Universitat Autònoma de Barcelona

Structure insights into the autoinhibitory mechanism of the deubiquitinating enzyme USP25 and into the SUMO E1-E2 protein-protein recognition

Bing Liu

Universitat Autònoma de Barcelona

Structure insights into the autoinhibitory mechanism of the deubiquitinating enzyme USP25 and into the SUMO E1-E2 protein-protein recognition

Bing Liu

Supervisor: Dr. David Reverter

A thesis submitted for the degree of

Biochemistry and Molecular Biology

Universitat Autònoma de Barcelona

Institut de Biotecnologia i de Biomedicina (IBB)

September 2018

The pursuit of science needs special courage

Acknowledgements

First of all, I would like to thank my mother and sisters for their love and support at every moment, despite a distance of ten thousand kilometers away. To my mother, I know you have been worried about my lives in Barcelona. I apologize that I cannot take care of you these years, albeit you never complain about that. To my two sisters, I thank you for the kindly help on my lives. I also would like to thank my girlfriend for your support when I was feeling frustrated and helpless and for the wonderful time we had together.

To my supervisor, David Reverter, I would like to sincerely thank you for your best training on my scientific research. Your knowledge and ideas of science always inspire me, and your patient guide let me have a glimpse into the beauty of protein structure at the first time.

I would like to thank my lab mates, Natalia, Jara, Pablo, Zhen, as well as members from our neighbor group, for your kindly help and discussion on my research. I am also thankful to members of the other groups in IBB, for the help on the lab work.

I would like to thank Maria Lois from CRAG-UAB, Marta Sureda-Gómez and Virginia Amador from IDIBAPS, for their direct contributions on my research publications.

Finally, thanks to Chinese Scholarship Council (CSC) grant for supporting my PhD study.

Bing Liu

September 2018

Index

Introduction

The ubiquitin system and deubiquitinating enzyme USP25

Post-translational modifications increase protein complexity

While it is estimated that human genome encodes 20,000-25,000 protein-coding genes (1), the total number of proteins in human proteome is estimated to be many times more, possibly ranging from 10,000 to almost 1 million. This is because a single gene can generate different variants of a particular protein through, for example, alternative splicing, mRNA editing, and genomic recombination. The complexity of human proteome is further increased by protein post-translational modifications (PTMs), including phosphorylation, glycosylation, lipidation, carbonylation, acetylation, methylation, and ubiquitination. These mechanisms can somehow explain why humans could be so much more complex with just 25,000 genes.

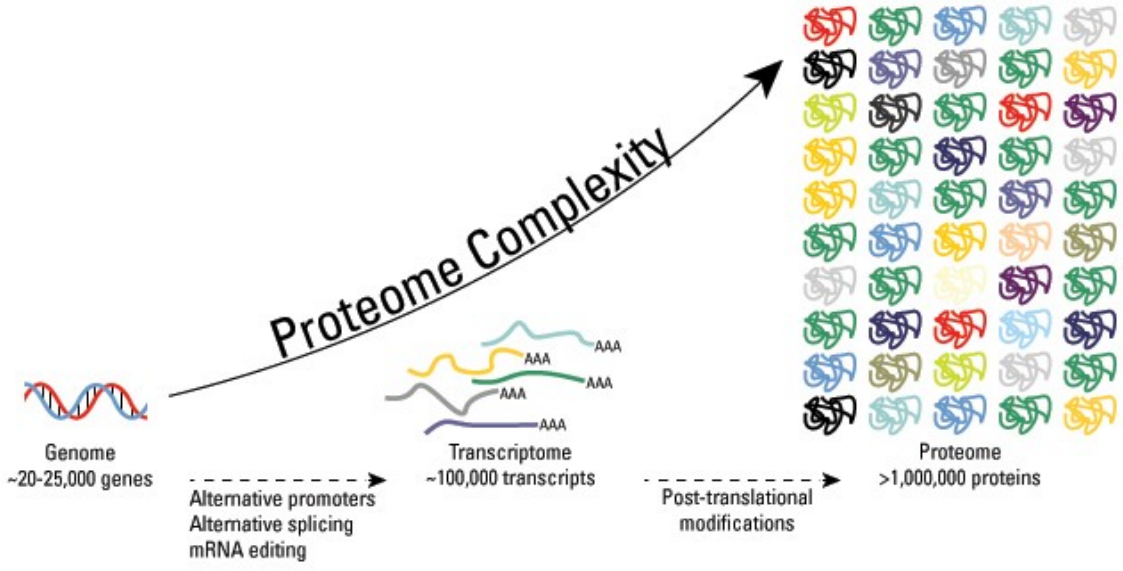


Figure 1. A view of post-translational modifications from protein complexity (2).

Ubiquitin and ubiquitination

In the late 1970s and 1980s, Avram Hershko, Aaron Ciechanover and Irwin Rose discovered and characterized the ATP-dependent, ubiquitin-mediated proteolysis. As a result of their pioneering studies, they were awarded with the 2004 Nobel Prize in Chemistry (3). After the landmark discovery of ubiquitin-mediated degradation by the 26S proteasome, the protein-based modifications became

prevalent. Today, it is quite clear that this function of ubiquitin was just the tip of an enormous iceberg. The recent global proteomic studies provide insight into the complexity of the ubiquitin system and about 1.3% of total cellular proteins are substrates of ubiquitination in human cells (4). In human societies, people communicate with each other using words and sentences that can trigger specific responses from other individuals. Similarly, proteins are modified with polymeric ubiquitin chains, in which the linkage between ubiquitin molecules encodes information. Thus, ubiquitin code is like the communication language between and within cells and determinates the fate of its substrate. The ubiquitin system is a universal means of protein regulation which controls a wide range of cellular processes including gene transcription, cell cycle, cell death, signal transduction, DNA repair, and autophagy (5-10).

Worth of the name, ubiquitin is quite ubiquitous in biology. Ubiquitin (Ub) is a highly conserved 76-residue protein (~8.5 kDa) that is present in all eukaryotes, from yeast to human. In human genome, there are four genes coding for ubiquitin: UBB, UBC, UBA52, and RPS27A (11). Ubiquitin-like proteins (Ubls) are a set of proteins which adopt the same β-grasp fold as ubiquitin. As one of the most known protein posttranslational modifiers, ubiquitin and Ubls are covalently attached to an ε-amino group of the substrate protein's lysine residues (Figure 2). This modification is carried out via a conserved three-step enzymatic cascade through the E1 activating enzyme, the E2 conjugation enzyme, and E3 protein ligase, resulting the formation of an isopeptide bond between the C terminus of ubiquitin and the lysine on the target protein (5, 12). First, E1 activates the C-terminal glycine residue of ubiquitin in an ATP-dependent manner, resulting in the formation of an intermediate ubiquitin adenylate, followed by the binding of the C-terminus of ubiquitin to the active cysteine residue of E1 through a thioester linkage. Second, the activated ubiquitin is transferred from E1 to E2, whose active cysteine residue forms a thioester bond with the C-terminus of ubiquitin. Finally, E3 ligase interacts with ubiquitin-loaded E2 and mediates the final ubiquitin transfer, forming an isopeptide bond between the lysine ε-amino group of the target protein and the C-terminal carboxyl group of ubiquitin. In addition, ubiquitin can also be attached on N-terminus amino group, cysteine, serine, and threonine residues of the target proteins (13-17).

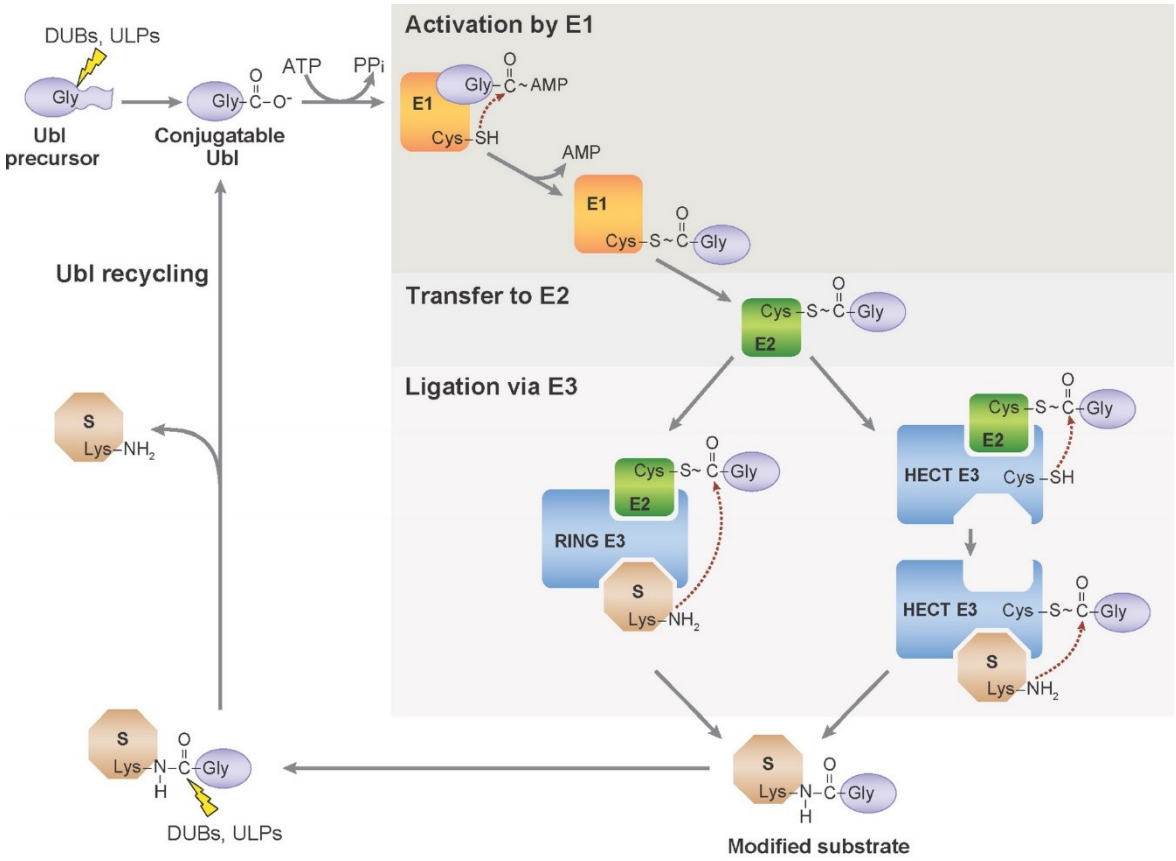


Figure 2. Ubl conjugation and deconjugation pathways (10). Precursor Ubls (including ubiquitin) are processed by either DUBs (deubiquitinating enzymes) or ULPs (Ubl-specific proteases) to expose a C-terminal glycine in the mature Ubl. The processed Ubl is conjugated to the target protein via a three-step enzymatic cascade through the E1 activating enzyme, E2 conjugation enzyme, and E3 protein ligase. The DUBs and ULPs can remove Ubls from substrates.

Ubiquitination enzymes

Specificity of ubiquitination to the thousands of human substrates depends on the sequential action of E1s, E2s, and E3s. There are 2 E1s, ~40 E2s, and more than 600 E3s in human ubiquitin system (18-23). E2s play the dominant role in the ubiquitin chain assembly. E3s are the most heterogeneous and critical components in the ubiquitination cascade, as they bind directly to their target proteins and strictly control the efficiency and specificity of the ubiquitination reaction (21). Currently, E3s can be classified in three different classes: HECT (homologous to the E6AP carboxyl terminus), RING (Really Interesting New Gene), and RBR (RING-between-RING), based on their structural composition and mechanism of ubiquitin transfer to the substrate (18, 20, 23-25) (Figure 2). RING E3s, the most abundant ubiquitin ligases, are characterized by the presence of a zinc-binding domain named RING or by a U-box domain, which has the same fold as RING but without zinc

coordination (20, 25-27). The conserved HECT domain is located at the C-terminus of the proteins and characterized as a bi-lobar architecture consisting of the N-terminal lobe that interacts with the E2 and a C-terminal lobe that contains the active cysteine which binds ubiquitin in a thioester linkage (24, 25). The HECT members catalyze ubiquitin transfer to the substrate in a two-step reaction: ubiquitin is first transferred to the catalytic cysteine of E3 and then transferred from E3 to the substrate, while RING or U-box ubiquitin ligases act as a scaffold that allows the transfer of ubiquitin from the E2 to the target protein (18, 20, 24, 25). The RBR E3s consist of two RING domains（RING1 and RING2）that are separated by an in-between-RING domain. RBR E3s catalyzed ubiquitin transfer to substrate through RING-HECT-hybrid mechanism that RING1 first recruits the ubiquitin loaded E2 and then the ubiquitin binds to the catalytic cysteine of RING2 in a thioester linkage, followed by the ubiquitin transfer to the substrate (28).

Non-canonical ubiquitination independent of E1 and E2

Given the central role of ubiquitin system to eukaryotic defense systems, it is not surprising that many bacterial pathogens exploit host ubiquitin systems for their own benefit. One such example is that some bacterial effector proteins are able to mimic and hijack the activity of eukaryotic E3 ligases (18, 29-31). Recent studies reveal that members of SidE effector family of the bacterial pathogen *Legionella pneumophila* are capable of catalyzing ubiquitination without the requirement of E1 and E2 enzymes of the host ubiquitination system (32, 33) (Figure 3). In this novel ubiquitination mechanism, ubiquitin is first activated by ADP-ribosylation at Arg42 and forms the ADPR-Ub intermediate. This intermediate then releases from the mART domain and diffuses to the ubiquitin-binding surface of the PDE domain. The ADPR-Ub is then catalyzed by the PED domain to release AMP and PR-Ub, concomitant with the attachment of ubiquitin to serine residues of substrate proteins via a phosphoribosyl linker (34, 35). This effect of SdiE family members can be antagonized by SidJ via cleaving the phosphodiester bond (36). This non-canonical ubiquitination mechanism expands the spectrum of ubiquitination system and the most exciting question now is whether this kind of ubiquitination also exist in eukaryotes.

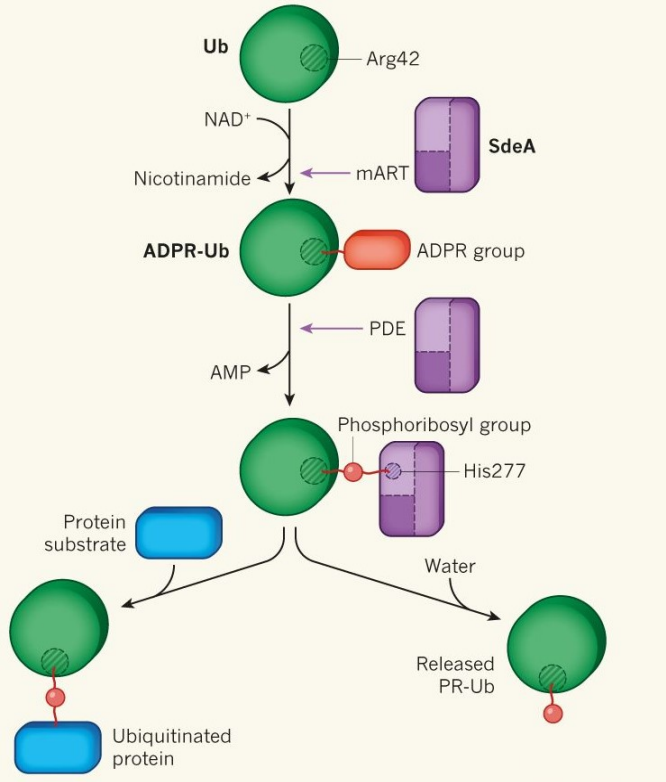


Figure 3. The Non-canonical ubiquitination mechanism used by bacteria (33). In the first step, the enzyme's mART domain processes NAD^+ and adds an adenosine diphosphate ribose (ADPR) group to Arg42 of ubiquitin. This generates ADPR-Ub in a reaction that releases nicotinamide. ADPR-Ub is then processed by the enzyme's PDE domain. The molecule AMP is released in a reaction that generates ubiquitin bound to a phosphoribosyl group (PR-Ub); the phosphoribosyl group, in turn, is covalently attached to the enzyme's His277. If a protein substrate enters the enzyme's active site, the enzyme catalyzes the attachment of PR-Ub to a serine residue on the protein substrate. If, instead, water enters the active site, PR-Ub is released.

The ubiquitin code

Ubiquitin has six lysine residues on its surface: Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63 (Figure 4A). A key feature of ubiquitin is that it can be ubiquitinated on any of its seven lysine residues, as well as the its N-terminus (Met1), leading to eight types of homotypic polyubiquitin chains and an almost infinite variety of heterotypic or branched polyubiquitin chains (9) (Figure 5B).

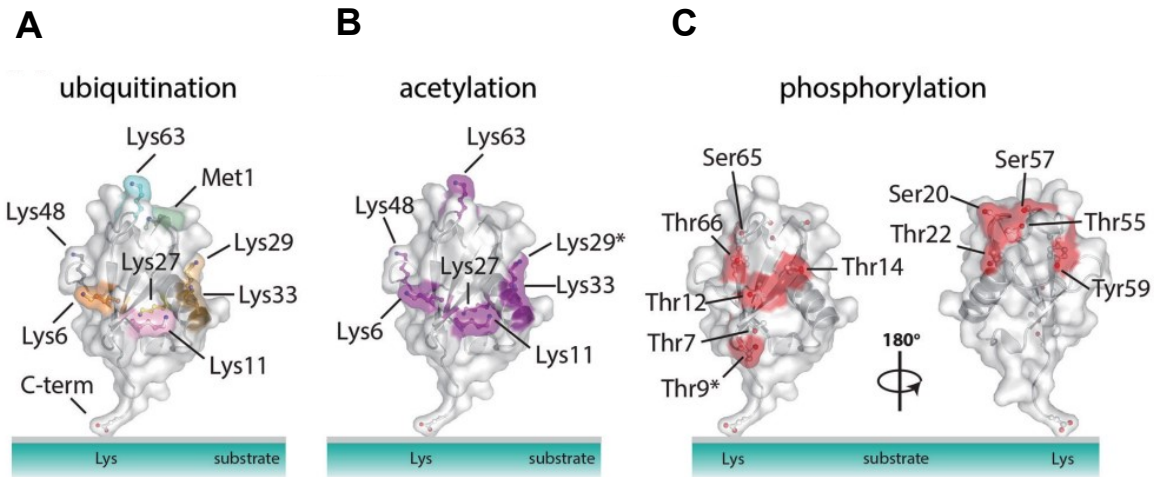


Figure 4. Modification sites on ubiquitin (7). (A) Structure of ubiquitin highlighting the eight sites of ubiquitination. (B) Six out of seven Lys residues on ubiquitin have been reported to be acetylated in proteomics datasets. An asterisk marks the seventh Lys, Lys29, for which acetylation has not been identified to date. (C) Identified phosphorylation sites of ubiquitin are displayed according to proteomic analysis. Red spheres indicate phosphorylatable hydroxyl groups on Ser/Thr and Tyr residues. The structure was rotated 180 degrees to show all phosphorylation sites. An asterisk on Thr9 indicates this site is ambiguously assigned.

All possible linkages types have been detected in cells through proteomics studies (37-41) and are thought to be the distinct signals recognized by specific ubiquitin receptors. Although many types of ubiquitination exist in cells, the most abundant forms are monoubiquitination, as well as Lys48-linked and Lys-63-linked polyubiquitination (42) (Figure 5A). The biological outcomes are diverse depending on the ubiquitin chain types. Monoubiquitination regulates processes such as membrane transport, histone regulation, and endocytosis (43, 44). Lys48-linked chains, the most predominant linkage type in cells (39), are related to protein degradation (5, 7), whereas Lys63-linke chains, the second most abundant linkage type in cells (39), play roles in a range of other processes such as DNA repair and protein trafficking (44). Linkages other than Lys48, including Lys11 and Lys29, also appear to serve as a signal for proteasome degradation (39, 45, 46). Additional linkage types such as Lys6, Lys27, and Lys33 have been assigned respective functions in ubiquitin-dependent signaling (47, 48). More recently, studies show that branched or heterotypic polyubiquitin chains including Lys11 branched, Lys48-Lys63 branched, and Lys63-Met1 mixed may serve as independent signals in cells and relate to multiply cellular processes (49-51). Ubiquitin can also be modified by other PTMs such as SUMOylation, phosphorylation, and acetylation, increasing the complexity of ubiquitin code (7, 47) (Figure 4B,C). The conformations of different polyubiquitin are either compact or open. Lys48-, Lys6-, and Lys11-linked chains adopt compact conformation in which the ubiquitin moieties interact via Ile44 or Il36, while Met- and Lys63-linked chains mostly display open conformations (9).

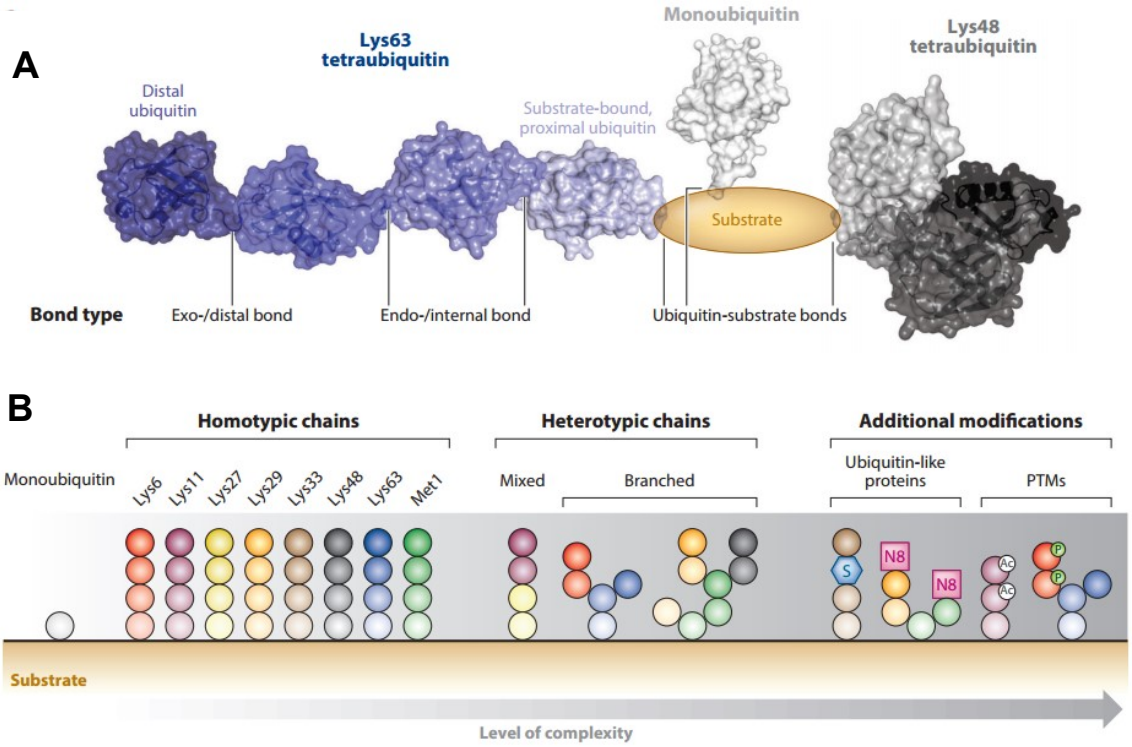


Figure 5. The complexity of ubiquitin modifications (52). (A) The three most common forms of ubiquitination—monoubiquitination, Lys48-linked polyubiquitination, and Lys63-linked polyubiquitination—are shown as structural models attached to a virtual substrate. Distinct chain topologies generate different structures, contributing to specific recognition. The nomenclature of the various bond types is introduced. (B) Schematic representation of the increasingly complex ubiquitin code. Monoubiquitin, the least complex modification, can be extended on lysine residues or its N-terminal methionine (Met1) residue, giving rise to eight homotypic polyubiquitin chains. Heterotypic chains contain more than one linkage type in mixed or branched (also known as forked) polymers. Additional layers of complexity arise through the cross talk between ubiquitin and ubiquitin-like proteins, such as SUMO (S) or NEDD8 (N8), as well as with other PTMs, including acetylation (Ac) and phosphorylation (P). Abbreviations: NEDD8, neural precursor cell expressed, developmentally downregulated 8; PTM, posttranslational modification; SUMO, small ubiquitin related modifier.

Deubiquitinating enzymes

Ubiquitination is reversible and proteases that remove ubiquitin signals are called deubiquitinating enzymes (DUBs). DUBs can remove ubiquitin by catalyzing the hydrolysis of isopeptide bond (Figure 2). Therefore, ubiquitin conjugation and deconjugating are balanced and tightly regulated by E3 and DUB. There are more than 100 DUBs encoded in the human genome (53). Based on the structure studies, their isopeptidase activities are conferred by distinct catalytic domains (52-55). Recently, the identification of a new family MINDY extends the DUBs to six distinct families (56). Five families, ubiquitin specific protease (USP), ubiquitin C-terminal hydrolase (UCH), the ovarian tumor protease (OTU), Machado-Joseph disease (MJD), and the motif interacting with ubiquitin (MIU) containing novel DUB family (MINDY), are cysteine proteases that contain a catalytic trad of Cys, His, and

Asp/Asn. The other family is Zn-dependent JAB1/MPN/MOV34 metalloprotease (JAMM) (52-55).

Functions of DUBs

DUBs play several roles in the ubiquitin pathways (Figure 2). First. Ubiquitin is always expressed either as a proprotein fused to one of two ribosomal proteins or as linear polyubiquitin which is consist of multiple copies of monoubiquitin. DUBs are required to process these proproteins or linear polyubiquitin to generate the mature ubiquitin monomer. Second, DUBs can remove ubiquitin chains from ubiquitinated proteins. In this role, DUBs reverse the ubiquitin signaling or stabilize proteins by rescue proteins from degradation. Third, Ubiquitin is recycled by DUBs and its homeostasis therefore be carefully controlled. Finally, DUBs can be used to trim polyubiquitin chains. (53, 55). Deubiquitination is a highly regulated process like ubiquitination that has been implicated in a range of cellular functions, such as cell cycle regulation, protein degradation, gene expression, DNA repair, kinase activation, and microbial pathogenesis (53).

USP family deubiquitinating enzymes

In human genome, USPs constitute the largest family of DUBs counting more than 50 members (53-55, 57) (Figure 6). Research on USPs has revealed many related specific biological pathways. For example, USP7/HAUSP regulates p53 and its E3 ubiquitin ligase MDM2 levels in cells (58), while USP28 control the stability of c-myc by removing degradative Lys48-linked polyubiquitin chains (59). USP14 is associated with the proteasome and can recycle ubiquitin by removing L48-linked ubiquitin chains from proteasome substrates (60). CYLD is a Lys63-specific and inhibits activation of the IKK kinase complex in the NF-κB pathway (61). USP44 deubiquitylates CDC20 and its activity is important for the regulation of the spindle checkpoint (55). However, there are still many USP enzymes that are not studied well.

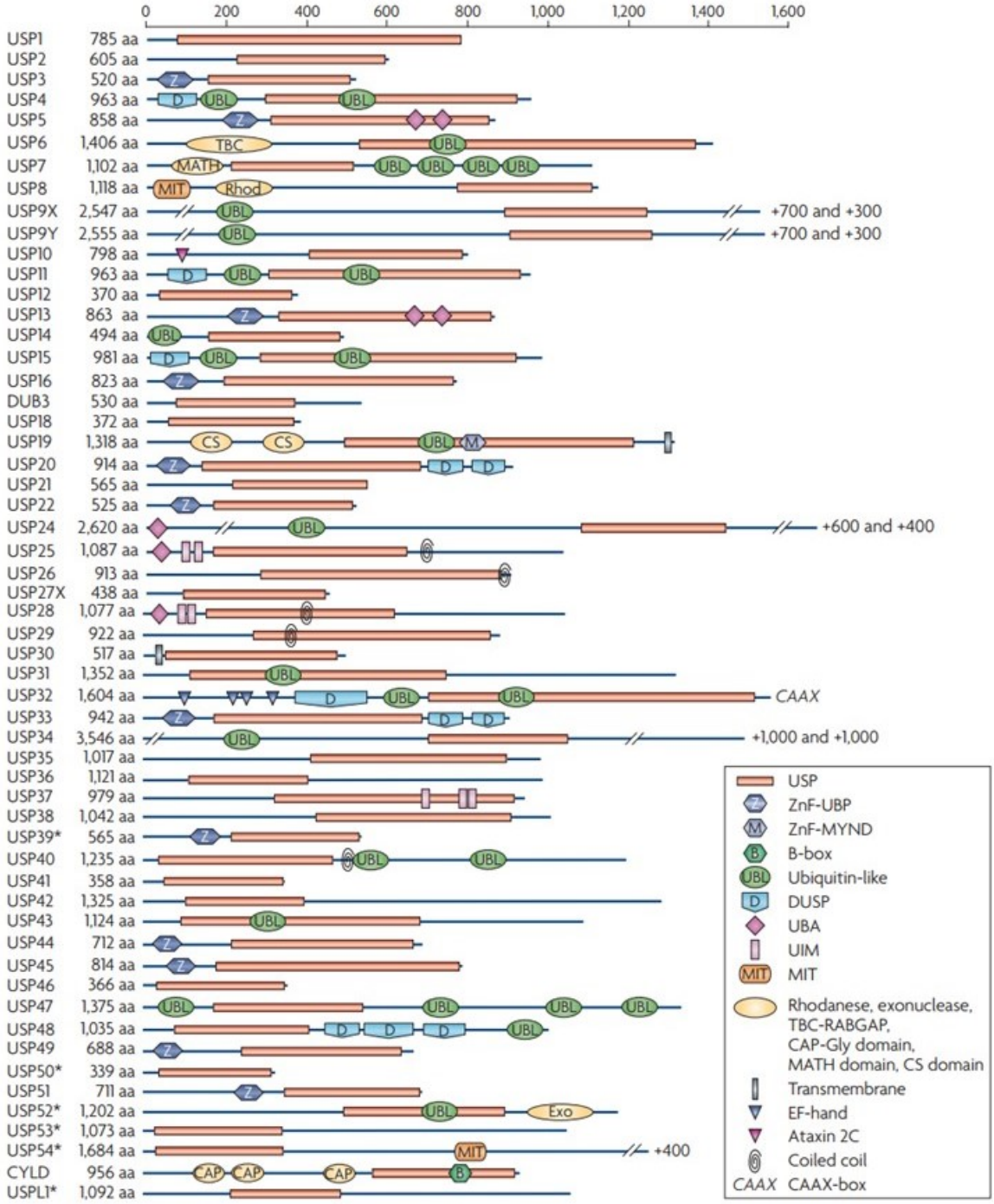


Figure 6. Domain structure of ubiquitin-specific proteases (55). The domain architecture of USPs reveals an abundance of predicted ubiquitin-binding domains, including the zinc finger ubiquitin-specific protease (ZnF-UBP) domain, the ubiquitin-interacting motif (UIM) and the ubiquitin-associated (UBA) domain. Also common are potential ubiquitin-like (UBL) domains, although many of these have not been validated at a structural level. Other domains are: Ataxin 2C, ataxin 2-like carboxy-terminal domain; CAP, CAP-Gly domain; CS, CHORD-SGT1 domain; DUSP, domain in USPs; EF-hand, Ca2+ binding motif; MATH, meprin and TRAF homology domain; TBC/RABGAP, domain in Tre-2, Bub2 and Cdc16/RAB GTPase activating protein; ZnF-MYND, MYND (myeloid, nervy and DEAF1)-type zinc fingers.

USP structure

All USPs contain a catalytic core of ~350 amino acids, in which the N-terminal Cys-box and a C-terminal His-box comprise the catalytic Cys and His residues, respectively. However, the size of USPs ranges from 339 (USP50) to 3546 (USP34) amino acids (Figure 6). This is because that many USPs not only contain a conserved catalytic domain but also different additional domains that can be involved in the protein-protein interaction and localization (55, 57) (Figure 6). Structure studies of various USPs, including USP2, USP4, USP7, USP8, USP12, USP14, USP18, USP21, USP30, USP46, show that the catalytic core of USPs adopts a conserved fold consisting of three subdomains known as Palm, Thumb, and Fingers (Figure 7A), in which the active site cysteine is located between the Palm and Thumb while the Fingers grip the “distal” ubiquitin (62-71). Only CYLD lacks the Finger subdomain (71). Crystal structure of the USP7-ubiquitin aldehyde complex shows that the C-terminal tail of ubiquitin inserts into the catalytic triad (Figure 7B).

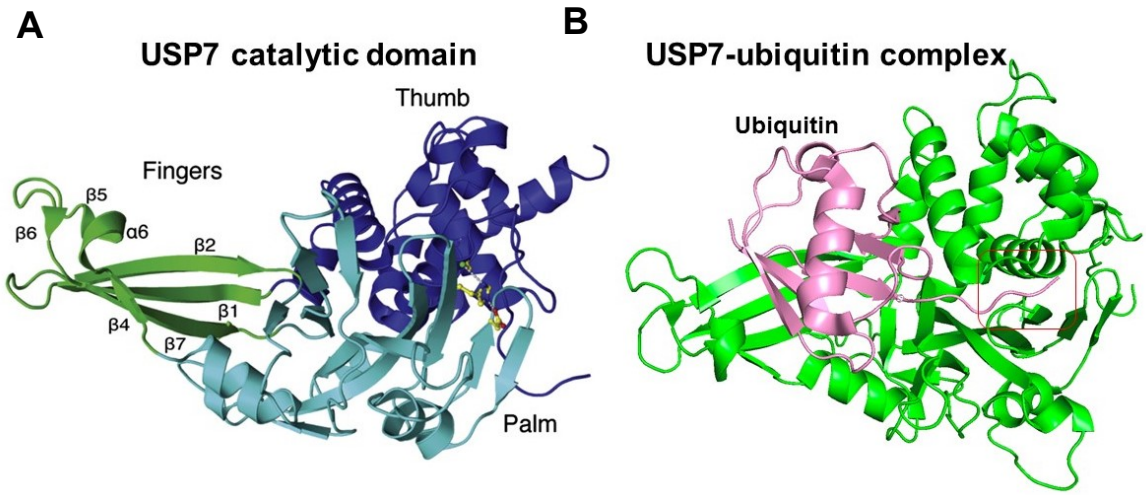


Figure 7. Crystal structure of USP7. (A) Structure of the catalytic domain of USP7 (PDB code 1NB8) is displayed in ribbon representation. Thumb, Palm and Fingers are labeled in blue, cyan and green respectively. Active site catalytic triad residues (Cys, His, and Asn) are shown in stick representation. (B) Ribbon representation of USP7-ubiquitin aldehyde complex (PDB code 1NBF). USP7 is labeled in green, while ubiquitin is shown in pink.

USP additional domains

These additional domains of USPs are either located into the conserved catalytic core or the N-terminal and C-terminal (Figure 6). These domains include many predicted ubiquitin-binding domains, such as the zinc finger ubiquitin-specific protease (ZnF-UBP) domain, the ubiquitin-interacting motif (UIM), ubiquitin-like (UBL) domain, and the ubiquitin-associated (UBA) domain (Figure 6). The catalytic activities of USPs are usually modulated through these additional domains, adding a layer to the regulation to these enzymes (55, 57). USP5 contains two UBA domains inserted in the catalytic core that can provide additional ubiquitin-binding sites and affect enzyme activity (72). The ZnF-UBP domain of USP5 is revealed to be associated with substrate targeting and specificity (73), and the crystal structure of full length USP5 reveals a previous unpredicted domain (nUBP), that is tightly bound to the catalytic core and essential for the activation of catalytic activity (74). A C-terminal 19 amino acids peptide binds the activation cleft in the catalytic domain and stabilizes the catalytically competent conformation, thus, enhances the activity of USP7 (75). This activation can be enhanced allosterically by the metabolic enzyme GMPS 12 (76). The binding of USP4 N-terminal DUSP-UBL domain promotes a change of a switching loop near the active site, hence enhances ubiquitin dissociation and makes it achieve full catalytic activity (64). The N-terminal tandem UIMs of USP25 is required for sufficient hydrolysis for ubiquitin chain, while the SIM, also locates at the N-terminal, interacts with SUMO proteins and promotes the SUMOylation at K99, leading to the inhibition of ubiquitin chain hydrolysis (77, 78).

USP25 is target of diverse PTMs

USP 25 is a member of USP family and shows strong activity in hydrolyzing K48 and K63 linked poly-ubiquitin chains, which are the most abundant linkage types in cells. There are three isoforms produce by alternative splicing, two of them (1055 and 1087 amino acids, respectively) are extensively expressed, while the longest isoform (named USP25m, 1125 amino acids) is only expressed in muscle tissues (79). USP25 is a target of different posttranslational modifications including phosphorylation, SUMOylation, and ubiquitination, concomitant with distinct outcomes (Figure 8). The SYK non-receptor tyrosine kinase can specifically

phosphorylate USP25 and decrease its cellular levels (80). USP25 enzymatic activity is suppressed after vaccinia-related kinase2 mediated phosphorylation at Thr680, Thr727, and Ser745, which controls the stability of the eukaryotic chaperonin TRIC (81) (Figure 8B). SUMO and ubiquitin regulate the catalytic activity of USP25 by conjugation at Lys99 with opposite effects: ubiquitination activates while SUMOylation inhibits the USP25 activity (77, 82).

Biological roles of USP25

USP25 has been associated with inflammation, immune response and cancer. For example, USP25 is involved in the Endoplasmic Reticulim-associated degradation (83), and acute ER stress regulates amyloid precursor protein processing through ubiquitin-dependent degradation by USP25 (84). USP25 is a negative regulator of IL-17-mediated signaling and inflammation through removal of ubiquitination in TRAF3, TRAF5, and TRAF6 and can regulate TLR4-dependent innate immune responses (85-87). Moreover, the type I interferon-IRF7 can activate the expression of USP25 gene, which is essential for innate immune signaling (88). Several evidences point to the involvement of USP25 in cancer. USP25 gene is found greater than threefold overexpression in human breast cancer tissue (89), and further study shows it as a putative tumor suppressor in human lung cancer (90). Another study in human non-small cell lung cancer reveals miR-200c inhibits invasion and metastasis by directly targeting USP25 and reducing the expression level (91). Recently, it has been shown that USP25 directly interacts with tankyrases through its C-terminal tail and promotes their deubiquitination and stabilization, thus regulating Wnt/β-catenin signaling pathway, making an important impact in cell proliferation and human cancer development (92). Usp25m is revealed as a protease that required for insulin-stimulated TUG cleavage and GLUT4 translocation in adipocytes (93). USP25 has a paralogue USP28 that shares high sequence similarity (51%). However, the functions and targets of these two proteins mostly do not overlap until now.

USP25 domain architecture

Despite these important roles in cellular pathways, structural insights of USP25 activity regulation are still not clear. The N-terminal domain of USP25 contains one ubiquitin associated domain (UBA), one SUMO interacting motif (SIM), and two ubiquitin interacting motifs (UIM1 and UIM2) (78). The UIM1 and UIM2 interact and recruit ubiquitin chains to USP25 catalytic domain, thus increase the polyubiquitin chain hydrolysis activity (Figure 8A). However, SUMO-modified USP25 maintains the ability to hydrolyze Ub-AMC, but exhibits lower activity during polyubiquitin chain degradation, which indicates that SUMO modification affects the recruitment of ubiquitin but not the catalytic domain activity (77) (Figure 8B). The same SUMOylation mediated activity inhibition also has been reported in USP28 (94). Although the basic elements of the catalytic domain of USP25 are conserved, there is a large domain (175 residues) inserted within the catalytic core (Figure 8). Also, previous studies show USP25 can form homodimers or oligomers in vivo and the catalytic domain (between residues 153 to 679) is relevant for this dimerization or oligomerization (77, 82). This indicates that the large inserted domain within the catalytic core may responsible for the dimerization or oligomerization. Finally, the C-terminal domain of USP25 is involved in phosphorylation and substrate recognition (80, 81), directly interacting with tankyrases to promote their deubiquitination and stabilization (92) (Figure 8A). Since regulation of the levels and activity of tankyrases is mechanistically important in controlling Wnt signaling, USP25 is identified as a positive regulator o Wnt signaling (92).

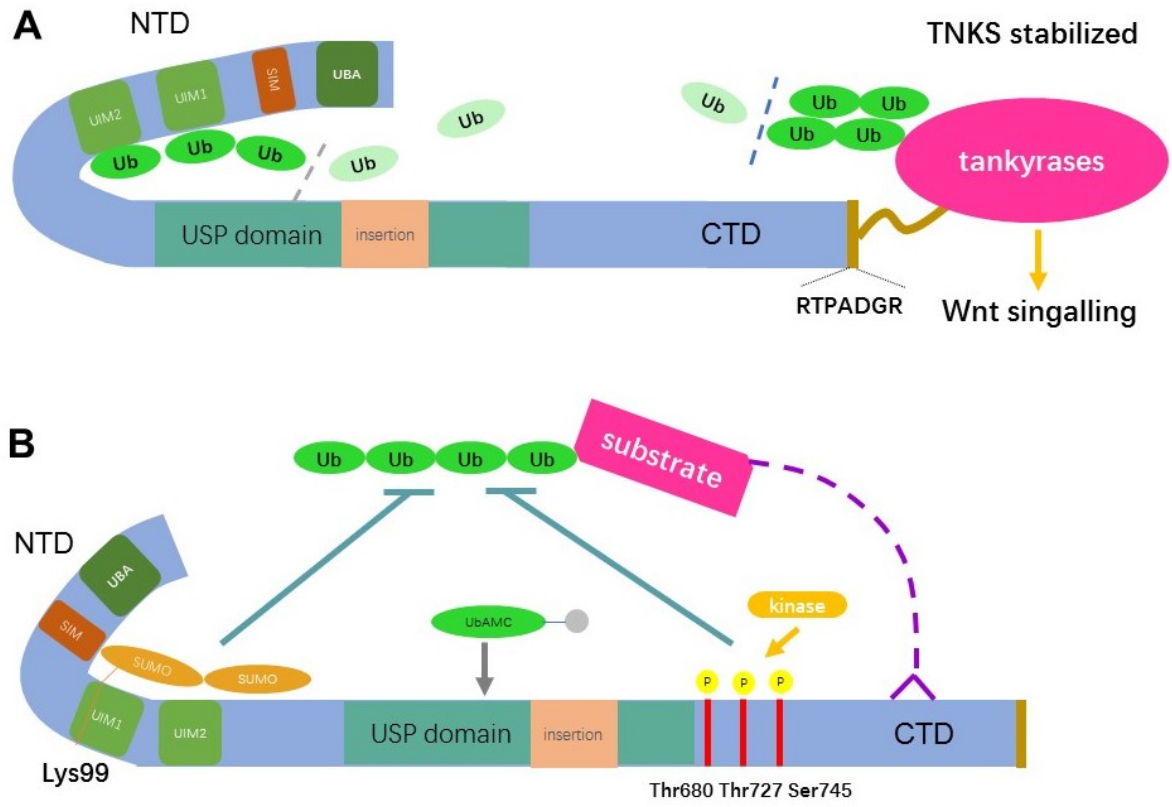


Figure 8. Scheme of USP25 domain architecture and its regulations by phosphorylation, ubiquitination, and SUMOylation. (A) USP25 are composed of the N-terminal domain (NTD), the catalytic domain (USP domain), and the C-terminal domain (CTD). NTD of USP25 contains three ubiquitin-binding domains (UBA, UIM1 and UIM2) and one SUMO interacting motif (SIM). UIM1 and UIM2 are important for the recruitment of ubiquitin chains to USP domain, and can promote the hydrolysis activity. The CTD tail (residue 1049-1055) directly interact with tankyrases to promote their deubiquitination and stabilization. The large unknown insertion domain is indicated in pink color. (B) SIM interaction with SUMO promotes the SUMOylation at Lys99 and inhibits the catalytic activity for polyubiquitin chain. CTD of USP25 is related to the substrate recognition and can be phosphorylated at Thr680, Thr727, and Ser745, which leads to the suppression of the enzymatic activity.

The SUMO system and E1-E2 protein-protein recognition

Ubiquitin-like proteins

Along with the discovery of ubiquitin, a set of ubiquitin-like proteins (Ubls) have been found as protein modifiers. Ubls including SUMO, Nedd8, Atg8, Atg12, ISG15, and FAT10, adopt the same β-grasp fold as ubiquitin. Similar to ubiquitin, most Ubls can be attached to substrates through the interplay of three types of proteins: E1 activating enzymes, E2 carrier enzymes, and E3 ligases (10, 13, 19, 95-98). In some cases, Ubls can be directly conjugated to substrates, such as RanGAP1, independent of E3 (99). Ubls attach to substrates to regulate their activity, localization, stability, or macromolecular interaction (96, 98). For example, both mono and chain SUMOylation are related to regulation of protein activity, complex formation, DNA repair, and transcriptional activation (100-102). Nedd8 conjugation to cullin proteins activates the ubiquitin conjugation. Atg8 and Atg12 play roles in autophagy, while ISG15 and FAT10 are associated with the innate immune response (96). Unlike ubiquitin, some Ubls, including SAMPs (small archaeal mofidier proteins) and Tbermus (TtuB, tRNA-two-thiouridine B), also exist in prokaryotes. They play roles not only in protein modification, but also in sulfur transfer, thus are thought to be the ancestors of eukaryotic ubiquitin/Ubls (103).

SUMO and SUMOylation

Small ubiquitin-related modifier (SUMO) protein is the most extensively studied Ubl. The sequence identity of SUMO and ubiquitin is only ~18%, but the folded structures of them are quite similar (101). SUMO is longer than ubiquitin and has an extra N terminal (~20 amino acids) with high flexibility, which primarily serve as acceptor in SUMO chain formation. SUMO is extensively expressed in all eukaryotes. Yeast and invertebrates have a single SUMO protein named Smt3, while vertebrates have several SUMO proteins. Mammalians such as human express four SUMO proteins: SUMO1, SUMO2, SUMO3, and SUMO4 (Table 1). SUMO2 and SUMO3 share 97% sequence identity and cannot be distinguished by antibodies (101, 102). However, SUMO1 is quite different from SUMO2/3 and only shares ~47% sequence identity with SUMO1 and SUMO2 (102). In contrast with other SUMO proteins, the roles and substrates of SUMO4 are still unclear now.

Despite the presence of SUMO4 mRNA in kidney, spleen, and lymph node, endogenous SUMO4 protein has not been detected (104, 105). In *A. thaliana*, there are eight encoding genes of SUMO isoforms (AtSUMO) (Table 1), but only SUMO1, SUMO2, SUMO3, an SUMO5 are expressed (106, 107). AtSUMO1 and AtSUMO2 (orthologs of human SUMO2/3) are essential in *A. thaliana* and with 83% sequence identity, while AtSUMO3 and AtSUMO5 only share 42% and 30% sequence identity respectively (108-111). Other than this, In vitro time course assays shows that AtSUMO1 and AtSUMO2 display much higher conjugation rate than AtSUMO3 and AtSUMO5 (110)

	proteins in *S. cerevisiae*	proteins in *H. sapiens*	proteins in *A. thaliana*
SUMO isoforms	Smt3	SUMO1 SUMO2 SUMO3 SUMO4	SUMO1 SUMO2 SUMO3 SUMO4 SUMO5 SUMO6 SUMO7 SUMO8
E1	Aos1 Uba2	Aos1 (SAE1) Uba2 (SAE2)	SAE1a SAE1b SAE2
E2	Ubc9	Ubc9	SCE1

Table 1. SUMOylation determinants in *Saccharomyces cerevisiae*, *Homo sapiens*, and *Arabidopsis thaliana*.

SUMO proteins are expressed immaturely with the C-terminal extensions. SUMO-specific proteases (SENPs) cleave the immature SUMO proteins to generate the mature forms, which expose the Gly-Gly motif (Figure 9). The mature SUMO is conjugated to target proteins through a three-step enzyme pathway as ubiquitin (Figure 9). SUMO E1 heterodimer SAE1/SAE2 activates the mature SUMO in an ATP-dependent step, leading to the formation of a thioester linkage between the C-terminal glycine of SUMO and the catalytic cysteine residue of SAE2. SUMO is then transferred to the E2 conjugating enzyme Ubc9, forming a thioester bond between the SUMO C-terminal and the catalytic cysteine residue of Ubc9. SUMO-loaded Ubc9 then transfer SUMO to substrates usually with the help of E3 ligases, forming an isopeptide bond between the C-terminal glycine of SUMO and a lysine residue in target protein. (101, 102). SUMO proteins are conjugated to substrates as monomers or polymers in one or multiple sites, serve as diverse signals leading to different biological outcomes. SUMOylation is a protein modification which can

be reversed by SUMO proteases (Figure 9). In human, SUMO-specific proteases (SENPs) include SENP1, SENP2, SENP3, SENP5, SENP6, and SENP7. SUMO proteases have two functions: cleave the C-terminal end SUMO precursors to generate mature forms, remove SUMO from target proteins (102).

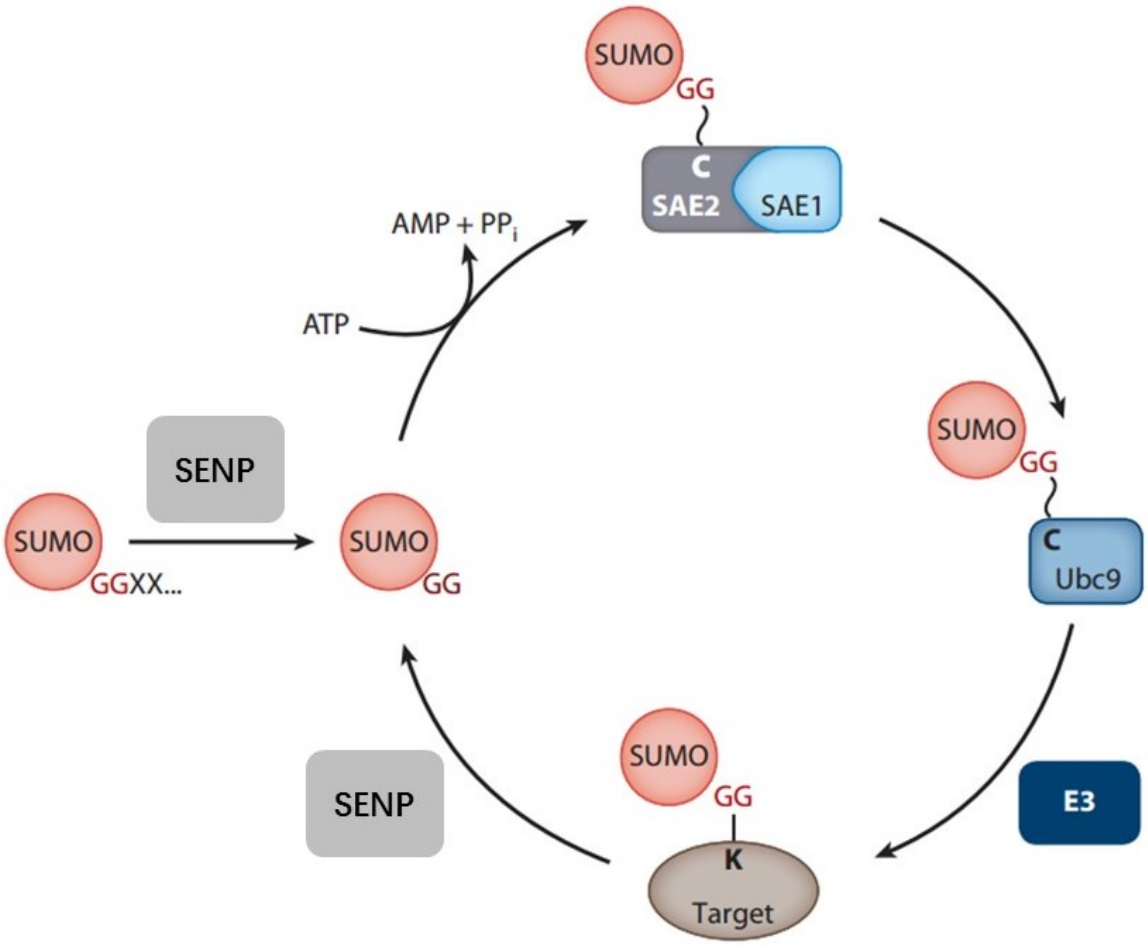


Figure 9. The cycle of reversible SUMOylation (102).

E1 E2 recognition

Ubl E1 is a heterodimer composed of two subunits. In SUMO system, they are SUMO-activating enzyme subunit 1 and 2 (also known as Aos1 and Uba2). Canonical E1s, including ubiquitin, NEDD8, and SUMO pathways, have a common architecture including a ubiquitin fold domain (UFD) at the C-terminal that serves to select cognate E2s (98, 112-115). However, the SUMO E1 SAE2 subunit has a C-terminal extension (~100 amino acids) that is absent in Nedd8 Uba3, ubiquitin Uba1, indicating that this C-terminal extension is not conserved in all E1 enzymes (112). The first specificity step in the pathway of Ubl conjugation is the interaction of the

Ubl modifier with its particular E1 activating enzyme (98, 110, 116). Next, the activated Ubl-thioester is transferred to the catalytic cysteine residue of the E2 conjugating enzyme (Ubc9 in SUMO). In this step, E1 must select cognate E2 from a range of structurally similar proteins, which represents a second specificity step in the pathway by the formation of the E1-E2 complex. The E1 UFD domain plays a major role in the specific E1-E2 interactions, as truncation of this domain leads to a significant decrease of the E2 transthiolation rate (117). This UFD-E2 interactions have been solved in ubiquitin, Nedd8, and SUMO systems (Figure 10A), showing the direct binding of E2 to the E1 UFD domain (114, 115, 118, 119). The UFD domain is connected to the E1 AAD domain (active adenylation domain) through a flexible hinge. In SUMO and Nedd8 systems, this disordered hinge is converted to the ordered state upon E2 binding, and it must undergo a rotation to bring the catalytic cysteine residues of E1 and E2 into proximity for thioester transfer (120, 121) (Figure 10B). In addition, crystal structure of ubiquitin E1-E2 complex in the thioester intermediate reveals dual interactions of E2 to the UFD domain and to the E1

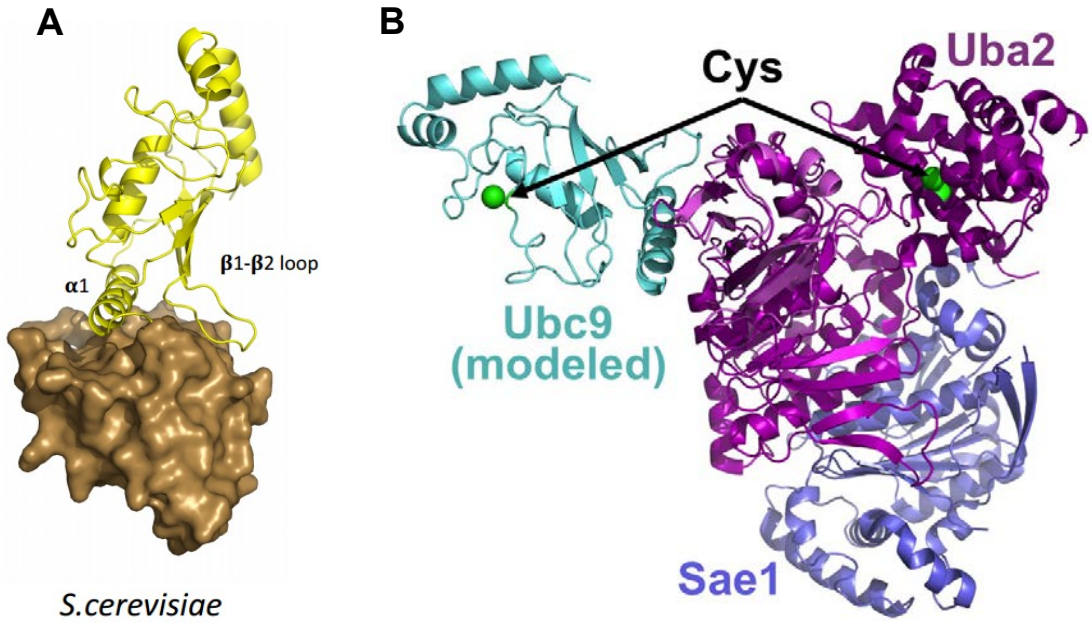


Figure 10. Crystal structure of E1 UFD and E2 complexe of *S.cerevisiae*. (A) Display of *S.cerevisiae* E1 (Uba2) UFD interaction with E2 (Ubc9). UFD domain and Ubc9 are represented in cartoon (yellow) and sphere (grey) respectively. (B) Structure of *S.cerevisiae* UFD-E2 (magenta, cyan), with UFD superimposed on the corresponding region of human Uba2 from the Sae1-Uba2-Sumo1~AMSN complex (PDB code 3KYC) (119). Sae1 is shown in blue, human Uba2 is shown in purple, and Sumo~AMSN is not shown for simplification. Catalytic cysteine residues of Ubc9 and Uba2 (shown in green) must undergo a rotation to bring them into proximity for thioester transfer.

catalytic cysteine domain, which occurs after the significant rotation of the UFD domain, providing the structural basis for the isoenergetic thioester transfer between the E1 and the E2 (114). This interaction is also proposed in the SUMO system by NMR analyses, although E1 UFD-E2 interactions display higher affinity (Kd= 1.2μM) (122) than E1 Cys-E2 interactions (Kd= 87μM) (123), supporting a major role of the E1 UFD domain in E2 recruitment.

The protein sequence variation of the E1 UFD domains in different Ubl systems is quite significant, especially in the binding region to the E2 enzymes (Figure 11A). Crystal structures of different UFD domains display an analogous β-grasp fold, and the UFD-E2 interactions are mediated by the same side of the β-sheet residues (113-115, 117, 124, 125). However, in the reported complex structures of ubiquitin, Nedd8, and SUMO systems, superposition of the E2 enzymes reveal distinct orientations of the UFD domain, which is a direct consequence of different contacts in each Ubl system (114, 115, 119). Despite of the high conservation (more than 50%) of SUMO E2 enzymes in human, yeast, and *A. thaliana*, their UFD domains show a little protein sequence homology, and it is even lower considering only the binding region to E2 (Figure 11). Even so, E1s can efficiently select the specific cognate E2 enzymes through the interactions between its UFD domain and the conserved binding surfaces in E2 enzymes. In this case, instead of sequence homology analysis, structural studies are required to identify the molecular determinants that mediate UFD-E2 interactions in evolutionary distant organisms.

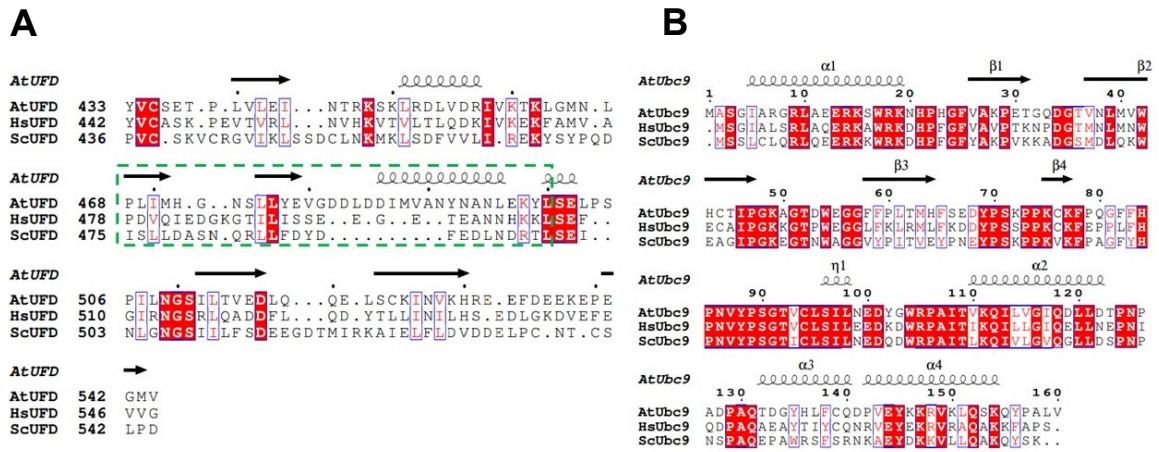


Figure 11. Sequence alignments of SUMO E1 UFD and E2$_{Ubc9}$ in *Arabidopsis thaliana* (At), *Homo sapiens* (Hs), and *Saccharomyces cerevisiae* (Sc). (A) Sequence alignment of AtUFD, HsUFD, and ScUFD. The region involved in the interactions with E2 are marked in a green square. (B) E2 sequence alignment of AtUbc9, HsUbc9, and ScUbc9. Second structures are indicated.

