ubl domain acts as an intermolecular modulating domain that inhibits the DUB activity of USP4. This inhibitory function of the integrated Ubl domain is induced by binding to the catalytic domain of USP4 with similar affinity as for Ub and thereby preventing Ub substrate binding. Interestingly, a binding partner of USP4, USP39 is able to bind to the same Ubl domain and relieve the inhibition.

Together the work presented in this thesis illustrates how complex and diverse the function and the regulation are of this relative small USP family. Hopefully this work gives interesting leads for future studies and helps contribute into the understanding of the intriguing DUB family, the USPs.
SAMENVATTING

Sinds de ontdekking van ubiquitine (Ub), wordt het steeds duidelijker dat Ub gemedi- eerde gebeurtenissen belangrijk zijn in de cel proliferatie. In de laatste decennia is er veel aandacht besteed aan de ubiquitinering route en onlangs is de rol van deubiquitineringsen- zymen (DUBs) in de omgekeerde route er- kend als belangrijke toezichthouders van deze processen. Er is ook een toenemende erken- ning van DUBs die gemuteerd zijn in mense- lijke kankers, waardoor ze interessant zijn als drug targets. Om deze DUBs beter te begrij- pen, concentreert dit proefschrift zich op de structurele and functionele aspecten van de grootste subgroep van DUBs, de ubiquitine- specificieke protease (USP) familie.

Hoofdstuk 2 geeft een genomisch en functioneel overzicht van de DUB familie en beschrijft de vijf verschillende subgroepen. Door voorbeelden te beschrijven van cellulaire rollen van USPs in de verschillende routes met bekende eiwit substraten, laat hoofdstuk 2 zien dat de USP familie leden, ondanks het kleine aantal, multifunctioneel zijn door meer dan één substraat te hebben. Verder wordt ook beschreven dat sommige belangrijke substra- ten zoals p53 en H2A/B extra regulatie verei- sen door meer dan één USP voor hun deubi- quitinering te hebben.

Om beter inzicht te krijgen in de moleculaire en biochemische rollen van USPs, is oplosbaar eiwit nodig voor zowel de krist- tallografische en biochemische studies. Om oplosbaar eiwit te verkrijgen, moet men een groot aantal DNA constructen produceren en screenen. In hoofdstuk 3 worden de ontwik- keling en het gebruik van de pET-NKI-LIC vectoren beschreven. Deze vectoren worden gebruikt als hoge-doorstroom methodes om DNA constructen te verkrijgen die kunnen leiden tot hoger eiwit expressie en betere eiwit oplosbaarheid. Ligatie-onafhankelijke klone- ring (LIC) door middel van de pET-NKI-LIC vectoren is tegenwoordig de standaard metho- de in ons lab en de implementatie van LIC op een groot aantal USPs resulteerde in een aantal oplosbare USP eiwitten.

Hoofdstuk 4 laat de enzymatische karakterisatie van de verkregen oplosbare USPs zien door gebruik te maken van de syntheti- sche substraten inclusief alle zeven lysine-ge- koppelde di-ubiquitines. De analyse laat zien dat de USPs zich anders gedragen in termen van DUB activiteit en dat gebaseerd op hun kinetic gedrag ze in drie groepen verdeeld kunnen worden: de inactieve, de gemiddelde en de zeer actieve groep. Bovendien geven we de eerste uitgebreide analyse waarin de voor- keur voor een Ub keten wordt vergeleken. We laten ook zien dat de USPs een gemiddelde activiteit vertonen tegenover de zeven di-ubi- quitine topoisomeren. De activiteit is variabel tussen de USPs en de Ub keten-type voorkeur in het geval van USP7 is toe te schrijven aan de bindingsaffiniteit (K_M) voor zijn substraat. Tenslotte laat hoofdstuk 4 zien dat het bestaan van intermoleculaire modulerende domeinen (HUBL en ZnF-UBP) en intramoleculaire modulatoren (GMPS en UAF1) in USPs helpen aan de regulatie van de DUB activiteit door voornamelijk de katalytische omzetting (k_cat) toe te laten nemen.

