






Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



Structural basis studies of Ubiquitin/SUMO- related protease activity

Ying Li

2022

Institut de Biotecnologia i Biomedicina (IBB)
Departament de Bioquímica i de Biologia Molecular
Universitat Autònoma de Barcelona



Structural basis studies of Ubiquitin/SUMO- related protease activity

Doctoral thesis for the Doctor of Philosophy Degree in
Biochemistry, Molecular Biology, and Biomedicine

Supervisor: Dr. David Reverter Cendrós

2022

Institut de Biotecnologia i Biomedicina (IBB)
Departament de Bioquímica i de Biologia Molecular
Universitat Autònoma de Barcelona

Nothing in life is to be feared, it is only to be understood.

- Marie Curie

Acknowledgements

First and foremost, I would like to express my gratitude to my family for their support and concern in providing me with the opportunity to see the wider world.

My deepest thanks, especially to my supervisor, Professor David Reverter, for providing me the opportunity to pursue a Ph.D., leading me into the world of protein crystals, and making me feel the beauty of scientific research. Thank you for your constant encouragement and guidance, without your consistent guidance I would not have been able to complete the task. You have influenced me more than anyone in Spain.

I would like to thank my lab partners: Nathalia, Jara, Lucia, Helena, Giovanni, Bing and Irina. I would like to express my heartfelt thanks to Nathalia, you have guided and helped me a lot over the past four years and I have learned a lot of very useful experiences and knowledge from you. Thanks to Jara, I am grateful to have such a good friend in my life, your kindness and patience have helped me clear a lot of haze, hope to see you in China in the future. Thanks for Lucía and Helena, your enthusiasm and laughter lightened the atmosphere in the lab, I learned much special information from you, it was a pleasure to work with you.

Thanks to all my Chinese friends in Spain. Thank for Weiqianq, I might not be able to complete my Ph.D. without you. You helped me with a lot of science and life advice. The best thing in a distant country is meeting you. I wish you all the best in your scientific endeavors.

I would also like to thank Susanna, Javier, Molood, Marcos, Marta, Cristina, Manuel, and Andera from the neighbor lab PDP group, as well as David, Pau, Paula, Eddi, Marta, and Sergi from the Enzimo group. Thanks also to the other IBB groups, as well as Maria and Jordi from CRAG-UAB. Thanks to the AlphaFold algorithm developed by DeepMind of Google, changed my Ph.D. grades and future career in time.

I would like to express my gratitude to the China Scholarship Council (CSC) for supporting my Ph.D. program.

Table of Contents

ABSTRACT	3
ABBREVIATIONS	5
INTRODUCTION	7
Ubiquitin-specific proteases	7
<i>Ubiquitination and Deubiquitination</i>	7
<i>Ubiquitin code</i>	8
<i>Deubiquitinating enzymes (DUBs)</i>	10
<i>Ubiquitin-binding domain</i>	11
<i>DUBs in Diseases</i>	12
<i>Ubiquitin-Specific Proteases (USPs)</i>	13
<i>Classical USP structure (USP7)</i>	14
<i>Atypical member of ubiquitin USP (USPL1)</i>	16
Small ubiquitin-like modifiers	16
<i>SUMO and SUMOylation</i>	16
<i>SEN/ULP family</i>	18
The effector protein NopD	21
<i>Rhizobium-Legume Symbiosis</i>	21
<i>Secretion system of Rhizobium</i>	22
<i>Bacterial Type III secretion system</i>	22
<i>Effector protein</i>	23
OBJECTIVES	25
CHAPTER I: STRUCTURAL BASIS FOR THE SUMO PROTEASE ACTIVITY OF THE ATYPICAL UBIQUITIN-SPECIFIC PROTEASE USPL1	27
Introduction	28
Results	31
<i>Covalent crosslinked complex between USPL1 and SUMO2 precursor</i>	31
<i>Overall structure of the USPL1-SUMO2 complex</i>	32
<i>Structural comparison with ubiquitin-specific USP members</i>	34
<i>SUMO2 interface of the C-terminal tail with USPL1</i>	37
<i>SUMO2 Interface with the thumb and fingers subdomains</i>	39
<i>Mutagenesis analysis of the specific interface contacts in USPL1</i>	41
Materials and Methods	45
Supplementary information	49
CHAPTER II: STRUCTURAL BASIS FOR THE SUMO2 ISOFORM SPECIFICITY OF SENP7	57
Introduction	58
Results	61
<i>Preparation of the covalent crosslink between SENP7 and SUMO2</i>	61
<i>Overall structure of the complex between SENP7-SUMO2</i>	62
<i>Interface between C-terminal tail of SUMO2 and SENP7</i>	64
<i>Interface between the globular domain of SUMO2 and SENP7</i>	65
<i>Activity assays with SENP7 point mutants.</i>	68
Materials and Methods	71

Supplementary information	75
CHAPTER III: STRUCTURAL ANALYSIS OF <i>BRADYRHIZOBIUM</i> NOPD PROVIDES INSIGHTS INTO THE DUAL PROTEASE ACTIVITY FOR UBIQUITIN AND SUMO	77
Introduction	78
Results	80
<i>NopD has a dual activity for SUMO and ubiquitin</i>	80
<i>Overall structures of NopD-SUMO2 and NopD-ubiquitin complex structures</i>	81
<i>The deubiquitinating activity of NopD does not depend on an N-terminal extension</i>	82
<i>Different binding interface with ubiquitin in NopD compared to XopD</i>	84
<i>Interface between C-terminal tail of ubiquitin and AtSUMO2 with NopD</i>	86
<i>Interface between ubiquitin and AtSUMO2 globular domain with NopD</i>	87
Materials and methods	91
Supplementary information	94
DISCUSSION	95
CONCLUSIONS	103
GENERAL EXPERIMENTAL METHOD	107
<i>Plasmid and cloning</i>	107
<i>Transformation to XL1-Blue strain competent cells</i>	109
<i>Colony PCR and plasmid sequencing</i>	110
<i>Transformation to Rosetta2 or BL21 competent cells</i>	110
<i>Protein Expression Test</i>	111
<i>Protein expression and purification</i>	111
<i>Protein concentration and Crystallization</i>	112
<i>Data Collection</i>	113
PUBLICATIONS	115
REFERENCES	117

ABSTRACT

Post-translational protein modifications by ubiquitin and SUMO regulate many major pathways in the cell. These modifications can be reversed by de-ubiquitinating enzymes such as ubiquitin-specific proteases (USPs) or deSUMOylases such as SUMO-specific proteases (SENPs). Proteolytic activity towards ubiquitin-modified substrates is common to all USP family members except for USPL1, which shows a unique preference for the small ubiquitin-like modifier SUMO. In humans, the deSUMOylating activity is mainly conducted by the SENP/ULP protease family, which is constituted of six members sharing a homologous catalytic globular domain. SENP6 and SENP7 are the most divergent members of the family and they show a unique SUMO2/3 isoform preference and a particular activity for dismantling polySUMO2 chains. NopD, an effector of the type III secretion system (T3SS) from *Bradyrhizobium*, is a multipurpose enzyme with specificity for ubiquitin and plant SUMO substrates.

In the present thesis, we first present the crystal structure of USPL1 bound to SUMO2. We find that USPL1 lacks major structural elements present in all canonical USPs members such as the so-called blocking loops, which facilitate SUMO binding instead of ubiquitin. Second, we reveal the crystal structure of the catalytic domain of human SENP7 bound to SUMO2, and described the specific contacts between SUMO2 and a unique insertion in SENP7 (named Loop1) that is responsible for the SUMO2 isoform specificity. We finally present the crystal structures of NopD bound to Arabidopsis AtSUMO2 and to ubiquitin, and reveal the molecular details for this dual activity of NopD for SUMO and ubiquitin. Mutagenesis analysis disclose the key determinants, such as the unique insertion loop in NopD, that are responsible for unusual dual NopD activity for SUMO and ubiquitin.

All results give insight into the structural details of these protein complexes' interface and contribute to enrich the knowledge of the distinct evolutionary directions followed by members of deubiquitinating and deSUMOylating families.

ABBREVIATIONS

°C	Degree Celsius
Amp	Ampicilin
ATP	Adenosine triphosphate
BL21	BL21 (DE3) strain competent cells
BME	β -mercaptoethanol
Chl	Chloramphenicol
DHA	2,5-dibromo hexanediamide
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
EDTA	Ethylenediaminetetraacetic acid
FL	Full-length
At	<i>Arabidopsis thaliana</i>
HEPES	4-(2-hydroxyethyl)-1 piperazine ethanesulfonic acid
IGEPAL	t-Octylphenoxy polyethoxyethanol
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kan	Kanamycin
KDa	kilodalton
kDa	Kilodalton
LB	Luria-Bertani
mM	milimolar
nM	Nanomolar
O/N	Overnight
PA	Propargylamine
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PTM	Post-translational modification
R2	Rosetta 2 strain competent cells
Rpm	Revolutions per minute
RQ	Resource Q
RS	Resource S
RT	Room temperature
SD	Superdex
SDS	Sodium dodecyl sulfate
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SUMO	Small Ubiquitin-like modifier
TEMED	N, N, N', N' - Tetramethylethylenediamine
Tris	Tris (hydroxymethyl) aminomethane
Tween-20	Polysorbate 20
WT	Wild-type
XL1B	XL1-Blue strain competent cells
μ M	micromolar
RF-PCR	Restriction Enzyme Free PCR
OD	Optical Density

INTRODUCTION

Ubiquitin-specific proteases

Ubiquitination and Deubiquitination

The ubiquitin-proteasome system (UPS) is the major pathway for the degradation of more than 80% of intracellular proteins. Ubiquitination is the major signal to label proteins for degradation, it controls their fates and contributes to the correct cellular homeostasis during all cell cycle phases [1,2]. After a particular protein receives a signal for degradation, such as phosphorylation, DNA damage, protein misfolding, or others, a cascade of ubiquitinating enzymes, namely the ubiquitin-activating enzymes E1, the ubiquitin conjugation enzymes E2, and the ubiquitin ligases E3 labels the protein target with a ubiquitin chain to be carried to the 26S-proteasome, which degrades ubiquitinated-target proteins and recycles ubiquitin for reuse (**Figure 1A**) [3]. Several E1, tens of E2, and hundreds of E3 enzymes have been found in humans, and they are responsible for the formation of the isopeptide bond between a lysine residue in the target and the C-terminus of ubiquitin. Cells degrade unnecessary proteins in this highly specific manner to prevent cell damage induced by incorrect interaction with other proteins or abnormal protein aggregation. In addition to labeling proteins for degradation by the UPS pathway, different types of ubiquitin tags can be specifically used by proteins in signaling pathways other than degradation [1,4].

Deubiquitination refers to the process by which ubiquitin molecules are removed from ubiquitin-labeled proteins under the proteolytic action of deubiquitinating enzymes, which cleaves off the isopeptide bond between the C-terminal tail of ubiquitin and the target lysine, thus changing the fate of ubiquitinated proteins and in some instances by preventing them from degradation (**Figure 1B**) [5,6]. Deubiquitination has an opposite function to the action of ubiquitin E3 ligases and their balance inside the cell is essential for protein homeostasis and the correct function of the cell. The human genome encodes about 100 deubiquitinase (DUB) genes [7], in contrast to more than 800 genes encoding ubiquitin E3 ligases. These numbers highlight the major role of the ubiquitin pathway in

INTRODUCTION

the cell, which must provide a high selectivity grade in the regulation of protein targets.

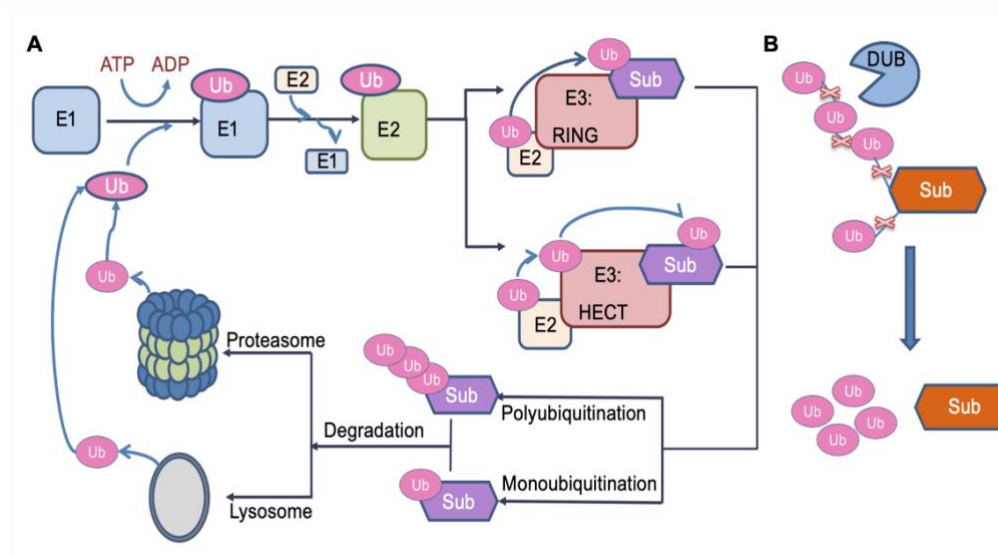


Figure 1. Protein ubiquitination and deubiquitination [8]. **A.** E1 starts the ubiquitination process along with ATP. The E1 enzyme passes the Ub protein to E2s. The E2s are then complex with Ub ligase E3s. In the final step, E3s catalyze Ub transferring to substrates. Proteins can be ubiquitinated with multiple types including monoubiquitination and polyubiquitination, and degraded in the lysosome or the proteasome, respectively. **B.** DUBs remove Ub molecules from target substrates.

Ubiquitin code

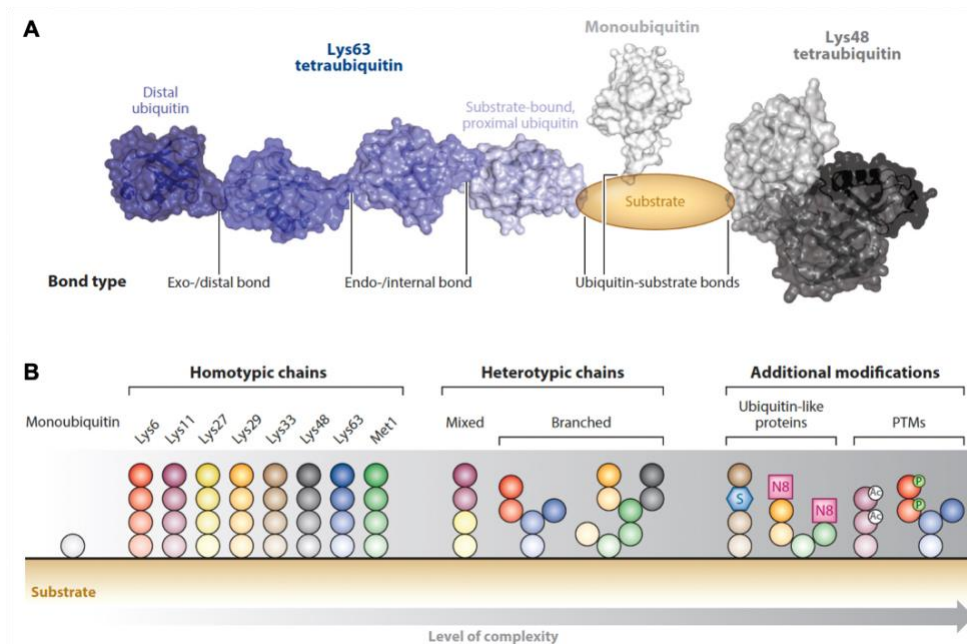


Figure 2. Depiction of the ubiquitin modifications complexity [6]. **A.** The three most common types of ubiquitination are depicted as structural models attached to a virtual substrate: monoubiquitination, Lys48-linked polyubiquitination, and Lys63-linked polyubiquitination. Distinct chain topologies provide different structures, which contribute to unique recognition. The various bond kinds' terminology is introduced. **B.** A schematic illustration of the ubiquitin code's increasing complexity.

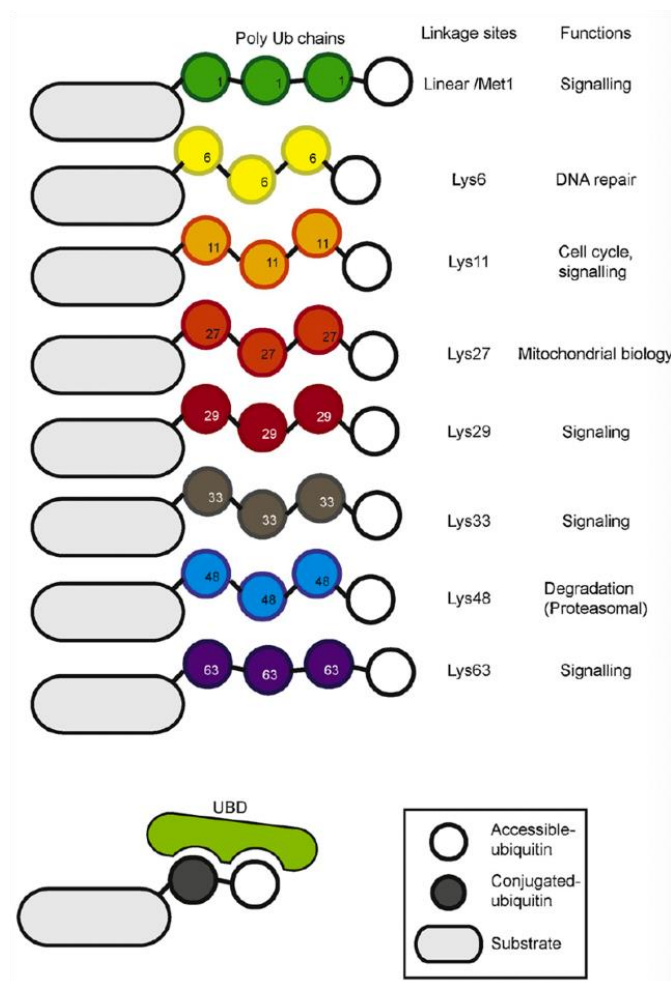


Figure 3. Various types of ubiquitination on a substrate are regulated by distinct enzymes [9]. Substrate ubiquitination with 8 possible linkage types (Met1/linear, Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63) of ubiquitin chains. Their major functions are indicated (right). Below, a ubiquitin-binding domain (UBD)-containing protein is recruited to a signaling complex by interacting with a distinct ubiquitin chain.

Ubiquitin can form different types of chain linkages, depending on the lysine residue used in ubiquitin as a target to build poly-ubiquitin chains. Additionally, four different types of protein ubiquitination in substrates enrich the diversity in ubiquitination, namely monoubiquitination, multi-monoubiquitination, homotypic polyubiquitination, and heterotypic polyubiquitination (**Figure 2A&B**) [10,11]. In homotypic polyubiquitination, multiple ubiquitins are linked to each other to form a chain and are connected to a certain lysine residue (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63) or to the amino-terminal methionine (**Figure 2B**). Various types of ubiquitination on a substrate are regulated by distinct enzymes (**Figure 3**).

INTRODUCTION

Also, protein ubiquitination could be accompanied by some other ubiquitin-like modifiers (UBL), such as SUMO, NEDD8, and ISG15. Heterotypic chains include mixed ubiquitin chains and branched ubiquitin chains (**Figure 2B**). In heterotypic polyubiquitination, by generating different connection groups on different ubiquitin molecules, a mixed ubiquitination sequence can be formed. Several modifications can also occur on each ubiquitin, resulting in branched polymers. Due to the diversity of ubiquitinated chain linkage styles, which has been named as “ubiquitin code”, particular DUBs have evolved to have specific preferences for the different types of ubiquitin chains [6,7,10,11].

Deubiquitinating enzymes (DUBs)

According to sequence and structure similarity, deubiquitylating enzymes can be divided into seven families, each presenting a unique structural fold: Ubiquitin-specific proteases (USPs), Ubiquitin C-terminal hydrolases (UCHs), Ovarian tumor proteases (OTUs), Machado–Joseph Disease protease family (MJDs), MINDY protease family (MIU-containing DUB family), JAMM family (Jad1/Pad/MPN domain-containing metalloenzymes), and the newly discovered ZUFSP/Mug105 family (zinc finger with UFM1-specific peptidase domain protein/C6orf113/ZUP1) [12,13] (**Figure 4 & Table 1**). Amid them, the JAMM family are zinc-dependent metalloproteinases, while the other six families are cysteine proteases, displaying a characteristic cysteine residue in the context of the active site catalytic triad. Except for MJDs, the other families are highly conserved in yeast and humans.

Species/DUBs	USP	OTU	JAMM	MJD	UCH	MINDY	ZUP1
<i>Homo sapiens</i>	56	17	12	4	4	5	1
<i>Danio rerio</i>	44	11	12	3	4	5	1
<i>Drosophila melanogaster</i>	23	5	10	1	3	1	0
<i>Schizosaccharomyces pombe</i>	11	1	4	0	2	1	1
<i>Saccharomyces cerevisiae</i>	11	3	7	0	1	1	0

Table 1. The seven classes of deubiquitinating enzymes (DUBs).

Among the 99 DUBs members, 11 are considered pseudo-enzymes due to the lack of key residues for deubiquitinating activity, nevertheless they still can allosterically activate other active deubiquitinating enzymes and other types of active enzymes, thereby executing a critical biological role [14].

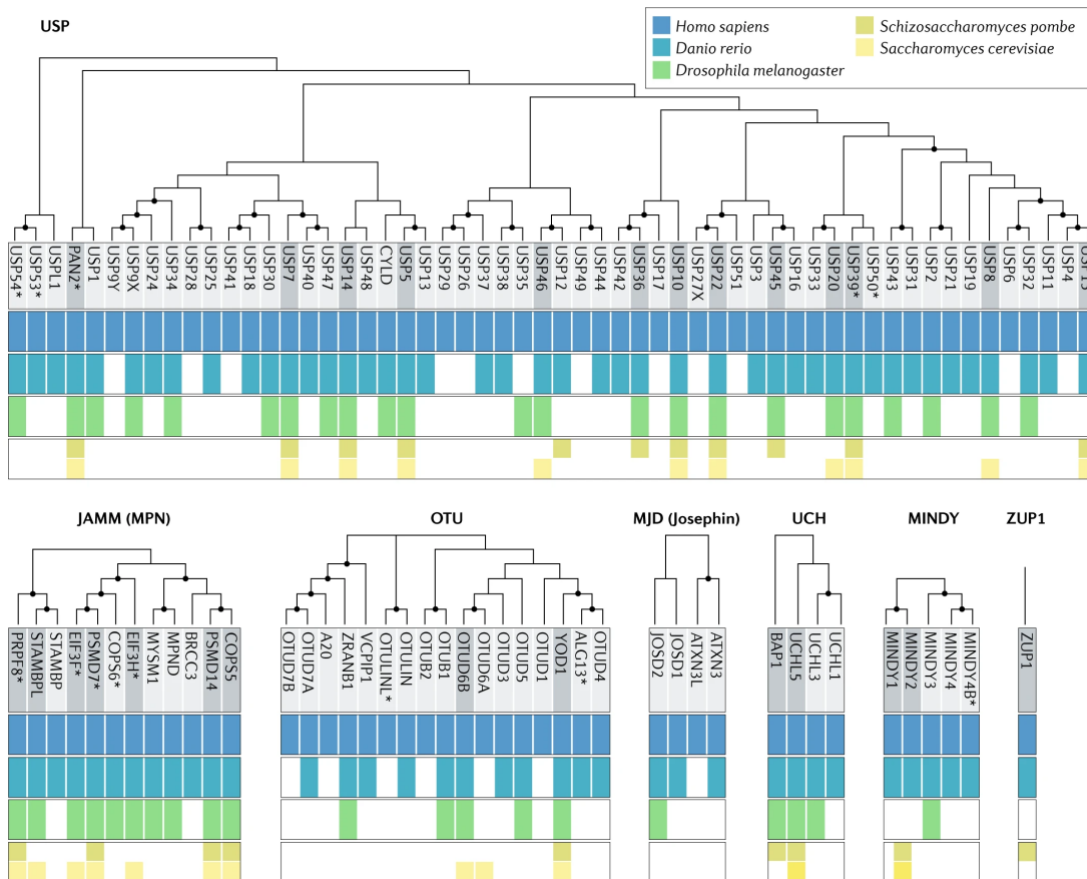


Figure 4. Phylogenetic conservation of DUBs [7]. Deubiquitylating enzymes (DUBs) are arranged according to a bootstrapped neighbour joining phylogenetic analysis of their catalytic domains, with the most reliable nodes (supported by bootstrap values of >50%) indicated by a black dot. *Predicted to be inactive on the basis of sequence or structural considerations.

Ubiquitin-binding domain

UBD (ubiquitin-binding domain) is a type of modular domain that can non-covalently bind to ubiquitin or ubiquitinated substrates, which is widely present in deubiquitinating enzymes. Almost 20 different UBDs families have been discovered so far. According to the structure of UBDs, they are divided into 5 categories: α helix, Zinc finger, PH domain, Ubc-like, and others. The most common in DUBs are UBA (ubiquitin-associated domain), UBL (ubiquitin-like modifiers), UIM (ubiquitin-interacting motif), ZnF-UBP (zinc finger ubiquitin-specific protease domain), and ZnF-A20s (A20-type zinc fingers) [7,15,16].

INTRODUCTION

Recent research has discovered a new UBD named CoCUN [15]. In the process of DUBs recognition and hydrolysis of ubiquitin chains, UBD plays an indispensable role. For example, OTUD1 can specifically hydrolyze the Lys63 ubiquitin chain, but the lack of the UIM domain decreases its specificity [16]. Also, the adjacent UBL domains of USP7 are indispensable to achieving complete deubiquitinating activity [17].

The mechanism by which deubiquitinating enzymes exert their function in the cell can be complex, in some cases it is not only deubiquitination, but it can also promote ubiquitination by recycling ubiquitin molecules, proofreading of the ubiquitination process, and decomposing ubiquitination inhibitors [18].

DUBs in Diseases

A large number of deubiquitinating enzymes are not only playing a role in the normal physiological activities of cells, but can also be relevant for the occurrence and development of tumorigenesis and other pathologies [19–21]. Target proteins for deubiquitinating enzymes include enzymes, transcription factors, signal transduction molecules, immune response proteins, viral proteins, epigenetic factors, and many other regulators of cell homeostasis, also including products of known oncogenes or tumor suppressor genes [22–29]. Therefore, disease treatment strategies with DUBs as molecular targets have a broad development value and promising clinical application prospects [30]. Many studies have shown DUBs involvement in the regulation of Wnt/ β -catenin signaling, TGF- β (transforming growth factor- β), Akt (Protein Kinase B), NF- κ B (nuclear factor kappa-light-chain-enhancer of activated B cells) and other cancer-related pathways [31–34].

Several DUBs are involved in the regulation of the Wnt/ β -catenin signaling. For example, overexpression of USP5 will cause an increase in the amount of transcription factor FoxM1, which will increase the content of β -catenin, resulting in faster cell proliferation and carcinogenesis in many tumors [35]. UCH37 can specifically bind and deubiquitinate transcription factor 7 (Tcf7) to activate Wnt signaling in human liver cancer cells [36]. In pediatric high-grade glioma, abnormal expression of UCH-L1 can promote tumor formation [37]. Abnormal activation of USP7 can lead to colorectal cancer [38]. USP14 overexpression is closely related to hepatocellular carcinoma [39]. USP4 is related to

colorectal cancer since the overexpression of USP4 exists in many types of tumors, thus it is considered to be a potential oncogene [40–43]. Most DUBs will promote the occurrence of tumors, but a small number of DUBs play the opposite role. For example, the tumor suppressor cylindromatosis (CYLD), which is expressed in tumor cells, inhibits the proliferation and spread of tumor cells. CYLD is not expressed in multiple myeloma, and inhibits the growth of multiple myeloma by acting on the Dvl substrate [44,45].

Most DUBs reduce the degradation of protein targets by protecting them from the proteasome, thereby a high concentration of TGF- β , promoted by specific DUBs, might lead to tumor occurrence and metastasis. Deubiquitinating enzyme USP10 promotes the metastasis of hepatocellular carcinoma metastasis by deubiquitinating and stabilizing Smad4 [46]. But there are special cases, CYLD can inhibit oral squamous cell in the metastasis of carcinoma, CYLD controls the downstream TGF- β pathway by regulating Smad7, thereby inhibiting the occurrence and development of cancer [47,48].

DUBs also play an important role in regulating the NF- κ B pathway. OTULIN/CYLD can interact with LUBAC (linear ubiquitin chain assembly complex) through the PUB (PNGase/UBA or UBX) domain of HOIP (a catalytic subunit of LUBAC), thereby inhibiting the NF- κ B pathway and exerting anti-tumor effects [49]. Studies have shown that alterations in the CYLD gene are closely related to HPV-associated cancers, it activates NF- κ B and is implicated in invasion and metastasis [50–52]. USP15 potentiates NF- κ B activation by differentially stabilizing TAB2 and TAB3 [53].

The Akt pathway promotes the survival and proliferation of cells in the organism, and its expression is increased in cancer cells [54]. The increased expression of USP4 will reduce the ubiquitination process of downstream PRL-3 and increase the content of PRL-3, thereby activating the PI3K/Akt pathway. The PI3K/Akt pathway is over-activated, causing tumorigenesis and other diseases [41]. In osteosarcoma cells, overexpression of USP22 leads to increased activity of the PI3K/Akt pathway and causes cancer [55]. USP22 plays a vital role in the development of chemoresistant hepatocellular carcinoma cells [56].

Ubiquitin-Specific Proteases (USPs)

Within the deubiquitinating enzymes, the ubiquitin-specific proteases family (USPs)

INTRODUCTION

contains the major number of members (more than 50) and its large structural diversity relies on the presence of multiple domains. All members are cysteine proteases and contain the characteristic catalytic triad in the active site of the catalytic USP domain. The conserved USP domain consists of three subdomains that were initially described as thumb, palm, and fingers of a human right hand in the USP7/HAUSP structure [57]. The catalytic triad of the active site is located between the palm and thumb subdomains, and the finger subdomain is responsible for the interaction with the ubiquitin substrate [58]. In addition to the catalytic domain, other domains of USPs contribute to a large diversity of functions, such as zinc finger domains, ubiquitin-like domains, ubiquitin-associated domain (UBA), ubiquitin-interacting motif (UIM). Some of them are associated with the specific recognition of substrates and poly-ubiquitin chains. Most USP members have no specific linkage-type preference *in vitro* [59], but *in vivo* there are some exceptions, for example, CYLD shows K63-linked specificity [60]; USP30 preferentially processes K6-linked chains [61]. Also, some of them show preference for different types of UBL such as USPL1 and USP18 for SUMO and ISG15, respectively [62,63].

Classical USP structure (USP7)

Research on deubiquitinating enzymes shows that although deubiquitinating enzymes are classified into several distinct families, the structural differences between members have caused their functional diversity. Here, we have selected a classical example.

USP7/HAUSP has been deeply studied from the structure/function standpoint and its molecular signaling pathway is well established and considered a potential therapeutic target for cancer [64]. USP7 is a multidomain protein with seven domains: the “canonical” USP catalytic domain (aa208-560), one N-terminal TRAF-like domain (Tumor necrosis factor Receptor-Associated Factor), and five consecutive C-terminal UBL domains (Ubiquitin-like domains) [65]. UBL domains can be also found in other USPs and are known to play a role in the regulation of the enzymatic activity [66]. USP7 can cleave K6, K11, K33, K48, and K63-linked ubiquitin chains [57,67].

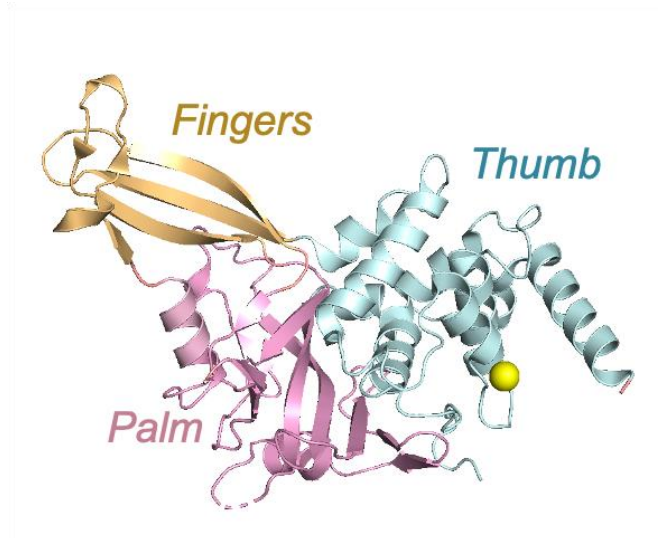


Figure 5. Right-hand subdomains of USP7. Crystal structure of the USP7 catalytic domain. Fingers domain (yellow), palm domain (pink), thumb domain (green) (PDB: 5JTV, 5FWI).

The initial crystal structure of the catalytic domain of USP7 defined the characteristic right hand-like structural fold of the USP family, containing the active site catalytic triad: Cys223, His464, and Asp481. Remarkably, in contrast to other USPs, the cysteine and histidine of the USP7 catalytic domain are in a non-productive configuration in the apoenzyme and need to be activated by structural rearrangement upon ubiquitin-binding [57,66] (**Figure 5**).

Interestingly, the active site of USP7 shows an inactive conformation in its apo form, but in the presence of ubiquitin, the active site is remodeled and the catalytic triad aligned [58] (**Figure 5**). UBL domains also play a role in the activation of the catalytic domain [68]. Deletion of the USP7 C-terminal UBL domains have an impact on the USP7 activity and then USP7 cannot bind ubiquitin chains [17,65]. According to the structure of the USP7CD- UBL4-5-Ubiquitin complex (**Figure 5**), the C-terminal extension after UBL-5 domain binds a cavity next to the active site and promotes its activation. The C-terminal peptide seems to stabilize the active conformation instead of significantly change the structure of the catalytic domain. It is accommodated in a hydrophobic cavity by several contacts, and can also bind the C-terminal Arg74 of ubiquitin [65]. This regulatory C-terminus extension of USP7 is essential for the enhancement of the USP7 catalytic activity [65,69].

Atypical member of ubiquitin USP (USPL1)

Most USP members are active towards ubiquitin, but there are two exceptions, such as USP18, which is active towards the double-headed ISG15, and USPL1, a distant member of the family specific for SUMO [62,70]. USPL1 is a deSUMOylating enzyme instead of a deubiquitinase. SUMO and ubiquitin display little homology (16% sequence identity), foreseeing the presence of a divergent binding mechanism between USPL1 and SUMO. the study of USPL1 would be significant for the USP family.

The full-length of USPL1 is composed by 1.092 residues that contains a USP-like catalytic domain in a middle region and long disordered protein extensions without the presence of evident globular domains. So far, the only function described for USPL1 takes place in the nucleus, where is a protein component of the Cajal Bodies (CBs), co-localizing with coilin [62]. USPL1 seems to have an impact in the formation and dynamics of CBs and in cell proliferation, as observed in the USPL1-depleted cells [62]. Whilst USPL1 deletion does not affect the overall SUMOylation in the cell, it causes significant frizzled protein mislocalization and damage in cell proliferation, which interestingly does not depend on its deSUMOylase catalytic activity [71]. The CBs are membrane-less compartments involved in the biogenesis of snRNP and snoRNP, maintenance of telomeres, and processing of histone mRNA [72]. USPL1 is a low-abundant component of the CBs that plays a role in the RNAPII transcription of snRNA that is essential for cell growth [71]. Moreover, SUMO has been involved in tumorigenesis, genetic variants of USPL1 are closely related to grade-3 breast cancer [73], as well as USPL1 may also be involved in the signaling pathway of Multiple myeloma [74].

Small ubiquitin-like modifiers

SUMO and SUMOylation

Small ubiquitin-like modifiers (SUMO) are post-translational modifiers that regulate a wide range of protein functions in cells [75,76]. They are members of the ubiquitin-like protein (Ubl) family. Although SUMO and ubiquitin have only about 18% sequence identity, their folded structures are very similar [77]. SUMO is widely expressed in all eukaryotes, and the human genome contains four SUMO proteins: SUMO1, SUMO2,

SUMO3, and SUMO4. Yeast and invertebrates have a SUMO protein called Smt3, vertebrates also have few SUMO. SUMO2 and SUMO3 have 97 percent sequence identity in humans. SUMO1 differs significantly from SUMO2/3, sharing only 47 percent identity with SUMO1 and SUMO2 [75,77].

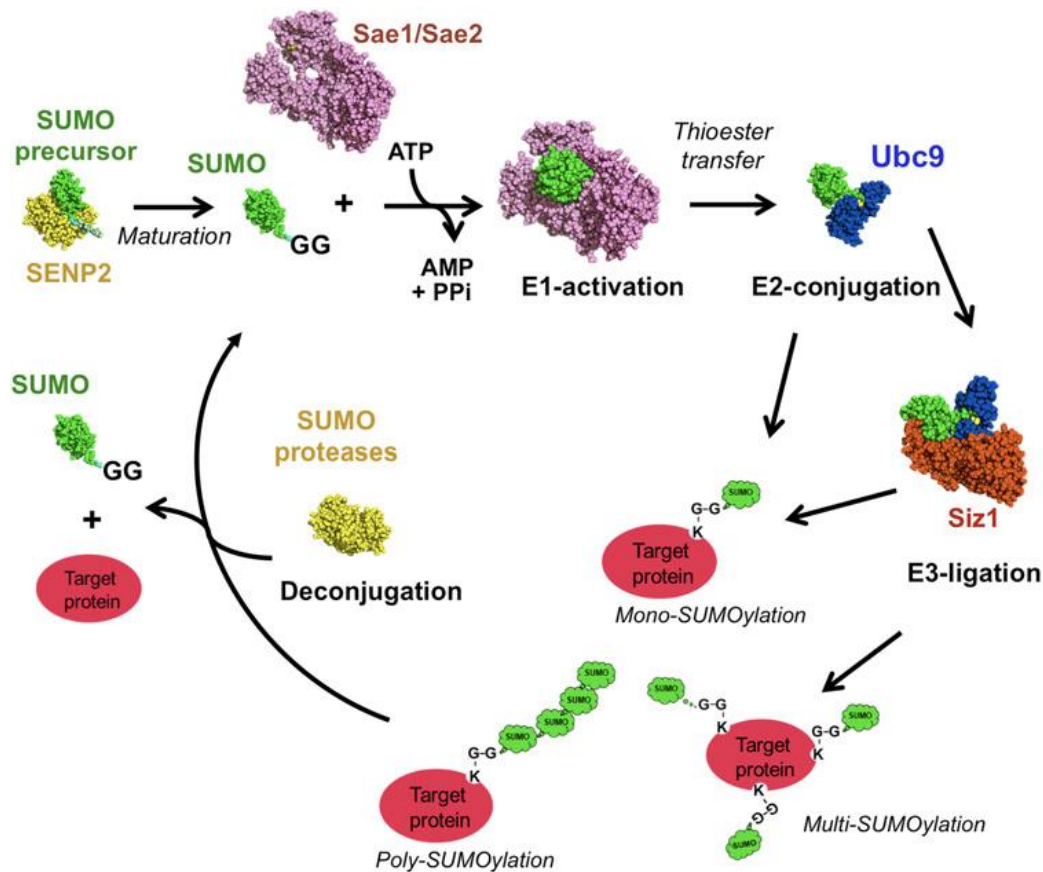


Figure 6. SUMO conjugation pathway [78]. Structures of the dedicated enzymes of the catalytic cascade that lead to the formation of SUMO conjugates are depicted in this diagram. By forming a high energy thioester-bond between the SUMO C-terminus and an internal Cys (Sae1–Sae2 PDBs 1Y8R, 3KYC), the ATP-dependent E1-activating enzyme activates the SUMO precursor (previously proteolytically processed leaving a Gly-Gly motif at the C-terminus). The SUMO-E1-thioester is then isoenergetically transferred to a Cys residue by the E2 conjugating enzyme (Ubc9 SUMO PDB 5JNE). The E2-SUMO-thioester can form covalent isopeptidic bonds by discharging SUMO on one or more lysine residues from target substrates. The action of an SUMO E3 ligase enzyme can stimulate E2-SUMO discharge (Siz1-SUMO-Ubc9 structure, PDB 5JNE). Poly-SUMO chains can be formed on target substrates as well. Finally, the action of a specific SUMO protease family (SENP2-SUMO PDB 2I00) can reverse SUMO conjugation by cleaving off SUMO precursors and SUMO conjugated substrates.

The most important function of ubiquitin among UbLs is the degradation of intracellular proteins by the ubiquitin-proteasome system (UPS) [1]. SUMOylation is involved in many cellular pathways and functions such as DNA replication, nuclear transport, and DNA damage control [79–81]. SUMO, ubiquitin, and all other UbL modifiers are covalently attached to target proteins via an isopeptide bond with an internal lysine residue, but

INTRODUCTION

first UbLs must be activated via a dedicated conjugation pathway involving the E1, E2, and E3 enzymes [78,82]. A conserved C-terminal diglycine motif (-Gly-Gly) is exposed by the action of specific proteases [83,84]. Afterward, the enzymatic cascade consists of a single heterodimeric E1-activating enzyme (Sae1/Sae2) that activates (by ATP) and loads SUMO onto the internal cysteine residue, forming a thioester bond [85–87]. The SUMO is then transferred to the E2-conjugating enzyme (Ubc9) via a thioester bond formed with the active site cysteine [88–93]. Finally, a charged E2 enzyme can directly transfer SUMO to one or more lysine residues of a target substrate via mono-, multi-, and poly-SUMOylation, which can be facilitated by an E3 ligase enzyme as well [94,95]. Interestingly, the covalent attachment of ubiquitin and SUMO to protein targets is reversible via de-ubiquitinating (DUBs) or de-SUMOylating proteases [96,97]. The SENP/Ulp family SUMO proteases productively cleave the isopeptide bond between the substrate and SUMO [97–99]. Thus, SUMO conjugation is a highly dynamic process governed by a balance between E3 ligase and the SENP/Ulp family (**Figure 6**).

SENP/ULP family

Until recently, SUMO proteases were only constituted by the 6 members of the SENP protease family in humans (SENP1, SENP2, SENP3, SENP5, SENP6, SENP7), SENP8 has specificity for Nedd8 instead of activity for SUMO. Ulp1 and Ulp2 were the first two SUMO proteases discovered in *Saccharomyces cerevisiae* [83]. Ulp1 and Ulp2 have evolved into two branches: the SENP1 and SENP2, SENP3 and SENP5 evolved from Ulp1, while SENP6 and SENP7 evolved from Ulp2 [97]. According to sequence homology, substrate specificity, and subcellular localization, human SENPs are classified into three subfamilies: SENP1 and SENP2, SENP3 and SENP5, SENP6 and SENP7 [97,98,100,101] (**Figure 7**).

To maintain the balance between protein SUMOylation and deSUMOylation, SENP carry double functions: processing immature SUMO, and uncoupling SUMO conjugates. SENP process the peptide bond in the C-terminal of SUMO to expose diglycine motif (mature SUMO), and cleave the isopeptide bond between SUMO and the lysine residue of substrate [97]. This is a critical step in triggering SUMO1/2/3 activity and completing the SUMO cycle.

All Ulp/SENP family belong to the C48 cysteine protease structural class (SENP1-SENP3

and SENP5-SENP7) [97]. However, two novel types of SUMO proteases have been described in the recent years without sequence/structural homology to the SENP/ULP family: the deSUMOylating peptidase 1 and 2 (DESI1 and DESI2) [102]; and the deubiquitinating enzyme USPL1, which is a member of the ubiquitin USP family but is specific for SUMO rather than ubiquitin [62]. An analogous finding was observed in the structural characterization of SENP8/DEN1, which is another member of the SUMO protease family (SENP/ULP family), that is specific for the Nedd8, another type of Ubl modifier [103–106]. There are two Smt3-specific proteases in *S. cerevisiae*: Ulp1 and Ulp2, which are capable of separating Smt3 from modified proteins and process Smt3 precursors to mature forms with a C-terminal [83,107].

Among all human SENPs, SENP1, SENP6, and SENP7 are localized to the nucleoplasm [108], whereas SENP3 and SENP5 are localized to the nucleolus. SENP2 has a nuclear export signal along with SENP1 to facilitate its shuttling in and out of the nucleus [109]. All SENPs possess isopeptidase activity, only SENP1, SENP2 and SENP5 can proteolytically process the precursor SUMO. All SENPs prefer SUMO-2/3 to SUMO-1 for deconjugation, except for SENP1 and SENP2. SENPs are unable to process pre-SUMO4 and thus most likely do not form SUMO conjugates [110].

