Investigation into the Specificities of Deubiquitinases and Ubiquitin-like Proteases

A Thesis

Submitted to the Department of Chemistry and Chemical Biology

TU Dortmund University

In Partial Fulfilment of the Requirements

for the Academic Degree of

Doctor of Natural Sciences

(Dr. rer. nat.)

Zhou Zhao

Jiangsu, China

Dortmund

2024

The work presented in this thesis was executed from September 2019 to July 2024 under the supervision of Dr. Malte Gersch at the Chemical Genomics Centre of the Max Planck Society and the Department of Chemical Biology of the Max Planck Institute of Molecular Physiology and at the Faculty of Chemistry and Chemical Biology of the TU Dortmund University.

The unique and persistent identifier for the author is ORCID: 0000-0001-9350-7238.

1st Referee: Dr. Malte Gersch

2nd Referee: Prof. Dr. Dr. h.c. Herbert Waldmann

Declarations

This thesis has contributed to the following publications:

Zhou Zhao, Rachel O'Dea, Kim Wendrich, Nafizul Kazi, & Malte Gersch. Native Semisynthesis of Isopeptide-linked Substrates for Specificity Analysis of Deubiquitinases and Ubl Proteases. Journal of the American Chemical Society 2023, 145, 20801-20812

Rachel O'Dea, Nafizul Kazi, Alicia Hoffmann-Benito, **Zhou Zhao**, Sarah Recknagel, Kim Wendrich, Petra Janning & Malte Gersch. Molecular Basis for Ubiquitin/Fubi Cross-reactivity in USP16 and USP36. Nature Chemical Biology 2023, 19, 1394–1405

Zhou Zhao, Vanessa Boll, Thomas Hermann, Kim Wendrich, Kay Hofmann & Malte Gersch. Polyubiquitin Probes with Internal Alkyl Bromide Warheads for the Investigation of Ubiquitin Linkage Specificity. *Manuscript in preparation*

Dr. Rachel O'Dea contributed to this thesis through preparation of FUBI-KG-TAMRA.

Kim Wendrich and Nafizul Kazi contributed to this thesis through purification of proteins.

Dr. Sonja Sievers and Dr. Philipp Lampe contributed to thesis by performing high-throughput screening.

Gian Marvin Kipka helped the language editing in the section of Kurzzusammenfassung which was originally translated by DeepL Translate.

Table of Contents

ABSTRACT

Ubiquitin as a macromolecule-based modification has been validated as the regulator of protein stability. Ubiquitin signals can be significantly diversified by forming polymerized chains as well as conjugating with some structurally similar modifiers (Ubiquitin-like proteins, Ubls). Therefore, cellular events can be modulated in a more complicated way through these modifications, which are reversible by specialized proteases called deubiquitinases (DUBs) and Ubiquitin-like proteases (ULPs). Some of these enzymes show specificity toward different modifiers, while some prefer to cleave linkage-specific polyubiquitin chains. To study the modifier specificity and linkage specificity, high quality chemical biology tools are the cornerstones.

Isopeptide-linked fluorescence polarization substrates are powerful tools for analyzing modifier specificity of DUBs/ULPs quantitatively. Since Ub/Ubl-based substrates can be sensitive to harsh reaction and purification conditions, a native semisynthetic method was developed based on recombinantly expressed proteins which were further functionalized with fluorophores and purified in aqueous buffer to ensure the homogeneity of substrates. Six substrates were prepared to assemble an assay panel for characterizing several DUBs/ULPs. USP16 and USP36 were unprecedently identified as the first proteases with triple modifier specificity.

Studies on linkage specificity were enabled by the development of novel diUb probes with internal warheads, which are based on alkyl bromide. Alkyl-bromide-based probes were successfully used for studying SnVTD, which is a Lys6-specific DUB. The complex structure was solved to reveal a unique recognition and activation mechanism. The probes were also applied for capturing two linkage-specific E3s.

Collectively, a high-quality fluorescence polarization assay platform was established and used for exploring modifier specificity of DUBs/ULPs. Novel diUb probes were developed to capture DUBs with linkage specificity. These tools will facilitate the understanding of DUBs/ULPs.