In hoofdstuk 5 wordt de driedi- mensionale kristal structuur van het minimale katalytische domein USP4-D1D2 getoond en wordt een nieuwe rol voor het Ubl domein dat geïntegreerd is in het katalytische domein van USP4, beschreven. De kristal structuur van het USP4 katalytisch domein heeft het gecon- serveerde UPS-lijkende vouwing met zijn typ- ische Ub bindingsplaats. Onze bevindingen laten zien dat het geïntegreerd Ubl domein als een intermoleculaire modulerende domein functioneert door de DUB activiteit van USP4 te inhiberen. Deze inhibitie functie van het geïntegreerde Ubl domein wordt tot stand ge- bracht door in het katalytische domein van USP4 te binden met eenzelfde affiniteit als voor Ub en hierdoor wordt Ub substraat bind- ing voorkomen. Verder is het interessant dat USP39, een bindingspartner van USP4, aan hetzelfde Ubl domein kan binden en hiermee de inhibitie van het Ubl domein opheft.

Tegenwoordig illustreert het werk gepre- senteerd in dit proefschrift, hoe complex en divers de functie en de regulatie van deze re-
latieve kleine USP familie zijn. Hopelijk geeft dit werk interessante aanknopingspunten voor toekomstige onderzoeken en draagt het bij in het begrijpen van deze intrigerende DUB familie, de USPs.
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>Ub</td>
<td>ubiquitin</td>
</tr>
<tr>
<td>Ubl</td>
<td>ubiquitin-like</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>E1</td>
<td>ubiquitin activating enzyme</td>
</tr>
<tr>
<td>E2</td>
<td>ubiquitin conjugating enzyme</td>
</tr>
<tr>
<td>E3</td>
<td>ubiquitin ligase</td>
</tr>
<tr>
<td>UBA1</td>
<td>ubiquitin-like modifier activating enzyme 1</td>
</tr>
<tr>
<td>MgATP</td>
<td>magnesium adenosine triphosphate</td>
</tr>
<tr>
<td>RING</td>
<td>Really Interesting New Gene</td>
</tr>
<tr>
<td>HECT</td>
<td>Homologous with E6-associated protein C-Terminus</td>
</tr>
<tr>
<td>URM1</td>
<td>ubiquitin related modifier 1</td>
</tr>
<tr>
<td>ATG12</td>
<td>autophagy related protein</td>
</tr>
<tr>
<td>Nedd8</td>
<td>neural precursor cell expressed, developmentally down-regulated 8</td>
</tr>
<tr>
<td>SUMO</td>
<td>small ubiquitin-like modifier</td>
</tr>
<tr>
<td>FAT10</td>
<td>ubiquitin like protein</td>
</tr>
<tr>
<td>ISG15</td>
<td>interferon-stimulated ubiquitin like protein</td>
</tr>
<tr>
<td>LC3</td>
<td>ubiquitin like protein</td>
</tr>
<tr>
<td>Cys</td>
<td>cysteine</td>
</tr>
<tr>
<td>Ser</td>
<td>serine</td>
</tr>
<tr>
<td>Thr</td>
<td>threonine</td>
</tr>
<tr>
<td>UBD</td>
<td>ubiquitin-binding domain</td>
</tr>
<tr>
<td>ZnF-UBP</td>
<td>zinc finger ubiquitin-specific protease</td>
</tr>
<tr>
<td>UIM</td>
<td>ubiquitin-interacting motif</td>
</tr>
<tr>
<td>UBA</td>
<td>ubiquitin-associated</td>
</tr>
<tr>
<td>UCH</td>
<td>ubiquitin C-terminal hydrolase</td>
</tr>
<tr>
<td>USP</td>
<td>ubiquitin-specific protease</td>
</tr>
<tr>
<td>OTU</td>
<td>ovarian tumor</td>
</tr>
<tr>
<td>JAMM</td>
<td>JAB1/MPN/MOV34</td>
</tr>
<tr>
<td>PCNA</td>
<td>proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>K</td>
<td>lysine</td>
</tr>
<tr>
<td>ERAD</td>
<td>endoplasmic-reticulum-associated-degradation</td>
</tr>
<tr>
<td>UFD</td>
<td>ubiquitin fusion degradation</td>
</tr>
<tr>
<td>BRCA1</td>
<td>breast cancer 1</td>
</tr>
<tr>
<td>BARD1</td>
<td>BRCA1-associated RING domain protein 1</td>