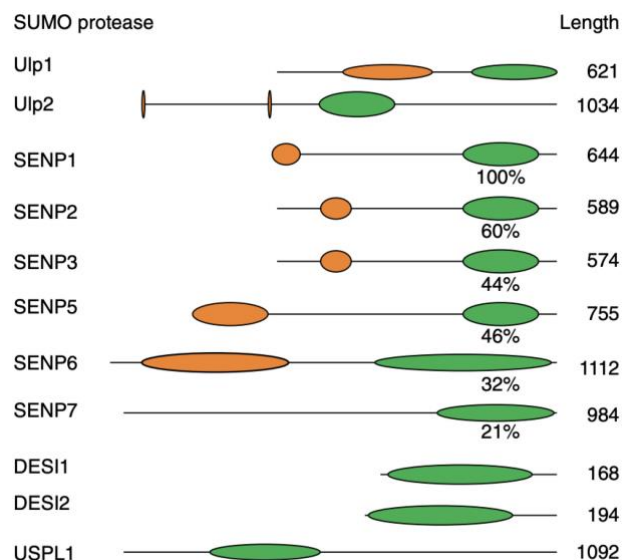


Figure 7. Structural organization of SUMO-specific-proteases/isopeptidases [99]. Domain organizations of Ulp/SENPs and Desi family members are depicted. The catalytic domain is represented by green ovals. Orange ovals represent the sequence determinants responsible for subcellular targeting. On the right side, the length of the proteins is shown as the total number of amino acids. Sequence identity with SENP1 is also shown for the catalytic domains of SENP family members.

INTRODUCTION

SENP3 and SENP5 are members of SENP subfamily with significant sequence homology and substrate specificities. They are both found in the nucleolus and have selectivity for SUMO-2 and SUMO-3 but not for SUMO-1 [111,112].

SENP5 may remove poly-SUMO2/3 from the Lys160 or Lys490 sites of PML (promyelocytic leukemia protein) using PML SUMOylation mutants as model substrates. However, SENP5 was unable to remove SUMO-1 from PML's Lys160 or Lys490 locations. Nonetheless, SENP5 may be able to remove SUMO-1, -2, and -3 from PML's Lys65 position. As a result, SENP5 has restricted SUMO-1 isopeptidase activity [111]. The manner in which SENP5 and SENP7 bind to the substrate and the key residues remain to be discovered.

SENPs have a catalytic triad (His- Asp-Cys) and a conserved C-terminal domain. SENP6 and SENP7 are the most divergent members of the SENP family, especially contacts between the interface of SENP and SUMO structures [100,113]. The catalytic domain of SENP6 and SENP7 have low sequence identity and loops insertion and prefer SUMO2/3 over SUMO1.

According to research, SENP7 interacts with SUMO1 and SUMO2/3, but it is more active against the SUMO2/3 isoform [114]. The structure of the catalytic domain of SENP7 shows that SENP7 has several unique or expanded secondary structure elements compared to SENP1 or SENP2, and the presence of an N-terminal α -helix, as well as Loop-1, -2, -3, and -4 of the four loops inserted. Structural modeling predicts that Loop-1 and Loop-2 are located on the surface of the protease, possibly interacting with SUMO in the binding site [115]. Deletion of loop 1 severely impairs the proteolytic activity of SENP7, and SENP7 Δ Loop 1 is less able to uncouple di-SUMO2/3 or poly-SUMO2/3 than SENP7 containing Loop 2 or Loop 3 deletions [115]. Both the spatial conformation of Loop 1 and the charge properties of Lys-691 are important for the efficient interaction of SENP7 with SUMO2 and thus for the correct cleavage of SUMO substrates. The Loop1 of SENP7 is indispensable for activity, while the Loop2 and Loop3 are inessential for the activity [114]. Loop 2 and Loop 3 are disordered and their removal did not produce any change in proteolytic activity, showing activity similar to wild-type [115]. The SENP7 catalytic domain shows a preference for SUMO deconjugation over processing. And, SENP7 has a higher deconjugation rate of di-SUMO2/3 or poly-SUMO2/3 than deconjugation of SUMO2/3-conjugated RanGAP1 [114].

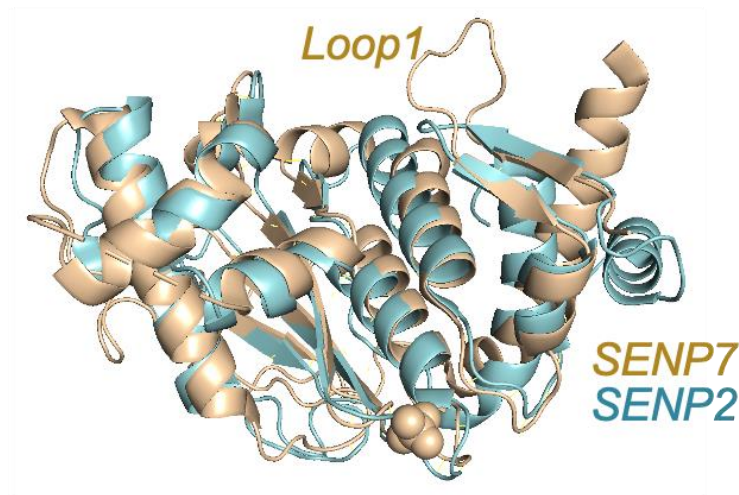


Figure 8. Structure of SENP7 catalytic domain [114]. Superposition of SENP7 and SENP2 catalytic domain in green and yellow, respectively. SENP7 Loop1 is indicated (PDB:1TH0 and 3EAY).

In vivo, SENP7 has specificity for SUMO2/3 as well. And the deletion of SENP7 results in the accumulation of SUMO-2 conjugates instead of SUMO-1 conjugates [116]. SENP7 is probably a crucial regulator of the PML (promyelocytic leukemia protein) turnover [117,118]. Recent studies have shown that SENP7 plays a critical role in maintaining CD8+ T cell metabolic fitness and effector functions [119]. SENPs are key enzymes in the control of balanced SUMOylation/deSUMOylation in the cardiovascular health and disease [120]. SENP7 maintains HP1 α cumulation at pericentric heterochromatin and regulates DNA repair by interacting with the chromatin remodeler CHD3 in mice [121,122]. The knockout/down experiments in vivo have shown an essential role for SENP7 in the mouse embryonic development [123]. SENP7 is transiently activated during the early stages of neuronal differentiation and is required for the vertebrate neuronal differentiation [124]. SENP7 has been implicated in breast cancer as a potential target for the cancer therapy [125,126].

The effector protein NopD

Rhizobium-Legume Symbiosis

Legumes are classified into six subfamilies, with 750 genera and nearly 20,000 plant species [127]. Most legumes fix nitrogen in the air with the help of the symbiotic bacteria Rhizobia for efficient nitrogen use [128]. The rhizobia-legume symbiotic system accounts for nearly 90% of the nitrogen in the natural environment. Different legumes form

INTRODUCTION

different root nodules, which are classified as determinate nodules or indeterminate nodules. The growth of determinate nodules is primarily determined by cell expansion, whereas the growth of indeterminate nodules is determined by cell division [129]. Legumes with indeterminate nodules are more common among them [129,130].

Rhizobium and legume genes recognize and interact with each other, resulting in the formation of nitrogen-fixing nodules [131]. The rhizobia-legume system has high specificity, a legume can only recognize some specific rhizobia, and rhizobia can only coexist with some specific legumes [132]. Rhizobia are associated with the plant through root hairs in most legumes, primarily through intercellular infection, crack entry, and root hair curling [129,133]. When the rhizobia receive the symbiotic signaling molecule flavonoids released by the host plant, the rhizobia initiates specific genes expression [134,135], secreting the Nod factor and related genes expression in the system, such as type III secretion effector proteins.

Secretion system of Rhizobium

Rhizobia secretes proteins into the extracellular space via various secretion systems, allowing it to exchange molecular signals and information with the host plant. T1SS, T2SS, T3SS, T4SS, T5SS, and T6SS are the six types of Secretion Systems found in Gram-negative bacteria. T1SS, T2SS, and T5SS secrete proteins from bacterial cells to the outside of the cell; T3SS, T4SS, and T6SS directly transfer bacterial effector proteins to the cytoplasm of the host cells, allowing for direct information exchange or modification of host cell substances [136,137]. The Type III secretion system is the most widely distributed and complicated secretion system in bacteria.

Bacterial Type III secretion system

Type III secretion system (T3SS) is located on the cell membrane of rhizobia. It is a multi-protein complex with over 20 proteins that is needle-shaped and evolved from flagella [138,139]. Flavonoids, plant symbiotic signaling molecules, induce the import of effector proteins from rhizobia into host cells, changing the signal pathway of host cells. This process is strictly regulated to ensure that various proteins are produced in an orderly manner [140]. **Figure 9** depicts the structure of the type III secretion system [141]. Only

a subset of rhizobia has Type III secretion systems [142,143].

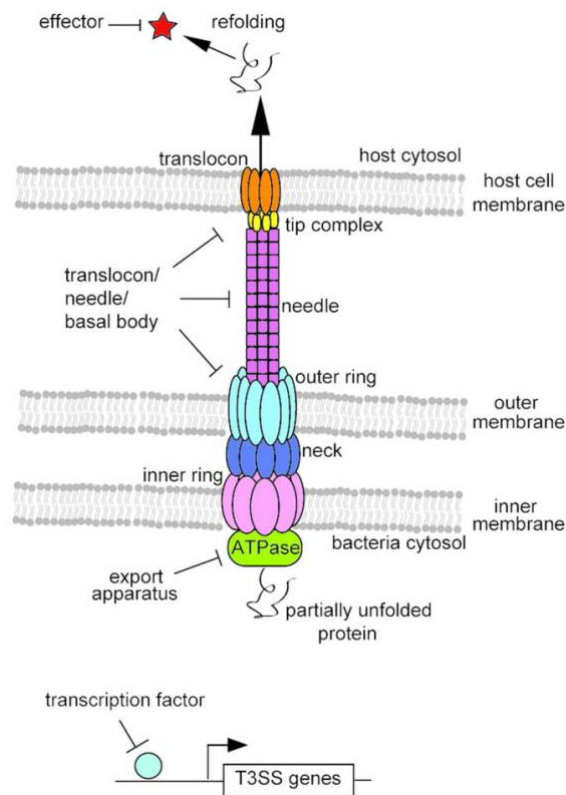


Figure 9. Potential targeting strategies for T3SS inhibitors [141]. Cartoon showing potential targeting strategies of the T3SS, including the T3SS apparatus, effector proteins and transcription factors.

Effector protein

Establishing effective symbiotic interactions between legumes and rhizobia is a complex process that necessitates multiple signal exchanges [144]. This is inseparable from the participation of different effector proteins. Effectors secreted by the T3SS of rhizobia are host-specific determinants of root nodule symbiosis. The effector protein of the type III secretion system in rhizobial is called Nops (Nodulation outer protein), and T3SS secretes many effector proteins, such as NopL, NopE, and NopC. Effector proteins of T3SSs can communicate with the host through three pathways: ubiquitination of degradative proteins or the 26S proteasome, altered RNA metabolism, and inhibition of protease activity associated with the plant defense signaling [145]. Mutations in effector proteins alter their host specificity as well.

NopD of *Bradyrhizobium sp. XS1150* was recently identified as a *Bradyrhizobium* type III

INTRODUCTION

effector with a functional N-terminal secretion signal sequence that processes SUMO and cleaves SUMO-conjugated proteins, a symbiosis-associated protein [146]. It is comprised of three domains: an N-terminal (1-390) domain, a tandem repeat (TR) domain with seven repeats (391-720), and a C-terminal protease domain (721-1017) [146]. The C-terminal protease domain of NopD *XS1150* is a SUMO protease belonging to the C48 cysteine peptidase family, similar to the sequence alignment of NopD of strain *HH103* and various other putative rhizobia effector proteins including XopD [146]. NopD can only bind to specific plant SUMO proteins with important SUMO motifs: A₃₅- R₂₉- M₇- L₆- H₅- Q₄- T₃- G₂- G₁, including AtSUMO1 and AtSUMO2 of *Arabidopsis thaliana*, GmSUMO of *Glycine max*, PvSUMO of *Phaseolus vulgaris*, these features are similar to XopD [147,148]. The cysteine residue C972 in NopD is required for its enzymatic activity as well [146].

OBJECTIVES

SUMOylation and ubiquitination are important protein post-translational modifications in all organisms. Intriguingly, as a member of USP (ubiquitin-specific protease), USPL1 has no activity for ubiquitin, while interacting with SUMO. SENP7 is the most divergent member of human SUMO-specific protease, which contains three loop insertions and has an activity for SUMO2/3. NopD is an effector from the Rhizobium type III secretory system, interestingly, it has activities for both SUMO and ubiquitin.

In this context, the general objective of this thesis has been to characterize these deSUMOylating enzyme or deubiquitinating enzymes and to study the interactions with either SUMO or ubiquitin by crystallography and biochemical assays.

Specific objectives were proposed for each chapter.

Chapter I: Structural basis for the SUMO protease activity of USPL1, an atypical ubiquitin USP family member.

- To express and purify the complex between USPL1 with human SUMO2.
- To crystallize the complex USPL1 -SUMO2.
- To solve the three-dimensional structure of the USPL1 -SUMO2 complex.
- To analyze the complex structure and characterize the role of the key residues for the deSUMOylating activity of USPL1.

Chapter II: Structural basis for SENP7 protease interactions with SUMO2.

- To express and purify the complex between SENP7 with human SUMO2.
- To crystallize the complex SENP7- SUMO2.
- To solve the three-dimensional structure of the complex SENP7- SUMO2 complex.
- To characterize the role of the critical residues of SENP7 in the SUMO2 specificity.

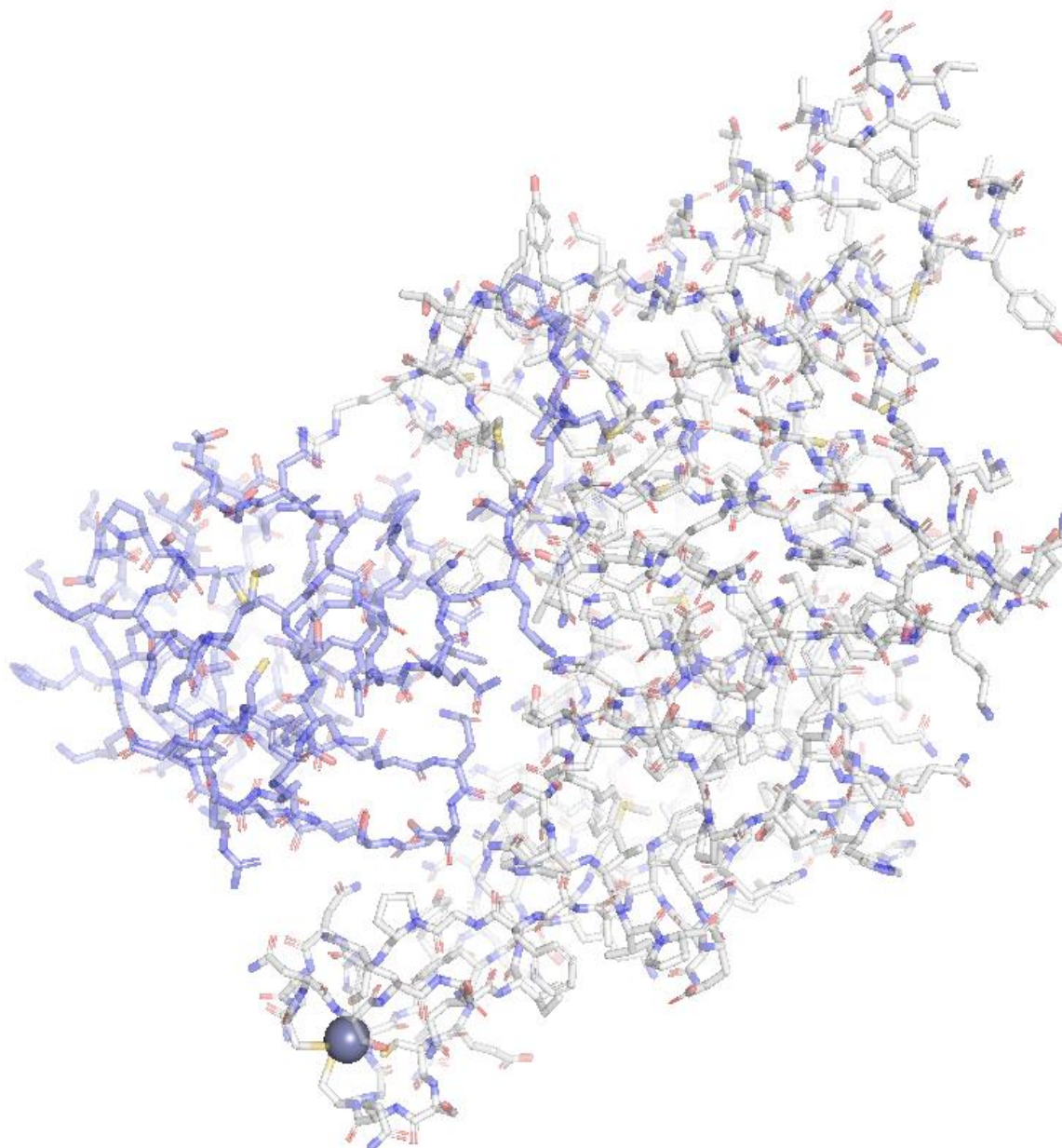
OBJECTIVE

Chapter III: Structural analysis of Bradyrhizobium NopD provides insights into the dual protease activity for ubiquitin and SUMO

- To express and purify different constructs of NopD catalytic domain.
- To crystallize the NopD-AtSUMO2 and NopD-Ubiquitin complexes.
- To solve the three-dimensional structures of the NopD-AtSUMO2 and NopD-Ubiquitin complexes.
- To analyze the interface residues of the NopD-AtSUMO2 and NopD-Ubiquitin complexes and to figure out the determinants for the dual activity of NopD.

CHAPTER I: STRUCTURAL BASIS FOR THE SUMO PROTEASE ACTIVITY OF THE ATYPICAL UBIQUITIN-SPECIFIC PROTEASE

USPL1



Introduction

Small ubiquitin-like modifiers (SUMO) belong to the family of ubiquitin-like proteins (UbLs), which are post-translational modifiers that regulate a wide plethora of protein functions in cells [75,76]. Among them, the most preeminent function of ubiquitin is the degradation of intracellular proteins by the ubiquitin-proteasome system (UPS) [1]. SUMOylation contributes to many cellular pathways and is involved in many functions, such as DNA replication, nuclear transport or DNA damage control [79–81]. SUMO, ubiquitin and all other UbL modifiers are covalently attached to target proteins by the formation of an isopeptide bond with an internal lysine residue, but before that, UbLs need to be activated by a dedicated conjugation pathway through an enzymatic cascade formed by E1, E2 and E3 enzymes [78,82]. The covalent attachment of ubiquitin and SUMO to protein targets is a reversible process through action of de-ubiquitinating (DUBs) or de-SUMOylating proteases [96,97].

DUBs can be divided into seven families, namely the Ubiquitin C-terminal hydrolases (UCHs), Ubiquitin-specific proteases (USPs), Machado–Joseph Disease protease family (MJDs), Ovarian tumor proteases (OTUs), MINDY protease family, JAMM family, and ZUFSP/Mug105 family, each of which presents a unique structural fold [7,12,13,149]. Most DUB families are cysteine proteases, except the JAMM family, which are metalloproteases. USPs constitute the largest DUB family constituted by 56 members in humans. The catalytic domain of USPs is composed of three subdomains, named thumb, palm and fingers, and the overall structure resembles the shape of a human right hand [6,58]. The active site of USPs is formed by a catalytic triad (Cys-His-Asp/Asn) located between the palm and thumb subdomains, which together with the fingers subdomain grasps ubiquitin for proteolytic cleavage [58]. Additionally, USPs are multidomain enzymes in which adjacent domains aid to achieve multiple functions, such as zinc finger domain (ZnF), ubiquitin-like domain (UbL), ubiquitin-related domain (UBA) or ubiquitin interaction motif (UIM) [6,149]. USPs are active towards ubiquitin, but there are two exceptions: USP18, which is active towards the double-headed ISG15; and USPL1, a distant member of the family specific for SUMO [62,70]. Whilst ISG15 and ubiquitin share a 30% sequence identity and the structural analysis showed a quite similar binding interface with USP18, SUMO and ubiquitin display little homology (16% sequence

identity), foreseeing the presence of a divergent binding mechanism between USPL1 and SUMO.

Until recently, SUMO proteases were only constituted by the 6 members of the SENP protease family in humans, which are similar to ULP1, the first discovered SUMO protease in *S.cerevisiae* [83] and all belong to the C48 cysteine protease structural class (SEN1-SEN3 and SEN5-SEN7) [97]. However, two novel types of SUMO proteases have been described in the recent years without sequence/structural homology to the SENP/ULP family: the deSUMOylating peptidase 1 and 2 (DES1 and DES2) [102]; and the deubiquitinating enzyme USPL1, which is a member of the ubiquitin USP family but is specific for SUMO, instead of ubiquitin [62]. An analogous finding was observed in the characterization of SENP8/DEN1, which is another member of the SUMO protease family (SEN1/ULP family), but specific for the Nedd8, another type of UbL modifier [103–106].

Full-length of USPL1 is composed of 1.092 residues containing a USP-like catalytic domain in a middle region and long, disordered protein extensions without the presence of evident globular domains. So far, the only function described for USPL1 takes place in the nucleus, where it is a component of the Cajal Bodies (CBs), co-localizing with coilin [62]. USPL1 seems to have an impact on the formation and dynamics of CBs and in cell proliferation, as observed in the USPL1-depleted cells [62]. Whilst USPL1 deletion does not affect the overall SUMOylation in the cell, it causes significant frizzled protein mislocalization and damage in cell proliferation, which interestingly does not depend on its deSUMOylase catalytic activity [71]. The CBs are membrane-less compartments involved in the biogenesis of snRNP and snoRNP, maintenance of telomeres, and processing of histone mRNA [72]. USPL1 is a low-abundant component of the CBs that plays a role in the RNAPII transcription of snRNA that is essential for cell growth [71]. Moreover, SUMO has been involved in tumorigenesis, genetic variants of USPL1 are closely related to grade-3 breast cancer [73], as well as USPL1 may also be involved in the signaling pathway of multiple myeloma [74].

USPL1 can interact with both SUMO1 and SUMO2/3, but it shows a much higher activity for the SUMO2/3 isoform [62]. The sequence of the catalytic domain of USPL1 is unequivocally related to USP members with deubiquitinase activity, probably forming the

canonical right hand-like structure with palm, thumb and fingers subdomains. The existing understanding of USPL1 is not abundant, and the determinants of the particular specificity for SUMO have not yet been determined.

In order to get insights into the interaction of USPL1 with SUMO, in this chapter we show the crystal structure of the complex between human USPL1 and the human precursor of SUMO2 at 1.8 Å resolution. Structural and biochemical analysis of the complex interface of USPL1-SUMO2 by mutagenesis analysis sheds light on key structural determinants of this unusual USP family member that has evolved to be specific for SUMO.

Results

Covalent crosslinked complex between USPL1 and SUMO2 precursor

The catalytic domain for recombinant expression of human USPL1 is based on the sequence alignment with members of the USP ubiquitin specific protease family and comprises residues Met212 through Leu514 ([Supplementary Figure 1](#)) [62]. The sequence identity of the catalytic domain among members of the USP family is typically around 15-50% and most are specific for cleaving off ubiquitin from protein targets or ubiquitin chains (or ISG15 in the case of USP18) [6]. However, USPL1 sequence identity is one of the lowest in the USP family [7], with identities ranging from 16% for USP7 to 14% for USP28, as examples of two USP members. The homology is basically observed in the secondary structure elements around the catalytic triad of the active site (Cys236-His456-Asp472) ([Supplementary Figure 1](#)). This sequence divergence with all members of the USP family reflects the existing relevant substitutions in USPL1, which evolved to cleave off SUMO rather than ubiquitin from protein targets.

The complex between USPL1 and SUMO2 was prepared by taking advantage of the formation of a covalent bond between the SUMO2 C-terminal glycine (Gly93) and the active site cysteine of USPL1 (Cys236), which stabilizes the complex, increases the binding affinity and thus the chances of crystallization. To prepare this covalent bond, we used the dehydroalanine strategy [150], which creates an electrophilic center highly reactive with nucleophiles such as internal cysteines in proteins ([Figure 1A](#)). First, in the human SUMO2 precursor, Gly93 was substituted for cysteine and Cys48 was substituted for serine, leaving only one internal cysteine at the position of the reactive Gly93. Next, the SUMO2-C48S/G93C precursor was incubated with 2,5-dibromohexanediamide, which can desulfurize the Cys93 of the SUMO precursor into dehydroalanine (DHA) at mild conditions. Finally, the SUMO2 precursor bearing DHA at position of Gly93 was incubated with USPL1 to create a covalent bond between active site nucleophile cysteine of USPL1 and the DHA electrophilic trap on the SUMO2 precursor ([Figure 1A](#)).

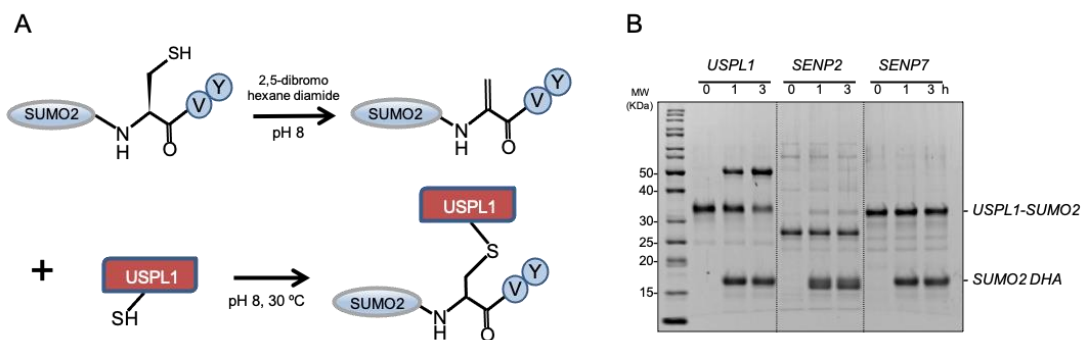


Figure 1. Covalent thioether bond formation of SUMO2-DHA with USPL1 catalytic domain. **A.** Schematic representation of the reaction to form the covalent thioether between the SUMO2-C48S/G93C precursor and USPL1 catalytic domain. **B.** Active site probe assay for USPL1, SENP2 and SENP7 with SUMO2-DHA. Time course assay of USPL1, SENP2 and SENP7 at 2 μ M using the SUMO2-DHA substrate at 6 μ M at 30°C for 3 hours. n=3 technical replicates. Source data are provided as a Source Data file.

A major concern of the DHA strategy in SUMO proteases is the possible steric hindrances in the C-terminal di-glycine motif after the substitution of Gly93 for DHA, which might clash with narrow protease binding pockets. Therefore, we initially checked the formation of the covalent crosslink after incubation of SUMO2 DHA precursor with USPL1, and two other well-characterized SUMO proteases, SENP2 and SENP7 (Figure 1B). Interestingly, only USPL1 is able to form the covalent crosslink with the SUMO2 DHA precursor, indicating a correct and specific binding between SUMO2 and USPL1. However, the two other well-characterized SUMO proteases, SENP2 and SENP7, are unable to form a covalent bond between the active site cysteine and DHA (Figure 1B). As known in the SENP/ULP protease family, the integrity of the C-terminal di-glycine motif is essential for C-terminal binding, which must be placed in a shallow tunnel formed by two tryptophan residues and the simple substitution of glycine for alanine precludes the interaction [84,151], as probably occurs in SENP2 and SENP7. On the other hand, since the reaction between SUMO2 DHA precursor and USPL1 is highly efficient, we envision a different binding mechanism for the interaction of the C-terminal tail of SUMO2 with the active site groove of USPL1, compared to the SENP protease family.

Overall structure of the USPL1-SUMO2 complex

The USPL1-SUMO2 complex was formed by incubation of USPL1 with the SUMO2 DHA precursor at 37 °C for 2 hours to form the covalent bond between the DHA group with the active site cysteine of USPL1 (Supplementary Figure 2). A few number of diffraction

quality crystals of the USPL1-SUMO2 complex were obtained in the initial screening in a condition containing 0.2 M potassium thiocyanate, 0.1 M sodium acetate pH 5.5, 8% w/v PEG20000 and 8% w/v PEG500MME, which was hard to reproduce in subsequent screenings. Molecular replacement with available USP models did not work due to the lower sequence identity with USPL1 (less than 15%), but fortunately we were able to solve the structure by using the recently reported USPL1 model from the alpha-fold server [152], which unambiguously resulted in a correct final solution. The crystals contained one USPL1-SUMO2 complex per asymmetric unit and diffracted to a resolution of 1.8 Å. The final electron density map model of the USPL1 catalytic domain includes most of the residues (Ser225 to Ile501), with the only disruption of a disordered loop connecting two alpha helices in the thumb subdomain (Leu285 to Lys295).

USPL1 adopts the typical fold of the catalytic domain of the USP family, resembling the shape of a right hand containing palm, fingers and thumb subdomains (**Figure 2B & Supplementary Figure 3**). A Zn²⁺ atom coordinated by four cysteines stabilizes the structure of the finger domain, as occurs in several USPs, forming a zinc-finger motif (ZnF). In addition to the general structural role of the ZnF motif, in USP21 it is also important for the binding of the distal ubiquitin in a linear diUb substrate [153]. SUMO2 is grabbed by the USPL1-like right hand in the complex, with the C-terminal tail of the SUMO2 precursor extended towards the catalytic triad of the USPL1 active site formed by Cys236, His456 and Asp472 (**Figure 2B&C**). Structural comparison between USPL1 in complex with SUMO2 with the apo form of USPL1 from the alpha-fold model indicates a very good overlapping of both structures, with a main-chain rmsd (root mean square deviation) of 0.8 Å (**Figure 2C**). The structural overlapping also revealed similar location of residues forming the active site catalytic triad, with a 3.8 and 3.6 Å distance between the Cys236 S γ and the His456 N δ 1; and 2.8 and 2.9 Å between His456 N ϵ 2 and Asp472 O δ 1, in the unbound and SUMO2-bound USPL1, respectively (**Figure 2C**). Such distances suggest that the USPL1 catalytic triad is already preformed and the protease might be active in the absence of the SUMO substrate, in contrast to the apo structures of other USPs, such as USP7, USP15 and USP18, in which binding of the ubiquitin substrate is necessary to rearrange the catalytic triad to an active conformation [58,70,154].

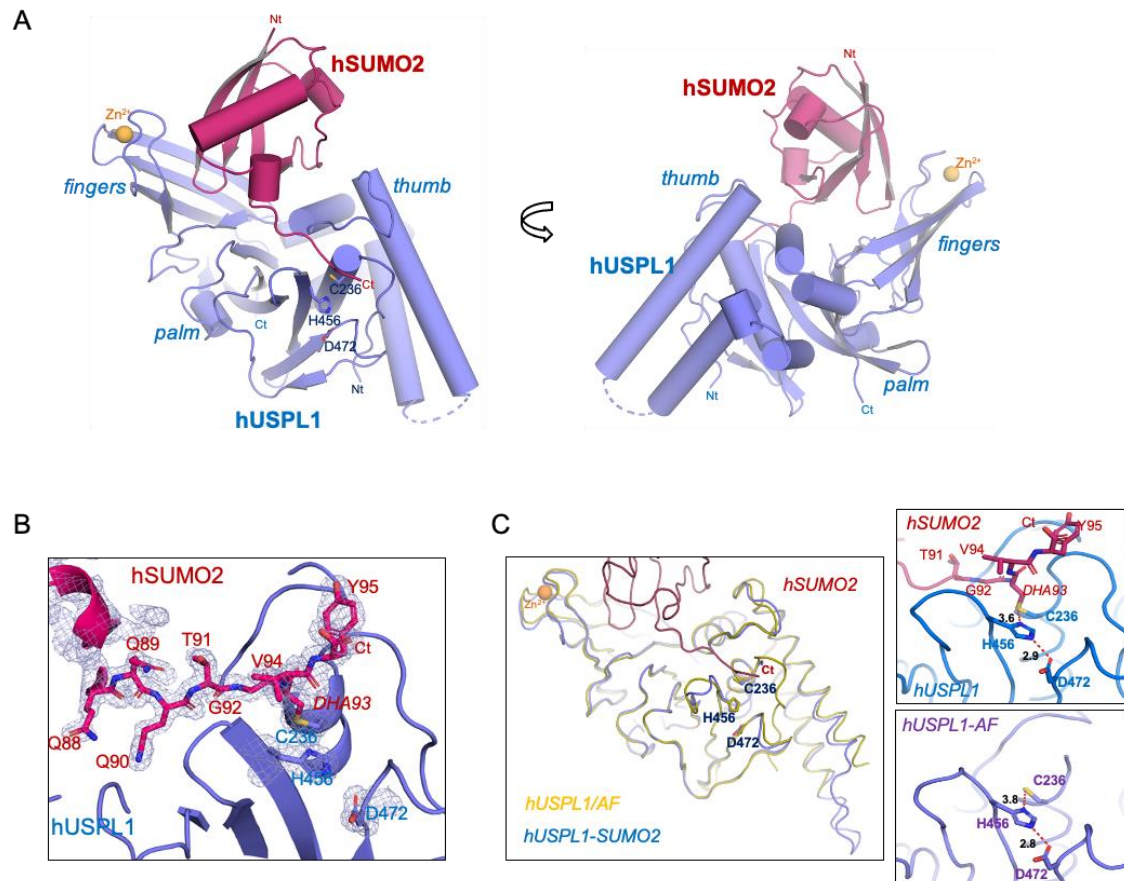


Figure 2. Crystal structure of the complex of human USPL1 with SUMO2 precursor. **A.** Two views of the USPL1-SUMO2 complex structure are shown in cartoon representation. USPL1 catalytic domain and SUMO2 precursor are shown in purple and red, respectively. USP right hand-like domains are labeled. The catalytic residues are labeled and depicted in stick representation. Zinc atom is shown in yellow. **B.** Detailed view of the 2Fo-Fc electron density map contoured at 1σ of the C-terminal tail of SUMO2 bound to USPL1. Same color code as in (a). **C.** Structural superimposition between USPL1-SUMO2 (blue) complex and the USPL1 AlphaFold (AF) model (yellow). Close-up views of the superimposition of active sites of USPL1 and AF model are shown in the right panel. Catalytic triad residues from USPL1 are labeled and shown in stick representation, whereas the corresponding residues from AF model are shown below. Hydrogen bonds are represented by dashed lines and the distances (Å) are also depicted.

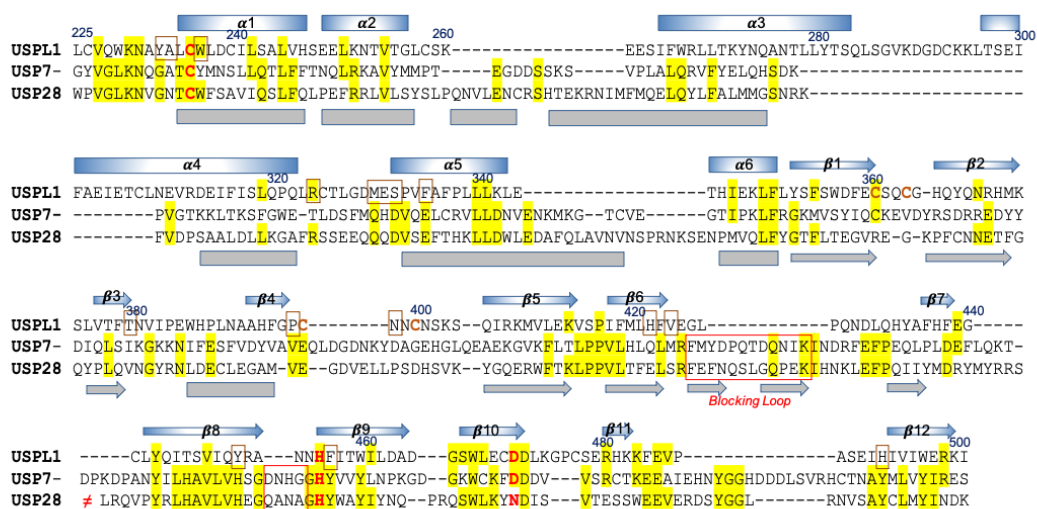
Structural comparison with ubiquitin-specific USP members

As mentioned above, the sequence identity between USPL1 and the other USP members with deubiquitinase activity is only around 15%, which differs from the higher sequence identities shown among the ubiquitin-specific USP members (Figure 3A & Supplementary Figure 1). This might be a consequence of the deSUMOylase activity displayed by USPL1 in contrast to the deubiquitinase activity for all other USPs. Structural overlapping between USPL1 with USP7 (rmsd 3,02 Å, 197 aligned, 14,21% identity) and with USP28 (rmsd 2,61 Å, 184 aligned, 14.67% identity) are low, only showing a good superposition in the secondary structure elements encompassing the catalytic triad in the active site in

the palm subdomain, namely the $\alpha 1$ helix (Cys236) and the $\beta 9$ and $\beta 10$ strands (His456 and Asp472) (Figure 3A&B). In general, the thumb and fingers subdomains of USPL1 show very little homology with USP7 and USP28, particularly in the different length and orientation of $\alpha 3$, $\alpha 4$ and $\alpha 5$ helices in the thumb subdomain, and in the beta sheet containing the ZnF motif in the fingers subdomain [58,155].

Interestingly, a major difference in USPL1 is the lack of two loops in the *palm* subdomain, namely so-called blocking loops, between strands $\beta 6$ and $\beta 7$, and between $\beta 8$ and $\beta 9$ next to the catalytic triad (Figure 3A-D). These two loops are highly conserved in the USP family due to their major role in ubiquitin binding (Supplementary Figures 1 & 6) and they are structurally rearranged upon ubiquitin binding to trap ubiquitin within the right hand-like structure [58]. However in USPL1 both loops are not required for binding to SUMO2. In fact, superposition of USPL1 with USP28 displays a $\sim 15^\circ$ angle rotation of SUMO with respect to ubiquitin, which would result in a collision with the larger *Blocking Loop* (Figure 3E). Thus, in addition to the unique contacts with fingers and thumb subdomains, the lack of the blocking loops in USPL1 contributes to the different position of SUMO on the surface of USPL1 compared to ubiquitin, which is particularly evident in the different orientation of C-terminal tail backbone (Figure 3F), and in the comparison of the electrostatic potential surfaces between USPL1 and USP28 (Figure 3G).

A



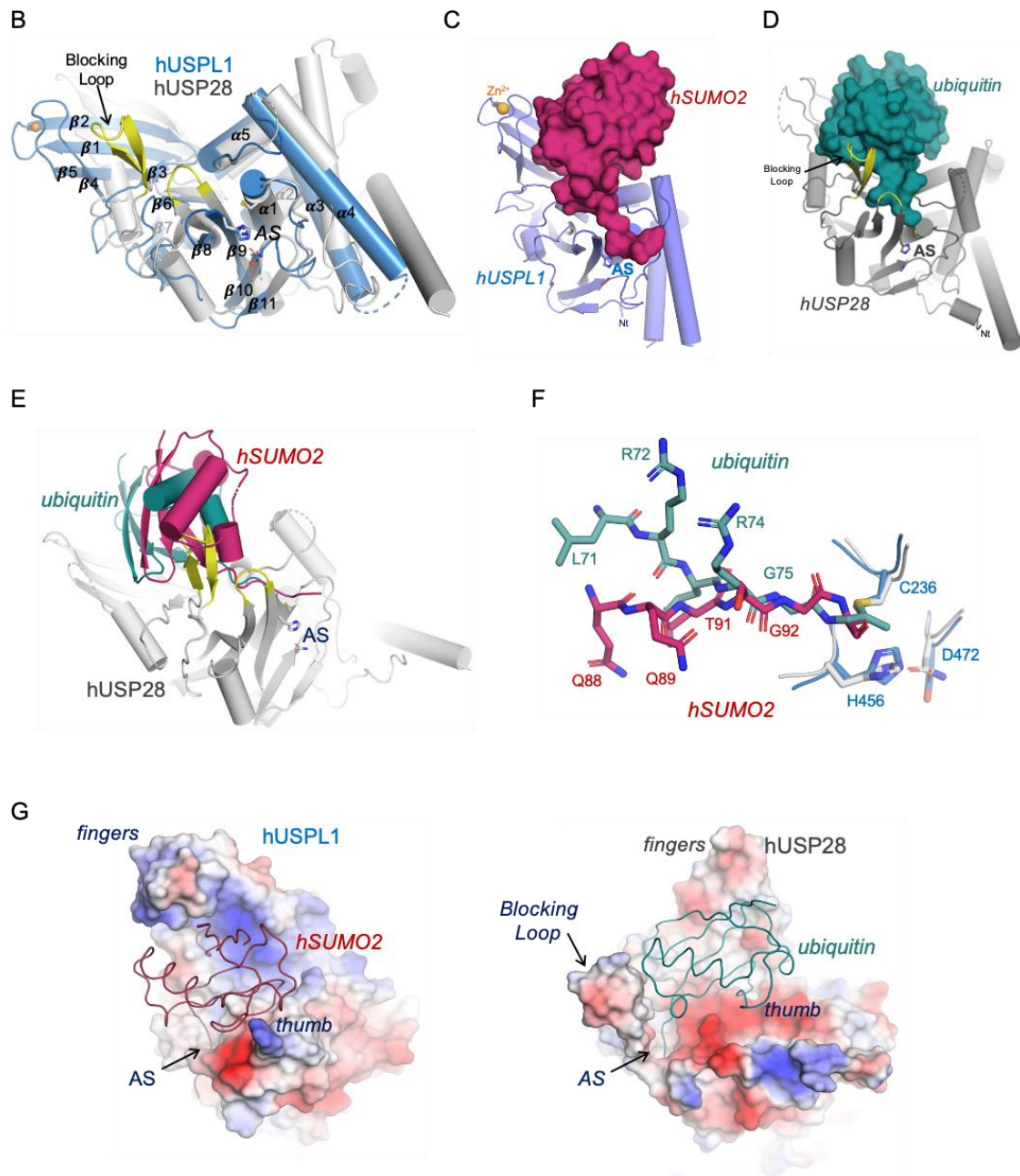


Figure 3. Comparative analysis of USPL1 with the ubiquitin USP28. **A.** Alignment of sequences corresponding to the catalytic domains for human USPL1, USP7 and USP28 based on the structural alignment of human USP7 (PDB code 4WPI) and USP28 (PDB code 6HEJ) [155,156]. Secondary structure elements are numbered (β -strands and α -helices) and indicated above the alignment for USPL1 (blue) and below the alignment for USP28 (grey). Gaps are denoted by -- and the large sequence insertion within USP28 is depicted \neq to indicate that the sequence is missing from the alignment. Side chain identity (100% conservation) is denoted in the alignment by a yellow background. Three catalytic residues are depicted in red. Active site blocking loops are denoted by red squares. All images were prepared with PyMOL [157]. **B.** Cartoon representation of the structural overlapping of the catalytic domains of USPL1 with USP28 (PDB code 6HEK) [155]. The blocking loops connect β -strand 6 and β -strand 7, and β -strand 8 and β -strand 9 of the palm domain are missing in USPL1 and shown in yellow. Active site residues (AS) are shown in stick representation. Zinc atom is shown in yellow. **C.** Surface representation of SUMO2 (red) in complex with USPL1, shown in cartoon representation (blue). Active site residues (AS) are shown in stick representation. Zinc atom is shown in yellow. **D.** Surface representation of ubiquitin (dark green) in complex with USP28 (PDB code 6HEK) [155], shown in cartoon representation (grey). Active site residues (AS) are shown in stick representation. Zinc atom is shown in yellow. Blocking loops (yellow) are indicated. **E.** Structural comparison of C-terminal tails of SUMO2 (red) and ubiquitin (green) in complex USPL1 and USP28, respectively. **F.** Structural overlapping of SUMO2-USPL1 and ubiquitin-USP28 displaying the collision of the blocking loop (yellow) of USP28 (grey) with SUMO2 (red). **G.** Electrostatic potential surface

representation for USPL1-SUMO2 and USP28-ubiquitin to highlight the differences between USPL2 and USP28 in the analogous surface. SUMO2 (red) and ubiquitin (dark green) are shown in a line representation.

SUMO2 interface of the C-terminal tail with USPL1

As expected by the divergent C-terminal sequences of SUMO (-FQQQTGG) and ubiquitin (-VLRLRGG), the contacts engaged by the SUMO C-terminal tail with the active site cavity of USPL1 constitute key signatures for the specificity of USPL1 for SUMO2. The electron density maps clearly show the covalent bond formed between DHA93 in SUMO2 and Cys236 in USPL1 (**Figure 2B**), in which the sidechain and not the C-terminal carboxylate is crosslinked to Cys236 *Sy*. This fact constrains the geometry of this region, as observed by the *cis* configuration of the Gly92-DHA93 peptide bond in the USPL1 complex (**Figure 4**), in which strong hydrogen bonds between the nitrogen and carbonyl oxygen of the Gly92-DHA93 peptide bond are established with the USPL1 backbone (each 2.9 Å distance) (**Figure 4**).