SUMOylation of RanGAP1

Most E2s do not show target specificity and consequently require E3s to promote the selective interactions with their substrates. However, the SUMO E2 enzyme Ubc9 is sufficient for the recognition and lysine modification of the known targets, which contain a SUMO consensus motif comprised of Ψ-Lys-X-Asp/Glu (Ψ is a hydrophobic residue, Lys is the acceptor lysine, and X is any residue) (98, 126). The structural basis for this kind of recognition was first elucidated in the crystal structure of mammalian Ubc9-RanGAP1 complex (99) (Figure 12). In this complex, the hydrophobic residue of RanGAP1 (Leu523, Ψ in the consensus motif) contacts a hydrophobic interface formed by Pro128-Ala129-Gln130 residues of human Ubc9. The acceptor Lys524 (Lys in the consensus motif) residue is positioned through hydrogen bond interactions with Asp127 and aliphatic contacts to Tyr87, bringing the primary amine and the catalytic cysteine of Ubc9 into proximity. Glu526 (Asp/Glu in the consensus motif) forms aliphatic contacts to Tyr87, as well as hydrogen bond contacts to Ser89 and Thr91. Since RanGAP1 SUMOylation with Ubc9 is independent of E3, it can be used as a model substrate of in vitro SUMOylation to study E1-E2 interactions.

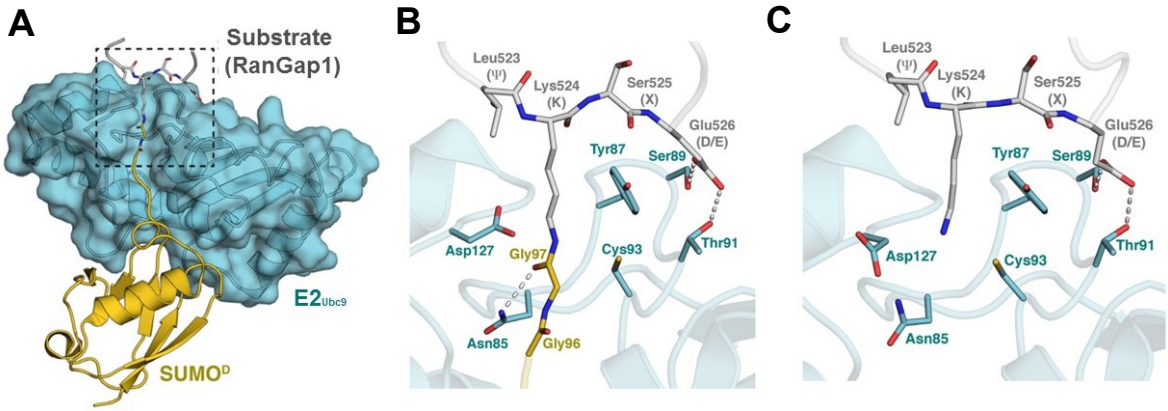


Figure 12. SUMO conjugation of RanGAP1 via $E2_{Ubc9}$ (98). (A) Overall view and (B) close-up view of $E2_{Ubc9}$/SUMO-RanGAP1 complex (PDB code 1Z5S) illustrating how the RANGAP1 substrate and SUMO are positioned in the $E2_{Ubc9}$ active site. This state represents a product complex after conjugation where SUMO, colored yellow, has been transferred to a lysine of RANGAP1 colored gray. $SUMO^D$ designates a SUMO protein in donor configuration. $E2_{Ubc9}$ is in cartoon representation colored cyan. Certain residues of the E2 active site are in stick representation. The consensus sequence for substrate recognition by $E2_{Ubc9}$ is indicated on top. (C) Close-up view of $E2_{Ubc9}$/RANGAP1 (PDB 1KPS) representing RanGAP1 binding prior to catalysis in the absence of SUMO.

USP25 in the cross talk of ubiquitin and SUMO pathways

Given that universal roles of both ubiquitin and SUMO on cellular processes, it is not surprising to find situations where these two PTMs communicate. Actually, a growing number of proteins have been identified as substrates of both ubiquitin and SUMO conjugation targets. The cross talk of ubiquitin and SUMO pathways, together with a range of other PTMs, increases the complexity of cell signaling. Many DUBs can undergo diverse PTMs, leading to inactivation, activation localization or degradation (52). USP25 and its homolog USP28 are such kind of substrates, whose DUB activities are adjusted by phosphorylation, ubiquitination, or SUMOylation. USP25 harbors three ubiquitin binding domains (UBA, UIM1, and UIM2) that modulate its DUB activity (Figure 8A). Monoubiquitination of Lys99 located at UIM1 activates the enzymatic activity (82). The USP25 N-terminal region also contains a SIM, which promotes the SUMOylation at Lys99. This SUMOylation impairs the hydrolysis activity of tetraubiquitin, but do not affect the processing of a monoubiquitin substrate (Ubiquitin-AMC) (Figure 8B). These phenomena suggest steric antagonism of SUMOylation against polyubiquitin chains recruitment through UIM1 and UIM2 (77). Finally, USP25 can also be phosphorylate at Thr680, Thr727, and Ser745, leading to the suppression of its enzymatic activity (81) (Figure 8B). These PTMs on USP25 expand the regulation mechanism of this important DUB.

References for introduction

1. Consortium IHGS. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431(7011):931.
2. Hemantaranjan A. Plant stress tolerance physiological and molecular strategies: *Scientific Publishers*; 2016.
3. Kresge N, Simoni RD, Hill RL. The discovery of ubiquitin-mediated proteolysis by Aaron Ciechanover, Avram Hershko, and Irwin Rose. *Journal of Biological Chemistry*. 2006;281(40):e32-e.
4. Clague MJ, Heride C, Urbé S. The demographics of the ubiquitin system. *Trends in cell biology*. 2015;25(7):417-26.
5. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annual review of biochemistry* **67**, 425-479 (1998).
6. Welchman RL, Gordon C, Mayer RJ. Ubiquitin and ubiquitin-like proteins as multifunctional signals. Nature reviews Molecular cell biology. 2005;6(8):599.
7. Swatek KN, Komander D. Ubiquitin modifications. *Cell research*. 2016;26(4):399.
8. Varshavsky A. The ubiquitin system, autophagy, and regulated protein degradation. *Annual review of biochemistry*. 2017;86:123-8.
9. Komander D, Rape M. The ubiquitin code. *Annual review of biochemistry*. 2012;81:203-29.
10. Kerscher O, Felberbaum R, Hochstrasser M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol*. 2006;22:159-80.
11. Kimura Y, Tanaka K. Regulatory mechanisms involved in the control of ubiquitin homeostasis. *The journal of biochemistry*. 2010;147(6):793-8.
12. Pickart CM. Mechanisms underlying ubiquitination. *Annual review of biochemistry*. 2001;70(1):503-33.
13. Hochstrasser M. Origin and function of ubiquitin-like proteins. *Nature*. 2009;458(7237):422.
14. Ciechanover A, Ben-Saadon R. N-terminal ubiquitination: more protein substrates join in. *Trends in cell biology*. 2004;14(3):103-6.
15. Cadwell K, Coscoy L. Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. *Science*. 2005;309(5731):127-30.

16. Ravid T, Hochstrasser M. Autoregulation of an E2 enzyme by ubiquitin-chain assembly on its catalytic residue. *Nature cell biology*. 2007;9(4):422.
17. Wang X, Herr RA, Chua W-J, Lybarger L, Wiertz EJ, Hansen TH. Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-I by viral E3 ligase mK3. *The Journal of cell biology*. 2007;177(4):613-24.
18. Zheng N, Shabek N. Ubiquitin ligases: structure, function, and regulation. *Annual review of biochemistry*. 2017;86:129-57.
19. Schulman BA, Harper JW. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nature reviews Molecular cell biology*. 2009;10(5):319.
20. Deshaies RJ, Joazeiro CA. RING domain E3 ubiquitin ligases. *Annual review of biochemistry.* 2009;78.
21. Ye Y, Rape M. Building ubiquitin chains: E2 enzymes at work. *Nature reviews Molecular cell biology*. 2009;10(11):755.
22. Li W, Bengtson MH, Ulbrich A, Matsuda A, Reddy VA, Orth A, et al. Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PloS one*. 2008;3(1):e1487.
23. Berndsen CE, Wolberger C. New insights into ubiquitin E3 ligase mechanism. *Nature structural & molecular biology*. 2014;21(4):301.
24. Rotin D, Kumar S. Physiological functions of the HECT family of ubiquitin ligases. *Nature reviews Molecular cell biology*. 2009;10(6):398.
25. Metzger MB, Hristova VA, Weissman AM. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci*. 2012;125(3):531-7.
26. Ohi MD, Vander Kooi CW, Rosenberg JA, Chazin WJ, Gould KL. Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nature Structural and Molecular Biology.* 2003;10(4):250.
27. Aravind L, Koonin EV. The U box is a modified RING finger—a common domain in ubiquitination. *Current Biology*. 2000;10(4):R132-R4.
28. Dove KK, Klevit RE. RING-between-RING E3 ligases: emerging themes amid the variations. *Journal of molecular biology*. 2017.
29. Ashida H, Kim M, Sasakawa C. Exploitation of the host ubiquitin system by

human bacterial pathogens. *Nature Reviews Microbiology*. 2014;12(6):399.

30. Hicks SW, Galán JE. Hijacking the host ubiquitin pathway: structural strategies of bacterial E3 ubiquitin ligases. *Current opinion in microbiology*. 2010;13(1):41-6.
31. Maculins T, Fiskin E, Bhogaraju S, Dikic I. Bacteria-host relationship: ubiquitin ligases as weapons of invasion. *Cell research*. 2016;26(4):499.
32. Qiu J, Sheedlo MJ, Yu K, Tan Y, Nakayasu ES, Das C, et al. Ubiquitination independent of E1 and E2 enzymes by bacterial effectors. *Nature*. 2016;533(7601):120.
33. Wong K, Gehring K. Deciphering the catalytic mechanism of bacterial ubiquitination. *Nature Publishing Group*; 2018.
34. Wang Y, Shi M, Feng H, Zhu Y, Liu S, Gao A, et al. Structural Insights into Non-canonical Ubiquitination Catalyzed by SidE. *Cell*. 2018;173(5):1231-43. e16.
35. Dong Y, Mu Y, Xie Y, Zhang Y, Han Y, Zhou Y, et al. Structural basis of ubiquitin modification by the Legionella effector SdeA. *Nature*. 2018:1.
36. Qiu J, Yu K, Fei X, Liu Y, Nakayasu ES, Piehowski PD, et al. A unique deubiquitinase that deconjugates phosphoribosyl-linked protein ubiquitination. *Cell research*. 2017;27(7):865.
37. Elia AE, Boardman AP, Wang DC, Huttlin EL, Everley RA, Dephoure N, et al. Quantitative proteomic atlas of ubiquitination and acetylation in the DNA damage response. *Molecular cell*. 2015;59(5):867-81.
38. Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, et al. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Molecular & Cellular Proteomics*. 2011:mcp. M111. 013284.
39. Xu P, Duong DM, Seyfried NT, Cheng D, Xie Y, Robert J, et al. Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell*. 2009;137(1):133-45.
40. Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, et al. A proteomics approach to understanding protein ubiquitination. *Nature biotechnology*. 2003;21(8):921.
41. Kirkpatrick DS, Hathaway NA, Hanna J, Elsasser S, Rush J, Finley D, et al. Quantitative analysis of in vitro ubiquitinated cyclin B1 reveals complex chain topology. *Nature cell biology*. 2006;8(7):700.

42. Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Molecular cell*. 2011;44(2):325-40.
43. Hicke L. Ubiquitin and proteasomes: protein regulation by monoubiquitin. *Nature reviews Molecular cell biology*. 2001;2(3):195.
44. Chen ZJ, Sun LJ. Nonproteolytic functions of ubiquitin in cell signaling. *Molecular cell*. 2009;33(3):275-86.
45. Finley D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual review of biochemistry*. 2009;78:477-513.
46. Liu J, Han C, Xie B, Wu Y, Liu S, Chen K, et al. Rhbdd3 controls autoimmunity by suppressing the production of IL-6 by dendritic cells via K27-linked ubiquitination of the regulator NEMO. *Nature immunology*. 2014;15(7):612.
47. Yau R, Rape M. The increasing complexity of the ubiquitin code. *Nature cell biology*. 2016;18(6):579.
48. Kulathu Y, Komander D. Atypical ubiquitylation—the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nature reviews Molecular cell biology*. 2012;13(8):508.
49. Meyer H-J, Rape M. Enhanced protein degradation by branched ubiquitin chains. *Cell*. 2014;157(4):910-21.
50. Ohtake F, Saeki Y, Ishido S, Kanno J, Tanaka K. The K48-K63 branched ubiquitin chain regulates NF-κB signaling. *Molecular cell*. 2016;64(2):251-66.
51. Emmerich CH, Ordureau A, Strickson S, Arthur JSC, Pedrioli PG, Komander D, et al. Activation of the canonical IKK complex by K63/M1-linked hybrid ubiquitin chains. *Proceedings of the National Academy of Sciences*. 2013;110(38):15247-52.
52. Mevissen TE, Komander D. Mechanisms of deubiquitinase specificity and regulation. *Annual review of biochemistry*. 2017;86:159-92.
53. Reyes-Turcu FE, Ventii KH, Wilkinson KD. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annual review of biochemistry*. 2009;78:363-97.
54. Clague MJ, Barsukov I, Coulson JM, Liu H, Rigden DJ, Urbé S. Deubiquitylases from genes to organism. *Physiological reviews*. 2013;93(3):1289-315.
55. Komander D, Clague MJ, Urbé S. Breaking the chains: structure and function

of the deubiquitinases. *Nature reviews Molecular cell biology*. 2009;10(8):550.
56. Rehman SAA, Kristariyanto YA, Choi S-Y, Nkosi PJ, Weidlich S, Labib K, et al. MINDY-1 is a member of an evolutionarily conserved and structurally distinct new family of deubiquitinating enzymes. *Molecular cell*. 2016;63(1):146-55.
57. Nijman SM, Luna-Vargas MP, Velds A, Brummelkamp TR, Dirac AM, Sixma TK, et al. A genomic and functional inventory of deubiquitinating enzymes. *Cell*. 2005;123(5):773-86.
58. Brooks CL, Gu W. p53 ubiquitination: Mdm2 and beyond. *Molecular cell*. 2006;21(3):307-15.
59. Popov N, Wanzel M, Madiredjo M, Zhang D, Beijersbergen R, Bernards R, et al. The ubiquitin-specific protease USP28 is required for MYC stability. *Nature cell biology*. 2007;9(7):765.
60. Guterman A, Glickman MH. Deubiquitinating enzymes are IN (trinsic to proteasome function). Current Protein and Peptide Science. 2004;5(3):201-10.
61. Sun S-C. Deubiquitylation and regulation of the immune response. *Nature Reviews Immunology*. 2008;8(7):501.
62. Hu M, Li P, Li M, Li W, Yao T, Wu J-W, et al. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell*. 2002;111(7):1041-54.
63. Renatus M, Parrado SG, D'Arcy A, Eidhoff U, Gerhartz B, Hassiepen U, et al. Structural basis of ubiquitin recognition by the deubiquitinating protease USP2. *Structure*. 2006;14(8):1293-302.
64. Clerici M, Luna-Vargas MP, Faesen AC, Sixma TK. The DUSP–Ubl domain of USP4 enhances its catalytic efficiency by promoting ubiquitin exchange. *Nature communications*. 2014;5:5399.
65. Hu M, Li P, Song L, Jeffrey PD, Chernova TA, Wilkinson KD, et al. Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *The EMBO journal*. 2005;24(21):3747-56.
66. Avvakumov GV, Walker JR, Xue S, Finerty PJ, Mackenzie F, Newman EM, et al. Amino-terminal dimerization, NRDP1-rhodanese interaction, and inhibited catalytic domain conformation of the ubiquitin-specific protease 8 (USP8). *Journal of Biological Chemistry*. 2006;281(49):38061-70.
67. Basters A, Geurink PP, Röcker A, Witting KF, Tadayon R, Hess S, et al.

Structural basis of the specificity of USP18 toward ISG15. *Nature structural & molecular biology*. 2017;24(3):270.

68. Ye Y, Akutsu M, Reyes-Turcu F, Enchev RI, Wilkinson KD, Komander D. Polyubiquitin binding and cross-reactivity in the USP domain deubiquitinase USP21. *EMBO reports*. 2011;12(4):350-7.
69. Sato Y, Okatsu K, Saeki Y, Yamano K, Matsuda N, Kaiho A, et al. Structural basis for specific cleavage of Lys6-linked polyubiquitin chains by USP30. *Nature Structural and Molecular Biology*. 2017;24(11):911.
70. Dharadhar S, Clerici M, van Dijk WJ, Fish A, Sixma TK. A conserved two-step binding for the UAF1 regulator to the USP12 deubiquitinating enzyme. *Journal of structural biology*. 2016;196(3):437-47.
71. Komander D, Lord CJ, Scheel H, Swift S, Hofmann K, Ashworth A, et al. The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B box module. *Molecular cell*. 2008;29(4):451-64.
72. Reyes-Turcu FE, Shanks JR, Komander D, Wilkinson KD. Recognition of polyubiquitin isoforms by the multiple ubiquitin binding modules of isopeptidase T. *Journal of Biological Chemistry*. 2008;283(28):19581-92.
73. Reyes-Turcu FE, Horton JR, Mullally JE, Heroux A, Cheng X, Wilkinson KD. The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell*. 2006;124(6):1197-208.
74. Avvakumov GV, Walker JR, Xue S, Allali-Hassani A, Asinas A, Nair UB, et al. Two ZnF-UBP domains in isopeptidase T (USP5). *Biochemistry*. 2012;51(6):1188-98.
75. Rougé L, Bainbridge TW, Kwok M, Tong R, Di Lello P, Wertz IE, et al. Molecular understanding of USP7 substrate recognition and C-terminal activation. *Structure*. 2016;24(8):1335-45.
76. Faesen AC, Dirac AM, Shanmugham A, Ovaa H, Perrakis A, Sixma TK. Mechanism of USP7/HAUSP activation by its C-terminal ubiquitin-like domain and allosteric regulation by GMP-synthetase. *Molecular cell*. 2011;44(1):147-59.
77. Meulmeester E, Kunze M, Hsiao HH, Urlaub H, Melchior F. Mechanism and consequences for paralog-specific sumoylation of ubiquitin-specific protease 25. *Molecular cell*. 2008;30(5):610-9.

78. Yang Y, Shi L, Ding Y, Shi Y, Hu H-Y, Wen Y, et al. Structural and Functional Investigations of the N-Terminal Ubiquitin Binding Region of Usp25. *Biophysical journal*. 2017;112(10):2099-108.
79. Valero R, Marfany G, González-Angulo O, González-González G, Puelles L, Gonzàlez-Duarte R. USP25, a novel gene encoding a deubiquitinating enzyme, is located in the gene-poor region 21q11. 2. *Genomics*. 1999;62(3):395-405.
80. Cholay M, Reverdy C, Benarous R, Colland F, Daviet L. Functional interaction between the ubiquitin-specific protease 25 and the SYK tyrosine kinase. *Experimental cell research*. 2010;316(4):667-75.
81. Lee D, Kim S, Lee J, Song H, Kim H-j. vaccinia-related kinase 2 controls eukaryotic chaperonin Tric/cct stability by inhibiting Ubiquitine-specific protease 25: p24-003-sp. *The Febs Journal*. 2015;282:297.
82. Denuc A, Bosch-Comas A, Gonzàlez-Duarte R, Marfany G. The UBA-UIM domains of the USP25 regulate the enzyme ubiquitination state and modulate substrate recognition. *PLoS One*. 2009;4(5):e5571.
83. Blount JR, Burr AA, Denuc A, Marfany G, Todi SV. Ubiquitin-specific protease 25 functions in Endoplasmic Reticulum-associated degradation. *PloS one*. 2012;7(5):e36542.
84. Jung ES, Hong H, Kim C, Mook-Jung I. Acute ER stress regulates amyloid precursor protein processing through ubiquitin-dependent degradation. *Scientific reports*. 2015;5:8805.
85. Zhong B, Liu X, Wang X, Chang SH, Liu X, Wang A, et al. Negative regulation of IL-17-mediated signaling and inflammation by the ubiquitin-specific protease USP25. *Nature immunology*. 2012;13(11):1110.
86. Zhong B, Liu X, Wang X, Liu X, Li H, Darnay BG, et al. Ubiquitin-specific protease 25 regulates TLR4-dependent innate immune responses through deubiquitination of the adaptor protein TRAF3. *Sci Signal*. 2013;6(275):ra35-ra.
87. Lin D, Zhang M, Zhang M-X, Ren Y, Jin J, Zhao Q, et al. Induction of USP25 by viral infection promotes innate antiviral responses by mediating the stabilization of TRAF3 and TRAF6. *Proceedings of the National Academy of Sciences*. 2015;112(36):11324-9.
88. Ren Y, Zhao Y, Lin D, Xu X, Zhu Q, Yao J, et al. The type I interferon-IRF7 axis mediates transcriptional expression of Usp25. *Journal of Biological Chemistry*.

2016:jbc. M116. 718080.
89. Deng S, Zhou H, Xiong R, Lu Y, Yan D, Xing T, et al. Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome subunits (PSs) in breast cancer tissue observed by the methods of RFDD-PCR and proteomics. *Breast cancer research and treatment*. 2007;104(1):21-30.
90. Yamada H, Yanagisawa K, Tokumaru S, Taguchi A, Nimura Y, Osada H, et al. Detailed characterization of a homozygously deleted region corresponding to a candidate tumor suppressor locus at 21q11-21 in human lung cancer. *Genes, Chromosomes and Cancer*. 2008;47(9):810-8.
91. Li J, Tan Q, Yan M, Liu L, Lin H, Zhao F, et al. miRNA-200c inhibits invasion and metastasis of human non-small cell lung cancer by directly targeting ubiquitin specific peptidase 25. *Molecular cancer*. 2014;13(1):166.
92. Xu D, Liu J, Fu T, Shan B, Qian L, Pan L, et al. USP25 regulates Wnt signaling by controlling the stability of tankyrases. *Genes & development*. 2017.
93. Habtemichael EN, Li DT, Alcázar-Román A, Westergaard XO, Li M, Petersen MC, et al. Usp25m protease regulates ubiquitin-like processing of TUG proteins to control GLUT4 glucose transporter translocation in adipocytes. *Journal of Biological Chemistry*. 2018:jbc. RA118. 003021.
94. Zhen Y, Knobel PA, Stracker TH, Reverter D. Regulation of USP28 de-ubiquitinating activity by SUMO conjugation. *Journal of Biological Chemistry*. 2014:jbc. M114. 601849.
95. Komander D. The emerging complexity of protein ubiquitination. *Biochemical Society Transactions*. 2009;37(5):937-53.
96. van der Veen AG, Ploegh HL. Ubiquitin-like proteins. *Annual review of biochemistry*. 2012;81:323-57.
97. Streich Jr FC, Lima CD. Structural and functional insights to ubiquitin-like protein conjugation. *Annual review of biophysics*. 2014;43:357-79.
98. Cappadocia L, Lima CD. Ubiquitin-like protein conjugation: structures, chemistry, and mechanism. *Chemical reviews*. 2017;118(3):889-918.
99. Bernier-Villamor V, Sampson DA, Matunis MJ, Lima CD. Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell*. 2002;108(3):345-56.
100. Gill G. SUMO and ubiquitin in the nucleus: different functions, similar

mechanisms? *Genes & development*. 2004;18(17):2046-59.

101. Johnson ES. Protein modification by SUMO. *Annual review of biochemistry*. 2004;73(1):355-82.
102. Flotho A, Melchior F. Sumoylation: a regulatory protein modification in health and disease. *Annual review of biochemistry*. 2013;82:357-85.
103. Maupin-Furlow JA. Prokaryotic ubiquitin-like protein modification. *Annual review of microbiology*. 2014;68:155-75.
104. Bohren KM, Nadkarni V, Song JH, Gabbay KH, Owerbach D. A M55V polymorphism in a novel SUMO gene (SUMO-4) differentially activates heat shock transcription factors and is associated with susceptibility to type I diabetes mellitus. *Journal of Biological Chemistry*. 2004;279(26):27233-8.
105. Guo D, Li M, Zhang Y, Yang P, Eckenrode S, Hopkins D, et al. A functional variant of SUMO4, a new IκBα modifier, is associated with type 1 diabetes. *Nature genetics*. 2004;36(8):837.
106. Novatchkova M, Budhiraja R, Coupland G, Eisenhaber F, Bachmair A. SUMO conjugation in plants. *Planta*. 2004;220(1):1-8.
107. Hammoudi V, Vlachakis G, Schranz ME, van den Burg HA. Whole-genome duplications followed by tandem duplications drive diversification of the protein modifier SUMO in Angiosperms. *New Phytologist*. 2016;211(1):172-85.
108. Saracco SA, Miller MJ, Kurepa J, Vierstra RD. Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. *Plant physiology*. 2007;145(1):119-34.
109. Tomanov K, Ziba I, Bachmair A. SUMO chain formation by plant enzymes. *Plant Proteostasis: Springer*; 2016. p. 97-105.
110. Castaño-Miquel L, Seguí J, Lois LM. Distinctive properties of Arabidopsis SUMO paralogues support the in vivo predominant role of AtSUMO1/2 isoforms. *Biochemical Journal*. 2011;436(3):581-90.
111. Benlloch R, Lois LM. Sumoylation in plants: mechanistic insights and its role in drought stress. *Journal of experimental botany*. 2018;69(19):4539-54.
112. Lois LM, Lima CD. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *The EMBO journal*. 2005;24(3):439-51.
113. Huang DT, Paydar A, Zhuang M, Waddell MB, Holton JM, Schulman BA.

Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. *Molecular cell*. 2005;17(3):341-50.

114. Olsen SK, Lima CD. Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. *Molecular cell*. 2013;49(5):884-96.
115. Huang DT, Hunt HW, Zhuang M, Ohi MD, Holton JM, Schulman BA. Basis for a ubiquitin-like protein thioester switch toggling E1–E2 affinity. *Nature*. 2007;445(7126):394.
116. Walden H, Podgorski MS, Huang DT, Miller DW, Howard RJ, Minor DL, et al. The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. *Molecular cell*. 2003;12(6):1427-37.
117. Tokgöz Z, Siepmann TJ, Streich FC, Kumar B, Klein JM, Haas AL. E1-E2 interactions in the ubiquitin and Nedd8 ligation pathways. *Journal of Biological Chemistry.* 2011:jbc. M111. 294975.
118. Huang DT, Ayrault O, Hunt HW, Taherbhoy AM, Duda DM, Scott DC, et al. E2-RING expansion of the NEDD8 cascade confers specificity to cullin modification. *Molecular cell*. 2009;33(4):483-95.
119. Wang J, Taherbhoy AM, Hunt HW, Seyedin SN, Miller DW, Miller DJ, et al. Crystal structure of UBA2ufd-Ubc9: insights into E1-E2 interactions in sumo pathways. *PloS one*. 2010;5(12):e15805.
120. Wang J, Lee B, Cai S, Fukui L, Hu W, Chen Y. Conformational transition associated with E1-E2 interaction in SUMO modifications. *Journal of Biological Chemistry*. 2009:jbc. M109. 000257.
121. Elgin ES, Sökmen N, Peterson FC, Volkman BF, Dağ Ç, Haas AL. E2-binding surface on Uba3 β-grasp domain undergoes a conformational transition. *Proteins: Structure, Function, and Bioinformatics*. 2012;80(10):2482-7.
122. Reiter K, Mukhopadhyay D, Zhang H, Boucher LE, Kumar N, Bosch J, et al. Identification of biochemically distinct properties of the SUMO conjugation pathway in Plasmodium falciparum. *Journal of Biological Chemistry*. 2013:jbc. M113. 498410.
123. 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. *Molecular cell*. 2007;27(2):228-37.

124. Lee I, Schindelin H. Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell*. 2008;134(2):268-78.
125. Durfee LA, Kelley ML, Huibregtse JM. The basis for selective E1-E2 interactions in the ISG15 conjugation system. *Journal of Biological Chemistry*. 2008;283(35):23895-902.
126. Sampson DA, Wang M, Matunis MJ. The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *Journal of Biological Chemistry*. 2001;276(24):21664-9.

Research objectives

1. Objectives of USP25 research

SUMOylation and ubiquitination on the N-terminal domain of USP25 have opposite effects: ubiquitination activates while SUMOylation inhibits the USP25 activity. Consistent with this, our previous results also revealed that SUMOylation at K99 of USP28 inhibited its ubiquitin chain activity. Moreover, although the basic elements of the catalytic domain of USP25 and USP28 are conserved, there is a large domain (175 residues) inserted within the catalytic core. Thus, structure insights are required to understand how these PTMs affect the catalytic activity of USP25 and USP28, as well as to understand the function of the large unknown insertion domain. We will conduct crystallization trials and biochemical assays to study these two proteins.

1.1 Design and construction of several USP25 and USP28 expression plasmids.
1.2 Expression and purification of different USP25 and USP28 constructs.
1.3 Crystallization trials and data collection by synchrotron radiation of USP25 or USP28 crystals.
1.4 Post-crystallization treatments to improve the diffraction quality of crystals.
1.5 Molecular replacement to solve the phase problem or if needed, use of experimental phasing methods such as SAD and MAD.
1.6 Analysis of crystal structure and characterization of the new structural motifs.
1.7 Biochemical and mutagenesis in vitro analysis to check the new functional features of USP25 or USP28.
1.8 Culture cell analysis to confirm the new structural and functional features in vivo.

2. Objectives of SUMO E1-E2 interaction research

The E1 UFD domain plays a major role in the E1-E2 protein-protein interaction. Despite the high homology between E2-conjugating proteins, the E1-E2 interaction provides specificity for SUMO E1 recruitment of the cognate E2 across species. Unlike E2, the sequence variation between the E1 UFD domains is quite significant, especially in the binding region to the E2 enzymes. Thus, the structural insights of how E1 UFD-E2 interaction interface provides the specificity across species is a main goal of the thesis. To investigate this, we will conduct crystallization trials, as well as and biochemical assays to study the E1 UFD-E2 complexes in human and *A. thaliana*.

2.1 Expression and purification of human and *A. thaliana* E1 UFD and UFDC (UFD with the C-terminal extension), as well as human E2 Ubc9 and *A. thaliana* E2 SCE1.

2.2 Purifications of complexes of human UFD-Ubc9 and UFDC-Ubc9, as well as *A. thaliana* UFD-SCE1 and UFDC-SCE1.

2.3 Crystallization trials and diffraction data collection by synchrotron radiation.

2.4 Post-crystallization treatments to improve the diffraction quality of crystals.

2.5 Solve the phase problem using molecular replacement, or by experimental phasing methods such as SAD or MAD.

2.6 Structure comparison of SUMO E1-E2 interaction interfaces in yeast, human, and *A. thaliana*.

2.7 Evolution analysis of the SUMO E1 UFD domain across species.

2.8 In vitro biochemical analysis of human and *A. thaliana* SUMO conjugation using RanGAP1 as a substrate to study the UFD-E2 interface residues.

2.9 Analysis of the different specificities provided by the E2-UFD domain interface in the SUMO pathway between human, yeast and *A. thaliana*.

Chapter 1

A quaternary tetramer assembly inhibits the deubiquitinating activity of USP25

Bing Liu[1], Marta Sureda-Gómez[2], Yang Zhen[1], Virginia Amador[2] & David Reverter[1*]

[1] Institut de Biotecnologia i de Biomedicina (IBB) and Dept. de Bioquímica i Biologia Molecular, Serra Hunter Fellow, Universitat Autònoma de Barcelona, Bellaterra, Spain.

[2] Institut de Investigacions Biomèdiques Agustí Pi i Sunyer (IDIBAPS). Barcelona, Spain.

Correspondence: david.reverter@uab.cat

Abstract

USP25 deubiquitinating enzyme is a key member of the ubiquitin system, which is essential for the control of the correct protein homeostasis in the cell. USP25 acts as a positive regulator of the Wnt/β-catenin signaling by promoting the deubiquitination and stabilization of tankyrases. USP25 sequence is characterized by the presence of a long insertion in the middle of the “conserved” catalytic domain. The crystal structure of USP25 displays an unexpected homotetrameric quaternary assembly that is directly involved in the inhibition of its enzymatic activity. The tetramer is assembled by the association of two dimers and includes contacts between the coiled-coil insertion domain and the ubiquitin-binding pocket at the catalytic domain, revealing a distinctive autoinhibitory mechanism. Biochemical and kinetic assays with dimer, tetramer and truncation constructs of USP25 support this mechanism, displaying in all cases a higher catalytic activity in the dimer assembly. Moreover, the strong stabilization of tankyrases in cultured cells by ectopic expression of a constitutive dimer form of USP25 verifies the biological relevance of this tetramerization/inhibition mechanism.

Introduction

The ubiquitin system is a universal means of protein regulation which controls a wide range of cellular processes [1-3]. The conjugation of ubiquitin to target proteins is conducted via a conserved multistep enzymatic cascade through E1, E2, and E3 enzymes, resulting the formation of an isopeptide bond between the C-terminus of ubiquitin and the lysine on the target protein [4]. Reversely, deubiquitinating enzymes (DUBs) can remove ubiquitin by catalyzing the hydrolysis of the isopeptide bond. Therefore, ubiquitin conjugating and deconjugating are balanced and tightly regulated by E3 ligases and DUBs. There are more than 100 DUBs encoded in the human genome [5]. Based on the structure studies, their isopeptidase activities are conferred by distinct catalytic domains [5-7]. Recently, the identification of a new family MINDY extends the DUBs to six distinct families [8]. In human genome, ubiquitin-specific proteases (USPs) constitute the largest family of DUBs counting more than 50 members and playing important roles in a wide variety of cellular processes [5-7,9].

Structural studies show USPs have a common conserved fold consisting of three subdomains known as *Palm*, *Thumb*, and *Fingers*, in which the active site cysteine is located between the *Palm* and *Thumb* while the *Fingers* grip the “distal” ubiquitin [10]. However, their catalytic activities are usually modulated through their different additional domains. For example, a C-terminal 19 amino acids peptide binds the activation cleft in the catalytic domain and stabilizes the catalytically competent conformation, thus, enhances the activity of USP7 [11]. This activation can be enhanced allosterically by the metabolic enzyme GMPS [12]. The binding of USP4 N-terminal DUSP-Ubl domain promotes a change of a switching loop near the active site, hence enhances ubiquitin dissociation and makes it achieve full catalytic activity [13].

USP25 shows activity in hydrolyzing K48 and K63 linked polyubiquitin chains [14,15] and is a target for different posttranslational modifications including phosphorylation, SUMOylation and ubiquitination. The SYK non-receptor tyrosine kinase can specifically phosphorylate USP25 and decrease its cellular levels [16]. USP25 enzymatic activity is suppressed after vaccinia-related kinase2 mediated phosphorylation at Thr680, Thr727 and Ser745, which controls the stability of the eukaryotic chaperonin TRIC [17]. SUMO and ubiquitin regulate the catalytic activity of

USP25 by conjugation at Lys99 with opposite effects: ubiquitination activates while SUMOylation inhibits the USP25 activity [15,18].

USP25 has been associated with inflammation, immune response and cancer. For example, it is associated with Endoplasmic Reticulum-associated degradation [19], and acute ER stress regulates amyloid precursor protein processing through ubiquitin-dependent degradation by USP25 [20]. USP25 is a negative regulator of IL-17-mediated signaling and inflammation through removal of ubiquitination in TRAF3, TRAF5 and TRAF6 and can regulate TLR4-dependent innate immune responses [21,22,23]. Moreover the type I interferon-IRF7 can activate the expression of USP25 gene, which is essential for innate immune signaling [24]. Several evidences point to the involvement of USP25 in cancer. USP25 gene is found greater than threefold overexpression in human breast cancer tissue [25], and further study shows it as a putative tumor suppressor in human lung cancer [26]. Another study in human non-small cell lung cancer reveals miR-200c inhibits invasion and metastasis by directly targeting USP25 and reducing the expression level [27]. Recently, it has been shown that USP25 directly interacts with tankyrases through its C-terminal tail and promotes their deubiquitination and stabilization, thus regulating Wnt/β-catenin signaling pathway, making an important impact in cell proliferation and human cancer development [28].

Despite these roles in cellular pathways, structural insights of USP25 activity regulation are still not clear. The N-terminal domain of USP25 contains one ubiquitin associated domain (UBA), one SUMO interacting motif (SIM) and two ubiquitin interacting motifs (UIM1 and UIM2) [29]. SUMO-modified USP25 maintains the ability to hydrolyze Ub-AMC, but exhibits lower activity during polyubiquitin chain degradation, which indicates that SUMO modification affects the recruitment of ubiquitin but not the catalytic domain activity [15]. Although the basic elements of the catalytic domain of USP25 are conserved, there is a large domain (175 residues) inserted within the catalytic core. Also, previous studies show USP25 can form homodimers or oligomers in vivo and the catalytic domain (between residues 153 to 679) is relevant for this dimerization or oligomerization [15,18]. Finally, the C-terminal domain of USP25 is involved in phosphorylation and substrate recognition [16,17], directly interacting with tankyrases to promote their deubiquitination and stabilization [28].

In this study, we present the crystal structure of USP25 catalytic domain. Strikingly, the USP25 structure displays a homotetramer quaternary assembly and the catalytic domain can be divided in three differentiated subdomains: a general palm-like catalytic domain (USP-like), a long coiled-coil (LCC) and an “inhibitory loop” (IL). Whereas the USP-like catalytic domain is similar with the other USPs, the LCC and IL domains are unique to USP25. In our USP25 structure, each IL-loop inserts into the catalytic domain from another molecule of the tetramer, thus preventing substrate binding and thereby impairs USP25 activity. Moreover, the biological relevance of this tetramerization/inhibition mechanism of USP25 has been confirmed in a cellular context with the stabilization of tankyrases, which directly regulate the Wnt/β-catenin signaling pathway. This tetramerization/inhibition mechanism of USP25 has not been described for other deubiquitinases and probably unveils a paradigmatic type of regulation in the USP family [30].

Results

USP25 recombinant expression produces two oligomer forms

Human USP25 is a modular protein composed by three domains: a N-terminal domain NTD (Met1 to Tyr159), a central USP-like domain (Asp160 to Glu714) and a C-terminal domain CTD (Lys715 to Arg1055) (**Fig. 1a** and **Supplementary Fig. 1**). Since the C-terminal domain does not directly participate in the catalytic activity and the expression levels of the full-length USP25 in *E. coli* were low [14], we produced a USP25 construct (USP25NCD) including the N-terminal and the USP-like catalytic domains (from Gln18 to Glu714).

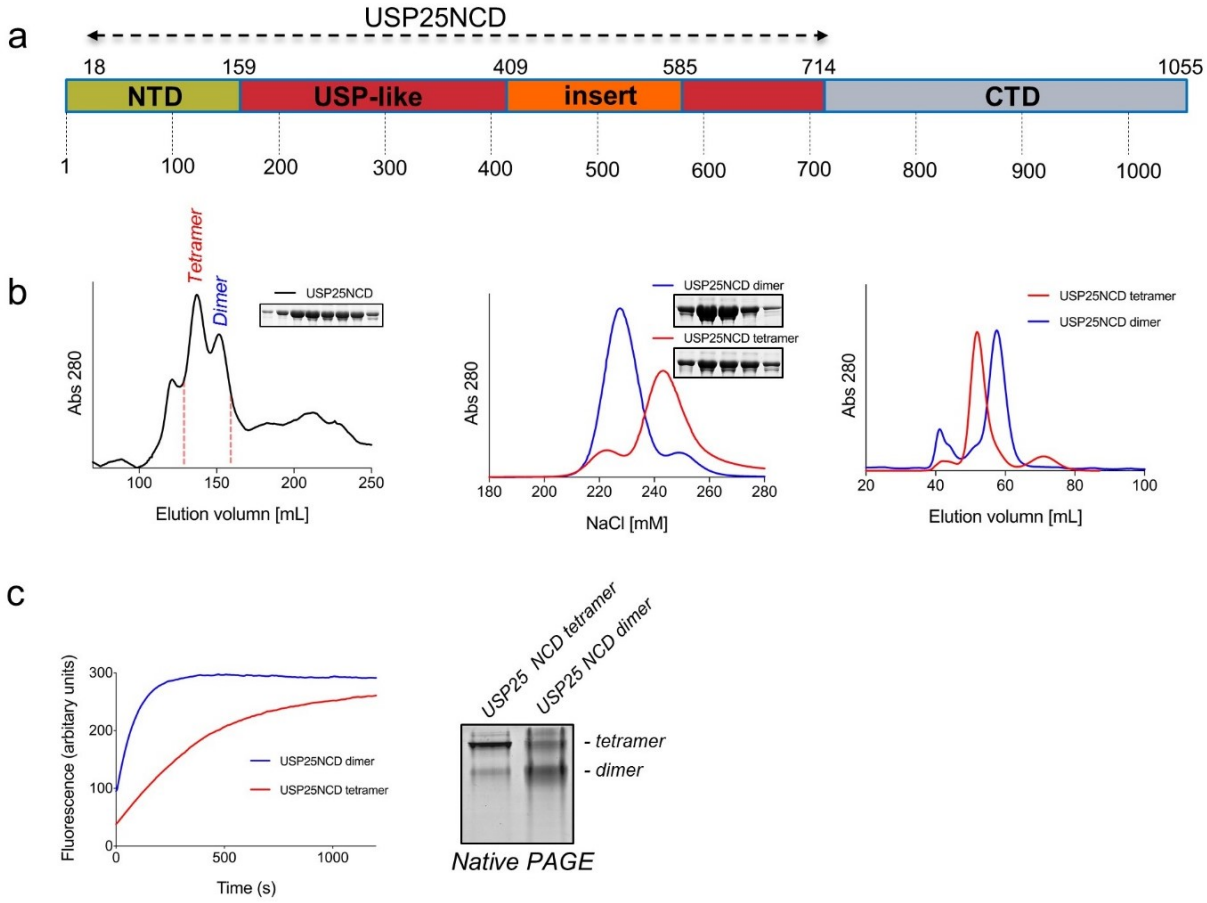


Figure 1. Recombinant USP25 is produced as a tetramer or a dimer in *E. coli*
(a) Scheme of the USP25 protein domains. The construct for crystallization is from residues 18 to 714 (USP25NCD). (b) *Left*, first preparative gel filtration purification of USP25 in Superdex 200 column, inset shows the SDS-PAGE of the indicated fractions. *Middle*, anionic-exchange purification of the two peak fractions of the previous gel filtration, inset shows the SDS-PAGE of the dimer and tetramer fractions. *Right*, analytical gel filtration purification of the dimer and tetramer fractions of USP25 in Superdex 16 column. (c) *Left*, plot of the deubiquitinating activity assays of dimer and tetramer with Ub-AMC as a substrate. *Right*, non-denaturing native PAGE of the dimer and tetramer fractions of USP25NCD.

A particular feature of the recombinant expression of human USP25 in *E. coli* is the presence two different stable oligomeric states, dimer and tetramer, which are stable after running two different size-exclusion and one ionic exchange chromatographies (**Fig. 1b**). USP25 oligomerization has also been reported by two groups in human cell lines using a dual-tag expression system [15,18], but the type of the quaternary assembly could not be determined. Mass spectroscopy analysis of recombinant USP25 did not reveal any particular feature, such as post-translational modifications, to explain the presence of these stable oligomer fractions.

However, activity assays using the standard fluorescence substrate for deubiquitinating activity, Ub-AMC, revealed marked differences between dimer and tetramer, being the dimer assembly of USP25 substantially more active than the tetramer (**Fig. 1c**). These activity differences are maintained despite the presence of cross-contamination between oligomers, as observed in non-denaturing PAGE analysis (**Fig. 1c**). To get more insight into these activities we conducted crystallization trials with these two oligomeric assemblies of USP25.

The crystal structure of USP25

Interestingly, only the tetramer assembly of USP25 gave rise to single crystals. Initial diffraction of the USP25 crystals was poor, never beyond 6 Å resolution, however, after conducting a lysine methylation protocol [31], diffraction was notably improved, and thus finally collecting a 3,28 Å resolution dataset. Apparently, the lysine methylation reaction did not affect the deubiquitinating activity of USP25 (**Supplementary Fig. 2b, 2c**). The crystals belonged to the I422 space group, and contained one molecule per asymmetric unit (**Supplemental Table 1**). The USP25 structure was solved by SAD using mercurial derivatives, three of which were located bound to three cysteine residues (one corresponding to the active site cysteine). After several steps of modeling and refinement using the structure of USP7 as a guide, the electron density maps clearly showed the elements of the catalytic USP-like domain (general “palm-like” structure), plus a long helical coiled-coil domain emanating from the middle of the USP-like domain (**Fig. 2a,b,c**), which corresponded to the long sequence insertion splitting the catalytic domain. The final electron density map model includes most of the chain for the catalytic domain (Tyr159 to Ala708) with a notable absence of the whole N-terminal domain (Gly18

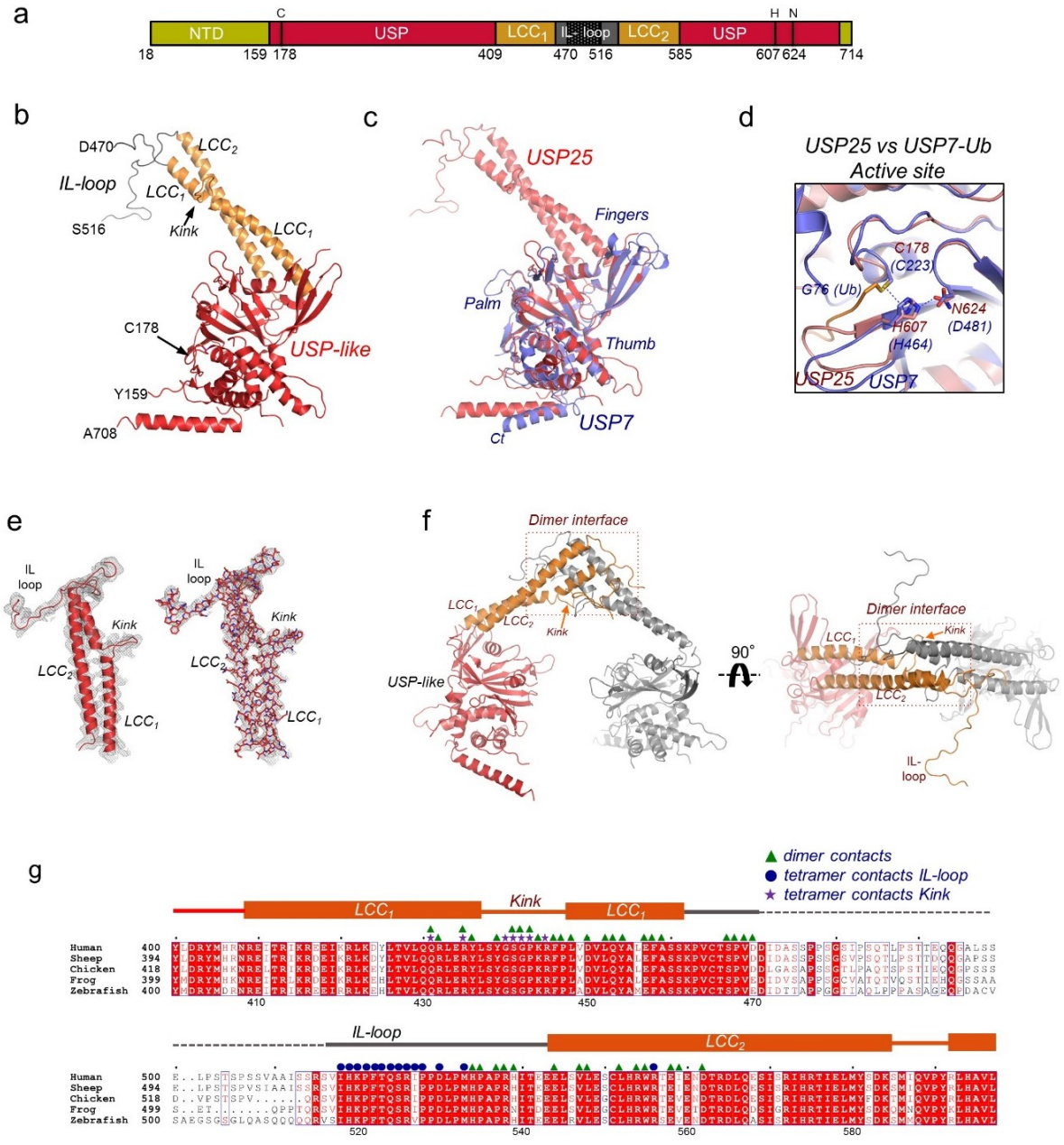


Figure 2. The crystal structure of USP25
(a) Scheme of the domain composition of the crystal USP25 construct. NTD (18-158), and the non-conserved sequence (dashed grey color, 471-515) of IL-loop are not observed in the final structure. (b) Ribbon representation of the different domains of the USP25 monomer. Red color shows the "conserved" USP-like domain, orange color the coiled-coil domain (LCC) and black color the "Inhibitory loop" domain (IL-loop). Active site cysteine and domain boundaries are labeled. (c) Structural superposition between USP25 (red) and USP7 (blue) structures (PDB code 5JTJ). "Palm"-like domains are labeled. (d) Close-up view of the superimposition of active sites of USP25 and USP7. USP25 and USP7-Ubiquitin aldehyde are shown in orange and blue, respectively. Catalytic triad residues from USP25 are labeled and shown in stick representation, whereas the corresponding residues from USP7 are shown in parentheses. Hydrogen bonds are represented by dashed lines. (e) Two representations of the electron-density maps of the LLC and IL-loop domains of USP25. (f) Ribbon representation of the two monomers composing the dimer structure observed in the crystal of USP25 in two different orientations. (g) Sequence alignment of the LCC-IL-loop domain of USP25 across species. Green triangles indicate contacts in the dimer interface. Blue circles and purple stars indicate tetramer contacts by the IL-loop or the "kink" region, respectively. All sequences are aligned online with Clustal Omega and formatted using ESPript [32].

to Pro158), and of some disordered loops connecting secondary structure elements in the USP-like domain (Asn205 to Lys214; Asn256 to Gln260; Ile346 to Ser354; Asp471 to Arg515; Thr671 to Gly677).

Structure of the USP-like catalytic domain of USP25

The USP-like domain of USP25 contains all characteristic elements of the USP family, which was initially compared to a "right hand" that could entrap ubiquitin [10], including "*Palm*", "*Fingers*" and "*Thumb*" subdomains (**Fig. 2c**). The distances between the catalytic triad residues indicate that the active site might be preformed in the absence of the ubiquitin substrate, with distances ranging around 3.2 Å between the sulfur of Cys178 and the imidazole ring of His607, and 3.0 Å between His607 and Asn624 (**Fig. 2d and Supplementary Fig. 3**), which is comparable to the USP7-Ub aldehyde complex structure (rmsd 1,67 Å for 268 aligned residues) [10]. The preformed active site in USP25 differs from other apo structures of USPs, in which the catalytic triad residues are too far for catalysis and binding of the "distal" ubiquitin substrate is required to rearrange the catalytic triad to an active conformation (ex. USP7, USP14, USP18) [11,33,34].

Unexpectedly, the long sequence insertion of the USP25 catalytic domain (between Asn408 and Met586) forms a coiled coil structure (LCC) composed by two long α-helices, which we named helix LCC_1 (from Arg409 to Ser460) and helix LCC_2 (from Leu540 to Ser585) (**Fig. 2b,e**). It is interesting to observe a break in the α-helix LCC_1 produced by a "kink" between Ser438 and Pro447 (**Fig. 2e**), which is involved in the assembly of the tetramer structure (as will be explained later). The two α-helices of the coiled-coil are connected by a long loop, which we named "inhibitory loop" (IL-loop) based on its role in the regulation of the enzymatic activity of USP25. The IL-loop is composed by a non-conserved sequence (**Fig. 2g**), absent in the USP25 crystal structure (from Asp471 to Arg515), and by a conserved sequence, which can be clearly observed in the electron density maps engaging contacts with the catalytic domain of a different molecule of the crystal. This interaction is essential for the tetramer assembly.

The tetramer assembly of USP25 is formed by the interaction of two dimers

The structure of USP25 reveals the presence of a homotetramer quaternary assembly composed by the interaction of four different USP25 molecules from the crystal lattice (**Fig. 3**). PISA server analysis also predicts that the tetramer structure of USP25 (A_4), with a buried surface area of 23730 $Å^2$, is formed by the assembly of two homodimers ($2xA_2$) (**Fig. 3**). Based on the high number of contacts, the dimer assembly might represent a minimal stable oligomer state of USP25. The structure of USP25 can thus explain the presence of the tetramer and dimer fractions observed during the purification by gel filtration (**Fig. 1**).

Each homodimer composing the USP25 tetramer is comprised by an extended contact interface that includes approximately half of the LCC coiled-coil domain, forming a four-helix bundle motif, which would act as a dimerization hub between two USP-like catalytic domains, resembling a "pair of cherries" with the knot at the end of their stems (**Fig. 2f** and **3**). PISA server analysis reveals a 2110 $Å^2$ dimer interface involving a total of 55 residues. Most residues are hydrophobic (ex. Leu448, Leu452, Phe458, Leu548, Trp555) (**Supplementary Fig. 4a**), but the dimer interface also includes 20 and 12 hydrogen bond and salt bridge contacts, respectively (see contact list in **Supplementary Fig. 5**). All contacts are highly conserved across species (**Fig. 2g** and **Supplementary Fig. 1a**). The dimer structure exposes two USP-like catalytic domains to the solvent.

The tetramer assembly depends on IL-loop contacts to the catalytic domain

The LCC-IL-loop insertion domain also participates in the tetramer interface, but in contrast to the dimer, the interaction occurs with the catalytic domain of the opposite dimer (**Fig. 3**). Two contact areas can be distinguished in the tetramer interface: between the extended IL-loop which is deeply inserted into the S1 ubiquitin-binding surface of the catalytic domain (**Fig. 4a,c** in green and **Supplementary Fig. 4c**); and between the LCC1 "*kink*" motif with residues from the loop connecting β9-β10 strands of the catalytic domain (**Fig. 4a** in blue and **Supplementary Fig. 4b**). So, the interaction of each IL-loop with the catalytic domain of the opposite dimer is essential for the formation of a stable tetramer assembly.

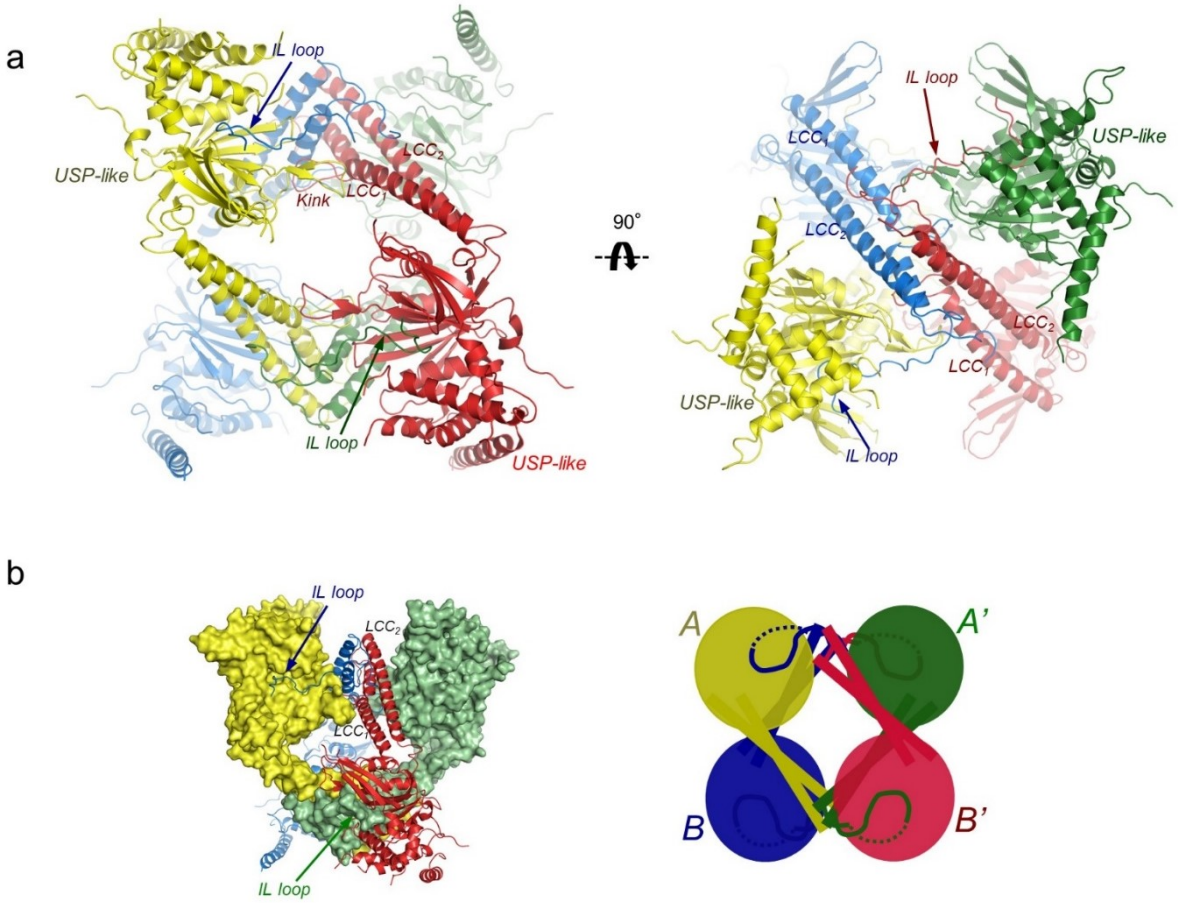


Figure 3. The USP25 structure is a homotetramer, formed by the assembly of two homodimers. (a) Ribbon representation of the USP25 tetramer. Each subunit is represented with a different color. LLC and IL-loop domains are labeled. *Right*, different orientation of the USP25 tetramer. (b) Mixed surface and ribbon representation of the homotetramer structure of USP25. *Right*, cartoon image of the tetramer assembly formed by the assembly of two homodimers (A, A' homodimer and B, B' homodimer).

The IL-loop was unambiguously observed in the electron density maps in contact with the S1 ubiquitin-binding pocket of the USP-like catalytic domain (**Fig. 4b**). Structural comparison with the USP7-Ubiquitin complex (PDB 5JTJ) and with the CYLD-diubiquitin K63 complex (PDB 3WXG) [35] indicate that the IL-loop and the "distal" ubiquitin share a similar binding surface (**Fig. 4b** and **Supplementary Fig. 3**), thus preventing substrate binding in the tetramer assembly. Such "tetramer-autoinhibition" mechanism would explain the different activities observed between the purified tetramer and dimer of USP25 (**Fig. 1c**).

The residues of the IL-loop involved in the tetramer interface, from Ile518 to Leu531, are mostly conserved across species in contrast to the rest of the IL-loop,

not observed in the electron density maps (**Fig. 2g**). Among the high number of contacts conducted by the IL-loop, Pro521 and Phe522 are completely buried in a deep crevice formed between the α-helices α6, α10 and the central β-sheet of the USP-like catalytic domain, establishing several hydrophobic interactions: Val183 and Leu187 from α-helix α6; Leu271 from α-helix α10; Phe299 and Tyr300 from α-helix α11; and Met653 from β-sheet β16 of the USP25 catalytic domain (**Fig. 4c,d** and **Supplementary Fig. 4c** and **5**).