KURZZUSAMMENFASSUNG

Ubiquitin als eine auf Makromolekülen basierende Modifikation wurde als Regulator der Proteinstabilität validiert. Ubiquitin-Signale können durch die Bildung polymerisierter Ketten und die Konjugation mit einigen strukturell ähnlichen Modifikationen (Ubiquitin-ähnliche Proteine, Ubls) erheblich diversifiziert werden. Daher können zelluläre Ereignisse durch diese Modifikationen, die durch spezialisierte Proteasen namens Deubiquitinasen (DUBs) und Ubiquitin-ähnliche Proteasen (ULPs) reversibel sind, auf kompliziertere Weise moduliert werden. Einige dieser Enzyme weisen eine Spezifität für verschiedene Modifikationen auf, während andere es vorziehen, bindungsspezifische Polyubiquitinketten zu spalten. Um die Spezifität der Modifikationen und die Bindungsspezifität zu untersuchen, sind hochwertige chemisch-biologische Werkzeuge die Eckpfeiler.

Isopeptid-verknüpfte Fluoreszenzpolarisationssubstrate sind leistungsstarke Werkzeuge zur quantitativen Analyse der Modifikationsspezifität von DUBs/ULPs. Da Ub/Ubl-basierte Substrate empfindlich auf harsche Reaktions- und Reinigungsbedingungen reagieren können, wurde eine native halbsynthetische Methode entwickelt, die auf rekombinant exprimierten Proteinen basiert, die zusätzlich mit Fluorophoren funktionalisiert und in wässrigem Puffer gereinigt wurden, um die Homogenität der Substrate sicherzustellen. Es wurden sechs Substrate hergestellt, um ein Assay-Panel zur Charakterisierung verschiedener DUBs/ULPs zusammenzustellen. USP16 und USP36 wurden als die ersten Proteasen mit dreifacher Modifikationsspezifität identifiziert.

Studien zur Bindungsspezifität wurden durch die Entwicklung neuartiger diUb-Sonden mit internen reaktiven Gruppen auf Alkylbromidbasis ermöglicht. Auf Alkylbromid basierende Sonden wurden erfolgreich zur Untersuchung von SnVTD, einer Lys6-spezifischen DUB, eingesetzt. Die Struktur des Komplexes wurde gelöst, um einen einzigartigen Erkennungs- und Aktivierungsmechanismus aufzudecken. Die Sonden wurden auch für die Erfassung von zwei bindungsspezifischen E3s verwendet.

Insgesamt wurde eine hochwertige Fluoreszenzpolarisations-Assay-Plattform eingerichtet und zur Erforschung der Modifikationsspezifität von DUBs/ULPs eingesetzt. Es wurden neuartige diUb-Sonden entwickelt, um DUBs mit Bindungsspezifität zu erfassen. Diese Instrumente werden das Verständnis von DUBs/ULPs erleichtern.

1. INTRODUCTION

1.1 The Ubiquitin System

Post-translational modifications are covalent attachment of moieties which are not only small functional groups but also macromolecules to the side chains of amino acids leading to the expanded chemical and structural diversity of proteins to precisely regulate relevant cellular events[1]. Among these modifications, ubiquitination is the most dominant regulatory factor of protein degradation and actively involved in almost all aspects of cellular processes through precise regulation in the ubiquitin system[2].

Ubiquitin, a small protein consisting of 76 amino acids, was initially discovered and purified as a polypeptide in bovine immunocytes in 1975[3]. Given that it is expressed ubiquitously in living cells, the polypeptide was named ubiquitous immunopoietic polypeptide (UBIP). A few years later, a heat-stable polypeptide was discovered to be a degradation signal in an adenosine triphosphate (ATP)-dependent manner. Therefore, this polypeptide was named ATP-dependent proteolysis factor 1 (APF-1)[4, 5]. Further biochemical characterization proved that UBIP and APF-1 are the same protein, namely ubiquitin[6]. Human ubiquitin is a highly conserved protein encoded by two polyubiquitin genes (*UBB* and *UBC*) and two fused genes (*UBA52* and *RPS27A*) which contain a single copy of ubiquitin in fusion with ribosomal proteins.