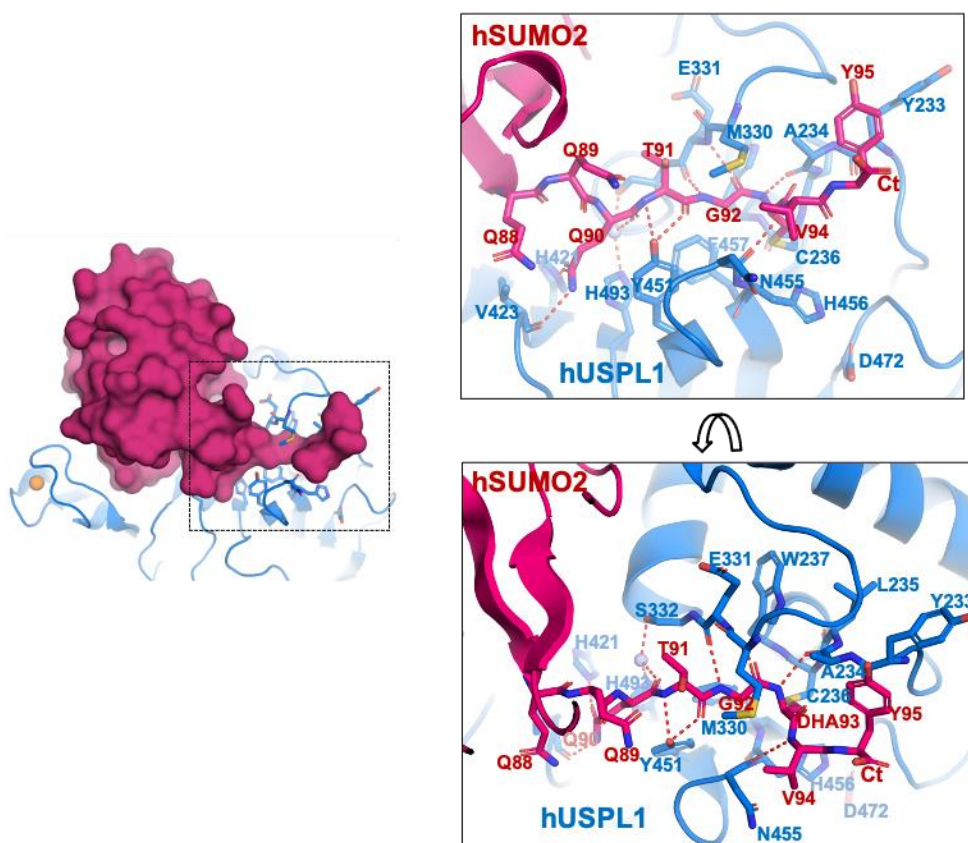


Figure 4. Atomic details of the C-terminal tail interaction of SUMO2 with USPL1. *Right*, two views of the stick representation of the main contacts of the C-terminal of SUMO2 in complex with the active site groove of USPL1. USPL1 catalytic domain and SUMO2 are shown in blue and red, respectively. SUMO2 and USPL1 interface residues are labeled.

Dashed lines indicate hydrogen bond contacts. *Left*, surface representation of SUMO2 (red) in complex with USPL1, shown in a blue cartoon. Interface residues are shown in stick representation. Zinc atom is shown in yellow.

The extended conformation of the SUMO2 tail (Gln90-Thr91-Gly92) is comparable to ubiquitin, but the specific backbone hydrogen bonds are different. Whereas Gly92 engages similar hydrogen bonds with USPL1 as in the ubiquitin-USP28 complex, Thr91 forms two hydrogen bonds with the side chain of Tyr451 O ζ (3.2 and 2.7 Å distance for the backbone N and O, respectively) (**Figure 4**). In ubiquitin USPs Tyr451 is occupied by a highly conserved histidine (**Supplementary Figure 4**), forming a similar hydrogen bond with the backbone oxygen of ubiquitin Arg74, but in this case the backbone nitrogen interacts with the β 8- β 9 blocking loop, absent in USPL1 (**Supplementary Figure 4**).

The backbone oxygen of Gln90 forms a hydrogen bond with a water molecule that is fixed by contacts with His493 and Ser332, both highly conserved in USPL1 (**Figure 4 & Supplementary Figure 1**). In ubiquitin the backbone oxygen of Leu73 (equivalent to Gln90 in USPL1) forms a hydrogen bond with the side chain of a tyrosine (replaced by Phe457 in USPL1) (**Supplementary Figure 4**). Such tyrosine, located one position after the active site histidine, is highly conserved in all ubiquitin-specific USPs (**Figure 3 & Supplementary Figure 1**), but has been substituted by phenylalanine in all USPL1 orthologs, which forms hydrophobic contacts with Gly92 but is unable to establish a hydrogen bond as in ubiquitin. Additionally, the side chain of Gln90 forms a hydrogen bond with a conserved His421 (3.0 Å distance) and with the backbone oxygen of Val423. Finally, the C-terminal tail of SUMO2 (Gly92) is sandwiched between Phe457 and Met330, substituted by a highly conserved glutamine in ubiquitin-specific USPs.

A relevant difference compared to the ubiquitin-specific USPs complexes is the extensive interface established by the two arginines and the two leucines in the ubiquitin tail (Leu71-Arg72-Leu73-Arg74), which participate in electrostatic and hydrophobic contacts on opposite sides of the ubiquitin tail. Particularly relevant is the interaction of the two ubiquitin leucines with the blocking loops, absent in USPL1, and the ubiquitin Arg72 and Arg74 with Glu258 and Gln254 in the ubiquitin-USP28 complex (**Supplementary Figure 4**) (PDB:6HEK) [155]. In contrast, in USPL1 only Gln90 from the equivalent Gln88-Gln89-Gln90-Thr91 tail is engaged in a specific interaction with His421, indicating that this

interface might be less relevant in USPL1 for SUMO binding.

The complex with USPL1 was formed with the SUMO2 precursor, which contains the complete or immature C-terminal tail formed by Val-Tyr extension after the proteolytic cleavage site (di-Glycine motif). Both residues at the SUMO2 C-terminal tail are well observed in the electron density maps, but probably their structural conformation is constrained by the covalent crosslink between DHA93 and the USPL1 Cys236.

SUMO2 Interface with the *thumb* and *fingers* subdomains

Two major specific contacts stand out in the interaction between the USPL1 thumb subdomain and the SUMO2 surface, namely Arg324 and Phe335. Arg324 is engaged in a strong well-oriented electrostatic interaction with Asp71 of SUMO2, with 2.7 and 3.1 Å distances between the NH1 and NH2 of the Arg324 guanidinium group and the OD1 and OD2 of the Asp71 carboxylate group, respectively ([Figure 5A](#)). Arg324 is well conserved in all USPL1 orthologs and this interaction is unique to USPL1, not present in ubiquitin-specific USPs. Interestingly, SUMO2 Asn68 and Asp72 were important determinants for the specificity of SENP7 for SUMO2/3 over SUMO1 [115,158], where they are substituted by alanine and histidine, respectively. In USPL1, the proteolytic activity against SUMO1 is low compared to SUMO2 ([Figure 6](#)), however, removal of the side chain of Arg324 did not increase the activity for SUMO1 substrates, as observed in activity assays with the USPL1 R324A point mutant ([Supplementary Figure 5](#)).

Phe335 is buried in a hydrophobic groove in the SUMO2 surface formed by Pro66, Phe87 and the aliphatic chains of Arg59 and Arg61 ([Figure 5A](#)). Phe335 is highly conserved in all USPL1 orthologs and not observed in other USPs. In ubiquitin-specific USPs this location is occupied by a conserved glutamate (Glu258 in USP28), which is engaged in a strong electrostatic bridge interaction with Arg72, from the ubiquitin C-terminal tail ([Supplementary Figure 4](#)). Both Phe335 in USPL1 and Glu258 in USP28 probably have an important impact on the binding and specificity of SUMO2 or ubiquitin, respectively.

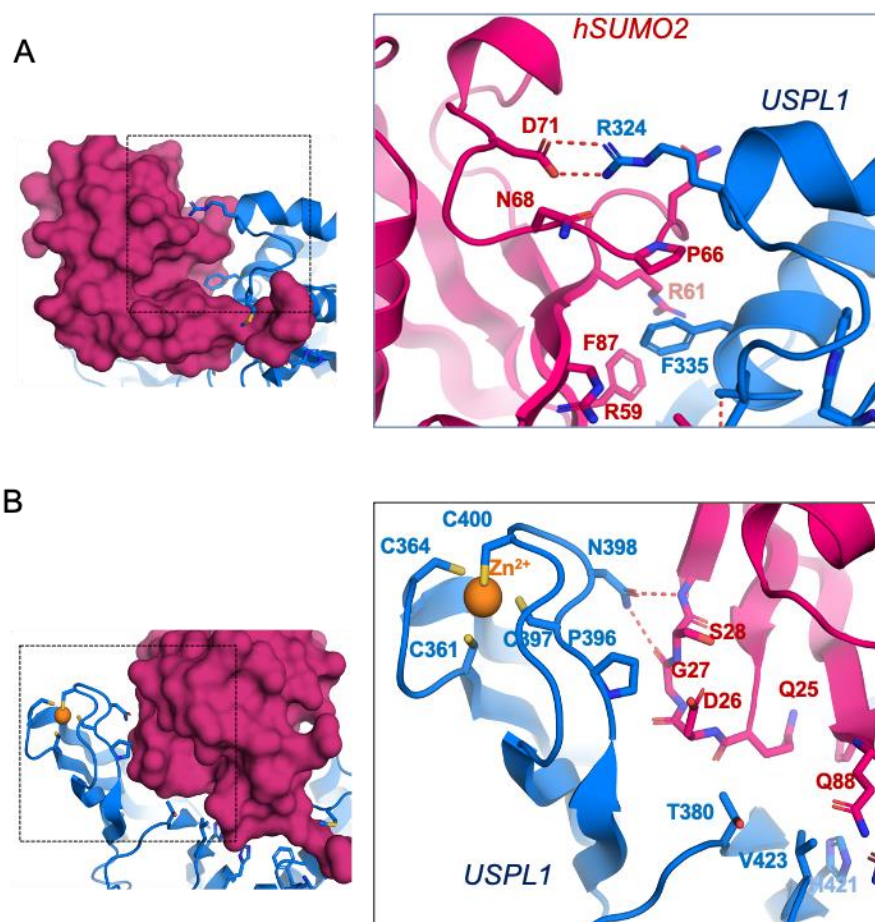


Figure 5. SUMO2 interface with the thumb and fingers subdomains of USPL1. **A.** *Right*, stick representation of the main contacts of the *thumb* subdomain of USPL1 with SUMO2. USPL1 catalytic domain and SUMO2 are shown in blue and red, respectively. SUMO2 and USPL1 interface residues are labelled. Dashed lines indicate hydrogen bond contacts. *Left*, surface representation of SUMO2 (red) in complex with USPL1, shown in a blue cartoon. Interface residues are shown in stick representation. Zinc atom is shown in yellow. **B.** *Right*, stick representation of the main contacts of the *fingers* subdomain of USPL1 with SUMO2. Same color code as in (a).

In the USPL1 fingers subdomain, a Zn²⁺ atom stabilizes the structure by the coordination of four conserved cysteine residues (Figure 5B), and as observed in other USPs, the cysteine coordination to Zn²⁺ is essential for the USP catalytic activity. In USPL1, the contacts are basically engaged by the β 1- β 2 hairpin loop of SUMO2, which nicely fits in the fingers subdomain surface. The only specific side chain contact is engaged by Asn398, located next to the Zn²⁺ site and conserved in all USPL1 orthologs, which forms two hydrogen bonds with the main chain oxygen and nitrogen of Gly27 and Ser28 in SUMO2, respectively (2.8 and 3.0 Å distances). In ubiquitin-USP complex structures, the fingers subdomain normally display a rigid body adjustment to fix ubiquitin, sometimes having an impact in chain specificity, as occurs in USP21 [153]. However, in the USPL1-SUMO2 complex, the orientation of the fingers subdomain is similar to the AlphaFold-2 model of

the apo form (Figure 2) and displays a different orientation compared to all known structures of ubiquitin USPs, showing an average 14 Å displacement between the Zn²⁺ atoms of the ZnF motifs (Supplementary Figure 6). As a consequence, the contacts observed at the interface of the fingers subdomain with SUMO2 are not observed in the other USP-ubiquitin complexes, basically due to the different orientation of the fingers subdomain with respect to SUMO2.

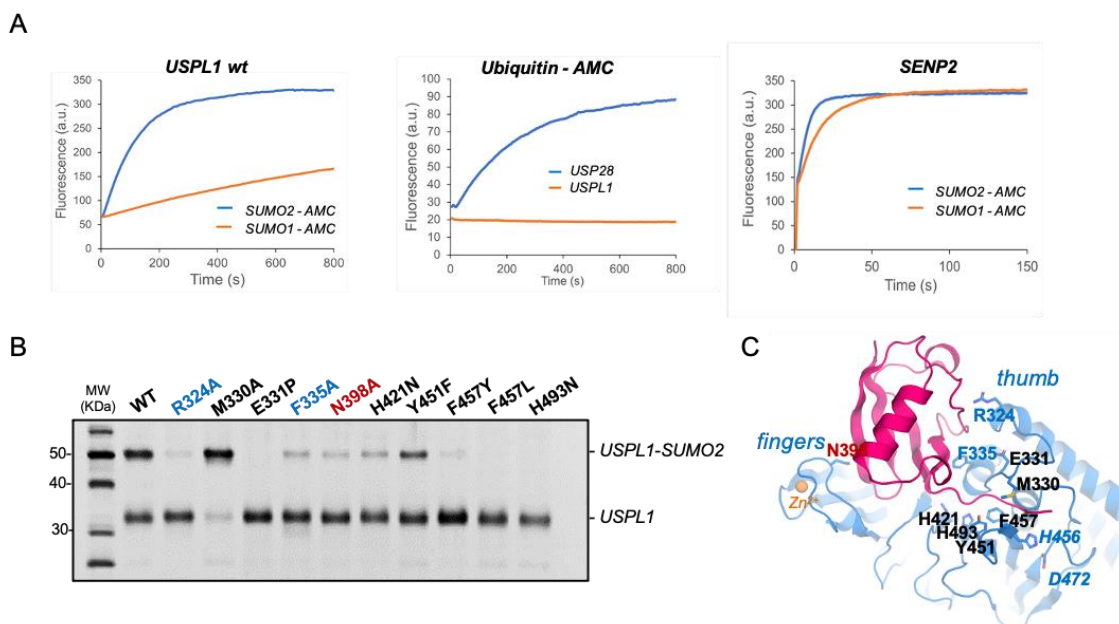
Mutagenesis analysis of the specific interface contacts in USPL1

USPL1 shows a proteolytic isoform preference for human SUMO2, in contrast to other SUMO peptidases such as SENP2 which shows similar activities for SUMO1 and SUMO2-AMC substrates (Figure 6A) [84]. As expected, USPL1 does not show any activity against ubiquitin-AMC substrate (Figure 6A) [62]. In order to characterize the binding interface in this kinetically trapped intermediate complex, specific interactions in the USPL1-SUMO2 interface have been mutated. The overall structural integrity of all USPL1 mutants are similar to the wild type form, as observed by comparing their purification profiles by ion-exchange chromatography and by the similar spectra displayed in the intrinsic Trp-fluorescence emission analysis (Supplementary Figure 7). The USPL1 mutant interface analysis has been conducted in the C-terminal tail, thumb and fingers subdomains (Figure 6C). Binding to SUMO2 has been checked by using a SUMO2 DHA chemical trap substrate, and the proteolytic activity has been checked against the SUMO2-AMC fluorescent substrate, as well as by RanGAP-SUMO2 and diSUMO2 substrates.

In the C-terminal tail interface, Phe457 has been replaced either by tyrosine, which is present in all USP members, or by leucine, which maintains the hydrophobic character of the interaction. In both F457Y and F457L mutants the binding reaction with SUMO2 DHA is strongly diminished and the catalytic activity against all tested substrates reduced (Figure 6B-F). Interestingly, despite that all other USP members contain a tyrosine in that position, the only presence of a O ζ from the tyrosine side chain seriously compromises the interaction of USPL1 with SUMO2 substrates. The SUMO C-terminus is sandwiched between Phe457 and Met330 (Figure 6C), however removal of Met330 side chain has a minor effect on the catalytic activity and even seems to increase the binding affinity for

SUMO2, as observed in the M330A mutant (Figure 6B). While Met330, which is normally substituted by glutamine in ubiquitin USPs, may still be involved in regulating USPL1 activity, it does not seem to be essential for binding and catalysis. However, the backbone interactions of the adjacent Glu331 with the SUMO C-terminal tail are essential for binding and catalysis, as observed in the E331P mutant, in which a proline substitution distorts the backbone orientation and removes a critical hydrogen bond (Figure 6B-F).

Removal of the O ζ from the tyrosine side chain in the Y451F mutant decreases binding and activity by four-fold compared to wild-type (Figure 6B-F), highlighting the role of the two hydrogen bonds with Thr91 of the C-terminal tail of SUMO2 (Figure 4). Finally, two conserved histidine residues were analyzed: H421N removes the specific hydrogen bond contact with Gln90, and H493N removes the interaction to a fixed water (Figure 4). Interestingly, whereas the H421N mutant affects partially binding and catalysis, the integrity of the His493 side chain, which bridges the C-terminal SUMO backbone through a water molecule, seems essential for binding and catalysis. The overall structural integrity of the H493N mutant seems correct, as observed in Trp-fluorescence stability analysis compared to wild-type (Supplementary Figure 7), thus highlighting the relevance of this water-bridged interaction in the overall hydrogen bond network of the C-terminal tail.



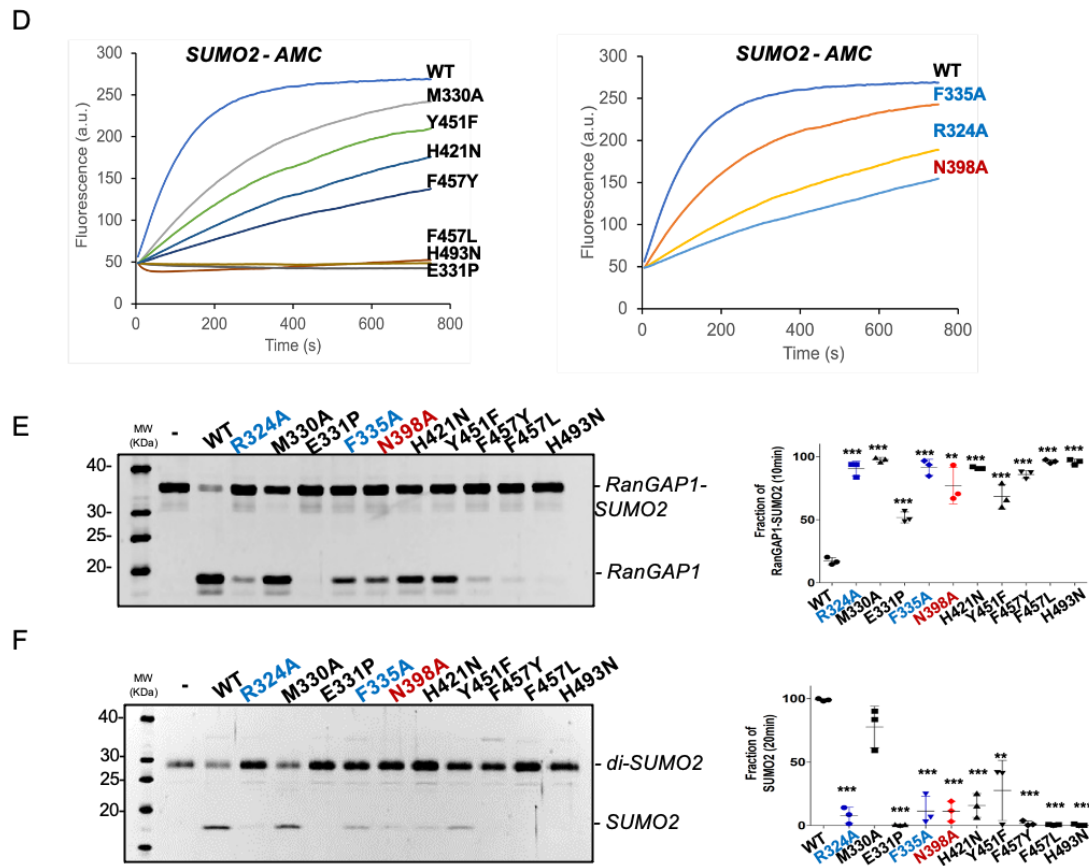


Figure 6. Functional analysis of the SUMO2-USPL1 interface. **A. Left**, activity assay of USPL1 with SUMO1-AMC and SUMO2-AMC. **Middle**, activity assay of Ubiquitin-AMC with USPL1 and USP28. **Right**, activity assay of SENP2 with SUMO1-AMC and SUMO2-AMC. Reaction were conducted in triplicate and the average curve is displayed. **B.** Binding interaction of USPL1 point mutants with SUMO2-DHA. Reaction assays were conducted with USPL1 wild-type and mutants for 3 hours. $n=3$ technical replicates. **C.** USPL1-SUMO2 complex structure shown in cartoon representation. USPL1 catalytic domain and SUMO2 precursor are shown in purple and red, respectively. USP right hand-like domains are labeled. Interface residues are labeled and shown in stick representation. Zinc atom is shown in yellow. **D. Left**, activity assays of USPL1 wild type and mutants of the C-terminal tail interface using SUMO2-AMC. **Middle**, similar activity assays of USPL1 wild type and mutants of the thumb and fingers subdomains. **Right**, table indicating the mean slope values plus/minus the standard deviation of the activity assays. $n=3$ technical replicates. Significance was measured by a two-tailed unpaired t-test relative to wild-type. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. **E. Left**, endpoint assays of USPL1 wild type and mutants using the RanGAP1-SUMO2 substrate. **Right**, plot of the RanGAP1-SUMO2 fraction after 10 minutes reaction. Data values represent the mean \pm SD, $n=3$ technical replicates. Significance was measured by a two-tailed unpaired t-test relative to wild-type. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Exact P values from left to right: <0.0001 , 0.0868 , <0.0001 , 0.0002 , <0.0001 , 0.0001 , 0.0065 , <0.0001 , <0.0001 , <0.0001 . **F. Left**, endpoint assays of USPL1 wild type and mutants using di-SUMO2 substrate. **Right**, plot representation of the product SUMO2 fraction after 20 minutes reaction. Data values represent the mean \pm SD, $n=3$ technical replicates. Significance was measured by a two-tailed unpaired t-test relative to wild-type. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Exact P values from left to right: <0.0001 , <0.0001 , 0.0003 , <0.0001 , 0.0021 , <0.0001 , 0.0009 , <0.0001 , <0.0001 , <0.0001 . Source data for figures 6a, b, d, e and f are provided as a Source Data file.

Two specific contacts to SUMO2 have been analyzed in the USPL1 thumb subdomain, Arg324 and Phe335, both highly conserved in the USPL1 orthologs ([Supplementary Figure 1](#)). The R342A mutant removes a strong and well-oriented electrostatic bridge with Asp71 of SUMO2 ([Figure 5A](#)), and the results indicate that binding and catalysis with SUMO2

substrates are seriously compromised, highlighting a major role for Arg324 (Figure 6B-F). Likewise, in a lesser degree compared to Arg324, the F335A mutant also disturbs binding and catalysis for SUMO2 substrates. Thus our *in vitro* proteolytic activities as well as the strong conservation in the family indicate that both Arg324 and Phe335 play a major role in the specific interaction of SUMO2 with USPL1.

The structural architecture of the fingers subdomain is maintained by four conserved cysteine coordination residues forming a ZnF motif (Figure 5B). Despite the extended interface of the fingers subdomain with SUMO, only the highly conserved Asn398 seems to establish specific contacts by the formation of two hydrogen bonds with SUMO2 backbone atoms (Figure 5B). Interestingly, removal of these interactions in the N398A mutant seriously compromises binding and catalysis with SUMO2 substrates. Altogether, the specific contacts of the thumb and fingers subdomain, Arg324, Phe335 and Asn398, have an essential contribution to fix the globular domain of SUMO2 in a correct orientation for catalysis.

Materials and Methods

Plasmids, Cloning and Point Mutation

HA-USPL1-pcDNA3.1 was a gift from Frauke Melchior (Addgene plasmid #85760; <http://n2t.net/addgene:85760>; RRID:Addgene_85760) [62]. The catalytic domain construct pET28a-USPL1CD was amplified by PCR using Phusion polymerase and cloned into the BamHI/NotI restriction enzymes sites of pET28a vector using ligation. The USPL1 point mutants constructs were designed by different primers and were created by the QuickChange site-directed mutagenesis kit (Stratagene). All primers are shown in [Supplementary Table2](#). pET28a- Δ 14-humanSUMO2 (14 amino acids deletion) was constructed at the Sloan-Kettering Institute in New York by David Reverter. The plasmid of pET28a- Δ 14-SUMO2GCVY(C48S) has been generated by Restriction Enzyme Free PCR [159].

Protein Expression and Purification

The USPL1 CD and Δ 14SUMO2 expression constructs were transformed into *E. coli* Rosetta (DE3) cells (Novagen). Bacteria were grown at 37 °C to OD₆₀₀=0.7~0.8, and IPTG was added to a final concentration of 0.5 mM. Bacteria were grown an additional 16 h at 20 °C and harvested by centrifugation. Cell suspensions were equilibrated in 350 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10 mM imidazole, 20% sucrose, 1 mM DTT, and 0.1% IGEPAL CA-630, and cells were broken by sonication. After removing cell debris by centrifugation, proteins were separated from lysate by nickel affinity chromatography using Ni Sepharose 6 Fast Flow (GE Healthcare) and eluted with lysis buffer including 20 mM Tris-HCl (pH 8.0), 350 mM NaCl, 300 mM imidazole, and 1 mM DTT. Fractions containing the target protein were collected, diluted to 50 mM NaCl, applied to an anion exchange resin (Resource Q; GE Healthcare), and eluted with a 0-1 M NaCl gradient from 0 to 35% in 20 mM Tris-HCl (pH 8.0) and 3 mM DTT. Concentrated the protein using Amicon Ultra-30K ultrafiltration device (Millipore) and snap-frozen in liquid nitrogen prior to storage at -80 °C.

Preparation of the USPL1CD- Δ 14-SUMO2 DHA complex

After purification of $\Delta 14$ -SUMO2GCVY (C48S) protein, 2mM DTT was added as a solid to a 500ul aliquot of protein solution to reduce any contaminant disulfide and gently shaking 15min. The buffer was changed to 20mM Tris8, 150mM NaCl by PD-10 (GE Healthcare) and concentrated to 10 mg/ml and kept at room temperature for the next step. A stock solution of 2,5-dibromo hexanediamide (DHA) was prepared by dissolving 35.5mg in 418 μ l DMF. A 10-fold molar DHA (25 μ mol) was added into $\Delta 14$ -SUMO2GCVY (2.5 μ mol). The reaction mixture was incubated at room temperature for 30min and then at 37°C for 10h. The insoluble dibromide was removed using centrifugation and further purified using a Resource Q column to get $\Delta 14$ -Sumo2 DHA protein. The mixture of USPL1CD and $\Delta 14$ -Sumo2 DHA (1:3 molar ratio) was incubated at 30°C for 3h. Anion exchange chromatography (Resource Q; GE Healthcare) and gel filtration chromatography (Superdex 75; GE Healthcare) were carried out to purify the USPL1CD- $\Delta 14$ SUMO2DHA complex.

Crystallization and Data Collection

The complex USPL1CD- $\Delta 14$ SUMO2DHA was concentrated to 8 mg/mL for crystallization screening in a buffer containing 20mM Tris 8.0, 170mM NaCl and 1mM DTT. Crystallization experiments were performed at 18°C by sitting drop vapor diffusion method and crystals grew up in a protein mixture with an equal volume of a condition solution containing 0.2 M potassium thiocyanate, 0.1 M sodium acetate 5.5, 8% w/v PEG20000 and 8% w/v PEG500MME. Crystals were harvested after 3 days and soaked 5-10 seconds in the crystallization buffer supplemented with 15% ethylene glycol, and then snap-frozen in liquid nitrogen to storage.

Diffraction data were collected to 1.8 Å resolution at beamline ID30B at the ESRF (Grenoble, France). Data processing was conducted by AutoProcessing with MxCUBE [160,161]. The space group was P2₁ and there was one complex per asymmetric unit. The structure of USPL1CD- $\Delta 14$ SUMO2DHA was solved by molecular replacement with USPL1 AlphaFold2 model as a search mode [152]. Following rounds of model building and refinement were carried out with Coot and Phenix [162,163]. The structure of USPL1CD-SUMO2 has been deposited in the Protein Data Bank under accession codes PDB 7P99.

SUMO-AMC hydrolysis assays

USPL1 wild type and mutants were incubated with ubiquitin-, SUMO1- or SUMO2-AMC at 30°C and measured the fluorescence emission using 345 nm excitation and 445 nm emission wavelengths using a Jasco FP-8200 spectrofluorometer. All measurements were carried out in triplicate with 1 nM USPL1 and 0.1 μM SUMO-AMC in a buffer containing 100mM NaCl, 20 mM Tris-HCl pH 8, 10 mM DTT.

In vitro de-SUMOylation assays

Protease activity was measured by incubating di-SUMO2 and NΔ419RanGAP1-SUMO2 (1 μm) with purified 20 nM of USPL1 wt and mutants at 37 °C in a buffer containing 40 mM Tris-HCl (pH 8.0), 250 mM NaCl, and 2 mM DTT. Reactions were stopped after 0, 5, 10 and 20 min with SDS-BME loading buffer and analyzed by gel electrophoresis (PAGE). Proteins were detected by staining with SYPRO Ruby protein gel stain (Bio-Rad). Products were detected and quantified using a Gel-Doc machine with associated integration software (ImageLab; Bio-Rad). Fraction of analyzed bands were plot as error bar graphs with SD.

Intrinsic Fluorescence Measurements

Intrinsic fluorescence spectra were recorded using a Jasco FP-8200 spectrofluorometer. Tryptophan emission spectra were obtained by setting the excitation wavelength at 295 nm and collecting emission in the 315–400 nm range. USPL1 (wild-type and mutants) were diluted to achieve 1 μM in 20 mM HEPES pH 7.5 buffer containing 250 mM NaCl and 10 mM DTT. The temperature was set at 30 °C. For thermostability at 30 °C proteins were kept at this temperature for up to 60 min and Trp spectra were recorded at different times.

Data availability: Structure reported has been deposited in the Protein Data Bank under accession code 7P99 ([10.2210/pdb7P99/pdb](https://www.rcsb.org/entry/7P99)). Other Protein Data Bank accession codes used in this study: 6HEK ([10.2210/pdb6HEK/pdb](https://www.rcsb.org/entry/6HEK)) (USP28-ubiquitin); 2HD5 ([10.2210/pdb2hd5/pdb](https://www.rcsb.org/entry/2HD5)) (USP2-ubiquitin); 5JTV ([10.2210/pdb5JTV/pdb](https://www.rcsb.org/entry/5JTV)) (USP7-ubiquitin); 5OHK ([10.2210/pdb5OHK/pdb](https://www.rcsb.org/entry/5OHK)) (USP30-ubiquitin). All other data supporting

the findings of this study are available within the article and its supplementary information files. Source data are provided with this paper.

USPL1-SUMO2	
Data collection	
Space group	P2 ₁
Unit cell parameters (Å)	50.71, 69.88, 53.64
Wavelength (nm)	0.97625
Resolution range (Å)	41.04 - 1.79
R _{merge}	0.08 (0.47)*
R _{pim}	0.06 (0.36)*
(I/σ(I))	7.8 (2.1)*
Completeness (%)	96.5 (97.8)*
Multiplicity	2.5 (2.5)*
CC (1/2)	0.99 (0.72)*
Structure refinement	
Resolution range (Å)	41.04 - 1.80
No. of unique reflections	33540
R _{work} / R _{free} (%)	17.6 / 19.7
No. of atoms	
Protein	2823
Water molecules	200
Zn ²⁺	1
Overall B factors (Å ²)	33.99
USPL1 (Å ²)	31.10
SUMO2 (Å ²)	42.66
Zn ²⁺ (Å ²)	24.22
Water molecules (Å ²)	38.29
Rms deviations	
Bonds (Å)	0.007
Angles (°)	0.845
Ramachandran favored (%)	97.94
Ramachandran allowed (%)	1.77
Ramachandran outliers (%)	0.29
PDB code	7P99

Table 1. Crystallographic statistics of the USPL1-SUMO2 complex. *Data from the last shell in parenthesis (1.70-1.79 Å).

Supplementary information

Supplementary Table 2 – List of the used primers.

USPL1CD fw	CGCGGATCCATGCCACTGGAGAGGAAATG
USPL1CD rv	AAAGCGGCCCGCTTAAAGTGGGAAGGCAGGC
Δ14SUMO2GCVY fw	CGGCCTGGTGCCGCGCGGCAGCCATAACGATCATATTAATTTGAA GGTG
Δ14SUMO2GCVY rv	GACGGAGCTCGAATTCCGATCCCTAGTAGACACATCCCGTCTGC TGTTGG
USPL1 Phe335 to Ala fw	AGGTGATATGGAAAGCCCTGTGGCTGCATTTCCCCTGCTCTT
USPL1 Phe335 to Ala rv	AAGAGCAGGGGAAATGCAGCCACAGGGCTTCCATATCACCT
USPL1 Arg324 to Ala fw	TTAGCCTTCAGCCCCAGCTTGCATGCACATTAGGTGATAT
USPL1 Arg324 to Ala rv	ATATCACCTAATGTGCATGCAAGCTGGGGCTGAAGGCTAA
USPL1 Met330 to Ala fw	AGATGCACATTAGGTGATGCGGAAAGCCCTGTGTTTGCAT
USPL1 Met330 to Ala rv	ATGCAAACACAGGGCTTCCGCATCACCTAATGTGCATCT
USPL1 Asn398 to Ala fw	CCATTTGGTCCATGTGCCAATTGCAACAGTAAATCACAAAT
USPL1 Asn398 to Ala rv	ATTTGTGATTTACTGTTGCAATTGGCACATGGACCAAAAATGG
USPL1 Tyr451 to Phe fw	AACTTCTGTAATTCAGTTTCGAGCAAATAATCATTTTATAACAT
USPL1 Tyr451 to Phe rv	ATGTTATAAAATGATTATTTGCTCGAAACTGAATTACAGAAGTT
USPL1 His493 to Asn fw	AGTTCCTGCTTCAGAGATAAATATTGTTATTTGGGAAAG
USPL1 His493 to Asn rv	CTTCCCAAATAACAATATTTATCTCTGAAGCAGGAACT
USPL1 His421 to Asn fw	ATCTCCCATATTCATGTTGAACTTTGTAGAAGGCTTACC
USPL1 His421 to Asn rv	GGTAAGCCTTCTACAAAGTTCAACATGAATATGGGAGAT
USPL1 Glu331 to Pro fw	TGCACATTAGGTGATATGCCAAGCCCTGTGTTTGCATTT
USPL1 Glu331 to Pro rv	AAATGCAAACACAGGGCTTGGCATATCACCTAATGTGCA
USPL1 Phe457 to Tyr fw	CAGTATCGAGCAAATAATCATTTATATAACATGGATTTTAGATGCT
USPL1 Phe457 to Tyr rv	AGCATCTAAAATCCATGTTATATAATGATTATTTGCTCGATACTG
USPL1 Phe457 to Leu fw	CAGTATCGAGCAAATAATCATCTTATAACATGGATTTTAGATGCT
USPL1 Phe457 to Leu rv	AGCATCTAAAATCCATGTTATAAGATGATTATTTGCTCGATACTG

A

```

1      10      20      30      40      50      60      70      80      90
HUMAN .MNDSPKICNGPLVIGPFDIICSSLHMVQVIGKNNFDSKXVPSDEYQPCRCRCKLKAHKTFRISFOQSVFLCEDLQCHFDGSKSLNLIIPDILE
MOUSE .MTDSSLKICNGPLVIGPFDIICSSLHMVQVIGKNNFDSKXVPSDEYQPCRCRCKLKAHKTFRISFOQSVFLCEDLQCHFDGSKSLNLIIPDILE
SHEEP .MTDSPKSGNGLVIVGFCADICSSLHMVQVIGKNNFDSKXVPSDEYQPCRCRCKLKAHKTFRISFOQSVFLCEDLQCHFDGSKSLNLIIPDILE
CHICK .MTDQKTTNGLRVIQGTGICGKSTLHMVQVIGKNNFDSKXVPSDEYQPCRCRCKLKAHKTFRISFOQSVFLCEDLQCHFDGSKSLNLIIPDILE
DANRE .MTDQKTTNGLRVIQGTGICGKSTLHMVQVIGKNNFDSKXVPSDEYQPCRCRCKLKAHKTFRISFOQSVFLCEDLQCHFDGSKSLNLIIPDILE

100     110     120     130     140
HUMAN .E.CHT.....PHKPKRRS.....LSSYKDSLLEANSKTRNYIAIDGGKV.....LN.SRHN...
MOUSE .D.CPT.....PKPKRR.....LTNCRNSPLPVSKTKSRIVTDSSEPI.....VM.GKYNG...
SHEEP .D.CHT.....LETNCELDPLADSKTKNHRGIDSEQV.....LN.SSHNG...
CHICK .N.LQV.....PSYDKRRNL.....CDI.SDLSPLSESNPKQARTNNVVNQQA.....IN.TDPVV...
DANRE .GCKRSISSLPDSDSDSCPKRPREELDLVLADVSEPCDAEVNDATLPDDTVKTQTFDQSSAPAEITISLDKTEQPVSIEDIKEQPISHDATEGQPI

150     160     170     180
HUMAN .....E.VYDET.....SNLPSD.....SGQONPIRTADSLERNE.....ILADTVDMA...
MOUSE .....E.VCDDF.....SASFPTD.....SAHQDPASTAASVEQSE.....ALADDVVVA...
SHEEP .....E.GEYDT.....SPLGFSLSGQNFVRTADSWEQNE.....AMADTVDMA...
CHICK .....E.KSCGSDSLCIPSRVCDVLENGQKSN.SEESINQKV.....DFTTSNNG...
DANRE .SIDQTEQPVSILHTEQPIVLDLIEVKPMSVDHSEQFICID.....NTKERPVSIVHTEEQSVLDFLQVFKPMSIAHTEEKPTSPDHIEEK

190     200     210     220     230
HUMAN .....T.KDPATV.....D.VSGTGKPSQN.....E.GCTSK.....L.MP.LESKCTSFQALCQVQNNAYALC
MOUSE .....A.TEDPATV.....S.....V.TSE.....L.M.PAKSRCLPLCQTCVQNNAYALC
SHEEP .....P.EEDATMV.....D.VSGTEGSPQN.....E.RCTSE.....L.M.P.LESKCTSFQALCQVQNNAYALC
CHICK .....H.GSPPK.....NFSSRQLNS.....THSSAS.....E.LLRDNNCSNTNTELCIQVQNNAYALC
DANRE .PMSIAHTEKPFVSLVHTEDQHLSDIQTEQPIIDPTEEQPEELPAVSDDEDVLCERKEDCVSSTQDEMEVLSLSSSLVPHSEFLWKNENNC

240     250     260     270     280     290     300     310     320
HUMAN .WLDCLLSALVHSEKNTVTG.....LCSKESIFWRILFKMNOANTLQYTSQLSGVKDGDCKKLTSEIFAETCNEVDEDFISLOPQPR
MOUSE .WLDCLLSALVHSEKNTVTG.....ACSRKCVFGRFLEKHQADELHHTHLHGVTGDECKKLTSEIFAETCNEVDEDFISLOPQPR
SHEEP .WLDCLLSALVHSEKNTVTG.....LCSKESIFWRILFKMNOANTLQYTSQLSGVKDGDCKKLTSEIFAETCNEVDEDFISLOPQPR
CHICK .WLDCLLSALVHSEKNTVTG.....E.DDEKCLLQRLTKLSQATVLENTCKRQKV.....KDLFPKAEILLNINQVLFQLOPQPR
DANRE .WLDCLLSALVHSEKNTVTG.....E.DDEKCLLQRLTKLSQATVLENTCKRQKV.....KDLFPKAEILLNINQVLFQLOPQPR

330     340     350     360     370     380     390     400     410     420
HUMAN .C.LGDMSPVFPFLLKLELHIEKLELQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK
MOUSE .C.LGDMSPVFPFLLKLELHIEKLELQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK
SHEEP .C.LGDMSPVFPFLLKLELHIEKLELQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK
CHICK .C.LGDMSPVFPFLLKLELHIEKLELQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK
DANRE .C.LGDMSPVFPFLLKLELHIEKLELQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK

430     440     450     460     470     480     490     500     510
HUMAN .H.VEGLPQRNDLQVADHFEGLQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK
MOUSE .H.VEGLPQRNDLQVADHFEGLQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK
SHEEP .H.VEGLPQRNDLQVADHFEGLQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK
CHICK .H.VEGLPQRNDLQVADHFEGLQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK
DANRE .H.VEGLPQRNDLQVADHFEGLQVTSVYQQA.NNHFTILLDADGSLWLEDDDLKGPCEAKRHRVTCVPPASPHIVIERKRSQVDEKAAACLPKMK

520     530     540     550     560     570     580     590     600     610
HUMAN .TNDCHALSNEKPVSLTSCVGDAAASAEASVTHPKDISVAPRILSQDVAVTHDHL.LSGPKGLVDN.IPLPTDEETOKTASVSLQSEAFLEENK
MOUSE .FNVQVSGEQPTCPALCSLAGTSEP.SVAHPTSMAGAQTIIPQAVARQSV.LSAGKGVDS.IPLPTDEETOKTASVSLQSEAFLEENK
SHEEP .TSDQVTFGNEKQASPAACSTGKAASAEPSSTHTPDVSMVPTLSQDEAVAPCHHS.LPDLKGSVDS.IPLPTDEETOKTASVSLQSEAFLEENK
CHICK .IEDSTANNVQLKSTVHLKCGFDDAVDNML.AEDHQEDTVRIPDKK.QQRVAECSSARCGLENLARGDLVTDIET.PLSESGKSMLEK
DANRE .PEVFNADDEHPPKLSDSVATDTCVISAL.....TVEDTASSIADTSIGSTLLDTEFGLTHKDIIVTLTVNS.....LSEPKNEPR

620     630     640     650     660     670     680     690     700
HUMAN .PVAEHTGLIKNTLLSQESLM..ASSVSAFCNEKLIQDFVDI..SFFSQV..VNTNMQS.....VQLNTEDEVNTSNNVNT.....DATGIGI
MOUSE .PVAEHTGLIKNTLLSQESLM..ASSVSAFCNEKLIQDFVDI..SFFSQV..VNTNMQS.....VQLNTEDEVNTSNNVNT.....DATGIGI
SHEEP .PVAEHTGLIKNTLLSQESLM..ASSVSAFCNEKLIQDFVDI..SFFSQV..VNTNMQS.....VQLNTEDEVNTSNNVNT.....DATGIGI
CHICK .KVVENNVNLTGSQKESAF.....SPNPTCGEVIATDLAMNDKMLSENSSMCLSLQELNPASTSSVPKKDFNPSDSSLAQSTGDRAMING
DANRE .PMR.....PGF..VSAFRHCFP.....E.....ASNLG

700     710     720     730     740     750     760     770
HUMAN .VKSIV.....EIEKDAQLKFLTPKTEQLKPER.VTSQVSNLKKKTTADSQTTTSSK...LQNQSLKENQKPFVGSWVKGISRGAEMPLCVSA
MOUSE .LKSIAATEK.....DSQTQLLPKTEKLDPEQPKSASNLKRRKTTASSKTVAAARS...AQNQPKRDEKQKRAFVGSWVKGLSRGGAPMPLCVLS
SHEEP .VKSIVTEGTVALERMAISLQKLVFTEKLNPEQHVTSQVSDWKEETATFSQTVAKP...LQNPFK..EKKPEETKQVGSWVKGLSRGGAPMPLCVLS
CHICK .E.HGL.....SSGQ.QICELPPEVENVV...QKSPDLTGASKTAVHQQVASSVDNNSFPDSDKQKRFVGSWVKGLSRGGAPMPLCVLS
DANRE .IKFT.....G.....S...RLSTPP.....IPKSSLVHXP.....EVAAAASVSKTHLQPTS.....LQKHPSPQSTPIRP

780     790     800     810     820     830     840     850     860
HUMAN .HN.....RNTITDLQPSVKGNNPQGGKTKGINQKASHVSKKARKSASKPPPIKPPAGPPSSNGTAAH..PHAAASEVLEKSGSTCGA
MOUSE .QS.....R.AVSDLQPSVKGNNPQGGKTKGISRRSRKASRKAHMEELSPRNSPPLSWTAAIT...QA.AENATSALLREGEQSRPA
SHEEP .HN.....RNTITDLQPSVKGNNPQGGKTKGISRRSRKASRKAHMEELSPRNSPPLSWTAAIT...QA.AENATSALLREGEQSRPA
CHICK .PK..NERSCKTSPVQLSDTWLVPKASNPQGGKTKGISRRSRKASRKAHMEELSPRNSPPLSWTAAIT...LSKFGQFSQGTCLPRSHSAKEGPTWKNKSGNTQGG
DANRE .FPPLFPAPKPKPSLQYDKHEDLPLKADMGKTKKLANSPKQKIS.....LPG

870     880     890     900     910     920     930     940
HUMAN .QLNHSSYG..NGISSANHEDLVEGQIKKRLKIKKIKLAAKMKSSPQSR.....TVRSNLEQVQDQSPNDCEIIDLKEL
MOUSE .PLRHRSPGNESAI SPARGDAEDQVHRLRLKIKKIKLAAKMKSSPQSR.....TVRSNLEQVQDQSPNDCEIIDLKEL
SHEEP .HLNHSNGNENGLVSPNHGDTIEGQIKKRLKIKKIKLAAKMKSSPQSR.....TVRSNLEQVQDQSPNDCEIIDLKEL
CHICK .GKTTQVHP..PSCNSVKAESDSDKTRKRLKIKKIKLAAKMKSSPQSR.....TVRSNLEQVQDQSPNDCEIIDLKEL
DANRE .GLN.....PSVKKTAGQEPISITTEALRLKIKKIKLAAKMKSSPQSR.....TVRSNLEQVQDQSPNDCEIIDLKEL

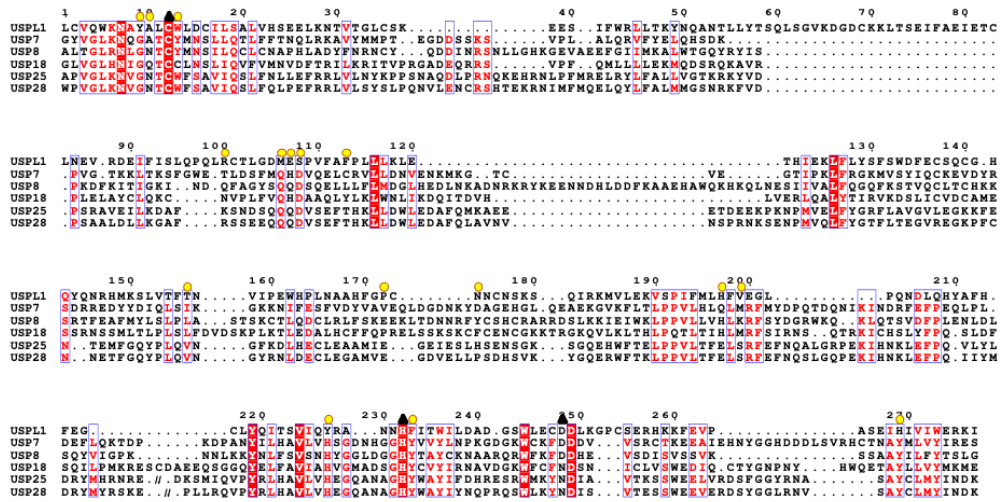
950     960     970     980     990     1000    1010    1020    1030
HUMAN .PYPIDIAS.ECACTVPGVSLYSSQTHEEILAEPLSPTP.VSTELSENCEGDFRYLQMGDGHIPP.PVSEFNDVSNTHLQDHNKYCSPKKNPC
MOUSE .QHQIDLADSKGCTTAPDASNNQSHEEILAEPLSPTP.SEPESGELRLRYLQMGDST.PA.QAPSEFNVSVLNTCLKQDHDYCSPKK.GORE
SHEEP .QHQIDLADSKGCTTAPDASNNQSHEEILAEPLSPTP.SEPESGELRLRYLQMGDGHIPP.PVSEFNVSVLNTCLKQDHDYCSPKK.GORE
CHICK .QYHIDAANSSEFVSNSGCTS..HNSNDEILAEPLSPTP.VASSEVPSKDECMYMEMVNSSAAAPAEAKTSSVSHAATSDRSYGVKDSNXY
DANRE .....STALSSEVSYSTTY.DSFDQFADLSPATTYNS.LVSPSTGLEMLNNGQNGEQ..NDAVATLAP.....EATLCSSTISPLD

1040    1050    1060    1070    1080    1090
HUMAN .VQPSDLTNNACVRLNLSLSPKTDIFDEFSSSALNALANDLDFHDFDEYLFENY
MOUSE .VDLHSMVDSACIRTLNLSLSPKTDIFDEFSSSALNALANDLDFHDFDEYLFENY
SHEEP .VQPSDLTNNACVRLNLSLSPKTDIFDEFSSSALNALANDLDFHDFDEYLFENY
CHICK .R...TVSKSSVKKTFESPTRERLELDL.SISAPTMAG.DMDVPHFDETLFETW
DANRE .EYMQ...SGMCHTALENA..DFNSLDIFF.....