To assess the role of this interface, several USP25 point mutants were generated in the IL-loop and in the S1 ubiquitin-binding surface (**Fig. 4e**, and **Supplemental Fig. 6**). P521S and F522G IL-loop point mutants, which are involved in hydrophobic contacts within the pocket, disrupt the tetramer assembly and their deubiquitinating activity is comparable to the USP25 dimer. Similar results are observed for P528G, P535L and S525P IL-loop point mutants, but in this instance the distorted conformation of the IL-loop probably perturbs the binding affinity. On the USP-like catalytic domain surface, both E373A and Q322A point mutants of the S1 ubiquitin-pocket disrupt three hydrogen bonds with the IL-loop and compromise the tetramer assembly, in particular E373A displays similar deubiquitinating activities as USP25 dimer. Finally, C651F and L271W point mutants were intended to lock the USP-like binding cleft, preventing the IL-loop binding. Whereas this was partially achieved by C651F, the L271W point mutant unexpectedly stabilized the tetramer with a consequent loss of deubiquitinating activity (**Fig. 4e**, and **Supplemental Fig. 6**).

Finally, to check a potential competition between the IL-loop and the ubiquitin substrate, USP25 deubiquitinating activities were measured in the presence of a peptide derived from the IL-loop binding sequence (HKPFTQSRIPPD) (**Supplemental Fig. 6e)**. The absence of inhibition produced by the IL-loop peptide suggests a model in which the structural context in the tetramer assembly is required for the proper binding and thus inhibition.

Overall, this reciprocal point mutant analysis of the binding interface verifies the role of the IL-loop in the tetramer assembly and in the regulation of the deubiquitinating activity.

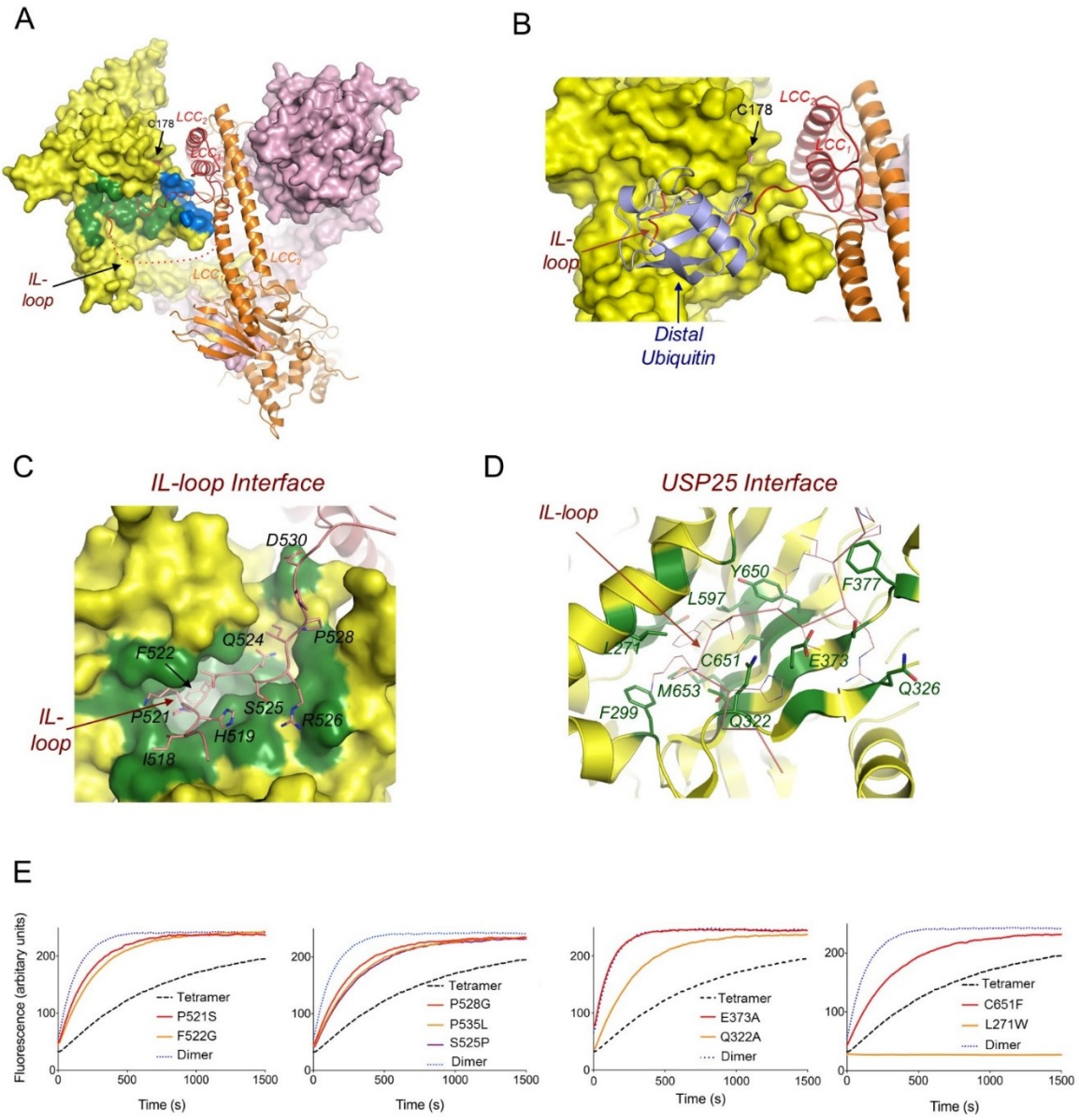


Figure 4. IL-loop contacts the catalytic domain of USP25 in the tetramer assembly
(a) Surface and cartoon representation of interface between the IL-loop domain (green), "kink" region (blue) and the catalytic domain in the tetramer assembly. (b) Structure overlapping of USP25 and USP7 structures (PDB code 5JTJ), depicting the S1 ubiquitin-binding domain in contact with either ubiquitin in complex with USP7 structure (light blue) or with the IL-loop domain in the USP25 tetramer (red ribbon). (c) Surface representation of the S1 ubiquitin-binding pocket (green) of the USP25 catalytic domain in complex with the IL-loop insertion with labeled residues depicted in stick representation. (d) Cartoon representation of the IL-loop binding interface with labeled residues depicted in stick representation (green). IL-loop is depicted as a thin line (red). (e) Deubiquitinating activity assays with Ub-AMC substrate of either USP25 wild type and IL-loop point mutants: purified tetramer and dimer, P521S, F522G, S525P, P528G and P535L (first and second panel); or USP25 wild type and S1 ubiquitin-binding residues point mutants: E373A, Q322A, L271W and C651F (third and fourth panel).

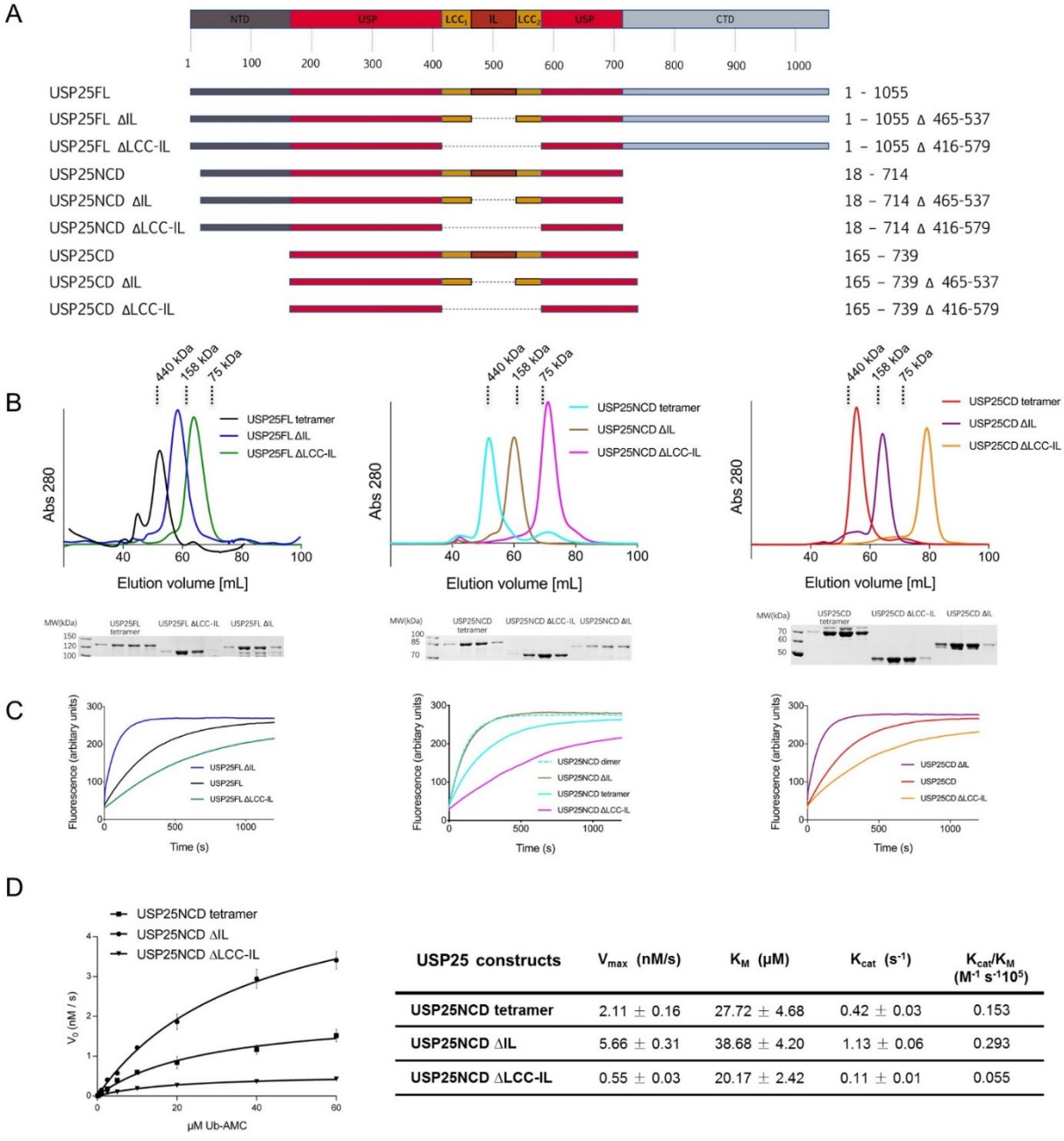
USP25 truncation analysis confirm the formation of dimer and tetramer

The role of the LCC-IL-loop domain in oligomerization was assessed by two truncation constructs of USP25: IL-loop truncation (*ΔIL*, deletion from residue Cys465 to residue Pro537) and LLC-IL-loop truncation (*ΔLCC-IL*, deletion from residue Arg416 to residue Leu579). These different constructs were checked in the context of the full-length of USP25 (*USP25 FL*), the C-terminal domain deletion (*USP25 NCD*) and the N- and C-terminal domains deletion (*USP25 CD*) (see scheme in **Fig. 5a**). All USP25 truncations did not display expression problems and were eluted as single peaks during their purification by gel filtration chromatography.

All USP25 truncation were run in parallel through an analytical size exclusion chromatography under identical experimental conditions, which permitted the correlation between the expected molecular weight and the gel filtration elution times (**Fig. 5b**). Interestingly, full-length *USP25FL*, *USP25NCD* and *USP25CD* elution times were compatible with the presence of a tetramer structure, whereas deletion of the IL-loop (*ΔIL*) and deletion of the LLC-IL-loop (*ΔLCC-IL*) resulted in elution times compatible with dimer and monomer structures, respectively (**Fig. 5b**). Additionally, dynamic light scattering analysis of the USP25 truncation constructs also displayed a good correlation between oligomers and their theoretical molecular weight (**Supplementary Fig. 7)**. These experiments confirm the role of the LLC coiled-coil domain in the formation of the dimer assembly, and the role of the IL-loop in the formation of the tetramer assembly.

All dimers display higher deubiquitinating activities than tetramers

Interestingly, all dimer constructs (*ΔIL* USP25FL, USP25NCD and USP25CD) displayed a higher hydrolysis of the Ub-AMC substrate in comparison to all tetramer (USP25FL, USP25NCD and USP25CD) and monomer (*ΔLCC-IL* USP25FL, USP25NCD and USP25CD) constructs (**Fig. 5c**). We also compared the purified dimer of USP25 (**Fig. 1**) and the truncation dimer construct, *ΔIL* USP25NCD (**Fig. 5c**), displaying both comparable deubiquitinating activities and similar purification profiles and native PAGE migrations (**Supplementary Fig. 7b**). The higher activity of the dimer over the tetramer was also confirmed using polyubiquitin chains, either K48 or K63-linked, and di-ubiquitin K48-linked substrates (**Supplementary Fig. 8).**



USP25 constructs	V_{max} (nM/s)	K_M (μM)	K_{cat} (s^{-1})	K_{cat}/K_M (M^{-1} $s^{-1}10^5$)
USP25NCD tetramer	2.11 ± 0.16	27.72 ± 4.68	0.42 ± 0.03	0.153
USP25NCD ΔIL	5.66 ± 0.31	38.68 ± 4.20	1.13 ± 0.06	0.293
USP25NCD ΔLCC-IL	0.55 ± 0.03	20.17 ± 2.42	0.11 ± 0.01	0.055

Figure 5. Biochemical and kinetic analysis of the USP25 assemblies.
(a) Cartoon representation of the different truncation constructs of USP25. Left columns indicate the names and the right columns the residue range for each construct. (b) Size exclusion chromatography profiles of the different USP25 truncation constructs run under identical experimental conditions in Superdex 16 column. Below, SDS-PAGE of the fractions of the gel filtration peaks. (c) Deubiquitinating activity assays of the different elution fractions of the gel filtration chromatography using Ub-AMC as a substrate. Activity assays were run in triplicate. (d) *Left*, steady-state "Michaelis-Menten" kinetic plot of the indicated oligomer constructs of USP25, including the purified tetramer, the truncated dimer construct (ΔIL) and the truncated monomer construct (ΔLCC-IL). *Right*, table indicating the kinetic constant values of the USP25 oligomers.

Quantitative activity analyses were also conducted by steady-state "*Michaelis-Menten*" kinetics with the purified tetramer and the "constitutive" truncated dimer and monomer of USP25 (**Fig. 5d**). The higher K_{cat}/K_M values displayed by the dimer (0.293 x 10^5 $M^{-1}s^{-1}$), compared to the tetramer (0.153 x 10^5 $M^{-1}s^{-1}$), supports our proposed model in which the deubiquitinating activity in the USP25 tetramer assembly is partially blocked. It is important to note here the presence of dimer cross-contamination in the purified tetramer sample, as observed in a native gel (see **Supplementary Fig. 7b**), which would increase the real activity of the tetramer sample. Interestingly, two USP25 point mutants that disrupt the tetramer assembly (F522G and E373A), display kinetic values close to the dimer of USP25 (**Supplementary Fig. 6d)**. Our results supports a model in which the presence of dimer (active) or tetramer (inactive) regulates the deubiquitinating activity of USP25 (**Fig. 7**).

Strikingly, the absence of the whole LCC coiled-coil insertion in the monomer truncation, *ΔLCC-IL USP25NCD*, produced a notable reduction of the deubiquitinating activity, displaying a very low K_{cat}/K_M (0.055 x 10^5 $M^{-1}s^{-1}$) (**Fig. 5d**). We attribute this catalytic impairment in the monomer to a wrong conformation of the catalytic domain, which probably requires contacts with the deleted LCC domain for a proper productive form of the protease (**Supplementary Fig. 9c)**. To assess this point, we generated a longer USP25 monomer including half of the LCC domain, named *ΔLCC-IL USP25NCD long (***Supplementary Fig. 9)**, which displays a deubiquitinating activity similar to the USP25 dimer, thus fulfilling all contacts needed for a correct arrangement of the catalytic domain in the monomer.

Ectopic expression of the constitutive USP25 dimer stabilizes tankyrases in cultured cells

We next checked the biological relevance of the tetramerization/inhibition mechanism of USP25 in cultured cells by checking the levels of endogenous tankyrases, which are bona fide USP25 substrates involved in the regulation of the Wnt/β-catenin signaling [28]. Ectopic expression of wild-type and *ΔIL* truncation, either USP25FL or USP25NCD, in HEK293T human cell lines revealed that the *ΔIL* USP25FL construct was able to increase significantly the protein levels of endogenous tankyrases after 48 hours of expression, indicating a higher

deubiquitinating activity of the constitutively dimer form of USP25 (**Fig. 6a**). Interestingly, the dimer assembly of USP25 lacking the C-terminal domain, *ΔIL USP25NCD*, which displayed similar in vitro deubiquitinating activities as USP25FL (**Fig. 5c, d**), did not increase the levels of takyrases, confirming the role of the C-terminal domain of USP25 in the specific interaction with tankyrases [28]. Also, an inactive USP25 active site mutant C178A produced similar levels of tankyrases as the GFP control cultures, highlighting the role of the deubiquinating activity of the ectopically expressed USP25 in the tankyrase stability (**Fig. 6a** and **Supplementary Fig. 10a**). These results were also confirmed by immunoprecipitation assays (**Supplementary Fig. 10b**), which retrieved higher amounts of tankyrases by the dimer (*ΔIL USP25FL*) than by the tetramer (*USP25FL*). As expected, the C-terminal deletion constructs were unable to immunoprecipitate tankyrases (**Supplementary Fig. 10b**). Also, cells treated with cyclohexamide, which inhibits novel protein synthesis, also revealed differences in the stability of tankyrases between dimer (*ΔIL USP25*) and tetramer (*FL USP25)* (**Fig. 6b**)*,* probably due to a higher deubiquitinating activity of the "constitutive" dimer form of USP25.

Finally, ectopic expression of USP25 point mutants of the IL-loop interface, which weakens the tetramer assembly and impairs the deubiquitinating activity (**Fig. 4e**, and **Supplemental Fig. 6**), showed results comparable to the expression of the "constitutive" dimer at different levels (**Fig. 6c**). Only the L271W point mutant, which already showed a notable loss of activity in vitro, exhibited similar tankyrase levels as in the GFP control cultures. Cyclohexamide treatment of USP25 P521S point mutant yielded similar tankyrase stability as the USP25 dimer (**Fig 6d**).

To confirm the ubiquitin-dependent stability of tankyrase [28], cells were treated with cyclohexamide and Bortezomib, which is a potent proteasome inhibitor [36]. Bortezomib-treated cells revealed higher endogenous tankyrase levels in all cultures analyzed, including the GFP control and the poorly active USP25 L271W point mutant (**Supplementary Fig. 10c)**, all confirming the ubiquitin-dependent degradation of tankyrase [28] and the role of the deubiquitinating activity of USP25 for the higher stability of tankyrase.

All experiments support our in vitro characterization of the USP25 structure and emphasize the relevance of the tetramerization/inhibition mechanism of USP25 in a cellular context.

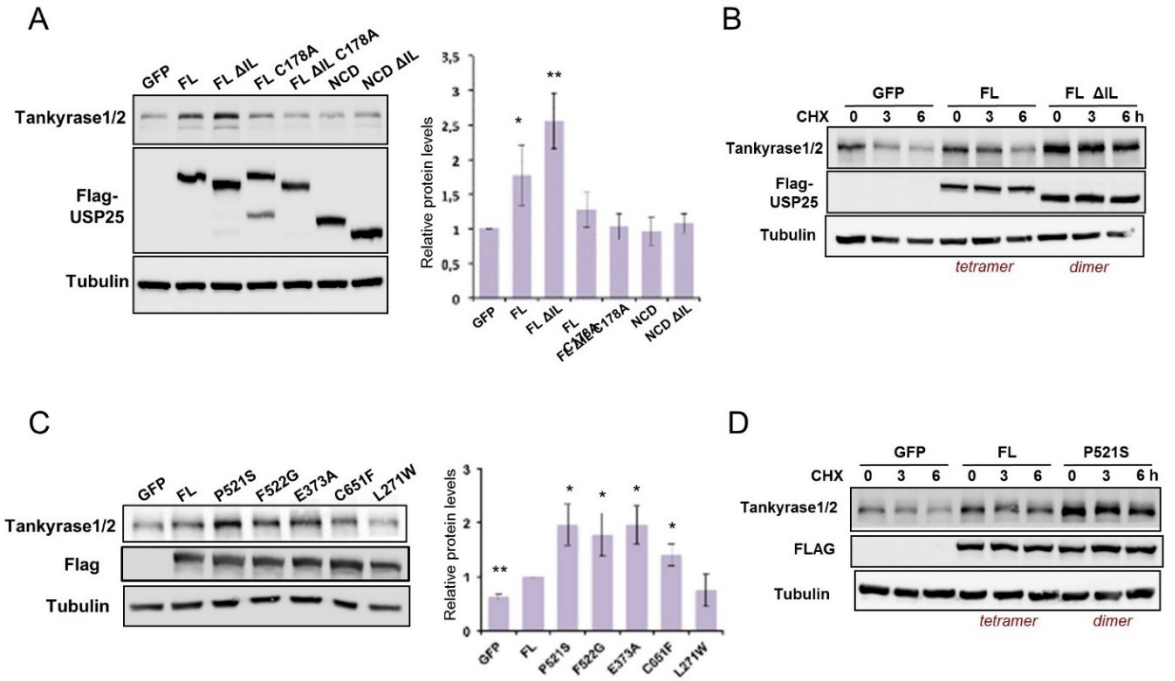


Figure 6. The USP25 dimer stabilizes endogenous tankyrase levels in HEK293T cells.

(a) *Left*, Flag-USP25FL, Flag-USP25FL ΔIL, Flag-USP25FL C178A, Flag-USP25FL ΔIL C178A, Flag-USP25NCD and Flag-USP25NCD ΔIL were transfected in HEK293T cells and the levels of endogenous tankyrases1/2 were analyzed by western-blot. GFP was transfected as a control. *Right*, plot of the quantification of tankyrase1/2 levels relative to tubulin. Data values are mean ± s.e.m and n=3 technical replicates. Significance was measured by a two-tailed unpaired t-test relative to GFP. **P*<0.05, ***P*<0.01. (b) HEK293T cells were transfected with Flag-USP25FL, Flag-USP25FL ΔIL, and GFP and cells were treated with 100 µg/ml of cyclohexamide (CHX) and collected at indicated times for western-blotting. Endogenous tankyrase1/2, Flag-USP25 and tubulin was checked and compared by WB. (c) *Left*, Flag-USP25FL and indicated point mutants were transfected in HEK293T cells and the levels of endogenous tankyrases1/2 were analyzed by western-blot. *Right*, plot of the quantification of tankyrase1/2 levels relative to tubulin. Data values are mean ± s.e.m and at least n=3 technical replicates. Significance was measured by a two-tailed unpaired t-test relative to FL. **P*<0.05, ***P*<0.01. (d) HEK293T cells were transfected with Flag-USP25FL, Flag-USP25FL P521S, and GFP and cells were treated with 100 µg/ml of cyclohexamide (CHX) and collected at indicated times for western-blotting.

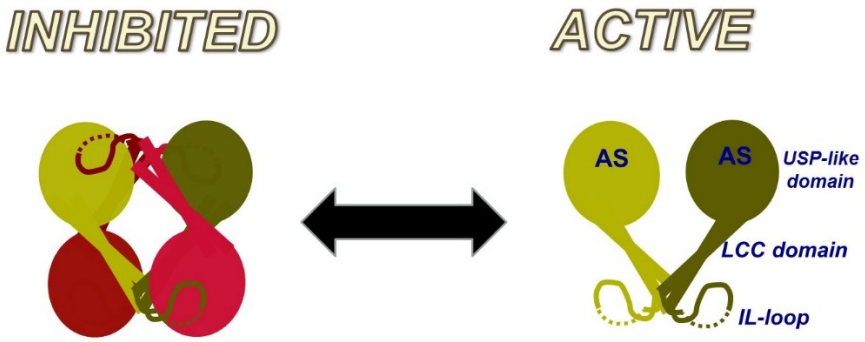


Figure 7. USP25 transition between dimer and tetramer.
Each cartoon color represents a USP25 monomer. The scheme depicts a potential switch between tetramer and dimer assemblies of USP25, or between inhibited and active forms of USP25. AS, active sit.

Discussion and conclusions

Two remarkable features are obtained from the crystal structure of USP25: the presence of two different stable quaternary assemblies, tetramer and dimer; and the implication of these two oligomer states in the deubiquitinating activity of USP25 by a unique autoinhibitory mechanism. This tetramerization-dependent inhibition mechanism of USP25 has never been described for other members of the DUB family and might represent a paradigmatic example of regulation of the deubiquitinating activity, as we demonstrated in vitro and in cultured cells with the stabilization of tankyrases.

Among all described regulatory mechanisms for USPs, such as phosphorylation, allosteric binding of modulators, or localization, only two included comparable "autoregulation" mechanisms. For example: USP7 is activated by the interaction of a conserved C-terminal extension with the catalytic domain, producing a rearrangement of the catalytic active site triad [11,12]; and in USP4 the reaction product is released from the active site pocket thanks to the binding of an N-terminal Ubiquitin-binding domain [13]. However, in USP25 the molecular mechanism is substantially different and the inhibition occurs by the contacts in the tetramer assembly.

USP25 is also regulated by posttranslational modifications on its N-terminal domain, such as inhibition by SUMO conjugation [15]. The N-terminal domain of USP25, disordered in our crystal structure, contains ubiquitin-binding domains that are responsible for the interaction with ubiquitin chains, which can be probably perturbed by SUMO attachment. Our assays indicate that the N-terminal domain is required for an efficient de-conjugation activity of ubiquitin chains, even in the case of the most "active" dimer (USP25CD ΔIL lane in **Supplementary Fig. 8**), in contrast to the Ub-AMC substrate, in which the N-terminal domain is dispensable. Therefore, there are at least two different layers of USP25 regulation at the protein level: binding of ubiquitin chains to the N-terminal domain; and by a switch in the oligomerization state of USP25, described in the present work.

USP25, and its homolog USP28 (51% sequence identity) (**Supplementary Fig. 1b**), are multidomain proteins characterized by the presence of long insertions in the middle of the USP-like catalytic domain, forming long coiled-coil structures

(*LCC-IL-loop)* responsible for quaternary assemblies. This oligomerization domain is unique to USP25 (and USP28) in the USP family, in which no quaternary assemblies have been reported [30]. Moreover, PISA server analysis predicts the formation of the homotetramer structure (A_4) by the assembly of two distinctive homodimer structures (2 x A_2), which we have validated by the characterization of several truncation constructs of USP25 (**Fig. 6a,b**). Interestingly, since the biophysical and enzymatic properties of the purified and truncated dimers are similar, we speculate whether a transition between these two oligomeric assemblies might be relevant in the regulation of the USP25 activity. Higher stability of endogenous tankyrases in HEK293T cells by the ectopic expression of a constitutive dimer (active) over a tetramer (inhibited) supports the biological relevance of this regulatory tetramer/inhibition mechanism of USP25.

This regulatory mechanism depends on the interaction of the IL-loop insertion with the S1 ubiquitin-binding region of the catalytic domain. Remarkably, the IL-loop residues involved in the interaction are highly conserved across species, in contrast to the poor conservation displayed by the "non-observable" part of the IL-loop. In all our activity assays the tetramer assembly always displays a lower deubiquitinating activity in comparison to the dimer. In our work, almost all point mutants involved in the interface are assembled as dimers and displayed higher deubiquitinating activity. However, neither high ubiquitin concentrations cannot disrupt the tetramer assembly (**Supplementary Fig. 2d,e**) nor increasing amounts the IL-loop peptide disturbs the USP25 deubiquitinating activity (**Supplementary Fig. 6e**), excluding a direct competition between ubiquitin substrate and the IL-loop peptide. It is worth mentioning P535L point mutant, which does not directly contact the ubiquitin-binding pocket but its substitution by leucine probably affects the IL-loop orientation and thus decreases its binding affinity. P535L mutant was checked due its occurrence in three independent cancer genomic studies (COSMIC database) [37]. Recently, USP25 has been associated to several types of cancer such as breast cancer, lung cancer and non-small cell lung cancer [25-27] and the P535L mutant would render USP25 more active than the wild type by destabilizing the tetramer assembly.

Our results reveal the presence of two stable assemblies of USP25, dimer and tetramer, or "active" and "inactive" enzyme, respectively. In the cellular context, the release of the inhibitory IL-loop from the ubiquitin-binding pocket would

destabilize the tetramer structure and promote the formation of the dimer (**Fig. 6**). In order to find relevant single point mutations prompting to this oligomeric transition, we replaced USP25 Tyr454 either to phenylalanine or to glutamic acid. Tyr454 is located in the LCC1 coiled-coil structure in contact with the "kink" motif and participates in the tetramer interface, being an excellent candidate to disassemble the tetramer (**Supplementary Fig. 11**). Moreover, the conserved homolog residue in USP28 (Tyr447) has been reported to be phosphorylated in a human proteomics phospho-site global analysis [38]. Interestingly, in vitro analysis reveals that Y454F do not affect the tetramer stability, but Y454E, a phosphomimetic tyrosine substitution, results in the formation of a stable dimer (**Supplementary Fig. 11**). Thus, even though the presence of high number of contacts in the tetramer, a single point mutant in the interface, such as Y454E, is capable to destabilize the tetramer and render USP25 to an "active" dimer. Phosphorylation of a single residue of the USP catalytic domain has already been shown to regulate the deubiquitinating activity in USP14 and USP37 [39,40], so we speculate whether it would be a plausible mechanism to switch between these two oligomer assemblies in USP25.

In addition to USP25, which is involved in the Wnt/β-catenin signaling by the regulation of the levels of tankyrases, the deubiquitinating activity of its homolog USP28 has been recently described to regulate the stability of P53 [41-43]. Intriguingly, USP28 shares a high degree of sequence and structural homology with USP25 (**Supplementary Fig. 1b)**, so it would be reasonable that the inhibitory regulatory mechanism described here for USP25 might be extensive to USP28, thus adding another layer of complexity to the regulation of P53.

Finally, the inhibition of the USP25 activity by the interaction with the IL-loop, which basically comprises 8 to 10 residues sequence stretch, suggests the possibility of identifying small molecules that could mimic this specific binding to the USP25 catalytic domain. Those compounds would represent novel therapeutic approaches for the treatment of pathologies derived from an abnormal proliferative functions, such as the Wnt/β-catenin signaling pathway [44].

Materials and methods

Plasmids, Cloning and Point Mutation

The full length of USP25 was cloned from pENTR-USP25 (purchased from Open Biosystems, Human ORFeome Collection). Different constructs of USP25 were amplified by PCR and cloned into the BamHI/NotI restriction enzymes sites of pET28-Smt3 vector using ligation dependent cloning, yielding an expression construct with an N-terminal SENP-cleavable Smt3 tag. Truncation mutants ΔIL and ΔLCC-IL were generated by overlapping extension PCR [45] and linked by a three-amino-acid linker peptide Gly-Ser-Gly. Point mutations were created using the QuickChange site-directed mutagenesis kit (Stratagene).

Protein Expression and Purification

pET28-Smt3 vectors harboring different constructs of USP25 were transformed into *Escherichia coli* strain Rosetta (DE3) cells and expression was induced with 0.5 mM IPTG and grown at 28 °C overnight. Proteins were purified from the soluble cell lysate by Ni^{2+} affinity chromatography using a lysis buffer containing 20 mM Tris-HCl, pH 8.0, 250 mM NaCl, and 1 mM β-mercaptoethanol. After Smt3 tag removal using SENP2, the proteins were further purified by gel filtration (Superdex 200 column, GE Healthcare) chromatography pre-equilibrated in 100 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM β-mercaptoethanol, followed by an anion exchange (Resource Q column, GE Healthcare) chromatography. Proteins were concentrate using Amicon Ultra-30K ultrafiltration device (Milipore) prior to the following experiments.

Methylation, Crystallization and Data Collection

USP25 construct containing residues 18-714 (USP25NCD) was expressed and purified by Ni^{2+} affinity chromatography. After SENP2 cleavage, it was purified by gel filtration chromatography using Superdex 200 column pre-equilibrated in 250 mM NaCl, 50 mM HEPES pH 7.5, 1 mM β-mercaptoethanol. Fractions containing the proteins of tetramer and dimer were pooled, followed by a lysine methylation step based on a published strategy [31]. In brief, borane-dimethylamine complex (Sigma-Aldrich) and formaldehyde (Sigma-Aldrich) were sequentially added into

protein solution and incubated overnight at 4 °C overnight. The methylation reaction was stopped by a final gel filtration chromatography on a Superdex 200 column pre-equilibrated in 200 mM NaCl, 20 mM Tris-HCl pH 7.5, 1 mM β-mercaptoethanol, followed by an anion exchange chromatography. USP25NCD tetramer and USP25NCD dimer were finally concentrated to 8 g/L for crystallization. Both USP25NCD tetramer and dimer could grow crystals, but only tetramer crystals could diffract well. Crystals were grown at 18°C by hanging-drop vapour diffusion method by mixing the 1 µl protein in 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM β-mercaptoethanol with an equal volume of reservoir solution containing 18% PEG3350 (w/v), 100 mM Bis-Tris Propane pH 8.5, 200 mM NaF. Full size crystals were obtained after 3 days, followed by a post-crystallization dehydration procedure, which involved transferring crystals to new solutions containing gradually increasing concentration of PEG3350 (18%-25%). Diffraction-quality crystals were soaked in buffers supplemented with 15% (v/v) ethylene glycol and flash-frozen in liquid nitrogen. Crystals have cell dimensions of a = 140.8 Å, b = 140.8 Å, c = 190.1 Å, α = β = γ = 90 °. The attempts to find molecular replacement solutions using all previous published USP structures as search models did not generate any reliable solutions. Therefore, experimental phasing was performed by soaking the crystals in a drop containing 1 mM $HgCl_2$ and incubated overnight at 18 °C and then harvested from the soaking condition. Native and derivative datasets were collected at ALBA synchrotron in Barcelona (BL13-XALOC beamline) [46]and processed with XDS [47] and scaled, reduced, and further analyzed using CCP4 [48].

Structure Determination and Refinement

SAD dataset was collected at 4,37 Å resolution using a wavelength of 0.9998 Å. Heavy atom sites and initial density map were calculated using PHENIX SAD Autosol [49]. The density maps clearly indicated the presence of 3 Hg^{2+} sites bound to free cysteine residues, which allowed the formation of a rough initial density map. Density modification and phase extension programs from CCP4 [48] were used to improve the initial density maps and build an initial model with Coot [50]. The initial model calculated with the Hg^{2+} dataset was used as a search model for molecular replacement in the native dataset collected at 3,28 Å resolution with Phaser-Phenix

[51]. The model was manually reconstructed using Coot [50] and further refine with PHENIX [49].

Analytical gel filtration and native PAGE Assay

All USP25 constructs were loaded on a Superdex 16 column pre-equilibrated in 100 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM β-mercaptoethanol and peak fractions were assayed by SDS-PAGE. Carves were generated by plotting the absorbance at 280 nM versus elution volumes and normalized to the same level. For running the native PAGE, all USP25 constructs were adjusted to 0.6 g/L and prepared in non-reducing non-denaturing loading buffer. All gels were stained with coomassie brilliant blue.

Deubiquitinating assays against diubiquitin and polyubiquitin chain substrates

Both diubiquitin (K48-linked) and polyubiquitin chain (K63-linked and K48-linked) hydrolysis reactions were performed in 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 5 mM DTT, 0.1% (v/v) Tween20 with different USP25 constructs at 30 °C or 37 °C. The concentrations of enzymes and substrates in reactions were determined by specific experiments. Deubiquitylation of diubiquitin chains with native and lysine methylated USP25NCD tetramer was performed in 0.01 mg/mL K48-linked diubiquitin substrates and 200 nM enzymes. Deubiquitylation of diubiquitin chains with USP25NCD point mutants was performed in 0.025 mg/mL diubiquitin (K48-linked) and 200 nM enzymes. Deubiquitylation of diubiquitin chains with USP25NCD truncation mutants was performed in 0.01 mg/mL diubiquitin (K48-linked) and 100 nM enzymes. Deubiquitylation of polyubiquitin chains was performed in 0.05 mg/mL K63-linked or K48-linked polyubiquitin chain substrates and 100 nM USP25 constructs. All reactions were stopped with SDS loading buffer at the indicated times and analyzed by SDS-PAGE followed by staining with SYPRO (Bio-Rad).

Dynamic light scattering

The average size of all USP25 constructs were measured by Dynamic Light Scattering (DLS) using a Malvern Zetasizer Nano-S90. Samples were measured in 100 µl buffer with 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM β-mercaptoethanol

at 25 °C in 0.1 mL Malvern disposable polystyrene cuvettes. Analyses were performed in triplicates with highly consistency and the representative results were shown. Due to aggregations, the intensity distributions which indicates how much light is scattered from various size “slices” or “bins” showed more than one peaks, and resulted in a high overall polydispersity index (PDI). However, when transforming the intensity distribution to a volume distribution the result only showed a single peak. The volume contributions from the aggregation peaks were then so small that they were no longer displayed. The size distribution by volume graphs for different USP25 constructs were generated using GraphPad Prism 7.0.

Hydrolysis and Kinetic Analysis with Ubiquitin-AMC Assay

Ub-AMC hydrolysis assays and kinetic analysis were set up in fluorescence cuvettes (Hellma Analytics) in 100 µl buffer with 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 5 mM DTT, 0.1% (v/v) Tween20 and measured with a fluorescence spectrometer (JASCO FP-8200) at excitation and emission wavelengths of 355 nM and 455 nM. Hydrolysis assays were performed at 37 °C with 200 nM USP25 constructs and 0.1 µM Ub-AMC in triplicates. Kinetic analysis (Michaelis-Menten kinetic measurements) were carried out using 5 nM USP25 constructs with a series of Ub-AMC substrate titrations at 37 °C. Initial rates of substrate hydrolysis were obtained using linear regression. Kinetic curves were obtained by plotting the measured enzyme initial rates versus the corresponding substrate concentrations, followed by the modeling using nonlinear regression fit with Michaelis-Menten equation. All data were processed using GraphPad Prism 7.0.

Cell culture, antibodies and reagents

The human embryonic kidney 293T cell line (HEK293T) (CRL-1573; ATCC) was used for ectopic expression of Flag-USP25 and its mutants. HEK293T cells were cultured in DMEM medium (Sigma-Aldrich, St Louis, MO), supplemented with 10% (v/v) FBS, 2 µM L-glutamine and 100 U/mL Penicillin/Streptomycin. Cells were growth in a 37ºC humidified incubator containing 5% CO_2. The commercial antibodies used for Western blotting analysis included the following: anti-TNKS1/2 (1:1000 dilutions; Santa Cruz Biotechnology; sc-8337); anti-USP25 (1:1000 dilution; Abcam, ab187156); anti-Flag (1:1000 dilutions; Sigma-Aldrich, F7425) and anti-

Tubulin (1:5000 dilutions; Sigma, T5168), Anti-Flag M2 affinity gel (Sigma-Aldrich, no. A2220). CHX (Sigma-Aldrich, no. 01810) was added to the cell culture medium in a final concentration of 100 µg/mL and cells were collected at the indicated times (0h, 3h and 6h) for western blotting. Bortezomib (Jansen Pharmaceuticals) was added to the cell culture medium in a final concentration of 0.5 µM and cells were collected after 6h for western blotting.

Cell culture vector constructs and transfection

pcDNA3.1-FlagUSP25FL, pcDNA3.1-FlagUSP25FL ΔIL, pcDNA3.1-FlagUSP25FL C178A, pcDNA3.1-FlagUSP25FL ΔIL C178A, pcDNA3.1-FlagUSP25 NCD, pcDNA3.1-FlagUSP25 NCD ΔIL and pcDNA3.1-FlagUSP25FL point mutants P521S, F522G, E373A, C651F and L271W vector constructs were generated by PCR using previous pET28-Smt3 USP25 constructs as DNA templates, the specific primers fused with a Flag tag at the N Termini and Phusion DNA polymerase (Thermo) following the manufacturer's instructions. The PCR products were cloned into pcDNA3.1 (Invitrogen) using the T4DNA ligase (Thermo) and sequenced previous to use. pcDNA3.1-Flag USP25 and its mutated vector constructs were transfected into HEK293T cells using the Lipotransfectin (NIVORLAB), according to manufacturer's instructions. Forty-eight hours later, transfected HEK293T cells were collected and lysed in Triton Lysis Buffer (TLB: 50 mM Tris-HCl pH 7.5, 150 µM NaCl, 1 µM EDTA, 50 µM NaF, 0.5% Triton X-100, plus protease inhibitors).

Western blot experiments

Total protein lysates derived from transfected HEK293T cells were used for western blot (WB) experiments using the human specific antibodies anti-TNKS1/2 (1:1000 dilutions; Santa Cruz Biotechnology; sc-8337); anti-USP25 (1:1000 dilution; Abcam, ab187156); anti-Flag (1:1000 dilutions; Sigma-Aldrich, F7425) and anti-Tubulin (1:5000 dilutions; Sigma, T5168). Membranes were developed with chemiluminescence substrate Pierce® ELC Western blotting substrate (Thermo Fisher Scientific, South Logan, UT), and visualized on a LAS4000 device (Fujifilm, Tokyo, Japan). Protein quantification was done with Image Gauge software (Fujifilm).

Immunoprecipitation Experiments

Total protein extracts obtained from HEK293T cells were transiently transfected with Flag-USP25 or its mutant expression plasmids using TLB as describer before. 5mg of protein extracts were used for IP using specific Anti-Flag M2 affinity gel. IP was performed overnight at 4ºC with slow rotation. After 24 hours, the supernatant was separated from the gel and washed three times with TLB, the Flag-fused proteins were eluted with SDS-loading buffer for 1h at room temperature. Immunoprecipitated proteins were analyzed by WB using the specific anti-Flag antibodies and anti-TNKS1/2.

Statistical analysis for table of tankyrase stability

Data are represented as mean ± SD of three independent experiments. Statistical tests were performed using SPSS v16.0 software (SPSS). Comparison between two groups of samples was evaluated by independent sample *t* test, and results were considered statistically significant when $p\text{-}val<0.05$**.**

Author contributions: BL conducted all protein expression, crystallization and kinetics experiments. BL and DR solved the crystal structure. VA and MS conducted in vivo cell culture analysis. DR and BL analyzed the results and wrote the paper. DR conceived the idea for the project.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Acknowledgements: This work was supported by grants from the “Ministerio de Economia y Competitividad” BFU2015-66417-P (MINECO/FEDER) to DR, and BFU2015-64879-R to VA. BL and YZ acknowledge the support from the China Scholarship Council (China).

Additional information:

Accession codes: Coordinates and reflection data files have been submitted to the protein data bank with accession number 5O71.

References

1. Komander, D. & Rape, M. The ubiquitin code. *Annual review of biochemistry* **81**, 203-229 (2012).
2. Yau, R. & Rape, M. The increasing complexity of the ubiquitin code. *Nature cell biology* **18**, 579 (2016).
3. Swatek, K.N. & Komander, D. Ubiquitin modifications. *Cell research* **26**, 399 (2016).
4. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annual review of biochemistry* **67**, 425-479 (1998).
5. Reyes-Turcu, F.E., Ventii, K.H. & Wilkinson, K.D. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annual review of biochemistry* **78**, 363-397 (2009).
6. Clague, M.J. et al. Deubiquitylases from genes to organism. *Physiological reviews* **93**, 1289-1315 (2013).
7. Komander, D., Clague, M.J. & Urbé, S. Breaking the chains: structure and function of the deubiquitinases. *Nature reviews. Molecular cell biology* **10**, 550 (2009).
8. Rehman, S.A.A. et al. MINDY-1 is a member of an evolutionarily conserved and structurally distinct new family of deubiquitinating enzymes. *Molecular cell* **63**, 146-155 (2016).
9. Nijman, S.M. et al. A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**, 773-786 (2005).
10. Hu, M. et al. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* **111**, 1041-1054 (2002).
11. Rougé, L. et al. Molecular understanding of USP7 substrate recognition and C-terminal activation. *Structure* **24**, 1335-1345 (2016).
12. Faesen, A.C. et al. Mechanism of USP7/HAUSP activation by its C-terminal ubiquitin-like domain and allosteric regulation by GMP-synthetase. *Molecular cell* **44**, 147-159 (2011).
13. Clerici, M., Luna-Vargas, M.P., Faesen, A.C. & Sixma, T.K. The DUSP–Ubl domain of USP4 enhances its catalytic efficiency by promoting ubiquitin exchange. *Nature communications* **5**(2014).

14. Zhen, Y., Knobel, P.A., Stracker, T.H. & Reverter, D. Regulation of USP28 deubiquitinating activity by SUMO conjugation. *Journal of Biological Chemistry* **289**, 34838-34850 (2014).
15. Meulmeester, E., Kunze, M., Hsiao, H.H., Urlaub, H. & Melchior, F. Mechanism and consequences for paralog-specific sumoylation of ubiquitin-specific protease 25. *Molecular cell* **30**, 610-619 (2008).
16. Cholay, M., Reverdy, C., Benarous, R., Colland, F. & Daviet, L. Functional interaction between the ubiquitin-specific protease 25 and the SYK tyrosine kinase. *Experimental cell research* **316**, 667-675 (2010).
17. Kim, S. et al. Vaccinia-related kinase 2 controls the stability of the eukaryotic chaperonin TRiC/CCT by inhibiting the deubiquitinating enzyme USP25. *Molecular and cellular biology* **35**, 1754-1762 (2015).
18. Denuc, A., Bosch-Comas, A., Gonzàlez-Duarte, R. & Marfany, G. The UBA-UIM domains of the USP25 regulate the enzyme ubiquitination state and modulate substrate recognition. *PLoS One* **4**, e5571 (2009).
19. Blount, J.R., Burr, A.A., Denuc, A., Marfany, G. & Todi, S.V. Ubiquitin-specific protease 25 functions in Endoplasmic Reticulum-associated degradation. *PloS one* **7**, e36542 (2012).
20. Jung, E.S., Hong, H., Kim, C. & Mook-Jung, I. Acute ER stress regulates amyloid precursor protein processing through ubiquitin-dependent degradation. *Scientific reports* **5**(2015).
21. Zhong, B. et al. Negative regulation of IL-17-mediated signaling and inflammation by the ubiquitin-specific protease USP25. *Nature immunology* **13**, 1110 (2012).
22. Zhong, B. et al. Ubiquitin-specific protease 25 regulates TLR4-dependent innate immune responses through deubiquitination of the adaptor protein TRAF3. *Sci. Signal.* **6**, ra35-ra35 (2013).
23. Lin, D. et al. Induction of USP25 by viral infection promotes innate antiviral responses by mediating the stabilization of TRAF3 and TRAF6. *Proceedings of the National Academy of Sciences* **112**, 11324-11329 (2015).
24. Ren, Y. et al. The Type I Interferon-IRF7 axis mediates transcriptional expression of Usp25 gene. *Journal of Biological Chemistry* **291**, 13206-13215 (2016).

25. Deng, S. et al. Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome subunits (PSs) in breast cancer tissue observed by the methods of RFDD-PCR and proteomics. *Breast cancer research and treatment* **104**, 21-30 (2007).
26. Yamada, H. et al. Detailed characterization of a homozygously deleted region corresponding to a candidate tumor suppressor locus at 21q11-21 in human lung cancer. *Genes, Chromosomes and Cancer* **47**, 810-818 (2008).
27. Li, J. et al. miRNA-200c inhibits invasion and metastasis of human non-small cell lung cancer by directly targeting ubiquitin specific peptidase 25. *Molecular cancer* **13**, 166 (2014).
28. Xu, D. et al. USP25 regulates Wnt signaling by controlling the stability of tankyrases. *Genes & development* (2017).
29. Yang, Y. et al. Structural and Functional Investigations of the N-Terminal Ubiquitin Binding Region of Usp25. *Biophysical journal* **112**, 2099-2108 (2017).
30. Mevissen, T.E. & Komander, D. Mechanisms of Deubiquitinase Specificity and Regulation. *Annual Review of Biochemistry* (2017).
31. Walter, T.S. et al. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617-1622 (2006).
32. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-W324 (2014).
33. Hu, M. et al. Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *The EMBO journal* **24**, 3747-3756 (2005).
34. Basters, A. et al. Structural basis for the specificity of USP18 towards ISG15. *Nature structural & molecular biology* **24**, 270 (2017).
35. Sato, Y. et al. Structures of CYLD USP with Met1-or Lys63-linked diubiquitin reveal mechanisms for dual specificity. *Nature Structural and Molecular Biology* **22**, 222 (2015).
36. Balsas, P. et al. SOX11 promotes tumor protective microenvironment interactions through CXCR4 and FAK regulation in mantle cell lymphoma. *Blood*, blood-2017-04-776740 (2017).
37. Forbes, S.A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic acids research* **45**, D777-D783 (2016).
38. Ross, K.E. et al. iPTMnet: Integrative Bioinformatics for Studying PTM

Networks. *Protein Bioinformatics: From Protein Modifications and Networks to Proteomics*, 333-353 (2017).

39. Xu, D. et al. Phosphorylation and activation of ubiquitin-specific protease-14 by Akt regulates the ubiquitin-proteasome system. *Elife* **4**, e10510 (2015).
40. Huang, X. et al. Deubiquitinase USP37 is activated by CDK2 to antagonize APC CDH1 and promote S phase entry. *Molecular cell* **42**, 511-523 (2011).
41. Cuella-Martin, R. et al. 53BP1 integrates DNA repair and p53-dependent cell fate decisions via distinct mechanisms. *Molecular cell* **64**, 51-64 (2016).
42. Lambrus, B.G. et al. A USP28–53BP1–p53–p21 signaling axis arrests growth after centrosome loss or prolonged mitosis. *J Cell Biol* **214**, 143-153 (2016).
43. Meitinger, F. et al. 53BP1 and USP28 mediate p53 activation and G1 arrest after centrosome loss or extended mitotic duration. *J Cell Biol* **214**, 155-166 (2016).
44. Guo, Y., Xiao, L., Sun, L. & Liu, F. Wnt/[beta]-catenin signaling: a promising new target for fibrosis diseases. *Physiological research* **61**, 337 (2012).
45. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59 (1989).
46. Juanhuix, J. et al. Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *Journal of synchrotron radiation* **21**, 679-689 (2014).
47. Kabsch, W. Xds. *Acta Crystallographica Section D: Biological Crystallography* **66**, 125-132 (2010).
48. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235-242 (2011).
49. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213-221 (2010).
50. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486-501 (2010).
51. McCoy, A.J. et al. Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674 (2007)

SUPPLEMENTARY FIGURES

A quaternary tetramer assembly inhibits the deubiquitinating activity of USP25

Bing Liu[1], Marta Sureda-Gómez[2], Yang Zhen[1], Virginia Amador[2] & David Reverter[1]

[1] Institut de Biotecnologia i de Biomedicina (IBB) and Dept. de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Spain.
[2] Institut de Investigacions Biomèdiques Agustí Pi i Sunyer (IDIBAPS). Barcelona, Spain.

Correspondence: david.reverter@uab.cat

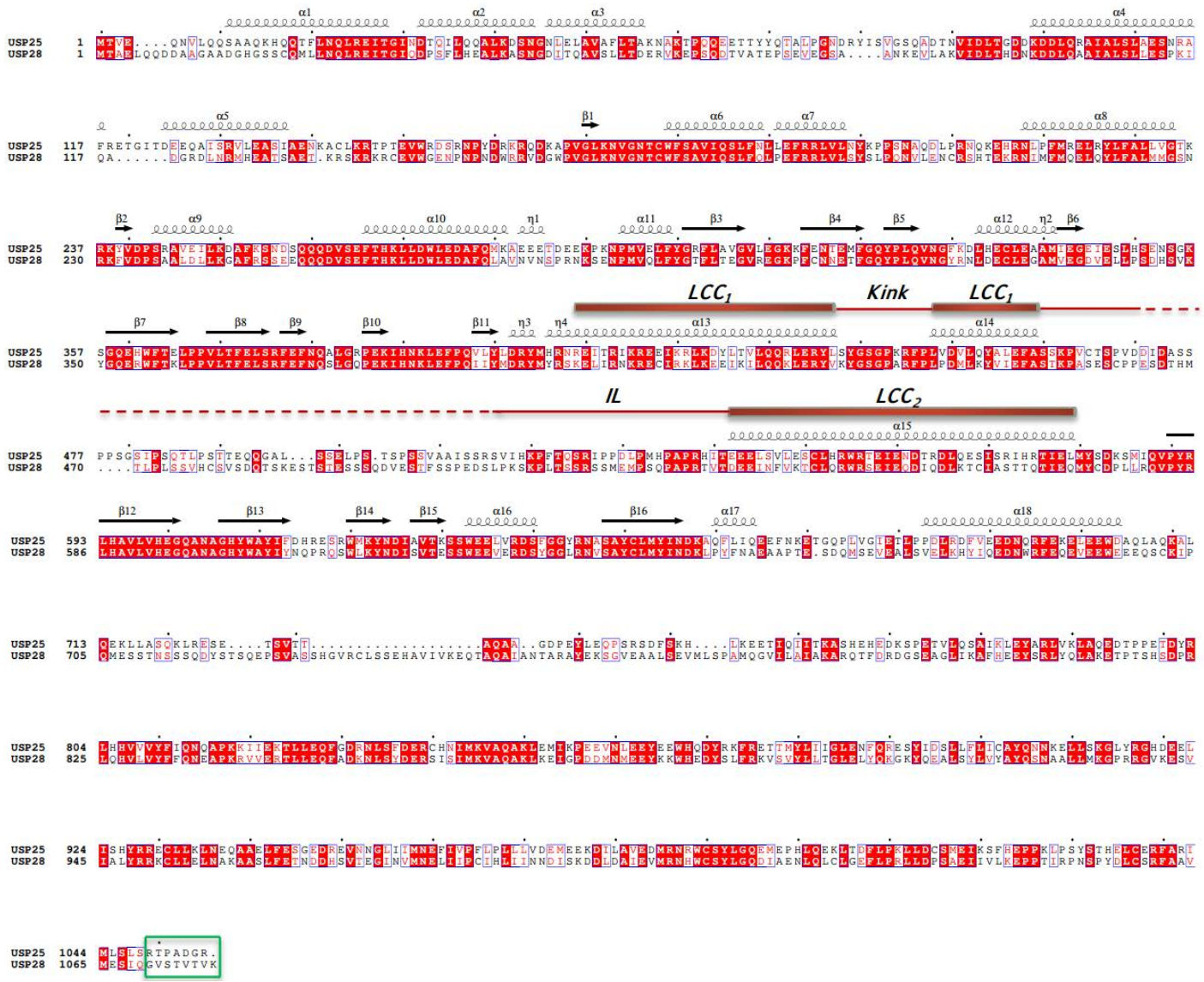
Table 1. Data collection and structure refinement statistics.

Data collection		
Beamline	ALBA-XALOC	ALBA-XALOC (Hg^{2+})
Space group	I 4 2 2	I 4 2 2
Wave length (Å)	0.9791	0,9998
Resolution (Å)	95.08-3.28 (3.29-3.28)	94.76-4.35 (4.35-4.37)
a, b, c (Å)	140.806, 140.806, 190.156	141.17,141.17,190,17
α, β, γ (°)	α = β = γ = 90	α = β = γ = 90
Unique reflections	14883	6588
Data redundancy	6.50 (6.40)	25.4 (26.5)
R_{merge}	0,07 (0.98)	0.19 (1.06)
CC (1/2)	0,99 (0.85)	0.99 (0.97)
I/σ	16.3 (2.3)	20,3 (6.9)
Completeness (%)	99.5 (97.7)	99.9 (95.7)
Refinement		
Resolution (Å)	95.08-3.28 (3.37-3.28)	
Non-anomalous reflections	14150	
R_{work}/R_{free}	0.202/0.275	
Number of all atoms	3963	
RMSD bond (Å)/Angle (°)	0.011/1.524	
Ramachandran plot		
Favored (%)	91.99	
Allowed (%)	6.71	
Disallowed (%)	1.30	

a


UBA
SIM
UIM1
UIM2
NTD
USP
+
LCC
+
IL
CTD
LCC1
Kink
LCC1
IL
LCC2
Human
Sheep
Chicken
Frog
Zebrafish

b



C

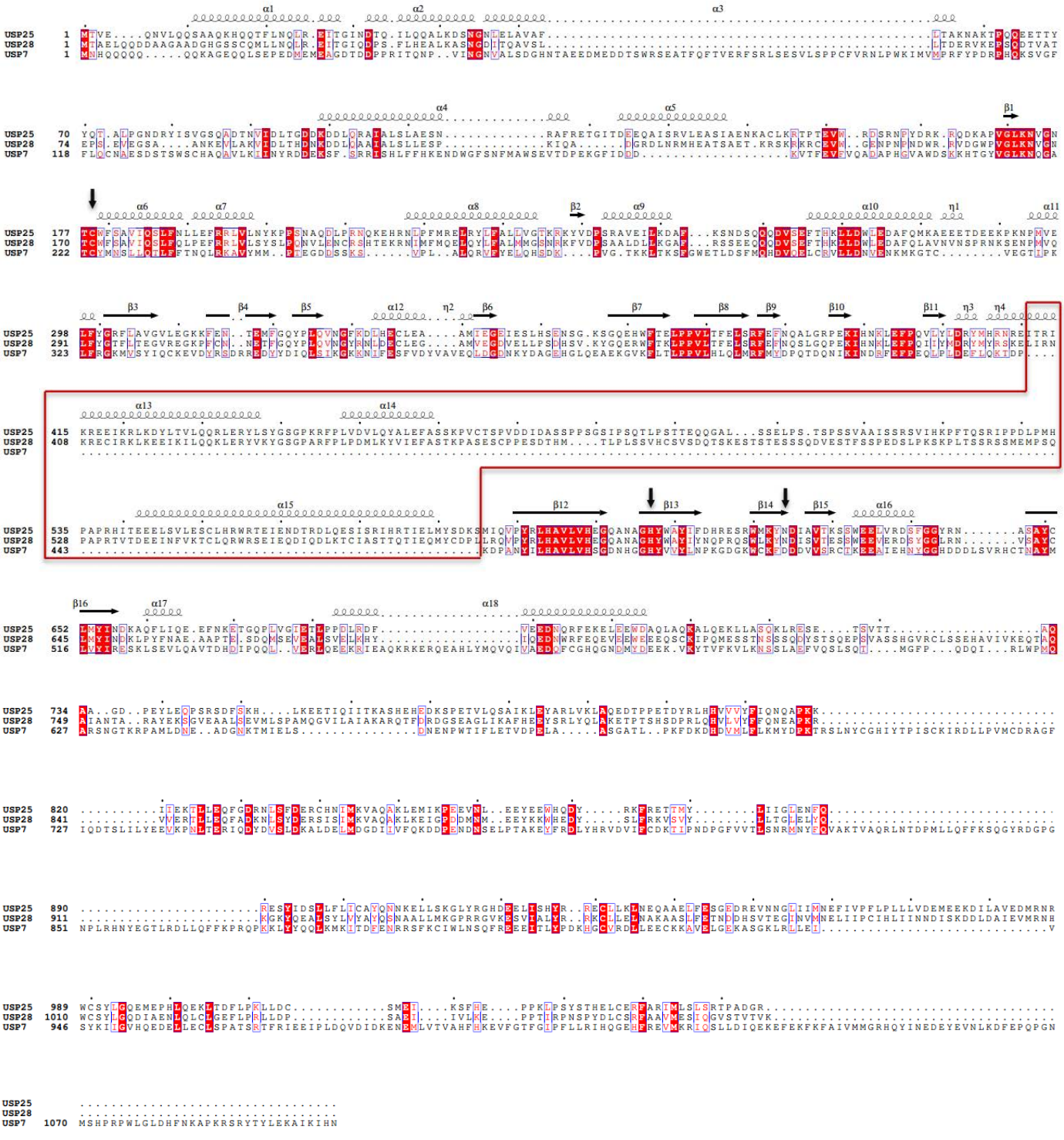


Figure S1. Sequence alignments of USP25.
(a) Sequence alignment of human USP25 with its homologs in sheep, chicken, frog and zebrafish. Conserved residues are shaded in red. Three different domains of USP25 are labeled: N-terminal domain (labeled as NTD), USP-like domain (labeled as USP+LCC+IL) and C-terminal domain (labeled as CTD). Motifs in NTD are indicated in grey cylinders: ubiquitin associated domain (UBA), SUMO interacting motif (SIM) and two ubiquitin interacting motifs (UIM1 and UIM2). Long coiled coils (LCC_1 and LCC_2) are indicated in red cylinders, whereas the "Kink" and inhibitory loop (IL) are indicated in red lines. Dashed red line indicates the non-conserved sequence (absent in the USP25 crystal structure) of IL. Catalytic triad residues (C178, H607 and N624) are highlighted with black arrows. Residues involved in the interactions between molecules B and A (tetramer contacts IL-loop), molecules B and A' (tetramer contacts Kink), and molecules B and B' (dimer contacts) are marked with blue dots, purple stars and green triangles, respectively. Molecules A, A', B and B' are indicated in **Fig. 3b** and **Fig. S5**. (b) Sequence alignment of USP25 with its homolog USP28. The labeling schemes for LCC_1, LCC_2, "Kink" and IL are the same as in (a). The C-terminal tails of USP25 (residues interacting with tankyrases) and USP28 are marked with a green square. (c) Sequence alignment of USP25 with USP28 and USP7. Catalytic triad residues are highlighted with black arrows. The long sequence insertions of USP25 and USP28 are circled with red lines. All sequences are aligned online with Clustal Omega and formatted using ESPript.