The ubiquitin system plays a decisive role in selective protein degradation. In this system, proteins which are targeted for degradation are labeled by ubiquitin covalently through a sophisticated enzymatic cascade. Briefly, the C-terminal glycine of ubiquitin is activated by the ubiquitin activating enzyme, E1, in an ATP-dependent way to form an active thioester intermediate followed by the transfer to the active cysteine residue of a ubiquitin conjugating enzyme E2. Subsequently, a ubiquitin ligase E3 is involved in the covalent attachment of ubiquitin. Once the first ubiquitin is placed on the protein, polyubiquitin chain can be formed and elongated leading to the degradation signals. The ubiquitinated protein is recognized by the proteasome for degradation while the ubiquitin chain is trimmed off before entering into the proteasome for recycling[7]. There are only two E1 activating enzymes and several E2s in the enzymatic cascade while E3s which are consisted of around 600 members control the final step of modification[8]. The ubiquitination is highly dynamic since it can be reversibly removed from the target protein by a group of dedicated proteases called deubiquitinases (DUBs). In the human genome, around 100 different DUBs have been reported with versatile functions[9]. Collectively,

the E1-E2-E3 cascade, proteasome and DUBs assemble an integrated system to modulate ubiquitination.

1.2 The Structural Features of Ubiquitin

Ubiquitin as a highly stable protein adopts a β-grasp fold consisting of three and a half α-helices and a compact β sheet containing five strands and starting from its initial methionine residue (Fig 1.1.A)[10]. The rest of the ubiquitin structure includes a short helix and seven disordered reverse turns. In contrary to its highly compressed body, ubiquitin has a flexible C-terminal tail, of which the last glycine residue was employed for conjugation to numerous substrates. Classic ubiquitin modification is enabled by an isopeptide bond between the Gly76 of ubiquitin and the lysine side chains on the protein of interest (Fig 1.1.B). Since lysine residues are usually abundant in protein, substrate protein can be modified with multi-monoubiquitin and the ubiquitin itself can also be further modified with other moieties, such as acetylation and phosphorylation[11, 12].

Ubiquitin has in total eight free amine functional groups from the first N-terminal methionine and the side chain of seven lysine residues (Fig 1.1.A). The eight free amines are the chemical basis for the elongation of ubiquitin chains distributed in the different subclasses. Notably, three of the lysine residues (Lys27, Lys29 and Lys33) are anchored in the α-helix while the N-terminal methionine and two lysine residues (Lys6 and Lys11) are buried in the two anti-parallel β strands. The other two lysine residues (Lys48 and Lys63) are located in disordered loop regions with more flexibility. During the ubiquitin chain elongation, eight possible homotypic polyubiquitin chains could be synthesized (Fig 1.1.C). The ubiquitin linkages have been extensively studied. As the most abundant linkage Lys48-linked polyubiquitin is a typical signal for proteasome degradation[13]. Other linkages are involved in many non-degradative functions, such as protein-protein interaction, DNA repair, and protein localization[14]. Beyond homotypic linkages, heterotypic chains can be formed as well to increase the complexity of polyubiquitin chain architecture and thereby expand the diversity of ubiquitin signals (Fig 1.1.D). Therefore, decoding the architecture of ubiquitin polymers is the key step for elaborating the functions of different ubiquitin codes. Advances in proteomics using quantitative mass spectrometry have significantly facilitated the understanding of branched ubiquitin chains in different context[15, 16]. With the Ub-clipping method, 10-20% of ubiquitin was found to be in branched form and the connection between ubiquitin modifications and mitophagy was deeply studied[17].

Ubiquitin signals rely on not only the polymerization and linkage topology, but also the intrinsic features of ubiquitin surface which are crucial for recognition and interaction. Over the past few decades, extensive efforts in structural biology have successfully advanced the interpretation of key motifs on the ubiquitin surface (Fig 1.1.E). Three hydrophobic patches dominated by residues Phe4, Ile36, and Ile44 respectively have been identified to mediate numerous interactions between ubiquitin and ubiquitin-binding domains in other proteins[1, 18, 19, 20]. For instance, mutations in the Ile44 patch are required to protect the trans-activators from destabilization of proteasomal ATPases in yeast[21]. Given that hydrophobic patches are unique in the solvated state, they can provide considerably significant contribution to the specific recognition and interaction.

Another feature of ubiquitin surface is the TEK-box which is composed of five hydrophilic residues (Thr12, Thr14, Glu34, Lys6, and Lys11) sitting at the intersection of the hydrophobic patches. TEK-box has been proven to be necessary for Lys11-linked ubiquitin chain formation which is important for regulating cell cycle precisely[22]. Following study also confirmed the TEK-box is required for Lys11-linked polyubiquitin formation by Ube2S[23].