```

USP

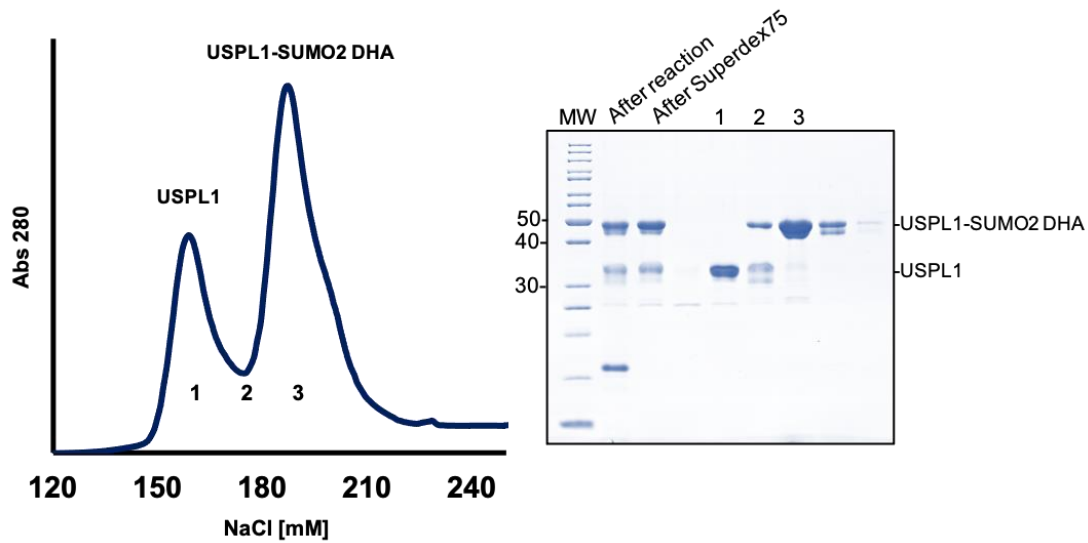
B



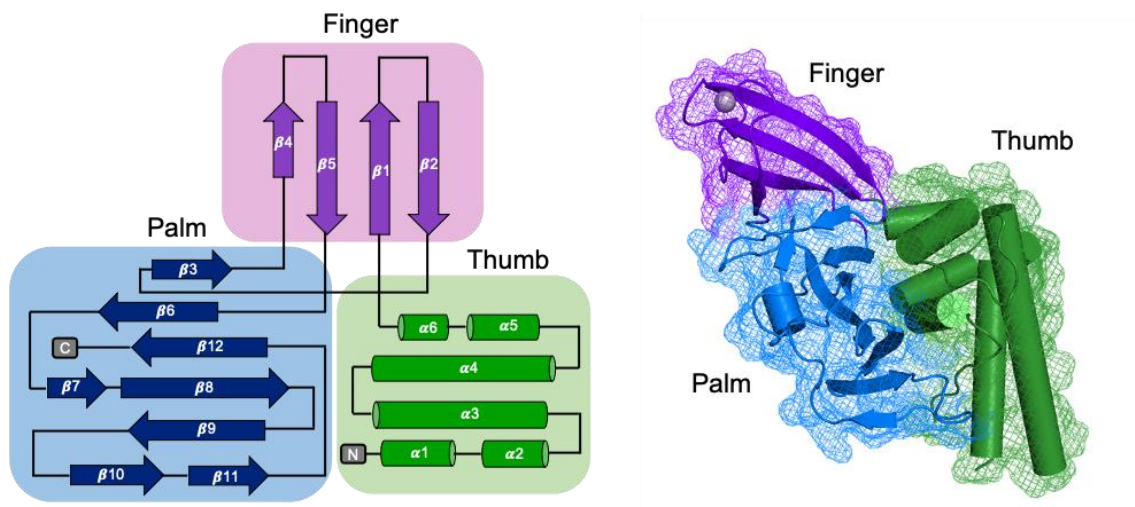
C



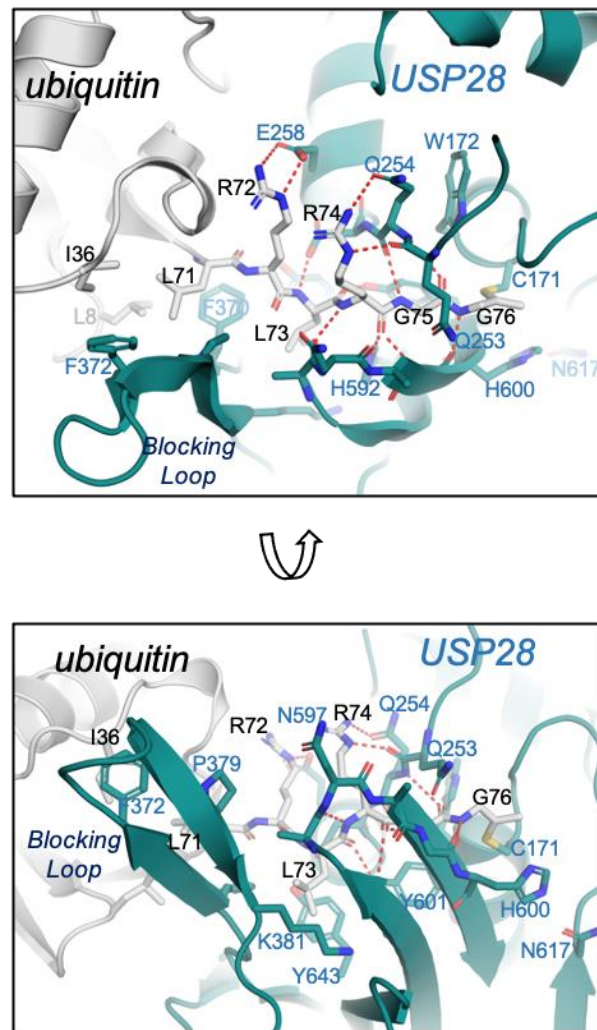
Supplementary Figure 1. A. Multiple sequence alignment of USPL1 with its orthologs in mouse, sheep, chicken and danio rerio (zebrafish). Red represents high conservation. USP-like catalytic domain is labeled by a red frame. **B.** Structural/sequential alignment of the USPL1 catalytic domain with the catalytic domains of USP7, USP8, USP28, USP18, USP25 and USP28. Blue triangles represents the catalytic triad and yellow circles the contact residues with SUMO2. **C.** Sequence alignment of SUMO1, SUMO2 and ubiquitin. All sequence alignments are from Clustal Omega and formatted by ESPript [164].



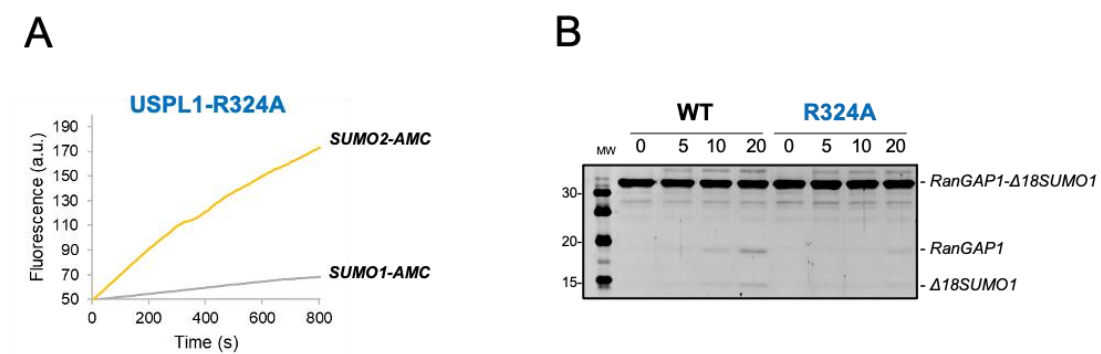
Supplementary Figure 2. Purification of the USPL1-SUMO2 complex. Left, anionic exchange chromatography profile of the purification of the USPL1-SUMO2 complex after the incubation of the SUMO2 DHA precursor with USPL1 catalytic domain at 30 °C for 2 hours. Right, SDS-PAGE of the indicated fractions of the anion exchange column.



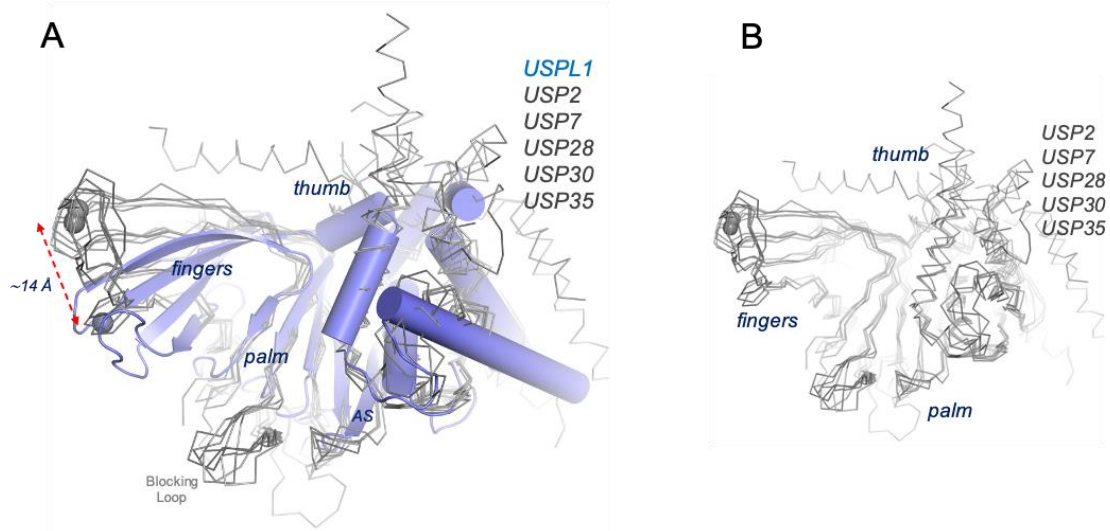
Supplementary Figure 3. Topology diagram of USPL1. The topology diagram of the USPL1 right hand-like subdomains: Finger (purple), Palm (blue), and Thumb (green). The cartoon representation of USPL1 is presented in the same colors.



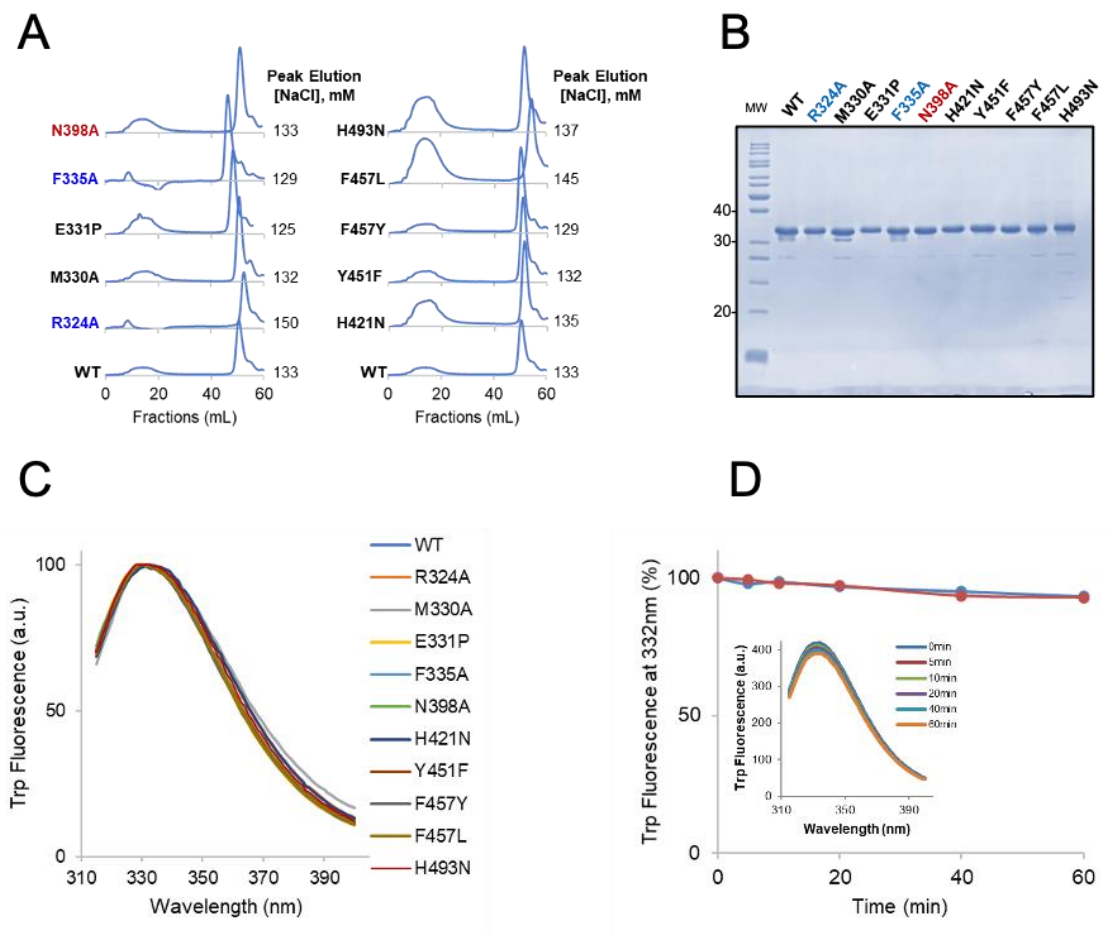
Supplementary Figure 4. USP28-ubiquitin interface details. Two views of the stick representation of the main contacts of the C-terminal of ubiquitin in complex with the active site groove of USP28 (PDB code 6HEK) [165]. USP28 catalytic domain and ubiquitin are shown in green and grey, respectively. Ubiquitin and USP28 interface residues are labelled. Dashed lines indicate hydrogen bond contacts.



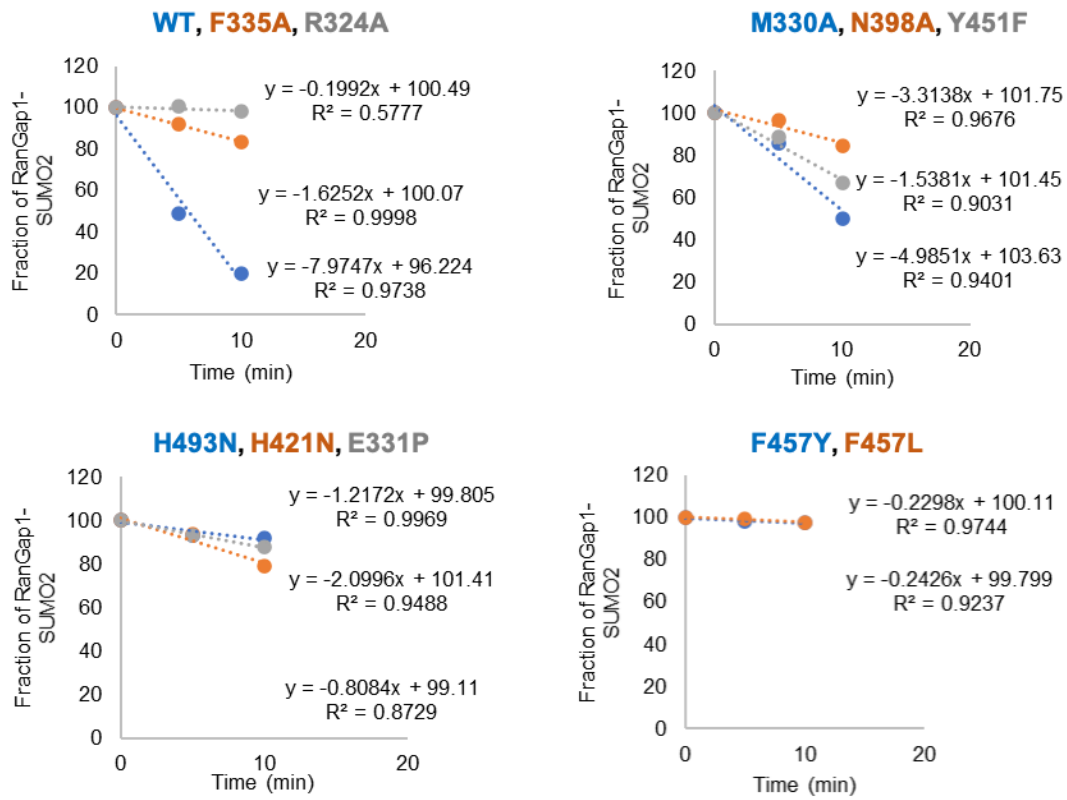
Supplementary Figure 5. SUMO1-AMC substrate analysis with mutant. **A.** Activity assay of USPL1 R324A mutant of the C-terminal tail interface using the SUMO2-AMC and SUMO1-AMC substrates. SUMO-AMC (0.25 μ M) was incubated with 1nM of USPL1 mutant at 30 $^{\circ}$ C, and released AMC was detected by fluorescence. **B.** Time course assays of 20 nM USPL1 wild type and R324A mutant using 1 μ M RanGAP1- Δ 18SUMO1 substrate at 37 $^{\circ}$ C.



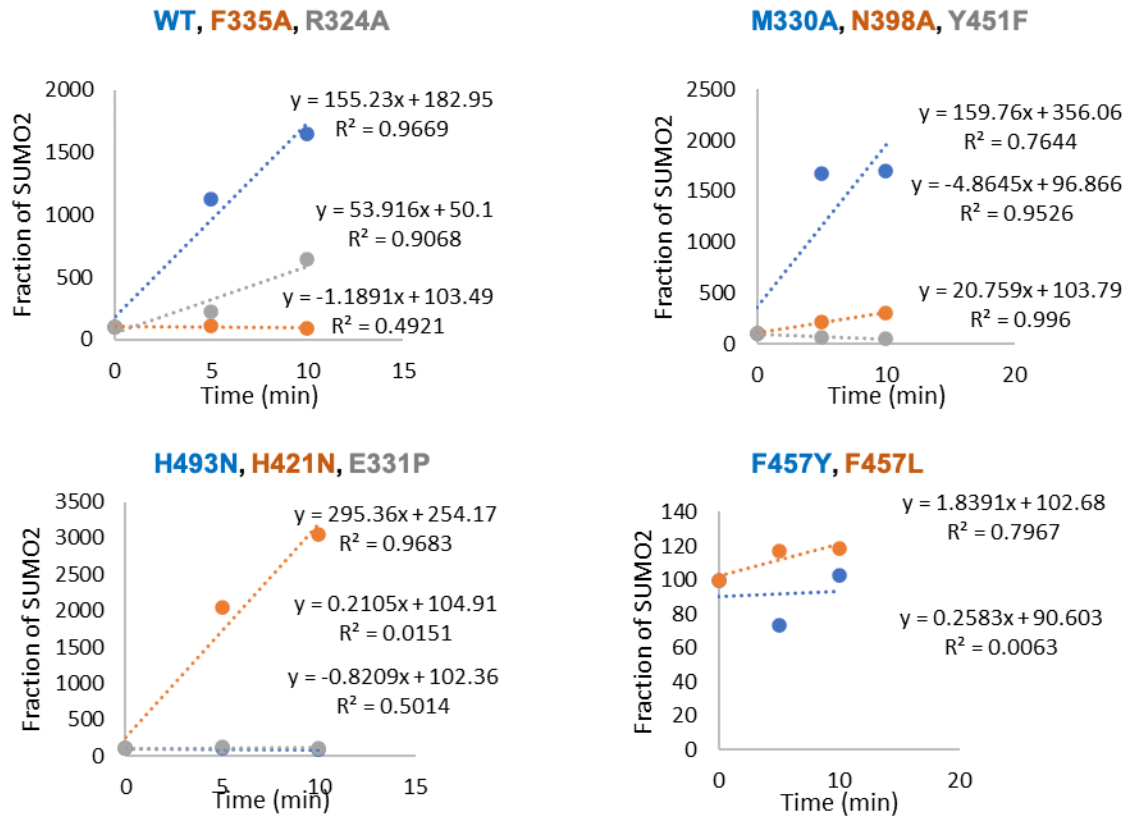
Supplementary Figure 6. Structural overlapping of USPL1 with ubiquitin USPs. **A.** Multiple structural overlapping of USPL1 with five structures of USP in complex with ubiquitin. USPL1 is shown in a blue cartoon representation and the five ubiquitin USPs in grey ribbon representation. Double red arrow indicates the average distance between the Zn^{2+} atom in USPL1 compared to the Zn^{2+} in ubiquitin USPs. USP subdomains are labelled. **B.** Multiple structural overlapping of five structures of USP in complex with ubiquitin. Ubiquitin has been removed from the picture. The USP structures correspond to USP2 (PDB code 2hd5), USP7 (PDB code 5jtv), USP28 (PDB code 6hek), USP30 (PDB code 5ohk), USP35 (PDB code 5txk), USP45 (PDB code 5l8h) and USPL1 (PDB code 7p99)[61,65,165–168].



Supplementary Figure 7. Purification and Trp fluorescence of USPL1 mutants. **A.** Resource Q elution profile of the purification of USPL1; NaCl concentration where the major peak eluted is shown. **B.** SDS-PAGE showing the collected RQ peaks. **C.** Intrinsic fluorescence showing that all mutants are well-folded. (d) Stability of WT (blue) and H493N mutant (red) at 30°C; the inset shows the raw Trp spectra used to calculate stability.

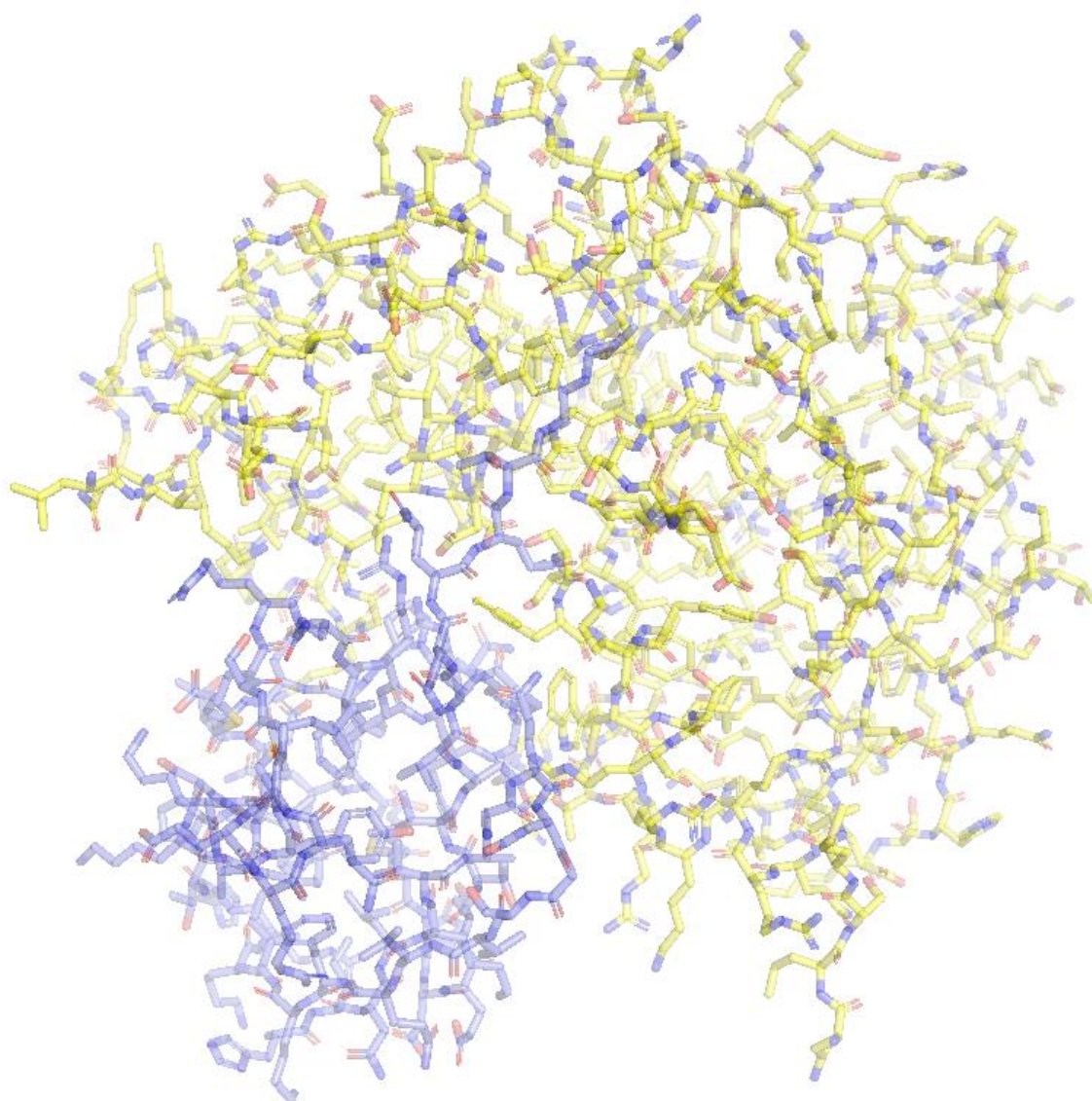


Supplementary Figure 8. Raw data for RanGAP1-SUMO2 deconjugation.



Supplementary Figure 9. Raw data for di-SUMO2 deconjugation.

CHAPTER II: STRUCTURAL BASIS FOR THE SUMO2 ISOFORM SPECIFICITY OF SENP7



Introduction

Small ubiquitin-like modifiers (SUMO) belong to ubiquitin-like proteins (Ubls) family, which are post-translational modifiers that regulate many protein functions in cells, such as DNA replication, nuclear transport or DNA damage control [75,76,79–81]. SUMO is widely expressed in all eukaryotes, there are four SUMO proteins in human: SUMO1, SUMO2, SUMO3, and SUMO4. Yeast and invertebrates have a SUMO protein called Smt3, vertebrates also have few SUMO. In human, SUMO2 and SUMO3 share 97% sequence identity, whereas SUMO1 only share 47% identity with SUMO2 and SUMO3 [75,77]. The SUMO4 precursor cannot be processed and do not form SUMO conjugates [110]. SUMO is activated by a dedicated conjugation pathway through an enzymatic cascade formed by E1, E2 and E3 enzymes [78,82]. The covalent attachment of SUMO to protein targets is a reversible process through action of de-SUMOylating proteases [96,97]. For example, the SUMO proteases of SENP/ULP family can efficiently cleave off the isopeptide bond between the substrate and SUMO [97,98,169].

Until recently, SUMO proteases were only constituted by the 6 members of the SENP/ULP protease family in humans (SENP1, SENP2, SENP3, SENP5, SENP6, SENP7). SENP8 (or NEDP1) is another member in humans that has specificity for Nedd8, a different type of Ubl modifier [103]. ULP1 was the first member of the family SUMO protease discovered in *Saccharomyces cerevisiae* [83]. ULP1 is related to human SENP1, SENP2, SENP3 and SENP5, while SENP6 and SENP7 are related ULP2, a second member in yeast [97]. Human SENPs can be divided into three subfamilies based on their sequence homology, substrate specificity and subcellular localization: SENP1 and SENP2; SENP3 and SENP5; and SENP6 and SENP7 [97,98,100,101]. In the last years, two new types of SUMO proteases have been revealed in human: the deSUMOylating peptidase 1 and 2 (DESI1 and DESI2) [102]; and the deubiquitinating enzyme USPL1, which is active for SUMO2 isoforms, instead of ubiquitin [62,170].

SENP/ULP family members can carry out double proteolytic activities: processing of the immature SUMO, and uncoupling SUMO conjugation. SENP/ULP members can process the α -peptide bond in the C-terminal of SUMO to expose diglycine motif (mature SUMO), and cleave off the isopeptide bond between SUMO and the lysine residue of substrate

[97]. SUMO precursor maturation is an essential step for the activation of SUMO1/2/3 precursors in humans to release a productive C-terminal diGly motif. All SENP/ULP members possess isopeptidase activity, but only SENP1, SENP2 and SENP5 can process the SUMO precursors. Among human SENPs, SENP1, SENP6, and SENP7 are localized to the nucleoplasm [108], whereas SENP3 and SENP5 are localized to the nucleolus [171]. SENP2 has a nuclear export signal along with SENP1 to facilitate its shuttling in and out of the nucleus [109].

All human SENP members are constituted by long unstructured protein chains followed by a conserved C-terminal catalytic domain that belong to the CE class of cysteine proteases [99], containing the active site catalytic triad (His- Asp-Cys). Among the human SENP family, SENP6 and SENP7 are the most dissimilar members, either by the lower sequence identity in their catalytic domains, as well as by the presence of long unstructured protein chain insertions in the middle of the catalytic domain. SENP6 and SENP7 members have been shown to display an isoform preference for SUMO2/3 conjugates [114], in contrast to SENP1 and SENP2 that are equally active for all SUMO isoforms [84,109,172,173].

The structure of the catalytic domain of SENP7 shows unique elements compared to SENP1 or SENP2, such as the absence of an N-terminal α -helix, as well as the presence of three loop insertions [114]. Only Loop 1 can be observed in the apo structure of SENP7 and its deletion severely impairs the proteolytic activity of SENP7, reducing the uncoupling of di-SUMO2/3 or poly-SUMO2/3 conjugates [114]. Whereas the Loop1 of SENP7 is indispensable for activity, the Loop2 and Loop3 are inessential for activity in the tested substrates, both are disordered and their removal did not produce any change in proteolytic activity, showing activities similar to wild-type in vitro [114,115]. SENP6 and SENP7 catalytic domains show a preference for SUMO deconjugation over processing of the precursors, and display a higher activity in dismantling poly-SUMO2/3 chains, in particular in SENP6 [114].

In vivo, SENP7 has been shown to be specific for SUMO2/3 over SUMO1, and the depletion of SENP7 results in the accumulation of SUMO-2 conjugates [116]. SENP7 has been proposed to be a crucial regulator of PML (promyelocytic leukemia protein)

turnover [117,118]. Recent studies have shown that SENP7 plays a critical role in maintaining CD8+ T cell metabolic fitness and effector functions [119]. SENP7 regulates heterochromatin integrity and DNA repair in mitotic cells [97,121,174]. SENP7 maintains HP1 α accumulation at pericentric heterochromatin and regulates DNA repair by interacting with the chromatin remodeler CHD3 in mice [121,122,174–177]. The knockout/down experiments in vivo have shown an essential role for SENP7 in mouse embryonic development [123]. SENP7 is transiently activated during early stages of neuronal differentiation and is required for vertebrate neuronal differentiation [124]. SENP7 has been implicated in breast cancer as a potential target for cancer therapy [125,126].

Existing comprehension of structure/function of SENP7 is limited to the apo form of the catalytic domain, and the molecular structure and determinants of the unique specificity for the SUMO2 isoform have not been yet defined.

To gain insight into the specific interaction of SENP7 with SUMO2, in this chapter we determined the crystal structure of the complex between human SENP7 and a chemically-modified human SUMO2 suicide substrate at the C-terminal glycine (SUMO2-PA), at 1.74 Å resolution. Structural and biochemical analysis of the complex interface of SENP7-SUMO2 by mutagenesis reveals key contact interface residues for its SUMO2/3 isoform specificity, contributing to enrich the knowledge of the SENP/ULP family.

Results

Preparation of the covalent crosslink between SENP7 and SUMO2

To get the structure of the complex between SENP7 and SUMO2, and after several unfruitful attempts in the past to form complexes with SUMO substrates, including SUMO precursors, we have successfully crosslinked the C-terminal carboxylate group of SUMO2 with the active site cysteine of SENP7 (Cys926) to form a stable covalent product complex. To do so the C-terminal carboxylate group of SUMO2 (Gly93) has been replaced with a propargylated group (SUMO-PA), leaving a highly reactive alkyne electrophilic trap that reacts with the cysteine nucleophile of the SENP7 active site (**Figure 1A**). This method uses the intein chemistry to form a terminal propargylated group and it was initially developed for ubiquitin [178,179], and later adapted to SUMO [180]. SUMO is produced as a fusion with intein after substitution of the terminal glycine for cysteine, which ultimately forms a highly reactive and stable C-terminal alkyne (SUMO-PA group) after a reaction with propargylamine.

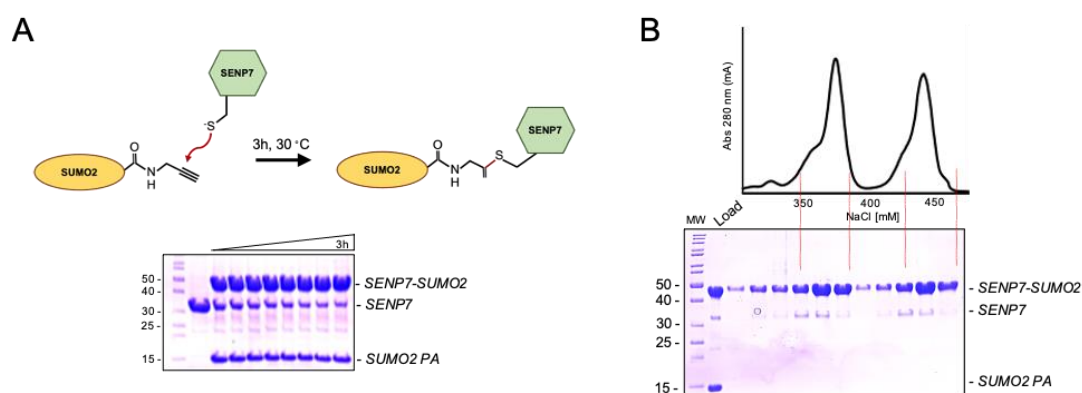


Figure 1. The formation of covalent thioether bond between SENP7 catalytic domain and SUMO2-PA. **A.** Schematic representation of the reaction to form the covalent thioether bond between the C-terminal of SUMO2-PA and SENP7 catalytic domain Cysteine. Time course assay of SENP7- SUMO2PA (molar ratio 1:3) was carried at 30 °C for 3 hours. **B.** Purification of the SENP7-SUMO2 complex. Above, Resource S cation exchange chromatography profile of the purification of the SENP7-SUMO2 complex after the incubation of the SUMO2-PA with SENP7 catalytic domain at 30 °C for 3 hours. Below, SDS-PAGE of the indicated fractions of the Resource S column.

To prepare complex between the SUMO2-PA and SENP7, we have used the catalytic domain of SENP7 comprising from Thr662 to Glu811 and from Lys861 to Ser984. We will refer to this catalytic fragment as SENP7 throughout the text, and the construct includes a truncation of an internal 50 residues-long disordered loop, previously named Loop3,

which did not play any role in the *in vitro* catalytic activity of the enzyme [114] and could represent a restraint in the crystallization experiments.

Overall structure of the complex between SENP7-SUMO2

The SENP7-SUMO2 complex was formed by incubation of SENP7 with the C-terminal modified SUMO2-propargylated at 30 °C for 2 hours to allow formation of a covalent bond between the SUMO terminal alkyne group and the SENP7 active site cysteine (Cys926) (Figure 1B). After purification by anionic exchange, the covalent complex between SENP7-SUMO2 was concentrated to 9.5 mg/ml and set for crystallization experiments. Good diffraction quality crystals of the SENP7-SUMO2 complex were obtained in the initial screening in a condition containing 0.1 M HEPES 7.0, 15 % w/v PEG 20,000. The crystals belonged to the monoclinic $P2_1$ space group, contained two complexes of SENP7-SUMO2 in the asymmetric unit and diffracted to 1.74 Å resolution (Table 1). The final electron density map model of SENP7 includes most of the sequence with the only absence of a gap between Arg748 and Pro758, previously referred to as Loop2 [114], and the aforementioned Loop3 (Glu811 to Lys861), which has been intentionally omitted in the construct of the catalytic domain of SENP7 (Figure 2A). The electron density maps clearly show the covalent bond formed between the SUMO C-terminal Gly93 and the SENP7 active site Cys926, confirming the specificity of the catalytic reaction between SENP7 and the SUMO2-PA suicide substrate.

The comparison of the overall structure of the SENP7 catalytic domain in the apo form (PDB 3EAY) and in complex with SUMO2 does not indicate major structural rearrangements upon substrate binding (rmsd values of 0.61 Å over 244 aligned residues). Only residues involved in the interface with SUMO display alternative side-chains rotation conformers (Figure 2B). Structural overlapping shows that the active site catalytic triad might already be formed in the absence of the SUMO substrate, as observed by the distances between the catalytic triad residues, His794, Cys926 and Asp873, with a 3.7 and 4.3 Å distance between the Cys926 S γ and the His794 N δ 1; and 2.5 and 2.7 Å between His794 N ϵ 2 and Asp873 O δ 1, in the unbound and SUMO2-bound SENP7, respectively (Figure 2C). Such distances suggest that SENP7 might be active in the absence of the SUMO substrate, as initially observed for SENP2 [84], and does not need

any substrate-induced rearrangement mechanism. Indeed, the lower structural similarity with the catalytic domain of SENP2 (rmsd values of 1.92 Å over 168 aligned residues), highlights the evolutionary distances between the SENP6/SENP7 and SENP1/SENP2 subfamilies (Figure 4 & Supplementary Figure 1). Our goal was to define the structural differences in the SUMO interface that might explain the activities by SENP7 for the SUMO2/3 isoform in contrast to SENP2 and SENP1, which is equally efficient for both SUMO1 and SUMO2/3 isoforms.

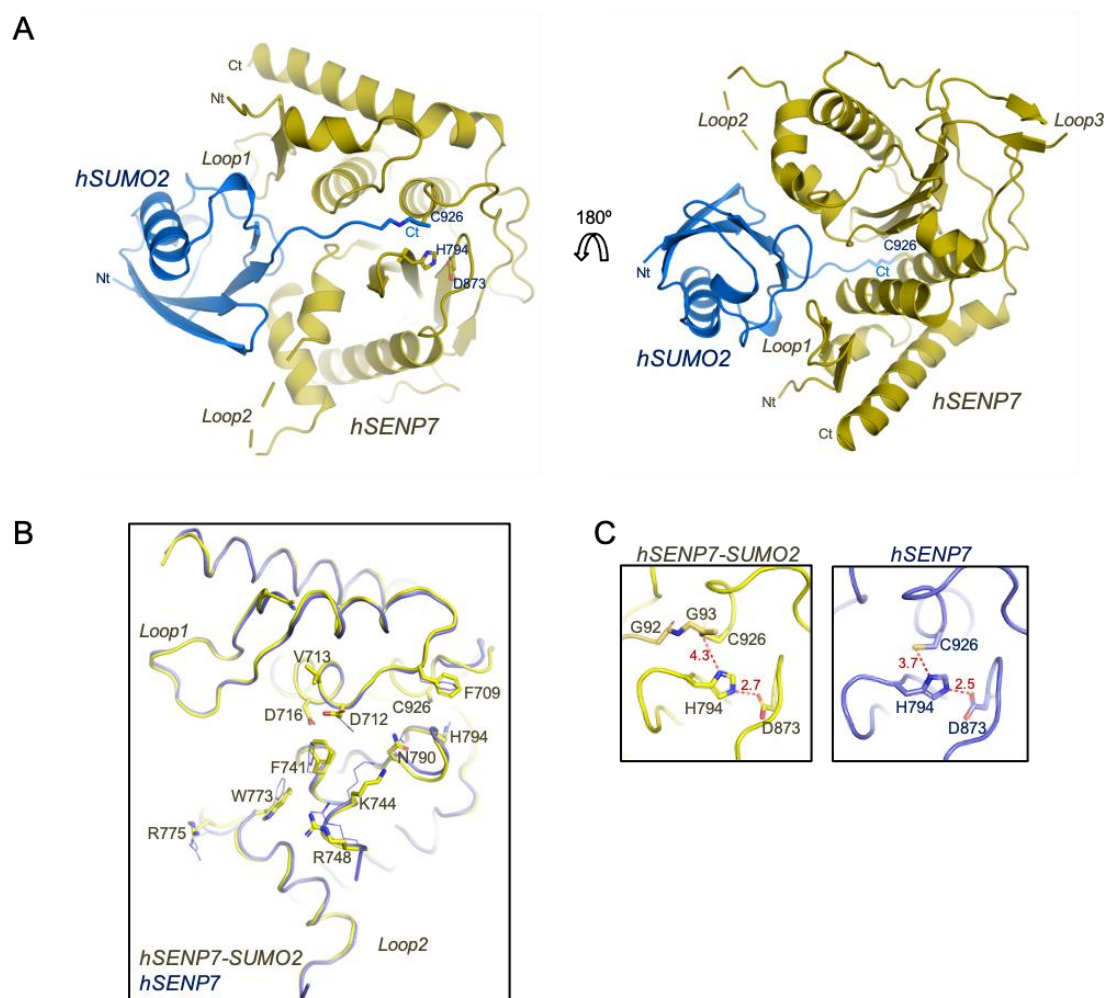


Figure 2. Crystal structure of the complex of human SENP7 with SUMO2. **A.** Cartoon representations of two aspects of the SENP7-SUMO2 complex structure. SENP7 catalytic domain and SUMO2 precursor are shown in yellow and blue, respectively. The catalytic residues are labelled and depicted in stick representation. Loop 1, Loop2 and Loop3 (not present) are labelled. N-terminal and C-terminal are marked. **B.** Structural superimposition between SENP7-SUMO2 (yellow) complex and the apo SENP7 catalytic domain (blue) (PDB code 3EAY). **C.** The superimposition of active sites of SENP7-SUMO2 and apo SENP7 is shown in detail. SENP7-SUMO2 catalytic triad residues are tagged and shown in sticks, whereas SENP7 catalytic triad residues are shown at right. Dashed lines indicate hydrogen bond contacts, and the distances (Å) are also described.