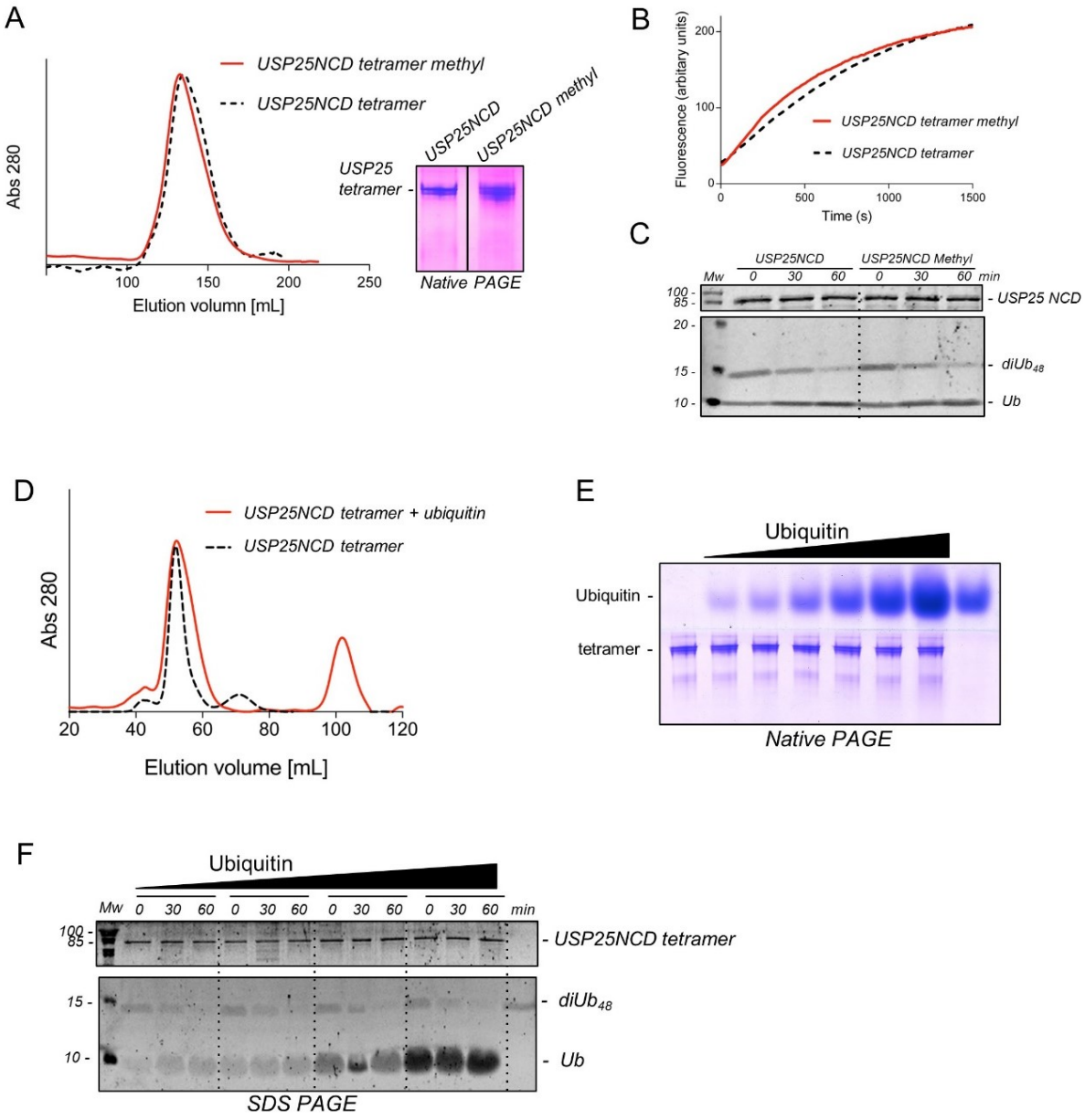


Figure S2. Biochemical characterization of methylated USP25 and of the competition with ubiquitin.

(A) Gel filtration profiles of the USP25NCD tetramer before and after conducting the lysine methylation protocol described in the methods. ***Right***, native PAGE of USP25NCD tetramer before and after conducting the lysine methylation protocol. (B) Time-course deubiquitinating activity of the USP25NCD, before and after conducting the lysine methylation protocol, using Ub-AMC fluorescent substrate. (C) PAGE analysis of a time-course deubiquitinating activity of the USP25NCD, before and after conducting the lysine methylation protocol, using $diUb_{48}$ as a substrate. (D) Gel filtration profiles of the USP25NCD in the presence of ubiquitin. (E) Native gel of the USP25NCD tetramer in the presence of increasing amounts of ubiquitin. (F) PAGE analysis of a time-course deubiquitinating activity of the USP25NCD, in the presence of increasing amounts of ubiquitin, using $diUb_{48}$ as a substrate.

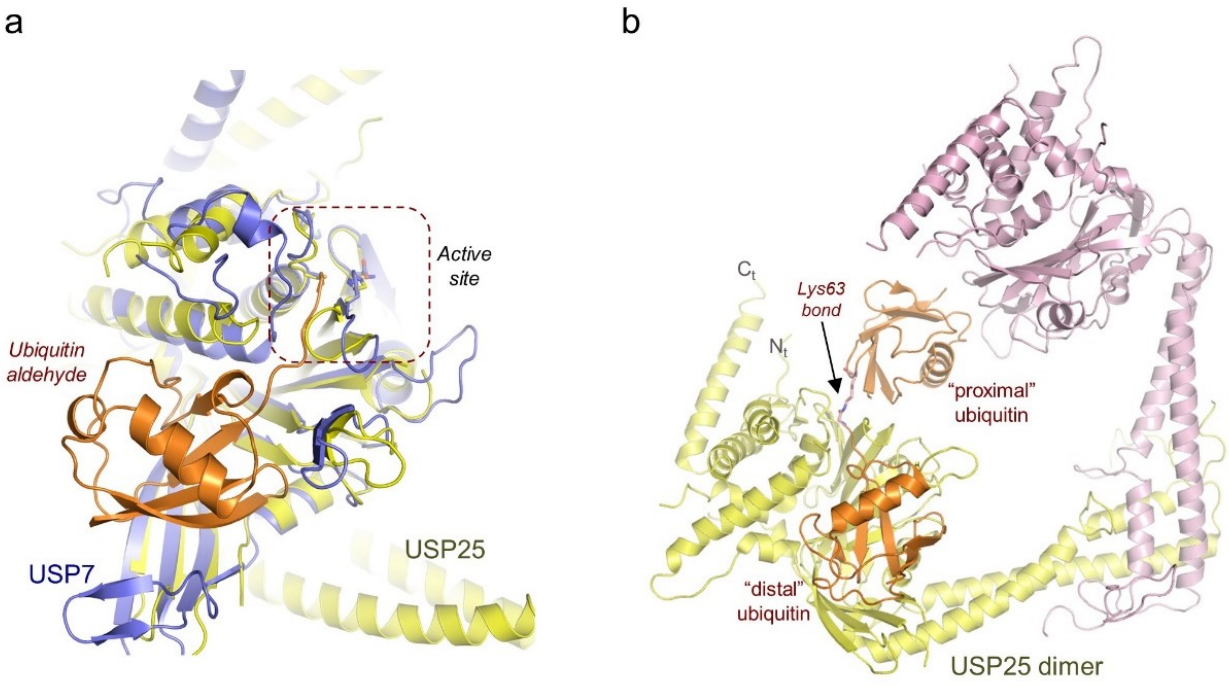


Figure S3. Structure comparison of USP25 with USP7-ubiquitin aldehyde and CYLD-K63 diubiquitin complexes.
(a) Superimposition of the catalytic region of USP25 and USP7-ubiquitin aldehyde complex. USP25 is shown in yellow. USP7-ubiquitin aldehyde complex (PDB code 5JTJ) are shown in blue and orange for each molecule. Active site catalytic triad residues are shown in stick representation. Zoomed-up details of the active site region depicted in **Fig. 2d**. (b) Structural model of K_{63}-diubiquitin substrate in complex with USP25 dimer, based on the superposition between USP25 and the CYLD-K63-diubiquitin structure (PDB code 3WXG). Rmsd of the superposition is 2,74 Å for 218 aligned residues. "Proximal" and "distal" ubiquitin are shown in orange.

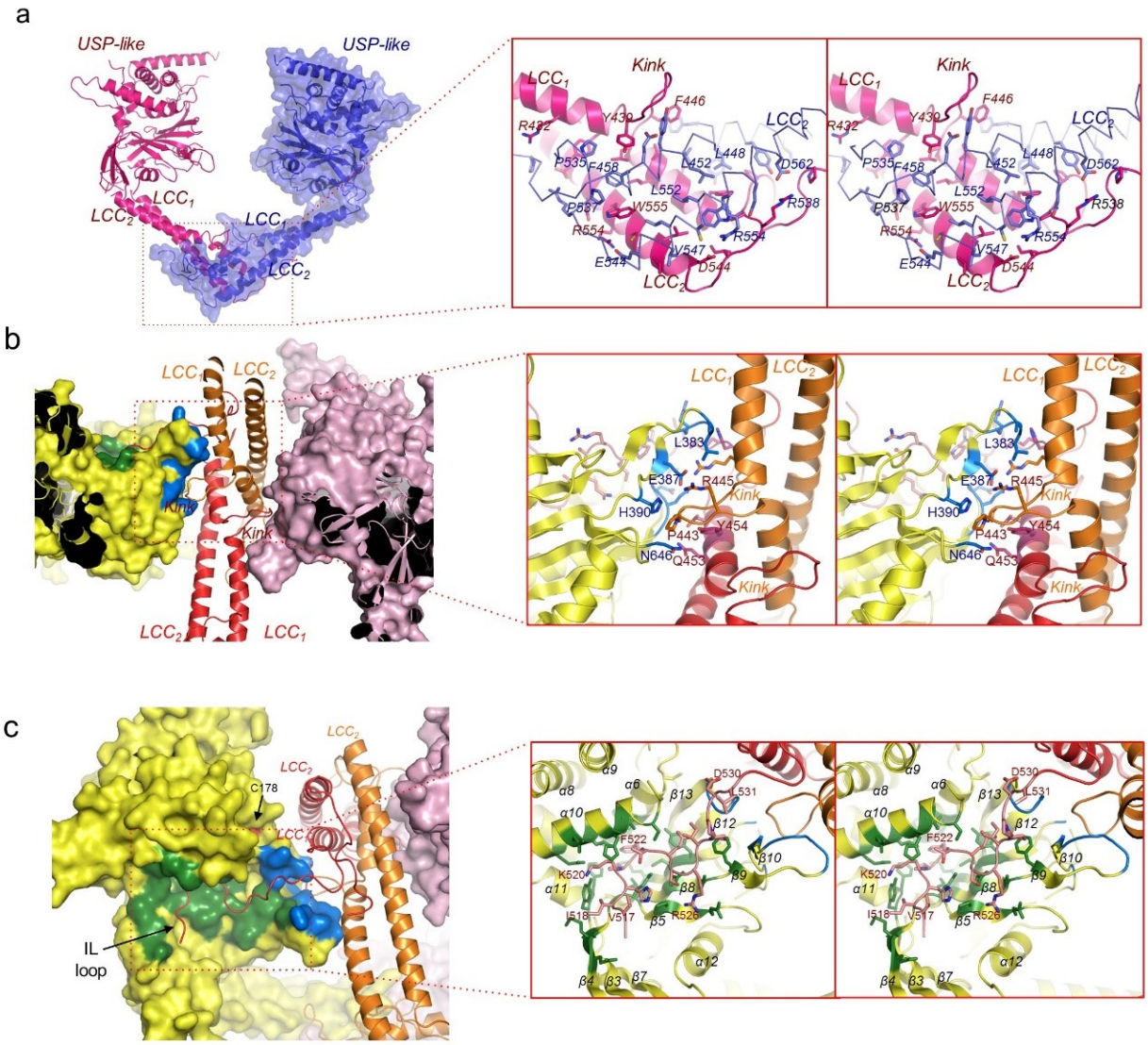


Figure S4. Detailed interactions of the residues composing the interfaces in the dimers and tetramer assemblies of USP25.
(a) *Left*, ribbon and surface representation of the two monomers composing the dimer structure observed in the crystal structure of USP25. *Right*, stereo representation of the interface between the LCC domains in the dimer assembly. Contact residues are labeled and shown in stick representation. (b) *Left*, overall structure of the "*kink*" interaction with the USP25 catalytic domain. The interaction surface is shown in blue. *Right*, stereo representation of the details of interaction. Contact residues are depicted in stick representation. "*Kink*" residues are labeled in orange, whereas the interaction residues in catalytic domain are labeled in blue. (c) *Left*, overall structure of IL-Loop interaction with the USP25 catalytic domain. The interaction surface is shown in green. *Right*, stereo representation of the details of interaction. Contact residues are depicted in stick representation. IL-loop residues are labeled in orange, whereas the interaction residues in catalytic domain are labeled in green.

a *Tetramer interaction (IL-loop)*

Chain B			Chain A			Distance(Å)
VAL	183	CG2	PHE	522	CG	3.13
LEU	187	CD2	PHE	522	CE2	3.68
ASP	262	OD2	ILE	527	CG2	3.32
SER	264	OG	GLN	524	N	2.68
SER	264	OG	THR	523	CB	3.31
THR	267	OG1	THR	523	CG2	3.4
HIS	268	NE2	LYS	520	CG	3.27
HIS	268	CA	THR	523	CG2	3.81
LEU	271	CD1	LYS	520	CB	3.49
LEU	271	CD1	PRO	521	N	3.71
LEU	271	CD1	PHE	522	N	3.78
ASP	272	OD1	LYS	520	CE	3.08
GLU	275	OE2	LYS	520	CE	3.21
PHE	299	CG	PRO	521	CG	3.98
TYR	300	OH	LYS	520	CD	2.91
TYR	300	CE2	ILE	518	CG2	3.47
GLU	318	OE1	SER	516	CB	3.17
MET	319	O	ILE	518	CG1	3.63
GLY	321	N	ILE	518	CG2	3.97
GLY	321	CA	HIS	519	O	3.2
GLN	322	NE2	THR	523	O	2.64
GLN	322	N	HIS	519	O	2.99
GLN	322	OE1	PRO	521	O	2.32
GLN	322	CD	LYS	520	O	3.55
PRO	324	CG	HIS	519	CD2	3.2
PRO	324	CB	ARG	526	NH2	3.76
GLN	326	CG	ARG	526	NH1	3.13
THR	371	OG1	PRO	521	O	3.86
GLU	373	OE2	GLN	524	CA	3.62
GLU	373	OE1	SER	525	OG	2.68
GLU	373	CG	THR	523	O	3.71
SER	375	OG	ARG	526	CB	3.42
SER	375	OG	SER	525	O	3.93
PHE	377	CB	ARG	526	O	3.67
PHE	377	CD1	PRO	528	CD	3.66
GLY	384	CA	HIS	534	NE2	3.14
ARG	385	CG	MET	533	SD	3.58
PRO	386	CG	PRO	528	CB	3.85
ASP	450	OD1	ASN	646	ND2	3.49
GLN	453	NE2	GLU	600	CG	3.65
GLN	453	OE1	ASN	646	ND2	2.72
SER	460	OG	ALA	605	CB	3.94
SER	516	CB	GLU	318	OE1	3.17
ILE	518	CG2	TYR	300	CE2	3.47
ILE	518	CG2	GLY	321	N	3.97
ILE	518	CG1	MET	319	O	3.63
HIS	519	O	GLY	321	CA	3.2
HIS	519	CD2	PRO	324	CG	3.2
HIS	519	O	GLN	322	N	2.99
LYS	520	CE	GLU	275	OE2	3.21
LYS	520	CD	TYR	300	OH	2.91
LYS	520	CG	HIS	268	NE2	3.27
LYS	520	CE	ASP	272	OD1	3.08
LYS	520	O	GLN	322	CD	3.55
LYS	520	CB	LEU	271	CD1	3.49
PRO	521	N	LEU	271	CD1	3.71
PRO	521	O	THR	371	OG1	3.86
PRO	521	O	CYS	651	SG	3.77
PRO	521	O	GLN	322	OE1	2.32
PRO	521	CG	PHE	299	CG	3.98
PHE	522	CZ	MET	653	SD	3.65
PHE	522	CE2	LEU	187	CD2	3.68
PHE	522	O	TYR	650	CE2	3.74
PHE	522	N	LEU	271	CD1	3.78
PHE	522	CG	VAL	183	CG2	3.13
PHE	522	CD1	LEU	597	CD1	3.4
PHE	522	CD1	CYS	651	SG	3.75
THR	523	CG2	HIS	268	CA	3.81
THR	523	O	GLU	373	CG	3.71
THR	523	CG2	THR	267	OG1	3.4
THR	523	O	GLN	322	NE2	2.64
THR	523	CB	SER	264	OG	3.31
GLN	524	CA	GLU	373	OE2	3.62
GLN	524	N	SER	264	OG	2.68
GLN	524	OE1	TYR	608	OH	3.62
GLN	524	CG	TYR	650	CD2	3.76
SER	525	OG	GLU	373	OE1	2.68
SER	525	O	SER	375	OG	3.93
ARG	526	O	PHE	377	CB	3.67
ARG	526	NH1	GLN	326	CG	3.13
ARG	526	CB	SER	375	OG	3.42
ARG	526	NH2	PRO	324	CB	3.76

Chain B			Chain A			Distance(Å)
ILE	527	CG2	ASP	262	OD2	3.32
ILE	527	CD1	ALA	603	O	3.45
PRO	528	CB	PRO	386	CG	3.85
PRO	528	CD	PHE	377	CD1	3.66
ASP	530	OD2	ASN	604	ND2	2.69
MET	533	SD	ARG	385	CG	3.58
HIS	534	NE2	GLY	384	CA	3.14
ARG	556	NH2	GLU	600	OE1	2.8
LEU	597	CD1	PHE	522	CD1	3.4
GLU	600	CG	GLN	453	NE2	3.65
GLU	600	OE1	ARG	556	NH2	2.8
ALA	603	O	ILE	527	CD1	3.45
ASN	604	ND2	ASP	530	OD2	2.69
ALA	605	CB	SER	460	OG	3.94
TYR	608	OH	GLN	524	OE1	3.62
ASN	646	ND2	ASP	450	OD1	3.49
ASN	646	ND2	GLN	453	OE1	2.72
TYR	650	CD2	GLN	524	CG	3.76
TYR	650	CE2	PHE	522	O	3.74
CYS	651	SG	PHE	522	CD1	3.75
CYS	651	SG	PRO	521	O	3.77
MET	653	SD	PHE	522	CZ	3.65

Dimer interaction (LCC domain)

Chain B			Chain B'			Distance(Å)
GLN	431	O	HIS	534	NE2	3.93
ARG	432	NH2	PRO	468	C	3.26
ARG	432	NH1	ASP	470	CB	3.65
ARG	432	NH2	SER	467	C	3.14
ARG	432	NH2	VAL	469	N	3.42
ARG	432	CG	HIS	534	CE1	3.63
ARG	435	CG	PRO	535	CD	3.46
ARG	435	CB	HIS	534	CD2	3.19
TYR	436	OH	SER	467	CB	3.71
TYR	436	CE2	PRO	535	CB	3.6
TYR	436	OH	HIS	534	O	3.39
TYR	439	CE1	PRO	535	CG	3.89
TYR	439	OH	PHE	458	CA	3.23
TYR	439	OH	GLU	457	C	3.67
TYR	439	CE2	TYR	454	O	3.48
GLY	440	C	GLU	457	OE1	3.91
GLY	440	CA	TYR	454	CE2	3.39
SER	441	OG	GLN	453	NE2	3.6
SER	441	N	TYR	454	CE1	3.67
SER	441	OG	GLU	457	OE2	2.77
GLY	442	N	TYR	454	OH	3.11
LYS	444	O	TYR	454	OH	2.65
PHE	446	CE2	TYR	454	CE2	3.8
PHE	446	CZ	PRO	447	CD	3.82
PHE	446	CG	PHE	446	CG	3.69
PHE	446	CE2	ASP	450	CB	3.52
PRO	447	CD	PHE	446	CZ	3.82
LEU	448	CD2	PHE	458	CE1	3.87
LEU	448	CD2	PRO	535	CB	3.68
ASP	450	CB	PHE	446	CE2	3.52
LEU	452	CD2	PHE	458	CD2	3.79
GLN	453	NE2	SER	441	OG	3.6
TYR	454	CE2	GLY	440	CA	3.39
TYR	454	O	TYR	439	CE2	3.48
TYR	454	CE1	SER	441	N	3.67
TYR	454	CE2	PHE	446	CE2	3.8
TYR	454	OH	GLY	442	N	3.11
TYR	454	OH	LYS	444	O	2.65
GLU	457	OE2	SER	441	OG	2.77
GLU	457	C	TYR	439	OH	3.67
GLU	457	OE1	GLY	440	C	3.91
PHE	458	CD2	LEU	452	CD2	3.79
PHE	458	CE2	TRP	555	CB	3.81
PHE	458	CE1	LEU	448	CD2	3.87
PHE	458	CA	TYR	439	OH	3.23
PHE	458	CZ	ILE	559	CD1	3.94
ALA	459	CB	TRP	555	CH2	3.94
SER	467	C	ARG	432	NH2	3.14
SER	467	CB	TYR	436	OH	3.71
SER	467	O	ASP	562	OD1	3.34
PRO	468	C	ARG	432	NH2	3.26
PRO	468	N	ASP	562	OD1	3.93
VAL	469	N	ARG	432	NH2	3.42
ASP	470	CB	ARG	432	NH1	3.65

Chain B			Chain B'			Distance(Å)
HIS	534	NE2	GLN	431	O	3.93
HIS	534	CE1	ARG	432	CG	3.63
HIS	534	CD2	ARG	435	CB	3.19
HIS	534	O	TYR	436	OH	3.39
PRO	535	CB	LEU	448	CD2	3.68
PRO	535	CD	ARG	435	CG	3.46
PRO	535	CB	TYR	436	CE2	3.6
PRO	535	CG	TYR	439	CE1	3.89
PRO	537	CA	GLU	558	OE1	3.8
PRO	537	CB	TRP	555	NE1	2.92
ARG	538	NH1	ASP	562	OD2	3.45
ARG	538	N	GLU	558	OE1	2.97
HIS	539	CE1	ARG	554	NE	3.13
HIS	539	O	TRP	555	NE1	3.64
GLU	544	OE1	ARG	554	NH2	3.19
GLU	544	CD	CYS	551	SG	3.12
GLU	544	CA	VAL	547	CG1	3.85
GLU	544	OE2	SER	550	OG	3.91
VAL	547	CG1	GLU	544	CA	3.85
LEU	548	CD2	LEU	548	CD2	2.88
LEU	548	CD1	TRP	555	CH2	3.45
SER	550	OG	GLU	544	OE2	3.91
CYS	551	SG	GLU	544	CD	3.12
ARG	554	NH2	GLU	544	OE1	3.19
ARG	554	NE	HIS	539	CE1	3.13
TRP	555	CH2	ALA	459	CB	3.94
TRP	555	NE1	HIS	539	O	3.64
TRP	555	CB	PHE	458	CE2	3.81
TRP	555	NE1	PRO	537	CB	2.92
TRP	555	CH2	LEU	548	CD1	3.45
GLU	558	OE1	ARG	538	N	2.97
GLU	558	OE1	PRO	537	CA	3.8
ILE	559	CD1	PHE	458	CZ	3.94
ASP	562	OD2	ARG	538	NH1	3.45
ASP	562	OD1	PRO	468	N	3.93
ASP	562	OD1	SER	467	O	3.34

Tetramer interaction (Kink)

Chain B			Chain A'			Distance (Å)
GLN	381	OE1	GLN	431	NE2	3.04
ALA	382	O	GLN	431	CA	3.57
ALA	382	O	GLU	434	CB	3.62
LEU	383	CD1	ARG	435	NE	3.76
LEU	383	CD2	GLU	434	CG	3.37
LEU	383	O	GLN	431	CG	3.19
GLY	384	N	GLN	431	CB	3.78
GLU	387	OE2	GLY	440	O	3.29
GLU	387	OE2	SER	441	O	3.63
GLU	387	CD	ARG	435	NH2	3.31
GLU	387	OE2	ARG	445	NH2	2.57
HIS	390	CE1	GLY	442	O	3.41
HIS	390	CE1	SER	441	O	3.14
HIS	390	CD2	PRO	443	CG	3.62
GLN	431	CA	ALA	382	O	3.57
GLN	431	CB	GLY	384	N	3.78
GLN	431	CG	LEU	383	O	3.19
GLN	431	NE2	GLN	381	OE1	3.04
GLU	434	CG	LEU	383	CD2	3.37
GLU	434	CB	ALA	382	O	3.62
ARG	435	NE	LEU	383	CD1	3.76
ARG	435	NH2	GLU	387	CD	3.31
GLY	440	O	GLU	387	OE2	3.29
SER	441	O	GLU	387	OE2	3.63
SER	441	OG	GLY	601	C	3.9
SER	441	O	HIS	390	CE1	3.14
GLY	442	O	HIS	390	CE1	3.41
PRO	443	CG	HIS	390	CD2	3.62
PRO	443	CG	ARG	645	O	3.89
PRO	443	CD	ASN	646	OD1	3.68
ARG	445	NH2	GLU	387	OE2	2.57
GLY	601	C	SER	441	OG	3.9
ARG	645	O	PRO	443	CG	3.89
ASN	646	OD1	PRO	443	CD	3.68

b

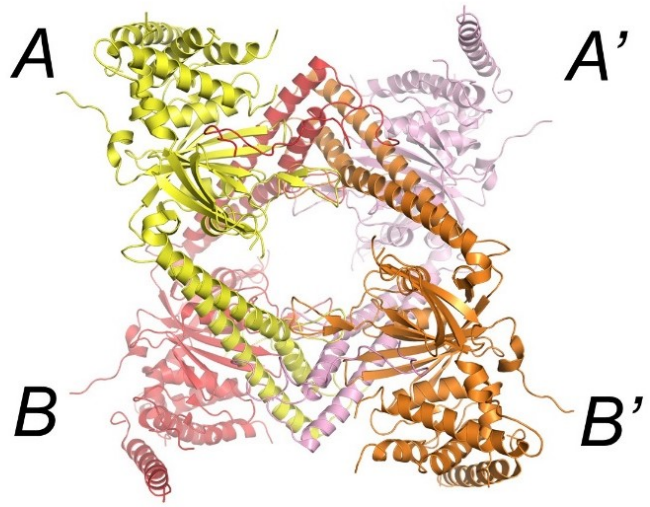
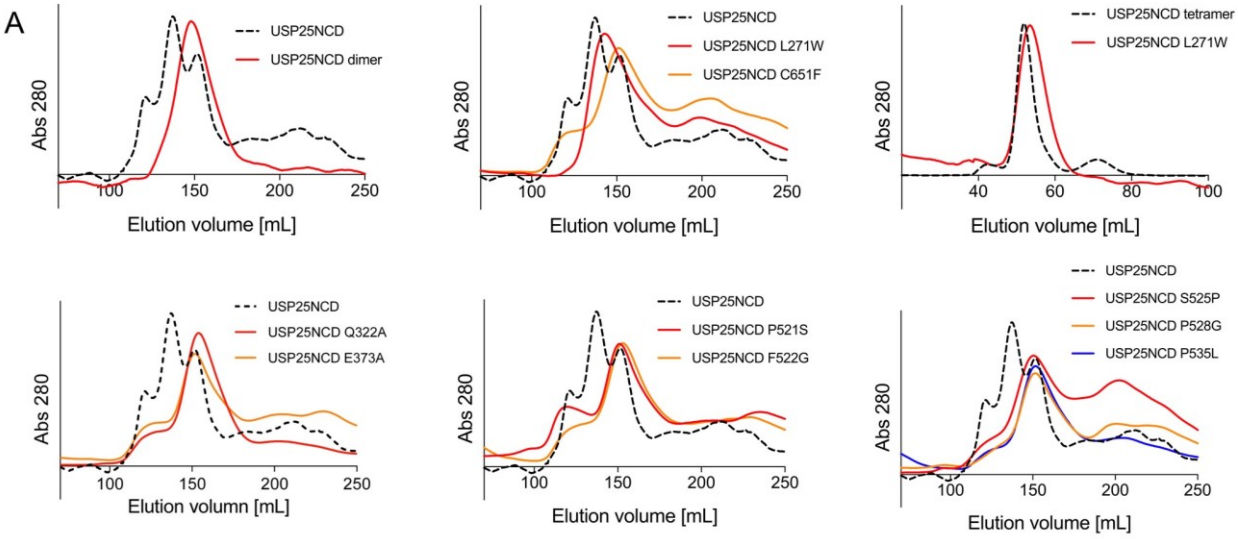
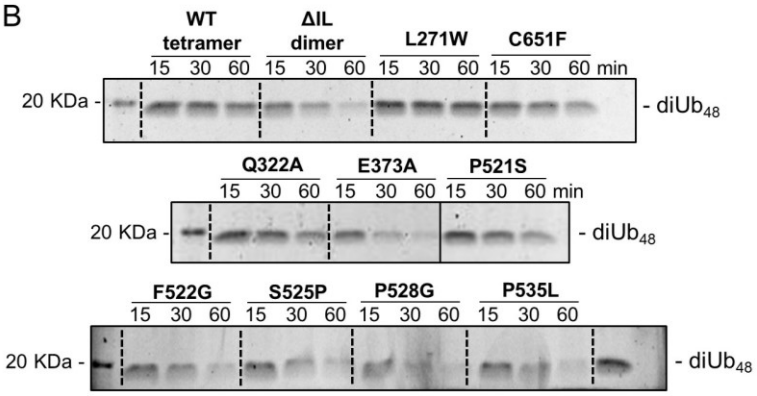
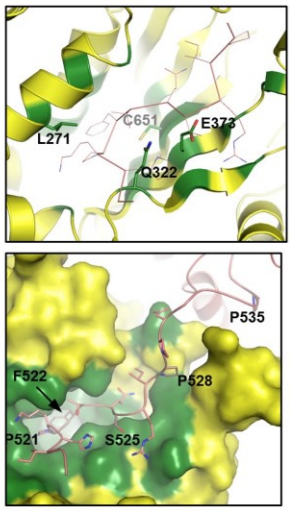
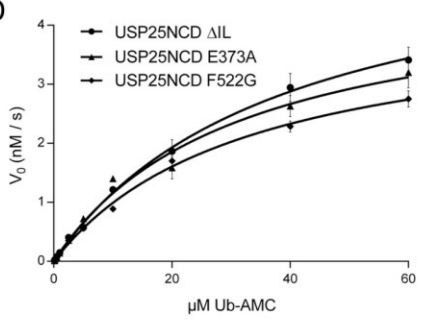


Figure S5. Contact list of residues involved in the USP25 tetramer assembly, cutoff 4.0 Å. (a) List of residues with interatomic distances less than 4.0 Å are shown. The interaction interface residues are generated by the online tool COCOMAPS. The interfaces between chains indicate tetramer, "*Il-loop*" and "*Kink*" (chains BA and BA', respectively) and dimer (chains BB') interactions. (b) Overall structure of the USP25 tetramer depicting its four chains labeled as A, A', B, and B'.









USP25NCD mutants	V_{max} (nM/s)	K_M (μM)	K_{cat} (s^{-1})	K_{cat}/K_M ($M^{-1}\ s^{-1}10^5$)
E373A	4.73 ± 0.37	31.01 ± 5.16	0.95 ± 0.07	0.305
F522G	4.23 ± 0.17	32.79 ± 2.76	0.85 ± 0.03	0.259
ΔIL dimer	5.66 ± 0.31	38.68 ± 4.20	1.13 ± 0.06	0.293

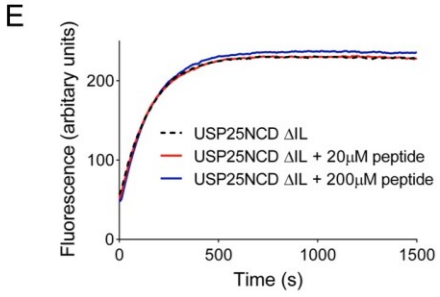
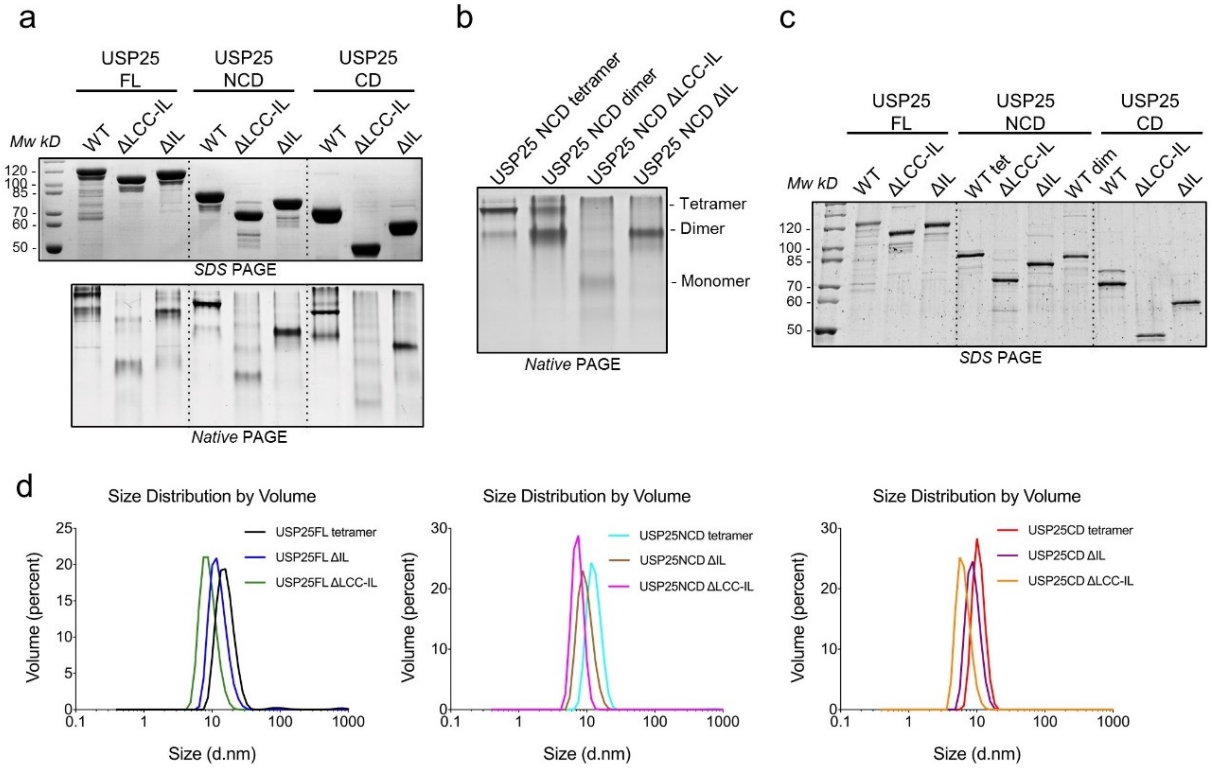


Figure S6. Oligomerization and de-ubiquitination analysis of the USP25 point mutants of the IL-loop interface.

(A) Gel filtration profiles of the purifications of the different USP25 NCD point mutants of the interface between the IL-loop and the S1 ubiquitin-binding surface (see also Ub-AMC activities in figure 4e). (B) SDS-PAGE of the time-course reaction of the deubiquitinating activities on K_{48}-linked diUb substrate with the different USP25 NCD point mutants of the interface. (C) ***Above***, cartoon representation of the IL-loop interaction, indicating in stick representation the point mutant residues in the USP-like surface. ***Below***, surface and stick representation of the IL-loop interaction. Point mutant residues are labeled and shown in stick representation. (D) ***Left***, plot of the comparison of the "Michaelis-Menten" curves for USP25 NCD ΔIL dimer and point mutants F522G and E373A. ***Right***, table of the "Michaelis-Menten" parameters of USP25 point mutants F522G, E373A and ΔIL dimer. (E) Effect of the synthetic IL-loop peptide, at 20 and 200 μM, on the deubiquitinating activity of the USP25 NCD ΔIL "constitutive" dimer.



USP25	PDI	Size by volume (d.nm)	Mean Count Rate (kcps)	Estimated MW (kDa)	MW (kDa)
FL tetramer	1.000	16.10 ± 4.64	233.4	443	489
FL ΔIL	0.401	12.67 ± 3.80	108.0	253	230
FL ΔLCC-IL	0.471	9.03 ± 2.55	125.9	115	104
NCD tetramer	1.000	12.98 ± 3.03	196.8	268	321
NCD ΔIL	0.672	9.66 ± 2.53	257.0	134	146
NCD ΔLCC-IL	0.551	7.44 ± 1.43	160.7	72.9	62
CD tetramer	1.000	10.80 ± 2.21	225.6	174	266
CD ΔIL	0.911	8.99 ± 2.11	166.7	114	118
CD ΔLCC-IL	0.866	6.46 ± 1.55	79.5	52.5	48

Figure S7. Gel pictures and dynamic light scattering of USP25 truncation constructs.
(a) SDS-PAGE (*above*) and native PAGE (below) of the different truncation constructs of USP25. (b) Native PAGE of the purified USP25 NCD tetramer and dimer, truncation monomer (USP25 NCD ΔLCC-IL), and truncation dimer (USP25 NCD ΔIL). Gels in A and B were stained with coomassie brilliant blue. (c) SDS-PAGE with the comparison of the amount of the USP25 truncation constructs used in the deubiquitinating activity assays in figure 5, stained with SYPRO-Ruby. (d) *Above*, size distribution by volume plots by dynamic light scattering of FL (right), NCD (middle) and CD (left) truncations of USP25. *Below*, table of the dynamic light scattering of the USP25 truncations.

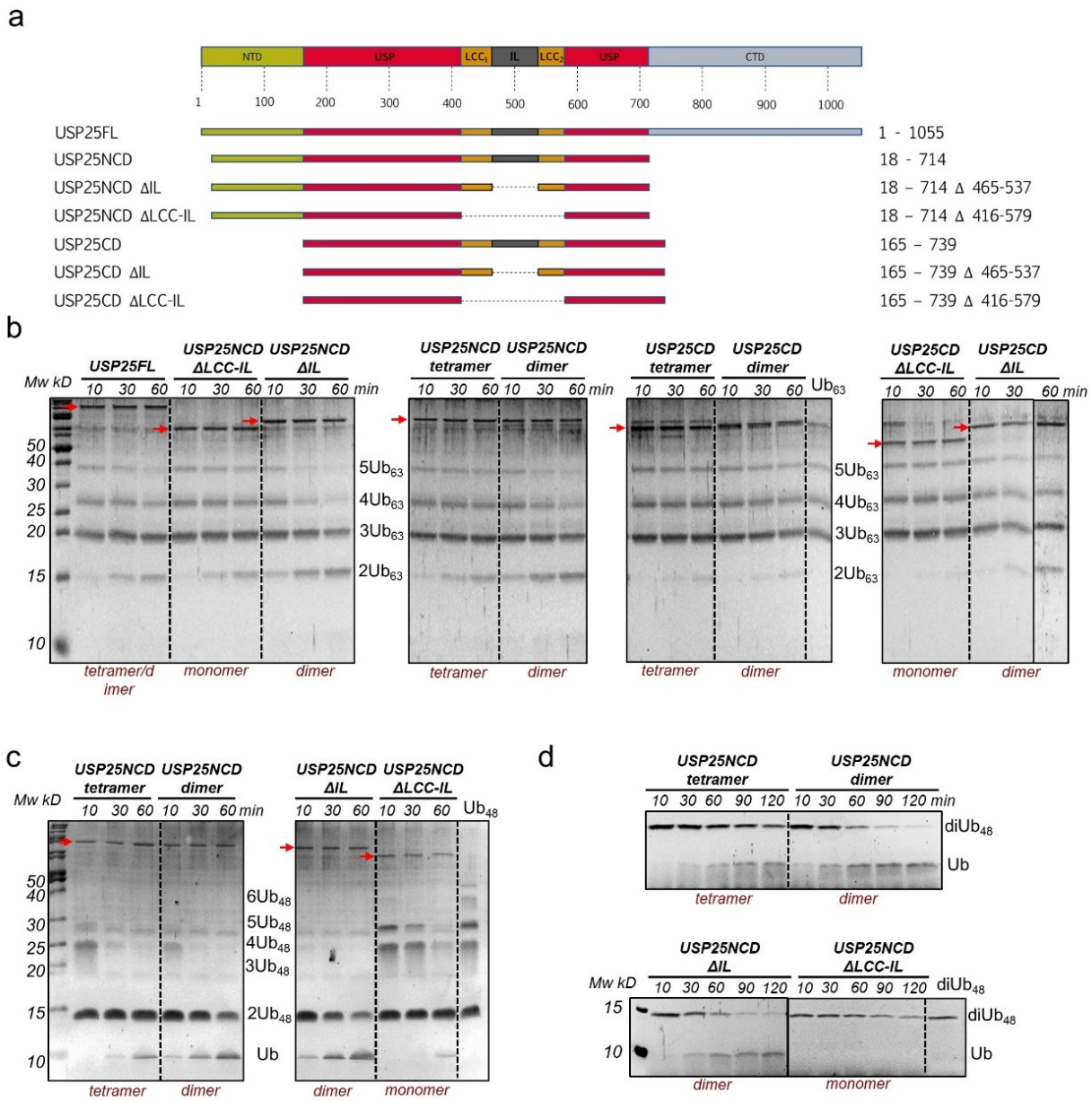
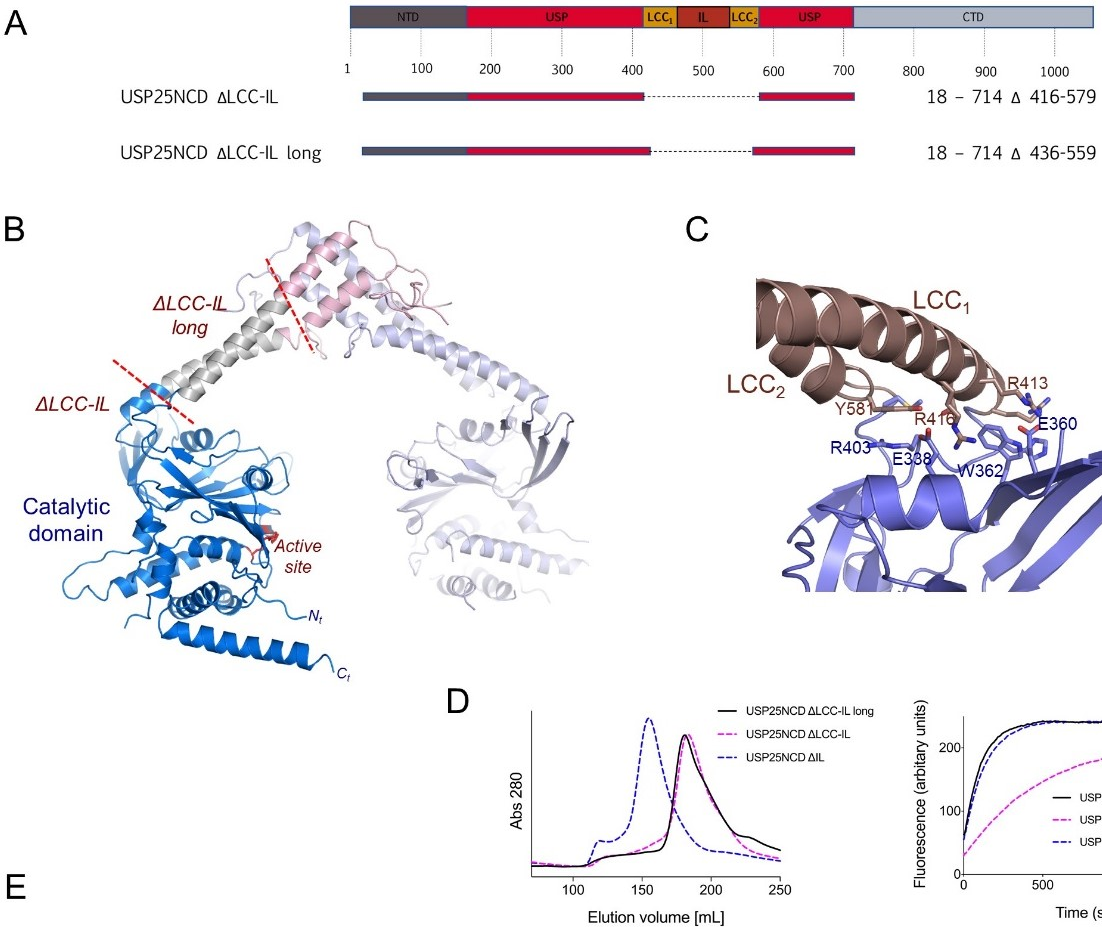


Figure S8. Activity assays of USP25 truncation constructs with polyubiquitin chains.

(a) Cartoon representation of the different truncation constructs of USP25. Left column indicate the names and the right column the residue range for each construct. (b) SDS-PAGE of the time course de-ubiquitinating assays with K_{63}-linked polyubiquitin substrate. Second column indicates the activities with purified tetramer and dimer USP25 fractions. Red arrows indicate USP25 truncation constructs. Reactions were stopped with SDS-loading buffer at indicated times. (c) SDS-PAGE of the time course de-ubiquitinating assays with K_{48}-linked poly-ubiquitin substrate. Left panels indicate the activities with purified tetramer and dimer USP25 fractions. Red arrows indicate USP25 truncation constructs. Reactions were stopped with SDS-loading buffer at indicated times. (d) SDS-PAGE of the time course de-ubiquitinating assays with K_{48}-linked diubiquitin substrate. Panels above indicate the activities with purified tetramer and dimer USP25 fractions.



USP25 NCD	PDI	Size by volume (d.nm)	Mean Count Rate (kcps)	Est. MW (kDa)	MW (kDa)
ΔLCC-IL long	0.260	7.54 ± 1.76	49.2	75.2	67
ΔLCC-IL	0.551	7.44 ± 1.43	160.7	72.9	62

Figure S9. Analysis of the two different USP25 monomer constructs.
(A) Scheme representation with boundaries for the two different monomer constructs of USP25. (B) Cartoon representation of the two different USP25 monomers, indicating the truncation of the long and short monomer. (C) Zoom-up view of the contacts between the coiled-coil LCC domain and the USP25 catalytic domain. Major contacts are labeled and shown in stick representation. (D) ***Left***, gel filtration purification profiles comparing the dimer (USP25 *ΔIL)* and the two different monomers constructs (USP25 *ΔLCC-IL and USP25 ΔLCC-IL long). **Right***, plots of the de-ubiquitinating activities using Ub-AMC for the dimer (USP25 *ΔIL)* and the two different monomers constructs (USP25 *ΔLCC-IL and USP25 ΔLCC-IL long).* (E) Dynamic-light scattering analysis of the two different monomer constructs (USP25 *ΔLCC-IL and ΔLCC-IL long).*

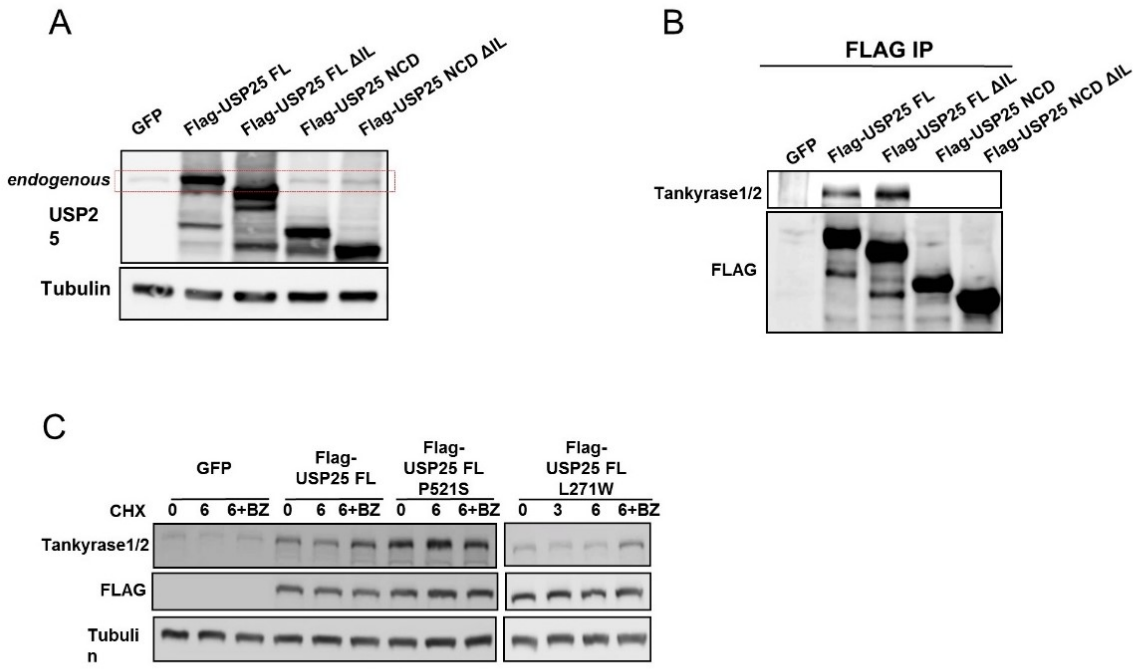


Figure S10. Endogenous USP25 levels and immunoprecipitation analysis.
(A) Flag-USP25FL, Flag-USP25FL ΔIL, Flag-USP25NCD and Flag-USP25NCD ΔIL were transfected in HEK293T cells and the levels of endogenous USP25 were analyzed by western-blot. GFP was transfected as a control. Dashed red rectangle indicates endogenous USP25. (B) Immunoprecipitation with anti-Flag resin of the HEK293T cells transfected with Flag-USP25FL, Flag-USP25FL ΔIL, Flag-USP25NCD and Flag-USP25NCD ΔIL plasmids. Endogenous tankyrases1/2 were analyzed by western-blot. (C) HEK293T cells were transfected with Flag-USP25FL, Flag-USP25FL P521S, Flag-USP25FL L271W and GFP and cells were treated with 100 μg/ml of cyclohexamide (CHX) and/or 0.5 μM of bortezomyb (BZ) and collected at 6 hours for western-blotting.

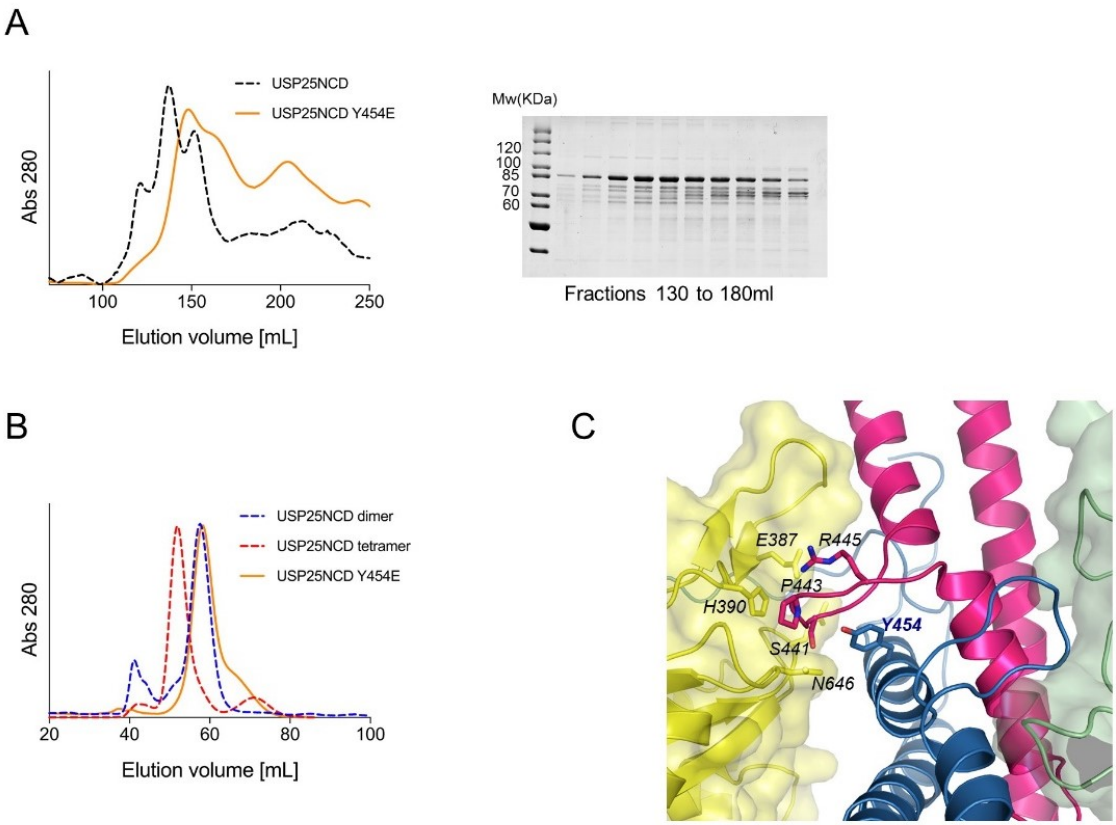


Figure S11. USP25 Y454E mutation results in the formation of a stable dimer.
(A) ***Left***, gel filtration chromatography of the purification of wild-type USP25 and USP25 Y454E constructs using Superdex200 column. ***Right***, SDS-PAGE of the fractions (5mL/fraction) of the gel filtration peak of USP25NCD Y454E. Gel stained with coomassie blue staining. (B) Analytical gel filtration chromatography analysis of the USP25NCD tetramer, USP25NCD dimer and USP25NCD Y454E using Superdex16 column. (C) Zoomed up representation of the contact residues around Tyr454. Stick model of interface residues of the "*kink*" motif and the catalytic USP-like domain.

Chapter 2

Structural analysis and evolution of specificity of the SUMO UFD E1-E2 interactions

Bing Liu[1], L. Maria Lois[2*] and David Reverter[1*]

[1]Institut de Biotecnologia i de Biomedicina. Departament de Bioquimica i Biologia Molecular. Serra Hunter Fellow. Universitat Autonoma de Barcelona, 08193 Barcelona, Bellaterra, Spain.
[2]Center for Research in Agricultural Genomics-CRAG. Edifici CRAG-Campus UAB, Bellaterra 08193 Barcelona, Spain.

*To whom the correspondence should be addressed: David Reverter, Universitat Autonòma de Barcelona, 08193 Bellaterra, Barcelona, Spain, Ph: +34 93 5868955, FAX: (93)5812011, E-mail: david.reverter@uab.cat and L. Maria Lois, Center for Research in Agricultural Genomics-CRAG, 08193 Bellaterra, Barcelona, Spain, Ph: +34 935636600 ext.3215, Fax. +34 93 5636601, email: maria.lois@cragenomica.es.

Abstract

SUMO belongs to the ubiquitin-like family (UbL) of protein modifiers. SUMO is conserved among eukaryotes and is essential for the regulation of processes such as DNA damage repair, transcription, DNA replication and mitosis. UbL modification of proteins occurs via a specific enzymatic cascade formed by the crosstalk between the E1-activating enzyme, the E2-conjugating enzyme and the E3-ligase. An essential discrimination step in all UbL modifiers corresponds to the interaction between E1 and E2 enzymes, which is mediated by the recruitment of the E2 to the UFD domain (ubiquitin-fold domain) of the E1 enzyme. To gain insights in the properties of this interface, we have compared the structures of the complexes between E1 UFD domain and E2 in human and yeast, revealing two alternative UFD platforms that interact with a conserved E2. Comparative sequence analysis of the E1 UFD domain indicates that the E2 binding region has been conserved across phylogenetic closely related species, in which higher sequence conservation can be found in the E2 binding region than in the entire UFD domain. These distinctive strategies for E1-E2 interactions through the UFD domain might be the consequence of a high selective pressure to ensure specificity of each modifier conjugation system.

Introduction

The post-translational modification pathway of proteins by UbLs (Ubiquitin-like modifiers) is characterized by the presence of specific enzymatic cascades (E1, E2 and E3 enzymes), which results in the formation of an isopeptidic bond between the UbL and the protein target[1,2]. A major characteristic of this process is the specificity provided between all components of the UbL pathway, resulting in the formation of protein-protein complementary interfaces[3,4]. SUMO (Small Ubiquitin-Like Modifier) is a UbL modifier that can alter the function of a myriad of target proteins inside the cell[5], being involved in processes such as DNA damage repair, transcription, DNA replication and mitosis[1,6,7].

The first specificity step in the pathway corresponds to the interaction of the UbL modifier with its particular E1-activating enzyme[8,9]. In the SUMO pathway, the E1-activating enzyme is a large multidomain heterodimer (Sae1-Uba2)[10] that initiates the process by adenylation of the SUMO C-terminus and the subsequently formation of a thioester bond with the active-site cysteine residue of E1[11]. Next, the activated UbL-thioester is transferred to the active-site cysteine residue of the E2-conjugating enzyme (Ubc9 in SUMO), which represents a second specificity step in the pathway by the formation of the E1-E2 complex. Crystal structures of several E1-activating enzymes[12-15] revealed the presence of a domain displaying an Ubiquitin-like Fold (UFD domain) in the E1-activating enzyme large subunit. The E1 UFD domain plays a major role in the binding of the E2 enzyme, providing specific contacts between E1 and E2 enzymes. This interaction was first observed in the crystal structure of the Nedd8 E1 in complex with E2[13], showing the direct binding of E2 to the E1 UFD domain. Recently, the crystal structure of the thio-ester transfer intermediate of ubiquitin E1-E2 complex[15] revealed a dual binding of E2 to the UFD domain and to the catalytic E1 Cys-domain, which occurs after an significant rotation of the UFD domain, providing the structural basis for the isoenergetic thio-ester transfer between the E1 and the E2 enzymes[15]. This interaction between E2 and the Cys domain of the E1 was proposed previously in the SUMO pathway by NMR analyses, although E1 UFD-E2 interactions display higher affinity (K_d = 1.2 μM)[16] than E1 Cys-E2 interactions (K_d = 87 μM)[17], supporting a major role of the E1 UFD domain in E2 recruitment.

Protein sequence variations in the E1 UFD domain of different UbL modifiers are quite significant, especially in the binding region to the E2 enzyme. All reported structures of UFD domains display an analogous β-grasp structure, and the interaction of the UFD domain with the E2-enzyme occurs through the same side of the β-sheet structure[13-15,18-20]. However, in all reported complex structures superposition of the E2 enzymes reveal distinct orientations of the UFD domain, which is a direct consequence of different contacts in each UbL system[15,18]. Conservation analysis according to sequence alignments showed that yeast and human SUMO UFD domains display little sequence homology (17% sequence identity), and it is even lower considering only the binding region to Ubc9. Notwithstanding this low conservation, both proteins can efficiently interact with a highly conserved surface in their cognate Ubc9. In this scenario, the identification of the molecular determinants that mediate E1 UFD and E2 interactions in evolutionary distant organisms cannot rely on sequence homology analysis. Instead, the elucidation of these molecular determinants requires specific structural studies of the interaction.