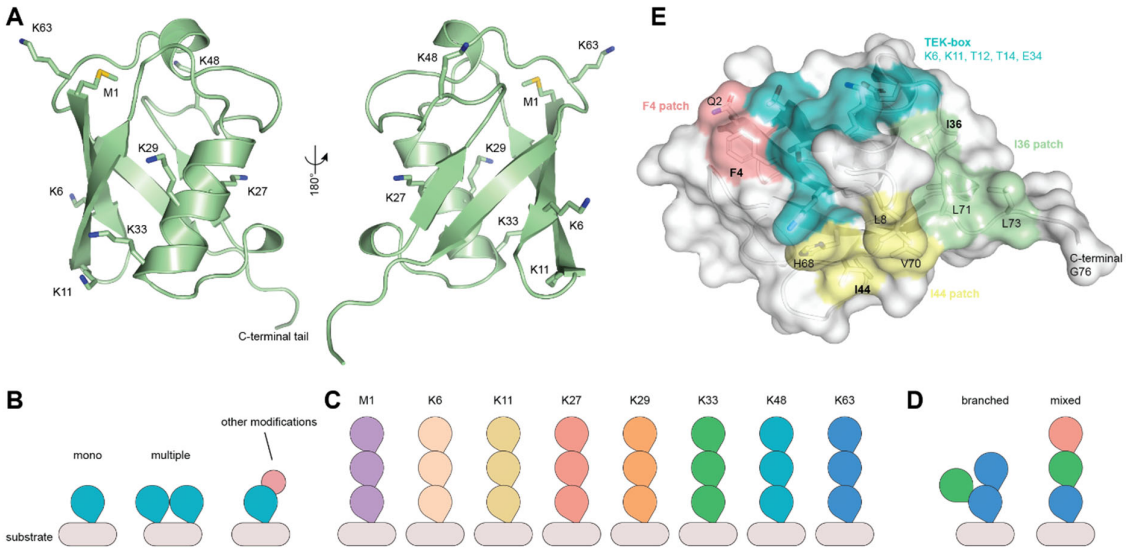


Fig 1.1. Structural overview of ubiquitin. **A**. Cartoon illustration of ubiquitin, of which residues for conjugation are highlighted as sticks (PDB ID: 1UBQ). C-terminal tail was annotated. **B**. Substrate is modified with ubiquitin at one or several sites and ubiquitin can also be further coupled to other modifications. **C**. Homotypic ubiquitin chains include eight different linkages which are highlighted in different colors. **D**. Heterotypic chains of ubiquitin. **E**. Patches and TEK-box of ubiquitin showed in surface.

The key-residue-oriented recognition motifs are the structural basis for the specific interaction. A deeper understanding of these motifs would be beneficial for further exploration of biological events driven by the interaction with ubiquitin.