Interface between C-terminal tail of SUMO2 and SENP7

The C-terminal tail of SUMO2 is buried in a SENP7 surface cleft that contains the catalytic triad, which is responsible for cleaving off the isopeptidic bond after the conserved diGly motif in the SUMO C-terminus. This interface is mainly stabilized by a quite number of hydrogen bonds established between the C-terminal tail of SUMO2 (Gln90-Thr91-Gly92-Gly93) and SENP7 residues (Figure 3A). The geometry of the contact interface between the SUMO2 C-terminal tail and SENP7 resembles the interface observed between SENP2 and SUMO2 (Figure 3A&B). A notable difference in SENP7 is the presence of Phe709 instead of a tryptophan, conserved in most members of the SENP/ULP family, sandwiching the C-terminal di-Gly motif. The substitution of Phe709 for tryptophan in SENP7 results in similar catalytic properties as the wild-type form [114], indicating that both bulky aromatic residues play a similar role to cover the SUMO C-terminal tail. Another difference in the C-terminal interface is the presence of Asn790 (in SENP7) instead of a histidine in most members of the SENP/ULP family (Supplementary Figure 1), however, both side chains can form an identical hydrogen bond with the carbonyl oxygen of Thr91.

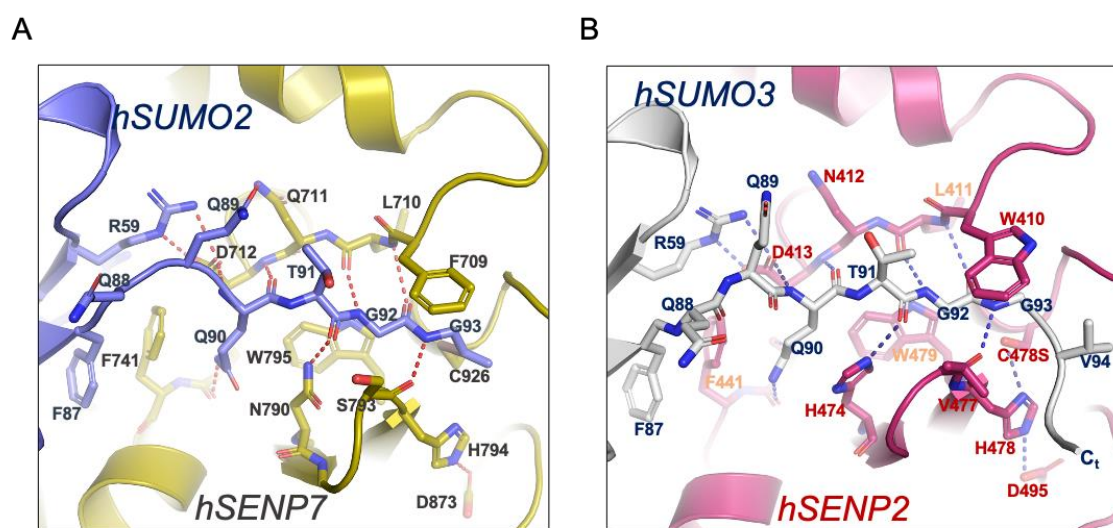


Figure 3. Comparison of atomic details of the SUMO C-terminal tail interaction in SENP7- hSUMO2 complex and SENP2- hSUMO3 complex. **A.** Stick depiction of the primary contacts of the C-terminal of SUMO2 in complex with the active site of SENP7. SENP7 catalytic domain and SUMO2 are shown in blue and yellow, respectively. SENP7-SUMO2 interface residues are labelled. Hydrogen bonds are represented by dashed lines. **B.** Stick representation of C-terminal of SUMO3 (gray) in complex with SENP2 (pink) (PDB code 2IO1). Main interface residues are shown in stick representation and labelled.

The only hydrogen bond interaction established by a side chain in the C-terminal of SUMO2 in the complex with SENP7 is conducted by Gln90. All the other contacts in the C-terminal tail of SUMO2 are basically established by main chain atoms and resemble the SENP2-SUMO2 complex (**Figure 3A&B**), emphasizing the relevance of this hydrogen bond network interface in all members of the SENP/ULP family in the proteolytic activity, which probably will not play an accountable role in the different SUMO isoforms activities observed between SENP1/2 and SENP6/7 subfamily members.

Interface between the globular domain of SUMO2 and SENP7

As initially reported in the yeast ULP1-SUMO crystal structure [151], there exists an extended quilt-like interface between the globular domain of SUMO2 and SENP7, in which the structural determinants for the specific interaction are probably located. Three conserved residues in the SENP/ULP family provide specific contacts at the base of the globular domain at the beginning of the SUMO C-terminal tail, namely Asp712, Phe741 and Trp773 (SENP7 nomenclature). Asp712 is engaged in an electrostatic interaction with Arg59 (SUMO2), Phe741 forms a hydrophobic stacking ring bond with Phe87 (SUMO2), and Trp773 engages a hydrophobic interaction with Gly64 (SUMO2). These three conserved interface residues have been shown to be essential in the SUMO proteolytic activity (**Supplementary Figure 1**) [151,181]. Interestingly, only in SENP8/NEDP1, which is a member of the SENP family but specific for another type of Ubl modifier, Nedd8, significant changes are observed in the phenylalanine and tryptophan positions, which are replaced by glutamate and proline, respectively (**Supplementary figure 1**).

The apo structure of SENP7 revealed the presence of three sequence insertions in the middle of the catalytic domain (**Figure 2&4**), which were named as Loop1, Loop2 and Loop3 and they were also conserved in SENP6 [114]. Similarly to the crystal structure of the apo SENP7 (PDB code 3EAY), Loop2 and Loop3 are not observed in the complex structure with SUMO2, and as shown by in vitro activity assays, they do not participate in the proteolytic activity against SUMO substrates, including polySUMO chains [114,115]. Only Loop1, which contains a well-structured polyproline helix, can be observed in the apo and SUMO2-complexed SENP7 crystal structures, and its structural conformation was been shown to be essential for the SENP6 and SENP7 proteolytic activity [115,158].

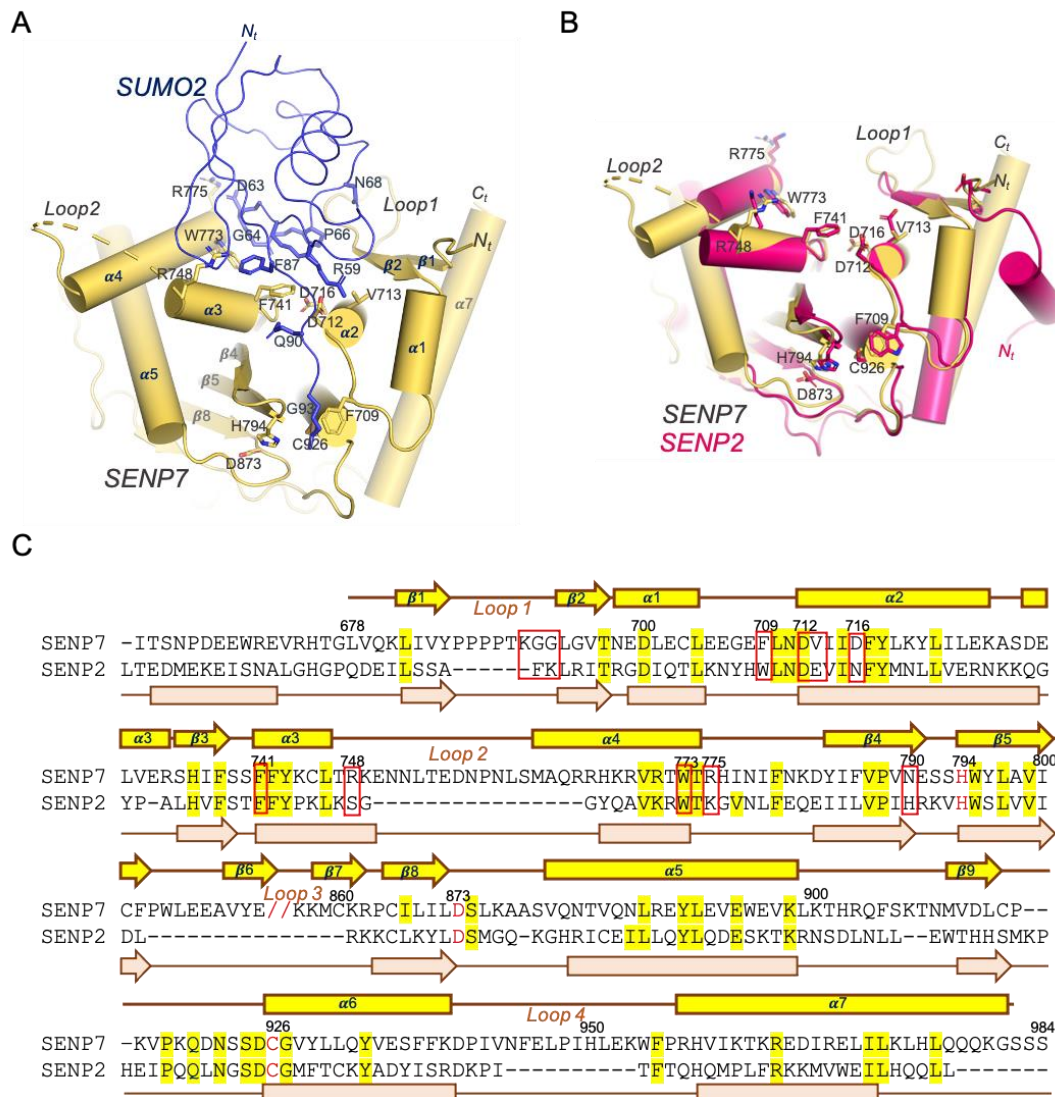


Figure 4. Comparative analysis of SENP7 with the SENP2. **A.** Interaction representation of SENP7 (yellow) in complex with SUMO2, shown in cartoon (blue). Active site residues and main interaction residues are shown in stick representation. All secondary structures α -helix and β -strands are numbered and labelled. **B.** Cartoon representation of the structural overlapping of the catalytic domains of SENP7 with SENP2 (PDB code 1TH0). The Loop1 connects β -strand 1 and β -strand 2 of SENP7 is shown. Active site residues (AS) are shown in stick representation. SENP7 and SENP2 are shown in yellow and pink, respectively. **C.** Alignment of sequences corresponding to human SENP7 and SENP2 catalytic domains based on the structural alignment of human SENP7 (PDB code 3EAY) and SENP2 (PDB code 1TH0). Secondary structural elements are numbered (β -strands and α -helices) and marked above the alignment for SENP7 (yellow) and below the alignment for SENP2 (orange). Gaps are identified by the dotted line and the massive sequence insertion Loop3 in SENP7 is depicted // to indicate that the sequence is missing in this alignment. A yellow backdrop indicates side chain identity with 100% conservation in the alignment. Red represents three catalytic residues. All images were prepared with PYMOL.

Three major contacts are observed in the SUMO2 interface with the SENP7 Loop1 region, namely Arg61, Pro66, Asn68 which are engaged with Gly693 and Leu694 (Figure 5A&B). Asn68 forms a hydrogen bond with the main chain carbonyl oxygen of Gly693, and Arg61 and Pro66 interacts in a hydrophobic pocket shaped by Leu664 and Gly693 in the SENP7 surface. Interestingly, Arg61 also forms a water-bridged hydrogen bond interaction with

Asp716 and Ser739, at the edge of Loop1, probably increasing the specificity of the SENP7 interaction for SUMO2. Remarkably, all these Loop1 residues in SENP7, together with Asp716 and Ser739, are conserved in SENP6, and might represent the particular signatures of the SENP6/7 subfamily for the SUMO2/3 isoform preference [114,115].

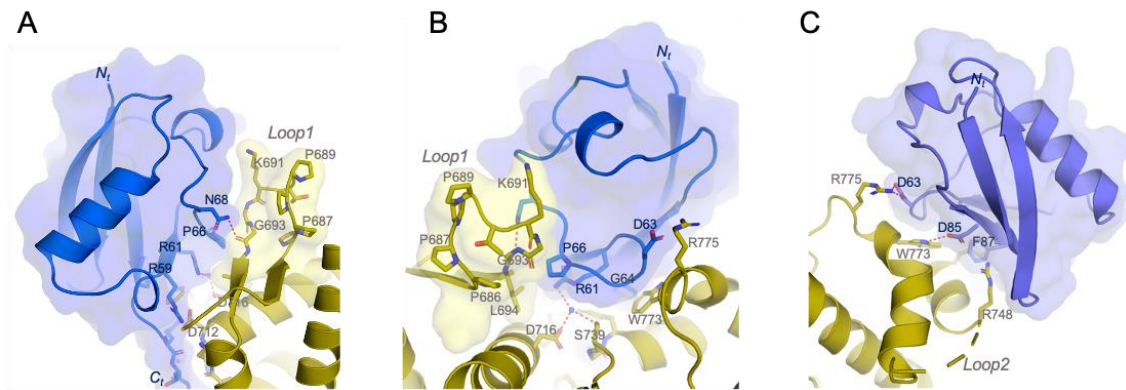


Figure 5. Interface between the globular domain of SUMO2 and SENP7. A. Stick representation of the main contacts of the Loop1 of SENP7 with SUMO2. SENP7 catalytic domain and SUMO2 are shown in blue and yellow, respectively. Interface residues are labelled. Dashed lines indicate hydrogen bond contacts. **B.** Another view with a stick representation of the main contacts of the Loop1 of SENP7 with SUMO2. Colors are same with a. **C.** Stick representation of the main contacts between SUMO and the Loop2 region of SENP7. Colors are same with a.

In the opposite face of SENP7 Loop1 interface, specific contacts are engaged with residues emerging from $\alpha 3$ and $\alpha 4$ helix that are connected through Loop2 (not observed in the crystal structure). Particularly relevant are the contacts engaged by Phe741, Trp773 and Arg775 (Figure 4&5C). Phe741 and Trp773 are engaged in hydrophobic interactions with the SUMO surface (Phe87 and Gly64) and these interactions are common to all members of the SENP/ULP family and essential in the proteolytic activity. Arg775 is engaged in an electrostatic interaction to SUMO2 Asp63 and its positive charged character is conserved in the other SENP/ULP family members. Also, despite that Arg748 is located at the beginning of Loop2 and is unique to SENP7, the final electron density maps models do not indicate strong specific interactions with SUMO2. In summary, the analysis of the SUMO2 interface with SENP7 indicates that the particular activities of the SENP6/7 subfamily for SUMO2/3 isoforms might be ascribed to the Loop1 region of the interface, where the unique contacts in the SENP7 are mostly located.

Activity assays with SENP7 point mutants.

The catalytic domain of SENP7 is specific for the SUMO2/3 isoforms [114,116], which we recapitulated here in Figure 6a using fluorogenic substrates (SUMO1-AMC and SUMO2-AMC), in contrast to SENP1 and SENP2 members that do not show any particular SUMO isoform preference activity [84,181]. The Loop1 region of SENP6 and SENP7 plays a major role in the proteolytic activity for the SUMO2 substrates, providing the determinants for the SUMO2 isoform selectivity [115]. The crystal structure of the SENP7-SUMO2 complex reported here reveals all contacts in the interface with Loop1, which could not be observed in the apo SENP7 structure (pdb 3EAY) [114]. We have mutated some of those contacts and conducted binding experiments using a SUMO2-PA suicide substrate, as well as in vitro proteolytic activity assays using either SUMO2-AMC and RanGAP1-SUMO2 conjugated substrate (Figure 6).

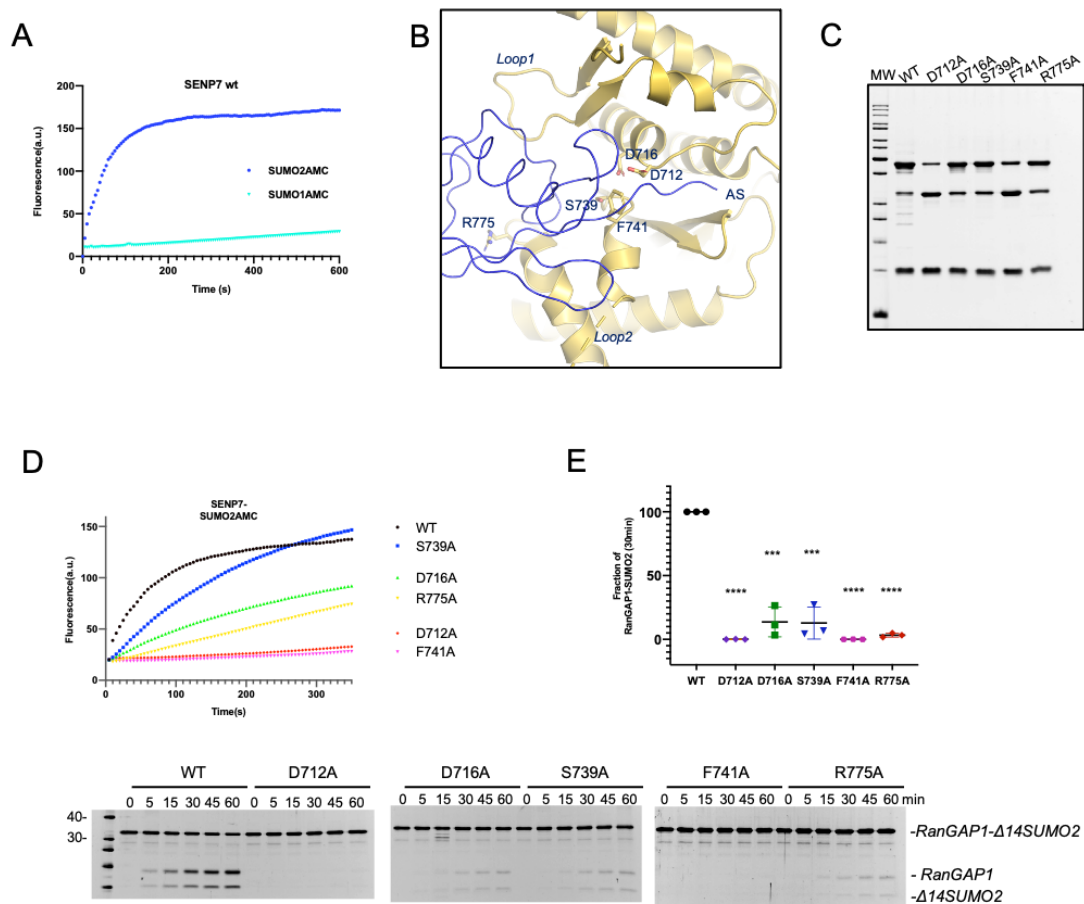


Figure 6. Functional analysis of the SENP7- hSUMO2 interface. **A.** Activity assays of SENP7 with SUMO1-AMC and SUMO2-AMC substrates. SUMO1-AMC and SUMO2-AMC (0.1 μ M) were incubated with SENP7, and released AMC was detected by fluorescence. **B.** Cartoon illustration of the SENP7-SUMO2 complex structure. Blue and yellow represent

the SENP7 catalytic domain and SUMO2, respectively. SENP7 interface residues are identified and represented as sticks. **C.** Binding interaction of SENP7 point mutations with a SUMO2 PA probe. The reaction test was performed at 30°C for 3 hours with SENP7 wild-type and mutants at 1μM using the SUMO2-PA substrate at 4μM. **D.** The SUMO2-AMC substrate was used to test the activity of SENP7 wild type and SENP7 point mutations of the interface. SUMO2-AMC (0.1 μM) was incubated with USPL1 mutants, and released AMC was identified by fluorescence. **E.** Endpoint assays of SENP7 wild type and point mutants at 50 nM using the RanGAP1-SUMO2 substrate at 5 μM. The fraction of the substrate RanGAP1-SUMO2 after 0, 5, 15, 30, 45, 60 minutes reaction was plotted. Data values denote standard deviation, n=3 technical replicates. A two-tailed unpaired t-test was used to determine significance in comparison to wild-type. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

In addition to the SUMO C-terminal contact interface, which is quite similar to the SENP2-SUMO2 C-terminal interaction (**Figure 3**), we can divide the extended contact interface into two regions on opposite sides of the SUMO2 globular domain, named as Loop1 and Loop2 regions (**Figure 6B**). Mutagenesis analysis revealed that the integrity of the structural conformation of the Loop1 region, which contains the polyproline helix, is essential for the SUMO2 proteolytic activity [115]. Here, the SENP7-SUMO2 complex discloses the specific contacts between the globular domain of SUMO2 and the Loop1 region (**Figures 5A&B**), which maintains the integrity of the Loop1 conformation.

In addition to the Loop1 binding to a surface SUMO patch formed by Pro66 and Asn68 [115], the complex structure reveals the presence of a water-bridged bond between SUMO Arg61 with SENP7 Asp716 and Ser739 (**Figure 5A&B**). Interestingly, the proteolytic activity of D716A and S739A is diminished approximately by 8-fold using the RanGAP1-SUMO2 substrate and by 3 to 6-fold using the fluorogenic SUMO2-AMC substrate (**Figure 6E&D**), highlighting the role of this water-bridge interaction for the productive positioning of the SUMO2 globular domain. However, in binding assays using SUMO2-PA, the interaction between these two SENP7 point mutants with SUMO2 is almost undistinguishable from the wild-type SENP7 under this experimental conditions (**Figure 6C**). These results contrast the neighbor Asp712, conserved in the SENP/ULP family and forming a strong salt bridge with SUMO Arg59, and in which its substitution for alanine completely abolishes the proteolytic activity and strongly reduces the binding capabilities of SENP7.

In the Loop2 region we have replaced two direct interactions with SUMO2, namely Phe741, next to the C-terminal tail, and Arg775, located at the opposite edge of the interface (**Figure 6B**). Phe741 is strictly conserved in the SENP/ULP family and its substitution for alanine completely abolishes the proteolytic activity and binding

capabilities of SENP7, highlighting the role of the stacking ring interaction with SUMO Phe59. In the case of Arg775, which forms a salt bridge with SUMO Asp63 (**Figure 5B&C**) and its positive charge character is conserved among the SENP/ULP family (**Supplementary Figure 1**), the proteolytic activity of R775A mutant is reduced approximately by 8-fold in both the RanGAP1-SUMO2 and in the SUMO2-AMC substrates (**Figure 6E&D**), and the binding to SUMO2-PA is also reduced approximately by 2-fold. It is worth noting the role of a single salt bridge between Arg775 and SUMO Asp63, which despite being located at the opposite edge of the interface regarding to the active site, it seems quite relevant in the modulation of the SUMO proteolytic activity of SENP7.

Materials and Methods

Plasmids, Cloning and Point Mutation

SENP7 Δ Loop-3 catalytic domain (aa 662-984 [Δ 811-861]) was constructed at the Sloan-Kettering Institute in New York by David Reverter. Restriction Enzyme Free PCR was used to clone the PTXB1-14-humanSUMO2-intein-chitin-binding domain (CBD) (14 amino acids deletion). [159]. The primers are all listed in the table. Following generation, all sequences were sequenced.

PTXB1- Δ 14SUMO2G fw	CTTAAGAAGGAGATATACATATGAACGATCATATTAATTTGAAGGTGGCG
PTXB1- Δ 14SUMO2G rv	CAACTAGTGCATCTCCCGTGATGCATCCCGTCTGCTGTTGGAACACATCA

Protein Expression and Purification

The SENP7 Δ Loop-3 and Δ 14SUMO2 expression constructs were expressed in *E. coli* Rosetta (DE3) cells (Novagen) for 5h at 30 °C after induction with 0.5 mM IPTG. SENP7 Δ Loop-3 suspensions were equilibrated in 350 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10 mM imidazole, 20% sucrose, 1 mM BME (β -mercaptoethanol), and 0.1% IGEPAL CA-630, and cells were sonicated to break. Nickel affinity chromatography was used to purify SENP7 Loop-3 proteins via the N-terminal his-tag from the pET28 vector. Proteins were further separated by gel filtration chromatography (Superdex 200 ; GE Healthcare) and eluted by an anion exchange resin with a 0-1 M NaCl gradient from 0 to 50% in 20 mM MES (pH 6.0) and 1 mM BME (Resource S; GE Healthcare).

Generation of SENP7 Δ Loop-3- Δ 14-SUMO2 PA complex

Prepare 60mL chitin beads (New England Biolabs, UK), previously washed with 5CV (column volumn) (1CV=200mL) cold lysis buffer (50mM HEPES 8.0, 50mM NaCl) in 4 degrees. Four liters Δ 14SUMO2-intein-CBD pellet is resuspended in 200mL lysis buffer. After being sonicated, the 200mL supernatant is filtered through a 0.45um nylon membrane filter and gently shaken overnight at 4 degrees. To remove unbound proteins, the column was washed in 4 degrees with 5CV lysis buffer. Then 200mL of lysis buffer containing 150mM MesNa was added and gently shaken for 48 hours at room

temperature. Collect the remaining product by adding 15mL more lysis buffer to the 200mL flow-through. Concentrate to 30mL with 3kd filter. After dialysis, added 150mM Propargylamine to 30mL product for 4-5h in room temperature. Dialysis the product overnight at 4 degrees. Concentrated to remove precipitate and applied to an anion exchange resin (Resource Q; GE Healthcare), and eluted with a 0-1 M NaCl gradient from 0 to 50% in 20 mM Tris-HCl (pH 8.0) and 1 mM BME (β -mercaptoethanol) to get Δ 14-SUMO2 PA protein. Incubate SENP7 Δ Loop-3 protein and Δ 14SUMO2PA (1:3 ratio) 3h at 30 degrees. Change the protein buffer to 50mM NaCl, 20mM MES 6.0, 1mM DTT by concentrate. Then applied to an anion exchange resin (Resource S; GE Healthcare) as described above.

Crystallization and Data Collection

The complex SENP7 Δ Loop-3- Δ 14SUMO2-PA were finally concentrated to 9.5 mg/mL for crystallization. Change the protein buffer to 60mM NaCl, 8mM MES 6.0, 1mM BME. Crystallization was performed at 18°C by sitting drop vapour diffusion method by mixing protein with an equal volume of a screen condition solution containing 0.1 M Sodium HEPES 7.0, 15 % w/v PEG 20,000. After 1 week, the crystals were harvested and soaked for 10 seconds in buffers containing 18% ethylene glycol before being stored in liquid nitrogen.

Diffraction data were collected to 1.9 Å resolution at ALBA synchrotron beamline BL13-XALOC (Barcelona, Spain) [182]. Crystallographic details are summarized in Table1.

Structure determination and refinement

The structure of SENP7 Δ Loop-3- Δ 14SUMO2-PA was solved by molecular replacement with SENP7 catalytic domain (PDB: 3EAY) and SUMO2 (1WM3) as search models. Following rounds of model building and refinement were carried out with Coot and Phenix [162,163] (Table1). The structure of SENP7 Δ Loop-3- Δ 14SUMO2-PA has been deposited in the Protein Data Bank with the code PDB 7R2E.

SUMO-AMC hydrolysis assays

SENP7 wild type and mutants were incubated with SUMO2-AMC at 37°C and fluorescence emission was measured using a Jasco FP-8200 spectrofluorometer at 345 nM excitation and 445 nM emission wavelengths. All measurements were carried out with 50 nM SENP7 and 0.1 μ M Ub-AMC in 250mM NaCl, 20 mM Tris-HCl pH 8, 5 mM DTT buffer.

In vitro de-SUMOylation assays

SENP7 protease activity was determined by incubating N Δ 419RanGAP1-SUMO2 (5 μ M) with purified 50 nM of SENP7 wild type and mutants at 37 °C in a buffer containing 20 mM Tris-HCl (pH 8.0), 250 mM NaCl, and 2 mM DTT. After 0, 5, 15, 30, 45, and 60 minutes, the reactions were stopped with SDS-BME loading buffer and gel electrophoresis was performed (PAGE). Gels were stained by SYPRO (Bio-Rad). Proteins were detected and quantified using a Gel-Doc machine with associated integration software (ImageLab; Bio-Rad). The fraction of analyzed bands was plotted as error bar graphs with standard deviation.

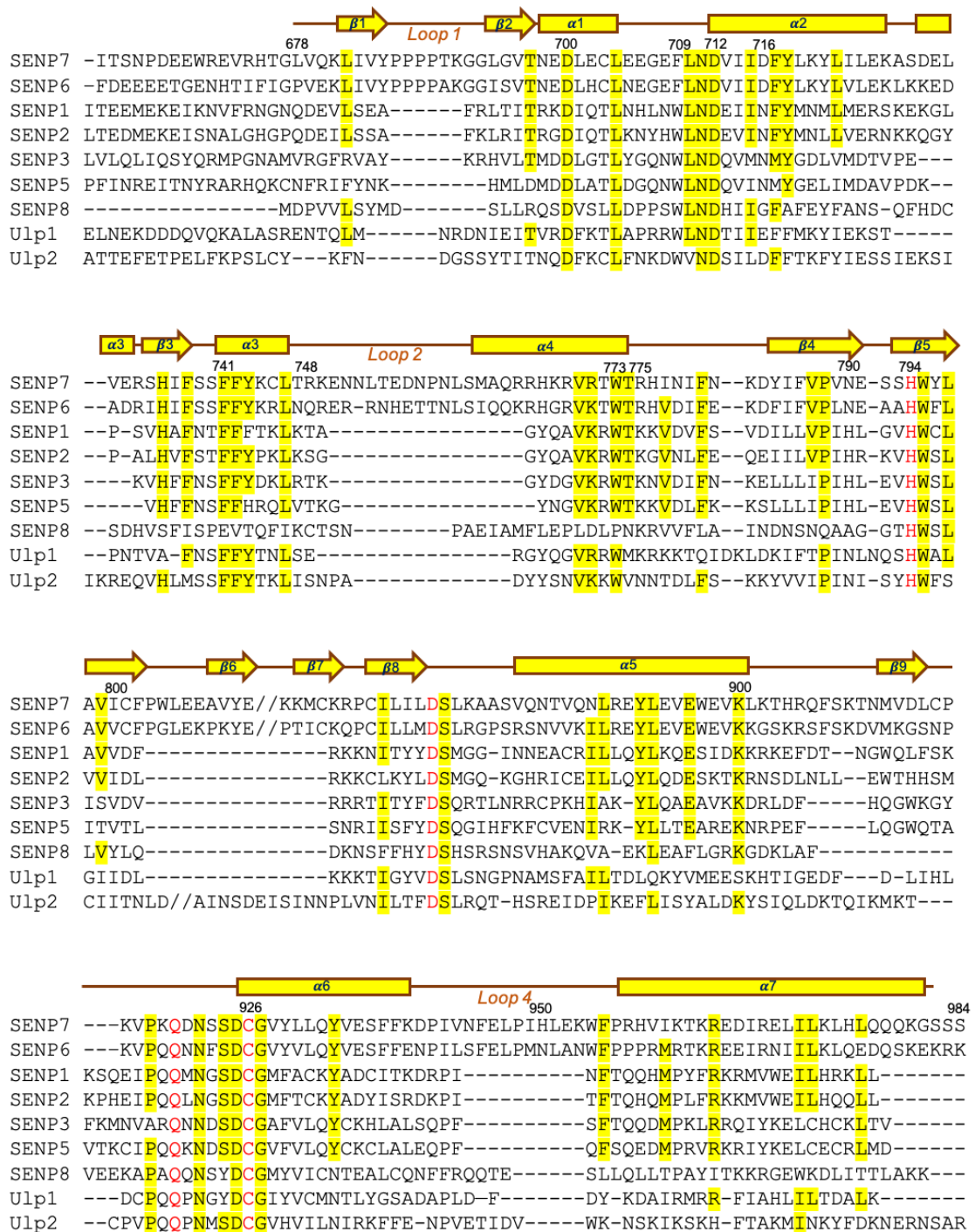
Accession number: Structure reported has been deposited in the Protein Data Bank (PDB ID: 7R2E).

Acknowledgements This work was supported by grants from the “*Ministerio de Ciencia, Innovación y Universidades*” PGC2018-098423-B-I00 to DR. YL acknowledges her scholarship of the China Scholarship Council program from the Chinese government. DR acknowledges support from the Serra Hunter program from Generalitat de Catalunya. X-ray experiments were performed at BL-13 Xaloc beamline at ALBA Synchrotron with the collaboration of ALBA staff.

SENP7-SUMO2	
Data collection	
Space group	P21
Unit cell parameters (Å)	36.41, 100.73, 93.04
Wavelength (nm)	0.97926
Resolution range (Å)	93.04 – 1.74
Rmerge	0.06 (0.63)
Rpim	0.04 (0.44)
(I/σ(I))	10.8 (1.5)
Completeness (%)	92 (55.7)
Multiplicity	3.3 (2.8)
CC (1/2)	0.99 (0.63)
Structure refinement	
Resolution range (Å)	50.36-1.74
No. of unique reflections	51889
Rwork / Rfree (%)	18.41 / 23.04
No. of atoms	
Protein	5323
Water molecules	329
Overall B factors (Å ²)	36.78
Rms deviations	
Bonds (Å)	0.008
Angles (°)	0.916
PDB code	7R2E

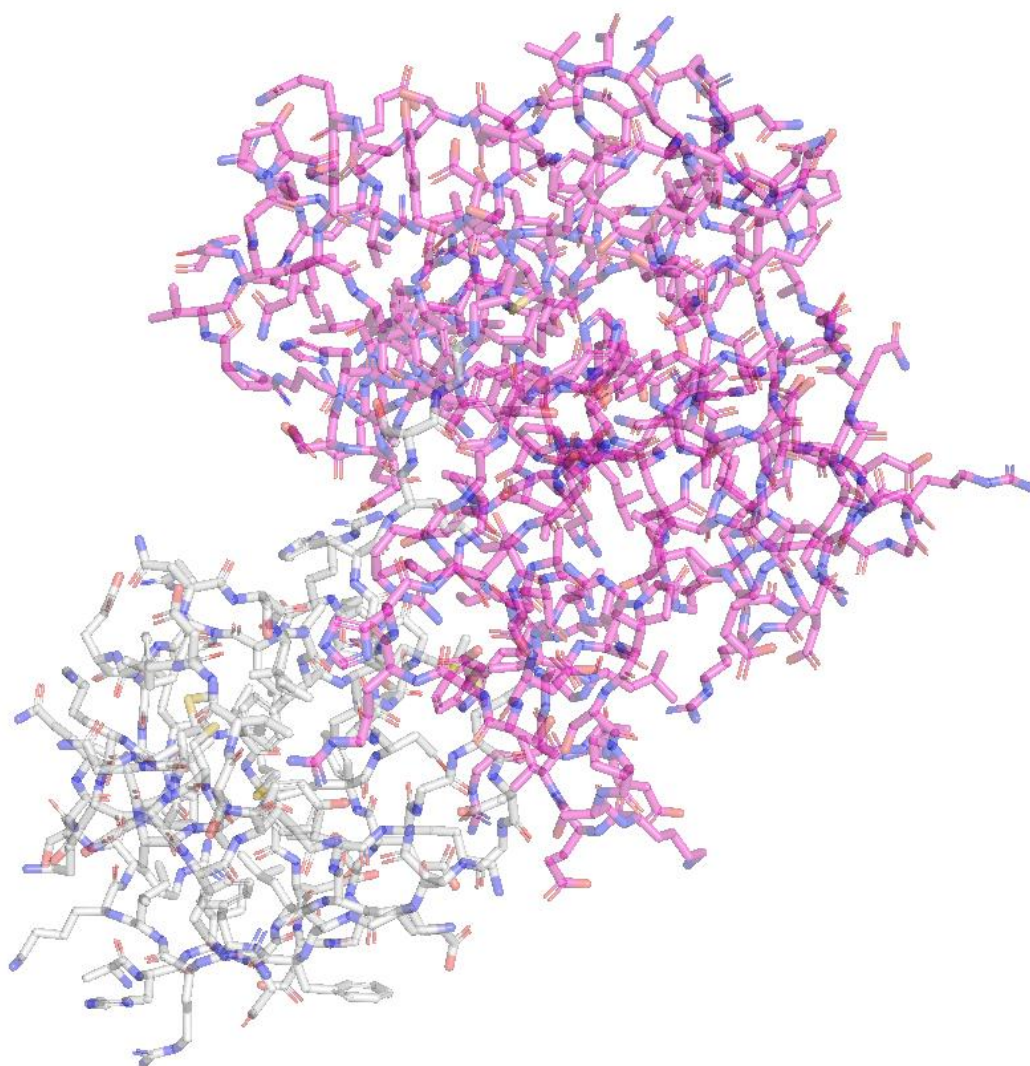
Table 1. Crystallographic statistics of the SENP7-SUMO2 complex.

Supplementary information



Supplementary Figure 1. Structural alignment of the SENP7 catalytic domain with the catalytic domains of SENP1, SENP2, SENP3, SENP4, SENP5, SENP6, SENP8, Ulp1 and Ulp2. SENP7 secondary structures (α -helix and β -strands) are numbered and labelled. A yellow backdrop indicates side chain identity with 100% conservation in the alignment. Red represents three catalytic residues. Loop1/2/3 are marked as well. Residues involved in the SUMO interface are marked with a red square.

CHAPTER III: STRUCTURAL ANALYSIS OF *BRADYRHIZOBIUM*
NOPD PROVIDES INSIGHTS INTO THE DUAL PROTEASE
ACTIVITY FOR UBIQUITIN AND SUMO



Introduction

Post-translational modification (PTM) is essential for the normal functions of proteins and an important part for signaling. Up to now, over 450 distinct protein modifications have been discovered, including ubiquitination, SUMOylation, acetylation and phosphorylation. Each post-translational modification can modify target protein activity, intracellular distribution, protein interactions and lifetime. As a member of PTM, ubiquitination modifies proteins and guides protein to proteolysis, relocalization or endocytosis by forming covalent bonds [183]. Ubiquitin is ubiquitous and highly conserved in organisms, including human beings, yeast and plants. As a 8.5 KDa protein, ubiquitin plays a critical role in various cellular processes [184]. The activating enzyme E1, conjugating enzyme E2 and ligating enzyme E3 constitute a three-steps enzymatic cascade to achieve ubiquitination.

In *Arabidopsis thaliana*, the transcriptome expresses ubiquitin-26S proteasome system (UPS) components, with the majority of ubiquitin ligases encoding genes [184]. UPS is one of the most complicated and extensive cellular regulatory process. E3s are critical factors for defining substrate specificity. Under the E3 enzyme, ubiquitin attaches to an amino of lysine of target protein. Ubiquitin can be linked as a monomer or in chains of various lengths by seven lysine residues. The density and function of ubiquitin chain is determined by its linkage type. The labeling of protein with Lys48 ubiquitin chain is the most well-studied in all different ubiquitin linkages. And the destiny of the protein with Lys48 linkage ubiquitin is as a substrate for 26S proteasome proteolysis. Protein ubiquitination has different functions in various organelles and is reversible.

Similarly, SUMOylation is a post-translational modification that can be reversed as well. SUMO sequence has a low sequence identity to ubiquitin, but its structure and binding mechanism are extremely similar. The C-terminal diGly residue of SUMO forms an isopeptide bond with the amino group of lysine in the catalytic domain of target protein. Under the action of SENPs/ULP, SUMO-bound substrates can remove SUMO and produce unmodified target proteins again, constituting a reversible SUMOylation cycle.

The human CE protease family consists of six SUMO-specific SENPs and NEDD8-specific

NEDP1/SENPA8, collectively referred to as Ubl proteases (ULPs) [185]. The sequence length, protein size, substrate affinity, and subcellular localization of each ULP protein varies. Bacteria and viruses also encode CE family enzymes as effectors to manipulate host signaling pathways [185]. The type III secretion system of Rhizobia injects effector proteins into host plant cells after interaction with the plant. Type III secretion system (T3SS) is located on the cell membrane of rhizobia. It is a multi-protein complex with over 20 proteins that is needle-shaped and evolved from flagella [138,139]. T3SS is a complex that spans the inner and outer bacterial membranes and is highly conserved among bacterial pathogens. The effector protein of T3SS rhizobial is called Nops (Nodulation outer protein) [186], and T3SS secretes many effector proteins, such as NopL, NopE, NopC and NopD.

As a member of CE, XopD is a virulence factor expressed and secreted by *Xanthomonas campestris* pv. *Vesicatoria* [187]. XopD is a multipurpose enzyme that can handle both SUMOylated and ubiquitinated substrates, and the binding sites of the two complexes are extremely distinct. XopD protein targets in Arabidopsis are transcription factors such as HFR1 (positive regulator of photomorphogenesis) [188] and DELLA proteins (negative regulator of gibberellin signalling) [189]. Several Rhizobium T3SS effectors, such as NopDs, which is one of T3SS effectors of Rhizobium, have sequence similarities to the C-terminal protease domain of XopD.

There are many NopD family proteins that have been verified, for example, the identification of *Sinorhizobium freundii* HH103 by mass spectrometry [190], NopD of *Bradyrhizobium* XS1150 [146]. However, the structure details of NopD and the binding mechanisms of SUMO/Ubl have not yet been studied.

In this chapter, we reveal that the effector protein of *Bradyrhizobium*, NopD, in addition to the deSUMOlyase activity, it also contains a K48 bond-specific deUbiquitinase action. We have determined the molecular mechanism for such dual deconjugation activities by solving the crystal structures complexes of NopD with either SUMO or ubiquitin at 1.7 or 2 Å resolution. Mutagenesis of the binding interface reveals particular residues in NopD relevant for this dual SUMO/ubiquitin specificity.

Results

NopD has a dual activity for SUMO and ubiquitin

The full-length protein of the *Bradyrhizobium* NopD effector contains 1017 residues, and it consists of a long N-terminal extension, which seems disordered based on structural and on the Alphafold-2 model predictions, followed by a globular C-terminal a protease domain with homology with the C48 cysteine protease family, which contains a conserved active site catalytic triad composed by cysteine, histidine and aspartic acid. This globular C-terminal domain has homology with the *Xanthomonas campestris* XopD (22.3% sequence identity for 184 residues), a bacterial effector protease with deSUMOylating activity. Based on a structural alignments with XopD, and on the previously published report on the deSUMOylating activity of NopD, we have delimited the conserved NopD C-terminal domain from Pro833 to Asn1016, which has homology with the ULP/SENp deSUMOylase family.

Interestingly in the homologous XopD, in addition to the deSUMOylating activity it was discovered to possess deubiquitinating activity. Such dual deconjugating activity in XopD is an unusual property in members of the SENp/ULP protease family, basically due to the dissimilar protease interface residues between SUMO and ubiquitin. Unexpectedly, an initial experiment with the NopD catalytic domain also revealed a dual activity for SUMO and ubiquitin, similar to XopD (Figure 1A). NopD is able to bind both *Arabidopsis* SUMO2-PA or ubiquitin-PA suicide probes after 30 minutes reaction, in contrast to lack of binding to human SUMO2, as previously reported [146]. The specificity of the interaction was demonstrated by the absence of crosslinked adducts when catalytic inactive active site cysteine to alanine point mutants were utilized (Figure 1A).

Next, the preference for the different types of poly-ubiquitin chains was assessed for *Bradyrhizobium* NopD using a di-ubiquitin chain assay kit (Figure 1B). The CE clan of bacterial effectors with deubiquitinating activity, present in human pathogens to interrupt eukaryotic host response processes, normally displays specificity for K63-linked chains over K48 and K11-linked chains [148]. However in *Xanthomonas campestris* XopD, a plant bacterial effector similar to NopD, the ubiquitin preference moves to K11, K29

and K48 chain linkages over K63 [148], indicating the different type of ubiquitin chain targeted by human and plant pathogen effectors. In *Bradyrhizobium* NopD, the deubiquitinating activity was similar to XopD, displaying a preference for K48 ubiquitin linkages over K63, and to a lesser extent to K11 linkages (Figure 1B).