</tr>
<tr>
<td>Ufd2</td>
<td>ubiquitin chain elongation enzyme</td>
</tr>
<tr>
<td>LUBAC</td>
<td>linear ubiquitin chain assembly complex</td>
</tr>
<tr>
<td>NF-κB</td>
<td>nuclear factor kappa-light-chain-enhancer of activated B cells</td>
</tr>
<tr>
<td>Ile</td>
<td>isoleucine</td>
</tr>
<tr>
<td>Kd</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>TRAF</td>
<td>TNF receptor associated factor</td>
</tr>
<tr>
<td>RAP80</td>
<td>receptor associated protein</td>
</tr>
<tr>
<td>NEMO</td>
<td>NF-κB essential modulator</td>
</tr>
<tr>
<td>ABIN</td>
<td>A20-binding inhibitor of NF-κB</td>
</tr>
<tr>
<td>UBAN</td>
<td>ubiquitin-binding in ABIN and NEMO</td>
</tr>
<tr>
<td>IKK</td>
<td>IκB kinase</td>
</tr>
<tr>
<td>TNF-α</td>
<td>tumor necrosis factor-α</td>
</tr>
<tr>
<td>SAGA</td>
<td>Spt-Ada-Gcn5 acetyltransferase</td>
</tr>
<tr>
<td>Akt</td>
<td>kinase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IkB</td>
<td>inhibitor of NF-κB</td>
</tr>
<tr>
<td>TNF-α</td>
<td>tumor necrosis factor α</td>
</tr>
<tr>
<td>IL-1</td>
<td>interleukin 1</td>
</tr>
<tr>
<td>IKK</td>
<td>IkB kinase</td>
</tr>
<tr>
<td>SCF</td>
<td>Skp, cullin, F-box containing</td>
</tr>
<tr>
<td>RIP</td>
<td>ribosome inactivating protein</td>
</tr>
<tr>
<td>HR</td>
<td>homologous recombination</td>
</tr>
<tr>
<td>UBE2T</td>
<td>ubiquitin conjugating enzyme</td>
</tr>
<tr>
<td>HERC2</td>
<td>HECT domain and RCC1-like domain containing protein 2</td>
</tr>
<tr>
<td>RNF8</td>
<td>Ring finger protein 8</td>
</tr>
<tr>
<td>RNF168</td>
<td>Ring finger protein 168</td>
</tr>
<tr>
<td>RIDDLE</td>
<td>radiosensitivity, immunodeficiency, dysmorphic features and learning difficulties</td>
</tr>
<tr>
<td>Ubc13</td>
<td>ubiquitin conjugating enzyme 13</td>
</tr>
<tr>
<td>53BP1</td>
<td>p53 binding protein 1</td>
</tr>
<tr>
<td>UPS</td>
<td>Ubiquitin-proteasome system</td>
</tr>
<tr>
<td>CYLD</td>
<td>cylindromatosis (turban tumor syndrome)</td>
</tr>
<tr>
<td>Wnt</td>
<td>wingless int</td>
</tr>
<tr>
<td>TGFβ</td>
<td>transforming growth factor beta</td>
</tr>
<tr>
<td>BCL-2</td>
<td>B-cell lymphoma 2</td>
</tr>
<tr>
<td>MCL1</td>
<td>myeloid cell leukemia 1</td>
</tr>
<tr>
<td>Mdm2</td>
<td>murine double mutant 2</td>
</tr>
<tr>
<td>COP1</td>
<td>caspase recruitment domain-containing protein 1</td>
</tr>
<tr>
<td>Pirh2</td>
<td>p53-induced ring-h2 domain</td>
</tr>
<tr>
<td>ARF-BP1</td>
<td>alternate reading frame-binding protein 1</td>
</tr>
<tr>
<td>HAUSP</td>
<td>herpesvirus associated ubiquitin-specific protease</td>
</tr>
<tr>
<td>PTEN</td>
<td>phosphatase and tensin homolog</td>
</tr>
<tr>
<td>Asp</td>
<td>aspartate</td>
</tr>
<tr>
<td>Asn</td>
<td>asparagines</td>
</tr>
<tr>
<td>DUSP</td>
<td>domain in USPs</td>
</tr>
<tr>
<td>Rad23</td>
<td>UV excision repair protein 23</td>
</tr>
<tr>
<td>Dsk2</td>
<td>ubl/ub associated containing protein</td>
</tr>
<tr>
<td>SG</td>
<td>structural genomics</td>
</tr>
<tr>
<td>RIKEN</td>
<td>Rikagaku Kenkyusho (institute of physical and chemical research)</td>
</tr>
<tr>
<td>PSI</td>
<td>protein structure initiative</td>
</tr>
<tr>
<td>SGC</td>
<td>structural genomics consortium</td>
</tr>
<tr>
<td>SPINE</td>
<td>structural proteomics in Europe</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
</tbody>
</table>
CURRICULUM VITAE