1.3 Ubiquitin-like Modifiers

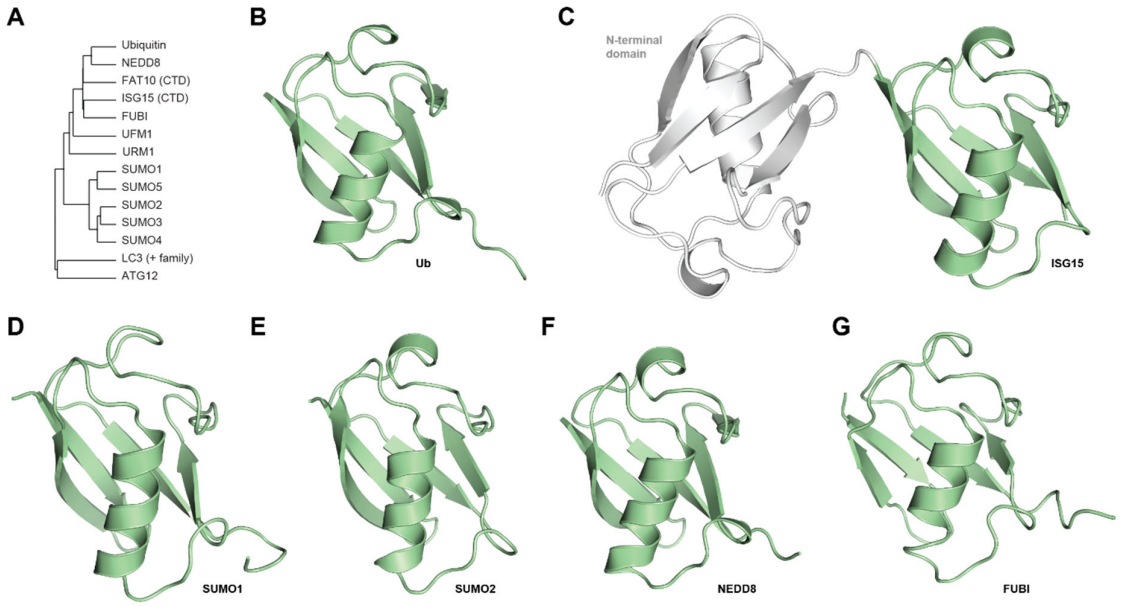


Fig 1.2. Clustering of ubiquitin-like modifiers and representative structures. **A**. Phylogenetic tree of Ub/Ubls. **B**. Structure of ubiquitin (PDB ID: 1UBQ). **C**. Structure of full-length ISG15 and its N-terminal is colored in grey (PDB ID: 6JH1). **D**. Structure of NEDD8 (PDB ID: 1NDD). **E**. Structure of SUMO1 (PDB ID: 2UYZ). **F**. Structure of SUMO2 (PDB ID: 1WM3). **G**. Structure of FUBI (PDB ID: 2L7R).

Following the discovery of ubiquitin, a number of small proteins related to ubiquitin called Ubiquitin-like proteins (Ubls) whose mature forms employ as well a C-terminal glycine for conjugation have been discovered (Fig 1.2.A)[24]. In similarity to the ubiquitin structure, Ubls generally adopt the β-grasp scaffold. Instead of the direct degradative signals, Ubls possess a wide range of functions in critically regulating enzyme activity, protein localization and interaction, immune response, and cell cycle[25]. The modifications of Ubls are usually realized by a E1-E2-E3 cascade where different enzymes are involved in a corresponding procedure to form an isopeptide bond typically with lysine residues of substrate protein[26, 27]. Some exceptions have also been reported that small molecules, such as lipids serve as substrates of Ubl modifications to regulate autophagy[28, 29]. The diversity of Ubls leading to various downstream signaling pathway is associated with physiological and pathological phenomenon. Comprehensive analysis of the modifications of Ubls will provide a deeper understanding of the biological functions of modifications and their impact in human diseases.

1.3.1 ISG15

Interferon-stimulated gene 15 (ISG15) is the first ubiquitin-like modifier that was discovered in the interferon-treated tumor cells but the functions was unknown at that time[30]. Its mRNA

was isolated from interferon-β-treated cells where ISG15 was highly expressed but it was not detectable in untreated cells[31]. Due to its cross-reaction with two ubiquitin antibodies, the structural homology to ubiquitin was observed[32]. Given that its mature form is around 15 kDa, this modifier was named as ISG15 though the premature ISG15 is a 17 kDa protein. ISG15 contains a tandem of two ubiquitin-like domains which are connected by a short and disordered linker and show different roles in protein modifications (Fig 1.2.B&C). It has been proven that the N-terminal domain is dispensable for ISG15 conjugation but it may get involved in the ligation step through interacting with E3 ligases in the context of viral infection[33].

Since ISG15 is one of the most abundant proteins upon stimulation by interferon and infections, it has been deemed to play a central role in the immune response. Over the past decades, numerous efforts have been put into understanding the role of ISG15 in regulating the host response to viral infection. Mature ISG15 has a protease-activated C-terminal LRLRGG motif to form a modification resembling ubiquitination referred to as ISGylation. ISGylation is not a direct signal for protein degradation, but evidence showed that elevated ISGylation level interferes with ubiquitination to negatively regulate protein degradation[34]. Besides ISGylation, increasing evidence showed that intracellular ISG15 prevents auto-inflammation[35]. Further efforts are highly needed to understand the role of ISG15 in viral infection and autoimmunity as well as the protein machineries for conjugating and processing ISG15[36].

1.3.2 SUMO

The SUMO (Small Ubiquitin-like MOdifier) family is composed of five members (SUMO1-5)[37]. Sequence alignment analysis showed that all the mature forms of SUMO family members have a ubiquitin-like scaffold (Fig 1.2.D&E) and a typical C-terminal diGly residues which are consistent with ubiquitin C-terminus (Fig 1.3). However, the C-terminal tails of SUMOs contain four residues completely different from the LRLR motifs in ubiquitin and ISG15. Although all members of SUMO share high sequence similarity, they still execute various functions in cells.