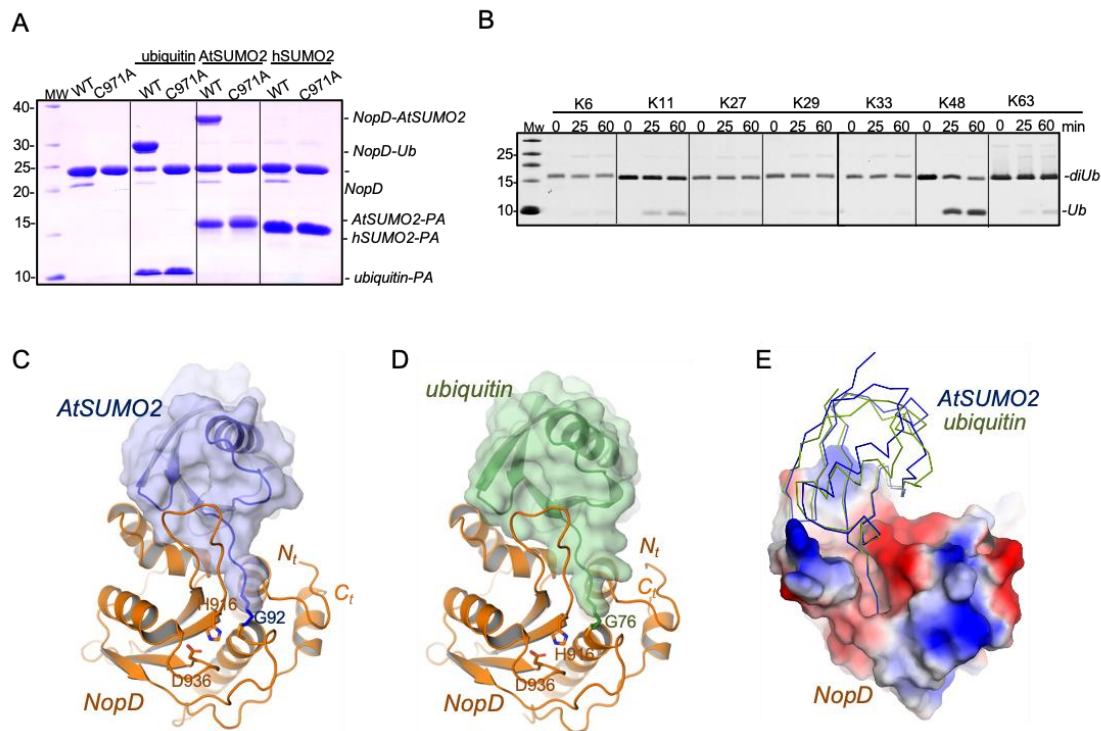


Figure 1. The dual activity of NopD for SUMO and Ubiquitin. **A.** NopD wild-type and active site mutant C971S constructs were tested against Ub/AtSUMO2/hSUMO2-PA suicide probes at 30°C for 2 hours. **B.** Linkage specificity analysis for NopD. A Sypro-stained SDS-PAGE gel shows diUb hydrolysis over time. The concentration of NopD and diUb is 600nM and 3 μM respectively. **C.** Cartoon representation of the NopD catalytic domain with AtSUMO2 is depicted. **D.** The NopD catalytic domain is portrayed as a cartoon with ubiquitin. The N/C-terminals are labeled N/C, respectively. **E.** Electrostatic potential surface representation for NopD-AtSUMO2 and NopD-ubiquitin. AtSUMO2 (blue) and ubiquitin (green) are shown in a line representation.

Overall structures of NopD-SUMO2 and NopD-ubiquitin complex structures

In order to form a complex between NopD catalytic domain and either *Arabidopsis* SUMO2 and ubiquitin, first the C-terminal carboxylate group of SUMO2 and ubiquitin was chemically modified to form a highly reactive and stable C-terminal alkyne (SUMO2-PA or ubiquitin-PA) after a reaction with propargylamine by using an intein approach [178–180]. The C-terminal carboxylate group of SUMO2-PA or ubiquitin-PA can thus be crosslinked with the active site cysteine of NopD (Cys711) to form a stable covalent product complex by incubating NopD (791-1016) with SUMO2-PA or ubiquitin-PA at 30

°C for 2 hours to generate the NopD-SUMO2 and NopD-ubiquitin complex. In an initial screening, a few diffraction quality crystals of the NopD-SUMO2 or NopD-ubiquitin complex were produced in a condition containing 0.1 M imidazole 7.0 and 50% MPD or 0.1 M imidazole 8.0 and 10% PEG8000, respectively. Molecular replacement with human ubiquitin (PDB 1UBQ), Arabidopsis SUMO2 and NopD models from Alphafold-2 predictions assisted to unravel the complex structures. Two complexes per asymmetric unit were found in the NopD-SUMO2 crystals, which belonged to the $P2_12_12_1$ space group and diffracted to a 1.7 Å resolution; and one complex per asymmetric unit in the NopD-ubiquitin crystals, which belonged to the $P4_1$ space group diffracted to a 2 Å resolution (Table 1). The final electron density map model of NopD included a continuous sequence, from Pro829 to Ala1011, in the complex with SUMO2, and from Gly828 to Leu1009 in the complex with ubiquitin, and NopD adopts the regular fold of CE cysteine protease family (Figure 1C&D). The electron density maps clearly show the covalent bond formed between the SUMO or ubiquitin C-terminal glycine and the NopD active site Cys971, confirming the specificity of the catalytic reaction between NopD with either SUMO2-PA or ubiquitin-PA suicide substrates.

The structural overlapping of NopD, in complex with SUMO2 or ubiquitin, with the apo form of NopD from the Alphafold-2 model is quite similar, with a mainchain rmsd (root mean square deviation) of 0.93 Å and 0.90 Å, respectively. Only residues involved in the interface with SUMO/ubiquitin display major changes, and this is particularly remarkable in the Loop insert located between strands $\beta 2$ and $\beta 3$, not present in XopD, that interacts with the SUMO/ubiquitin C-terminal tail. Probably, the active site catalytic triad is already formed in the absence of the SUMO/ubiquitin substrate, as observed by the distances between the catalytic triad residues, Cys971, His916 and Asp936 in the Alphafold-2 model, suggesting that NopD might be active in the apo form, as initially observed in SENP2 [84], and does not need any substrate-induced rearrangement mechanism to activate the protease.

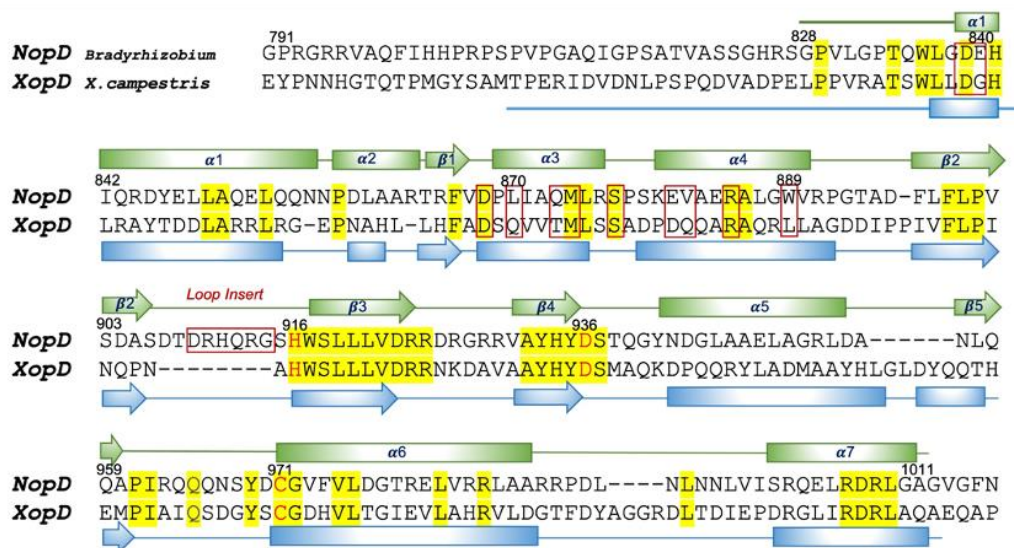
The deubiquitinating activity of NopD does not depend on an N-terminal extension

The structural/functional analysis of the complex between XopD and ubiquitin (PDB 5JP3)

revealed a novel interface between the N-terminal extension and the binding surface around the ubiquitin Ile44-patch [148]. Thus, according to the sequence alignment with XopD (Figure 2D), we initially produced a NopD catalytic domain construct with a longer N-terminal extension to check whether it is necessary for the proteolytic activity against ubiquitin, as occurs in XopD. We designed two NopD constructs, namely NopD (791-1016), and NopD (833-1016) to assess their catalytic activity against ubiquitin substrates (Figure 2F&G).

In XopD, the N-terminal extension is essential for activity against ubiquitin, and full deletion or single point mutagenesis of residues in this region, namely Pro332, Val325 and Asp327, caused failed binding to ubiquitin-PA [148]. However in NopD, the corresponding N-terminal region is not observed in the complex structure and does not seem to play any significant role in the catalytic activity, displaying similar activities for both NopD N-terminal extensions constructs (Figure 2F&G). Thus in contrast to XopD, since the corresponding N-terminal extension in NopD does not play any role in the reaction with AtSUMO2 or ubiquitin, we envisage a different ubiquitin-binding interface different to XopD, as confirmed by the crystal structure of the complex.

A



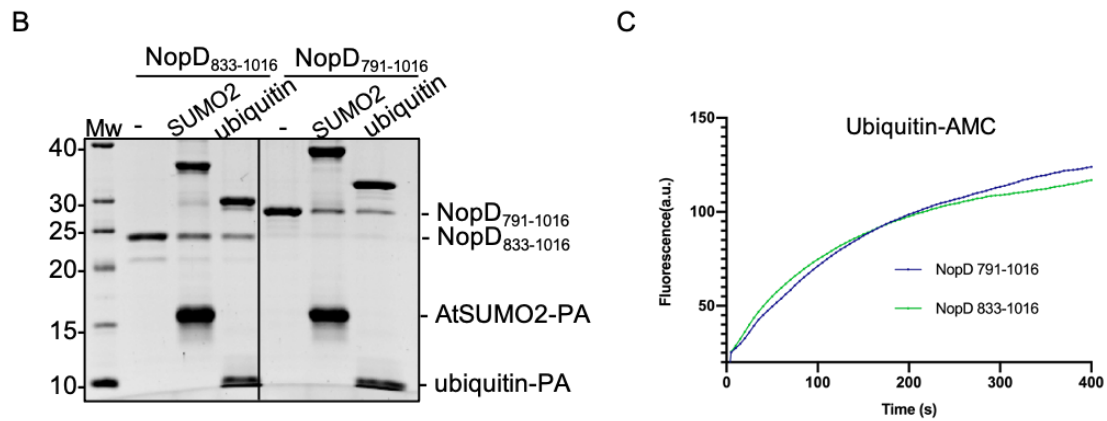


Figure 2. The deubiquitinating activity of NopD does not depend on an N-terminal extension. **A.** Alignment of sequences corresponding to the catalytic domains for NopD and XopD based on structural alignment. Red represents high conservation. All sequence alignments are from Clustal Omega and formatted by ESPript [164]. **B.** NopD constructs with longer N-terminal or without longer N-terminal were tested against AtSUMO2/Ubiquitin suicide probes at 30°C for 2 hours. The bands of NopD complex, NopD₇₉₁₋₁₀₁₆, NopD₈₃₃₋₁₀₁₆, AtSUMO2-PA, ubiquitin-PA are labeled. **C.** The Ubiquitin-AMC substrate was used to test the activity of NopD₈₃₃₋₁₀₁₆ and NopD₇₉₁₋₁₀₁₆. Ub-AMC (0.1 μM) was incubated with NopD (5nM), and released AMC was identified by fluorescence.

Different binding interface with ubiquitin in NopD compared to XopD

Indeed, the crystal structure of the complex of NopD with ubiquitin confirms the different strategy of ubiquitin binding, compared to XopD. Structural overlapping of NopD and XopD shows a different orientation of ubiquitin, showing a displacement of around 5-6 Å, which is basically produced by the different binding interface between NopD and XopD, the latter interacting to the N-terminal extension (**figure 3A-C**). Interestingly in NopD, unlike XopD, the complex structures with SUMO2 and ubiquitin display a similar orientation and some binding interface contacts (**Figure 1C&D**).

The NopD-ubiquitin complex reveals the presence of a strong well-oriented electrostatic interaction between ubiquitin's C-terminal Arg72 and NopD Glu840. This interaction is unique to NopD, since the equivalent position in XopD contains a glycine (**Figure 3D**), and in XopD ubiquitin's Arg72 interacts with the N-terminal extension residue Asp327 of XopD (**Figure 2D**). Glu840 probably might be a particular acquisition of NopD to interact with ubiquitin in *Rhizobium*. Ubiquitin's C-terminal tail contains two arginine residues, Arg72 and Arg74, whilst Arg74 is not engaged in any specific contact in the NopD-ubiquitin complex, Arg72 has usually been revealed as an essential and specific contact point in other deubiquitinating enzymes (for example in USPs). In vitro activity and binding assays confirm the major role of the Arg72-Glu840 interaction in the ubiquitin complex (**Figure**

6).

The other unique characteristic of NopD-ubiquitin complex, not observed in XopD, is the presence of a binding loop insert between $\beta 2$ and $\beta 3$ (Figure 2A). The only other member of the SENP/ULP family with a similar insertion is the in the Den1/NEDP1/SEN8 member, which is specific for Nedd8 instead of SUMO [191]. This Loop insert participates in the backbone interaction with the C-terminal tail of ubiquitin and AtSUMO2 complexes, and it is essential in the *in vitro* proteolytic activity of NopD for both SUMO and ubiquitin (Figure 6). The binding of ubiquitin to this Loop insert in NopD, instead of to the N-terminal extension, explains the different ubiquitin orientation observed between the NopD and XopD complex structures (Figure 3C).

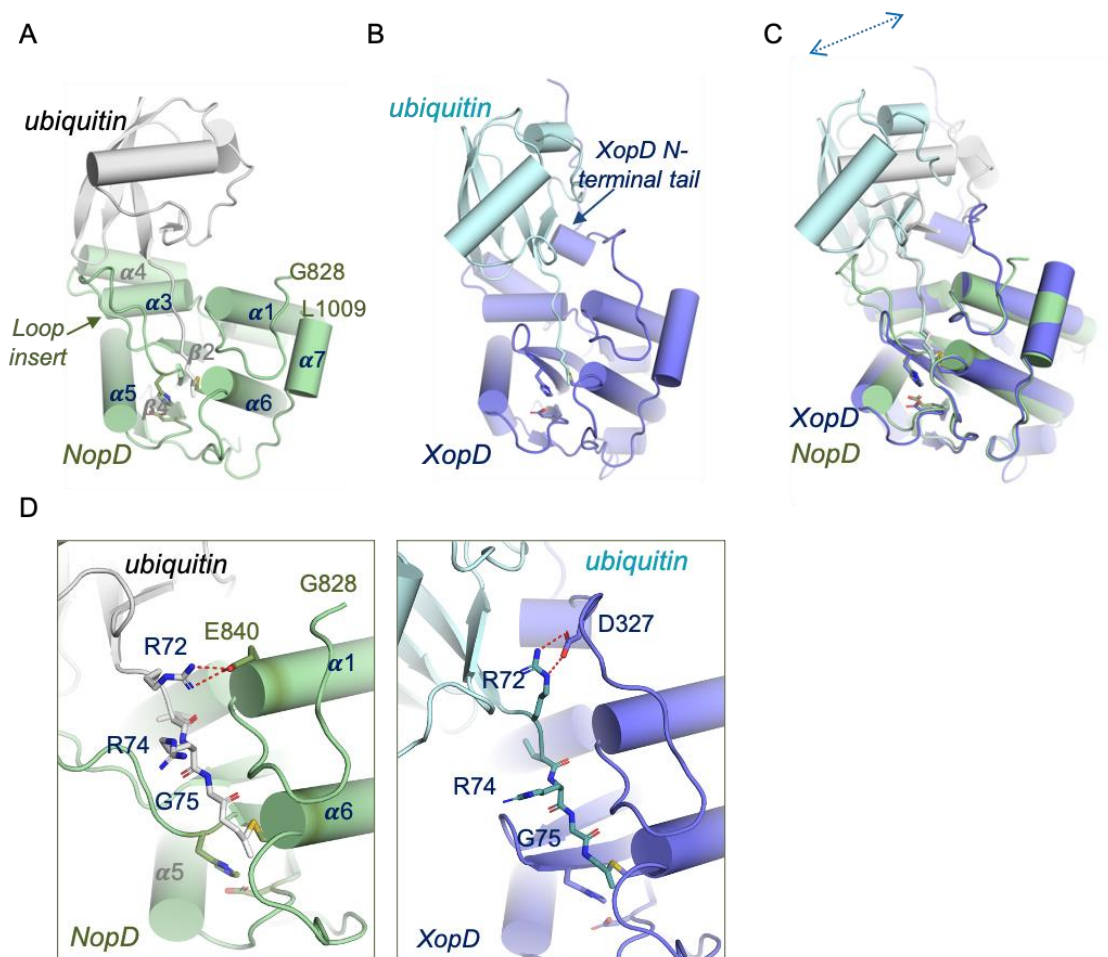


Figure 3. The deubiquitinating activity of NopD does not depend on an N-terminal extension. A-C. Superposition of the NopD-Ubiquitin and XopD-Ubiquitin (PDB: 5JP3) structures in ribbon representation with NopD green color and XopD purple color. The catalytic residue is shown in magenta stick representation. Loop insert is labeled and marked. D. The Close-up views of the C-terminal tail of ubiquitin in NopD-ubiquitin complex (green) and XopD-ubiquitin complex (purple). Residues from ubiquitin are labeled and shown in stick representation. Hydrogen bonds are represented by dashed lines and the distances (\AA) are also depicted.

Interface between C-terminal tail of ubiquitin and AtSUMO2 with NopD

The C-terminal tails of ubiquitin and SUMO2 are buried in a NopD surface cleft that contains the catalytic triad, which is responsible for cleaving off the isopeptidic bond after the diGly motif in both SUMO and ubiquitin C-terminus. It is interesting to observe a quite number of conserved contacts despite the low-identity C-terminal tails of AtSUMO2 (-MLHQTGG) and ubiquitin (-VLRLRGG). Similar to other SENP/ULP complex structures, this interface is mainly stabilized by a quite number of backbone hydrogen bonds established between the C-terminal tail and NopD residues (**Figure 4**). The geometry and distances of these hydrogen bond contacts between the C-terminal tails of AtSUMO2 and ubiquitin with NopD are very similar, and almost all possible backbone hydrogen bonds are established. Interestingly, the guanidinium sidechain and carbonyl mainchain of Arg913, in the Loop insert of NopD, contribute to the backbone network with two hydrogen bonds in both ubiquitin and SUMO2 complexes. Trp836 in NopD sandwiches the C-terminal diGly motif, as usually observed in all complex structures of members in the SENP/ULP family.

Thr90 in AtSUMO2 or the equivalent Arg74 in ubiquitin are not engaged in any interaction, in contrast to the contribution of all other residues of the C-terminal tail of AtSUMO2 and ubiquitin. Glu89 in AtSUMO2 forms a hydrogen bond with Gln873, which is not present in the equivalent Leu73 in ubiquitin. In vitro activity and binding assays with NopD Q873N point mutant confirms the role of this hydrogen bond in the AtSUMO2 binding, but not for ubiquitin (**Figure 6**). His88 in AtSUMO2 forms an electrostatic interaction with Asp839 in NopD, but in ubiquitin the equivalent Arg72 forms an electrostatic interaction with Asp840 in NopD (as described before in **Figure 3D**). Finally, Leu87 in AtSUMO2 and Leu71 in ubiquitin are engaged in similar hydrophobic interaction with the Loop insert of NopD.

A previous study indicated that NopD is specific for plant SUMO isoforms, such as AtSUMO1 and AtSUMO2, GmSUMO (soy bean), PvSUMO (common bean), all containing a similar C-terminal sequence (-MLHQTGG), in contrast to the lack of activity for human SUMO isoforms, with different C-terminal sequences (-YQEQTGG in SUMO1 or -FQQQTGG in SUMO2). Our AtSUMO2-NopD structure might explain the SUMO isoform

specificity in plants by the unique contacts established by Leu88 and His89 with NopD residues (**Figure 4**), in particular by the electrostatic interaction of His88 in AtSUMO2, instead of the hydrophobic interaction engaged by the equivalent Tyr or Phe in human SUMO1 and SUMO2, respectively.

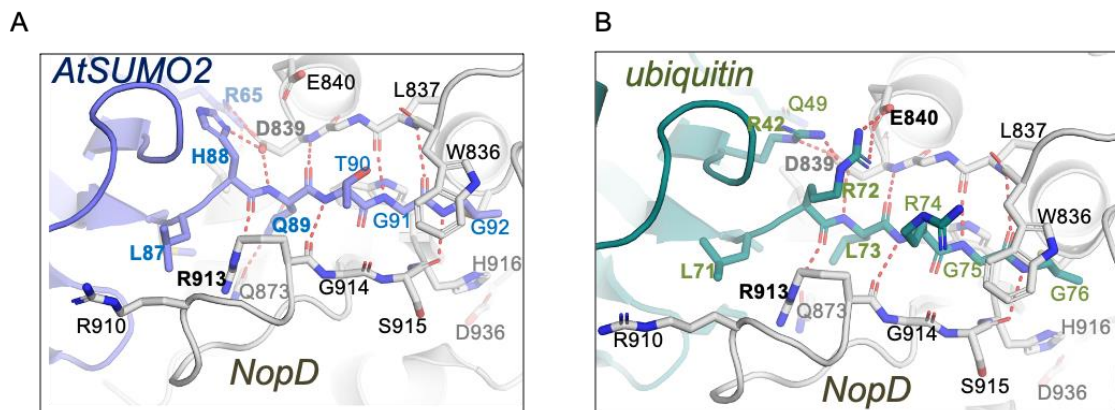


Figure 4. The significance of NopD C-terminal for its activity. **A.** Close-up view of the C-terminal of NopD-AtSUMO2 interaction. NopD (gray) and AtSUMO2 (purple) are shown and marked. The important residues in the contact area of substrate and enzyme are shown as sticks and marked. **B.** Close-up view of the C-terminal of NopD-Ubiquitin interaction. The ubiquitin (green) is shown and other labels are the same as described as A.

Interface between ubiquitin and AtSUMO2 globular domain with NopD

As initially reported in the yeast ULP1-SUMO crystal structure, an extended interface between the globular domain of SUMO (and ubiquitin) with NopD is observed in the SENP/ULP family [151]. To analyze such extended interface, we have divided the interface in three orthogonal views, as indicated in **Figure 5**.

In the first view of the interface both AtSUMO2 and ubiquitin are engaged to NopD with an acidic surface composed by Asp839, Glu840 and Asp868 (**Figure 5A**). In the AtSUMO2 complex, Asp839 and Asp868 form an electrostatic interactions with His88 and Arg65, respectively. However in the ubiquitin complex this interface is formed by different contacts: Asp839 is engaged in an electrostatic interaction with Arg42, and Asp868 forms a hydrogen bond with Gln49 (in addition to the aforementioned electrostatic interaction between Glu840 and the ubiquitin C-terminal Arg92). Mutagenesis analysis underlines the role of Asp839 in both SUMO and ubiquitin interaction, such position is conserved and essential in the SENP/ULP family (**Figure Sup 1**). However, as observed in the

structures, activity assays with the E840A and D868A point mutants display opposite results in AtSUMO2 and ubiquitin. Whereas the electrostatic interaction of Glu840 to Arg74 is essential in ubiquitin, but not relevant in AtSUMO2; the Asp868 electrostatic interaction to Arg65 is essential in AtSUMO2, but not relevant for ubiquitin (Figure 6).

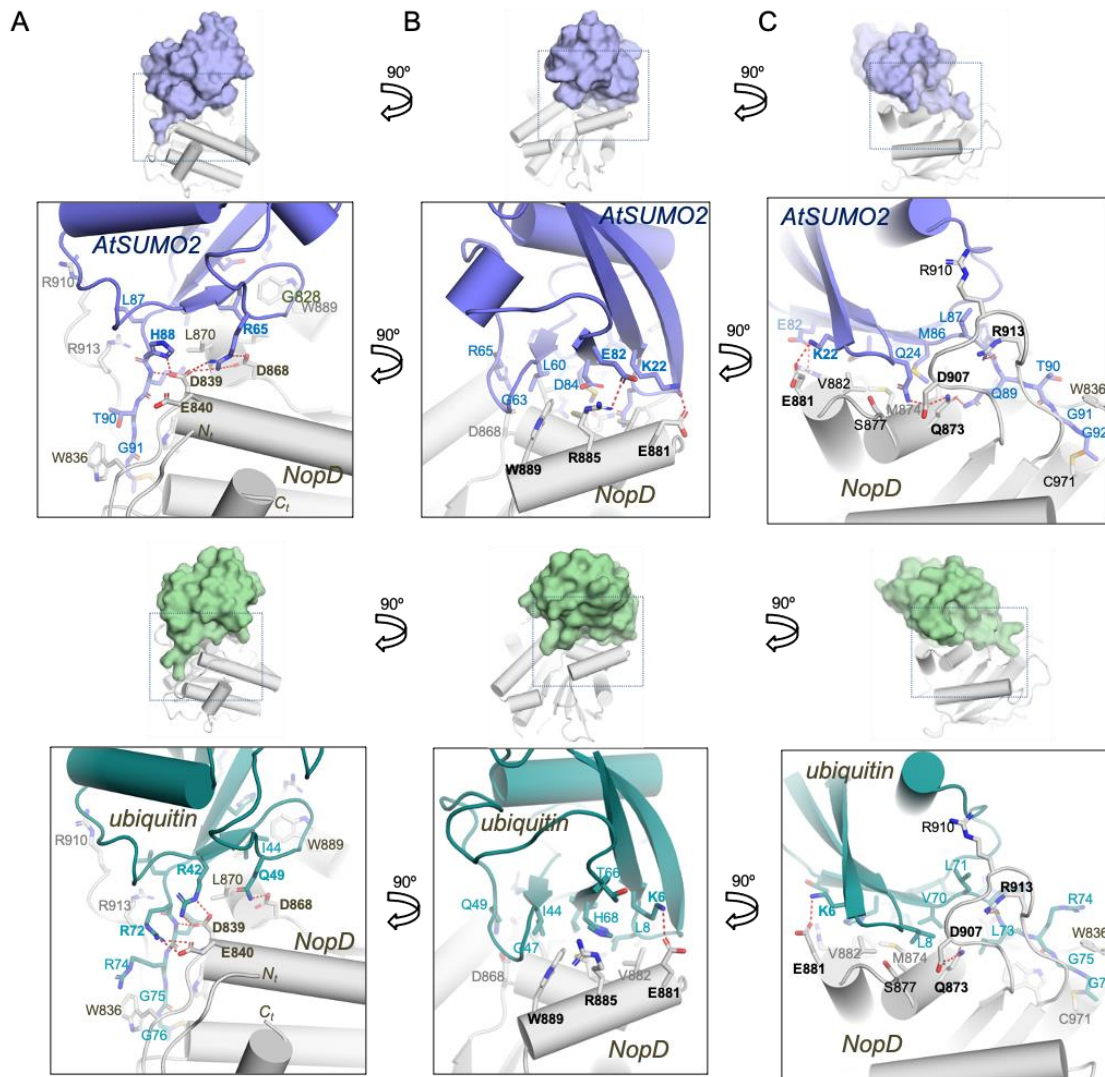


Figure 5. Interface between ubiquitin and AtSUMO2 globular domain with NopD. A-C. Close-up view of the NopD-AtSUMO2 and NopD-ubiquitin interface in three orthogonal. NopD (gray), ubiquitin (green) and AtSUMO2 (purple) are shown and marked. The important residues in the contact area are shown as sticks and marked.

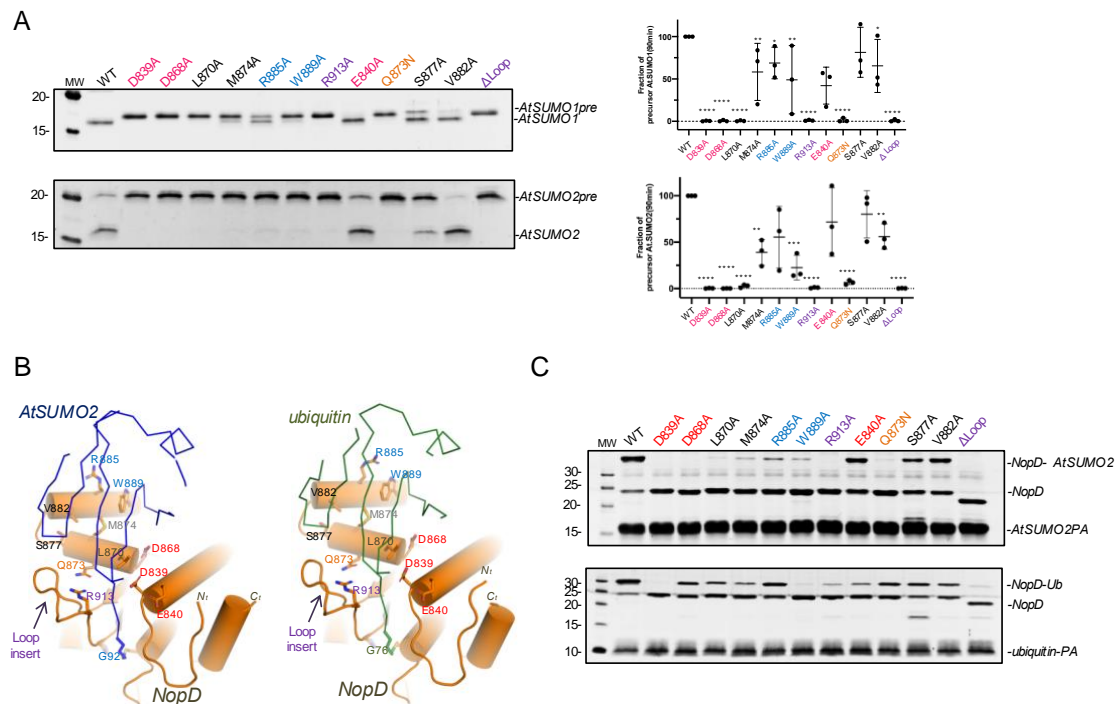


Figure 6. Molecular analysis of NopD SUMO/Ub specificity. **A.** End point assays to detect activities for NopD at 200 nM using the substrates precursor AtSUMO1 and precursor AtSUMO2 at 1 μM. **B.** NopD-AtSUMO2/ubiquitin complex structure shown in cartoon representation, AtSUMO2 and ubiquitin are shown in a line representation. NopD catalytic domain, AtSUMO2 and ubiquitin are shown in orange, blue and green, respectively. Interface residues are labeled and shown in stick representation. **C.** Binding interaction of NopD point mutants with AtSUMO2-PA/ Ub-PA probes (propargylamine-derived probes). Reaction assays were performed with NopD wild-type and mutants at 1 μM using the PA substrate at 4 μM at 30°C for 2 hours.

In the second view of the interface, orthogonal to the C-terminal tail (**Figure 5B**), the major contacts emerging from helix $\alpha 4$ are Trp889, Arg885 and Glu881. In ubiquitin Trp889 is located in a hydrophobic pocket formed by Ile44, Gly47 and His68 (all three sidechains at distances around 3.5 Å); whereas in AtSUMO2 Trp889 is located in a similar pocket formed by Leu60, Gly63 and Asp84 (similar distance 3.5 Å). Mutagenesis analysis highlights the essential role of Trp889 in binding and activity assays for both SUMO and ubiquitin (**Figure 6**). Such position is conserved and essential in the human SENP/ULP family members (**Figure Sup 1**). Arg885 is engaged in an electrostatic interaction with Glu82 only in AtSUMO2, whereas in ubiquitin this position is replaced by His68 and does not interact with Arg885 (**Figure 4b**). Binding and activity assays with the NopD R885A point mutant confirms the role of Arg885 in the SUMO reaction, being irrelevant in ubiquitin (**Figure 6**). Finally Glu881 is at contact distance to the equivalent Lys6 or Lys22 in ubiquitin or AtSUMO2, but the electron density maps do not indicate a strong interaction. It is worth mentioning that in XopD, this region (helix $\alpha 4$) only participates in

the SUMO2 complex, however in the XopD complex with ubiquitin it adopts a different orientation due to the interaction with the N-terminal extension (**Figure 3**).

In the third view of the interface (**Figure 5C**), contacts with both AtSUMO2 and ubiquitin are established by the unique Loop insert in NopD and from residues emerging from helix $\alpha 3$ – loop – $\alpha 4$. In ubiquitin Leu8 and Val70 are placed in a NopD pocket formed by Leu870, Gln873 and Met874 from helix $\alpha 3$ and Val882 from helix $\alpha 4$, whereas in AtSUMO2 the equivalent Gln24 and Met86 interact with the same NopD surface region. Asp907 from the unique Loop insert of NopD forms a hydrogen bond with Gln24 in AtSUMO2, but not with the equivalent Leu8 in ubiquitin. Ser877 in the $\alpha 3$ - $\alpha 4$ loop forms a strong hydrogen bond with the mainchain carbonyl of Gln24 or Leu8 in AtSUMO2 or ubiquitin, respectively (around 2.7 Å distance), however the NopD S877A point mutant does not display a strong effect in the binding and activity assays in either AtSUMO2 or ubiquitin (**Figure 6**).

In summary, the detailed analysis of the AtSUMO2 and ubiquitin interfaces with NopD reveals a similar orientation of the globular domain of AtSUMO2 or ubiquitin on the NopD surface, however in each case, particular contacts are unique to each interaction. In particular it is remarkable the presence of Glu840, which is necessary to fix the C-terminal Arg74 of ubiquitin, absent in XopD and irrelevant for the AtSUMO2 binding. The *Rhizobium* NopD effector protease has evolved with particular adaptations to acquire a dual activity for SUMO and ubiquitin modifiers simultaneously, despite the different surface residues present in this two UbL modifiers. This unusual characteristic seems to be common in other plant effector proteins, such as XopD, but interestingly by using different binding mechanisms.

Materials and methods

Plasmids, Cloning and Point Mutation

The plasmid of NopD was cloned from pMx-NopD (purchased from Thermo Fisher Scientific). Different constructs of PET28a- NopD is amplified by PCR using Restriction Enzyme Free PCR [159]. The NopD point mutats constructs were generated by different primers and were created by the QuickChange site-directed mutagenesis kit (Stratagene). The plasmid of PTXB1-AtSUMO2G was cloned from PET28b-*Arabidopsis thaliana* SUMO2 (a gift from Maria Lois).

Protein expression and purification

The NopDCD, *Arabidopsis thaliana* SUMO2 (AtSUMO2) and human Ubiquitin expression constructs were transformed into *E. coli* Rosetta (DE3) cells (Novagen). *E. coli* were grown at 37 °C to OD₆₀₀=0.7~0.8, and IPTG was added to a final concentration of 0.5 mM. Bacteria were grown an additional 5 h at 30 °C and harvested by centrifugation . Cell suspensions were equilibrated in 350 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10 mM imidazole, 20% sucrose, 1 mM DTT, and 0.1% IGEPAL CA-630, and cells were broken by sonication. After removed cell debris by centrifugation, proteins were separated from lysate by nickel affinity chromatography using Ni Sepharose 6 Fast Flow (GE Healthcare) and eluted with lysis buffer including 20 mM Tris-HCl (pH 8.0), 350 mM NaCl, 300 mM imidazole, and 1 mM DTT. Proteins were separated by gel filtration (Superdex 75; GE Healthcare). Fractions containing the target protein were collected, diluted to 50 mM NaCl, applied to an anion exchange resin (Resource S; GE Healthcare), and eluted with a 0-1 M NaCl gradient from 0 to 50% in 20 mM HEPES (pH 8.0) and 1 mM DTT. Concentrated the protein using Amicon Ultra-30K ultrafiltration device (Milipore) and snap-frozen in liquid nitrogen prior to storage at -80 °C.

Preparation of the NopD- AtSUMO2/Ub complex

The method to get AtSUMO2-PA or Ub-PA protein is noted in **Chapter II**. Incubate NopD CD protein and AtSUMO2-PA/ Ub-PA (1:4 ratio) 3h at 30 °C. Change the protein buffer to 50mM NaCl, 20mM HEPES 7.42, 1mM DTT by concentrate. Then applied to an anion

exchange resin (Resource S; GE Healthcare) as described above.

Crystallization and Data Collection

NopD-AtSUMO2PA and NopD-UbPA were concentrated to 12 mg/mL for crystallization screening. Crystallization experiments were performed at 18°C by sitting drop vapor diffusion method. NopD-AtSUMO2PA crystals grew up in a protein mixture with an equal volume of a condition solution containing 0.1 M Imidazole 7.0 and 50% v/v MPD. NopDUbPA crystals grew up in a protein mixture with an equal volume of a condition solution containing 0.1 M Imidazole 8.0 and 10% w/v PEG8000. Crystals were harvested after 1-2 weeks and soaked 5-10 seconds in the crystallization buffer supplemented with 15% ethylene glycol, and then snap-frozen in liquid nitrogen to storage.

Diffraction data were collected at beamline BL13-XALOC at the ALBA synchrotron (Barcelona, Spain). NopD-AtSUMO2PA and NopD-UbPA get a resolution: 1.8 Å and 1.9 Å, respectively. Resolution Data processing was conducted by AutoProcesing with MxCUBE [160,161]. The space group of NopD-AtSUMO2PA is $P2_1$ and there was two complexes per asymmetric unit. The space group of NopD-UbPA was $P4_1$. Structures were solved by molecular replacement with NopD AlphaFold2 model as a search mode [152]. Following rounds of model building and refinement were carried out with Coot and Phenix [162,163].

In vitro de-SUMOylation assays

Protease activity was measured by incubating precursor of *Arabidopsis thaliana* SUMO1/2 with purified 200 nM of NopD wild type and mutants at 30 °C in a buffer containing 20 mM Tris-HCl (pH 8.0), 250 mM NaCl, and 2 mM DTT. SDS-BME loading buffer was used to terminate the reactions after 2 hours, and gel electrophoresis was used to examine the results (PAGE). SYPRO staining was used to identify proteins (Bio-Rad). A Gel-Doc machine and integration software were used to detect and quantify the products (ImageLab; Bio-Rad).

SUMO-AMC hydrolysis assays

NopD wild type and mutants were incubated with ubiquitin-AMC at 30°C and fluorescence emission was measured at 345 nM excitation and 445 nM emission wavelengths using a Jasco FP-8200 spectrofluorometer. All measurements were carried out with 5 nM NopD and 0.1 μ M Ub-AMC in 250mM NaCl, 20 mM Tris-HCl pH 8, 2 mM DTT buffer.

Table 1. Crystallographic statistics of the NopD- AtSUMO2/ Ubiquitin complex.

	NopD- AtSUMO2	NopD- Ubiquitin
Data collection		
Space group	P2 ₁ 2 ₁ 2 ₁	P4 ₁
Unit cell parameters (Å)	80.85, 86.83, 90.49	86.50, 86.50, 46.67
Wavelength (nm)	0.97918	0.97926
Resolution range (Å)	49.52 – 1.69	43.25 – 2.04
R _{merge}	0.07 (0.83)	0.07 (1.42)
R _{pim}	0.03 (0.48)	0.03 (0.60)
(I/ σ (I))	14.2 (1.8)	13.3 (1.2)
Completeness (%)	92.7 (48.9)	94.2 (55.0)
Multiplicity	6.1 (3.7)	6.8 (6.6)
CC (1/2)	0.99 (0.49)	0.99 (0.51)
Structure refinement		
Resolution range (Å)	49.52-1.50	43.25-1.94
No. of unique reflections	76875	22770
R _{work} / R _{free} (%)	17.84 / 20.36	17.62 / 20.59
No. of atoms		
Protein	8363	2055
Water molecules	312	62
Overall B factors (Å ²)	31.02	58.12
Rms deviations		
Bonds (Å)	0.015	0.007
Angles (°)	1.140	1.055

Supplementary information

790
 NopD LGPRGRRVAQFIHHRPSPVPGAQIGPSATVASSGHRSGPVLGPTQWLGDEHIQRDYELLAQELQQNNPDLAAR
 XopD YPNNHGTQTPMGYSAMTPERIDVDNLPSPQD--VADPELPPVRATSWLLDGHRLRAYTDDLARRLRG-EPNAHL-
 SENP1 EFPEITEEMEKEIKNVFRNGNQDEVLSEAFRLTITRKDIQTLNHLNWLNDIINFYMNML--MERSKEKGLP-S
 SENP2 DLLELTEDMEKEISNALGHGPDQDEILSSAFKLRITRGDIQTLKNYHWLNDEVINFYMNLL--VERNKKQGY-P-A
 SENP3 SRKGLVLQLIQSYQRMPGNAMVRGFRVAYKRHVLTMDDLGTLYGQNWLNQVMNMYGDLV--MDTVPE-----K
 SENP5 DFSNRKPFINREITNYRARHQKCNFRIFYNKHMLDMDLATLDGQNWLNQVINMYGELI--MDAVPDK-----
 SENP6 DEEETGENHTIFIGPVEKLIVYPPPPAKGGISVTNEDLHCLNEGEFLNDVIIDFYLYLYL--VLEKLLKEDADR
 SENP7 TSNPDEEWRVRHTGLVQKLIIVYPPPTKGLGVTNEDLECLEEGEF LNDVIIDFYLYLYL--ILEKASDELVER
 SENP8 -----MDPVVLSYMSLLRQSDVLSLLDPPSWLNDHIIIGFAFEYF--ANS-QFHDCSDH

864
 NopD TRFVDPLIAQMILRS-----PSKEVAERALGWRP---GTAD-FLFLPVSDASDTRHRGRSHWSL-
 XopD LHFADSQVVTMLSS-----ADPDQARAQRLLAG---DDIPPIVFLPINQPN-----AHWSL-
 SENP1 VHAFTFFFTKLTAKTA-----GYQAVKRWTKKVDVFSVD-IILVPIHLGV-----HWCLA
 SENP2 LHVFTSTFFYPKIKSG-----GYQAVKRWTKGVNLFQE-IILVPIHRKV-----HWSLV
 SENP3 VHFNSFFYDKLRTK-----GYDGVKRWTKNVDIFNKE-LLLIPIHLEV-----HWSLI
 SENP5 VHFNSFFHRLVTKG-----YNGVKRWTKKVDLFFKS-LLLIPIHLEV-----HWSLI
 SENP6 IHISSFFYKRLNQRER-RNHETNLSIQQRHGRVKTWTRHVDIFEKD-FIFVPLNEAA-----HWFLA
 SENP7 SHIFSSFFYKCLTRKENNLTEDNPNLSMAQRHRKRVRTWTRHINIFNKD-YIFVPLNESS-----HWYLA
 SENP8 VSFISPEVTQFIKCTSN-----PAEIAMFLEPLDL-PNKR-VVFLAINDNSNQAA--GGTHWSLL

920
 NopD -----LLVDRDRGRVAYHYDSTQGYNDGLAAELAGRLDAN-----LQQAP--
 XopD -----LLVDRRNKDAVAAYHYDSMAQKDPQRYLADMAAYHLGLDYQQTHEMP--
 SENP1 VVDF-----RKNITYYDSMGG-----INNEACRILLQYLKQESIDKKRKEFDT--NGWQLFS
 SENP2 VIDL-----RKKCLKYLD SMGQ-----KGHRICEILLQYLQDES KTKRNSDLNLL--EWT HHS
 SENP3 SVDV-----RRRTITYFDSQRTL-----NRRCPKHIAK-YLQAEAVKKDRDLDF-----HOGWKG
 SENP5 TVTL-----SNRIISFYDSQGIH-----FKFCVENIRK-YLLTEAREKNRPEF-----LQGQWT
 SENP6 VVCFPGLEKPKYE//PTICKQPCILLMDSLGRG-----SRSNVVKILREYLEVEWEVKKGSKRSFSKDVMMKGSN
 SENP7 VICFPWLEEAVYE//KMKCRPCILLILDSLKAA-----SVQNTVQNLREYLEVEWEVKKLTKHRQFSKTNMVDLC
 SENP8 VYLQ-----DKNSFFHYDSHSRS-----NSVHAKQVA-EKLEAFLGRKGDKLAFL-----

962
 NopD -----IRQQNSYDCGVFVLDGTRELVRRRLAARRPDL----NLNNLVISRQELRDLRDLGAGVGFN-----
 XopD -----IAIQSDGYSCGDHVLGTGIEVLAHRVLDGTFDYAGGRDLTDIEPDRGLIRDRLAQAEQAPAESSIRQVP-
 SENP1 KKSQEI PQQNGSDCGMFACKYADCITKDRPI-----NFTQQHMPYFRKRMVWEILHRKLL-----
 SENP2 MKPHEI PQQLNGSDCGMFTCKYADYISRDKPI-----TFTQHQPPLFRKRMVWEILHQQL-----
 SENP3 YFKMNVARQNNDSDCGAFVLYQYCKHLALSQPF-----SFTQQDMPKLRRIYKELCHCKLTV-----
 SENP5 AVTKCI PQQKNSDCGVFVLYQYCKLALAQPF-----QFSQEDMPRVKRIYKELCECRLMD-----
 SENP6 P---KVPQNNFSDCGVYLYQYVESFFENPILSFE LPMNLAN-WFPPPRMRTKREIIRNIIILKLQEDQSKEKRR
 SENP7 P---KVPKQDNSSDCGVYLYQYVESFFKDPVNFELPHLEK-WFPRHVIKTKREDIRELIIILKLHLQQQKGS
 SENP8 -VEEKAPAQQNSYDCGMVICNTEALCQNFRRQTE-----SLLQLLTPAYITKKRGEWKDLITTLAKK---

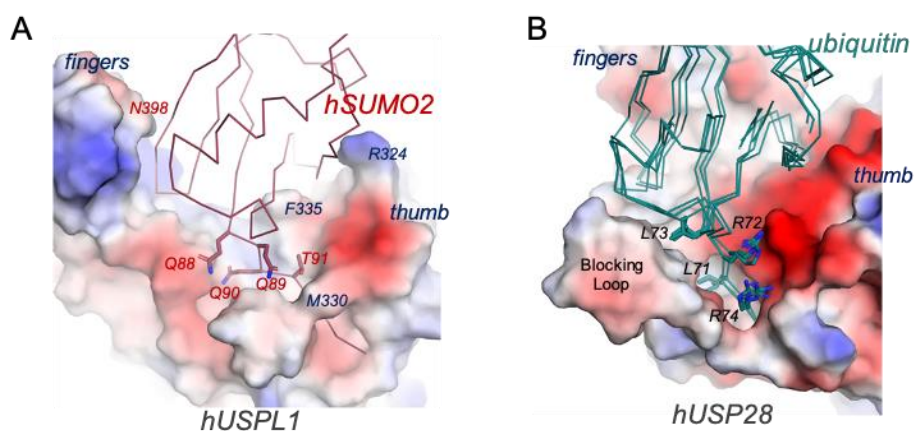
Supplementary Figure 1. Structural alignment of the NopD catalytic domain with the catalytic domains of XopD, SENP1, SENP2, SENP3, SENP4, SENP5, SENP6, SENP7 and SENP8. A yellow backdrop indicates side chain identity with 100% conservation in the alignment. Red represents three catalytic residues.