Here we present a detailed structural comparison analysis of the two complexes between SUMO E1 UFD domain and Ubc9 from yeast[21] and human. We also present a novel structure of the human complex solved in a different space group than the recently deposited[22]. Our results indicate that human and yeast UFD domains interact with a conserved surface of Ubc9, in each case by maintaining the same chemical character of the interface contacts despite the lack of sequence homology. Sequence alignment of these two E2 binding region discloses unique consensus motifs that have been maintained across species from the same kingdom (in *Metazoa*) or in the same order (in *Saccharomycetales*). Phylogenetic and homology analysis revealed that the region involved in Ubc9 binding displays a slightly higher conservation degree than the UFD domain between phylogenetically closely related organisms, although it also displays higher variability, highlighting the relevance of this interface in the protein-protein specificity for each type of UbL modification.

Results

Complex between human SUMO E1 UFD domain and SUMO E2

The interaction between the ubiquitin-fold domain (UFD) of the E1-activating enzyme and the E2-conjugating enzyme has been revealed as a crucial discrimination step in the conjugation pathway of UbL modifiers[14,18,23,24]. The interface between E1 and E2 enzymes is unique and is required to confer specificity between cognate enzymes of each UbL conjugation pathway. Notably, differences in the interface are more significant in the UFD domain than in the E2-conjugating enzyme, which is a highly conserved enzyme. Within the same UbL family, evolutionary distant species display a low degree of sequence conservation between E1 UFD domains. In SUMO pathway, yeast and human UFD domains were shown to display only a 17% sequence identity[25]. To get structural insights for the different interaction between yeast and human, we have determined the crystal structure of the complex between human E1 Uba2 UFD domain and human E2 Ubc9 at 2.2 Å resolution and compared this structure to the analogous complex in yeast (PDB code 3ONG)[25]. During the preparation of the manuscript, another structure of the human E1 UFD-Ubc9 was also published in a different crystallographic space group, supporting our results[22] (PDB ID 4WSV).

Human E1 ubiquitin-fold domain (UFD) was designed based on the structure of the full-length human SAE1-SAE2 E1-activating enzyme (PDB code 1Y8Q)[24]. We alternatively prepared a UFD construct including the C-terminal flexible extension, however this longer UFD was unstable and displayed proteolysis after complex formation with Ubc9. Both native gel electrophoresis and gel filtration chromatography indicated the formation of the complex between human E1 UFD domain and Ubc9 (see Supplementary Fig. S1). After initial unfruitful crystallization trials with the purified complex, a lysine methylation protocol was conducted to induce crystallization. Suitable crystals diffracted beyond 2.2 Å resolution, contained one complex of UFD/Ubc9 per asymmetric unit and belonged to the tetragonal space group system (Table 1), which is different to the monoclinic crystals recently deposited[22] (PDB code 5W5V).

Table 1. Summary of crystallographic analysis

Data collection	
Beamline	ALBA-XALOC
Space group	$P4_32_12$
Wavelength (Å)	0.97946
Resolution (Å)	46.44-2.20 (2.32-2.20)*
a, b, c (Å)	129.61, 129.61, 66.60
α, β, γ (°)	α = β = γ = 90
Unique reflections	29309
Data redundancy	10.9 (11.0)
R_{merge}	0.043 (0.602)
I/σ	26.4 (4.3)
Completeness (%)	99.8 (98.4)
Refinement	
Resolution (Å)	46.44 - 2.20
Unique reflections	29254
R_{work}/R_{free}	0.22/0.24
Number of all atoms	2100
Number of waters	38
RMSD bond (Å)/Angle (°)	0.008/1.11
Average B factor (protein/water)	60.69/56.26
Ramachandran plot	
Favored (%)	97.64
Allowed (%)	2.36
Disallowed (%)	0.00

* Highest resolution shell is shown in parenthesis.

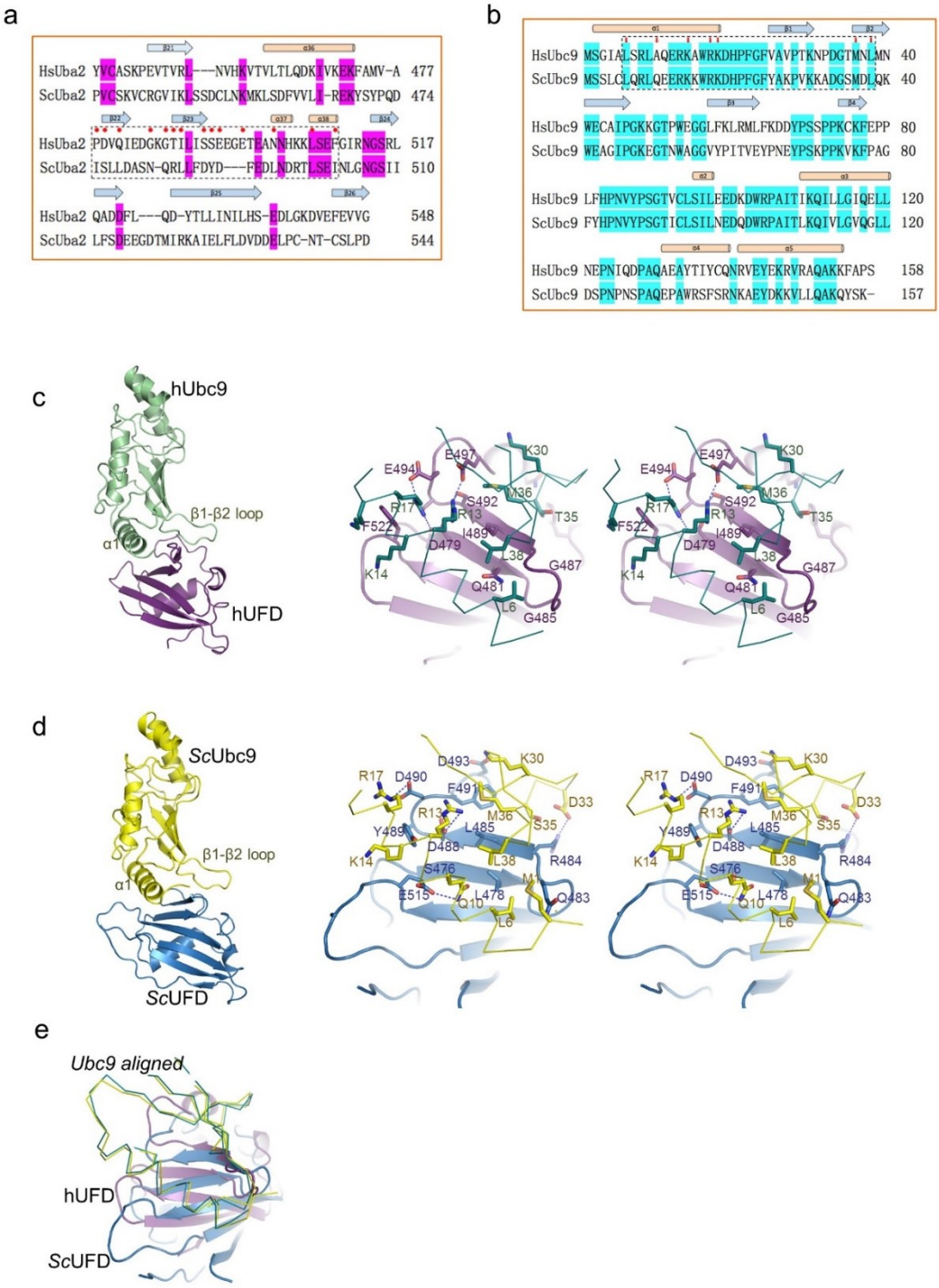


Figure 1. ***Structural alignment and comparison of the interfaces between UFD and Ubc9 from human and yeast complexes.*** (a) Structural alignment of the UFD domains from yeast (*ScUba2*) and human (*HsUba2*). Red circles indicate contact residues to Ubc9. Dotted rectangle represents the binding region to Ubc9. (b) Structural alignment of yeast and human Ubc9. Small red arrows indicate contact residues to UFD. Dotted rectangle represents the binding region to UFD. Secondary structure is depicted above sequence. (c) *Left*, ribbon representation of the complex of human Ubc9 and the UFD domain. *Right*, stereo representation of the interface residues between Ubc9 (green line) and UFD (purple ribbon). Major contacts are labeled and represented in stick configuration. (d) *Left*, ribbon representation of the complex of yeast Ubc9 and the UFD domain. *Right*, stereo representation of the interface residues between Ubc9 (yellow line) and UFD (blue ribbon). Major contacts are labeled and represented in stick configuration. (e) Structural superposition of Ubc9 in the human and yeast complex with UFD.

An overall structural comparison of free and bound Ubc9 structures displayed little variation (0.71 Å rmsd Cα deviation), although differences were observed in the binding region to UFD, displaying rmsd Cα deviations between 1.50 Å and 3.66 Å in the residues forming the β1-β2 loop (Lys30 to Met36). Similarly, free and bound E1 UFD structures are almost identical (overall 0.72 Å rmsd deviation). According to the PISA server[26] the human complex buries a surface of 1493 A^2 and involves 26 and 19 residues of UFD and Ubc9, respectively, which is comparable to the analogous complex in yeast (1557 A^2 interface)[25].

Yeast and human interface comparison

In contrast to the conservation proposed according to sequence homology between human and yeast UFD, 17% of identity, this homology is even lower, 11%, when a structural alignment is performed (Fig. 1a). Interestingly, differences between yeast and human are increased, 7% of sequence identity, when comparing only the residues forming the interface of the E1 UFD domain with Ubc9, which are basically formed by a different set of residues capable to interact with a conserved Ubc9 surface (Fig. 1b). However, despite this low sequence homology between yeast and human UFD, the interaction occurs through the same surface[22,25], forming an interdigitated complex between the α1 helix and the β1-β2 loop of Ubc9 that sits on the β-sheet surface of UFD (Fig. 1c). As a consequence of this low conservation, structural superposition of Ubc9 reveals a rotation of the E1 UFD domain between both complexes (Fig. 1e). Different orientations of UFD domains in complex with E2 can also be observed in other UbL systems, as revealed by the structures of ubiquitin and Nedd8 UFD in complex with E2 (Uba1-Ubc4 complex PDB code 4II2; and Uba3-Ubc12, pdb code 2NVU,[13,15]). All these structures suggest plasticity in the UFD interface that has evolved to specifically interact with its cognate E2-conjugating enzyme in each UbL pathway.

The binding surface of Ubc9 in yeast and human is highly conserved, and is basically formed by similar backbone and side-chain interactions in the α1 helix and the β1-β2 loop of Ubc9 (yeast and human Ubc9 share 56% sequence identity). Specific side-chain interactions in Ubc9 include a hydrophobic patch formed by Leu6, Met36, and Leu38, and a basic patch formed by Arg13, Lys14, Arg17 and Lys18 (Fig. 2a, 2b). Interestingly, all these conserved residues in Ubc9 engage

specific contacts with the non-conserved UFD surfaces of yeast and human. A major contact difference in Ubc9 corresponds to Ala10, which is substituted by Gln10 in yeast, forming a polar interaction with Glu515 of the yeast UFD domain (Fig. 1d). Previous point mutational analysis on these two patches of Ubc9, namely the basic α1 helix and the hydrophobic β1-β2 loop, showed impairment in the binding to the E1 activating enzyme, indicating a major role of these two regions in the transfer of SUMO between E1 and E2 proteins[25,27].

The UFD interacting surface is extended and mostly formed by residues emanating from β22 and β23 strands and connecting loops (Fig. 1c, 1d). A general feature of this interaction, widespread so far in all characterized UbL pathways, is the interaction through the same β-sheet surface (Fig. 2). But in contrast to the Ubc9, the region involved in UFD is poorly conserved between yeast and human. For simplicity, we have designated this region as *Low Homology region involved in E2 Binding 2* (LHEB2). The LHEB2 can be divided in two regions, each establishing interactions with the basic and hydrophobic patches of Ubc9 (Fig. 2a, 2b). The first contact region in human UFD is composed by Asp479, Ser492, Ser493, Glu494 and Glu497, which engage polar and charged contacts with the basic patch of Ubc9, composed by α1 helix residues Arg13 and Arg17. The LHEB2 in human is highly conserved in all *metazoan* species analyzed (see later in Fig. 3). In contrast, in yeast the basic patch of Ubc9 interacts with Asp488, Tyr489 and Asp490, which is also a highly conserved sequence in all *Saccharomycetales* species analyzed (see later in Fig. 3). Interestingly, in the human structure Phe522 is buried in an aliphatic pocket formed by Arg17, Lys14 and Lys18 (see Fig. 1c), whereas in yeast a similar Ubc9 pocket buries Tyr489 (see Fig. 1d), a residue located at the center of the LHEB2 sequence instead of the LHEB2 C-terminal position that occupies its human counterpart.

The second contact region of UFD interacts with the hydrophobic patch of Ubc9. This interface is composed by backbone and side chain interactions emanating from the β22-β23 connecting loop and β23 strand of UFD. In human Ubc9 Leu6 interacts with the β22-β23 loop formed by Gly485 and Gly487 (Fig. 1c). In contrast, in yeast, the composition and length of this loop is different and Ubc9 Leu6 interacts with Leu478 (Fig. 1d). In this region we can observe the highest structural homology between human and yeast. In human three backbone hydrogen bonds are formed

between Gly487 and Ile489 with Ubc9 Met36 and Asn37, by only two in yeast, between Leu485 and Ubc9 Met36. Additionally, the side chain of Ile489 in human (or Leu485 in yeast) is buried in both cases in the Ubc9 hydrophobic pocket formed by Met36 and Leu38 (Fig. 1c, 1d). Finally, it is worth mentioning that the specific contacts established by yeast Arg484 and Phe491 are absent in metazoan sequences but highly conserved in *Saccharomycetales*.

Interface comparison with other UbL E1-E2 complexes

Comparison of the UFD-E2 interfaces in ubiquitin and Nedd8 (Uba1-Ubc4 complex, PDB code 4II2; and Uba3-Ubc12, pdb code 1Y8X)[13,15] indicate that only the hydrophobic patch in Ubc9 is partially conserved in the ubiquitin E2 (Ubc4), formed in this instance by Leu3 and Leu30, which can be aligned with Ubc9 Leu6 and Leu38 (Fig. 2). However, the basic patch in the α1 helix of Ubc9, formed by Arg13, Lys14, Arg17 and Lys18, represents a specific feature of the SUMO pathway and are not present neither in Nedd8 (Ubc12) nor in ubiquitin (Ubc4), which are replaced by acidic and aliphatic residues. Ubc4 and Ubc12 also display a shorter β1-β2 loop compared to Ubc9 (Fig. 2). These differences in the E2-conjugating enzyme between UbLs modifiers result in the presence of non-complementary surfaces with E1 UFD domains, which are indeed the basis for the enzyme specificity among each UbL pathway. For instance, SUMO E1 UFD domain contains specific polar contacts (Ser492, Ser493, Glu494 and Glu497 in human or Asp488, Asp490 and Asp493 in yeast) to interact with the basic patch of Ubc9, however, in ubiquitin and Nedd8 these positions are substituted by aliphatic residues (Ala954 and Phe956 in ubiquitin or Leu415, Val418 and Ile421 in Nedd8) (Fig. 2).

As mentioned before, human Ile489 and yeast Leu485 adopt a similar conformation and engage identical backbone hydrogen bonds with the β1-β2 loop of Ubc9. Interestingly, despite the lack of sequence conservation, the equivalent residues in the ubiquitin UFD domain, namely Ser950-Leu951, also engage analogous backbone hydrogen bonds contacts with the β1-β2 loop of the E2 enzyme (Ubc4)[15]. Thus, this backbone interaction represents a unique conserved structural element maintained in distant UbL systems such as SUMO and ubiquitin. This interaction does not occur in Nedd8 (Ubc12-Uba3, pdb code 1Y8X,[13]), in which the E2 enzyme sits across a similar region of the UFD domain but with a different

angle compared to ubiquitin, and the contacts between both UbL systems are poorly conserved and thus not complementary[18,15].

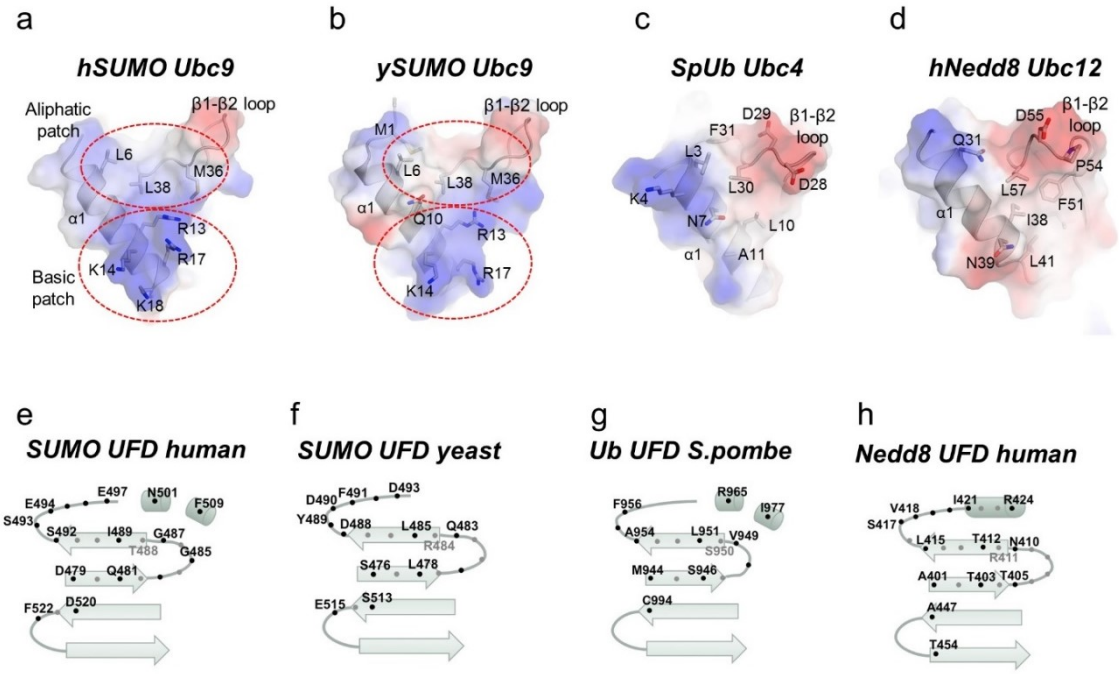


Figure 2. ***Comparison of the UFD-E2 interface from different UbL systems.*** (a) Transparent electrostatic representation of the interface of human SUMO E2 (*hSUMO Ubc9*) with the UFD domain. Major contacts are labeled and represented in stick configuration. Basic and aliphatic surface patches are indicated by dotted circles. (b) Transparent electrostatic representation of the interface of yeast SUMO E2 (*ySUMO Ubc9*) with the UFD domain. Major contacts are labeled and represented in stick configuration. Basic and aliphatic surface patches are indicated by dotted circles. (c) Transparent electrostatic representation of the interface of *S.pombe* ubiquitin E2 (*SpUb Ubc4*) with the UFD domain. Major contacts are labeled and represented in stick configuration. (d) Transparent electrostatic representation of the interface of human Nedd8 E2 (*hNedd8 Ubc12*) with the UFD domain. Major contacts are labeled and represented in stick configuration. (e) Schematic representation of the human SUMO E1 UFD domain contacts with the E2 enzyme. (f) Schematic representation of the yeast SUMO E1 UFD domain contacts with the E2 enzyme. (g) Schematic representation of the *S.pombe* ubiquitin E1 UFD domain contacts with the E2 enzyme. (h) Schematic representation of the human Nedd8 E1 UFD domain contacts with the E2 enzyme. Black and grey spots indicate the orientation of the side chain in the structure regarding the β-sheet plane.

Evolutionary conservation of the SUMO E1 UFD interfaces

Structural comparison between human and yeast indicate that SUMO UFD domains are composed by a different set of residues that can engage productive interactions with a conserved E2 enzyme. In order to evaluate the biological relevance of these two alternative structural interfaces involved in E2 interactions, we have analyzed the conservation of the LHEB2 sequence of the E1 UFD domains across species. In human, the LHEB2 region is comprised by residues between Pro478 and Phe509 (Fig. 1a, 3e), whilst in yeast is composed by residues between

Pro472 and Ile502 (Fig. 1b, 3e). We searched for human E1 UFD domain orthologs at the EggNOG database and focused on metazoan and fungal species, for which structural information of the E1 UFD-E2 interactions is available. Since fungal E1 displayed a high level of sequence divergence, we focused on species belonging to the *Saccharomycetales* order, which include *Saccharomyces cerevisae*. The phylogenetic analyses of both domains, E1 UFD and LHEB2, show that clustering of the E1-UFD sequences in the phylogenetic tree reflects the taxonomical relationships of the species represented (Fig. 4a). On the contrary, when the same analysis was performed with the LHEB2 domain, tree distribution is not consistent with taxonomic lineages (Fig. 4b), suggesting that this region presents higher variability than the UFD domain where is contained.

In addition, we analyzed the distribution of homology between pairs. *Saccharomycetales* UFD or LHEB2 sequences were compared with the corresponding human (conservation to outlier) or yeast sequences (conservation within the group). Similarly, metazoan UFD or LHEB2 sequences were compared with the corresponding yeast (conservation to outlier) or human sequences (conservation within the group). The homology pair distributions were plotted onto box plots (Supplemental Fig. S2). In general, when sequences were compared with an outlier, in all groups except in *Saccharomycetales*, the median of the obtained distribution was higher in the LHEB2 sequence analysis than in the UFD. At the same time, the box length, whose limits indicate the 25th and 75th percentiles, is higher for the LHEB2 than for the UFD homology pair distributions, suggesting that this region also presents higher variability. These results support the phylogenetic analyses indicating that, in general, the LHEB2 sequence is more conserved than the UFD domain within each evolutionary group, although the individual sequences contained in each group display higher variability. Similar results were obtained when sequences were compared with a reference sequence within each group, *Saccharomycetales* or metazoan.

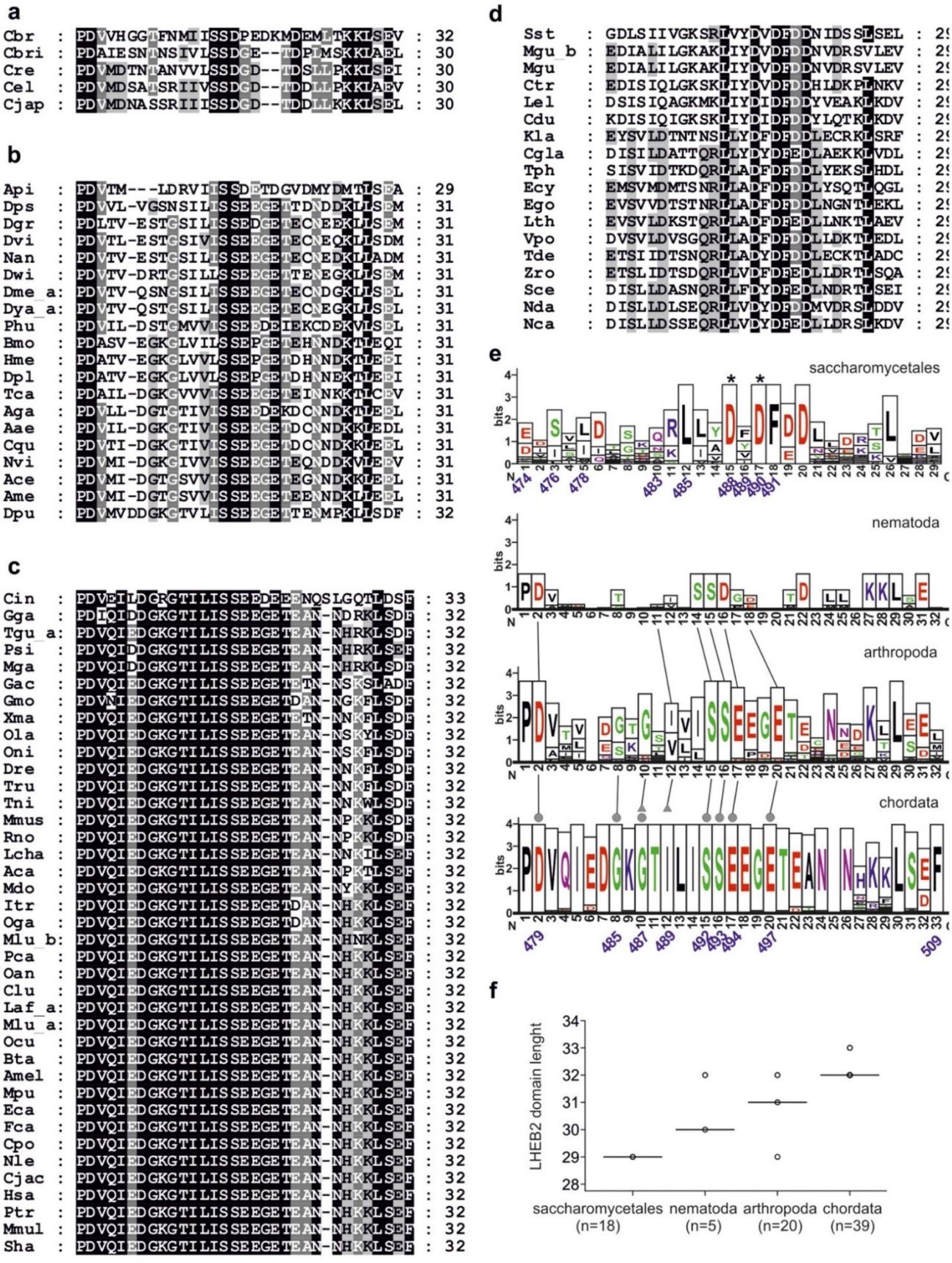

a
Cbr : PDVVHGGTFNMIISSDPEDKMDEMLTKKLSEV : 32
Cbri : PDAIESNTNSIVLSSDGE--TDPLMSKKLAEL : 30
Cre : PDVMDTNTANVVLSSDGD--TDSLLPKKLSEI : 30
Cel : PDVMDSATSRIIVSSDGD--TDDLLPKKLAEV : 30
Cjap : PDVMDNASSRIIISSDGD--TDDLLKKKLSEL : 30
b
Api : PDVTM---LDRVIISSDETDGVDMYDMTLSEA : 29
Dps : PDVVL-VGSNSILISSEEGETTDNDDKLLSEM : 31
Dgr : PDLTV-ESTGSILISSEDGETECNEEKLLSEM : 31
Dvi : PDVTL-ESTGSIVISSEEGETECNEQKLLSDM : 31
Nan : PDVTV-ESTGSILISSEEGETECNEDKLLADM : 31
Dwi : PDVTV-DRTGSILLSSEEGETTENEGKLLSEM : 31
Dme_a: PDVTV-QSNGSILISSEEGETECNDGKLLSEL : 31
Dya_a: PDVTV-QSTGSILISSEEGETECNEGKLLSEL : 31
Phu : PDVIL-DSTGMVVISSEEDEIEKCDEKVLSEL : 31
Bmo : PDASV-EGKGLVILSSEPGETEHNNDKTLEQI : 31
Hme : PDATV-EGKGLVVLSSEPGETDHNNDKTLEEI : 31
Dpl : PDATV-EGKGLVVLSSEPGETDHNNEKTLEEI : 31
Tca : PDAIL-DGKGVVVISSEEGETEINNKKTLCEV : 31
Aga : PDVLL-DGTGTIVISSEEDEKDCNNDKTLEEL : 31
Aae : PDVIL-DGKGTIVISSEEGETDCNNDKKLEDL : 31
Cqu : PDVTI-DGKGTIVISSEEGETDCNNDKKLEEL : 31
Nvi : PDVMI-DGKGIVVISSEEGETEQNDDKVLEEV : 31
Ace : PDVMI-DGTGSVVISSEEGETEGNNDKKLEEL : 31
Ame : PDVMI-DGTGSVVISSEEGETEENNDKLLEEL : 31
Dpu : PDVMVDDGKGTVLISSEEGETTENMPKLLSDF : 32
c
Cin : PDVEILDGRGTILISSEEDEEENQSLGQTLDSF : 33
Gga : PDIQIDDGKGTILISSEEGETEAN-NDRKLSDF : 32
Tgu_a: PDVQIEDGKGTILISSEEGETEAN-NHRKLSDF : 32
Psi : PDVQIDDGKGTILISSEEGETEAN-NHRKLSEF : 32
Mga : PDVQIDDGKGTILISSEEGETEAN-NHRKLSDF : 32
Gac : PDVQIEDGKGTILISSEEGETETN-NSKSLADF : 32
Gmo : PDVNIEDGKGTILISSEEGETDAN-NGKFLSDF : 32
Xma : PDVQIEDGKGTILISSEEGETETN-NNKFLSDF : 32
Ola : PDVQIEDGKGTILISSEEGETEAN-NSKYLSDF : 32
Oni : PDVQIEDGKGTILISSEEGETEAN-NSKFLSDF : 32
Dre : PDVQIEDGKGTILISSEEGETEAN-NNKFLSDF : 32
Tru : PDVQIEDGKGTILISSEEGETEAN-NNKFLSDF : 32
Tni : PDVQIEDGKGTILISSEEGETEAN-NNKWLSDF : 32
Mmus : PDVQIEDGKGTILISSEEGETEAN-NPKKLSDF : 32
Rno : PDVQIEDGKGTILISSEEGETEAN-NPKKLSDF : 32
Lcha : PDVQIEDGKGTILISSEEGETEAN-NNKILSEF : 32
Aca : PDVQIEDGKGTILISSEEGETEAN-NPKTLSEF : 32
Mdo : PDVQIEDGKGTILISSEEGETEAN-NYKKLSEF : 32
Itr : PDVQIEDGKGTILISSEEGETDAN-NHKKLSEF : 32
Oga : PDVQIEDGKGTILISSEEGETDAN-NHKKLSEF : 32
Mlu_b: PDVQIEDGKGTILISSEEGETEAN-NHNKLSEF : 32
Pca : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Oan : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Clu : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Laf_a: PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Mlu_a: PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Ocu : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Bta : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Amel : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Mpu : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Eca : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Fca : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Cpo : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Nle : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Cjac : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Hsa : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Ptr : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Mmul : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
Sha : PDVQIEDGKGTILISSEEGETEAN-NHKKLSEF : 32
d
Sst : GDLSIIVGKSRIVYDVDFDDNIDSSLSEL : 29
Mgu_b : EDIALILGKAKIIYDVDFDDNVDRSVLEV : 29
Mgu : EDIALILGKAKIIYDVDFDDNVDRSVLEV : 29
Ctr : EDISIQLGKSKIIYDVDFDDHLDKPLNKV : 29
Lel : DSISIQAGKMKIIYDIDFDDYVEAKLKDV : 29
Cdu : KDISIQIGKSKIIYDIDFDDYLQTKLKDV : 29
Kla : EYSVLDTNTNSILYDFDFDDLECRKLSRF : 29
Cgla : DISILDATTQRILYDYDFEDLAEKKLVDL : 29
Tph : SISVIDTKDQRILADFDFDDLYEKSLHDL : 29
Ecy : EMSVMDMTSNRILADFDFDDLYSQTLQGL : 29
Ego : EVSVVDTSTNRILADFDFDDLNGNTLEKL : 29
Lth : EVSVLDKSTQRILADFDFEDLLNKTLAEV : 29
Vpo : DVSVLDVSGQRILADFDFDDLLDKTLEDL : 29
Tde : ETSIIDTSNQRILADYDFDDLECKTLADC : 29
Zro : ETSLIDTSDQRILVDFDFEDLLDRTLSQA : 29
Sce : DISLLDASNQRILFDYDFEDLNDRTLSEI : 29
Nda : DISLLDLSEQRILVDYDFDDMVDRSLDDV : 29
Nca : DISLLDSSEQRILVDYDFEDLLDRSLKDV : 29
e
saccharomycetales
nematoda
arthropoda
chordata
bits
f
LHEB2 domain lenght
saccharomycetales (n=18)
nematoda (n=5)
arthropoda (n=20)
chordata (n=39)

Figure 3. ***Conservation analysis of Sae2 LHEB2 domain in metazoan and Saccharomycetales*****.** Amino acid sequence alignment of Sae2 LHEB2 domain orthologs from nematoda (a), arthopoda (b), chordate (c) and saccharomycetales (d). Residue shading correspond to 95% (white letter and dark background), 75% (white letter and gray background), and 55% (black letter and light gray background) of sequence identity in (a),(b) and (c). In the case of saccharomycetales, the shading types correspond to 90%, 70% and 50% of sequence identity, respectively. Metazoan multiple sequence alignments of LHEB2 sequences were performed using Clustal Omega software and including yeast Uba2 (NP_010678) as outlier. Saccharomycetales multiple sequence alignments of LHEB2 sequences were performed using Muscle software and including human Sae2 (Q9UBT2) as outlier. (e) Graphical representation of LHEB2 domain consensus sequences determined from amino acid sequence alignments shown in (a), (b), (c) and (d). The overall height of the stack indicates the sequence conservation at that position, while the symbol height within the stack indicates the relative frequency of each amino acid within that position. The positions of yeast Sae2 residues involved in Ubc9 interaction according to the previously resolved structure (3ONG) are indicated in blue below the sequences graph. Asterisks indicate residues shown to have a major contribution to E1-E2 interactions in mutagenesis analysis[25]. The positions of human Sae2 residues involved in Ubc9 interactions according to the resolved structure are indicated in blue below the chordata consensus sequence graph. Grey circles indicate residues establishing contacts with Ubc9 1-helix, while grey triangles indicate residues interacting with Ubc9 residues located at the Ubc9 1 2-loop. Conserved residues across phyla are indicated by lines. (f) Distribution of LHEB2 sequence length displayed by orthologs within each phylogenetic group analyzed was plotted on a box plot graph. Data points are represented by circles. Outliers are represented by dots. The number of data points analyzed in each phylogenetic group is indicated below the x-axis.

The homology analysis of the residues involved in Ubc9 contacts are even higher conserved and minor differences were identified among phyla. In chordates, the highest conserved group, all the contacts described in the human structure are present in all species analyzed within this phylum (Fig. 3c). In arthropoda, sequence comparison also display little divergence, but in this case variations can be found in the composition of the β22-β23 connecting loop of UFD, presenting different loop lengths but still conserving Gly485 and Gly487 in many species of the phylum (Fig. 3b). In nematode, the five species analyzed also display conservation in the major E2 contact residues, but in contrast to chordate and arthropoda, the β12-β13 connecting loop displays little homology and Gly485 and Gly487 have been substituted (Fig. 3a).

Therefore, all major specific UFD contacts with Ubc9 are basically conserved in *metazoa*, including contacts with the hydrophobic and the basic patches of the Ubc9 surface. For instance, human Asp479, Ser492, Ser493, Glu494 and Glu497 (occasionally replaced by aspartic), which interact with Ubc9 Arg13 and Arg17, are highly conserved in all species analyzed. Similarly, the hydrophobic interaction of the human UFD Ile489, which interacts with Ubc9 Met36 and Leu38, is also conserved but can be occasionally substituted by valine in some species. The major differences in the UFD interface in *metazoa* are located in the β22-β23 connecting

loop, formed in human by Gly485 and Gly487, which interact with Ubc9 Leu6. Whereas in chordate this connecting loop is highly conserved, in arthropod and particularly in nematode, this loop displays different lengths and amino acid composition. It is worth mentioning here that the extension and composition of this particular loop in UFD domains is highly divergent between E1-activating enzymes specific to different UbL modifiers, suggesting that this domain has evolved to interact with its cognate E2-enzyme.

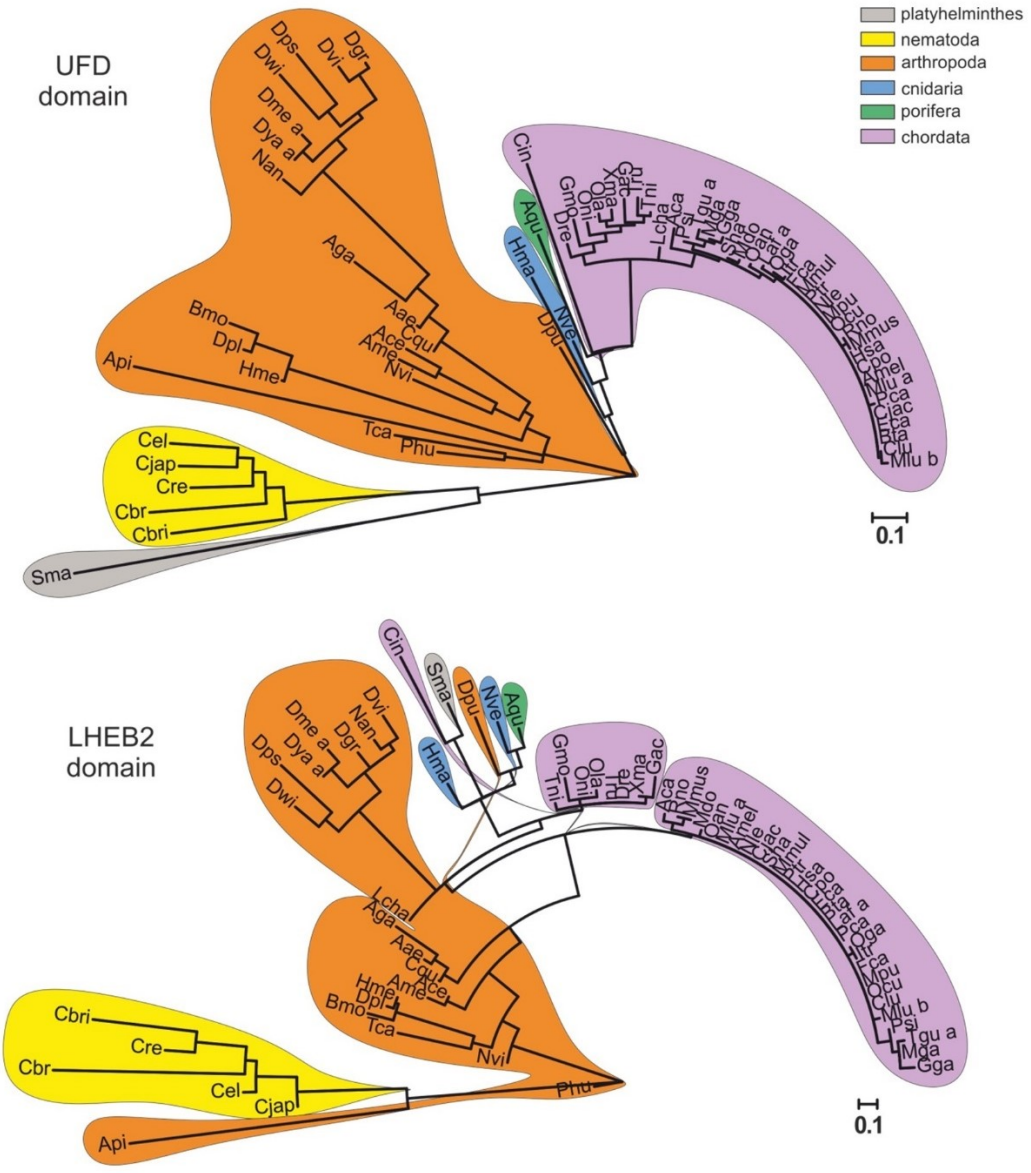


Figure 4. ***Phylogenetic analysis of UFD and LHEB2 domains from Metazoa and Saccharomycetales.*** Maximum likelihood phylogenetic trees depicting the evolutionary relationships among 86 Sae2 UFD (a) or Sae2 LHEB2 (b) domain sequences from 68 metazoa species species using sequence alignments shown in Fig. S3 (UFD) and Fig. 3 (LHEB2). Sequences belonging to the same phylum are enclosed in colored areas. Tree scales are shown below each tree.

In contrast to *metazoa*, sequence comparison of the E2 binding region of the yeast UFD domain with members of the *Fungi* kingdom is highly diverse, and further structural analyses will be required for establishing the molecular basis of E1-E2 interactions in those divergent groups. In the order of *Saccharomycetales*, the major specific contacts described in the structure of *S.cerevisiae* UFD-Ubc9 complex[25] are conserved. In this instance, as shown before in the structural comparison, the consensus binding sequence in *Saccharomycetales* is completely different to *metazoan*. Yeast UFD residues Asp488, Asp490 and Asp493, which interact with the basic patch of Ubc9, are highly conserved in *Saccharomycetales* (Fig. 3d, 3e), as well as Leu478, Leu485 and Phe491, which interact with the hydrophobic patch of Ubc9. Other specific contacts in the structure, such as Arg484 and Tyr489, are replaced by residues with similar chemical properties, such as lysine for Arg484 and phenylalanine, isoleucine or valine for Tyr489 (Fig. 3d, 3e).

During the essential SUMO conjugation pathway, protein-protein interactions have evolved to maintain specificity of the modifier and the targets. The SUMO conjugating enzyme constitutes the link between modifier specificity, which is selected by the E1 activating enzyme, and the protein substrate specificity, mediated by the cooperation between the E2 and the E3 ligase enzymes. In the E1-E2 interactions, a region in the E1 UFD domain, the LHEB2 sequence, plays a major role in E2 recruitment to the E1. In evolutionary distant groups, the LHEB2 sequences are poorly conserved according to sequence homology and length. On the contrary, in closely related phylogenetic groups, the conservation of the LHEB2 sequence is higher, highlighting the relevance of this interaction between E1 and E2 structures. We speculate that these distinctive strategies for E1-E2 interactions through the E1 UFD domain are the consequence of a high selective pressure to ensure modifier specificity. Future structural analyses of other evolutionary distant groups, such as plants or protozoa, will most probably uncover novel molecular determinants mediating E1-E2 interactions.

Conclusions

In summary, structural comparison of yeast and human SUMO UFD-E2 complexes and sequence alignment of SUMO UFD domains, reveal the presence of at least two complementary types of interfaces, which are conserved across species from the same kingdom (*metazoan*) or in the case of the *Fungi* kingdom, in the same order (*Saccharomycetales*). Despite the low level of sequence homology in the UFD domains among these distant species of different kingdoms, these two types of interfaces maintain the structural and chemical properties necessary to interact with a conserved E2 binding surface. Structural and sequential comparisons have also revealed at least two different types of consensus sequences in the E1 UFD domain, which we named LHEB2 sequences, that can complement the conserved surface in the E2 enzyme. Interestingly, sequence conservation in the E2 binding region is higher than in the overall UFD domain, suggesting the presence of an evolutionary pressure to maintain the contacts with the E2-conjugating enzyme, which are essential for the correct function of each UbL pathway.

Materials and methods

Protein expression and purification

Expression constructs were generated by a standard PCR-based cloning method. The full length human Ubc9 and E1 UFD domain (residues 447-547) were cloned to pET28a tagged with 6×His at the N-terminal. *Escherichia coli* BL21(DE3) plysS containing the expression vector were grown in Luria Bertani medium with chloramphenicol (17 μg/mL) and kanamycine (50 μg/mL) at 37 °C until the OD600 reached to 0.8. Expression was induced by 0.1mM IPTG, followed by overnight culturing at 28 °C. Recombinant proteins were purified by nickel-nitrilotriacetic acid agarose resin (Qiagen) and dialyzed against 250 mM NaCl, 20 mM Tris-HCl (pH 8.0), 1 mM β-mercaptoethanol in the presence of thrombin protease overnight at 4 °C to remove the 6×His tag. Proteins were further purified by gel filtration chromatography on a Superdex75 column (GE Healthcare), which was pre-equilibrated in 250mM NaCl, 20mM Tris-HCl pH 7.5, 1mM β-mercaptoethanol.

Protein complex preparation and methylation

Ubc9 and UFD complex was made by mixing equimolar amounts of proteins and purified by gel filtration chromatography using a Superdex75 column. Ubc9 and UFD were co-eluted in a single peak and confirmed by SDS-PAGE. After gel filtration, the complex was dialyzed against 250mM NaCl, 50mM HEPES pH 7.5, 1mM β-mercaptoethanol for lysine methylation based on a published strategy[28]. In brief, borane-dimethylamine complex (Sigma-Aldrich) and formaldehyde (Sigma-Aldrich) were sequentially added into protein solution and incubated overnight at 4 °C. The methylation reaction was stopped by a final gel filtration chromatography on a Superdex75 column pre-equilibrated in 200mM NaCl, 20mM Tris-HCl (pH 7.5), 1mM β-mercaptoethanol. Purified protein complex was concentrated to 30 g/L using an Amicon Ultra-10K ultrafiltration device (Millipore) prior to crystallization.

Crystallization and data collection

Crystals were grown by the sitting-drop vapor diffusion method by mixing the protein complex (30 g/L) with an equal volume of reservoir solution containing 16% PEG6000 (w/v), 100mM MES pH 6.5, and 5% MPD (v/v), at 18 °C. Crystals

appeared after 24 hours and continued to grow to full size in one week. Big crystals were soaked in mother liquor supplemented with gradually increasing concentration of 5%, 10%, 20% (v/v) MPD for 60 seconds each time and flash frozen in liquid nitrogen. Diffraction data were collected to 2.20 Å resolution at ALBA synchrotron in Barcelona (BL13-XALOC beamline). The crystals belong to the space group $P4_32_12$ and the unit cell has a dimension of a = 129.61 Å, b = 129.61 Å, and c = 66.60 Å. Data were processed with XDS[29] and scaled, reduced, and further analyzed using CCP4[30]. More details are shown in Table 1.

Structure determination and refinement

The structure was determined by molecular replacement method using the full length human Ubc9 (protein data bank code 1U9B) as a search model for one molecular in the asymmetric unit in PHASER[31]. Initial electron density was manually improved to build up the final model using Coot[32], and the refinement was performed using Phenix[33]. Refinement statistics are shown in Table1. The structure has been deposited in the PDB data bank with the code 5FQ2.

Phylogenetic sequence comparison

Human Sae2 orthologs were search in EggNOG database (http://eggnogdb.embl.de/#/app/home) and sequences from *Metazoa* and *Saccharomycetales* were selected for homology analysis. Detailed information about sequence homology analysis methods is indicated in figure legends. Briefly, when sequences were highly divergent, multiple sequence alignments were performed using Muscle tool (http://www.ebi.ac.uk/Tools/msa/muscle/). When sequence displayed higher conservation level, homology analyses were performed using Clustal Omega program (http://www.clustal.org/omega/). Phylogenetic distances were calculated by the maximum likelihood method and the JTT model included in the Seaview v4 software package[34], and unrooted trees exported. Phylogenetic trees were drawn using the online iTOL software (http://itol.embl.de/). Consensus sequences were calculated using WebLogo software (http://weblogo.berkeley.edu/)[35]. Multiple sequence alignments were edited, analyzed and shaded using GeneDoc software[36] (http://iubio.bio.indiana.edu/soft/molbio/ibmpc/genedoc-readme.html). Data distribution was plotted on box plots using “BoxPlotR: a web-tool for generation of box plots”.

Acknowledgements

We thank Martí Bernardo for fruitful discussions. This work was supported by grants from the Ministerio de Economia y Competitividad BFU2015-66417-P (MINECO/FEDER) to DR, and from the Generalitat de Catalunya (2009SGR 09626) and from the European Commission (ERC-2014-PoC-671839) to LML. B.L. acknowledges his scholarship to the Chinese Research Council program from the Chinese government. X-ray experiments were performed at the BL-13 beamline at the ALBA synchrotron in collaboration with the ALBA staff.

Author Contributions

BL and DR conducted all crystallographic experiments presented in Figures 1 and 2. LML performed and analyzed the sequence alignments presented in Figures 3 and 4. DR and LML conceived the idea for the project, analyzed the results and wrote the paper.

Additional Information

The crystal structure of the complex was deposited in the PDB data bank with the code 5FQ2.

Conflict of interest

The authors declare no competing financial interests.

References

1 Johnson, E. S. Protein modification by SUMO. *Annual review of biochemistry* **73**, 355-382 (2004).

2 Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458**, 422-429, doi:10.1038/nature07958 (2009).

3 Schulman, B. A. & Harper, J. W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat. Rev. Mol. Cell. Biol.* **10**, 319-331 (2009).

4 Streich, F. C., Jr. & Lima, C. D. Structural and functional insights to ubiquitin-like protein conjugation. *Annual review of biophysics* **43**, 357-379, doi:10.1146/annurev-biophys-051013-022958 (2014).

5 Gareau, J. R. & Lima, C. D. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **11**, 861-871 (2010).

6 Geiss-Friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* **8**, 947-956 (2007).

7 Hay, R. T. SUMO: a history of modification. *Molecular cell* **18**, 1-12 (2005).

8 Walden, H. *et al.* The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. *Molecular cell* **12**, 1427-1437 (2003).

9 Castaño-Miquel, L., Seguí, J. & Lois, L. M. Distinctive properties of Arabidopsis SUMO paralogues support the in vivo predominant role of AtSUMO1/2 isoforms. *Biochemical Journal* **436**, 581-590 (2011).

10 Desterro, J. M., Rodriguez, M. S., Kemp, G. D. & Hay, R. T. Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *Journal of Biological Chemistry* **274**, 10618-10624 (1999).

11 Olsen, S. K., Capili, A. D., Lu, X., Tan, D. S. & Lima, C. D. Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature* **463**, 906-912 (2010).

12 Walden, H. *et al.* The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. *Molecular Cell* **12**, 1427-1437 (2003).

13 Huang, D. T. *et al.* Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. *Molecular Cell* **17**, 341-350 (2005).

14 Lee, I. & Schindelin, H. Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* **134**, 268-278 (2008).

15 Olsen, S. K. & Lima, C. D. Structure of a Ubiquitin E1-E2 Complex: Insights to E1-E2 Thioester Transfer. *Molecular Cell* **49**, 884-896 (2013).

16 Reiter, K. *et al.* Identification of Biochemically Distinct Properties of the Small Ubiquitin-related Modifier (SUMO) Conjugation Pathway in Plasmodium falciparum. *Journal of Biological Chemistry* **288**, 27724-27736, doi:10.1074/jbc.M113.498410 (2013).

17 Wang, J. *et al.* The intrinsic affinity between E2 and the Cys domain of E1 in ubiquitin-like modifications. *Molecular Cell* **27**, 228-237 (2007).

18 Huang, D. T. *et al.* Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. *Nature* **445**, 394-398 (2007).

19 Durfee, L. A., Kelley, M. L. & Huibregtse, J. M. The basis for selective E1-E2 interactions in the ISG15 conjugation system. *Journal of Biological Chemistry* **283**, 23895-23902 (2008).

20 Tokgöz, Z. *et al.* E1-E2 interactions in ubiquitin and Nedd8 ligation pathways. *Journal of Biological Chemistry* **287**, 311-321 (2012).

21 Wang, J. *et al.* Crystal Structure of UBA2 ufd-Ubc9: Insights into E1-E2 Interactions in Sumo Pathways. *PloS one* **5**, e15805 (2010).

22 Reiter, K. H. *et al.* Characterization and Structural Insights into Selective E1-E2 Interactions in the Human and Plasmodium falciparum SUMO Conjugation Systems. *Journal of Biological Chemistry*, doi:10.1074/jbc.M115.680801 (2015).

23 Walden, H., Podgorski, M. S. & Schulman, B. A. Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. *Nature* **422**, 330-334 (2003).

24 Lois, L. M. & Lima, C. D. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *The EMBO journal.* **24**, 439-451 (2005).

25 Wang, J. *et al.* Crystal Structure of UBA2ufd-Ubc9: Insights into E1-E2 Interactions in SUMO Pathways. *PloS one* **5**, e15805 (2010).

26 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774-797 (2007).

27 Bencsath, K. P., Podgorski, M. S., Pagala, V. R., Slaughter, C. A. & Schulman, B. A. Identification of a multifunctional binding site on Ubc9p required for Smt3p conjugation. *Journal of Biological Chemistry* **277**, 47938-47945 (2002).

28 Walter, T. S. *et al.* Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617-1622 (2006).

29 Kabsch, W. XDS. *Acta Crystallographica Section D* **66**, 125-132, doi:doi:10.1107/S0907444909047337 (2010).

30 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235-242 (2011).

31 Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallographica Section D: Biological Crystallography* **60**, 432-438 (2004).

32 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126-2132 (2004).

33 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213-221 (2010).

34 Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* **27**, 221-224, doi:10.1093/molbev/msp259 (2010).

35 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Research* **14**, 1188-1190 (2004).

36 Nicholas, K. B. & Nicholas, H. B. J. GeneDoc: a tool for editing and annotating multiple sequence alignments. (1997).

SUPPLEMENTARY FIGURES

Structural analysis and evolution of specificity of the SUMO UFD E1-E2 interactions

Bing Liu[1], L. Maria Lois[2]* and David Reverter[1]*

1Institut de Biotecnologia i de Biomedicina and Departamento de Bioquimica i Biologia Molecular, Universitat Autonoma de Barcelona, 08193 Barcelona, Bellaterra, Spain.

2Center for Research in Agricultural Genomics-CRAG. Edifici CRAG-Campus UAB, Bellaterra 08193 Barcelona, Spain

*To whom the correspondence should be addressed: David Reverter, Universitat Autonòma de Barcelona, 08193 Bellaterra, Barcelona, Spain, Ph: +34 93 5868955, FAX: (93)5812011, E-mail: david.reverter@uab.cat and L. Maria Lois, Center for Research in Agricultural Genomics-CRAG, 08193 Bellaterra, Barcelona, Spain, Ph: +34 93 5636600 ext.3215, Fax. +34 93 5636601, email: maria.lois@cragenomica.es

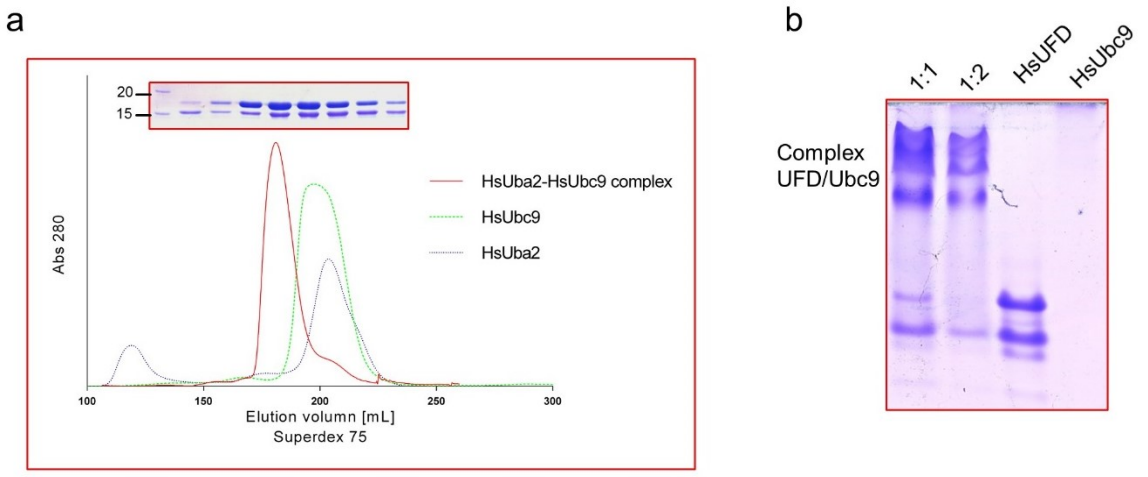


Figure S1.- *Purification of the complex between human E1 UFD domain and Ubc9.* (a) Size exclusion chromatography profiles of human Ubc9, human E1 UFD domain and the complex with a buffer containing 100mM NaCl, 20mM Tris 8, 1mM BME. *Inset*, SDS-PAGE of the peak fractions. (b) Native gel electrophoresis of the complex formation between human UFD domain and human Ubc

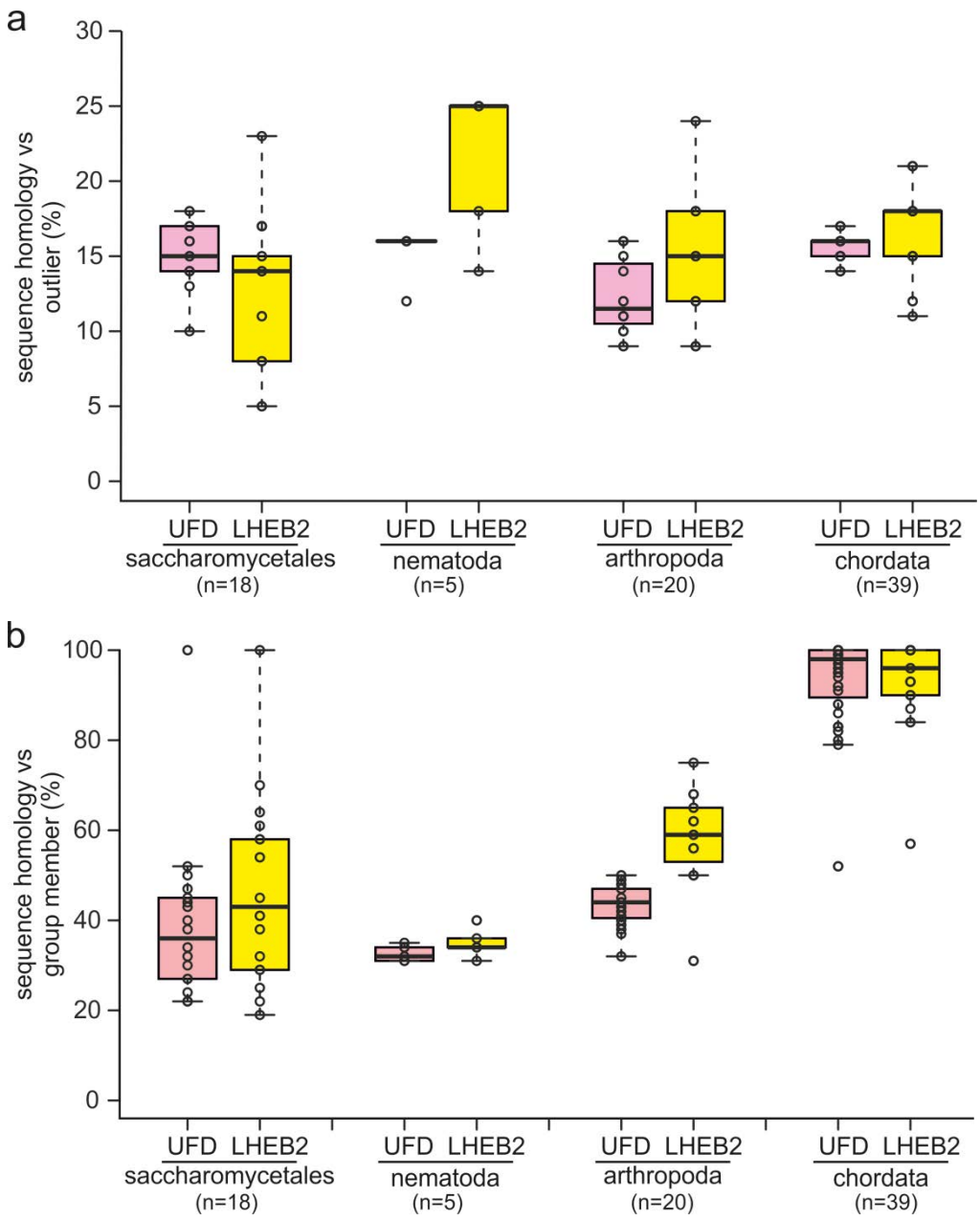
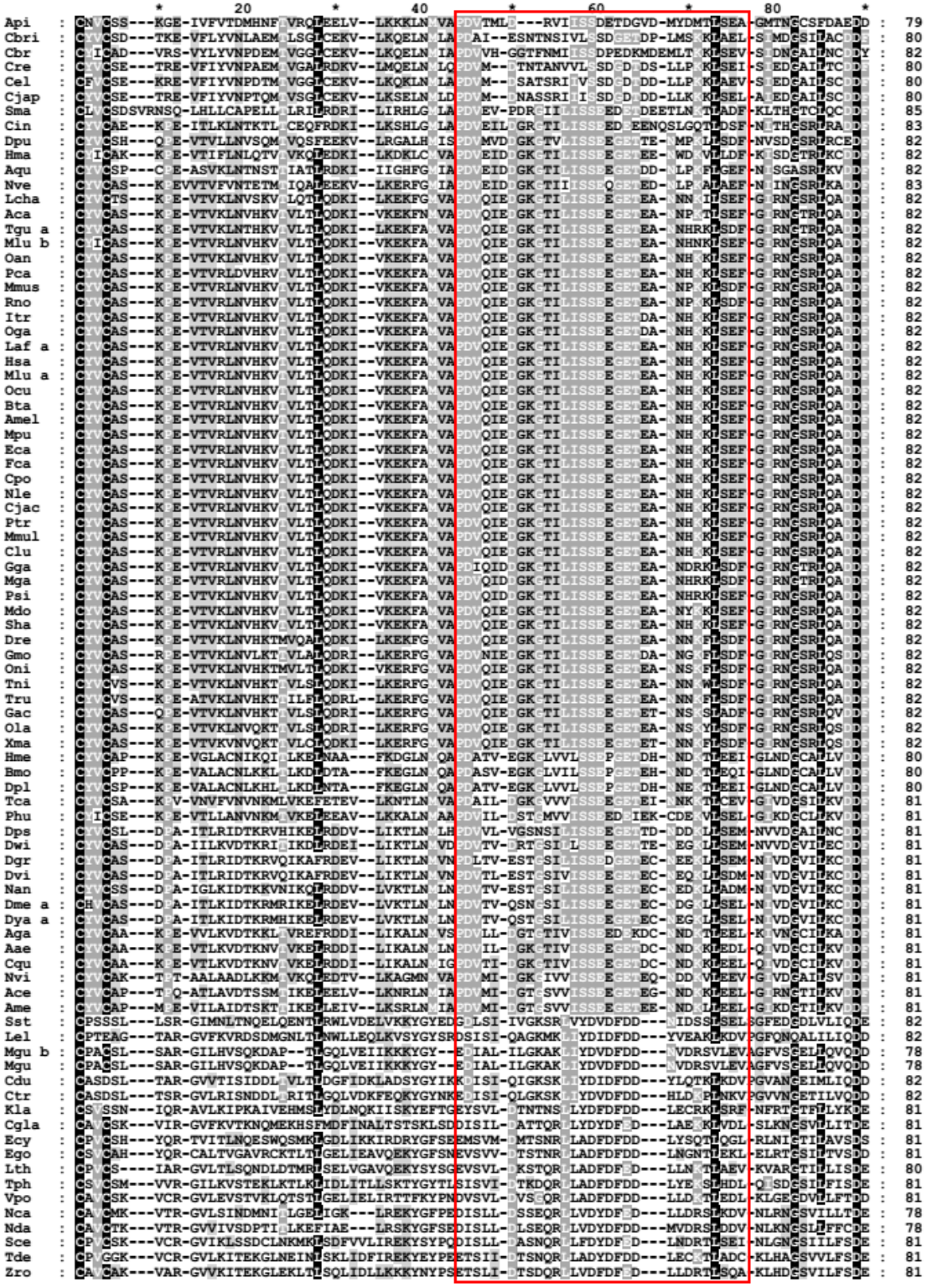


Figure S2. *Analysis of pair homology distribution between UFD or LHEB2 sequences within each phylogenetic group analyzed.* (a) Distribution of identity percentage of UFD (pink boxes) or LHEB2 (yellow boxes) domains between each sequence in the analyzed group and an outlier sequence (human, in the case of saccharomycetales, and yeast, in the case of nematoda, arthropoda and chordata). (b) Distribution identity percentage of UFD (pink boxes) or LHEB2 (yellow boxes) domains between each sequence in the analyzed group and a sequence from the same group (yeast, in the case of saccharomycetales, and human, in the case of nematoda, arthropoda and chordata). Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; data points are plotted as open circles.