Mark Luna-Vargas was born on the 17th of March, 1979 in Amsterdam. In 1997 he finished his high-school and received his VWO-gymnasium diploma at the Katholieke College Amsterdam. In the same year he started his studies at the Chemistry faculty of the University of Amsterdam. After receiving his Master’s degree in Biochemistry in 2003, he started working as a technician in the group of Prof. Dr. Titia Sixma. During the period 2004-2011 he worked on this thesis under the supervision of Prof. Dr. Titia Sixma at the Netherlands Cancer Institute. In January 2012 he will move to New York where he will join the group of Prof. Dr. Ming-Ming Zhou at Mount Sinai School of Medicine. There he will conduct research into the molecular interactions and regulation of histone-directed chromatin biology in the cell nucleus.
PHD PORTFOLIO

SUMMARY OF PHD TRAINING AND TEACHING

Name PhD student: Mark Patrick Alexander Luna-Vargas
Erasmus Department: Genetics
Research School: Onderzoeksschool Oncologie Amsterdam
PhD period: 2004-2011
Promotor/Supervisor: Prof. dr. Titia K. Sixma

1. PhD training
   General courses
   Radiation course 2004
   Scientific writing and presenting in English 2009

   Specific courses
   Functional Genomics, NKI, Amsterdam, The Netherlands 2005
   Protein Structure and Function, NKI, Amsterdam, The Netherlands 2006
   Bio Crystallography Practical Course, ITQB, Oeiras, Portugal 2008

   Seminars & Workshops
   Bioinformatics, IGBMC, Strasbourg, France 2005
   Seminar Staff-Meeting, NKI, Amsterdam, The Netherlands 2010

   (Inter)national conferences
   2004-2009 Annual meeting Dutch studygroups of NWO, Lunteren, The Netherlands (poster presentation)
   2005-2006 OOA PhD student retreat, Texel, The Netherlands
   2005&2007 The Ubiquitin Family conference, Cold Spring Harbor, NY, USA (poster presentation)
   2006-2011 Rubicon Network of Excellence, ubiquitin conference, Florence, Rome (Italy), Barcelona (Spain) (poster presentation) and Malta (oral presentation)

2. Teaching
   2007-2008 Supervising master’s thesis research project of Master student
LIST OF PUBLICATIONS


* These authors contributed equally to this work


Van Leuken RJ, Luna-Vargas MPA, Sixma TK, Wolthuis RM, Medema RH. USP39 is essential for mitotic spindle checkpoint integrity and controls mRNA-levels of AuroraB. Cell Cycle 2008; 17:2710-9


Review

ACKNOWLEDGEMENTS

Finally, I’ve made it. The last chapter of my thesis: the acknowledgements. It has been a (very) long journey/adventure/struggle, but one that I would not have enjoyed/endured/experienced without the help/guidance/support from certain people I would like to thank.

First of all my biggest thanks for Titia, who first hired me as a technician and after one year gave me the opportunity to start a PhD in her group. Throughout my PhD I have seen the small crystallography group from H2 grow into a big and flourishing department under your guidance and supervision. I admire your endless enthusiasm and your vast knowledge and your attention for details. I have learned a lot from you and I know it has not been easy for both of us, but thank you for your listening ear and support! I wish and your family all the best for the future!