DISCUSSION

In all chapters, we solved four Ubiquitin/SUMO-related protein complexes structures: USPL1-SUMO2, SENP7-SUMO2, NopD-AtSUMO2, and NopD-ubiquitin. We solved their three-dimensional crystal structures, analyzed the structures, and characterized the role of the key residues of the interface in each structure.

In **Chapter I**, USPL1 is a distant member of the family with the canonical right-hand scaffold of a ubiquitin USP catalytic domain, but exposing key surface elements unique for SUMO binding, a low-identity UbL in comparison to ubiquitin.

All members of the human USP family of DUBs show specificity towards ubiquitinated substrates, with the exception of USP18, which is specific towards ISG15, a double-headed UbL modifier similar to ubiquitin, and USPL1, which shows specificity towards SUMO, a different UbL modifier. USPs are multidomain proteins that utilize different strategies to ensure their specific functions in the cell. In some cases the presence of additional domains adjacent to the USP catalytic domain enhances substrate binding or provides specificity towards a particular type of polyubiquitin chain. Commonly, in all members of the USP family, the structural fold of the catalytic domain, formed by the characteristic palm, thumb and fingers subdomains reminiscent of a human right hand, is maintained, exposing all the key surface elements necessary for interaction with ubiquitin.



FigureD 1. Comparison of the C-terminal tails of SUMO2 and ubiquitin. A. Electrostatic potential surface representation for the USPL1 in complex with SUMO2 (red line). The C-terminal tail of SUMO (QQQTGG) is labeled and shown in stick

DISCUSSION

representation. Main interface contacts of thumb and fingers subdomains are labeled. **B.** Electrostatic potential surface representation for the USP28 (PDB code 6HEK) in complex with ubiquitin (green lines) from complexes with USP2 (PDB code 2hd5), USP7 (PDB code 5JTV), and USP30 (PDB code 5OHK) [61,65,155,166]. The C-terminal tail of ubiquitin (LRLRGG) is labeled and shown in stick representation.

Possibly, a major difference between USPL1 and the other USP members can be found in the interface formed by the dissimilar C-terminal tail of ubiquitin (LRLRGG) and SUMO (QQQTGG) (**FigureD 1**). In addition to different hydrogen bond contacts engaged by the backbone, the presence of two arginines and two leucines in opposite sides of the ubiquitin tail generates unique interactions with acidic and hydrophobic surface patches in all ubiquitin USPs (**FigureD 1**), including USP18, in which ISG15 has a similar tail as ubiquitin. Interestingly, such hydrophobic patch in all ubiquitin-specific USPs is formed by the so-called blocking loops, which are structurally rearranged upon binding to the ubiquitin substrate. Interestingly, in all USPL1 orthologs the two blocking loops sequences have been deleted and are absent in the structure. Thus, in contrast to the high number of specific contacts engaged by the -LRLR- motif of the ubiquitin C-terminal tail, in the USPL1-SUMO2 structure only the Gln90 side chain of the equivalent -QQQT- motif forms specific interactions with USPL1, and all other contacts in the C-terminal tail of SUMO are established by backbone atoms (**Figure ChapterI 4**).

In this chapter we want to emphasize the plasticity of the conserved catalytic domain of USPs, which despite being designed to bind ubiquitin with high specificity, has evolved to interact with SUMO in USPL1. SUMO and ubiquitin share a 16% sequence identity, basically observed in the structural elements of the globular domain. However, the non-conserved residues in their surface participate in unique and specific protein-protein interactions with USP family members. USPL1 is a paradigmatic example of divergent evolution in the USP family, in which the aforementioned surface substitutions facilitate the interaction with SUMO, but maintain the right hand-like subdomain scaffold of the catalytic domain.

The catalytic domain of USPL1 is embedded in the middle of a structurally disordered full-length protein, which does not appear to have any other obvious well-folded domain in the sequence, as seen in the predicted model from AlphaFold-2 [152]. This type of domain organization of the full-length protein is often present in the SENP/ULP protease family,

in which long disordered protein extensions are usually found adjacent to the catalytic domain [97]. As for the SENP/ULP family, we cannot rule out a functional role for these long non-catalytic disordered extensions of the full-length USPL1, such as in the recruitment of SUMO substrates or in the regulation of the proteolytic activity of the catalytic domain, perhaps by internal protein-protein interactions. USPL1 has been reported to be involved in the regulation of the RNA polymerase-II-mediated snRNA transcription in the Cajal Bodies (CBs), a membrane-less compartment in the nucleus where USPL1 has been localized. However, the essential function of USPL1 in the CBs biology seems to be independent of the SUMO protease activity, perhaps the long disordered extensions of USPL1 participate in protein-protein interactions of this membrane-less compartment. In any case, the relevance of the SUMO protease activity of USPL1 in the cellular context still needs to be disclosed.

We have unraveled the structural determinants for the unique specificity of USPL1 for SUMO2 in **Chapter I**, a paradigmatic example of divergent evolution in the ubiquitin USP family to interact with a distant UbL family member.

In **Chapter II**, the structure of the complex between SENP7 and SUMO2 reveals all the contacts in the extended quilt-like interface between SENP7 catalytic domain and the chemically-modified SUMO2 suicide substrate (SUMO2-PA).

SENP6 and SENP7 are the most divergent members of the human SENP/ULP SUMO protease family, structurally containing long insertions in the middle of the catalytic domain, and showing a isoform preference for SUMO2/3, in contrast to SENP1 and SENP2 that are equally active for SUMO1 and SUMO2/3 conjugates. Furthermore, the catalytic domain of SENP7 (and SENP6) cleaves preferentially isopeptidic bonds in conjugated substrates, and it is particularly active in the disassembly of polySUMO2 chains, over the cleavage of the α -peptide during the maturation process of SUMO precursors [114,116].

The full-length of all SENP/ULP family members contain long extensions structurally disordered before the conserved catalytic domain that displays a regular globular structure belonging to the C48 cysteine protease class. SENP6 and SENP7 are the longest members of the SENP/ULP family, with 1112 and 984 residues, respectively, with the

DISCUSSION

catalytic domain located in the terminal part of the polypeptidic chain. SENP6 and SENP7 also contain a long disordered insertion in the middle of the catalytic domain, namely Loop3, with 150 and 50 residues long, respectively. It has been speculated on the role of the disordered extensions in the SENP/ULP family, which probably may be involved in substrate regulation activities inside the cell. However, in all conducted in vitro experiments the presence of the catalytic domain, without the disordered extensions, suffices for a full proteolytic activity against standard substrates, either for the cleavage of single SUMO2 conjugates or in the dismantling of polySUMO2 chains.

Interestingly, examples of cellular functions attributed to the disordered extensions in the SENP6 and SENP7 members have been reported in the recent years. In SENP7 seven potential SIM motifs (SUMO-Interacting motifs), located in the disordered region outside the catalytic domain, have been involved in homologous recombination by providing specific regions for the recruitment of polySUMOylated KAP1 (Karyopherin 1) to be further deSUMOylated [121]. In SENP6 eight SIM motifs are responsible to recruit chromatin proteins through the polySUMO2 chain to be subsequently cleaved off by the peptidase [192]. Also in ULP2, an equivalent member of the family in yeast with higher affinity for polySUMOylated substrates, SIM motifs are also involved in the recruitment of specific substrates [193]. In all cases there seems to be a common role of SIM motifs located in the disordered protein regions to provide an specific anchor to recruit polySUMO2 chain substrates. Nevertheless, as mentioned before, such long disordered regions are dispensable for the catalytic activity of all SENP/ULP family members in vitro, even for the dismantling of polySUMO2 chains, a property that resides exclusively in the globular catalytic domain of the SENP/ULP family members.

The major role of Loop1 in the function of SENP6 and SENP7 for cleaving off SUMO2 conjugated substrates has been verified [114,158]. Here we can disclose the details of the contact interface between SENP7 Loop1 with the globular domain of SUMO2, being able to visualize the direct contacts of SUMO2 Asn68 and Pro66 in the binding interface with Loop1 [115]. In contrast, the corresponding region in SENP1 and SENP2, formed only by a short β 1- β 2 hairpin loop (Figure ChapterII 4), does not play any relevant role in their proteolytic activity for SUMO substrates. On the opposite side of the SUMO interface,

most contacts between SENP7 and SUMO2 are mostly conserved in the other members of the SENP/ULP family, including the contacts engaged by the C-terminal tail, and by Asp712, Phe741, Trp773 and Arg775, most of them proven to be essential in the deSUMOylating activity [151,181].

Thus we hypothesized that the Loop1 insertion observed in SENP6 and SENP7 might be a consequence of an ultimate adaptation of these peptidases to provide a particular specificity for the SUMO2/3 isoforms. SENP7 (and SENP6) are essentially deSUMOylating peptidases of conjugated substrates, showing very little proteolytic activity in the maturation process of the three human SUMO precursors (in clear contrast to SENP1 and SENP2). In the particular case of SENP6, the *in vitro* deconjugation activity dismantling polySUMO2 chains is quite remarkable [114,194]. SENP7 (and SENP6) have been essentially localized throughout the nucleoplasm inside the cell and have been proposed to exert a major role in the regulation of the levels of polySUMO2 chains. The regulation of polySUMO2 chains levels has been demonstrated in experiments with SENP6-depleted mammalian cell cultures, where the accumulation of polySUMO2/3 conjugated in substrates led to a reduced cell survival and mitotic progression problems [195,196].

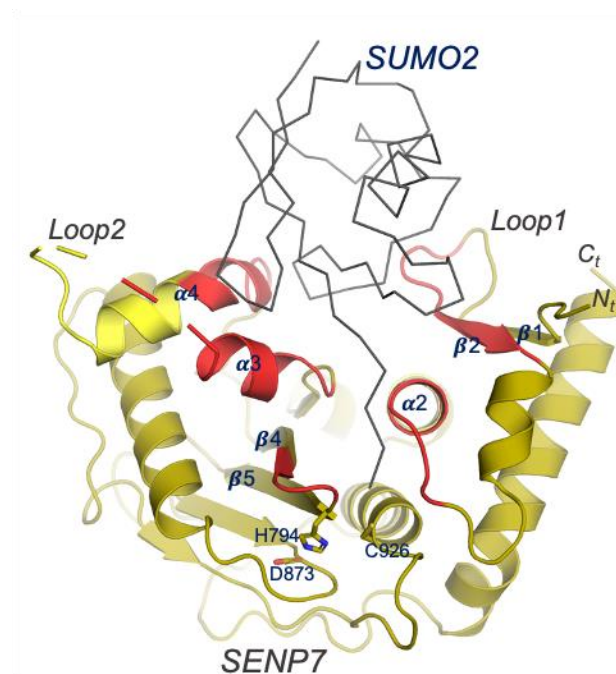


Figure 2. Interface between SENP7 and SUMO2. Cartoon representations of the SENP7-SUMO2 complex structure. SENP7 catalytic domain and SUMO2 precursor are shown in yellow and blue, respectively. SENP7 interface regions are colored in red. The catalytic residues are labelled and depicted in stick representation. Loop 1, Loop2 are labelled. N- and C-terminal are marked.

DISCUSSION

In summary, the specificity of the different members of the human SENP/ULP deSUMOylating family for the SUMO isoforms can be attributed to a few regions in the interface with the SUMO globular domain (**Figure 2 & Chapter II Sup 1**): in the N-terminal region in one side of the interface, with residues emerging from the $\beta 1$ - $\beta 2$ loop (Loop1 in SENP7) and from the $\alpha 2$ helix; and in the opposite side of the interface, with residues from $\alpha 3$ and $\alpha 4$ (connected by the Loop2 in SENP7) and from the $\beta 4$ - $\beta 5$ hairpin loop. Probably, in addition to single residue changes during evolution, loop insertions between conserved secondary structure elements has been a common feature in the adaptation of the deSUMOylase family towards particular activities. Such as the Loop1 insertion between $\beta 1$ - $\beta 2$ in SENP6 and SENP7, which confers specificity for SUMO2/3 isoforms; or the loop insertion between $\beta 4$ - $\beta 5$ in SENP8/DEN1/NEDP1, which confers specificity for the C-terminal tail of a different Ubl modifier, Nedd8 [104,191]. Interestingly, low homology members of the same CE protease clan present in pathogenic bacteria have evolved to act as bacterial effector proteases by developing specific insertions in the binding interface elements to provide particular activities, in this instance towards ubiquitin substrates, instead of SUMO [148], thus highlighting the high adaptability of the SENP/ULP proteolytic fold to cope with particular functions.

In **Chapter III**, we crystallized and solved the structures of the NopD- AtSUMO2 and NopD-ubiquitin complexes, and analyzed the crystal structures and study the interface residues involved in each of NopD complexes.

The CE protease clan in humans, the SENP/ULP family, consists of six SUMO-specific and one Nedd8-specific members. In human the conserved catalytic domain has evolved to distinguish between SUMO isoforms and it has also been adapted for a different Ubl modifier, Nedd8, but interestingly not for ubiquitin despite the high identity with Nedd8. Secreted effector proteins in infectious bacteria also contain CE proteases to perturb the host response processes. However in bacteria, in addition to a deSUMOylase activity, the CE protease clan has evolved to acquire deubiquitinating, or even an unusual acetyltransferase activities, such as in *Yersinia pestis* YopJ [197]. Thus, it seems that the conserved catalytic fold of the CE protease family can evolve in different scenarios to cleave off different types of Ubl modifiers.

It is interesting how in bacteria, the CE catalytic fold has adopted distinct structural features to cleave off host-cell ubiquitin, compared to the known human SUMO-specific proteases. Also, so far all infectious human pathogens show a preference for targeting K63-linkage ubiquitin chains, which are involved in inflammatory signaling cascades. In contrast, the plant pathogen XopD, from *Xanthomonas campestris*, prefers the K48-linkage ubiquitin chains over K63- and K6-linkages, indicating that the versatility of the CE fold is also able to modulate the cleavage of different types of ubiquitin chains. XopD has also the unusual property to possess a dual activity for both SUMO and ubiquitin simultaneously. Structural analysis revealed that the XopD activity for ubiquitin was acquired by the formation of a novel protein-protein interface, conducted by an N-terminal extension that interacts with the ubiquitin C-terminal Arg72 and with the promiscuous I44 ubiquitin patch. In contrast, the XopD binding to plant SUMO was conducted through the standard SENP/ULP interface.

In this Chapter we reveal that NopD, an effector protease from *Rhizobium* involved in symbiotic plant nodulation, possesses also unusual dual deubiquitinating and deSUMOylating activities. Like XopD, NopD also prefers the cleavage of K48-linkage ubiquitin, in contrast to K63-linkage preference of animal pathogens, probably indicating the use of the different ubiquitin linkages in signaling defense pathways between plants and animals. Curiously, the structure/functional analysis of NopD in complex with ubiquitin reveals a different binding mechanism compared to XopD, serving as a clear example of convergent evolution to cleave off ubiquitin by using different protein-protein interface. NopD interacts with ubiquitin and SUMO through the same protein interface, in contrast to XopD, which absolutely relied on the presence of an N-terminal extension to bind ubiquitin.

A notable structural characteristic of NopD, absent in XopD, is the presence of a Loop insert between $\beta 3$ and $\beta 4$ strands in the catalytic domain that fixes the C-terminal tail of both SUMO and ubiquitin. This evolutive trait in NopD resembles an analogous insert in SENP8/NEDP1, which contributed to the specificity for Nedd8 over ubiquitin or SUMO. In NopD, *in vitro* functional analysis reveals the essential role of this Loop insert for both SUMO and ubiquitin activities. A second unique characteristic of NopD to favor ubiquitin

DISCUSSION

binding, not present in XopD, is the electrostatic interaction between Glu840 and Arg72 of the C-terminal tail of ubiquitin. In XopD, Arg72 interacts with a glutamic acid from the N-terminal extension, however in NopD, Arg72 is engaged with Glu840 (glycine in XopD). The ubiquitin C-terminal tail Arg72 has usually been employed as a major binding point signature in different families of deubiquitinases, such as in USPs, OTUs [69,153,198,199]. Interestingly, in our functional analysis of the NopD-SUMO2 complex the E840A point mutant did not play any role in the deSUMOylating activity, in contrast to its major role in ubiquitin binding, underling Glu840 as an evolutive trait in Rhizobium NopD to be able to bind ubiquitin.

Whereas XopD forms a different interface to bind ubiquitin, NopD uses the same interface to bind both SUMO and ubiquitin, only presenting a few number of different contacts between both interfaces. This cross-reactivity of NopD for SUMO and ubiquitin using a similar interface is quite unusual in the CE clan. Only in some other human pathogen effectors, such as ChlaDUB1 and RickCE, cross-reactivity can be found for ubiquitin and Nedd8 using a similar interface, but in this case both modifiers share a 58% sequence identity.

CONCLUSIONS

Chapter I. Structural basis for the SUMO protease activity of the atypical ubiquitin-specific protease USPL1.

- The USPL1-SUMO2 crystal structure reveals that USPL1 is a distant member of the family with the canonical right-hand scaffold of a ubiquitin USP catalytic domain, but exposes key surface elements unique for SUMO binding.
- In all USPL1 orthologs the two blocking Loops are absent in the structure, which are responsible for forming the hydrophobic patch in all ubiquitin-specific USPs.
- All contacts in the C-terminal tail of SUMO are established by backbone atoms, only the Gln90 side chain of the -QQQT- motif of SUMO2 forms specific interactions with USPL1.
- In all three USPL1 subdomains, palm, fingers and thumb, the key residues for the specific interaction with ubiquitin have been deleted or, in some cases, replaced by residues specific for interacting with SUMO2.

Chapter II. Structural basis for the SUMO2 isoform specificity of SENP7.

- The crystal structure of SENP7-SUMO2 has been solved at 1.74 Å resolution and reveals that the active site catalytic triad is already formed in the absence of the SUMO substrate.
- The details of all interface contacts between SENP7 and SUMO2 are disclosed. The SUMO2 C-terminal tail interface is basically formed by mainchain hydrogen bonds with SENP7.
- The interface analysis of the globular part of SUMO2 with SENP7 reveals the contacts of the insertion Loop1 in SENP7, which confers isopeptidase specificity for human SUMO2/3 isoforms.

Chapter III. Structural analysis of bradyrhizobium NopD provides insights into the dual protease activity for ubiquitin and SUMO.

- In addition to plant SUMO, the rhizobium NopD effector has also activity for ubiquitin.
- The crystal structure of the NopD-AtSUMO2 and NopD-Ubiquitin complexes reveal the interface details for the unusual dual activity of NopD.
- The ubiquitin-binding mechanism differs from the homolog *X. campestris* XopD, where the N-terminal extension is very important. In NopD, a unique Loop insertion is essential for its binding ability and activities for both substrates, SUMO and ubiquitin.
- Mutagenesis and activity assays confirm the role of key interface residues in the dual NopD activities for SUMO and ubiquitin, such as the crucial role of Glu840 in NopD to fix the C-terminal Arg72 of ubiquitin.

GENERAL EXPERIMENTAL METHOD

Plasmid and cloning

We designed the constructs by other papers and prediction program Phyre2 or Alpha-fold program. We obtained the protein gene by Addgene (<http://n2t.net/addgene:85760>) or ThermoFisher (<https://www.thermofisher.com/order/gene-design/index.html>).

All primers were designed by SnapGene. Normally Primers length are around 40bp per primer. There are some ways to get the construct. After Amplification PCR, the target gene can be amplified by PCR using Phusion polymerase and cloned into the restriction enzymes sites (e.g. BamHI/NotI) of pET28a vector using ligation. Or constructs can be generated by Restriction Enzyme Free PCR. The point mutats constructs were designed by different primers and were created by the QuickChange site-directed mutagenesis kit (Stratagene). Here, we used Restriction Enzyme Free PCR method to generate most constructs.

Amplification PCR is the first PCR that we need to carry out when we get the primers from the company. After that, the PCR results are loaded in an 1% agarose gel (w/v), then cut the target gene band and purify the gene fragment by GeneJET PCR Purification Kit (Thermo Scientific). All procedures are carried in Prime Thermal Cycler. The specific steps and required materials are as follows.

GENERAL EXPERIMENTAL METHOD

Cycle Step			Component	
<i>Amplification PCR</i>	TEMP	TIME	10ul	10X Buffer
Initial Denaturation	98°C	30 seconds	1ul	10mM dNTPs
30 cycles	98°C	10 seconds	0.5ul	Forward primer
			0.5ul	Reverse primer
	55°C	30 seconds	100ng	Plasmid vector
	72°C	30 seconds per kb	0.5ul	Phusion
Final Extension	72°C	10 minutes	H ₂ O	added to 50ul
Hold	10°C	hold		

Restriction Enzyme Free PCR (RF-PCR) is a very simple method that can insert DNA fragment into any location in a plasmid without restriction enzyme [159]. Procedure is simple and only run two different PCR. After two round RF-PCR, there is no unwanted extra residues in the circular plasmid. The primers can be designed in URL: <https://www.rf-cloning.org/>.

Restriction Enzyme Free PCR 1 / STEP	Amplification PCR
--------------------------------------	-------------------

Cycle Step			Component	
<i>Restriction Enzyme Free PCR 2 / STEP</i>	TEMP	TIME	5ul	5X Buffer
Initial Denaturation	98°C	2 minutes	0.5ul	10mM dNTPs
35 cycles	98°C	30 seconds	250ng	Gene fragment
			25ng	Plasmid vector
	55°C	1 minutes	0.75ul	DMSO
	72°C	6 minutes (whole plasmid)	0.25ul	Phusion
Final Extension	72°C	10 minutes	H ₂ O	added to 25ul
Hold	10°C	hold		

Mutagenesis PCR were constructed by QuikChange mutagenesis kit (Stratagene) and Phusion High-Fidelity DNA Polymerase.

Cycle Step			Component	
Mutagenic PCR	TEMP	TIME	10ul	10X Buffer
Initial Denaturation	98°C	30 seconds	1ul	10mM dNTPs
16 cycles	98°C	10 seconds	0.1ul	Forward primer
			0.1ul	Reverse primer
	50°C	30 seconds	30ng	Plasmid vector
	72°C	30 seconds per kb	0.5ul	Phusion
			1.5ul	DMSO
Final Extension	72°C	10 minutes	H2O	added to 50ul
Hold	10°C	hold		

Transformation to XL1-Blue strain competent cells

Take the XL1B competent cells (*E. coli*) and place them in an ice bath. The recommended dosage for one transformation of competent cells is 50-100 μ l, which can be adjusted according to the actual situation. After the competent cells are thawed, add the target DNA plasmid to the competent cell suspension. The plasmid volume cannot exceed 10% of competent cells volume. Gently mix and ice bath for 20 minutes. Heat shock at 42°C for 90 seconds in water bath, quickly transfer the centrifuge tube to an ice bath, and let stand on ice for 2-3 minutes. Add 1 mL LB medium (without antibiotics) to each centrifuge tube, mix well and place at 37°C shaker at 250 rpm for 1 hour to recover the bacteria. Centrifuge the cell in 4°C at 3000rpm for 5min. On the agar medium (warm a little bit in 37°C before using), spread the cells evenly with a sterile spreading rod, and place the plate on the table until the liquid is absorbed. Invert the plate and incubate at 37°C for 12-16 hours.

Colony PCR and plasmid sequencing

Colony PCR is a regular amplification PCR by using Taq DNA Polymeras but instead of using plasmid, 2ul colony is needed. According to the plasmid, we use T7 oligonucleotides primer to carry colony PCR. After PCR, load the sample to 1% agarose gel (w/v) to check the DNA band size before sending to sequencing.

Cycle Step			Compotent	
<i>Mutagenic PCR</i>	TEMP	TIME	1ul	10X Buffer
Initial Denauration	95°C	5 mimutes	0.2ul	10mM dNTPs
28 cycles	94°C	30 seconds	0.1ul	Forward primer
			0.1ul	Reverse primer
	50°C	30 seconds	2ul	colony
	72°C	1 mimutes per kb	0.5ul	Taq DNA Polymeras
Final Extension	72°C	10 minutes	H ₂ O	added to 10ul
Hold	94°C	hold		

Incubate the positive colonies in 10mL LB at 37°C shaker at 250 rpm for O/N. we use GeneJET Plasmid Miniprep Kit (Thermo Fisher) to extract plasmid. And then check the concentration of the plasmid by NanoDrop, and send the plasmid to the Servei de Genòmica Bioinformàtica (UAB) to check the DNA sequence by Sanger sequencing reactions. To compare the sequence of plasmid and Genebank, ExpASy-translate tool and Protein- Blast of NCBI is used for confirming the DNA sequence.

Transformation to Rosetta2 or BL21 competent cells

The procedure is same as the method of transformation to XL1-Blue strain competent cells (E.coli). Rosetta 2 or BL21 competent cells are better for protein expression than XL1B. The Rosetta series strains are derived from the BL21 series host bacteria, and Rosetta2 (DE3) is derived from Rosetta (DE3). This strain contains a pRARE2 plasmid. In addition to providing tRNA with six rare codons of AUA, AGG, AGA, CUA, CCC, and GGA contained in the original Rosetta(DE3) host, The tRNA for the seventh rare codon CGG is

also provided. At the same time, the pRARE2 plasmid is chloramphenicol resistant. By providing rare codons, the Rosetta2 (DE3) vector enables the host bacteria to provide more "universal" protein expression compared to other E. coli, thereby improving the expression level of the target protein.

Protein Expression Test

After obtaining colonies from Rosetta 2 or BL21. Incubate the colonies at 37 °C in 5mL LB at 37°C shakers at 250 rpm for O/N. Take 2mL E.coli culture to 15mL LB (with antibiotic), cells grow at 37 °C 250 rpm until OD= 0.6- 0.8, then add 0.5mM IPTG. Incubate cultures at 250 rpm at 30°C 5h or 20°C O/N, and centrifuge at 4°C at 3000rpm to harvest cells. Suspend pellet in 1mL 20% sucrose, 50mM Tris 8 buffer. Add lysis buffer: 150 µl NaCl, 15 µl 10% IGEPAL, 7.5 µl 2M Imidazole, and 1mM BME or DTT. Cells were broken by sonicator at 35% amplitude for 1min (2s on/ 3s off) at 22°C. Centrifuge the samples at 4°C 10000rpm for 15min to get the supernatant. Most proteins here have a histidine tag. Incubate the supernatant with Ni Sepharose 6 Fast Flow (GE Healthcare) for 20min at 4°C. Imidazole and Histidine-tag compete for Nickle, and as the concentration of imidazole increases, the protein bound to Nickle will be eluted. To achieve fractional elution of proteins, the purpose of protein separation is achieved. 1mL Buffer A (350mM NaCl, 20mM Tris 8, 10mM Imidazole, 1mM BME) was added to elute other proteins or proteins that were not bound to the nickel column and centrifuge at 6000rpm for 1min. Repeat the above steps twice. Then 70 µl Buffer B (350mM NaCl, 20mM Tris 8, 300mM Imidazole, 1mM BME) was added, gently mixed, and centrifuged at 6000rpm for 1min. Check the protein expression status by loading samples on a 12.5% SDS-PAGE gel, and stained by Coomassie brilliant blue.

Protein expression and purification

To get more protein for crystallization, prepare 8 liters of LB by sterilization. Bacteria were grown at 37 °C to OD₆₀₀=0.6~0.8, and IPTG was added to a final concentration of 0.5 mM. Bacteria were grown in the good condition that gets from the expression test and harvested by centrifugation at 4°C at 5000rpm for 15min. Cell suspensions were equilibrated in 350 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10 mM imidazole, 20% sucrose, 1 mM DTT, and 0.1% IGEPAL CA-630, and cells were broken by sonication. After removing

GENERAL EXPERIMENTAL METHOD

cell debris by centrifugation, proteins were separated from the lysate by nickel affinity chromatography using Ni Sepharose 6 Fast Flow and eluted with lysis buffer including 20 mM Tris-HCl (pH 8.0), 350 mM NaCl, 300 mM imidazole, and 1 mM DTT. Fractions containing the target protein were collected, diluted to 50 mM NaCl, applied to an anion exchange resin (Resource Q; GE Healthcare), and eluted with a 0-1 M NaCl gradient from 0 to 35% in 20 mM Tris-HCl (pH 8.0) and 3 mM DTT.

Protein concentration and Crystallization

Concentrated the protein using Amicon Ultra-30K ultrafiltration device (Millipore), change the protein buffer to lower buffer concentration, for example, 50mM NaCl, 5mM Tris 8. To obtain the crystal structure of the protein, it is necessary to prepare a large amount of purified protein (around 10 mg), the concentration of which is usually above 10 mg/ml and based on this, the crystallization conditions are screened.

After obtaining a high-purity protein solution, the next step is the cultivation of crystals. Similar to the formation of crystals of other compounds, protein crystals are formed slowly in saturated solutions. The conditions for growing crystals of each protein are different, and many variables affect the formation of crystals, including pH, ion concentration, protein concentration, temperature, metal ions, isoelectric point, etc.

There are two methods for crystallization here: sitting drop and hanging drop. The sitting drop method is usually used to complete protein crystallization by a robot. Using the Phoenix instrument, the protein and screen solution (1:1) can be quickly added to a 96-well plate, and the plate is covered with a film to form a closed environment for protein crystal. In the hanging drop method, droplets are prepared on a siliconized microscope coverslip by mixing 0.5-2 μ l of protein solution and an equal amount of precipitant solution. The coverslip is placed over the groove of a plate, and a portion of the groove is filled with about 500 μ l of the desired precipitant solution. Before the coverslip was placed, the chamber was sealed with grease around the groove. In the hanging drop method, if the surface tension of the protein solution is small, it will spread on the surface of the coverslip. At this time, the sitting drop method is more advantageous. Normally crystals are obtained by the sitting drop method first, and then more and larger crystals are obtained in batches by the pendant drop method. The rest protein is snap-frozen in

liquid nitrogen before storage at -80 °C.

Protein crystals contain an average of 50 percent solvent, they are easy to dry and break down if exposed to air. To protect protein crystals and prevent the formation of ice crystals, usually prepare cryo-protectants such as ethylene glycol, and glycerol when picking crystals. Crystals are typically mounted on 50-100 μm nylon rings, soaking in cryo-protectant 15% ethylene glycol for 5-15 seconds and then snap-frozen by immersion in liquid nitrogen.

Data Collection

Diffraction data were collected to a good resolution at beamline ID30B at the ESRF (Grenoble, France) or beamline BL13-XALOC at the ALBA synchrotron (Barcelona, Spain). Data processing was conducted by AutoProcesing with MxCUBE [160,161]. The structure was solved by molecular replacement with the Alphafold2 model as a search mode [152]. Following rounds of model building and refinement were carried out with Coot and Phenix [162,163]. Then the structure has been deposited in the Protein Data Bank under specific PDB codes to wait for the release.

PUBLICATIONS

Li, Y., Varejão, N., & Reverter, D. (2022). Structural basis for the SUMO protease activity of the atypical ubiquitin-specific protease USPL1. **Nature communications**, 13(1), 1819. <https://doi.org/10.1038/s41467-022-29485-0>.

Li, Y., & Reverter, D. (2021). Molecular Mechanisms of DUBs Regulation in Signaling and Disease. **International journal of molecular sciences**, 22(3), 986. <https://doi.org/10.3390/ijms22030986>.

Li, Y., Liu, Bing., & Reverter, D. (2020) A Novel Regulatory Mechanism to Regulate the Deubiquitinating Activity of USP25 by Oligomerization. **Journal of Cellular Signaling**. 2020; 1(4): 151-154.

Varejão, N., Lascorz, J., Li, Y., & Reverter, D. (2020). Molecular mechanisms in SUMO conjugation. **Biochemical Society transactions**, 48(1), 123–135. <https://doi.org/10.1042/BST20190357>.

REFERENCES

1. Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **1998**, *67*, 425–479, doi:10.1146/annurev.biochem.67.1.425.
2. Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **2009**, *78*, 477–513, doi:10.1146/annurev.biochem.78.081507.101607.
3. Pickart, C.M.; Eddins, M.J. Ubiquitin: structures, functions, mechanisms. *Biochim. Biophys. Acta* **2004**, *1695*, 55–72, doi:10.1016/j.bbamcr.2004.09.019.
4. Deng, L.; Meng, T.; Chen, L.; Wei, W.; Wang, P. The role of ubiquitination in tumorigenesis and targeted drug discovery. *Signal Transduct. Target. Ther.* **2020**, *5*, 11, doi:10.1038/s41392-020-0107-0.
5. Wing, S.S. Deubiquitinating enzymes--the importance of driving in reverse along the ubiquitin-proteasome pathway. *Int. J. Biochem. Cell Biol.* **2003**, *35*, 590–605, doi:10.1016/s1357-2725(02)00392-8.
6. Mevissen, T.E.T.; Komander, D. Mechanisms of Deubiquitinase Specificity and Regulation. *Annu. Rev. Biochem.* **2017**, *86*, 159–192, doi:10.1146/annurev-biochem-061516-044916.
7. Clague, M.J.; Urbé, S.; Komander, D. Breaking the chains: deubiquitylating enzyme specificity begets function. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 338–352, doi:10.1038/s41580-019-0099-1.
8. Cai, J.; Culley, M.K.; Zhao, Y.; Zhao, J. The role of ubiquitination and deubiquitination in the regulation of cell junctions. *Protein Cell* **2018**, *9*, 754–769, doi:10.1007/s13238-017-0486-3.
9. Gómez-Díaz, C.; Ikeda, F. Roles of ubiquitin in autophagy and cell death. *Semin. Cell Dev. Biol.* **2019**, *93*, 125–135, doi:10.1016/j.semcdb.2018.09.004.
10. Komander, D.; Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **2012**, *81*, 203–229, doi:10.1146/annurev-biochem-060310-170328.
11. Kwon, Y.T.; Ciechanover, A. The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy. *Trends Biochem. Sci.* **2017**, *42*, 873–886, doi:10.1016/j.tibs.2017.09.002.
12. Haahr, P.; Borgermann, N.; Guo, X.; Typas, D.; Achuthankutty, D.; Hoffmann, S.; Shearer, R.; Sixma, T.K.; Mailand, N. ZUFSP Deubiquitylates K63-Linked Polyubiquitin Chains to Promote Genome Stability. *Mol. Cell* **2018**, *70*, 165–174.e6, doi:10.1016/j.molcel.2018.02.024.
13. Amerik, A.Y.; Hochstrasser, M. Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta* **2004**, *1695*, 189–207,

- doi:10.1016/j.bbamcr.2004.10.003.
14. Walden, M.; Masandi, S.K.; Pawłowski, K.; Zeqiraj, E. Pseudo-DUBs as allosteric activators and molecular scaffolds of protein complexes. *Biochem. Soc. Trans.* **2018**, *46*, 453–466, doi:10.1042/BST20160268.
 15. Nepravishita, R.; Ferrentino, F.; Mandaliti, W.; Mattioni, A.; Weber, J.; Polo, S.; Castagnoli, L.; Cesareni, G.; Paci, M.; Santonico, E. CoCUN, a Novel Ubiquitin Binding Domain Identified in N4BP1. *Biomolecules* **2019**, *9*, doi:10.3390/biom9070284.
 16. Mevissen, T.E.T.; Hospenthal, M.K.; Geurink, P.P.; Elliott, P.R.; Akutsu, M.; Arnaudo, N.; Ekkebus, R.; Kulathu, Y.; Wauer, T.; El Oualid, F.; et al. OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. *Cell* **2013**, *154*, 169–184, doi:10.1016/j.cell.2013.05.046.
 17. Faesen, A.C.; Dirac, A.M.G.; Shanmugham, A.; Ovaa, H.; Perrakis, A.; Sixma, T.K. Mechanism of USP7/HAUSP activation by its C-terminal ubiquitin-like domain and allosteric regulation by GMP-synthetase. *Mol. Cell* **2011**, *44*, 147–159, doi:10.1016/j.molcel.2011.06.034.
 18. Sowa, M.E.; Bennett, E.J.; Gygi, S.P.; Harper, J.W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **2009**, *138*, 389–403, doi:10.1016/j.cell.2009.04.042.
 19. Kaushal, K.; Antao, A.M.; Kim, K.-S.; Ramakrishna, S. Deubiquitinating enzymes in cancer stem cells: functions and targeted inhibition for cancer therapy. *Drug Discov. Today* **2018**, *23*, 1974–1982, doi:10.1016/j.drudis.2018.05.035.
 20. He, M.; Zhou, Z.; Shah, A.A.; Zou, H.; Tao, J.; Chen, Q.; Wan, Y. The emerging role of deubiquitinating enzymes in genomic integrity, diseases, and therapeutics. *Cell Biosci.* **2016**, *6*, 62, doi:10.1186/s13578-016-0127-1.
 21. Kwon, S.-K.; Saindane, M.; Baek, K.-H. p53 stability is regulated by diverse deubiquitinating enzymes. *Biochim. Biophys. acta. Rev. cancer* **2017**, *1868*, 404–411, doi:10.1016/j.bbcan.2017.08.001.
 22. Burska, U.L.; Harle, V.J.; Coffey, K.; Darby, S.; Ramsey, H.; O'Neill, D.; Logan, I.R.; Gaughan, L.; Robson, C.N. Deubiquitinating enzyme Usp12 is a novel co-activator of the androgen receptor. *J. Biol. Chem.* **2013**, *288*, 32641–32650, doi:10.1074/jbc.M113.485912.
 23. Imamura, T.; Oshima, Y.; Hikita, A. Regulation of TGF- β family signalling by ubiquitination and deubiquitination. *J. Biochem.* **2013**, *154*, 481–489, doi:10.1093/jb/mvt097.
 24. Lopez-Castejon, G.; Edelmann, M.J. Deubiquitinases: Novel Therapeutic Targets in Immune Surveillance? *Mediators Inflamm.* **2016**, *2016*, 3481371, doi:10.1155/2016/3481371.
 25. Gu, Z.; Shi, W. Manipulation of viral infection by deubiquitinating enzymes: new players in host-virus interactions. *Future Microbiol.* **2016**, *11*, 1435–1446,

- doi:10.2217/fmb-2016-0091.
26. Mungamuri, S.K.; Qiao, R.F.; Yao, S.; Manfredi, J.J.; Gu, W.; Aaronson, S.A. USP7 Enforces Heterochromatinization of p53 Target Promoters by Protecting SUV39H1 from MDM2-Mediated Degradation. *Cell Rep.* **2016**, *14*, 2528–2537, doi:10.1016/j.celrep.2016.02.049.
 27. McClurg, U.L.; Robson, C.N. Deubiquitinating enzymes as oncotargets. *Oncotarget* **2015**, *6*, 9657–9668, doi:10.18632/oncotarget.3922.
 28. Hetfeld, B.K.J.; Helfrich, A.; Kapelari, B.; Scheel, H.; Hofmann, K.; Guterman, A.; Glickman, M.; Schade, R.; Kloetzel, P.-M.; Dubiel, W. The zinc finger of the CSN-associated deubiquitinating enzyme USP15 is essential to rescue the E3 ligase Rbx1. *Curr. Biol.* **2005**, *15*, 1217–1221, doi:10.1016/j.cub.2005.05.059.
 29. Mennerich, D.; Kubaichuk, K.; Kietzmann, T. DUBs, Hypoxia, and Cancer. *Trends in cancer* **2019**, *5*, 632–653, doi:10.1016/j.trecan.2019.08.005.
 30. Jacomin, A.-C.; Taillebourg, E.; Fauvarque, M.-O. Deubiquitinating Enzymes Related to Autophagy: New Therapeutic Opportunities? *Cells* **2018**, *7*, doi:10.3390/cells7080112.
 31. Nguyen, H.H.; Kim, T.; Nguyen, T.; Hahn, M.-J.; Yun, S.-I.; Kim, K.K. A Selective Inhibitor of Ubiquitin-Specific Protease 4 Suppresses Colorectal Cancer Progression by Regulating β -Catenin Signaling. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* **2019**, *53*, 157–171, doi:10.33594/000000127.
 32. Eichhorn, P.J.A.; Rodón, L.; González-Juncà, A.; Dirac, A.; Gili, M.; Martínez-Sáez, E.; Aura, C.; Barba, I.; Peg, V.; Prat, A.; et al. USP15 stabilizes TGF- β receptor I and promotes oncogenesis through the activation of TGF- β signaling in glioblastoma. *Nat. Med.* **2012**, *18*, 429–435, doi:10.1038/nm.2619.
 33. Qu, Z.; Zhang, R.; Su, M.; Liu, W. USP13 serves as a tumor suppressor via the PTEN/AKT pathway in oral squamous cell carcinoma. *Cancer Manag. Res.* **2019**, *11*, 9175–9183, doi:10.2147/CMAR.S186829.
 34. Lork, M.; Verhelst, K.; Beyaert, R. CYLD, A20 and OTULIN deubiquitinases in NF- κ B signaling and cell death: so similar, yet so different. *Cell Death Differ.* **2017**, *24*, 1172–1183, doi:10.1038/cdd.2017.46.
 35. Chen, Y.; Li, Y.; Xue, J.; Gong, A.; Yu, G.; Zhou, A.; Lin, K.; Zhang, S.; Zhang, N.; Gottardi, C.J.; et al. Wnt-induced deubiquitination FoxM1 ensures nucleus β -catenin transactivation. *EMBO J.* **2016**, *35*, 668–684, doi:10.15252/embj.201592810.
 36. Han, W.; Lee, H.; Han, J.-K. Ubiquitin C-terminal hydrolase37 regulates Tcf7 DNA binding for the activation of Wnt signalling. *Sci. Rep.* **2017**, *7*, 42590, doi:10.1038/srep42590.
 37. Sanchez-Diaz, P.C.; Chang, J.C.; Moses, E.S.; Dao, T.; Chen, Y.; Hung, J.Y. Ubiquitin carboxyl-terminal esterase L1 (UCHL1) is associated with stem-like

- cancer cell functions in pediatric high-grade glioma. *PLoS One* **2017**, *12*, e0176879, doi:10.1371/journal.pone.0176879.
38. An, T.; Gong, Y.; Li, X.; Kong, L.; Ma, P.; Gong, L.; Zhu, H.; Yu, C.; Liu, J.; Zhou, H.; et al. USP7 inhibitor P5091 inhibits Wnt signaling and colorectal tumor growth. *Biochem. Pharmacol.* **2017**, *131*, 29–39, doi:10.1016/j.bcp.2017.02.011.
 39. Huang, G.; Li, L.; Zhou, W. USP14 activation promotes tumor progression in hepatocellular carcinoma. *Oncol. Rep.* **2015**, *34*, 2917–2924, doi:10.3892/or.2015.4296.
 40. Yun, S.-I.; Kim, H.H.; Yoon, J.H.; Park, W.S.; Hahn, M.-J.; Kim, H.C.; Chung, C.H.; Kim, K.K. Ubiquitin specific protease 4 positively regulates the WNT/ β -catenin signaling in colorectal cancer. *Mol. Oncol.* **2015**, *9*, 1834–1851, doi:10.1016/j.molonc.2015.06.006.
 41. Xing, C.; Lu, X.-X.; Guo, P.-D.; Shen, T.; Zhang, S.; He, X.-S.; Gan, W.-J.; Li, X.-M.; Wang, J.-R.; Zhao, Y.-Y.; et al. Ubiquitin-Specific Protease 4-Mediated Deubiquitination and Stabilization of PRL-3 Is Required for Potentiating Colorectal Oncogenesis. *Cancer Res.* **2016**, *76*, 83–95, doi:10.1158/0008-5472.CAN-14-3595.
 42. Zhou, Y.; Liang, P.; Ji, W.; Yu, Z.; Chen, H.; Jiang, L. Ubiquitin-specific protease 4 promotes glioblastoma multiforme via activating ERK pathway. *Onco. Targets. Ther.* **2019**, *12*, 1825–1839, doi:10.2147/OTT.S176582.
 43. Wang, Y.; Zhou, L.; Lu, J.; Jiang, B.; Liu, C.; Guo, J. USP4 function and multifaceted roles in cancer: a possible and potential therapeutic target. *Cancer Cell Int.* **2020**, *20*, 298, doi:10.1186/s12935-020-01391-9.
 44. van Andel, H.; Kocemba, K.A.; de Haan-Kramer, A.; Mellink, C.H.; Piwowar, M.; Broijl, A.; van Duin, M.; Sonneveld, P.; Maurice, M.M.; Kersten, M.J.; et al. Loss of CYLD expression unleashes Wnt signaling in multiple myeloma and is associated with aggressive disease. *Oncogene* **2017**, *36*, 2105–2115, doi:10.1038/onc.2016.368.
 45. Tauriello, D.V.F.; Haegebarth, A.; Kuper, I.; Edelman, M.J.; Henraat, M.; Canninga-van Dijk, M.R.; Kessler, B.M.; Clevers, H.; Maurice, M.M. Loss of the tumor suppressor CYLD enhances Wnt/ β -catenin signaling through K63-linked ubiquitination of Dvl. *Mol. Cell* **2010**, *37*, 607–619, doi:10.1016/j.molcel.2010.01.035.
 46. Yuan, T.; Chen, Z.; Yan, F.; Qian, M.; Luo, H.; Ye, S.; Cao, J.; Ying, M.; Dai, X.; Gai, R.; et al. Deubiquitinating enzyme USP10 promotes hepatocellular carcinoma metastasis through deubiquitinating and stabilizing Smad4 protein. *Mol. Oncol.* **2020**, *14*, 197–210, doi:10.1002/1878-0261.12596.
 47. Ge, W.-L.; Xu, J.-F.; Hu, J. Regulation of Oral Squamous Cell Carcinoma Proliferation Through Crosstalk Between SMAD7 and CYLD. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* **2016**, *38*, 1209–1217, doi:10.1159/000443069.