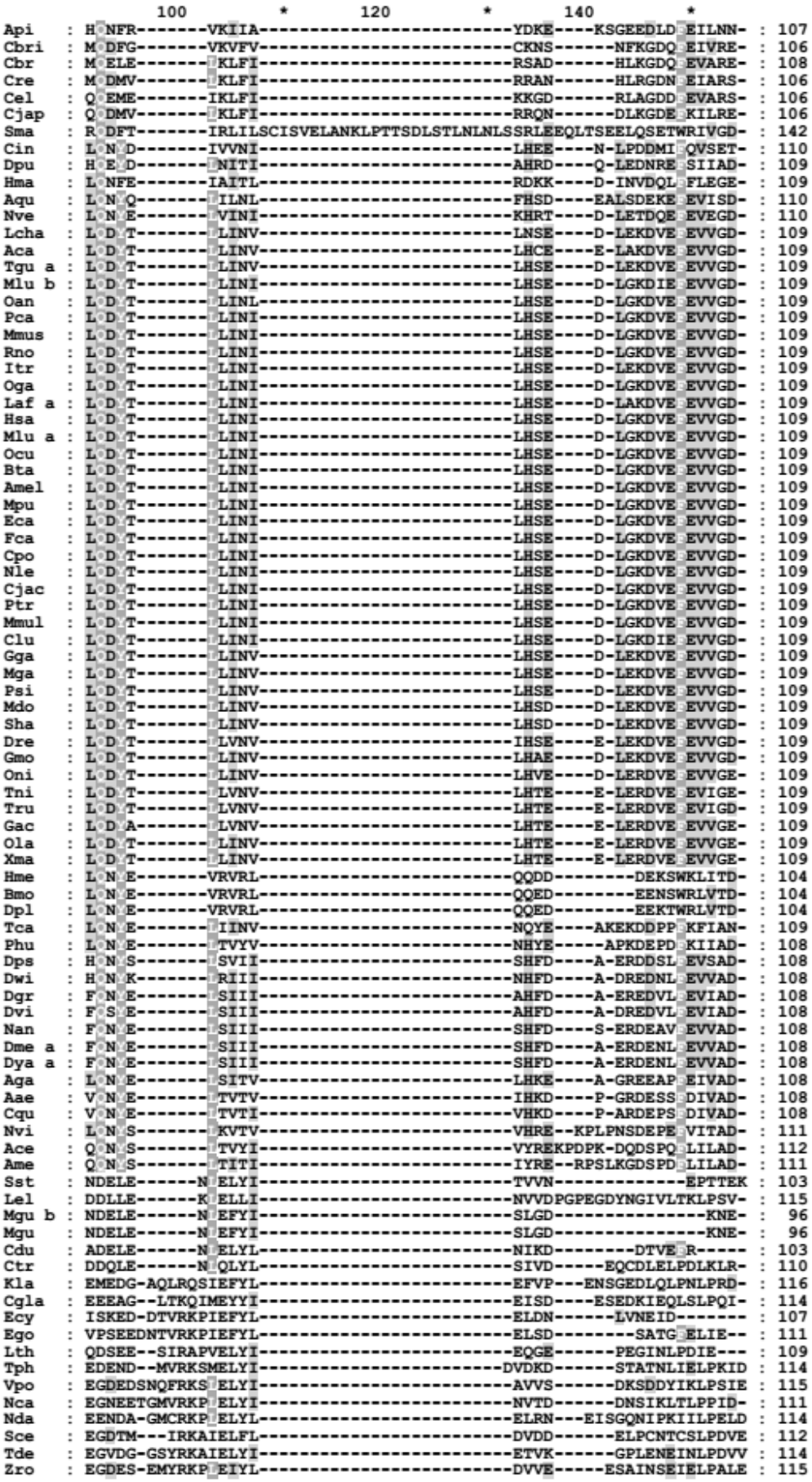

100 * 120 * 140 *
Api : H NFR-------VKIIA----------YDKE----KSGEEDLD EILNN- : 107
Cbri : M DFG-------VKVFV----------CKNS------NFKGDQ EIVRE- : 106
Cbr : M ELE-------KLFI----------RSAD------HLKGDQ EVARE- : 108
Cre : M DMV-------KLFI----------RRAN------HLRGDN EIARS- : 106
Cel : Q EME-------IKLFI----------KKGD------RLAGDD EVARS- : 106
Cjap : Q DMV-------KLFI----------RRQN------DLKGDE KILRE- : 106
Sma : R DFT-------IRLILSCISVELANKLPTTSDLSTLNLNLSSRLEEQLTSEELQSETWRIVGD- : 142
Cin : L N D-------IVVNI----------LHEE----N-LPDDMI QVSET- : 110
Dpu : H E D-------NITI----------AHRD----Q-LEDNRE SIIAD- : 109
Hma : L NFE-------IAITL----------RDKK----D-INVDQL FLEGE- : 109
Aqu : L N Q-------ILNL----------FHSD----EALSDEKE EVISD- : 110
Nve : L N E-------VINI----------KHRT----D-LETDQE EVEGD- : 110
Lcha : L D T-------LINV----------LNSE----D-LEKDVE EVVGD- : 109
Aca : L D T-------LINV----------LHCE----E-LAKDVE EVVGD- : 109
Tgu a : L D T-------LINV----------LHSE----D-LEKDVE EVVGD- : 109
Mlu b : L D T-------LINI----------LHSE----D-LGKDIE EVVGD- : 109
Oan : L D T-------LINL----------LHSE----D-LGKDVE EVVGD- : 109
Pca : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Mmus : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Rno : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Itr : L D T-------LINI----------LHSE----D-LEKDVE EVVGD- : 109
Oga : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Laf a : L D T-------LINI----------LHSE----D-LAKDVE EVVGD- : 109
Hsa : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Mlu a : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Ocu : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Bta : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Amel : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Mpu : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Eca : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Fca : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Cpo : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Nle : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Cjac : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Ptr : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Mmul : L D T-------LINI----------LHSE----D-LGKDVE EVVGD- : 109
Clu : L D T-------LINI----------LHSE----D-LGKDIE EVVGD- : 109
Gga : L D T-------LINV----------LHSE----D-LEKDVE EVVGD- : 109
Mga : L D T-------LINV----------LHSE----D-LEKDVE EVVGD- : 109
Psi : L D T-------LINV----------LHSE----D-LEKDVE EVVGD- : 109
Mdo : L D T-------LINV----------LHSD----D-LEKDVE EVVGD- : 109
Sha : L D T-------LINV----------LHSD----D-LEKDVE EVVGD- : 109
Dre : L D T-------LVNV----------IHSE----E-LEKDVE EVVGD- : 109
Gmo : L D T-------LINV----------LHAE----D-LEKDVE EVVGD- : 109
Oni : L D T-------LINV----------LHVE----D-LERDVE EVVGE- : 109
Tni : L D T-------LVNV----------LHTE----E-LERDVE EVIGE- : 109
Tru : L D T-------LVNV----------LHTE----E-LERDVE EVIGD- : 109
Gac : L D A-------LVNV----------LHTE----E-LERDVE EVVGE- : 109
Ola : L D T-------LINV----------LHTE----E-LERDVE EVVGE- : 109
Xma : L D T-------LINV----------LHTE----E-LERDVE EVVGE- : 109
Hme : L N E-------VRVRL----------QQDD--------DEKSWKLITD- : 104
Bmo : L N E-------VRVRL----------QQED--------EENSWRLVTD- : 104
Dpl : L N E-------VRVRL----------QQED--------EEKTWRLVTD- : 104
Tca : L N E-------IINV----------NQYE----AKEKDDPP KFIAN- : 109
Phu : L N E-------TVYV----------NHYE-----APKDEPD KIIAD- : 108
Dps : H N S-------SVII----------SHFD----A-ERDDSL EVSAD- : 108
Dwi : H N K-------RIII----------NHFD----A-DREDNL EVVAD- : 108
Dgr : F N E-------SIII----------AHFD----A-EREDVL EVIAD- : 108
Dvi : F S E-------SIII----------AHFD----A-DREDVL EVIAD- : 108
Nan : F N E-------SIII----------SHFD----S-ERDEAV EVVAD- : 108
Dme a : F N E-------SIII----------SHFD----A-ERDENL EVVAD- : 108
Dya a : F N E-------SIII----------SHFD----A-ERDENL EVVAD- : 108
Aga : L N E-------SITV----------LHKE----A-GREEAP EIVAD- : 108
Aae : V N E-------TVTV----------IHKD----P-GRDESS DIVAD- : 108
Cqu : V N E-------TVTI----------VHKD----P-ARDEPS DIVAD- : 108
Nvi : L N S-------KVTV----------VHRE--KPLPNSDEPE VITAD- : 111
Ace : Q N S-------TVYI----------VYREKPDPK-DQDSPQ LILAD- : 112
Ame : Q N S-------TITI----------IYRE--RPSLKGDSPD LILAD- : 111
Sst : NDELE------N ELYI----------TVVN----------EPTTEK : 103
Lel : DDLLE------K ELLI----------NVVDPGPEGDYNGIVLTKLPSV- : 115
Mgu b : NDELE------N EFYI----------SLGD----------KNE- : 96
Mgu : NDELE------N EFYI----------SLGD----------KNE- : 96
Cdu : ADELE------N ELYL----------NIKD--------DTVE R----- : 103
Ctr : DDQLE------N QLYL----------SIVD-----EQCDLELPDLKLR- : 110
Kla : EMEDG-AQLRQSIEFYL----------EFVP---ENSGEDLQLPNLPRD- : 116
Cgla : EEEAG--LTKQIMEYYI----------EISD----ESEDKIEQLSLPQI- : 114
Ecy : ISKED-DTVRKPIEFYL----------ELDN------LVNEID------- : 107
Ego : VPSEEDNTVRKPIEFYL----------ELSD--------SATG ELIE-- : 111
Lth : QDSEE--SIRAPVELYI----------EQGE------PEGINLPDIE--- : 109
Tph : EDEND--MVRKSMELYI----------DVDKD------STATNLIELPKID : 114
Vpo : EGDEDSNQFRKS ELYI----------AVVS------DKSDDYIKLPSIE : 115
Nca : EGNEETGMVRKP ELYI----------NVTD------DNSIKLTLPPID- : 111
Nda : EENDA-GMCRKP ELYL----------ELRN---EISGQNIPKIILPELD : 114
Sce : EGDTM---IRKAIELFL----------DVDD------ELPCNTCSLPDVE : 112
Tde : EGVDG-GSYRKAIELYI----------ETVK------GPLENEINLPDVV : 114
Zro : EGDES-EMYRKP EIYL----------DVVE-----ESAINSEIELPALE : 115

Figure S3. *Protein sequence alignment of the Sae2 UFD domain from metazoa and Saccharomycetales*. Human Sae2 homolog sequences were retrieved from EggNOG database. After removal of incomplete sequences or sequences exceeding in length, 86 sequences were retained. Selected sequences were aligned using MUSCLE at EMBL-EBI. The sequence alignment fragment corresponding to the LHEB2 domain is enclosed in a red rectangle. Residue shading correspond to 90% (white letter and dark background), 70% (white letter and gray background), and 50% (black letter and light gray background) of sequence identity.

Phylum	Symbol	Sequence code	Species
Arthropoda	Aae	>7159.AAEL010641-PA	Aedes aegypti (yellow fever mosquito), species, mosquitos
Arthropoda	Ace	>12957.ACEP_00011683-PA	Atta cephalotes (Leafcutter ant)
Arthropoda	Aga	>7165.AGAP008637-PA	Anopheles gambiae (African malaria mosquito), species, mosquitos
Arthropoda	Ame	>7460.GB12173-PA	Apis mellifera (honey bee), species, bees
Arthropoda	Api	>7029.ACYPI006138-PA	Acyrthosiphon pisum (pea aphid), species, aphids
Arthropoda	Cqu	>7176.CPIJ007700-PA	Culex quinquefasciatus (southern house mosquito), species, mosquitos
Arthropoda	Dgr	>7222.FBpp0149193	Drosophila grimshawi (species, flies)
Arthropoda	Dme_a	>7227.FBpp0076457	Drosophila melanogaster (fruit fly), species, flies
Arthropoda	Dpl	>13037.EHJ64540	Danaus plexippus (Monarch butterfly)
Arthropoda	Dps	>7237.FBpp0276753	Drosophila pseudoobscura (species, flies)
Arthropoda	Dpu	>6669.DappuP300104	Daphnia pulex (common water flea), species, crustaceans
Arthropoda	Dvi	>7244.FBpp0226639	Drosophila virilis (species, flies)
Arthropoda	Dwi	>7260.FBpp0249132	Drosophila willistoni (species, flies)
Arthropoda	Dya_a	>7245.FBpp0265382	Drosophila yakuba (species, flies)
Arthropoda	Dan	>7217.FBpp0113673	Drosophila ananassae (species, flies)
Arthropoda	Tca	>7070.TC002294-PA	Tribolium castaneum (red flour beetle), species, beetles
Arthropoda	Bmo	>7091.BGIBMGA012116-TA	Bombyx mori (domestic silkworm), species, moths
Arthropoda	Hme	>34740.HMEL017289-PA	Heliconius melpomene
Arthropoda	Nvi	>7425.NV10104-PA	Nasonia vitripennis (jewel wasp), species, wasps &c
Arthropoda	Phu	>121225.PHUM280660-PA	Pediculus humanus
Chordata	Aca	>28377.ENSACAP00000001485	Anolis carolinensis (Green anole) (American chameleon)
Chordata	Amel	>9646.ENSAMEP00000017826	Ailuropoda melanoleuca (giant panda), species, carnivores
Chordata	Bta	>9913.ENSBTAP00000003333	Bos taurus (cattle), species, even-toed ungulates
Chordata	Cjac	>9483.ENSCJAP00000020075	Callithrix jacchus (white-tufted-ear marmoset), species, primates
Chordata	Clu	>9615.ENSCAFP00000010792	Canis lupus familiaris (dog), subspecies, carnivores
Chordata	Cpo	>10141.ENSCPOP00000009465	Cavia porcellus (Guinea pig)
Chordata	Dre	>7955.ENSDARP00000102182	Danio rerio (zebrafish), species, bony fishes
Chordata	Eca	>9796.ENSECAP00000000109	Equus caballus (horse), species, odd-toed ungulates
Chordata	Fca	>9685.ENSFCAP00000020786	Felis catus (domestic cat), species, carnivores
Chordata	Gac	>69293.ENSGACP00000004237	Gasterosteus aculeatus (three-spined stickleback), species, bony fishes
Chordata	Gga	>9031.ENSGALP00000038502	Gallus gallus (chicken), species, birds
Chordata	Gmo	>8049.ENSGMOP00000004426	Gadus morhua (Atlantic cod), species, bony fishes
Chordata	Hsa	>9606.ENSP00000246548	Homo sapiens
Chordata	Itr	>43179.ENSSTOP00000002332	Ictidomys tridecemlineatus (thirteen-lined ground squirrel), species, rode
Chordata	Laf_a	>9785.ENSLAFP00000011542	Loxodonta africana (African savanna elephant), species, placentals
Chordata	Mga	>9103.ENSMGAP00000006132	Meleagris gallopavo (turkey), species, birds
Chordata	Mlu_a	>59463.ENSMLUP00000002773	Myotis lucifugus (little brown bat), species, bats
Chordata	Mlu_b	>59463.ENSMLUP00000018571	Myotis lucifugus (little brown bat), species, bats
Chordata	Mmul	>9544.ENSMMUP00000007742	Macaca mulatta (Rhesus monkey), species, primates
Chordata	Mmus	>10090.ENSMUSP00000099807	Mus musculus
Chordata	Mpu	>9669.ENSMPUP00000007408	Mustela putorius furo (domestic ferret), subspecies, carnivores
Chordata	Nle	>61853.ENSNLEP00000013124	Nomascus leucogenys ((northern white-cheeked gibbon), species, primates)
Chordata	Ocu	>9986.ENSOCUP00000001260	Oryctolagus cuniculus (rabbit), species, rabbits & hares
Chordata	Ola	>8090.ENSORLP00000015752	Oryzias latipes (Japanese medaka), species, bony fishes
Chordata	Oni	>8128.ENSONIP00000013939	Oreochromis niloticus (Nile tilapia), species, bony fishes
Chordata	Pca	>9813.ENSPCAP00000009534	Procavia capensis (Cape rock hyrax), species, placentals
Chordata	Ptr	>9598.ENSPTRP00000018512	Pan troglodytes (chimpanzee), species, primates
Chordata	Rno	>10116.ENSRNOP00000028672	Rattus norvegicus
Chordata	Sha	>9305.ENSSHAP00000002779	Sarcophilus harrisii (Tasmanian devil), species, marsupials
Chordata	Tgu_a	>59729.ENSTGUP00000009909	Taeniopygia guttata (zebra finch), species, birds
Chordata	Tni	>99883.ENSTNIP00000005165	Tetraodon nigroviridis (spotted green pufferfish), species, bony fishes
Chordata	Xma	>8083.ENSXMAP00000016139	Xiphophorus maculatus (southern platyfish), species, bony fishes
Chordata	Cin	>7719.ENSCINP00000030774	Ciona intestinalis ((vase tunicate), species, tunicates)
Chordata	Lcha	>7897.ENSLACP00000007856	Latimeria chalumnae (coelacanth), species, coelacanths
Chordata	Mdo	>13616.ENSMODP00000014955	Monodelphis domesticus
Chordata	Oan	>9258.ENSOANP00000015209	Ornithorhynchus anatinus (platypus), species, monotremes
Chordata	Oga	>30611.ENSOGAP00000007933	Otolemur garnettii (Small-eared galago) (Garnett's greater bushbaby)
Chordata	Psi	>13735.ENSPSIP00000010107	Pelodiscus sinensis
Chordata	Tru	>31033.ENSTRUP00000016721	Takifugu rubripes (Japanese pufferfish) (Fugu rubripes)
cnidaria	Nve	>45351.NEMVEDRAFT_v1g161397-PA	Nematostella vectensis ((starlet sea anemone), species, sea anemones)
cnidaria	Hma	>6085.XP_002165708	Hydra magnipapillata
Nematoda	Cbr	>135651.CBN10217	Caenorhabditis brenneri (Nematode worm)
Nematoda	Cbri	>6238.CBG13604	Caenorhabditis briggsae (nematodes)
Nematoda	Cel	>6239.W02A11.4	Caenorhabditis elegans (roundworm), species, nematodes
Nematoda	Cjap	>281687.CJA09929	Caenorhabditis japonica
Nematoda	Cre	>31234.CRE28085	Caenorhabditis remanei (Caenorhabditis vulgaris)
Platyhelminthes	Sma	>6183.Smp_166220__mRNA	Schistosoma mansoni (flatworns)
Porifera	Aqu	>400682.PAC_15728046	Amphimedon queenslandica (species, sponges)

Supplementary Table S1. *Phylum, symbol, sequence code and species used in the alignment of the Sae2 UFD domain from metazoa*.

Chapter 3

Structure insights into specificities of SUMO E1-E2 interactions between *A. thaliana* and human

Bing Liu[1], L. Maria Lois[2] and David Reverter[1*]

[1]Institut de Biotecnologia i de Biomedicina. Departament de Bioquimica i Biologia Molecular. Serra Hunter Fellow. Universitat Autonoma de Barcelona, 08193 Barcelona, Bellaterra, Spain.

[2]Center for Research in Agricultural Genomics-CRAG. Edifici CRAG-Campus UAB, Bellaterra 08193 Barcelona, Spain.

*To whom the correspondence should be addressed: David Reverter, Universitat Autonòma de Barcelona, 08193 Bellaterra, Barcelona, Spain, Ph: +34 93 5868955, FAX: (93)5812011, E-mail: david.reverter@uab.cat

Abstract

SUMO belongs to the ubiquitin-like family (Ubl) of protein modifiers. It is conserved among eukaryotes and is essential in processes such as DNA damage repair, transcription, DNA replication and mitosis. Ubl modification of proteins occurs via a specific enzymatic cascade formed by the crosstalk between the E1-activating enzyme, the E2-conjugating enzyme and the E3-ligase. An essential discrimination step in all Ubl modifiers corresponds to the interaction between E1 and E2 enzymes, which is mediated by the recruitment of the E2 to the UFD domain (ubiquitin-fold domain) of the E1 enzyme. Comparative sequence analysis of the E1 UFD domain indicates that the E2 binding region is quite conserved across phylogenetic closely related species, especially only considering the binding region. To gain more insights in the properties of this interface in a distant organism, we solved the crystal structure of SUMO E2 SCE1 and its complex with E1 UFD in *A. thaliana*. Despite common determinants, the interaction interface between the E1 UFD and E2 in *A. thaliana* is distinct compared with human and yeast in the structure analysis. Although the E2s possesses high sequence identity, *A. thaliana* E1 and E2 SCE1 failed to interact with cognate human E2 Ubc9 and E1 partners, respectively, in the heterologous reactions to conjugate the substrate RanGAP1. Substituting one or up to four residues in the UFD binding region of SCE1 with residues from human Ubc9 also failed in the RanGAP1 conjugation assays. Moreover, structure analysis and biochemical assays reveled that surface residues outside UFD interaction interface of *A. thaliana* SCE1 affect the RanGAP1 conjugation, even though these residues are not involved in the E1 binding or RanGAP1 interaction. The surface residues mutant presented a lower stability in the temperature and urea induced unfolding analysis, probably indicating an important role of these residues in the activity of the E1 SUMO-thioester transfer to the E2.

Introduction

Small ubiquitin-related modifier (SUMO) protein is one of the most extensively studied ubiquitin-like proteins (Ubls). SUMO is extensively expressed in all eukaryotes. Yeast and invertebrates have a single SUMO protein named Smt3, while vertebrates have several SUMO proteins. Mammalians such as human express four SUMO proteins: SUMO1, SUMO2, SUMO3, and SUMO4. SUMO2 and SUMO3 share 97% sequence identity and cannot be distinguished by antibodies (1, 2). However, SUMO1 is quite different from SUMO2/3 and only shares ~47% sequence identity with SUMO1 and SUMO2 (2). In *A. thaliana*, there are eight encoding genes of SUMO isoforms (AtSUMO), but only SUMO1, SUMO2, SUMO3, an SUMO5 are expressed (3, 4). AtSUMO1 and AtSUMO2 (orthologs of human SUMO2/3) are essential in *A. thaliana* and with 83% sequence identity, while AtSUMO3 and AtSUMO5 only share 42% and 30% sequence identity respectively (5-8).

SUMO is conjugated to target proteins through a three-step enzyme cascade as ubiquitin via SUMO E1 activating enzyme, E2 conjugating enzyme, and E3 ligase (1, 2). In this proposed model for the SUMO/Ubl conjugation pathway, the E2 conjugating enzyme interacts with the E1 activating enzyme by means of the UFD domain (9-12). This UFD-E2 interactions have been solved in ubiquitin, Nedd8, and SUMO systems, showing the direct binding of E2 to the E1 UFD domain (11-16). The UFD domain is in the C-terminal end and connected to the E1 active adenylation domain through a flexible hinge. In SUMO and Nedd8 systems, this disordered hinge is converted to the ordered state upon E2 bindings, and it must undergo a rotation to bring the catalytic cysteine residues of E1 and E2 into proximity for thioester transfer (17, 18). Moreover, crystal structure of ubiquitin E1-E2 complex revels a direct interaction of E1 and E2 through their catalytic cysteine domains, which occurs after UFD-E2 binding and the significant rotation of the UFD domain, providing structure

insights for the E1 E2 thioester transfer. This interaction was also proposed in the SUMO pathway by NMR analyses. However, UFD-E2 interactions (Kd= 1.2μM) display much higher affinity than the catalytic cysteine domain interactions (Kd= 87μM), consistent with a major role of E1 UFD domain as a E2 binding platform (19, 20).

Structural and functional data indicate that the UFD-E2 interaction represents the first contact between the E1-activating enzyme and the E2-conjugating enzyme, providing an important discrimination step for the pathway (21). Therefore, the UFD-E2 interaction has been revealed essential to provide specificity between the different systems of ubiquitin-like modifiers (i.e. SUMO, Nedd8, ubiquitin…). Moreover, in the SUMO system, this interaction can also provide specificity across species, despite the high homology between E2-conjugating proteins, Ubc9, which in some instances reaches around 70% sequence identity. Unlike E2 enzymes, UFD domains show a little protein sequence homology across species, and it is even lower when considering just the E2 binding region or LHEB2 domain (Low Homology region involved in E2 Binding 2) (22). However, the UFD domains are quite conserved across phylogenetic closely related species, especially only taking the LHEB2 domain into account (16, 22).

The analysis of the different specificities provided by the UFD-E2 domain interface in the SUMO pathway between human, yeast and *A. thaliana* systems, despite the strong similarity between E2-conjugating enzymes, is a major objective of this research. For example, despite the 65% sequence identity between human and *A. thaliana* E2 conjugating enzymes, human Ubc9 cannot complement SCE1 (*A. thaliana'*s Ubc9) in our in vitro conjugating systems using RanGAP1 as model substrate, and vice versa. A major goal of this work is to figure out the structural determinants for such exquisite specificity in the interface of the E2-E1 UFD interaction. We have been able to solve the crystal structure of the *A. thaliana* E2-conjugating enzyme (SCE1), alone at very high

resolution (1.2 Å) and in complex with the *A. thaliana* UFD domain of the E1-conjugating enzyme at 1.8 Å resolution. We have compared the *A. thaliana* UFD-E2 domain complex interface with the human and yeast systems and performed biochemical and mutagenesis analysis to understand the role of the interface in the specificity of the SUMO conjugation pathway.

Results

Protein expression and complex formation

The recombinant expression in *E. coli* of both proteins was successful and notable protein amounts could be retrieved after the final purification steps. *A. thaliana* E2-conjugating enzyme (AtSCE1) was expressed as a full-length protein (Met1 to Val 158) in a pET28 vector, which includes a N-terminal His-tag extension to facilitate its purification. The UFD domain of the *A. thaliana* E1-activating enzyme was expressed in two different versions: a short version including only the UFD domain (ubiquitin-like fold) from Glu437 to Thr549 (AtUFD); and a longer version including the C-terminal extension, from Glu437 to Glu625 (AtUFDC). AtUFD E1 domain constructs, AtUFD and AtUFDC, were both recombinantly produced at high yields in *E. coli* with a pET28 vector, which includes a N-terminal His-tag for affinity purification.

Complex formation between AtSCE1 and AtUFD domains was firstly conducted by mixing equimolar concentration amounts of both components and running a size-exclusion chromatography. Unexpectedly, in *A. thaliana* the pH of the gel-filtration running buffer turned out to be determinant for the complex formation after the analysis of the size-exclusion profile results (***Fig. 1***). In contrast to the human complex we purified previously (16), which was run with a low salt buffer at pH 8.0, the formation of the complex in *A. thaliana* was only observed in a low salt buffer at pH 7.0 (either with Tris or HEPES buffer), whereas at pH 8.0 the complex was unstable and both protein did not co-elute from the size-exclusion column (***Fig. 1***). The fractions of the complex between the AtSCE1 and AtUFD domains were pooled and concentrated to set up crystallization screens. Interestingly, only the complex including the C-terminal extension of the UFD E1 domain (AtUFDC) produced crystals, however the extension was not observed in the final electron density maps after refinement.

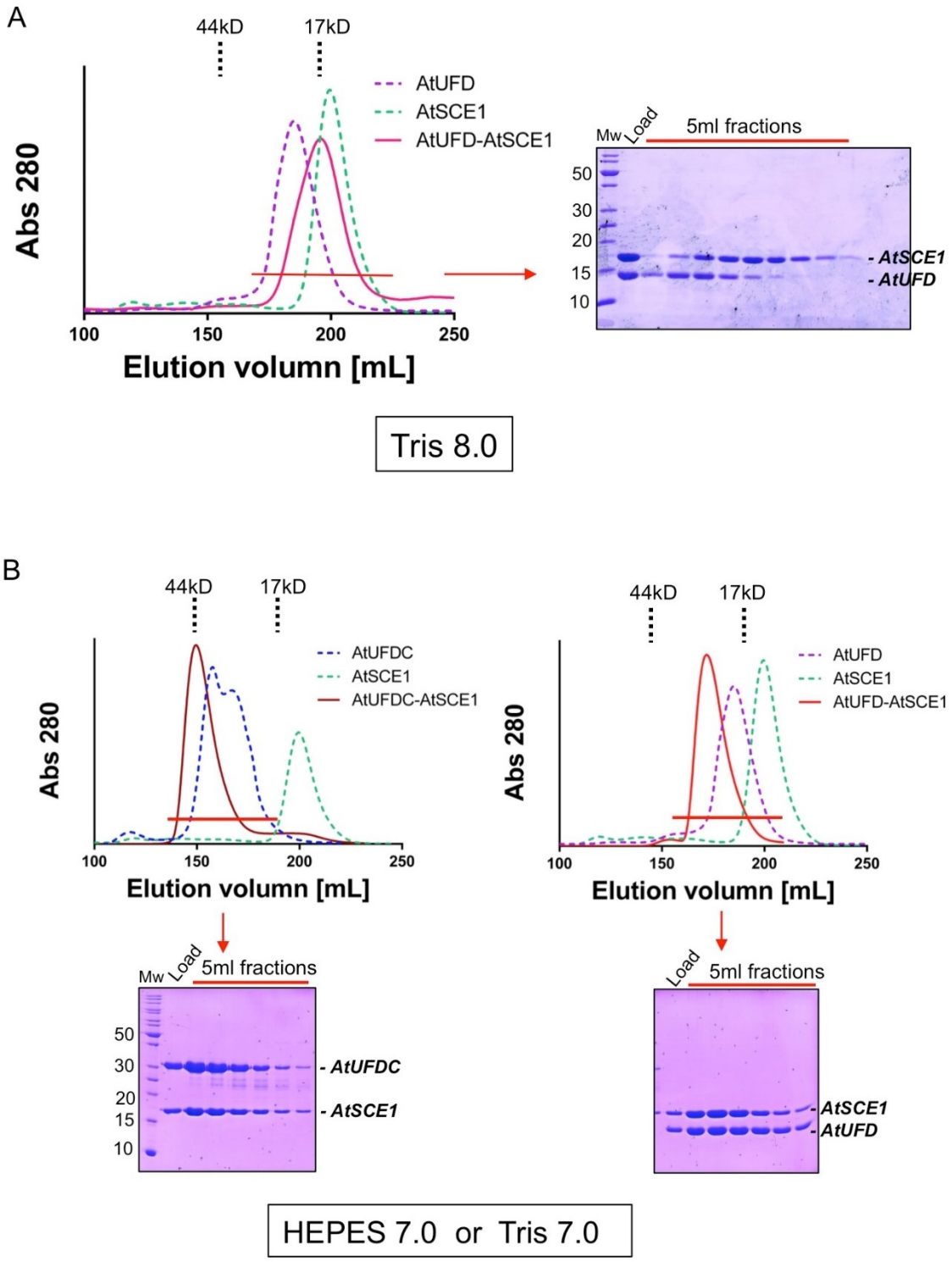


Fig 1. ***Purification of the complex between AtSCE1 and AtUFD E1 domain****. A*, *Left*, gel filtration purification at pH 8.0 after mixing equimolar concentrations of AtUFD and AtSCE1 (red line). Dotted lines indicated the elution volumes of the individual components. *Right*, PAGE analysis of the fractions of the elution peak. *B*, Gel filtration purification at pH 7.0 after mixing equimolar concentrations of AtSCE1 with either AtUFD (*left*), or AtUFDC (*right*). Dotted lines indicated the elution volumes of the individual components *Below*, PAGE analysis of the fractions of the elution peak.

Crystal structure of *A. thaliana* SUMO E2 alone and in complex with UFD E1 domain

The structure of the *A. thaliana* E2-conjugating enzyme, SCE1, has been solved by x-ray diffraction at very high resolution (1,2 Å) (***Table 1***). This resolution allows the visualization of many structural details in the electron density, such as the presence of double conformations for some residues of the SCE1 structure. Intriguingly, in addition to these side chains double conformations, the final electron density reveals a 6 residues helical stretch in the C-terminal region of SCE1, between Tyr134 and Asp140, that seems flexible in our structure and can adopt at least two different conformations, which is also revealed by the high B-factor values of the atoms for this region. Despite this particular feature in the AtSCE1 crystal structure, altogether the structure is highly similar to other crystal structures of human Ubc9 (65% identity and rmsd=0,88 Å with human Ubc9 for 155 aligned residues) (***Fig. 2***). The major differences in the backbone overlapping with human Ubc9, are the β1-β2 loop, which is involved in the UFD E1 domain binding, and a small loop at the end of β4 strand, in which 3 residue substitutions (Pro-Gln-Gly in *A. thaliana,* for Glu-Pro-Pro in human) adopt a different backbone orientation. As mention later in the text, this region might have a role in the paralog specificity of the E2s in our conjugation assays (***Fig. 2***). The AtSCE1 structure has been deposited in the PDB data bank with the code 6GV3.

We have also been able to solve the crystal structure of AtSCE1 in complex with the UFD domain of the *A. thaliana* E1-activating enzyme at 1,8 Å resolution (***Fig. 3***). After several unfruitful crystallization assays, crystals of the complex grew after incubation for several months at 18°C, displaying a very good X-ray diffraction quality, which avoid long periods of optimization steps to reproduce the crystals. The structure of the complex was solved by molecular replacement using the previous AtSCE1 structure as a search model. As detailed later in the text, the interface region in the AtSCE1 E2-conjugating enzyme comprises

contacts in the α1 helix and in the β1-β2 loop, as expected from previous structures of the complex in human and in yeast (14-16). The structure of the UFD-AtSCE1 complex has been deposited in the PDB data bank with the code 6GUM. In table 1 we have summarized the collection and statistical details of the crystals of AtSCE1 alone and in complex with AtUFD domain.

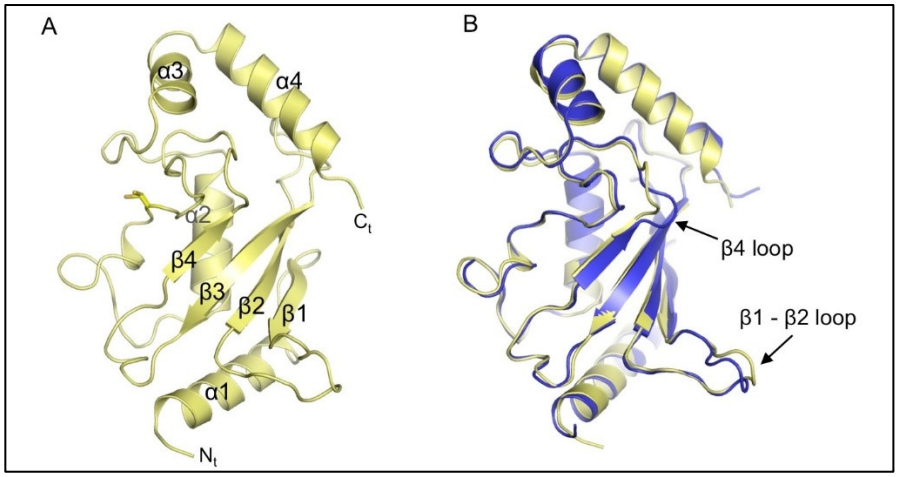


Fig 2. ***Crystal structure of the A. thaliana E2-conjugating enzyme AtSCE1****. A*. Cartoon representation of the structure of AtSCE1. *B*, Structural alignment of AtSCE1 and human Ubc9 (PDB code 1A3S) (23). Backbone differences are depicted.

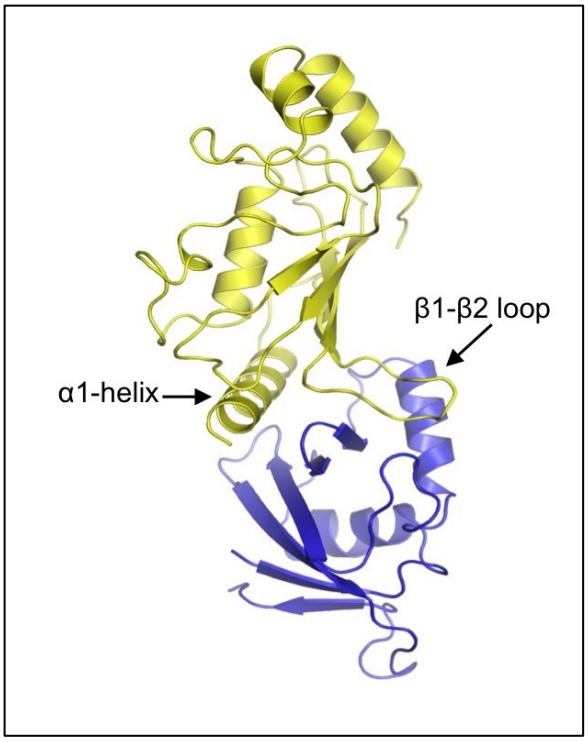


Fig 3. ***Crystal structure of the complex between A. thaliana E2-conjugating enzyme SCE1 and the UFD domain of the E1-activating enzyme****.* Cartoon representation of the structure of AtSCE1 in yellow and the AtUFD E1 domain in blue.

Table 1. ***Summary of crystallographic analysis***

Data collection		
	AtSCE1	**AtSCE1 - AtUFD**
Beamline	ALBA-XALOC	ALBA-XALOC
Space group	$P2_12_12_1$	$P2_1$
Wave length (Å)	0.97934	0.97923
Resolution (Å)	50.38-1.201 (1.201-1.205)	58.82-1.792 (1.792-1.798)
a, b, c (Å)	32.90, 50.38, 93.78	36.16, 58.82, 70.23
α, β, γ (°)	α = β = γ = 90	α = γ = 90, β = 92.05
Unique reflections	49524	26999
Data redundancy	6.0 (6.2)	3.4 (3.3)
R_{merge}	0.14 (1.45)	0.094 (0.97)
CC(1/2)	0.996 (0.642)	0.992 (0.581)
I/σ	9.5 (2.2)	8.5 (1.7)
Completeness (%)	99.9 (93.6)	99.9 (93.6)
Refinement		
Resolution (Å)	44.38 – 1.20	45.08 – 1.792
Non-anomalous reflections	49227	26964
R_{work}/R_{free}	0.19/0.21	0.17/0.22
Number of all atoms	1310	2212
Number of waters	188	68
RMSD bond (Å)/Angle (°)	0.014/1.50	0.017/1.62
Ramachandran plot		
Favored (%)	98.17	99.24
Allowed (%)	0.61	0.76
Disallowed (%)	1.21	0

* Highest resolution shell is shown in parenthesis.

Structural analysis of the complex interface between AtSCE1 and the UFD E1 domain

The complex between AtSCE1 and the UFD domain in *A. thaliana* display a similar structure interface alike the human and yeast counterparts (PDB codes 5FQ2 and 3ONG)(14, 16), in which the α1 helix and the β1-β2 loop of the AtSCE1 E2-conjugating enzyme sit on one side of the central β-sheet of the UFD domain (***Fig. 3***). As described in previous comparisons of the complex, structural superposition of the E2 enzymes between human, yeast and *A. thaliana* reveal a rotation of the UFD domain, which probably account for the particular specificity of each SUMO conjugation system (16).

As mentioned before, the binding surface in the *A. thaliana* E2-conjugating enzyme, AtSCE1, is quite similar to the known structures of the complex in the human and yeast counterparts, and it is formed by a high number of contacts between the α1 helix and the β1-β2 loop of AtSCE1 (around 65% sequence identity between AtSCE1 and human Ubc9) (***Fig. 4 & 5***). Specific contacts include Val37, Leu39, Met40 and Glu67 from the β1-β2 loop and the basic patch in the α1 helix involving Arg7, Arg14, Lys15, Arg18 and Lys19. Differences between human and *A. thaliana* in the E2 contacts are little, only the presence of an Arg7 (leucine in human), Lys28 (valine in human) and Val37 (Met36 in human). Interestingly, despite this highly conserved surface in the E2 between human and *A. thaliana*, each E2-conjugating enzyme only interacts productively with the UFD domain of its cognate E1-activating enzyme. One of the aims of this project was to figure out the structural basis for such exquisite specificity in the SUMO conjugation pathways between species, and to figure out whether this can be attributed to the interface between the E1 and the E2 enzymes. Previous mutational analysis of Ubc9 contact residues, such as the basic patch in α1 helix, revealed the major role of this interaction in the binding to the UFD E1 domain, which was proved to be essential for the transfer of SUMO between E1 and E2 enzymes in the conjugation pathway.

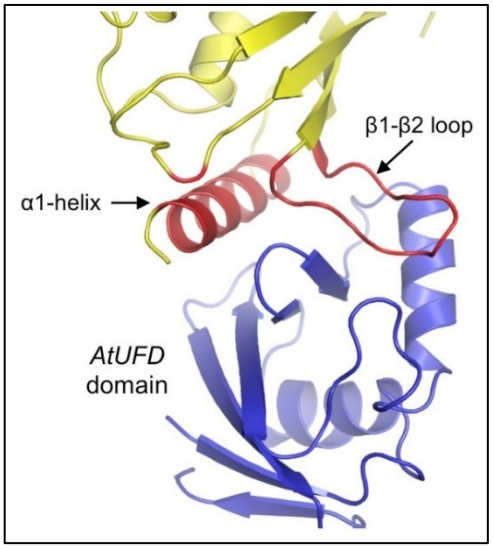


Fig 4. ***Interface of the complex between A. thaliana E2-conjugating enzyme SCE1 and the UFD domain of the E1-activating enzyme***. Cartoon representation of the structure of AtSCE1 in yellow and the AtUFD E1 domain in blue. Interface elements α1 helix and the β1-β2 loop of SCE1 are depicted in red.

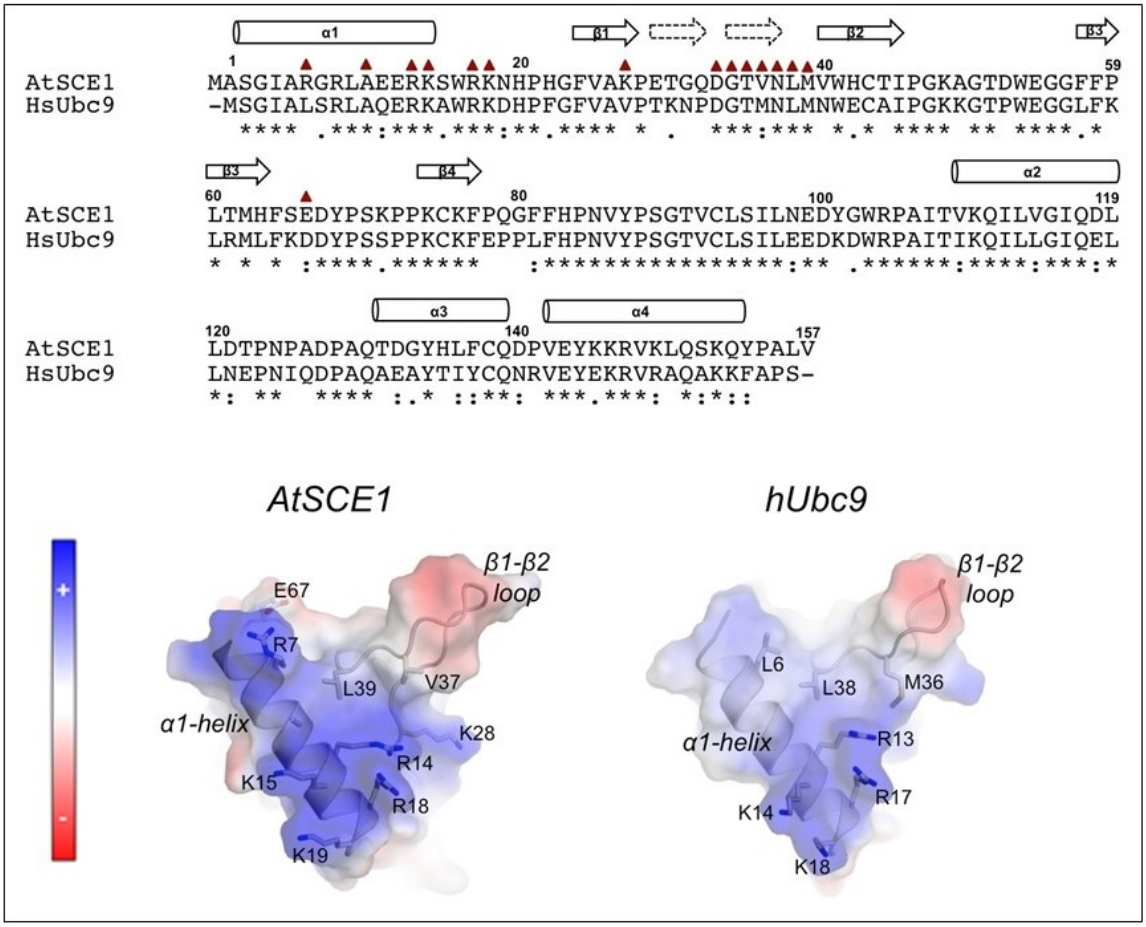


Fig 5. ***Analysis of the AtSCE1 surface interaction with the AtUFD domain***. *A* Structural sequence alignment of *A. thaliana* AtSCE1 and human Ubc9. AtSCE1 domain secondary structure is shown on top. Red triangles indicate AtSCE1 contacts with AtUFD domain. *B*, cartoon and transparent electrostatic surface representation of the structure of AtSCE1 and human Ubc9. Contact residues are labeled and shown in stick representation.

The structure of the AtUFD domain of *A. thaliana* E1-activating enzyme shows a similar fold as the human or yeast counterparts: i.e. comparison with the human UFD E1 domain shows a rmsd 1,17 Å for 91 aligned residues with a 30% sequence identity. A major structural difference of AtUFD domain is the presence of a longer α-helix which participates in the binding to the β1-β2 loop surface of the AtSCE1 counterpart (***Fig. 6***). The presence of this novel α-helix is a consequence of a sequence insertion in the middle of the UFDs domains only observed in the *A. thaliana* UFD. In figure 7 we can observe that this longer α-helix, in addition to the residues emanating from central β-sheet of the AtUFD domain, participates in the contact region with the E2 enzyme. The physical and chemical features of the contacts between the UFD domain and the E2 protein are basically conserved, only different tilts between the two proteins in the complex can be observed in comparison with the human or yeast complexes.

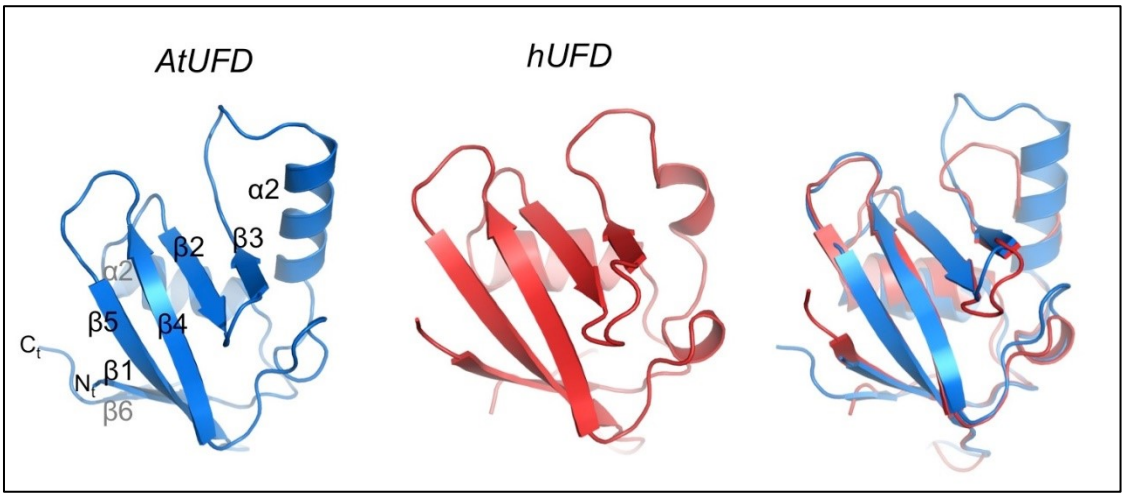


Fig 6. ***Comparison of the sturctures of the UFD E1 domains from A. thaliana and human****. Left,* cartoon representation of the structure of AtUFD (blue). Secondary structure elements are depicted. *Middle*, cartoon representation of the structure of human hUFD (red). *Right*, structural overlapping of AtUFD (blue) and hUFD (red).

Analysis of the contact region between AtUFD and AtSCE1

We have previously defined the interacting region between E2 and UFD E1 domain as *Low Homology region involved in E2 Binding 2* (LHEB2) (22), because it displays a lower homology in comparison to the rest of the UFD domain: i.e. UFD sequence identity between human and *A. thaliana* is 31%, but is reduced to 11% when considering only the LHEB2 binding region. The LHEB2 can be divided in two regions, each establishing interactions with either the α1 helix or the β1-β2 loop of AtSCE1. Interestingly, there is a unique contact in *A. thaliana* engaged between AtUFD Asn474 and AtSCE1 Glu67, region far from the consensus binding module α1-helix/β1-β2 loop of the E2-conjugating enzyme.

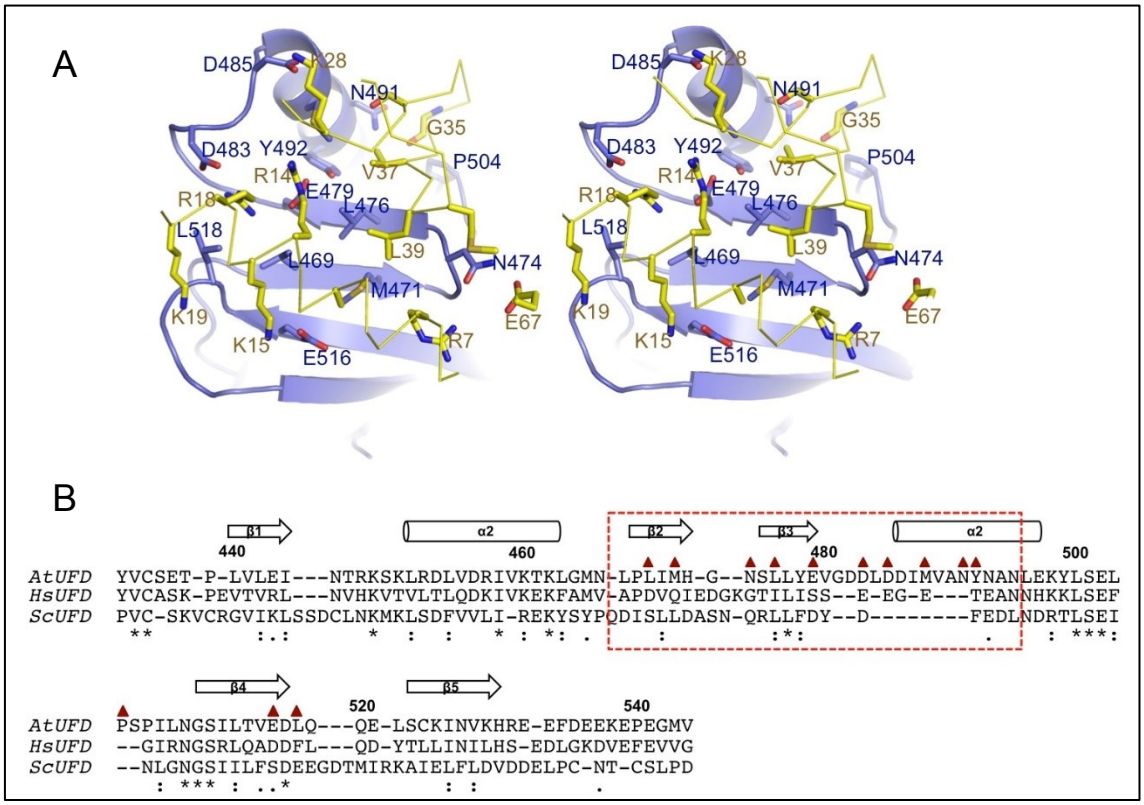


Fig 7. ***Structural details of the complex interface between the AtUFD domain and AtSCE1****. A,* stereo cartoon representation of the contacts between AtSCE1 (yellow thin ribbon) and AtUFD E1 domain (blue cartoon). Side-chain contact residues are labeled and shown in stick representation. *B*, structural sequence alignment of the UFD domains of *A. thaliana*, human and yeast. AtUFD domain secondary structure is shown on top. Red triangles indicate AtUFD contacts with SCE1. Dotted red rectangle indicates the predicted LHEB2 region.

The first contact region in the AtUFD domain is composed by Glu516, Glu479 and Asp483, which contact the charged groups of the AtSCE1 α1-helix (Arg14, Lys15, Arg18 and Lys19) and by AtUFD Leu469 and Leu518, which are buried in a hydrophobic patch formed by the aliphatic side chain atoms of the aforementioned AtSCE1 α1 helix basic residues (***Fig. 7***). Interestingly, Phe522 occupies this region in human, whereas in yeast is occupied by Tyr489, all of them buried in a similar aliphatic pocket formed between the α1 helix basic residues. The other contacts present in the AtSCE1 α1 helix include the unique Arg7 (leucine in human), and Ala11, which interact with the AtUFD Met471 (glutamine or leucine in human and yeast, respectively).

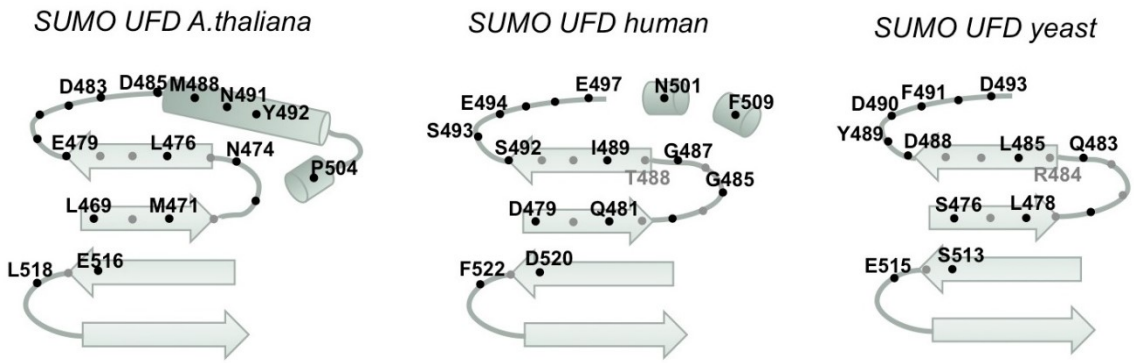


Fig 8. ***Comparison of the UFD SUMO E1 domain contacts from A. thaliana, human and yeast****. Left,* schematic representation of the *A. thaliana* SUMO E1 contacts with the E2 enzyme. *Middle*, schematic representation of the human SUMO E1 contacts with the E2 enzyme (PDB code 5FQ2). *Right*, schematic representation of the yeast SUMO E1 contacts with the E2 enzyme (PDB code 3ONG).

The second region of AtUFD includes contacts with the β1-β2 loop surface of the AtSCE1. This interface is composed by backbone and side chain interactions emanating from the β2-β3 connecting loop and the longer α2 helix of AtUFD. Leu476, Asp485, Asn491, Tyr492 and Pro504 side chains of AtUFD domain are the major contacts with AtSCE1. In the human and yeast complexes this region is formed by different contacts (see ***Fig. 7 & 8***), but some of the interactions are conserved, such as the central Leu476 (isoleucine and leucine

in human and yeast, respectively) and the backbone hydrogen bond contacts between AtUFD Asn474 and Leu476, with AtSCE1 Val37 and Leu39. It is worth mentioning in the *A. thaliana* complex the contact region with the longer α3 helix in AtUFD domain, which is not present in either the human or the yeast counterparts.

Comparison between *A. thaliana* and human SUMO conjugation.

A major question in the field deals with the compatibility between enzymes of the SUMO conjugation pathway form different species, in particular we have been interested in the analysis of the E2-conjugating enzyme, which shares 65% sequence identity between human and *A. thaliana*, but they cannot be exchanged between systems in our in vitro conjugation assays using RanGAP1 as conjugation substrate (***Fig. 9B***), consistent with the same experiments conducted between human and *P. falciparum* (20). Moreover, structural alignment of UFD-E2 complexes between human and *A. thaliana* indicates that the residues involved in the contact region, basically the α1-helix/β1-β2 loop module of the E2-conjugating enzyme, are almost conserved between human and plants, thus indicating the role of other determinants other than just the initial binding between E2 and UFD E1 domains.