Pim, the ´engine´ of the department, sitting next to you starting in H2 and ending in B8 has been a-never-dull-moment experience I will never forget! I will miss our laughs and talks. Thanks for being there and helping me prepping all those damned difficult USPs.

Dancing with Alex. I never thought we would have teamed up to tackle the USP project, but I’m glad we did. I admire how focused you can be and how you manage everything with the dancing, the science and having a family. I wish you all the best in Germany!!

Tassos, you and your group I have also seen evolve throughout the years. Thanks for all your help with the crystallography. From being a temperamental and thundering Greek god to a more relaxed and diplomatic group-leader, I enjoyed our discussions about (Greek) football and your passion for food. Maybe one day we will meet again in the Arena where Ajax and Panathinaikos will play each other again in the Champions League. Of course this time Ajax will win :)

I always tell people around me that working in the lab is like playing in the ‘speeltuin’. Not only the environment contributed to that feeling but of course the people I had the pleasure of working and discussing with. Herrie, the Muts-factory, one day I will come and visit you in Hungary and see your magnificent ‘castle’. Patrick, I think you and I are the last ones remaining (except for Pim and Herrie) from Titia’s ‘klasje 2004’. I wish you all the best with the protein facility, it is in good hands! Fra, sempre con un grande sorriso, will never forget your ‘ajuuto!’ and ‘Gateaux Pimonde’ on Reunion Island. Hang on, don’t give up, you’re almost there!!

And next time we meet, your Dutch will be as perfetto as my Italian :) Alex aka Sasha, thanks for all your help with the infamous Biacore! I will miss our talks about boxing and judo. Maybe one day your sons will be the new Dutch judoka champions :) Rick, thanks for all the discussions about analyzing biacore data and enzyme kinetics. Wish you all the best for the future! Kash, everyone in the lab says your default mode is being grumpy, but luckily I haven’t experienced that ‘Mode’ much :) Thanks for being there when I needed to talk about science and life and I’m glad you will be standing next to me as my paranimf. You better not turn into default :P Judith, your finishing-my-phd-in-four-years plan did not turn up what you were hoping for, but I am sure you will finish it way faster than I did. At least nothing explodes anymore in the lab and everything is Judith-proof :) Mariano, mio amico italiano, I was hoping I could have practiced my Italian with you more often, but I guess I have to come and visit you and your new group in Napoli one day and discuss about women in the Italian way :) Flora, when are we going hiking again? We should organize a reunion-hike on Reunion Island! Danny, is ADO going to repeat season ‘10-’11 again someday? Really hope you will get your BAP1/ASX-structure :) Marcell(in)o, I will miss our discussions about USPs, our travels and the different restaurants we have been to :D I have a feeling our paths will cross again soon. Good luck with continuing the USP project! Michael, the last member to join the Sixma lab. Wish you all the best! Bierje? To all the other people in the lab: Eli, John, Robbie, Leonie, Krista, Jens, Magda, Caroline, Christophino –take care next time! I won’t be there to drag you to your hotelroom :) – and finally Tati –my ‘little sister’ always taking care of things in the lab and helping everyone out. Will miss the fun and teasing in the lab :) : it was great having you as my colleagues and thank you for all your help and advice.

Of course being in the lab for so long, I have met
and worked with many other people who left the lab: Valerie, Puck, Joyce, Gretel, Sari, Ganesh, Victor, Chris, Annet, Andrea, Azusa, Vangelis & Valeria, Kostas, Serge, Oli, Mark H., Koen, Angelina, Mobien, Suzan, Diederick, Cristiane, Wijnand, Dene, Eirini (gyfsaki) and Bernat: thank you all!! I enjoyed working with you.

Special thanks to the Ovaa-lab: Huib, Farid, Paul, Remco, Reggy and Dharjath. Thank you all for the chemical tools/reagents. After a slow start, we all can finally benefit from each other's skills.


Dear Ale, even though I will be going to the other side of the ocean, you will always be with me. Mmmmbeeeeee!!!

I’ve made it!
Mark