48. Zhao, Y.; Thornton, A.M.; Kinney, M.C.; Ma, C.A.; Spinner, J.J.; Fuss, I.J.; Shevach, E.M.; Jain, A. The deubiquitinase CYLD targets Smad7 protein to regulate transforming growth factor β (TGF- β) signaling and the development of regulatory T cells. *J. Biol. Chem.* **2011**, *286*, 40520–40530, doi:10.1074/jbc.M111.292961.
49. Takiuchi, T.; Nakagawa, T.; Tamiya, H.; Fujita, H.; Sasaki, Y.; Saeki, Y.; Takeda, H.; Sawasaki, T.; Buchberger, A.; Kimura, T.; et al. Suppression of LUBAC-mediated linear ubiquitination by a specific interaction between LUBAC and the deubiquitinases CYLD and OTULIN. *Genes Cells* **2014**, *19*, 254–272, doi:10.1111/gtc.12128.
50. An, J.; Mo, D.; Liu, H.; Veena, M.S.; Srivatsan, E.S.; Massoumi, R.; Rettig, M.B. Inactivation of the CYLD deubiquitinase by HPV E6 mediates hypoxia-induced NF-kappaB activation. *Cancer Cell* **2008**, *14*, 394–407, doi:10.1016/j.ccr.2008.10.007.
51. Hajek, M.; Sewell, A.; Kaech, S.; Burtness, B.; Yarbrough, W.G.; Issaeva, N. TRAF3/CYLD mutations identify a distinct subset of human papillomavirus-associated head and neck squamous cell carcinoma. *Cancer* **2017**, *123*, 1778–1790, doi:10.1002/cncr.30570.
52. Cui, Z.; Kang, H.; Grandis, J.R.; Johnson, D.E. CYLD Alterations in the Tumorigenesis and Progression of Human Papillomavirus-Associated Head and Neck Cancers. *Mol. Cancer Res.* **2020**, doi:10.1158/1541-7786.MCR-20-0565.
53. Zhou, Q.; Cheng, C.; Wei, Y.; Yang, J.; Zhou, W.; Song, Q.; Ke, M.; Yan, W.; Zheng, L.; Zhang, Y.; et al. USP15 potentiates NF- κ B activation by differentially stabilizing TAB2 and TAB3. *FEBS J.* **2020**, *287*, 3165–3183, doi:10.1111/febs.15202.
54. Jiang, N.; Dai, Q.; Su, X.; Fu, J.; Feng, X.; Peng, J. Role of PI3K/AKT pathway in cancer: the framework of malignant behavior. *Mol. Biol. Rep.* **2020**, *47*, 4587–4629, doi:10.1007/s11033-020-05435-1.
55. Zhang, D.; Jiang, F.; Wang, X.; Li, G. Downregulation of Ubiquitin-Specific Protease 22 Inhibits Proliferation, Invasion, and Epithelial-Mesenchymal Transition in Osteosarcoma Cells. *Oncol. Res.* **2017**, *25*, 743–751, doi:10.3727/096504016X14772395226335.
56. Zhang, J.; Luo, N.; Tian, Y.; Li, J.; Yang, X.; Yin, H.; Xiao, C.; Sheng, J.; Li, Y.; Tang, B.; et al. USP22 knockdown enhanced chemosensitivity of hepatocellular carcinoma cells to 5-Fu by up-regulation of Smad4 and suppression of Akt. *Oncotarget* **2017**, *8*, 24728–24740, doi:10.18632/oncotarget.15798.
57. Pozhidaeva, A.; Bezsonova, I. USP7: Structure, substrate specificity, and inhibition. *DNA Repair (Amst.)* **2019**, *76*, 30–39, doi:10.1016/j.dnarep.2019.02.005.
58. Hu, M.; Li, P.; Li, M.; Li, W.; Yao, T.; Wu, J.-W.; Gu, W.; Cohen, R.E.; Shi, Y. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in

- complex with ubiquitin aldehyde. *Cell* **2002**, *111*, 1041–1054, doi:10.1016/s0092-8674(02)01199-6.
59. Ritorto, M.S.; Ewan, R.; Perez-Oliva, A.B.; Knebel, A.; Buhrlage, S.J.; Wightman, M.; Kelly, S.M.; Wood, N.T.; Virdee, S.; Gray, N.S.; et al. Screening of DUB activity and specificity by MALDI-TOF mass spectrometry. *Nat. Commun.* **2014**, *5*, 4763, doi:10.1038/ncomms5763.
60. Sato, Y.; Goto, E.; Shibata, Y.; Kubota, Y.; Yamagata, A.; Goto-Ito, S.; Kubota, K.; Inoue, J.; Takekawa, M.; Tokunaga, F.; et al. Structures of CYLD USP with Met1- or Lys63-linked diubiquitin reveal mechanisms for dual specificity. *Nat. Struct. Mol. Biol.* **2015**, *22*, 222–229, doi:10.1038/nsmb.2970.
61. Gersch, M.; Gladkova, C.; Schubert, A.F.; Michel, M.A.; Maslen, S.; Komander, D. Mechanism and regulation of the Lys6-selective deubiquitinase USP30. *Nat. Struct. Mol. Biol.* **2017**, *24*, 920–930, doi:10.1038/nsmb.3475.
62. Schulz, S.; Chachami, G.; Kozackiewicz, L.; Winter, U.; Stankovic-Valentin, N.; Haas, P.; Hofmann, K.; Urlaub, H.; Ovaa, H.; Wittbrodt, J.; et al. Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. *EMBO Rep.* **2012**, *13*, 930–938, doi:10.1038/embor.2012.125.
63. Malakhov, M.P.; Malakhova, O.A.; Kim, K. II; Ritchie, K.J.; Zhang, D.-E. UBP43 (USP18) specifically removes ISG15 from conjugated proteins. *J. Biol. Chem.* **2002**, *277*, 9976–9981, doi:10.1074/jbc.M109078200.
64. Bhattacharya, S.; Chakraborty, D.; Basu, M.; Ghosh, M.K. Emerging insights into HAUSP (USP7) in physiology, cancer and other diseases. *Signal Transduct. Target. Ther.* **2018**, *3*, 17, doi:10.1038/s41392-018-0012-y.
65. Rougé, L.; Bainbridge, T.W.; Kwok, M.; Tong, R.; Di Lello, P.; Wertz, I.E.; Maurer, T.; Ernst, J.A.; Murray, J. Molecular Understanding of USP7 Substrate Recognition and C-Terminal Activation. *Structure* **2016**, *24*, 1335–1345, doi:10.1016/j.str.2016.05.020.
66. Wolberger, C. Mechanisms for regulating deubiquitinating enzymes. *Protein Sci.* **2014**, *23*, 344–353, doi:10.1002/pro.2415.
67. Faesen, A.C.; Luna-Vargas, M.P.A.; Geurink, P.P.; Clerici, M.; Merckx, R.; van Dijk, W.J.; Hameed, D.S.; El Oualid, F.; Ovaa, H.; Sixma, T.K. The differential modulation of USP activity by internal regulatory domains, interactors and eight ubiquitin chain types. *Chem. Biol.* **2011**, *18*, 1550–1561, doi:10.1016/j.chembiol.2011.10.017.
68. Faesen, A.C.; Luna-Vargas, M.P.A.; Sixma, T.K. The role of UBL domains in ubiquitin-specific proteases. *Biochem. Soc. Trans.* **2012**, *40*, 539–545, doi:10.1042/BST20120004.
69. Kim, R.Q.; van Dijk, W.J.; Sixma, T.K. Structure of USP7 catalytic domain and three Ubl-domains reveals a connector α -helix with regulatory role. *J. Struct. Biol.* **2016**, *195*, 11–18, doi:10.1016/j.jsb.2016.05.005.

70. Basters, A.; Geurink, P.P.; Röcker, A.; Witting, K.F.; Tadayon, R.; Hess, S.; Semrau, M.S.; Storicci, P.; Ovaa, H.; Knobloch, K.-P.; et al. Structural basis of the specificity of USP18 toward ISG15. *Nat. Struct. Mol. Biol.* **2017**, *24*, 270–278, doi:10.1038/nsmb.3371.
71. Hutten, S.; Chachami, G.; Winter, U.; Melchior, F.; Lamond, A.I. A role for the Cajal-body-associated SUMO isopeptidase USPL1 in snRNA transcription mediated by RNA polymerase II. *J. Cell Sci.* **2014**, *127*, 1065–1078, doi:10.1242/jcs.141788.
72. Hebert, M.D.; Poole, A.R. Towards an understanding of regulating Cajal body activity by protein modification. *RNA Biol.* **2017**, *14*, 761–778, doi:10.1080/15476286.2016.1243649.
73. Bermejo, J.L.; Kabisch, M.; Dünnebier, T.; Schnaidt, S.; Melchior, F.; Fischer, H.-P.; Harth, V.; Rabstein, S.; Pesch, B.; Brüning, T.; et al. Exploring the association between genetic variation in the SUMO isopeptidase gene USPL1 and breast cancer through integration of data from the population-based GENICA study and external genetic databases. *Int. J. cancer* **2013**, *133*, 362–372, doi:10.1002/ijc.28040.
74. Gao, H.; Wang, H.; Yang, W. Identification of key genes and construction of microRNA-mRNA regulatory networks in multiple myeloma by integrated multiple GEO datasets using bioinformatics analysis. *Int. J. Hematol.* **2017**, *106*, 99–107, doi:10.1007/s12185-017-2216-2.
75. Flotho, A.; Melchior, F. Sumoylation: a regulatory protein modification in health and disease. *Annu. Rev. Biochem.* **2013**, *82*, 357–385, doi:10.1146/annurev-biochem-061909-093311.
76. van der Veen, A.G.; Ploegh, H.L. Ubiquitin-like proteins. *Annu. Rev. Biochem.* **2012**, *81*, 323–357, doi:10.1146/annurev-biochem-093010-153308.
77. Johnson, E.S. Protein modification by SUMO. *Annu. Rev. Biochem.* **2004**, *73*, 355–382, doi:10.1146/annurev.biochem.73.011303.074118.
78. Varejão, N.; Lascorz, J.; Li, Y.; Reverter, D. Molecular mechanisms in SUMO conjugation. *Biochem. Soc. Trans.* **2020**, *48*, 123–135, doi:10.1042/BST20190357.
79. Jentsch, S.; Psakhye, I. Control of nuclear activities by substrate-selective and protein-group SUMOylation. *Annu. Rev. Genet.* **2013**, *47*, 167–186, doi:10.1146/annurev-genet-111212-133453.
80. García-Rodríguez, N.; Wong, R.P.; Ulrich, H.D. Functions of Ubiquitin and SUMO in DNA Replication and Replication Stress. *Front. Genet.* **2016**, *7*, 87, doi:10.3389/fgene.2016.00087.
81. Seeler, J.-S.; Dejean, A. SUMO and the robustness of cancer. *Nat. Rev. Cancer* **2017**, *17*, 184–197, doi:10.1038/nrc.2016.143.
82. Cappadocia, L.; Lima, C.D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem. Rev.* **2018**, *118*, 889–918,

- doi:10.1021/acs.chemrev.6b00737.
83. Li, S.J.; Hochstrasser, M. A new protease required for cell-cycle progression in yeast. *Nature* **1999**, *398*, 246–251, doi:10.1038/18457.
 84. Reverter, D.; Lima, C.D. A basis for SUMO protease specificity provided by analysis of human Senp2 and a Senp2-SUMO complex. *Structure* **2004**, *12*, 1519–1531, doi:10.1016/j.str.2004.05.023.
 85. Walden, H.; Podgorski, M.S.; Schulman, B.A. Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. *Nature* **2003**, *422*, 330–334, doi:10.1038/nature01456.
 86. Lois, L.M.; Lima, C.D. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *EMBO J.* **2005**, *24*, 439–451, doi:10.1038/sj.emboj.7600552.
 87. Schulman, B.A.; Harper, J.W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 319–331, doi:10.1038/nrm2673.
 88. Johnson, E.S.; Blobel, G. Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *J. Biol. Chem.* **1997**, *272*, 26799–26802, doi:10.1074/jbc.272.43.26799.
 89. Olsen, S.K.; Lima, C.D. Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. *Mol. Cell* **2013**, *49*, 884–896, doi:10.1016/j.molcel.2013.01.013.
 90. Wang, J.; Hu, W.; Cai, S.; Lee, B.; Song, J.; Chen, Y. The intrinsic affinity between E2 and the Cys domain of E1 in ubiquitin-like modifications. *Mol. Cell* **2007**, *27*, 228–237, doi:10.1016/j.molcel.2007.05.023.
 91. Wang, J.; Taherbhoy, A.M.; Hunt, H.W.; Seyedin, S.N.; Miller, D.W.; Miller, D.J.; Huang, D.T.; Schulman, B.A. Crystal structure of UBA2(ufd)-Ubc9: insights into E1-E2 interactions in Sumo pathways. *PLoS One* **2010**, *5*, e15805, doi:10.1371/journal.pone.0015805.
 92. Lv, Z.; Rickman, K.A.; Yuan, L.; Williams, K.; Selvam, S.P.; Woosley, A.N.; Howe, P.H.; Ogretmen, B.; Smogorzewska, A.; Olsen, S.K. S. pombe Uba1-Ubc15 Structure Reveals a Novel Regulatory Mechanism of Ubiquitin E2 Activity. *Mol. Cell* **2017**, *65*, 699–714.e6, doi:10.1016/j.molcel.2017.01.008.
 93. Hann, Z.S.; Ji, C.; Olsen, S.K.; Lu, X.; Lux, M.C.; Tan, D.S.; Lima, C.D. Structural basis for adenylation and thioester bond formation in the ubiquitin E1. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 15475–15484, doi:10.1073/pnas.1905488116.
 94. Pichler, A.; Fatouros, C.; Lee, H.; Eisenhardt, N. SUMO conjugation - a mechanistic view. *Biomol. Concepts* **2017**, *8*, 13–36, doi:10.1515/bmc-2016-0030.
 95. Streich, F.C.J.; Lima, C.D. Structural and functional insights to ubiquitin-like protein conjugation. *Annu. Rev. Biophys.* **2014**, *43*, 357–379,

- doi:10.1146/annurev-biophys-051013-022958.
96. Komander, D.; Clague, M.J.; Urbé, S. Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 550–563, doi:10.1038/nrm2731.
 97. Hickey, C.M.; Wilson, N.R.; Hochstrasser, M. Function and regulation of SUMO proteases. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 755–766, doi:10.1038/nrm3478.
 98. Yeh, E.T.H. SUMOylation and De-SUMOylation: wrestling with life's processes. *J. Biol. Chem.* **2009**, *284*, 8223–8227, doi:10.1074/jbc.R800050200.
 99. Nayak, A.; Müller, S. SUMO-specific proteases/isopeptidases: SENPs and beyond. *Genome Biol.* **2014**, *15*, 422, doi:10.1186/s13059-014-0422-2.
 100. Mukhopadhyay, D.; Dasso, M. Modification in reverse: the SUMO proteases. *Trends Biochem. Sci.* **2007**, *32*, 286–295, doi:10.1016/j.tibs.2007.05.002.
 101. Kumar, A.; Zhang, K.Y.J. Advances in the development of SUMO specific protease (SENP) inhibitors. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 204–211, doi:10.1016/j.csbj.2015.03.001.
 102. Shin, E.J.; Shin, H.M.; Nam, E.; Kim, W.S.; Kim, J.-H.; Oh, B.-H.; Yun, Y. DeSUMOylating isopeptidase: a second class of SUMO protease. *EMBO Rep.* **2012**, *13*, 339–346, doi:10.1038/embor.2012.3.
 103. Gan-Erdene, T.; Nagamalleswari, K.; Yin, L.; Wu, K.; Pan, Z.-Q.; Wilkinson, K.D. Identification and characterization of DEN1, a deneddylase of the ULP family. *J. Biol. Chem.* **2003**, *278*, 28892–28900, doi:10.1074/jbc.M302890200.
 104. Reverter, D.; Wu, K.; Erdene, T.G.; Pan, Z.-Q.; Wilkinson, K.D.; Lima, C.D. Structure of a complex between Nedd8 and the Ulp/Senp protease family member Den1. *J. Mol. Biol.* **2005**, *345*, 141–151, doi:10.1016/j.jmb.2004.10.022.
 105. Mendoza, H.M.; Shen, L.-N.; Botting, C.; Lewis, A.; Chen, J.; Ink, B.; Hay, R.T. NEDP1, a highly conserved cysteine protease that deNEDDylates Cullins. *J. Biol. Chem.* **2003**, *278*, 25637–25643, doi:10.1074/jbc.M212948200.
 106. Wu, K.; Yamoah, K.; Dolios, G.; Gan-Erdene, T.; Tan, P.; Chen, A.; Lee, C.-G.; Wei, N.; Wilkinson, K.D.; Wang, R.; et al. DEN1 is a dual function protease capable of processing the C terminus of Nedd8 and deconjugating hyper-neddylated CUL1. *J. Biol. Chem.* **2003**, *278*, 28882–28891, doi:10.1074/jbc.M302888200.
 107. Takahashi, Y.; Mizoi, J.; Toh-E, A.; Kikuchi, Y. Yeast Ulp1, an Smt3-specific protease, associates with nucleoporins. *J. Biochem.* **2000**, *128*, 723–725, doi:10.1093/oxfordjournals.jbchem.a022807.
 108. Gong, L.; Millas, S.; Maul, G.G.; Yeh, E.T. Differential regulation of sentrinized proteins by a novel sentrin-specific protease. *J. Biol. Chem.* **2000**, *275*, 3355–3359, doi:10.1074/jbc.275.5.3355.

REFERENCE

109. Hang, J.; Dasso, M. Association of the human SUMO-1 protease SENP2 with the nuclear pore. *J. Biol. Chem.* **2002**, *277*, 19961–19966, doi:10.1074/jbc.M201799200.
110. Owerbach, D.; McKay, E.M.; Yeh, E.T.H.; Gabbay, K.H.; Bohren, K.M. A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochem. Biophys. Res. Commun.* **2005**, *337*, 517–520, doi:10.1016/j.bbrc.2005.09.090.
111. Gong, L.; Yeh, E.T.H. Characterization of a family of nucleolar SUMO-specific proteases with preference for SUMO-2 or SUMO-3. *J. Biol. Chem.* **2006**, *281*, 15869–15877, doi:10.1074/jbc.M511658200.
112. Di Bacco, A.; Ouyang, J.; Lee, H.-Y.; Catic, A.; Ploegh, H.; Gill, G. The SUMO-specific protease SENP5 is required for cell division. *Mol. Cell. Biol.* **2006**, *26*, 4489–4498, doi:10.1128/MCB.02301-05.
113. Mikolajczyk, J.; Drag, M.; Békés, M.; Cao, J.T.; Ronai, Z.; Salvesen, G.S. Small ubiquitin-related modifier (SUMO)-specific proteases: profiling the specificities and activities of human SENPs. *J. Biol. Chem.* **2007**, *282*, 26217–26224, doi:10.1074/jbc.M702444200.
114. Lima, C.D.; Reverter, D. Structure of the human SENP7 catalytic domain and poly-SUMO deconjugation activities for SENP6 and SENP7. *J. Biol. Chem.* **2008**, *283*, 32045–32055, doi:10.1074/jbc.M805655200.
115. Alegre, K.O.; Reverter, D. Swapping small ubiquitin-like modifier (SUMO) isoform specificity of SUMO proteases SENP6 and SENP7. *J. Biol. Chem.* **2011**, *286*, 36142–36151, doi:10.1074/jbc.M111.268847.
116. Shen, L.N.; Geoffroy, M.-C.; Jaffray, E.G.; Hay, R.T. Characterization of SENP7, a SUMO-2/3-specific isopeptidase. *Biochem. J.* **2009**, *421*, 223–230, doi:10.1042/BJ20090246.
117. Tatham, M.H.; Geoffroy, M.-C.; Shen, L.; Plechanovova, A.; Hattersley, N.; Jaffray, E.G.; Palvimo, J.J.; Hay, R.T. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat. Cell Biol.* **2008**, *10*, 538–546, doi:10.1038/ncb1716.
118. Lallemand-Breitenbach, V.; Jeanne, M.; Benhenda, S.; Nasr, R.; Lei, M.; Peres, L.; Zhou, J.; Zhu, J.; Raught, B.; de Thé, H. Arsenic degrades PML or PML-RAR α through a SUMO-triggered RNF4/ubiquitin-mediated pathway. *Nat. Cell Biol.* **2008**, *10*, 547–555, doi:10.1038/ncb1717.
119. Wu, Z.; Huang, H.; Han, Q.; Hu, Z.; Teng, X.-L.; Ding, R.; Ye, Y.; Yu, X.; Zhao, R.; Wang, Z.; et al. SENP7 senses oxidative stress to sustain metabolic fitness and antitumor functions of CD8⁺ T cells. *J. Clin. Invest.* **2022**, *132*, doi:10.1172/JCI155224.
120. Hotz, P.W.; Müller, S.; Mandler, L. SUMO-specific Isopeptidases Tuning Cardiac SUMOylation in Health and Disease. *Front. Mol. Biosci.* **2021**, *8*, 786136, doi:10.3389/fmolb.2021.786136.

121. Garvin, A.J.; Densham, R.M.; Blair-Reid, S.A.; Pratt, K.M.; Stone, H.R.; Weekes, D.; Lawrence, K.J.; Morris, J.R. The deSUMOylase SENP7 promotes chromatin relaxation for homologous recombination DNA repair. *EMBO Rep.* **2013**, *14*, 975–983, doi:10.1038/embor.2013.141.
122. Maison, C.; Romeo, K.; Bailly, D.; Dubarry, M.; Quivy, J.-P.; Almouzni, G. The SUMO protease SENP7 is a critical component to ensure HP1 enrichment at pericentric heterochromatin. *Nat. Struct. Mol. Biol.* **2012**, *19*, 458–460, doi:10.1038/nsmb.2244.
123. Huang, C.-J.; Wu, D.; Jiao, X.-F.; Khan, F.A.; Xiong, C.-L.; Liu, X.-M.; Yang, J.; Yin, T.-L.; Huo, L.-J. Maternal SENP7 programs meiosis architecture and embryo survival in mouse. *Biochim. Biophys. Acta. Mol. cell Res.* **2017**, *1864*, 1195–1206, doi:10.1016/j.bbamcr.2017.03.005.
124. Juarez-Vicente, F.; Luna-Pelaez, N.; Garcia-Dominguez, M. The Sumo protease Semp7 is required for proper neuronal differentiation. *Biochim. Biophys. Acta* **2016**, *1863*, 1490–1498, doi:10.1016/j.bbamcr.2016.03.028.
125. Bialik, P.; Woźniak, K. SUMO proteases as potential targets for cancer therapy. *Postepy Hig. Med. Dosw. (Online)* **2017**, *71*, 997–1004, doi:10.5604/01.3001.0010.6667.
126. Karami, S.; Lin, F.-M.; Kumar, S.; Bahnassy, S.; Thangavel, H.; Quttina, M.; Li, Y.; Ren, J.; Bawa-Khalfe, T. Novel SUMO-Protease SENP7S Regulates β -catenin Signaling and Mammary Epithelial Cell Transformation. *Sci. Rep.* **2017**, *7*, 46477, doi:10.1038/srep46477.
127. Lewis, G.P. Legumes of the world. *Edinburgh J. Bot.* **2005**, *62*, doi:https://doi.org/10.1017/S0960428606190198.
128. Gojon, A.; Krouk, G.; Perrine-Walker, F.; Laugier, E. Nitrate transceptor(s) in plants. *J. Exp. Bot.* **2011**, *62*, 2299–2308, doi:10.1093/jxb/erq419.
129. J.I, S.; J.K, A.; E.K, J. From North to South: A latitudinal look at legume nodulation processes. *South African J. Bot.* **2013**, *89*, 31–41, doi:doi.org/10.1016/j.sajb.2013.06.011.
130. Liu, W.Y.Y.; Ridgway, H.J.; James, T.K.; James, E.K.; Chen, W.-M.; Sprent, J.I.; Young, J.P.W.; Andrews, M. Burkholderia sp. induces functional nodules on the South African invasive legume *Dipogon lignosus* (Phaseoleae) in New Zealand soils. *Microb. Ecol.* **2014**, *68*, 542–555, doi:10.1007/s00248-014-0427-0.
131. Mahmud, K.; Makaju, S.; Ibrahim, R.; Missaoui, A. Current Progress in Nitrogen Fixing Plants and Microbiome Research. *Plants (Basel, Switzerland)* **2020**, *9*, doi:10.3390/plants9010097.
132. Wang, D.; Yang, S.; Tang, F.; Zhu, H. Symbiosis specificity in the legume: rhizobial mutualism. *Cell. Microbiol.* **2012**, *14*, 334–342, doi:10.1111/j.1462-5822.2011.01736.x.
133. Goormachtig, S.; Capoen, W.; Holsters, M. Rhizobium infection: lessons from the

- versatile nodulation behaviour of water-tolerant legumes. *Trends Plant Sci.* **2004**, *9*, 518–522, doi:10.1016/j.tplants.2004.09.005.
134. Long, S.R. Rhizobium symbiosis: nod factors in perspective. *Plant Cell* **1996**, *8*, 1885–1898, doi:10.1105/tpc.8.10.1885.
 135. Spaink, H.P. Root nodulation and infection factors produced by rhizobial bacteria. *Annu. Rev. Microbiol.* **2000**, *54*, 257–288, doi:10.1146/annurev.micro.54.1.257.
 136. Costa, T.R.D.; Felisberto-Rodrigues, C.; Meir, A.; Prevost, M.S.; Redzej, A.; Trokter, M.; Waksman, G. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **2015**, *13*, 343–359, doi:10.1038/nrmicro3456.
 137. Tillotson, G.S.; Tillotson, J. Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis. *Expert Rev. Anti. Infect. Ther.* **2009**, *7*, 691–693, doi:doi.org/10.1586/eri.09.47.
 138. Ghosh, P. Process of protein transport by the type III secretion system. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 771–795, doi:10.1128/MMBR.68.4.771-795.2004.
 139. Shaulov, L.; Gershberg, J.; Deng, W.; Finlay, B.B.; Sal-Man, N. The Ruler Protein EscP of the Enteropathogenic Escherichia coli Type III Secretion System Is Involved in Calcium Sensing and Secretion Hierarchy Regulation by Interacting with the Gatekeeper Protein SepL. *MBio* **2017**, *8*, doi:10.1128/mBio.01733-16.
 140. Büttner, D.; Bonas, U. Common infection strategies of plant and animal pathogenic bacteria. *Curr. Opin. Plant Biol.* **2003**, *6*, 312–319, doi:10.1016/s1369-5266(03)00064-5.
 141. McShan, A.C.; De Guzman, R.N. The bacterial type III secretion system as a target for developing new antibiotics. *Chem. Biol. Drug Des.* **2015**, *85*, 30–42, doi:10.1111/cbdd.12422.
 142. De Lyra, M. do C.C.P.; Lopez-Baena, F.J.; Madinabeitia, N.; Vinardell, J.M.; Espuny, M. del R.; Cubo, M.T.; Belloguin, R.A.; Ruiz-Sainz, J.E.; Ollero, F.J. Inactivation of the Sinorhizobium fredii HH103 rhcJ gene abolishes nodulation outer proteins (Nops) secretion and decreases the symbiotic capacity with soybean. *Int. Microbiol.* **2006**, *9*, 125–133.
 143. Krause, A.; Doerfel, A.; Göttfert, M. Mutational and transcriptional analysis of the type III secretion system of Bradyrhizobium japonicum. *Mol. Plant. Microbe. Interact.* **2002**, *15*, 1228–1235, doi:10.1094/MPMI.2002.15.12.1228.
 144. Miwa, H.; Okazaki, S. How effectors promote beneficial interactions. *Curr. Opin. Plant Biol.* **2017**, *38*, 148–154, doi:10.1016/j.pbi.2017.05.011.
 145. Block, A.; Li, G.; Fu, Z.Q.; Alfano, J.R. Phytopathogen type III effector weaponry and their plant targets. *Curr. Opin. Plant Biol.* **2008**, *11*, 396–403, doi:10.1016/j.pbi.2008.06.007.
 146. Xiang, Q.-W.; Bai, J.; Cai, J.; Huang, Q.-Y.; Wang, Y.; Liang, Y.; Zhong, Z.;

- Wagner, C.; Xie, Z.-P.; Staehelin, C. NopD of *Bradyrhizobium* sp. XS1150 Possesses SUMO Protease Activity. *Front. Microbiol.* **2020**, *11*, 386, doi:10.3389/fmicb.2020.00386.
147. Chosed, R.; Tomchick, D.R.; Brautigam, C.A.; Mukherjee, S.; Negi, V.S.; Machius, M.; Orth, K. Structural analysis of *Xanthomonas* XopD provides insights into substrate specificity of ubiquitin-like protein proteases. *J. Biol. Chem.* **2007**, *282*, 6773–6782, doi:10.1074/jbc.M608730200.
148. Pruneda, J.N.; Durkin, C.H.; Geurink, P.P.; Ovaa, H.; Santhanam, B.; Holden, D.W.; Komander, D. The Molecular Basis for Ubiquitin and Ubiquitin-like Specificities in Bacterial Effector Proteases. *Mol. Cell* **2016**, *63*, 261–276, doi:10.1016/j.molcel.2016.06.015.
149. Li, Y.; Reverter, D. Molecular Mechanisms of DUBs Regulation in Signaling and Disease. *Int. J. Mol. Sci.* **2021**, *22*, doi:10.3390/ijms22030986.
150. Chalker, J.M.; Gunnoo, S.B.; Boutureira, O.; Gerstberger, S.C.; Fernández-González, M.; Bernardes, G.J.L.; Griffin, L.; Hailu, H.; Schofield, C.J.; Davis, B.G. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2011**, *2*, 1666–1676, doi:10.1039/C1SC00185J.
151. Mossessova, E.; Lima, C.D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **2000**, *5*, 865–876, doi:10.1016/s1097-2765(00)80326-3.
152. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, doi:10.1038/s41586-021-03819-2.
153. Ye, Y.; Akutsu, M.; Reyes-Turcu, F.; Enchev, R.I.; Wilkinson, K.D.; Komander, D. Polyubiquitin binding and cross-reactivity in the USP domain deubiquitinase USP21. *EMBO Rep.* **2011**, *12*, 350–357, doi:10.1038/embor.2011.17.
154. Ward, S.J.; Gratton, H.E.; Indrayudha, P.; Michavila, C.; Mukhopadhyay, R.; Maurer, S.K.; Caulton, S.G.; Emsley, J.; Dreveny, I. The structure of the deubiquitinase USP15 reveals a misaligned catalytic triad and an open ubiquitin-binding channel. *J. Biol. Chem.* **2018**, *293*, 17362–17374, doi:10.1074/jbc.RA118.003857.
155. Gersch, M.; Wagstaff, J.L.; Toms, A. V; Graves, B.; Freund, S.M. V; Komander, D. Distinct USP25 and USP28 Oligomerization States Regulate Deubiquitinating Activity. *Mol. Cell* **2019**, *74*, 436-451.e7, doi:10.1016/j.molcel.2019.02.030.
156. Pfoh, R.; Lacdao, I.K.; Georges, A.A.; Capar, A.; Zheng, H.; Frappier, L.; Saridakis, V. Crystal Structure of USP7 Ubiquitin-like Domains with an ICP0 Peptide Reveals a Novel Mechanism Used by Viral and Cellular Proteins to Target USP7. *PLoS Pathog.* **2015**, *11*, e1004950, doi:10.1371/journal.ppat.1004950.
157. Schrödinger, L. and W.D. PyMOL. Retrieved from <http://www.pymol.org/pymol> 2020.

158. Alegre, K.O.; Reverter, D. Structural insights into the SENP6 Loop1 structure in complex with SUMO2. *Protein Sci.* **2014**, *23*, 433–441, doi:10.1002/pro.2425.
159. van den Ent, F.; Löwe, J. RF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods* **2006**, *67*, 67–74, doi:10.1016/j.jbbm.2005.12.008.
160. Monaco, S.; Gordon, E.; Bowler, M.W.; Delagenière, S.; Guijarro, M.; Spruce, D.; Svensson, O.; McSweeney, S.M.; McCarthy, A.A.; Leonard, G.; et al. Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. *J. Appl. Crystallogr.* **2013**, *46*, 804–810, doi:10.1107/S0021889813006195.
161. Gabadinho, J.; Beteva, A.; Guijarro, M.; Rey-Bakaikoa, V.; Spruce, D.; Bowler, M.W.; Brockhauser, S.; Flot, D.; Gordon, E.J.; Hall, D.R.; et al. MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *J. Synchrotron Radiat.* **2010**, *17*, 700–707, doi:10.1107/S0909049510020005.
162. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 486–501, doi:10.1107/S0907444910007493.
163. Adams, P.D.; Afonine, P. V.; Bunkóczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.-W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 213–221, doi:10.1107/S0907444909052925.
164. Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, W320–4, doi:10.1093/nar/gku316.
165. Gersch, M.; Wagstaff, J.L.; Toms, A. V.; Graves, B.; Freund, S.M.V.; Komander, D. Distinct USP25 and USP28 Oligomerization States Regulate Deubiquitinating Activity. *Mol. Cell* **2019**, *74*, 436–451.e7, doi:10.1016/j.molcel.2019.02.030.
166. Renatus, M.; Parrado, S.G.; D’Arcy, A.; Eidhoff, U.; Gerhartz, B.; Hassiepen, U.; Pierrat, B.; Riedl, R.; Vinzenz, D.; Worpenberg, S.; et al. Structural basis of ubiquitin recognition by the deubiquitinating protease USP2. *Structure* **2006**, *14*, 1293–1302, doi:10.1016/j.str.2006.06.012.
167. Leznicki, P.; Natarajan, J.; Bader, G.; Spevak, W.; Schlattl, A.; Abdul Rehman, S.A.; Pathak, D.; Weidlich, S.; Zoephel, A.; Bordone, M.C.; et al. Expansion of DUB functionality generated by alternative isoforms - USP35, a case study. *J. Cell Sci.* **2018**, *131*, doi:10.1242/jcs.212753.
168. Dharadhar, S.; Clerici, M.; van Dijk, W.J.; Fish, A.; Sixma, T.K. A conserved two-step binding for the UAF1 regulator to the USP12 deubiquitinating enzyme. *J. Struct. Biol.* **2016**, *196*, 437–447, doi:10.1016/j.jsb.2016.09.011.
169. Kunz, K.; Piller, T.; Müller, S. SUMO-specific proteases and isopeptidases of the SENP family at a glance. *J. Cell Sci.* **2018**, *131*, doi:10.1242/jcs.211904.

170. Li, Y.; Varejão, N.; Reverter, D. Structural basis for the SUMO protease activity of the atypical ubiquitin-specific protease USPL1. *Nat. Commun.* **2022**, *13*, 1819, doi:10.1038/s41467-022-29485-0.
171. Yun, C.; Wang, Y.; Mukhopadhyay, D.; Backlund, P.; Kolli, N.; Yergey, A.; Wilkinson, K.D.; Dasso, M. Nucleolar protein B23/nucleophosmin regulates the vertebrate SUMO pathway through SENP3 and SENP5 proteases. *J. Cell Biol.* **2008**, *183*, 589–595, doi:10.1083/jcb.200807185.
172. Goeres, J.; Chan, P.-K.; Mukhopadhyay, D.; Zhang, H.; Raught, B.; Matunis, M.J. The SUMO-specific isopeptidase SENP2 associates dynamically with nuclear pore complexes through interactions with karyopherins and the Nup107-160 nucleoporin subcomplex. *Mol. Biol. Cell* **2011**, *22*, 4868–4882, doi:10.1091/mbc.E10-12-0953.
173. Shen, L.; Tatham, M.H.; Dong, C.; Zagórska, A.; Naismith, J.H.; Hay, R.T. SUMO protease SENP1 induces isomerization of the scissile peptide bond. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1069–1077, doi:10.1038/nsmb1172.
174. Romeo, K.; Louault, Y.; Cantaloube, S.; Loiodice, I.; Almouzni, G.; Quivy, J.-P. The SENP7 SUMO-Protease Presents a Module of Two HP1 Interaction Motifs that Locks HP1 Protein at Pericentric Heterochromatin. *Cell Rep.* **2015**, *10*, 771–782, doi:10.1016/j.celrep.2015.01.004.
175. Bannister, A.J.; Zegerman, P.; Partridge, J.F.; Miska, E.A.; Thomas, J.O.; Allshire, R.C.; Kouzarides, T. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **2001**, *410*, 120–124, doi:10.1038/35065138.
176. Cheutin, T.; McNairn, A.J.; Jenuwein, T.; Gilbert, D.M.; Singh, P.B.; Misteli, T. Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **2003**, *299*, 721–725, doi:10.1126/science.1078572.
177. Peng, J.C.; Karpen, G.H. Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genet.* **2009**, *5*, e1000435, doi:10.1371/journal.pgen.1000435.
178. Ekkebus, R.; van Kasteren, S.I.; Kulathu, Y.; Scholten, A.; Berlin, I.; Geurink, P.P.; de Jong, A.; Goerdal, S.; Neefjes, J.; Heck, A.J.R.; et al. On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. *J. Am. Chem. Soc.* **2013**, *135*, 2867–2870, doi:10.1021/ja309802n.
179. Hemelaar, J.; Galardy, P.J.; Borodovsky, A.; Kessler, B.M.; Ploegh, H.L.; Ovaa, H. Chemistry-based functional proteomics: mechanism-based activity-profiling tools for ubiquitin and ubiquitin-like specific proteases. *J. Proteome Res.* **2004**, *3*, 268–276, doi:10.1021/pr0341080.
180. Sommer, S.; Weikart, N.D.; Linne, U.; Mootz, H.D. Covalent inhibition of SUMO and ubiquitin-specific cysteine proteases by an in situ thiol-alkyne addition. *Bioorg. Med. Chem.* **2013**, *21*, 2511–2517, doi:10.1016/j.bmc.2013.02.039.
181. Shen, L.N.; Dong, C.; Liu, H.; Naismith, J.H.; Hay, R.T. The structure of SENP1-SUMO-2 complex suggests a structural basis for discrimination between SUMO

- paralogues during processing. *Biochem. J.* **2006**, *397*, 279–288, doi:10.1042/BJ20052030.
182. Juanhuix, J.; Gil-Ortiz, F.; Cuní, G.; Colldelram, C.; Nicolás, J.; Lidón, J.; Boter, E.; Ruget, C.; Ferrer, S.; Benach, J. Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrotron Radiat.* **2014**, *21*, 679–689, doi:10.1107/S160057751400825X.
183. Ikeda, F.; Dikic, I. Atypical ubiquitin chains: new molecular signals. “Protein Modifications: Beyond the Usual Suspects” review series. *EMBO Rep.* **2008**, *9*, 536–542, doi:10.1038/embor.2008.93.
184. Vierstra, R.D. The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 385–397, doi:10.1038/nrm2688.
185. Ronau, J.A.; Beckmann, J.F.; Hochstrasser, M. Substrate specificity of the ubiquitin and Ubl proteases. *Cell Res.* **2016**, *26*, 441–456, doi:10.1038/cr.2016.38.
186. Staehelin, C.; Krishnan, H.B. Nodulation outer proteins: double-edged swords of symbiotic rhizobia. *Biochem. J.* **2015**, *470*, 263–274, doi:10.1042/BJ20150518.
187. Hotson, A.; Chosed, R.; Shu, H.; Orth, K.; Mudgett, M.B. Xanthomonas type III effector XopD targets SUMO-conjugated proteins in planta. *Mol. Microbiol.* **2003**, *50*, 377–389, doi:10.1046/j.1365-2958.2003.03730.x.
188. Tan, C.M.; Li, M.-Y.; Yang, P.-Y.; Chang, S.H.; Ho, Y.-P.; Lin, H.; Deng, W.-L.; Yang, J.-Y. Arabidopsis HFR1 is a potential nuclear substrate regulated by the Xanthomonas type III effector XopD(Xcc8004). *PLoS One* **2015**, *10*, e0117067, doi:10.1371/journal.pone.0117067.
189. Lu, Y.-T.; Li, M.-Y.; Cheng, K.-T.; Tan, C.M.; Su, L.-W.; Lin, W.-Y.; Shih, H.-T.; Chiou, T.-J.; Yang, J.-Y. Transgenic plants that express the phytoplasma effector SAP11 show altered phosphate starvation and defense responses. *Plant Physiol.* **2014**, *164*, 1456–1469, doi:10.1104/pp.113.229740.
190. Rodrigues, J.A.; López-Baena, F.J.; Ollero, F.J.; Vinardell, J.M.; Espuny, M.D.R.; Bellogín, R.A.; Ruiz-Sainz, J.E.; Thomas, J.R.; Sumpton, D.; Ault, J.; et al. NopM and NopD are rhizobial nodulation outer proteins: identification using LC-MALDI and LC-ESI with a monolithic capillary column. *J. Proteome Res.* **2007**, *6*, 1029–1037, doi:10.1021/pr060519f.
191. Shen, L.; Liu, H.; Dong, C.; Xirodimas, D.; Naismith, J.H.; Hay, R.T. Structural basis of NEDD8 ubiquitin discrimination by the deNEDDylating enzyme NEDP1. *EMBO J.* **2005**, *24*, 1341–1351, doi:10.1038/sj.emboj.7600628.
192. Wagner, K.; Kunz, K.; Piller, T.; Tascher, G.; Hölper, S.; Stehmeier, P.; Keiten-Schmitz, J.; Schick, M.; Keller, U.; Müller, S. The SUMO Isopeptidase SENP6 Functions as a Rheostat of Chromatin Residency in Genome Maintenance and Chromosome Dynamics. *Cell Rep.* **2019**, *29*, 480–494.e5, doi:10.1016/j.celrep.2019.08.106.

193. de Albuquerque, C.P.; Suhandynata, R.T.; Carlson, C.R.; Yuan, W.-T.; Zhou, H. Binding to small ubiquitin-like modifier and the nucleolar protein Csm1 regulates substrate specificity of the Ulp2 protease. *J. Biol. Chem.* **2018**, *293*, 12105–12119, doi:10.1074/jbc.RA118.003022.
194. Liebelt, F.; Jansen, N.S.; Kumar, S.; Gracheva, E.; Claessens, L.A.; Verlaan-de Vries, M.; Willemstein, E.; Vertegaal, A.C.O. The poly-SUMO2/3 protease SENP6 enables assembly of the constitutive centromere-associated network by group deSUMOylation. *Nat. Commun.* **2019**, *10*, 3987, doi:10.1038/s41467-019-11773-x.
195. Hattersley, N.; Shen, L.; Jaffray, E.G.; Hay, R.T. The SUMO protease SENP6 is a direct regulator of PML nuclear bodies. *Mol. Biol. Cell* **2011**, *22*, 78–90, doi:10.1091/mbc.E10-06-0504.
196. Mukhopadhyay, D.; Dasso, M. The fate of metaphase kinetochores is weighed in the balance of SUMOylation during S phase. *Cell Cycle* **2010**, *9*, 3194–3201, doi:10.4161/cc.9.16.12619.
197. Mittal, R.; Peak-Chew, S.Y.; Sade, R.S.; Vallis, Y.; McMahon, H.T. The acetyltransferase activity of the bacterial toxin YopJ of *Yersinia* is activated by eukaryotic host cell inositol hexakisphosphate. *J. Biol. Chem.* **2010**, *285*, 19927–19934, doi:10.1074/jbc.M110.126581.
198. Avvakumov, G. V; Walker, J.R.; Xue, S.; Finerty, P.J.J.; Mackenzie, F.; Newman, E.M.; Dhe-Paganon, S. Amino-terminal dimerization, NRDP1-rhodanese interaction, and inhibited catalytic domain conformation of the ubiquitin-specific protease 8 (USP8). *J. Biol. Chem.* **2006**, *281*, 38061–38070, doi:10.1074/jbc.M606704200.
199. Messick, T.E.; Russell, N.S.; Iwata, A.J.; Sarachan, K.L.; Shiekhhattar, R.; Shanks, J.R.; Reyes-Turcu, F.E.; Wilkinson, K.D.; Marmorstein, R. Structural basis for ubiquitin recognition by the Otu1 ovarian tumor domain protein. *J. Biol. Chem.* **2008**, *283*, 11038–11049, doi:10.1074/jbc.M704398200.