The first analyses that we have performed consist in SUMO conjugation assays with the purified components of human and *A. thaliana* systems: E1, E2, SUMOs and a canonical substrate, RanGAP1, which do not need the presence of E3s for an efficient conjugation. In figure 9A, the results of the combination of SUMOs between human and *A. thaliana* systems are displayed, indicating that whereas *A. thaliana* SUMO1 & 2 can be both efficiently conjugated by E1 and E2 in both systems, *A. thaliana* E1 and E2 were not able to conjugate human SUMO1 & 2 efficiently to RanGAP1 substrate. These results show that SUMOs can be exchanged between systems, at least for human enzymes, indicating

that probably the observed incompatibility raises from the E2 interaction to E1 and/or to the SUMO-thioester transfer between E1 and E2.

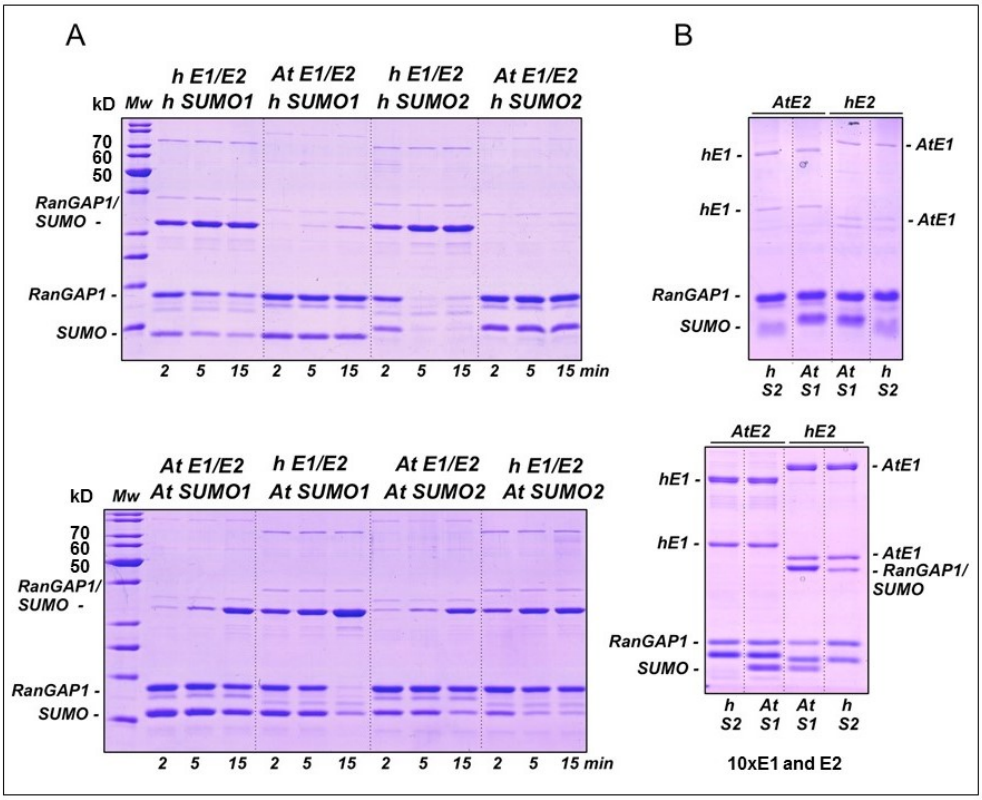


Fig 9. ***SUMO conjugation assays exchanging SUMOs between human and A. thaliana systems****. A, above,* PAGE analysis of the conjugation assays using human SUMO1 & 2 with E1 and E2 enzymes from human and *A. thaliana*. *Below*, PAGE analysis of the conjugation assays using *A. thaliana* SUMO1 & 2 with E1 and E2 enzymes from human and *A. thaliana. B*, *above*, end-point reaction after 120 minutes of the conjugation assays combining SUMO1 and SUMO2 from human and *A. thaliana*, respectively, with E1 and E2 enzymes from human and *A. thaliana*. *Below*, same reaction as above panel containing 10x of E1 and E2.

However, when similar SUMO conjujation reactions were performed combining human and *A. thaliana* E1 and E2 enzymes, SUMOylated RanGAP1 could only be detected in trace amount in AtSCE1 and human E2 combination after 120min reaction (***Fig. 9B***). In reactions containing 10x E1 and E2, the SUMOylated RanGAP1 was readily detectable, and there was more SUMOylated RanGAP1 in reaction containing *A. thaliana* E1 than human SUMO2, indicating that *A. thaliana* E1 displayed a preference for *A. thaliana*

SUMO1 than human SUMO2. More efforts using different susbtrates and E3s also failed to get detectable SUMOylated RanGAP1.

These results indicate despite that the E1-SUMO activation could be achieved combining human or plant SUMOs, the whole SUMO conjugation process relies on compatible interactions between E1 and E2 enzymes, which are necessary for the SUMO-thioester transfer from the E1 to the E2-conjugating enzyme.

Pull-down assays of E2 and UFD domain between human and *A. thaliana*

Since SUMO was not supposed to be an important issue for cross-conjugation SUMO assays between species (at least between human enzymes and *A. thaliana* SUMOs), we next checked the direct binding of the E2-conjugating enzymes with the UFD domains by pull-down assays. This interaction has been proposed to be an essential recognition step and the first recruitment of E2 on the E1-conjugation enzyme. Our crystal structures indicate that E2 surface in contact with the UFD E1 domain is highly conserved across species (see ***Fig. 5***), only few modifications can be observed: Arg7-Gly8 (Leu-Ser in human), but there is not any salt bridge by Arg7 in the *A. thaliana* complex; Lys28 (Val in human), Val37 (Met in human) and Glu67 (Asp in human) (see ***Fig. 5***). Interestingly Glu67 is distantly located in the E2 structure to the α1-helix/β1-β2 loop module and forms a new hydrogen bond that is not present in human. Accordingly, we have produced 4 different constructs in AtSCE1 combining the aforementioned mutations with the aim to resemble as much as possible human Ubc9 and recover some conjugation activity using AtSCE1 in the human system.

Since our both constructs of the UFD domain can interact with the E2 enzymes, we decided to utilize the His-tagged UFD domain for the pull-down assays, which includes the C-terminal extension of the E1 (UFDC) for a better clearance of the protein bands. In Figure 10A, AtSCE1 pull-down loading control

for all mutants is displayed, showing in all cases similar E2 protein amounts. Our results indicate that the *A. thaliana* His-UFDC domain is able to pull-down with similar efficiency all AtSCE1 mutant constructs except the one that contains four mutations (R7L, G8S, K28V and V37M). Probably major constraints are encountered by the V37M mutation, since a similar triple mutant lacking V37M seems the bind AtUFDC with similar efficiency as the wild-type AtSCE1. Interestingly, AtUFDC domain is able to pull-down human E2, indicating that the E2-E1 UFD domain interaction is compatible.

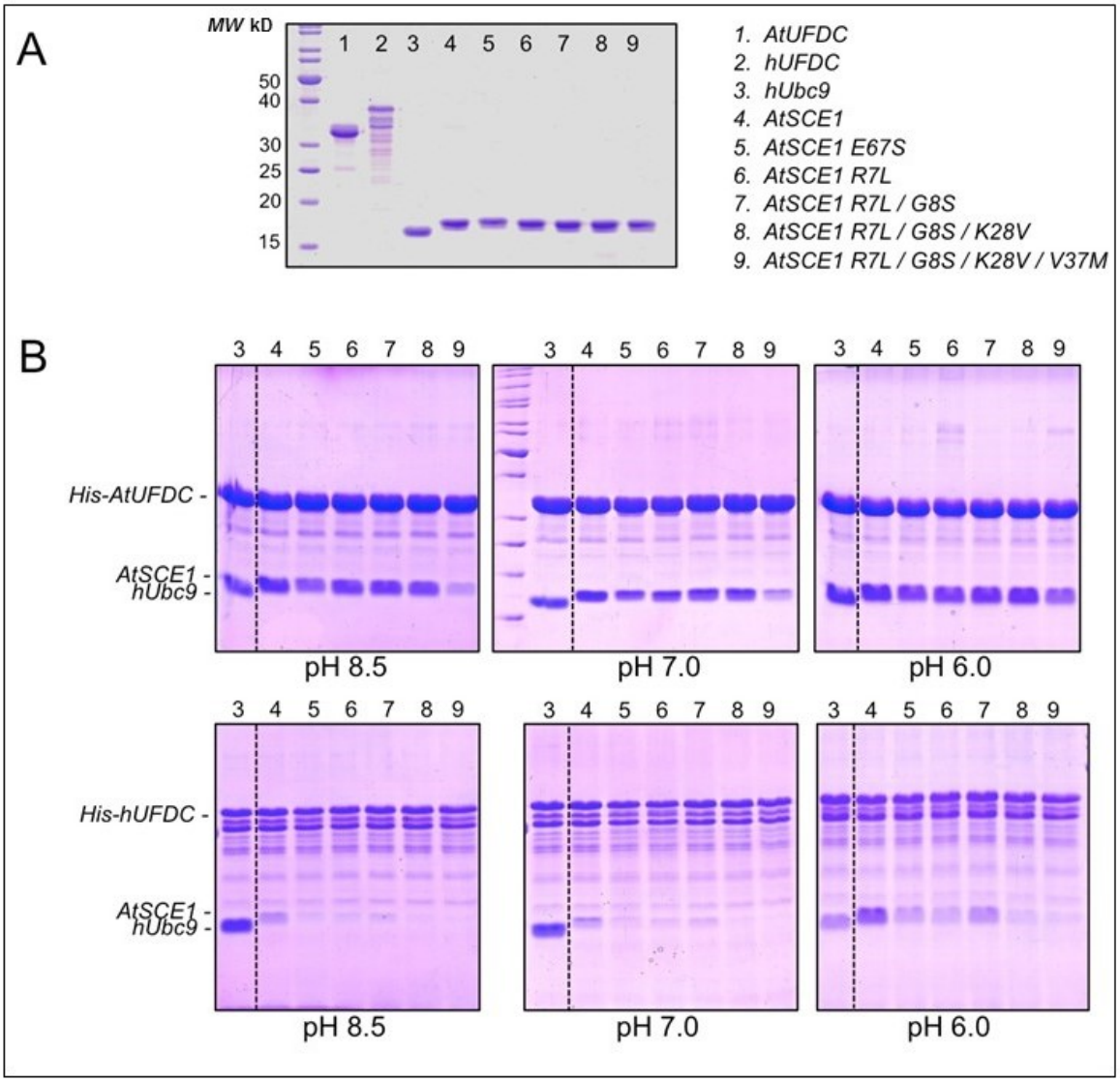


Fig 10. **Pull-down assays of AtSCE1 wt and mutants with human and *A. thaliana* UFDC domains**. *A*, PAGE of the pull-down protein amounts loading control of AtSCE1 mutant constructs. *B*, PAGE pull-down results at three different pHs using *A. thaliana* His-AtUFDC (above) or human His-hUFDC (*below*).

However, human His-UFDC domain is only able to efficiently pull-down its cognate human E2, but not any of the *A. thaliana* SCE1 mutant constructs (***Fig. 10***). Only the wild type and single point mutants of AtSCE1 can be recovered with low efficiency from the pull-down assay (the binding seems a little improved at low pHs). These results indicate that whereas the *A. thaliana* UFD can interact with human E2 efficiently, the human UFD domain have more problems to interact with *A. thaliana* E2, despite the fact that the E2 binding surface of the four-mutant construct of *A. thaliana* E2 is almost identical to the human E2 (see ***Fig.5***). Major binding problems occur when point mutants inside the β1-β2 loop (K28V and V37M) are checked, although these mutants were chosen to resemble the human E2 enzyme.

pH-dependence of the SUMO conjugation activity in human and *A. thaliana* systems

As mentioned before in purification of the complex by gel filtration (***Fig. 1***), we observed a clear pH dependence in the complex formation in the *A. thaliana* system. However, in the previous pull-down analysis (***Fig. 10***), we did not observe major differences in the binding between AtSCE1 and AtUFD domain at different pHs. These results probably suggest a better binding at low pH for the *A. thaliana* complex, which can only be appreciated in a more sensible technique such as gel filtration.

We next conducted a pH-dependence analysis of SUMO conjugation assays at a pH range between pH 6.0 and pH 8.5 for the human and *A. thaliana* system (***Fig. 11***). Interestingly, the SUMO conjugation reaction is strongly dependent of the pH, which is more prominent in the *A. thaliana* system. Interestingly, at pH 7.0 the conjugation in the *A. thaliana* system is hardly appreciated, whereas for the human system is quite strong (***Fig. 11B***).

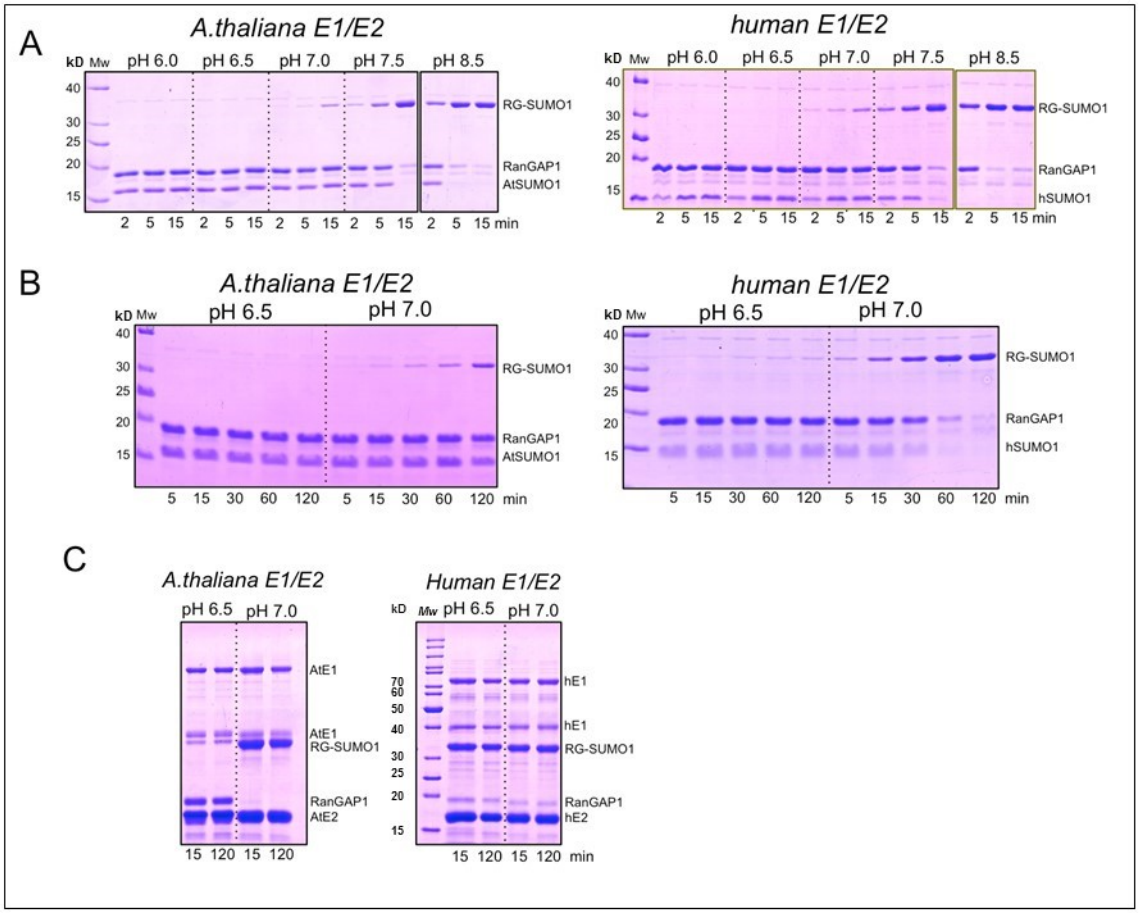


Fig 11. ***pH-dependence of the SUMO conjugation reaction in human and A. thaliana***. *A*, Representative PAGE of the time-course SUMO conjugation reaction of *A. thaliana* (*left*) and human (*right*) at different pHs. *B*, Comparison of the time-course SUMO conjugation reaction of *A. thaliana* (*left*) and human (*right*) at different pH 6.5 and pH 7.0. *C*, Similar to *B* but with 10x E1 and E2 enzyme concentration.

Since our pull-down assays did not conclude with strong binding differences in the complex between E2 and the UFD E1 domain, we believe that this pH-dependence of the SUMO conjugation probably depends on the catalysis of the isopeptide bond formation, which strongly depends on the pH for the lysine substrate nucleophilic attack on the E2-thioester bond (24).

SUMO conjugation assays with mutants of the *A. thaliana* SCE1

Interestingly, sequential and structural comparison between human and *A. thaliana* E2-conjugation enzymes, display the presence of different charged residues in the surface of the E2 next to the β1-β2 loop binding module (***Fig. 12***). In particular, whereas in human there is a salt bridge interaction between Glu78, Arg61 and Glu42, in *A. thaliana* lacks such charged-dependent contacts and those residues are substituted by Pro, Thr and His, respectively (***Fig. 12***). Interestingly, in *A. thaliana* His43 forms a salt bridge with Glu30, not present in human, which could be relevant for the correct conformation of the β1-β2 loop during the binding to the UFD domain.

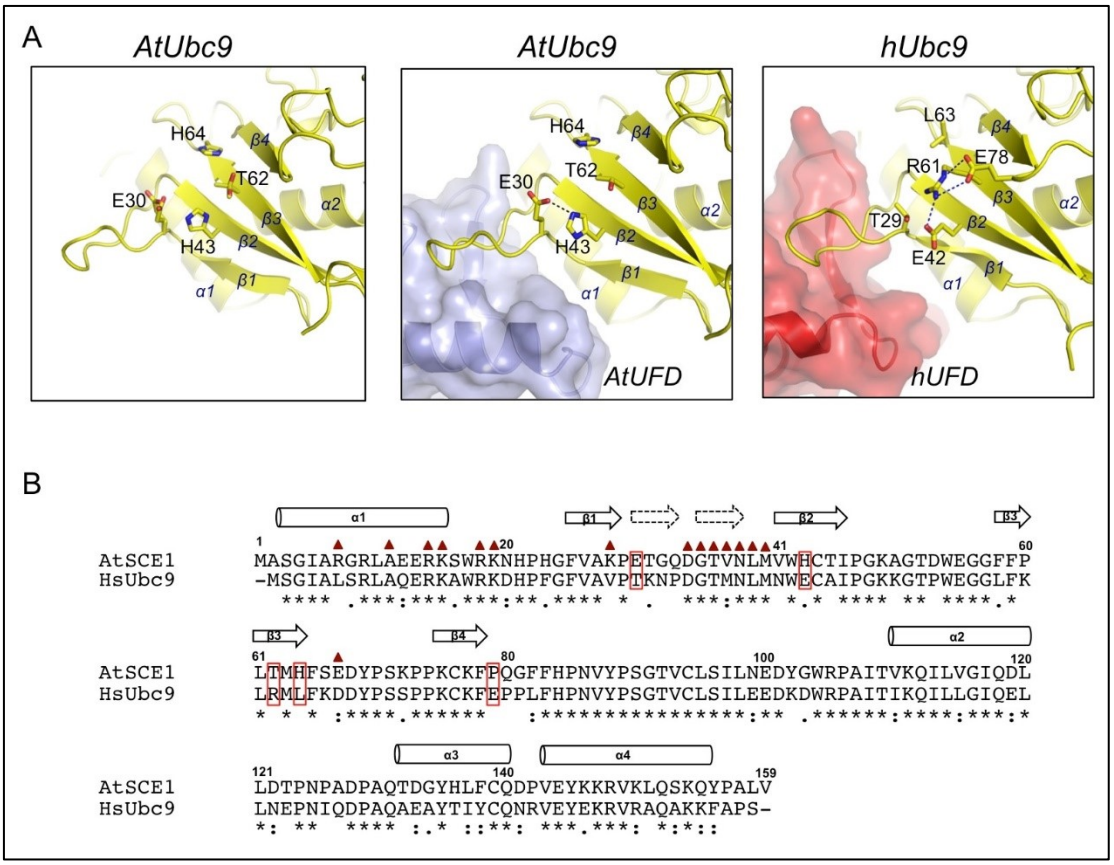


Fig 12. ***Sequential and structural comparison of the E2 surface between human and A. thaliana****. A*, Cartoon (E2) and transparent surface representation (UFD domain) of the structure of AtSCE1 alone (*left*), in complex with AtUFD (*middle*) and human Ubc9 in complex with hUFD (*right*). Non-conserved surface residues are labeled and shown in stick representation. *B*, Structural sequence alignment of *A. thaliana* AtSCE1 and human Ubc9. AtSCE1 domain secondary structure is shown on top. Red triangles indicate AtSCE1 contacts with AtUFD domain. Red rectangles indicate proposed mutagenesis residues.

We next checked the activity of two sets of AtSCE1 mutants (***Fig. 13***): five mutants of the interface with the UFD E1 domain (R7L; R7L/G8S; R7L/G8S/K28V; R7L/G8S/K28V/V37M; E67S), in this instance were only chosen for mutagenesis those unique to *A. thaliana* AtSCE1; two mutants of the E2 β-sheet surface (H43E; T62R/H64L/P79E) that are unique to the *A. thaliana* E2. In both cases all *A. thaliana* AtSCE1 mutants were replace by the corresponding human residues. One of the aims of the project was to find the structural determinants for the specificity between human and *A. thaliana* E2 enzymes.

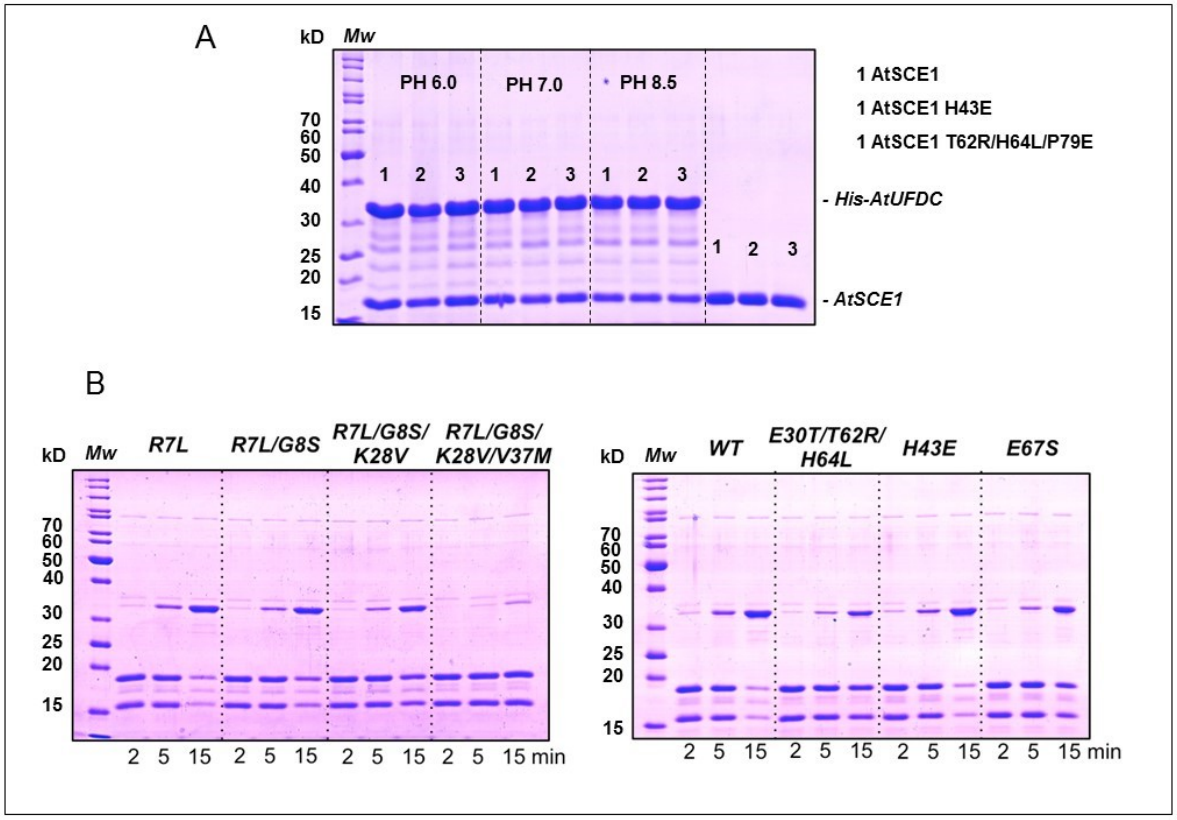


Fig 13. ***Pull down and SUMO conjugation assays of AtSCE1 WT and mutants.*** *A*, Pull-down assays of AtSCE1 wild type (WT) and point mutants of β-sheet surface residues (H43E; T62R/H64L/P79E). *B*, SUMO conjugation assays of the mutants of the *A. thaliana* AtSCE1. PAGE of the time-course of the SUMO conjugation activity in the *A. thaliana* system of different AtSCE1 mutants compared to the WT.

Regarding to the residues in the binding interface with the UFD domain (we only selected the non-conserved with human), Arg7 and Gly8 (Leu and Ser in human), present at the beginning of the E2 α1-helix, did not have any effect on the SUMO conjugation and displayed similar activities as the wild-type (***Fig. 13***). Regarding to Lys28 (Val in human) and Val37 (Met in human), both present in the β1-β2 binding region, only Val37 seems to have a strong effect on the conjugation activity and is probably caused by a binding defect (as observed in the pull-down assays, see ***Fig.10***). Glu67 (Asp in human), which forms a unique salt bridge in *A. thaliana*, had only a little effect in the SUMO conjugation reaction.

Pull-down assays did not show any difference between the WT and point mutants H43E and T62R/H64L/P79E (***Fig. 13A***). On the other hand, the SUMO conjugation reaction with the other 2 mutants (outside the UFD interface) indicate a decrease in the activity for the triple mutant, Thr62, His64 and Pro79 (Arg, Leu and Glu in human, respectively) (***Fig. 13B***), which probably indicate a role in the activity of the E1 SUMO-thioester transfer to the E2. This reduced activity points to the role of other E2 surface residues, outside the UFD binding region, involved in the E1-E2 activity, and might represent other structural determinants responsible for the specificity of the SUMO conjugation systems across species (between human and *A. thaliana*). Unfortunately, the AtSCE1 triple mutant that resembles the human E2 enzyme, cannot recapitulate the human SUMO conjugation system, even in 10x E1 and E2 concentration (***Fig. 14***).

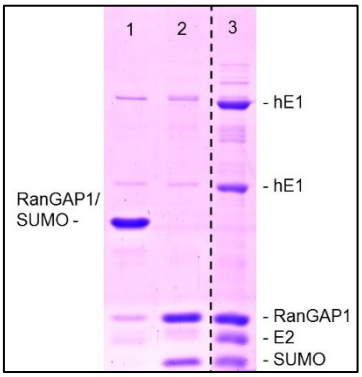


Fig 14. ***SUMO conjugation assays of AtSCE1 T62R/H64L/P79E in human system.*** End-point reaction after 120 minutes of the conjugation assays containing: human E1 and Ubc9 (1), human E1 and AtSCE1 T62R/H64L/P79E (2), 10x human E1 and AtSCE1 T62R/H64L/P79E (3).

Protein denaturation analysis of AtSCE1 mutants

Despite AtSCE1 T62R/H64L/P79E showed the same binding affinity as the wild type in the pull-down assays, it displayed a reduced RanGAP1 conjugation activity. Given that these three residues are out the UFD binding interface and locate at the surface of the protein, we inferred that mutations of these surface residues could affect the overall stability. First, we compared the center of spectral mass values of wild type AtSCE1, R7L/G8S/K28V/V37M, and T62R/H64L/P79E in PBS buffer, and all of them were similar, around 344 nM, suggesting the consistent structural integrities of these three proteins. However, in the temperature-induced protein unfolding assays, T62R/H64L/P79E displayed less stability (Tm 41.5 °C) than wild type AtSCE1 (Tm 48.4 °C), while R7L/G8S/K28V/V37M (Tm 49.1 °C) was as stable as the wild type (***Fig. 15A***). In the further urea-induced protein unfolding assays, wild type AtSCE1 and R7L/G8S/K28V/V37M displayed the half transition concentration of 2.6 M and 2.4 M urea respectively, while T62R/H64L/P79E curve presented a completely different trend and the half transition concentration of 1.6 M urea, indicating an important decrease of stability (***Fig. 15B***).

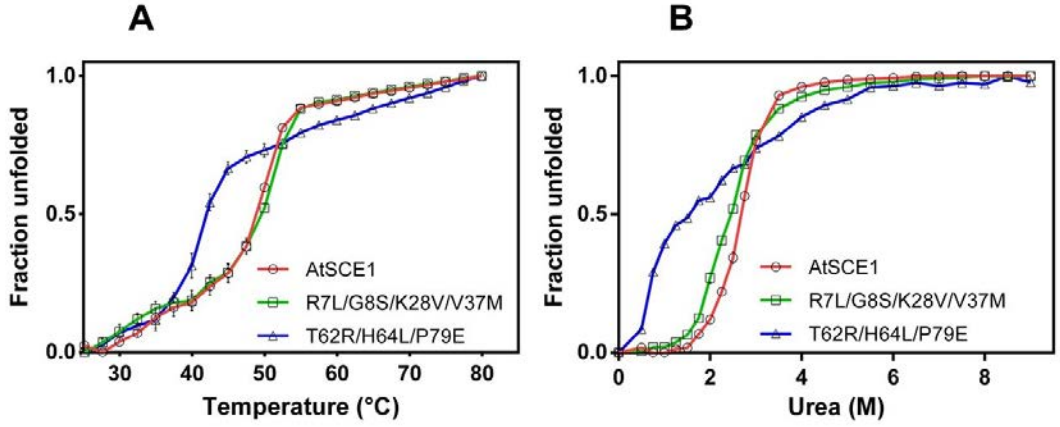


Fig 15. ***Comparison of the temperature (A) and urea (B) induced protein unfolding curves of AtSCE1, R7L/G8S/K28V/V37M, and T62R/H64L/P79E.***

Discussion and conclusions

The SUMO pathway is conserved from yeast to human. In the SUMO conjugation enzymatic cascade, the first specificity step is the interaction of the SUMO with its particular E1 activating enzyme. Next, the activated SUMO-thioester is transferred to the catalytic cysteine residue of the E2 conjugating enzyme. In this step, E1 must select cognate E2 from a range of structurally similar proteins, representing a second specificity step by the formation of the E1-E2 complex. The E1 UFD domain has been proved to play major role in E1-E2 interaction. Previous study has solved the crystal structure of UFD-E2 complex in yeast and human (14-16), indicating two alternative UFD interaction interfaces. However, it is still not clear how UFD-E2 interaction provide specificity across species, despite the high conservation between E2-conjugating proteins and the little homology between E1 UFD domains. Moreover, *A. thaliana* UFD domain is quite different to the human and yeast counterpart, thus the *A. thaliana* E1-E2 interaction cannot rely on sequence homology analysis and a structural analysis is required. Here, we performed crystallization trials and biochemical assays to investigate the UFD-E2 complex of *A. thaliana* and compared this interface with human.

The overall crystal structure of AtSCE1 is quite similar to human Ubc9, with only two differences: one is the β1-β2 loop, which is involved in the UFD domain recognition; the other is a small loop at the end of β4 strand, which adopts a different backbone orientation (***Fig. 2***). Interestingly, human Ubc9 displays a salt bridge interaction between the surface residues next to the β1-β2 loop binding module while *A. thaliana* SCE1 lacks such contacts (***Fig. 12***). Instead, *A. thaliana* His43 forms a salt bridge with Glu30, which is not present in human (***Fig. 12***). Comparison of the overall electrostatic potential of the interaction interfaces on AtSCE1 and hUbc9 revealed potential charge differences (***Fig 5b***),

possible because of the divergent residues in the α1 helix β1-β2 loop of AtSCE1 (***Fig. 2***).

In our final AtUFDC-SCE1 crystal structure, the C-terminal end was not observed even though the expressed protein contained this part, indicating a high flexibility of this region. The structure UFD-AtSCE1 complex displays a similar interaction interface as human and yeast, in which the SCE1 α1 helix and the β1-β2 loop sit on the central β-sheet of the UFD domain. Despite the main contact residues of AtSCE1 are quite conserved, there are still three contact residues, Arg7 (leucine in human), Lys28 (valine in human) and Val37 (Met36 in human), different from the human complex, as well as the UFD domain orientation also differs, which is consistent with the same alignment we made between human and yeast previously (16). These is also a unique contact between AtSCE1 Glu67 and the UFD domain, which is not found in human or yeast. Another main feature is that AtUFD domain has a long α3 helix not found in human or yeast (***Fig. 6,7***), which is also involved in the interactions with SCE1 (***Fig. 8***). These contact residue substitutions between *A. thaliana* SCE1 and human Ubc9 are accompanied by the compensatory residue substitutions within their cognate E1 UFD interface, suggesting the co-evolution of the two proteins to provide the E1-E2 interaction specificities across species.

In order to study the *A. thaliana* SUMO conjugation in vitro, we first tried to use human RanGAP1 as a substrate, since the E2 binding region with RanGAP1 is quite conserved between *A. thaliana* and human. Our biochemical assays revealed that AtSCE1 recognized the Ψ-Lys-X-Asp/Glu consensus motif and was able to conjugate SUMO to RanGAP1. Using RanGAP1 as a model substrate, we first checked if SUMOs can be interchanged between *A. thaliana* and human. The results showed that human SUMO1 & 2 can efficiently be conjugated in *A. thaliana*, although *A. thaliana* SUMO1 & 2 were less efficiently be conjugated in human. Given the high conservation of AtSCE1 active site compared with human, this difference might be because *A. thaliana* E1 cannot

effectively activate human SUMOs (***Fig. 9***). Previous study also revealed that *P. falciparum* and human E1 can recognize and activate both *P. falciparum* and human SUMOs (20). These results indicate the interchangeable possibilities for SUMOs across species. In contrast to SUMOs, AtSCE1 and human Ubc9 were not effectively interchangeable between systems in our in vitro conjugation assays (***Fig. 9B*** up). When using buffer containing 10x *A. thaliana* E1 and human E2 after incubation for 120 min at 37 °C, the SUMO conjugated RanGAP1 could be detected but still weak (***Fig. 9B*** up), while in reaction containing 10x human E1 and *A. thaliana* E2, the SUMO conjugated RanGAP1 was completely undetectable (***Fig. 9B*** *below*).

Thanks to the crystal structures of UFD-E2 complex in both *A. thaliana* and human, we could construct four chimeric AtSCE1 mutants substituting the different residues in the interface with residues from human to determine the E2 specificity. Another mutant we made is E67S, a particular E2 contact residue to UFD domain not found in human or yeast. We first performed pull-down assays to check the binding affinity at three different pHs and the results displayed no clear differences between wild type AtSCE1 and the point mutants, except the mutant containing four residue mutations (R7L/G8S/K28V/V37M), suggesting a major role of V37 in UFD-SCE1 interactions. Efforts to combine these AtSCE1 mutants with human E1 in conjugation assays failed, suggesting that the UFD-E2 specificity does not only rely on these interaction interface residues. We then ask whether these mutations affect their own conjugation. Among these AtSCE1 mutants in ***Fig. 10***, only tetra-mutants containing V37M displayed a significant decrease of RanGAP1 conjugation in further conjugation assays using *A. thaliana E1* (***Fig. 13B***), indicating an important role of this residue in the SUMO conjugation.

We also prepared other two chimeric AtSCE1 mutants, H43E and T62R/H64L/P79E, in which the mutated residues are in the E2 surface out of the UFD binding interface. Unfortunately, efforts to combine these two AtSCE1

mutants with human E1 in conjugation assays failed again (***Fig. 14***) even using 10x E1 and E2, which indicated that the UFD-E2 specificity does not only rely on these residues. We then checked the conjugation activities of these two mutants using *A. thaliana* E1 (***Fig. 13B***) and the results showed that T62R/H64L/P79E displayed substantial decrease of the conjugation activity. However, in the pull-down assays T62R/H64L/P79E did not present any difference compared with wild type AtSCE1 (***Fig. 13A***). Given that AtSCE1 T62R/H64L/P79E had no binding problems with AtUFD, it was surprising that this mutant had an insufficient activity in the conjugation assays. We then conducted protein denaturation assays to investigate whether these mutations affect the overall protein stability. In both temperature and urea induced unfolding assays, T62R/H64L/P79E displayed an important decrease of stability, while R7L/G8S/K28V/V37M did not. Thus, the insufficient conjugation activity of T62R/H64L/P79E probably because of its lower stability.

Previous efforts of interchanging SUMO E2s between *P. falciparum* and human revealed residues 44-81, which are not in the UFD binding or RanGAP1 interaction, could affect the conjugation activity (20). The UFD domain is connected to the E1 active adenylation domain through a flexible hinge. Up on E2 binding, this disordered hinge undergoes a rotation to bring catalytic cysteine residues of E1 and E2 into proximity for thioester transfer (17, 18). Moreover, the E1 E2 catalytic cysteine domains interactions also proved to be important for the SUMO conjugation (15, 19). Thus, in the E1-E2 thioester transfer, E2 is involved in diverse protein-protein interactions, which requires the structural integrity and stability. Our results revealed that, the surface residues of *A. thaliana*, T62/H64/P79, which are not in the UFD binding or RanGAP1 interaction, were responsible for the structure stability and could affect the SUMO conjugation. Given that the corresponding residues in human are quite different, these surface residues might also contribute to the specificity of SUMO systems across species other than the UFD binding interface residues.

Materials and methods

Plasmids, Cloning and Point Mutation

Expression constructs were generated by a standard PCR-based cloning method. *A. thaliana* E1 heterodimers, SUMO1 and 2, SCE1, UFD and UFDC domain, as well as human E1 heterodimers, SUMO1 and 2, Ubc9, UFD and UFDC domain were cloned to pET28a tagged with 6×His at the N-terminal. Point mutations were created using the QuickChange site-directed mutagenesis kit (Stratagene).

Recombinant protein expression and purification

Escherichia coli BL21(DE3) plysS containing the expression vector were grown in Luria Bertani medium with chloramphenicol (17 μg/mL) and kanamycine (50 μg/mL) at 37 °C until the OD600 reached to 0.8. Expression was induced by 0.1mM IPTG, followed by overnight culturing at 28 °C. Recombinant proteins were purified by nickel-nitrilotriacetic acid agarose resin (Qiagen) and dialyzed against 250 mM NaCl, 20 mM Tris-HCl (pH 8.0), 1 mM β-mercaptoethanol in the presence of thrombin protease overnight at 4 °C to remove the 6×His tag. Proteins were further purified by gel filtration chromatography on a Superdex75 column (GE Healthcare), which was pre-equilibrated in 250mM NaCl, 20mM Tris-HCl pH 7.5, 1mM β-mercaptoethanol.

Protein complex preparation and crystallization

AtSCE1 and UFD or UFDC complexes were made by mixing equimolar amounts of proteins and purified by gel filtration chromatography using a Superdex75 column. AtSCE1 and UFD or UFDC were co-eluted in a single peak in buffer containing 100 mM NaCl, HEPES pH 7.0, 1 mM β-mercaptoethanol, and confirmed by SDS-PAGE. Purified protein complex was concentrated to 15 g/L using an Amicon Ultra-10K ultrafiltration device (Millipore) prior to

crystallization. Crystals were grown at 18 °C by the sitting-drop vapor diffusion method by mixing the protein with an equal volume of reservoir solution containing 0.1 M Bis-tris pH 5.5, 25% w/v PEG3350 for AtSCE1, or containing 15% PEG6000, 5% glycerol for UFDC-AtSCE1 complex. AtSCE1 crystals appeared after one week, while UFDC-AtSCE1 crystals appeared after several months. Big crystals were soaked in mother liquor supplemented with gradually increasing concentration of 5%, 10%, 20% (v/v) Glycerol and flash frozen in liquid nitrogen.

Data collection, structure determination and refinement

Diffraction data were collected to 1.20 Å resolution for AtSCE1 and 1.8 Å for AtSCE1 at ALBA synchrotron in Barcelona (BL13-XALOC beamline) (25). Data were processed with XDS (26) and scaled, reduced, and further analyzed using CCP4 (27). More details are shown in Table 1. The structure of AtSCE1 was determined by molecular replacement method using the full length human Ubc9 (PDB code 1U9B) as a search model in PHASER (28). The structure of UFDC-AtSCE1 complex was determined by molecular replacement method using the the solved AtSCE1 as a search model in PHASER (28). Initial electron density was manually improved to build up the final model using Coot (29), and the refinement was performed using Phenix (30). Refinement statistics are shown in Table1.

In vitro SUMOylation assays

SUMO conjugation assays were performed in reactions containing 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.1% Tween 20, 50 mM NaCl, 1m M DTT, 1 mM ATP, 150 nM E1 heterodimer, 300 nM E2, 10 mM RanGAP1, and 10 mM SUMO protein. For reactions that contained 10x E1 and E2, 1.5 mM E1 and 3 mM were used. For pH-dependence experiments, the SUMO conjugation reactions were performed in 20 mM Bis-tris propane pH 6.0, 7.0 and 8.5 respectively, instead

of using HEPS pH 7.5. Reaction were incubated at 30 °C and aliquots were mixed with loading buffer for analysis by SDS-PAGE, followed by staining with coomassie brilliant blue.

In vitro pull-down assays

Ni Sepharose 6 Fast Flow beads (GE healthcare) were washed with assay buffer containing 100 mM NaCl, 20 mM Bis-tris propane pH 6.0/7.0/8.5, 0.1% Tween 20, 1 mM DTT and incubated with 30 μg His-AtUFDC or His-hUFDC for 1 hour. The beads were then incubated with the same buffer containing 2% BSA for 1 hour to block the unspecific binding, followed by washing three times with the assay buffer. 10 μg of the untagged E2 was mixed in 300 μl of assay buffer and added in the beads, followed by incubation for 1 hour. The beads were then washed three times and elute with assay buffer containing 500 mM imidazole. Elution aliquots were mixed with loading buffer and analyzed by SDS-PAGE with coomassie brilliant blue staining.

Temperature and urea induced unfolding assays

The tryptophan fluorescence spectra of temperature induced unfolding measurements were taken at excitation wavelength 280 nM and emission 340 nM in PBS buffer containing 0.15 g/L proteins. Data collection was at temperature from 25 °C to 80 °C with 1 °C steps and 1 minute between steps for equilibration in triplicates. Fraction unfolded was obtained by plotting the normalized fluorescence versus the corresponding temperatures. The urea induced unfolding were performed in PBS buffer by mixing 0.15 g/L proteins with increasing concentrations of urea (0 – 9M) at 25 °C overnight to reach equilibrium. Tryptophan emission spectra of urea induced unfolding were obtained by setting the excitation wavelength at 280 nM and collecting emission in the 300-400 nM range These spectra were quantified as the center of spectral mass (v) according to equation $v = \Sigma v_i F_i / \Sigma F_i$ where F_i stands for the fluorescence

emission at a given wavelength (v_i) and the summation is carried out over the range of appreciable values of F (31, 32). Fraction unfolded was obtained by plotting the normalized center of spectral mass versus the corresponding urea concentrations. All measurements were conducted using Jasco FP-8200 spectrofluorimeter. Carves were generated using GraphPad Prism 7.0.

References

1. Johnson ES. Protein modification by SUMO. *Annual review of biochemistry*. 2004;73(1):355-82.
2. Flotho A, Melchior F. Sumoylation: a regulatory protein modification in health and disease. *Annual review of biochemistry*. 2013;82:357-85.
3. Novatchkova M, Budhiraja R, Coupland G, Eisenhaber F, Bachmair A. SUMO conjugation in plants. *Planta*. 2004;220(1):1-8.
4. Hammoudi V, Vlachakis G, Schranz ME, van den Burg HA. Whole-genome duplications followed by tandem duplications drive diversification of the protein modifier SUMO in Angiosperms. *New Phytologist*. 2016;211(1):172-85.
5. Saracco SA, Miller MJ, Kurepa J, Vierstra RD. Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. *Plant physiology*. 2007;145(1):119-34.
6. Castaño-Miquel L, Seguí J, Lois LM. Distinctive properties of Arabidopsis SUMO paralogues support the in vivo predominant role of AtSUMO1/2 isoforms. *Biochemical Journal.* 2011;436(3):581-90.
7. Tomanov K, Ziba I, Bachmair A. SUMO chain formation by plant enzymes. *Plant Proteostasis: Springer*; 2016. p. 97-105.
8. Benlloch R, Lois LM. Sumoylation in plants: mechanistic insights and its role in drought stress. *Journal of experimental botany*. 2018;69(19):4539-54.
9. Lois LM, Lima CD. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *The EMBO journal*. 2005;24(3):439-51.
10. Huang DT, Paydar A, Zhuang M, Waddell MB, Holton JM, Schulman BA. Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. *Molecular cell*. 2005;17(3):341-50.

11. Huang DT, Hunt HW, Zhuang M, Ohi MD, Holton JM, Schulman BA. Basis for a ubiquitin-like protein thioester switch toggling E1–E2 affinity. *Nature*. 2007;445(7126):394.
12. Olsen SK, Lima CD. Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. *Molecular cell*. 2013;49(5):884-96.
13. Huang DT, Ayrault O, Hunt HW, Taherbhoy AM, Duda DM, Scott DC, et al. E2-RING expansion of the NEDD8 cascade confers specificity to cullin modification. *Molecular cell*. 2009;33(4):483-95.
14. Wang J, Taherbhoy AM, Hunt HW, Seyedin SN, Miller DW, Miller DJ, et al. Crystal structure of UBA2ufd-Ubc9: insights into E1-E2 interactions in sumo pathways. *PloS one*. 2010;5(12):e15805.
15. Reiter KH, Ramachandran A, Xia X, Boucher LE, Bosch J, Matunis MJ. Characterization and structural insights into selective E1-E2 interactions in the human and Plasmodium falciparum SUMO conjugation systems. *Journal of Biological Chemistry*. 2016;291(8):3860-70.
16. Liu B, Lois LM, Reverter D. Structural analysis and evolution of specificity of the SUMO UFD E1-E2 interactions. *Scientific reports*. 2017;7:41998.
17. Wang J, Lee B, Cai S, Fukui L, Hu W, Chen Y. Conformational transition associated with E1-E2 interaction in SUMO modifications. *Journal of Biological Chemistry*. 2009:jbc. M109. 000257.
18. Elgin ES, Sökmen N, Peterson FC, Volkman BF, Dağ Ç, Haas AL. E2-binding surface on Uba3 β-grasp domain undergoes a conformational transition. *Proteins: Structure, Function, and Bioinformatics*. 2012;80(10):2482-7.
19. 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. *Molecular cell*. 2007;27(2):228-37.
20. Reiter K, Mukhopadhyay D, Zhang H, Boucher LE, Kumar N, Bosch J, et al. Identification of biochemically distinct properties of the SUMO conjugation

pathway in Plasmodium falciparum. *Journal of Biological Chemistry*. 2013:jbc. M113. 498410.

21. Cappadocia L, Lima CD. Ubiquitin-like protein conjugation: structures, chemistry, and mechanism. *Chemical reviews*. 2017;118(3):889-918.
22. Castaño-Miquel L, Mas A, Teixeira I, Seguí J, Perearnau A, Thampi BN, et al. SUMOylation inhibition mediated by disruption of SUMO E1-E2 interactions confers plant susceptibility to necrotrophic fungal pathogens. *Molecular plant*. 2017;10(5):709-20.
23. Giraud MF, Desterro JM, Naismith JH. Structure of ubiquitin-conjugating enzyme 9 displays significant differences with other ubiquitin-conjugating enzymes which may reflect its specificity for sumo rather than ubiquitin. *Acta Crystallographica Section D: Biological Crystallography*. 1998;54(5):891-8.
24. Yunus AA, Lima CD. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nature Structural and Molecular Biology*. 2006;13(6):491.
25. Juanhuix J, Gil-Ortiz F, Cuní G, Colldelram C, Nicolás J, Lidón J, et al. Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *Journal of synchrotron radiation*. 2014;21(4):679-89.
26. Kabsch W. Xds. *Acta Crystallographica Section D: Biological Crystallography*. 2010;66(2):125-32.
27. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D*. 2011;67(4):235-42.
28. Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. *Acta Crystallographica Section D*. 2004;60(3):432-8.
29. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*. 2004;60(12):2126-32.

30. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*. 2010;66(2):213-21.
31. Silva JL, Miles EW, Weber G. Pressure dissociation and conformational drift of the. beta. dimer of tryptophan synthase. *Biochemistry*. 1986;25(19):5780-6.
32. Foguel D, Silva JL. Cold denaturation of a repressor-operator complex: the role of entropy in protein-DNA recognition. *Proceedings of the National Academy of Sciences*. 1994;91(17):8244-7.

General conclusions

1 Conclusions of USP25 research

1.1 Crystal structure of USP25 displays a homotetramer quaternary assembly and the catalytic domain can be divided in three differentiated subdomains: a general palm-like catalytic domain (USP-like), a long coiled-coil (LCC) and an “inhibitory loop” (IL).

1.2 Although USP-like catalytic domain of USP25 adopts a similar fold with other USPs, the LCC and IL domains are unique to USP25.

1.3 Each IL-loop inserts into the catalytic domain from another molecular of the tetramer and the tetramer assembly depends on IL-loop contacts to the catalytic domain.

1.4 The residues of the IL-loop involved in the tetramer interface are mostly conserved across species. In contrast, the rest residues of the IL-loop are with high sequence variation and not observed in the electron density maps.

1.5 Truncation mutants of IL form dimer and display higher activity than wild type tetramer in Ub-AMC, diubiquitin, and polyubiquitin chain hydrolysis, consistent with the inhibition role of IL-loop.

1.6 Truncation mutants of LCC-IL form monomer, consistent with its role in the dimer formation.

1.7 Point mutations of the residues involved in the IL-loop contacts to catalytic domain results in the disruption of the tetramer assembly and display higher enzymatic activity, verifying the role IL-loop in the tetramer assembly and in the deubiquitinating activity.

1.8 In vitro, either ubiquitin or the IL-loop cannot dissociate the tetramer, discarding a competition mechanism. USP25 dimer is active, whereas USP25 tetramer is inactive.

1.9 Ectopic expression of the constitutive USP25 dimer stabilizes tankyrases in HEK293T cells, confirming the auto inhibition mechanism.

2 Conclusions of SUMO E1-E2 protein-protein recognition

2.1 Crystal structures of the E1 UFD-E2 complexes reveal the presence of three different types of binding interfaces in human, yeast, and *A. thaliana*.

2.2 However, despite the low E1 UFD structural and sequence homology, structural comparison reveals common determinants in the E2 binding interfaces between human, yeast, and *A. thalia*na.

2.3 Phylogenetic analysis reveals that UFD domains are quite conserved across closely related species, especially only taking the E2 binding region into account, revealing a high selective pressure to ensure specificity.

2.4 Structural superpositions of the E2 enzymes between human, yeast and *A. thaliana* complexes reveal different orientations of the UFD domain, which probably account for the specificity of each SUMO conjugation system to interact with its cognate E2-conjugating enzyme.

2.5 The contact residue substitutions between human, yeast, and *A. thaliana* E2s are accompanied by the compensatory residue substitutions within their cognate E1 UFD interface, suggesting the co-evolution of the two proteins directed by the E1-E2 interaction.

2.6 Human SUMO1 & 2 can efficiently be conjugated by the *A. thaliana* enzymes, although *A. thaliana* SUMO1 & 2 are less efficiently be conjugated by human enzymes, indicating SUMO interchangeability across species.

2.7 *A. thaliana* E2 (SCE1) cannot conjugate SUMO to RanGAP1 in the human system, even after a complete substitution of *A. thaliana* residues in the UFD binding interface by human residues.

2.8 Mutation of *A. thaliana* E2 (SCE1) surface residues T62/H64/P79, outside the UFD binding, results in the decrease of protein stability and a decline of SUMO conjugation. These surface residues might also contribute to specificity across species in addition to the UFD binding interface.

General discussion

Ubiquitination and SUMOylation are two of the most studied post-translational modifications (PTMs). In this thesis dissertation, we have studied structural and functional characteristics of protein-protein interactions belonging to these two PTM pathways: structures of the deubiquitinase USP25, and of the SUMO E1 UFD-E2 protein complexes. USP25 and USP28 are important deubiquitinating enzymes whose activities are regulated by multiply PTMs including phosphorylation, ubiquitination, and SUMOylation, while the E1 UFD-E2 interaction is a major specificity step in the SUMO conjugation pathway by the recognition of E1 and E2. The structural and functional studies of USP25 reveal a novel autoinhibitory mechanism, adding a new layer of complexity to the USP25 regulation and expanding the spectrum of different types of deubiquitinating enzymes (DUBs) regulations. On the other hand, the crystal structures of UFD-E2 complexes in human and *A. thaliana* indicate the different interaction interfaces of SUMO E1 recruitment of E2, providing insights into the specificities of SUMO E1-E2 protein-protein interactions.

Our previous study revealed that SUMOylation at K99 of USP28 inhibited its ubiquitin chain hydrolysis activity (1), consistent with a similar study on USP25 (2). Thus, we were interested in the structure basis for USP25 and USP28 activity regulations. We conducted crystallization trials of USP25 and USP28, but only USP25 gave rise to single diffractable crystals. At the beginning, the visually well-formed crystals of USP25 diffracted poorly to a resolution that is too low to be suitable for structure determination. We tried *in situ* proteolysis and post-crystallization treatments including annealing, dehydration, soaking and cross-linking, to improve the diffraction quality. However, none of these treatments could improve the diffraction quality (3, 4). Then we tried another strategy, lysine methylation (5), for protein crystallization, and the USP25 crystals could diffract to 3.28 Å and the structure could be finally solved by single-anomalous diffraction (SAD) after soaking mercury derivatives.

In our final structure of USP25, the N-terminal domain (NTD, residues 18-158)

was not observed, probably due to its high flexibility and the lack of a structured domain. The NTD is important for the recruitment of polyubiquitin chains due to the presence of two ubiquitin-interacting motifs (IUMs). Interestingly, loss of this domain resulted in the insufficient hydrolysis of K48-linked ubiquitin chains but not the single synthetic Ub-AMC substrate (2). Consistent with this, we also found that the NTD loss lead to impaired hydrolysis activity of K63-linked ubiquitin chains. Sequence analysis showed that USP25 UIM1 is a double side ubiquitin binding motif (6), while UIM2 was not, probably indicating their distinct roles. Thus, structure insights of UIM1 and UIM2 recruitment of ubiquitin chains are required, despite the tandem UIMs of USP25 might confer the K48-linked ubiquitin chain substrate preference (7).

The crystal structure of USP25 displays an unexpected homotetrameric quatemary assembly, which is assembled by the association of two dimers and includes contacts between the IL-loop and the ubiquitin-binding pocket at the catalytic domain. Biochemical analysis conforms that the IL loop play roles in the autoinhibition of its enzymatic activity. So, we have discovered two different structural assemblies for USP25, dimer (active) and tetramer (inactive). This tetramerization-dependent inhibition mechanism of USP25 has never been described in the regulatory mechanisms for other members in the DUB family and might represent a novel regulation mechanism of deubiquitinating activity (8).

We next speculated whether USP28 also has the same inhibition mechanism, which shares more than 50% sequence identity with USP25, and both display hydrolysis activity for K48- and K63- linked ubiquitin chains. In our primary attempt, we made the same truncation mutant as USP25, but the USP28 mutants displayed expression problems of degradation. Given that the IL-loop domains of USP25 and USP28 are not well conserved, the degradation problems might be due to the wrong amino acid truncation of USP28. Moreover, the IL-loop contact residues of USP25 are highly conserved across species, but for USP28, this conservation is not so clear. In fact, differences between USP25 and USP28 also rely on substrate recognition. Evidences have been shown that USP28 is required to MYC stability and involved

in p53 activation (9, 10), while USP25 have not been found related to these regulations. In contrast, the C-terminal tail of USP25 (residues RTPADGR), which is highly conserved across species, is involved in the binding to tankyrases, promoting their deubiquitination and stabilization (11). The corresponding C-terminal tail residues of USP28 are different from USP25 and completely not conserved across species, suggesting that USP28 probably do not have this directly binding and regulation role for tankyrases. Other USP25 biological roles have been review in the introduction part. All in all, despite of high protein sequence identity and same type of ubiquitin chain hydrolysis activity, the biological roles of USP25 and USP28 are quite diverse. Since their domain architectures are very similar, and USP28 also displays as an oligomer in recombinant expression, more efforts are worth to take to confirm whether this novel autoinhibitory mechanism exist in USP28.

Our in vitro protein expression data showed an equilibrium of active (tetramer) and inactive (dimer) state of USP25. The point mutation of Y454E, a phosphomimetic tyrosine substitution, results in the formation of a stable dimer. Thus, even though the presence of high number of contacts in the tetramer, a single point mutant in the interface, such as Y454E, is capable to destabilize the tetramer and render USP25 to an “active” dimer state. Phosphorylation of a single residue of the USP catalytic domain has already been shown to regulate the deubiquitinating activity in USP14 and USP37 (12, 13), so we speculate whether it would be a plausible mechanism to switch between these two oligomer assemblies in USP25. It would also be reasonable to find determinants that control the tetramer and dimer states transformation in cells. Finally, only 8 to 10 residues on IL loop bind to USP25, suggesting the possibility to find lead compounds to mimic this specific binding and inhibit the USP25 activity, since USP25 is involved in several pathologies.

Compared with the ubiquitin pathway, relatively few components of SUMO conjugation machinery have been described. Important, Ubc9 is the only known E2-conjugating enzyme and can recognize the Ψ-Lys-X-Asp/Glu consensus motif and directly conjugate SUMO to substrate (14). The SUMO E1 is a heterodimer

composed of two subunits and select the cognate E2 through the ubiquitin fold domain (UFD) (15). This E1 UFD-E2 interaction can also provide specificity across species, despite the high homology between E2-conjugating proteins. Previous study has solved the crystal structure of UFD-E2 complex in yeast (16), showing the direct binding of E2 to the E1 UFD domain. However, how the UFD-E2 interaction interface provides different specificities across species is still not clear.

In this thesis, we solved the crystal structure of UFD-E2 complexes in both human and *A. thaliana*. In the final structure, the C-terminal extension of E1 UFD was not apparent in electron density maps, despite its presence in our protein and crystalline preparations, suggesting that this portion of amino acids are not directly involved in SUMO E1-E2 interaction. The UFD domains only show a little protein sequence homology across species, but our primary phylogenetic analysis revealed that UFD domains are highly conserved across closely related species, especially only taking the E2 binding region into account. Structural comparison of the UFD-E2 complexes between human, yeast, and *A. thalia*na revealed the presence of three different types of binding interfaces, even though common determinants existed. In addition, the contact residue substitutions between human, yeast, and *A. thaliana* E2s are accompanied by the compensatory residue substitutions with their cognate E1 UFD interface, suggesting the co-evolution of the two proteins directed by the E1-E2 interaction.

In vitro SUMO conjugation assays showed that SUMO1 & 2 are interchangeable between human and *A. thalia*na, consistent with the similar study on *P. falciparum* and human (17). However, AtSCE1 and human Ubc9 were not interchangeable. Thanks to the crystal structures of UFD-E2 complexes, we could construct chimeric AtSCE1 mutants substituting the different residues in the binding interface with residues from human to determine the E2 specificity. We also prepared other two chimeric AtSCE1 mutants, in which the mutated residues are in the E2 surface out of the UFD binding interface. Unfortunately, all the chimeric AtSCE1 failed to conjugate SUMO in human system. Further pull-down and

conjugation assays using these chimeric ATSCE1 suggested a major role of V37 in UFD-SCE1 interactions. Interestingly, mutation of *A. thaliana* E2 (SCE1) surface residues T62/H64/P79, which is outside the UFD binding, displayed no binding problems to UFD domain in pull-down assays, but lead to the decrease of protein stability and a reduced SUMO conjugation activity. Previous efforts of interchanging SUMO E2s between P. falciparum and human revealed residues 44-81, which are not in the UFD binding or RanGAP1 interaction, could affect the conjugation activity (17). Given that the corresponding residues of T62/H64/P79 in human are quite different, these surface residues might also contribute to the specificity of SUMO systems across species.

References for general discussion

1. Zhen Y, Knobel PA, Stracker TH, Reverter D. Regulation of USP28 de-ubiquitinating activity by SUMO conjugation. *Journal of Biological Chemistry*. 2014:jbc. M114. 601849.
2. Meulmeester E, Kunze M, Hsiao HH, Urlaub H, Melchior F. Mechanism and consequences for paralog-specific sumoylation of ubiquitin-specific protease 25. *Molecular cell*. 2008;30(5):610-9.
3. Heras B, Martin JL. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D: Biological Crystallography*. 2005;61(9):1173-80.
4. Dong A, Xu X, Edwards AM, Chang C, Chruszcz M, Cuff M, et al. In situ proteolysis for protein crystallization and structure determination. *Nature methods*. 2007;4(12):1019.
5. Walter TS, Meier C, Assenberg R, Au K-F, Ren J, Verma A, et al. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure*. 2006;14(11):1617-22.
6. Hirano S, Kawasaki M, Ura H, Kato R, Raiborg C, Stenmark H, et al. Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nature Structural and Molecular Biology*. 2006;13(3):272.
7. Kawaguchi K, Uo K, Tanaka T, Komada M. Tandem UIMs confer Lys48 ubiquitin chain substrate preference to deubiquitinase USP25. *Scientific reports*. 2017;7:45037.
8. Mevissen TE, Komander D. Mechanisms of deubiquitinase specificity and regulation. *Annual review of biochemistry.* 2017;86:159-92.
9. Meitinger F, Anzola JV, Kaulich M, Richardson A, Stender JD, Benner C, et al. 53BP1 and USP28 mediate p53 activation and G1 arrest after centrosome loss or extended mitotic duration. *J Cell Biol.* 2016;214(2):155-66.
10. Popov N, Wanzel M, Madiredjo M, Zhang D, Beijersbergen R, Bernards R, et

al. The ubiquitin-specific protease USP28 is required for MYC stability. *Nature cell biology*. 2007;9(7):765.

11. Xu D, Liu J, Fu T, Shan B, Qian L, Pan L, et al. USP25 regulates Wnt signaling by controlling the stability of tankyrases. *Genes & development*. 2017.
12. Huang X, Summers MK, Pham V, Lill JR, Liu J, Lee G, et al. Deubiquitinase USP37 is activated by CDK2 to antagonize APCCDH1 and promote S phase entry. *Molecular cell*. 2011;42(4):511-23.
13. Xu D, Shan B, Lee B-H, Zhu K, Zhang T, Sun H, et al. Phosphorylation and activation of ubiquitin-specific protease-14 by Akt regulates the ubiquitin-proteasome system. *Elife*. 2015;4:e10510.
14. Bernier-Villamor V, Sampson DA, Matunis MJ, Lima CD. Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell*. 2002;108(3):345-56.
15. Lois LM, Lima CD. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *The EMBO journal*. 2005;24(3):439-51.
16. Wang J, Taherbhoy AM, Hunt HW, Seyedin SN, Miller DW, Miller DJ, et al. Crystal structure of UBA2ufd-Ubc9: insights into E1-E2 interactions in sumo pathways. *PloS one*. 2010;5(12):e15805.
17. Reiter K, Mukhopadhyay D, Zhang H, Boucher LE, Kumar N, Bosch J, et al. Identification of biochemically distinct properties of the SUMO conjugation pathway in Plasmodium falciparum. *Journal of Biological Chemistry*. 2013:jbc. M113. 498410.

Summary in English

Ubiquitination and SUMOylation are of the most studied post-translational modifications (PTMs). Here, we focus on USP25, USP28, and the SUMO E1-E2 protein-protein recognition in these two PTM pathways. USP25 and USP28 have important roles in cellular processes, and their enzymatic activities are regulated by diverse PTMs including SUMOylation, ubiquitination, and phosphorylation. SUMO E1-E2 protein-protein interaction is a major discrimination step in the conjugation pathway. In this thesis, the main goals include the elucidation of the structural basis for the activity regulation of USP25 and USP28, as well as to decipher the structural determinants for the specificity provided by the E1 UFD-E2 interaction.

We have solved the crystal structure of human USP25 (18 – 714). Unexpectedly, USP25 displays a homotetrameric quaternary assembly that is directly involved in the inhibition of its enzymatic activity, revealing a novel tetramerization/inhibition mechanism. In vitro biochemical and kinetic assays with dimer, tetramer and truncation constructs of USP25 support this mechanism, displaying in all cases a higher catalytic activity in the dimer assembly. Moreover, the strong stabilization of tankyrases in cultured cells by the ectopic expression of the USP25 dimer verifies the biological relevance of this novel tetramerization/inhibition mechanism.

Regarding to the E1 UFD-E2 interaction, we have solved the crystal structure of the E1 UFD-E2 complex in both human and *A. thaliana*. Despite the low sequence homology displayed by the UFD binding interface, structural comparison between complexes reveals common determinants in the interfaces between human, yeast, and *A. thalia*na. Structural comparison also reveals a strong conservation in the E2 binding interface across species, despite the strong specificity displayed in SUMO conjugation assays for each organism. Interestingly, E2 residues outside the UFD interface had impact on SUMO conjugation, suggesting the contribution of determinants other than the primary UFD binding interface in the specificity of the conjugation system.

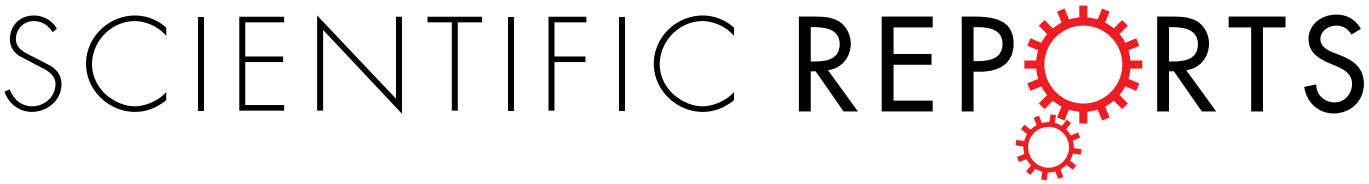
Summary in Catalán

La ubiquitinació i la SUMOilació són dos de les modificacions post-traduccionals més estudiades (PTM). En aquesta tesi, ens hem centrat en estudiar USP25, USP28 i en el reconeixement proteïna-proteïna dels enzims SUMO E1-E2 d'aquestes dues vies PTM. USP25 i USP28 són de-ubiquitinases que tenen papers importants en processos cel·lulars i les seves activitats enzimàtiques estan regulades per diversos PTMs, incloent SUMOilació, ubiquitinació i fosforilació. La interacció proteïna-proteïna de SUMO E1-E2 és un pas essencial de discriminació en la via de la conjugació. En aquesta tesi, els principals objectius inclouen l'elucidació de les bases estructurals de la regulació de l'activitat de USP25 i USP28, així com poder desxifrar els determinants estructurals en l'especificitat proporcionada per la interacció SUMO E1-E2.

Hem resolt l'estructura cristal·lina de la USP25 humana (18 - 714). Sorprenentment, USP25 mostra una estructura quaternària homotetramèrica que està directament implicat en la inhibició de la seva activitat enzimàtica, i revela un nou mecanisme de tetrameització/inhibició. Assaigs bioquímics i cinètics in vitro amb construccions de dímer, tetràmer de USP25 donen suport a aquest mecanisme, mostrant en tots els casos una major activitat catalítica en el dímer. A més, la forta estabilització de les *tankirases* en cèl·lules en cultiu mitjançant l'expressió ectòpica del dímer de USP25 verifica la rellevància biològica d'aquest nou mecanisme de tetrameització/inhibició.

Pel que fa a la interacció E1 UFD - E2, hem resolt l'estructura cristal·lina del complex E1 UFD - E2 tant en humans com en *A. thaliana*. Tot i la baixa homologia de seqüència presentada en la superfície d'interacció de UFD entre especies, la comparació estructural entre complexos revela determinants comuns en les interfícies entre humans, llevats i *A. thaliana*. La comparació estructural també revela una forta conservació en la superfície d'interacció de les E2 entre espècies, tot i la forta especificitat mostrada en assajos de conjugació de SUMO en a cada organisme. Curiosament, hem trobat residus de la E2 fora de la superfície d'interacció amb UFD que tenen un impacte en la conjugació de SUMO, el que suggereix la presència d'altres determinants estructurals diferents als de la superfície d'interacció entre E2 i E1 UFD en l'especificitat de la conjugació de SUMO.

Publications



OPEN

Structural analysis and evolution of specificity of the SUMO UFD E1-E2 interactions

Received: 16 November 2016
Accepted: 04 January 2017
Published: 06 February 2017

Bing Liu[1], L. Maria Lois[2] & David Reverter[1]

SUMO belongs to the ubiquitin-like family (UbL) of protein modifiers. SUMO is conserved among eukaryotes and is essential for the regulation of processes such as DNA damage repair, transcription, DNA replication and mitosis. UbL modification of proteins occurs via a specific enzymatic cascade formed by the crosstalk between the E1-activating enzyme, the E2-conjugating enzyme and the E3-ligase. An essential discrimination step in all UbL modifiers corresponds to the interaction between E1 and E2 enzymes, which is mediated by the recruitment of the E2 to the UFD domain (Ubiquitin-Fold Domain) of the E1 enzyme. To gain insights in the properties of this interface, we have compared the structures of the complexes between E1 UFD domain and E2 in human and yeast, revealing two alternative UFD platforms that interact with a conserved E2. Comparative sequence analysis of the E1 UFD domain indicates that the E2 binding region has been conserved across phylogenetic closely related species, in which higher sequence conservation can be found in the E2 binding region than in the entire UFD domain. These distinctive strategies for E1-E2 interactions through the UFD domain might be the consequence of a high selective pressure to ensure specificity of each modifier conjugation system.

The post-translational modification pathway of proteins by UbLs (Ubiquitin-like modifiers) is characterized by the presence of specific enzymatic cascades (E1, E2 and E3 enzymes), which results in the formation of an isopeptidic bond between the UbL and the protein target[1,2]. A major characteristic of this process is the specificity provided between all components of the UbL pathway, resulting in the formation of protein-protein complementary interfaces[3,4]. SUMO (Small Ubiquitin-Like Modifier) is a UbL modifier that can alter the function of a myriad of target proteins inside the cell[5], being involved in processes such as DNA damage repair, transcription, DNA replication and mitosis[1,6,7].

The first specificity step in the pathway corresponds to the interaction of the UbL modifier with its particular E1-activating enzyme[8,9]. In the SUMO pathway, the E1-activating enzyme is a large multidomain heterodimer (Sae1-Uba2)[10] that initiates the process by adenylation of the SUMO C-terminus and the subsequently formation of a thioester bond with the active-site cysteine residue of E1[11]. Next, the activated UbL-thioester is transferred to the active-site cysteine residue of the E2-conjugating enzyme (Ubc9 in SUMO), which represents a second specificity step in the pathway by the formation of the E1-E2 complex. Crystal structures of several E1-activating enzymes[12–15] revealed the presence of a domain displaying an Ubiquitin-like Fold (UFD domain) in the E1-activating enzyme large subunit. The E1 UFD domain plays a major role in the binding of the E2 enzyme, providing specific contacts between E1 and E2 enzymes. This interaction was first observed in the crystal structure of the Nedd8 E1 in complex with E2[13], showing the direct binding of E2 to the E1 UFD domain. Recently, the crystal structure of the thio-ester transfer intermediate of ubiquitin E1-E2 complex[15] revealed a dual binding of E2 to the UFD domain and to the catalytic E1 Cys-domain, which occurs after an significant rotation of the UFD domain, providing the structural basis for the isoenergetic thio-ester transfer between the E1 and the E2 enzymes[15]. This interaction between E2 and the Cys domain of the E1 was proposed previously in the SUMO pathway by NMR analyses, although E1 UFD-E2 interactions display higher affinity ($K_d = 1.2\,\mu M$)[16] than E1 Cys-E2 interactions ($K_d = 87\,\mu M$)[17], supporting a major role of the E1 UFD domain in E2 recruitment.

[1]Institut de Biotecnologia i de Biomedicina, Departament de Bioquimica i Biologia Molecular, Serra Hunter Fellow, Universitat Autonoma de Barcelona, 08193 Barcelona, Bellaterra, Spain. [2]Center for Research in Agricultural Genomics-CRAG, Edifici CRAG-Campus UAB, Bellaterra 08193 Barcelona, Spain. Correspondence and requests for materials should be addressed to L.M.L. (email: maria.lois@cragenomica.es) or D.R. (email: david.reverter@uab.cat)

Protein sequence variations in the E1 UFD domain of different UbL modifiers are quite significant, especially in the binding region to the E2 enzyme. All reported structures of UFD domains display an analogous β-grasp structure, and the interaction of the UFD domain with the E2-enzyme occurs through the same side of the β-sheet structure[13–15,18–20]. However, in all reported complex structures superposition of the E2 enzymes reveal distinct orientations of the UFD domain, which is a direct consequence of different contacts in each UbL system[15,18]. Conservation analysis according to sequence alignments showed that yeast and human SUMO UFD domains display little sequence homology (17% sequence identity), and it is even lower considering only the binding region to Ubc9. Notwithstanding this low conservation, both proteins can efficiently interact with a highly conserved surface in their cognate Ubc9. In this scenario, the identification of the molecular determinants that mediate E1 UFD and E2 interactions in evolutionary distant organisms cannot rely on sequence homology analysis. Instead, the elucidation of these molecular determinants requires specific structural studies of the interaction.

Here we present a detailed structural comparison analysis of the two complexes between SUMO E1 UFD domain and Ubc9 from yeast[21] and human. We also present a novel structure of the human complex solved in a different space group than the recently deposited[22]. Our results indicate that human and yeast UFD domains interact with a conserved surface of Ubc9, in each case by maintaining the same chemical character of the interface contacts despite the lack of sequence homology. Sequence alignment of these two E2 binding region discloses unique consensus motifs that have been maintained across species from the same kingdom (in *Metazoa*) or in the same order (in *Saccharomycetales*). Phylogenetic and homology analysis revealed that the region involved in Ubc9 binding displays a slightly higher conservation degree than the UFD domain between phylogenetically closely related organisms, although it also displays higher variability, highlighting the relevance of this interface in the protein-protein specificity for each type of UbL modification.

Results

Complex between human SUMO E1 UFD domain and SUMO E2. The interaction between the ubiquitin-fold domain (UFD) of the E1-activating enzyme and the E2-conjugating enzyme has been revealed as a crucial discrimination step in the conjugation pathway of UbL modifiers[14,18,23,24]. The interface between E1 and E2 enzymes is unique and is required to confer specificity between cognate enzymes of each UbL conjugation pathway. Notably, differences in the interface are more significant in the UFD domain than in the E2-conjugating enzyme, which is a highly conserved enzyme. Within the same UbL family, evolutionary distant species display a low degree of sequence conservation between E1 UFD domains. In SUMO pathway, yeast and human UFD domains were shown to display only a 17% sequence identity[25]. To get structural insights for the different interaction between yeast and human, we have determined the crystal structure of the complex between human E1 Uba2 UFD domain and human E2 Ubc9 at 2.2 Å resolution and compared this structure to the analogous complex in yeast (PDB code 3ONG)[25]. During the preparation of the manuscript, another structure of the human E1 UFD-Ubc9 was also published in a different crystallographic space group, supporting our results[22] (PDB ID 4W5V).

Human E1 ubiquitin-fold domain (UFD) was designed based on the structure of the full-length human SAE1-SAE2 E1-activating enzyme (PDB code 1Y8Q)[24]. We alternatively prepared a UFD construct including the C-terminal flexible extension, however this longer UFD was unstable and displayed proteolysis after complex formation with Ubc9. Both native gel electrophoresis and gel filtration chromatography indicated the formation of the complex between human E1 UFD domain and Ubc9 (see Supplementary Fig. S1). After initial unfruitful crystallization trials with the purified complex, a lysine methylation protocol was conducted to induce crystallization. Suitable crystals diffracted beyond 2.2 Å resolution, contained one complex of UFD/Ubc9 per asymmetric unit and belonged to the tetragonal space group system (Table 1), which is different to the monoclinic crystals recently deposited[22] (PDB code 4W5V).

An overall structural comparison of free and bound Ubc9 structures displayed little variation (0.71 Å rmsd Cα deviation), although differences were observed in the binding region to UFD, displaying rmsd Cα deviations between 1.50 Å and 3.66 Å in the residues forming the β1-β2 loop (Lys30 to Met36). Similarly, free and bound E1 UFD structures are almost identical (overall 0.72 Å rmsd deviation). According to the PISA server[26] the human complex buries a surface of 1493 A^2 and involves 26 and 19 residues of UFD and Ubc9, respectively, which is comparable to the analogous complex in yeast (1557 A^2 interface)[25].

Yeast and human interface comparison. In contrast to the conservation proposed according to sequence homology between human and yeast UFD, 17% of identity, this homology is even lower, 11%, when a structural alignment is performed (Fig. 1a). Interestingly, differences between yeast and human are increased, 7% of sequence identity, when comparing only the residues forming the interface of the E1 UFD domain with Ubc9, which are basically formed by a different set of residues capable to interact with a conserved Ubc9 surface (Fig. 1b). However, despite this low sequence homology between yeast and human UFD, the interaction occurs through the same surface[22,25], forming an interdigitated complex between the α1 helix and the β1-β2 loop of Ubc9 that sits on the β-sheet surface of UFD (Fig. 1c). As a consequence of this low conservation, structural superposition of Ubc9 reveals a rotation of the E1 UFD domain between both complexes (Fig. 1e). Different orientations of UFD domains in complex with E2 can also be observed in other UbL systems, as revealed by the structures of ubiquitin and Nedd8 UFD in complex with E2 (Uba1-Ubc4 complex PDB code 4II2; and Uba3-Ubc12, pdb code 2NVU refs 13 and 15). All these structures suggest plasticity in the UFD interface that has evolved to specifically interact with its cognate E2-conjugating enzyme in each UbL pathway.

The binding surface of Ubc9 in yeast and human is highly conserved, and is basically formed by similar backbone and side-chain interactions in the α1 helix and the β1-β2 loop of Ubc9 (yeast and human Ubc9 share 56% sequence identity). Specific side-chain interactions in Ubc9 include a hydrophobic patch formed by Leu6, Met36, and Leu38, and a basic patch formed by Arg13, Lys14, Arg17 and Lys18 (Fig. 2a,b). Interestingly, all these conserved residues in Ubc9 engage specific contacts with the non-conserved UFD surfaces of yeast and human.

Data collection	
Beamline	ALBA-XALOC
Space group	$P4_32_12$
Wavelength (Å)	0.97946
Resolution (Å)	46.44–2.20 (2.32–2.20)*
a, b, c (Å)	129.61, 129.61, 66.60
α, β, γ (°)	α = β = γ = 90
Unique reflections	29309
Data redundancy	10.9 (11.0)
R_{merge}	0.043 (0.602)
I/σ	26.4 (4.3)
Completeness (%)	99.8 (98.4)
Refinement	
Resolution (Å)	46.44–2.20
Unique reflections	29254
R_{work}/R_{free}	0.22/0.24
Number of all atoms	2100
Number of waters	38
RMSD bond (Å)/Angle (°)	0.008/1.11
Average B factor (protein/water)	60.69/56.26
Ramachandran plot	
Favored (%)	97.64
Allowed (%)	2.36
Disallowed (%)	0.00

Table 1. Summary of crystallographic analysis. *Highest resolution shell is shown in parenthesis.

A major contact difference in Ubc9 corresponds to Ala10, which is substituted by Gln10 in yeast, forming a polar interaction with Glu515 of the yeast UFD domain (Fig. 1d). Previous point mutational analysis on these two patches of Ubc9, namely the basic α1 helix and the hydrophobic β1-β2 loop, showed impairment in the binding to the E1 activating enzyme, indicating a major role of these two regions in the transfer of SUMO between E1 and E2 proteins[25,27].

The UFD interacting surface is extended and mostly formed by residues emanating from β22 and β23 strands and connecting loops (Fig. 1c,d). A general feature of this interaction, widespread so far in all characterized UbL pathways, is the interaction through the same β-sheet surface (Fig. 2). But in contrast to the Ubc9, the region involved in UFD is poorly conserved between yeast and human. For simplicity, we have designated this region as *Low Homology region involved in E2 Binding 2* (LHEB2). The LHEB2 can be divided in two regions, each establishing interactions with the basic and hydrophobic patches of Ubc9 (Fig. 2a,b). The first contact region in human UFD is composed by Asp479, Ser492, Ser493, Glu494 and Glu497, which engage polar and charged contacts with the basic patch of Ubc9, composed by α1 helix residues Arg13 and Arg17. The LHEB2 in human is highly conserved in all *metazoan* species analyzed (see later in Fig. 3). In contrast, in yeast the basic patch of Ubc9 interacts with Asp488, Tyr489 and Asp490, which is also a highly conserved sequence in all *Saccharomycetales* species analyzed (see later in Fig. 3). Interestingly, in the human structure Phe522 is buried in an aliphatic pocket formed by Arg17, Lys14 and Lys18 (see Fig. 1c), whereas in yeast a similar Ubc9 pocket buries Tyr489 (see Fig. 1d), a residue located at the center of the LHEB2 sequence instead of the LHEB2 C-terminal position that occupies its human counterpart.

The second contact region of UFD interacts with the hydrophobic patch of Ubc9. This interface is composed by backbone and side chain interactions emanating from the β22-β23 connecting loop and β23 strand of UFD. In human Ubc9 Leu6 interacts with the β22-β23 loop formed by Gly485 and Gly487 (Fig. 1c). In contrast, in yeast, the composition and length of this loop is different and Ubc9 Leu6 interacts with Leu478 (Fig. 1d). In this region we can observe the highest structural homology between human and yeast. In human three backbone hydrogen bonds are formed between Gly487 and Ile489 with Ubc9 Met36 and Asn37, but only two in yeast, between Leu485 and Ubc9 Met36. Additionally, the side chain of Ile489 in human (or Leu485 in yeast) is buried in both cases in the Ubc9 hydrophobic pocket formed by Met36 and Leu38 (Fig. 1c,d). Finally, it is worth mentioning that the specific contacts established by yeast Arg484 and Phe491 are absent in metazoan sequences but highly conserved in *Saccharomycetales*.

Interface comparison with other UbL E1-E2 complexes. Comparison of the UFD-E2 interfaces in ubiquitin and Nedd8 (Uba1-Ubc4 complex, PDB code 4II2; and Uba3-Ubc12, pdb code 1Y8X)[13,15] indicate that only the hydrophobic patch in Ubc9 is partially conserved in the ubiquitin E2 (Ubc4), formed in this instance by Leu3 and Leu30, which can be aligned with Ubc9 Leu6 and Leu38 (Fig. 2). However, the basic patch in the α1 helix of Ubc9, formed by Arg13, Lys14, Arg17 and Lys18, represents a specific feature of the SUMO pathway and are not present neither in Nedd8 (Ubc12) nor in ubiquitin (Ubc4), which are replaced by acidic and

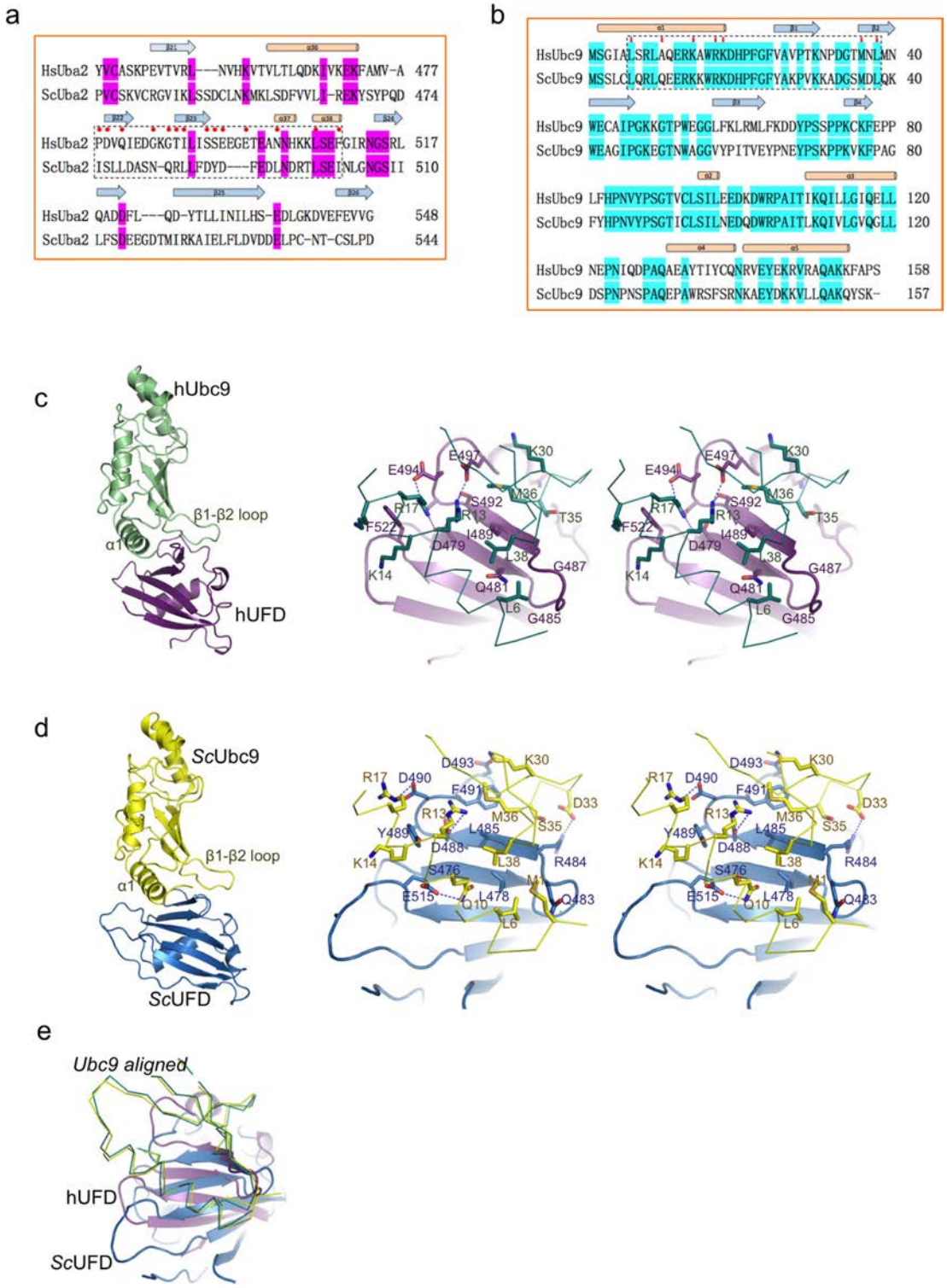


Figure 1. Structural alignment and comparison of the interfaces between UFD and Ubc9 from human and yeast complexes. (**a**) Structural alignment of the UFD domains from yeast (*ScUba2*) and human (*HsUba2*). Red circles indicate contact residues to Ubc9. Dotted rectangle represents the binding region to Ubc9. (**b**) Structural alignment of yeast and human Ubc9. Small red arrows indicate contact residues to UFD. Dotted rectangle represents the binding region to UFD. Secondary structure is depicted above sequence. (**c**) *Left*, ribbon representation of the complex of human Ubc9 and the UFD domain. *Right*, stereo representation of the interface residues between Ubc9 (green line) and UFD (purple ribbon). Major contacts are labeled and represented in stick configuration. (**d**) *Left*, ribbon representation of the complex of yeast Ubc9 and the UFD domain. *Right*, stereo representation of the interface residues between Ubc9 (yellow line) and UFD (blue ribbon). Major contacts are labeled and represented in stick configuration. (**e**) Structural superposition of Ubc9 in the human and yeast complex with UFD.

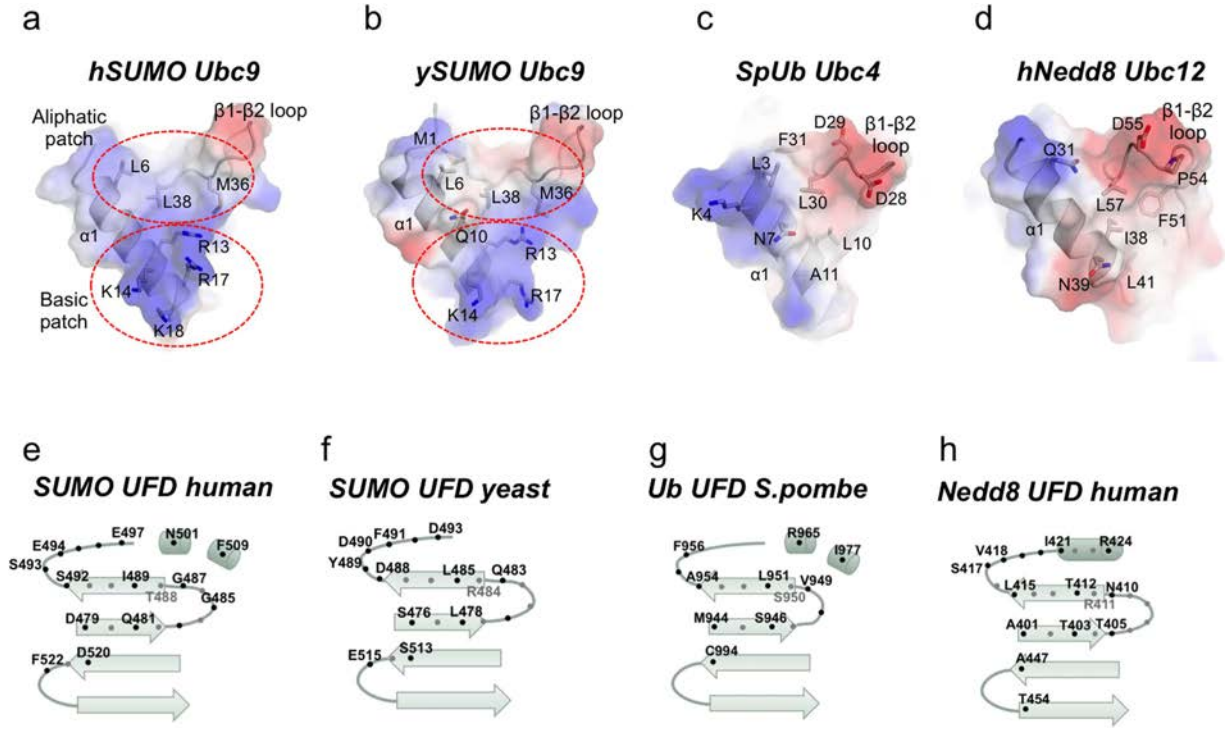


Figure 2. Comparison of the UFD-E2 interface from different UbL systems. (**a**) Transparent electrostatic representation of the interface of human SUMO E2 (*hSUMO Ubc9*) with the UFD domain. Major contacts are labeled and represented in stick configuration. Basic and aliphatic surface patches are indicated by dotted circles. (**b**) Transparent electrostatic representation of the interface of yeast SUMO E2 (*ySUMO Ubc9*) with the UFD domain. Major contacts are labeled and represented in stick configuration. Basic and aliphatic surface patches are indicated by dotted circles. (**c**) Transparent electrostatic representation of the interface of *S.pombe* ubiquitin E2 (*SpUb Ubc4*) with the UFD domain. Major contacts are labeled and represented in stick configuration. (**d**) Transparent electrostatic representation of the interface of human Nedd8 E2 (*hNedd8 Ubc12*) with the UFD domain. Major contacts are labeled and represented in stick configuration. (**e**) Schematic representation of the human SUMO E1 UFD domain contacts with the E2 enzyme. (**f**) Schematic representation of the yeast SUMO E1 UFD domain contacts with the E2 enzyme. (**g**) Schematic representation of the *S.pombe* ubiquitin E1 UFD domain contacts with the E2 enzyme. (**h**) Schematic representation of the human Nedd8 E1 UFD domain contacts with the E2 enzyme. Black and grey spots indicate the orientation of the side chain in the structure regarding the β-sheet plane.

aliphatic residues. Ubc4 and Ubc12 also display a shorter β1-β2 loop compared to Ubc9 (Fig. 2). These differences in the E2-conjugating enzyme between UbLs modifiers result in the presence of non-complementary surfaces with E1 UFD domains, which are indeed the basis for the enzyme specificity among each UbL pathway. For instance, SUMO E1 UFD domain contains specific polar contacts (Ser492, Ser493, Glu494 and Glu497 in human or Asp488, Asp490 and Asp493 in yeast) to interact with the basic patch of Ubc9, however, in ubiquitin and Nedd8 these positions are substituted by aliphatic residues (Ala954 and Phe956 in ubiquitin or Leu415, Val418 and Ile421 in Nedd8) (Fig. 2).

As mentioned before, human Ile489 and yeast Leu485 adopt a similar conformation and engage identical backbone hydrogen bonds with the β1-β2 loop of Ubc9. Interestingly, despite the lack of sequence conservation, the equivalent residues in the ubiquitin UFD domain, namely Ser950-Leu951, also engage analogous backbone hydrogen bonds contacts with the β1-β2 loop of the E2 enzyme (Ubc4)[15]. Thus this backbone interaction represents a unique conserved structural element maintained in distant UbL systems such as SUMO and ubiquitin. This interaction does not occur in Nedd8 (Ubc12-Uba3, pdb code 1Y8X ref. 13), in which the E2 enzyme sits across a similar region of the UFD domain but with a different angle compared to ubiquitin, and the contacts between both UbL systems are poorly conserved and thus not complementary[18,15].

Evolutionary conservation of the SUMO E1 UFD interfaces. Structural comparison between human and yeast indicate that SUMO UFD domains are composed by a different set of residues that can engage productive interactions with a conserved E2 enzyme. In order to evaluate the biological relevance of these two alternative structural interfaces involved in E2 interactions, we have analyzed the conservation of the LHEB2 sequence of the E1 UFD domains across species. In human, the LHEB2 region is comprised by residues between Pro478 and Phe509 (Figs 1a and 3e), whilst in yeast is composed by residues between Pro472 and Ile502 (Figs 1b and 3e). We searched for human E1 UFD domain orthologs at the EggNOG database and focused on metazoan and fungal species, for which structural information of the E1 UFD-E2 interactions is available. Since fungal E1 displayed a

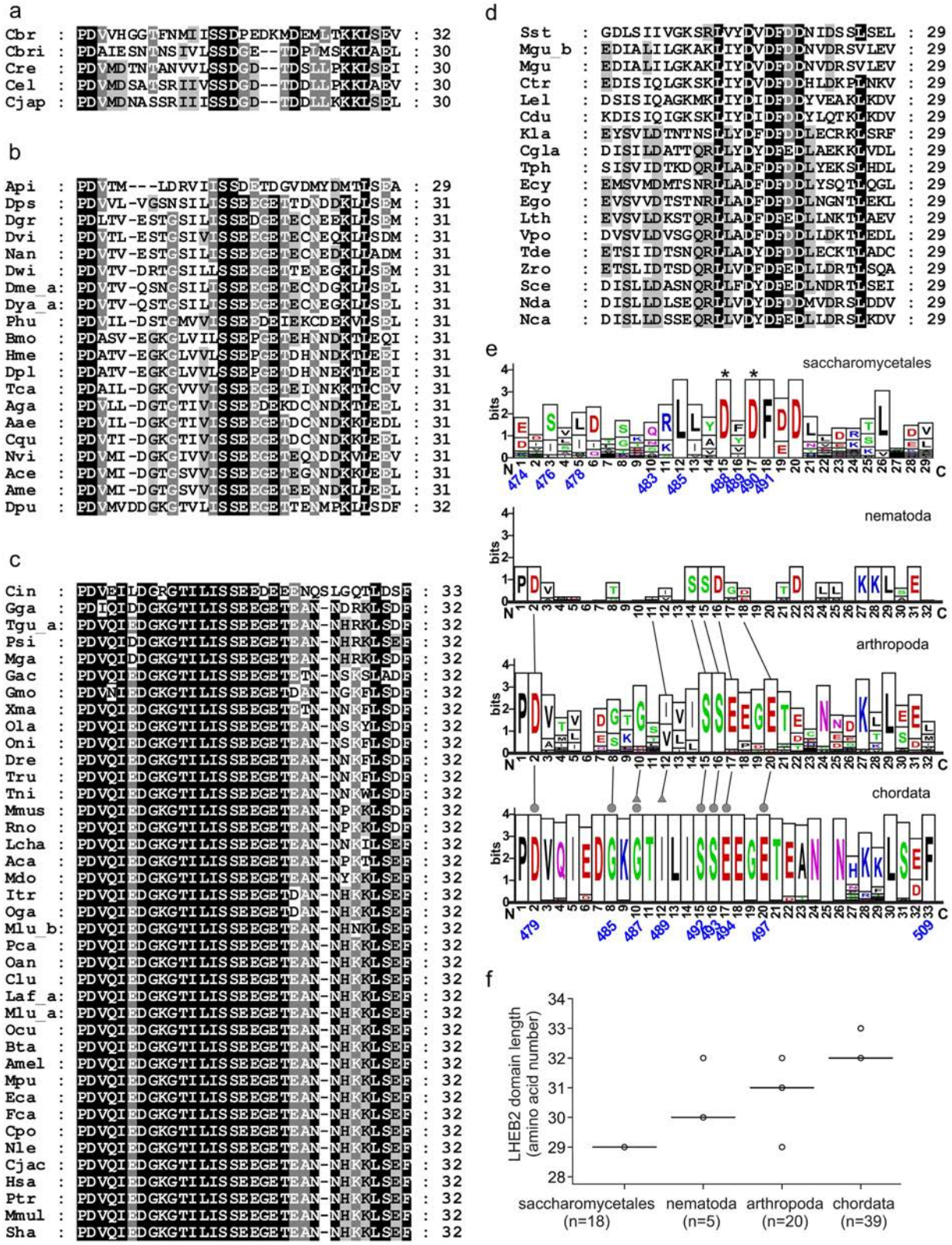


Figure 3. Conservation analysis of Sae2 LHEB2 domain in metazoan and Saccharomycetales. Amino acid sequence alignment of Sae2 LHEB2 domain orthologs from nematoda (**a**), arthopoda (**b**), chordate (**c**) and saccharomycetales (**d**). Residue shading correspond to 95% (white letter and dark background), 75% (white letter and gray background), and 55% (black letter and light gray background) of sequence identity in (**a**,**b** and **c**). In the case of saccharomycetales, the shading types correspond to 90%, 70% and 50% of sequence identity, respectively. Metazoan multiple sequence alignments of LHEB2 sequences were performed using Clustal Omega software and including yeast Uba2 (NP_010678) as outlier. Saccharomycetales multiple sequence alignments of LHEB2 sequences were performed using Muscle software and including human Sae2

(Q9UBT2) as outlier. (**e**) Graphical representation of LHEB2 domain consensus sequences determined from amino acid sequence alignments shown in (**a**,**b**,**c** and **d**). The overall height of the stack indicates the sequence conservation at that position, while the symbol height within the stack indicates the relative frequency of each amino acid within that position. The positions of yeast Sae2 residues involved in Ubc9 interaction according to the previously resolved structure (3ONG) are indicated in blue below the sequences graph. Asterisks indicate residues shown to have a major contribution to E1-E2 interactions in mutagenesis analysis[25]. The positions of human Sae2 residues involved in Ubc9 interactions according to the resolved structure are indicated in blue below the chordata consensus sequence graph. Grey circles indicate residues establishing contacts with Ubc9 α1-helix, while grey triangles indicate residues interacting with Ubc9 residues located at the Ubc9 β1β2-loop. Conserved residues across phyla are indicated by lines. (**f**) Distribution of LHEB2 sequence length displayed by orthologs within each phylogenetic group analyzed was plotted on a box plot graph. Data points are represented by circles. Outliers are represented by dots. The number of data points analyzed in each phylogenetic group is indicated below the x-axis.

high level of sequence divergence, we focused on species belonging to the *Saccharomycetales* order, which include *Saccharomyces cerevisae*. The phylogenetic analyses of both domains, E1 UFD and LHEB2, show that clustering of the E1-UFD sequences in the phylogenetic tree reflects the taxonomical relationships of the species represented (Fig. 4a). On the contrary, when the same analysis was performed with the LHEB2 domain, tree distribution is not consistent with taxonomic lineages (Fig. 4b), suggesting that this region presents higher variability than the UFD domain where is contained.

In addition, we analyzed the distribution of homology between pairs. *Saccharomycetales* UFD or LHEB2 sequences were compared with the corresponding human (conservation to outlier) or yeast sequences (conservation within the group). Similarly, metazoan UFD or LHEB2 sequences were compared with the corresponding yeast (conservation to outlier) or human sequences (conservation within the group). The homology pair distributions were plotted onto box plots (Supplemental Fig. S2). In general, when sequences were compared with an outlier, in all groups except in *Saccharomycetales*, the median of the obtained distribution was higher in the LHEB2 sequence analysis than in the UFD. At the same time, the box length, whose limits indicate the 25th and 75th percentiles, is higher for the LHEB2 than for the UFD homology pair distributions, suggesting that this region also presents higher variability. These results support the phylogenetic analyses indicating that, in general, the LHEB2 sequence is more conserved than the UFD domain within each evolutionary group, although the individual sequences contained in each group display higher variability. Similar results were obtained when sequences were compared with a reference sequence within each group, *Saccharomycetales* or metazoan.

The homology analysis of the residues involved in Ubc9 contacts are even higher conserved and minor differences were identified among phyla. In chordates, the highest conserved group, all the contacts described in the human structure are present in all species analyzed within this phylum (Fig. 3c). In arthropoda, sequence comparison also display little divergence, but in this case variations can be found in the composition of the β22-β23 connecting loop of UFD, presenting different loop lengths but still conserving Gly485 and Gly487 in many species of the phylum (Fig. 3b). In nematode, the five species analyzed also display conservation in the major E2 contact residues, but in contrast to chordate and arthropoda, the β12-β13 connecting loop displays little homology and Gly485 and Gly487, present in chordata and arthropoda, have been substituted (Fig. 3a).

Therefore, all major specific UFD contacts with Ubc9 are basically conserved in *metazoa*, including contacts with the hydrophobic and the basic patches of the Ubc9 surface. For instance, human Asp479, Ser492, Ser493, Glu494 and Glu497 (occasionally replaced by aspartic), which interact with Ubc9 Arg13 and Arg17, are highly conserved in all species analyzed. Similarly, the hydrophobic interaction of the human UFD Ile489, which interacts with Ubc9 Met36 and Leu38, is also conserved but can be occasionally substituted by valine in some species. The major differences in the UFD interface in *metazoa* are located in the β22-β23 connecting loop, formed in human by Gly485 and Gly487, which interact with Ubc9 Leu6. Whereas in chordate this connecting loop is highly conserved, in arthropod and particularly in nematode, this loop displays different lengths and amino acid composition. It is worth mentioning here that the extension and composition of this particular loop in UFD domains is highly divergent between E1-activating enzymes specific to different UbL modifiers, suggesting that this domain has evolved to interact with its cognate E2-enzyme.

In contrast to *metazoa*, sequence comparison of the E2 binding region of the yeast UFD domain with members of the *Fungi* kingdom is highly diverse, and further structural analyses will be required for establishing the molecular basis of E1-E2 interactions in those divergent groups. In the order of *Saccharomycetales*, the major specific contacts described in the structure of *S. cerevisiae* UFD-Ubc9 complex[25] are conserved. In this instance, as shown before in the structural comparison, the consensus binding sequence in *Saccharomycetales* is completely different to *metazoan*. Yeast UFD residues Asp488, Asp490 and Asp493, which interact with the basic patch of Ubc9, are highly conserved in *Saccharomycetales* (Fig. 3d,e), as well as Leu478, Leu485 and Phe491, which interact with the hydrophobic patch of Ubc9. Other specific contacts in the structure, such as Arg484 and Tyr489, are replaced by residues with similar chemical properties, such as lysine for Arg484 and, phenylalanine, isoleucine or valine for Tyr489 (Fig. 3d,e).

In the essential SUMO conjugation pathway, protein-protein interactions have evolved to maintain specificity of the modifier and the targets. The SUMO conjugating enzyme constitutes the link between modifier specificity, which is selected by the E1 activating enzyme, and the protein substrate specificity, mediated by the cooperation between the E2 and the E3 ligase enzymes. In the E1-E2 interactions, a region in the E1 UFD domain, the LHEB2 sequence, plays a major role in E2 recruitment to the E1. In evolutionary distant groups, the LHEB2 sequences are poorly conserved according to sequence homology and length. On the contrary, in closely related phylogenetic

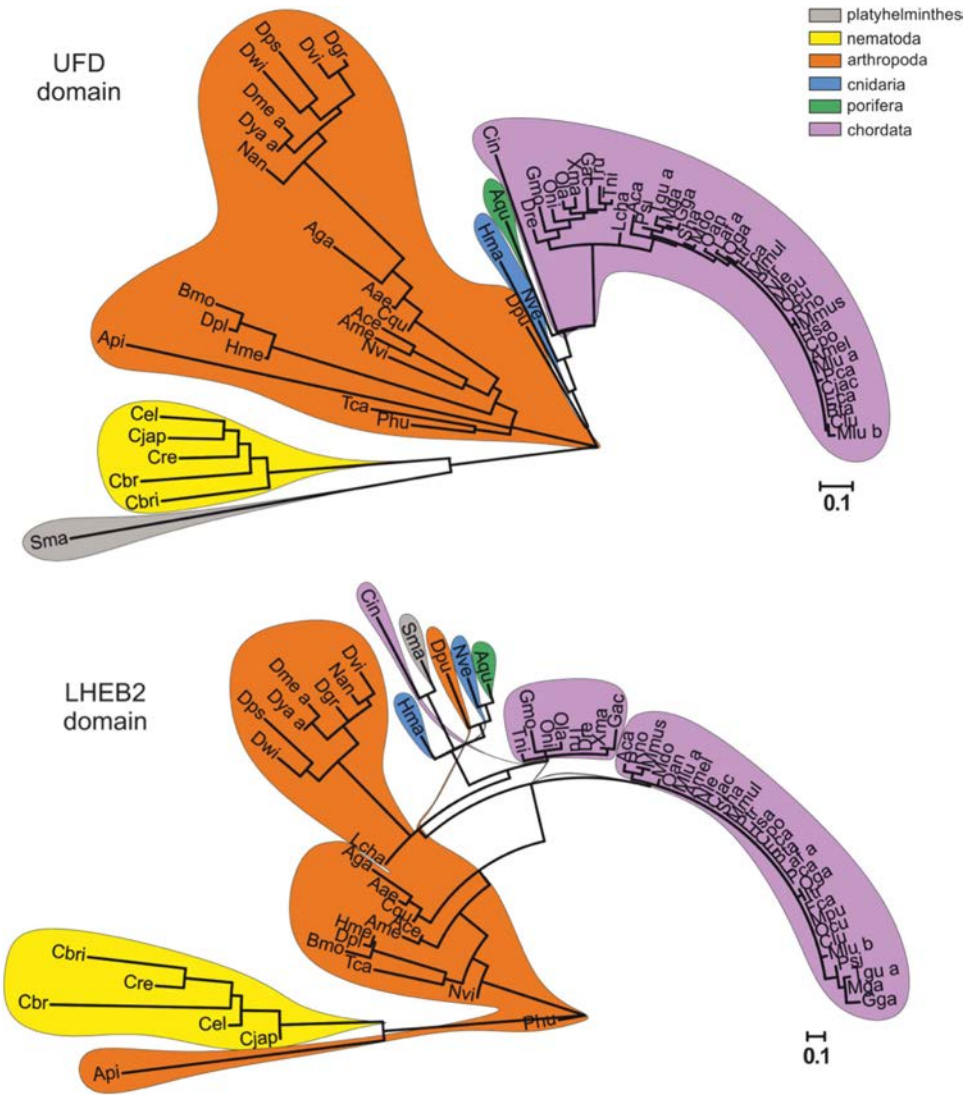


Figure 4. Phylogenetic analysis of UFD and LHEB2 domains from Metazoa and Saccharomycetales. Maximum likelihood phylogenetic trees depicting the evolutionary relationships among 86 Sae2 UFD (**a**) or Sae2 LHEB2 (**b**) domain sequences from 68 metazoa species species using sequence alignments shown in Fig. S3 (UFD) and Fig. 3 (LHEB2). Sequences belonging to the same phylum are enclosed in colored areas. Tree scales are shown below each tree.

groups, the conservation of the LHEB2 sequence is higher, highlighting the relevance of this interaction between E1 and E2 structures. We speculate that these distinctive strategies for E1-E2 interactions through the E1 UFD domain are the consequence of a high selective pressure to ensure modifier specificity. Future structural analyses of other evolutionary distant groups, such as plants or protozoa, will most probably uncover novel molecular determinants mediating E1-E2 interactions.

Conclusions

In summary, structural comparison of yeast and human SUMO UFD-E2 complexes and sequence alignment of SUMO UFD domains, reveal the presence of at least two complementary types of interfaces, which are conserved across species from the same kingdom (*metazoan*) or in the case of the *Fungi* kingdom, in the same order (*Saccharomycetales*). Despite the low level of sequence homology in the UFD domains among these distant species of different kingdoms, these two types of interfaces maintain the structural and chemical properties necessary to interact with a conserved E2 binding surface. Structural and sequential comparisons have also revealed at least two different types of consensus sequences in the E1 UFD domain, which we named LHEB2 sequences, that can complement the conserved surface in the E2 enzyme. Interestingly, sequence conservation in the E2 binding region is higher than in the overall UFD domain, suggesting the presence of an evolutionary pressure to maintain the contacts with the E2-conjugating enzyme, which are essential for the correct function of each UbL pathway.

Materials and Methods

Protein expression and purification. Expression constructs were generated by a standard PCR-based cloning method. The full length human Ubc9 and E1 UFD domain (residues 447–547) were cloned to pET28a tagged with 6x His at the N-terminal. *Escherichia coli* BL21(DE3) plysS containing the expression vector were grown in Luria Bertani medium with chloramphenicol (17 μg/mL) and kanamycine (50 μg/mL) at 37 °C until the OD600 reached to 0.8. Expression was induced by 0.1 mM IPTG, followed by overnight culturing at 28 °C. Recombinant proteins were purified by nickel-nitrilotriacetic acid agarose resin (Qiagen) and dialyzed against 250 mM NaCl, 20 mM Tris-HCl (pH 8.0), 1 mM β-mercaptoethanol in the presence of thrombin protease overnight at 4 °C to remove the 6x His tag. Proteins were further purified by gel filtration chromatography on a Superdex75 column (GE Healthcare), which was pre-equilibrated in 250 mM NaCl, 20 mM Tris-HCl pH 7.5, 1 mM β-mercaptoethanol.

Protein complex preparation and methylation. Ubc9 and UFD complex was made by mixing equimolar amounts of proteins and purified by gel filtration chromatography using a Superdex75 column. Ubc9 and UFD were co-eluted in a single peak and confirmed by SDS-PAGE. After gel filtration, the complex was dialyzed against 250 mM NaCl, 50 mM HEPES pH 7.5, 1 mM β-mercaptoethanol for lysine methylation based on a published strategy[28]. In brief, borane-dimethylamine complex (Sigma-Aldrich) and formaldehyde (Sigma-Aldrich) were sequentially added into protein solution and incubated overnight at 4 °C. The methylation reaction was stopped by a final gel filtration chromatography on a Superdex75 column pre-equilibrated in 200 mM NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM β-mercaptoethanol. Purified protein complex was concentrated to 30 g/L using an Amicon Ultra-10 K ultrafiltration device (Millipore) prior to crystallization.

Crystallization and data collection. Crystals were grown by the sitting-drop vapor diffusion method by mixing the protein complex (30 g/L) with an equal volume of reservoir solution containing 16% PEG6000 (w/v), 100 mM MES pH 6.5, and 5% MPD (v/v), at 18 °C. Crystals appeared after 24 hours and continued to grow to full size in one week. Big crystals were soaked in mother liquor supplemented with gradually increasing concentration of 5%, 10%, 20% (v/v) MPD for 60 seconds each time and flash frozen in liquid nitrogen. Diffraction data were collected to 2.20 Å resolution at ALBA synchrotron in Barcelona (BL13-XALOC beamline). The crystals belong to the space group $P4_32_12$ and the unit cell has a dimension of a = 129.61 Å, b = 129.61 Å, and c = 66.60 Å. Data were processed with XDS[29] and scaled, reduced, and further analyzed using CCP4[30]. More details are shown in Table 1.

Structure determination and refinement. The structure was determined by molecular replacement method using the full length human Ubc9 (protein data bank code 1U9B) as a search model for one molecule in the asymmetric unit in PHASER[31]. Initial electron density was manually improved to build up the final model using Coot[32], and the refinement was performed using Phenix[33]. Refinement statistics are shown in Table1. The structure has been deposited in the PDB data bank with the code 5FQ2.

Phylogenetic sequence comparison. Human Sae2 orthologs were search in EggNOG database (http://eggnogdb.embl.de/#/app/home) and sequences from *Metazoa* and *Saccharomycetales* were selected for homology analysis. Detailed information about sequence homology analysis methods is indicated in figure legends. Briefly, when sequences were highly divergent, multiple sequence alignments were performed using Muscle tool (http://www.ebi.ac.uk/Tools/msa/muscle/). When sequence displayed higher conservation level, homology analyses were performed using Clustal Omega program (http://www.clustal.org/omega/). Phylogenetic distances were calculated by the maximum likelihood method and the JTT model included in the Seaview v4 software package[34], and unrooted trees exported. Phylogenetic trees were drawn using the online iTOL software (http://itol.embl.de/). Consensus sequences were calculated using WebLogo software (http://weblogo.berkeley.edu/)[35]. Multiple sequence alignments were edited, analyzed and shaded using GeneDoc software[36] (http://iubio.bio.indiana.edu/soft/molbio/ibmpc/genedoc-readme.html). Data distribution was plotted on box plots using "BoxPlotR: a web-tool for generation of box plots".

References

1. Johnson, E. S. Protein modification by SUMO. *Annual review of biochemistry* **73,** 355–382 (2004).
2. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458,** 422–429, doi: 10.1038/nature07958 (2009).
3. Schulman, B. A. & Harper, J. W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat. Rev. Mol. Cell. Biol.* **10,** 319–331 (2009).
4. Streich, F. C. Jr. & Lima, C. D. Structural and functional insights to ubiquitin-like protein conjugation. *Annual review of biophysics* **43,** 357–379, doi: 10.1146/annurev-biophys-051013-022958 (2014).
5. Gareau, J. R. & Lima, C. D. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **11,** 861–871 (2010).
6. Geiss-Friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* **8,** 947–956 (2007).
7. Hay, R. T. SUMO: a history of modification. *Molecular cell* **18,** 1–12 (2005).
8. Walden, H. *et al.* The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. *Molecular cell* **12,** 1427–1437 (2003).
9. Castaño-Miquel, L., Seguí, J. & Lois, L. M. Distinctive properties of Arabidopsis SUMO paralogues support the *in vivo* predominant role of AtSUMO1/2 isoforms. *Biochemical Journal* **436,** 581–590 (2011).
10. Desterro, J. M., Rodriguez, M. S., Kemp, G. D. & Hay, R. T. Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *Journal of Biological Chemistry* **274,** 10618–10624 (1999).
11. Olsen, S. K., Capili, A. D., Lu, X., Tan, D. S. & Lima, C. D. Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature* **463,** 906–912 (2010).
12. Walden, H. *et al.* The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1. *Mol Cell* **12,** 1427–1437 (2003).

13. Huang, D. T. *et al.* Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. *Mol Cell* **17,** 341–350 (2005).
14. Lee, I. & Schindelin, H. Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* **134,** 268–278 (2008).
15. Olsen, S. K. & Lima, C. D. Structure of a Ubiquitin E1-E2 Complex: Insights to E1-E2 Thioester Transfer. *Molecular Cell* **49,** 884–896 (2013).
16. Reiter, K. *et al.* Identification of Biochemically Distinct Properties of the Small Ubiquitin-related Modifier (SUMO) Conjugation Pathway in Plasmodium falciparum. *Journal of Biological Chemistry* **288,** 27724–27736, doi: 10.1074/jbc.M113.498410 (2013).
17. Wang, J. *et al.* The intrinsic affinity between E2 and the Cys domain of E1 in ubiquitin-like modifications. *Mol Cell* **27,** 228–237 (2007).
18. Huang, D. T. *et al.* Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. *Nature* **445,** 394–398 (2007).
19. Durfee, L. A., Kelley, M. L. & Huibregtse, J. M. The basis for selective E1-E2 interactions in the ISG15 conjugation system. *Journal of Biological Chemistry* **283,** 23895–23902 (2008).
20. Tokgöz, Z. *et al.* E1-E2 interactions in ubiquitin and Nedd8 ligation pathways. *Journal of Biological Chemistry* **287,** 311–321 (2012).
21. Wang, J. *et al.* Crystal Structure of UBA2 ufd-Ubc9: Insights into E1-E2 Interactions in Sumo Pathways. *PloS one* **5,** e15805 (2010).
22. Reiter, K. H. *et al.* Characterization and Structural Insights into Selective E1-E2 Interactions in the Human and Plasmodium falciparum SUMO Conjugation Systems. *Journal of Biological Chemistry*, doi: 10.1074/jbc.M115.680801 (2015).
23. Walden, H., Podgorski, M. S. & Schulman, B. A. Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. *Nature* **422,** 330–334 (2003).
24. Lois, L. M. & Lima, C. D. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *EMBO J.* **24,** 439–451 (2005).
25. Wang, J. *et al.* Crystal Structure of UBA2ufd-Ubc9: Insights into E1-E2 Interactions in SUMO Pathways. *PloS one* **5,** e15805 (2010).
26. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372,** 774–797 (2007).
27. Bencsath, K. P., Podgorski, M. S., Pagala, V. R., Slaughter, C. A. & Schulman, B. A. Identification of a multifunctional binding site on Ubc9p required for Smt3p conjugation. *J Biol Chem* **277,** 47938–47945 (2002).
28. Walter, T. S. *et al.* Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14,** 1617–1622 (2006).
29. Kabsch, W. X. D. S. *Acta Crystallographica Section D* **66,** 125–132, doi: 10.1107/S0907444909047337 (2010).
30. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67,** 235–242 (2011).
31. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallographica Section D: Biological Crystallography* **60,** 432–438 (2004).
32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60,** 2126–2132 (2004).
33. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66,** 213–221 (2010).
34. Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* **27,** 221–224, doi: 10.1093/molbev/msp259 (2010).
35. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Research* **14,** 1188–1190 (2004).
36. Nicholas, K. B. & Nicholas, H. B. J. GeneDoc: a tool for editing and annotating multiple sequence alignments. (1997).

Acknowledgements

We thank Martí Bernardo for fruitful discussions. This work was supported by grants from the Ministerio de Economia y Competitividad BFU2015-66417-P (MINECO/FEDER) to DR, and from the Generalitat de Catalunya (2009SGR 09626) and from the European Commission (ERC-2014-PoC-671839) to LML. B.L. acknowledges his scholarship to the China Scholarship Council program from the Chinese government. X-ray experiments were performed at the BL-13 beamline at the ALBA synchrotron in collaboration with the ALBA staff.

Author Contributions

B.L. and D.R. conducted all crystallographic experiments presented in Figures 1 and 2. L.M.L. performed and analyzed the sequence alignments presented in Figures 3 and 4. D.R. and L.M.L. conceived the idea for the project, analyzed the results and wrote the paper.

Additional Information

Accession codes: The crystal structure of the complex was deposited in the PDB data bank with the code 5FQ2.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Liu, B. *et al.* Structural analysis and evolution of specificity of the SUMO UFD E1-E2 interactions. *Sci. Rep.* **7**, 41998; doi: 10.1038/srep41998 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.