SUMO1 was initially found as a modification of mammalian Ran GTPase-activating protein RanGAP1 to regulate its interaction with RanBP2 to form a stable nuclear complex[38, 39, 40]. SUMO1 modifications regulate a wide variety of protein localization[41]. Polo-like kinase 1 (PLK1), a key regulator in mitosis, was identified to be a substrate of SUMO1. Once modified by SUMO1 on the residue Lys492, PLK1 was transported into nucleus with increased protein

stability[42]. Although SUMO1 does not function as ubiquitin to determine the protein degradation directly, modifications by SUMO1 compete with ubiquitin to block the degradation signal on cyclin-dependent kinase 6 (CDK6) in glioblastoma[43]. The stabilized CDK6 thereby contributes to cancer progression.

SUMO2 and SUMO3 were identified and cloned as close family members of SUMO1[44]. Their sequences share higher similarity in comparison to SUMO1. The sequence similarity affects their functions. For example, SUMO2 and SUMO3 regulate the production of interferon in a negative manner, while SUMO1 has no such kind of function[45]. The discrepancy between the mature forms of SUMO2 and SUMO3 only occurs in the N-termini (Fig 1.3) which determine that they share the same functions in many cellular processes. SUMO2 has been reported as the most abundant SUMO isoform and is only located in the nucleus while SUMO3 is widely spread throughout the whole cell[46]. The high expression level of SUMO2 protein is critical for normal embryo development[47].

SUMO4 has a sequence similarity of 86% to SUMO2 and is only detected in the kidney[48]. SUMO4 is able to modify IκBα to suppress the transcriptional activity of NF-κB. Mutation in SUMO4 (M55V) leads to enhanced NF-κB transcriptional activity and results in the activation of downstream genes which may be connected to the pathogenesis of type 1 diabetes[49]. A wide range of substrates were identified through a proteomic analysis which paved the road for further SUMO4-orientied research[50].

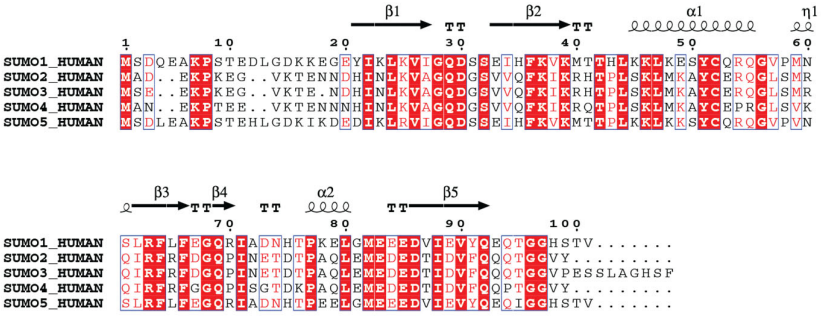


Fig 1.3. Sequence alignment of SUMOs. TT, α and β denote turn, α-helix and β-strand respectively.

SUMO5 is the least studied SUMO isoform which was the most recently identified isoform with strict tissue specificity in testes and peripheral blood leukocytes[51]. More exploration is needed to understand its functions different from other isoforms.

It is obvious that the functions of five SUMO paralogues which have been uncovered in human cells are merely a tip of the iceberg. Given that SUMO modifications regulate nuclear transport,

gene expression, DNA damage response and cell cycle, extensive investigation into SUMO signaling will open new avenues for target discovery and drug development.

1.3.3 NEDD8

NEDD8 (Neural precursor cell expressed developmentally down-regulated protein 8) displays approximately 60% sequence identity and 80% similarity with ubiquitin and thereby exists as the closest relative of ubiquitin among the Ubl family (Fig 1.2.A)[52]. Mature NEDD8 has also 76 amino acids, of which the C-terminal diGly is used for covalent attachment which is referred to as NEDDylation (Fig 1.2F). NEDD8 has a broad spectrum of substrates, among which cullins are the most studied and best-characterized[53]. Cullin family includes seven members that each in cooperation with a RING (Really Interesting New Gene) protein to form a core module of cullin-RING ubiquitin ligases (CRLs)[54]. The CRL activity can be regulated by several mechanisms, including NEDDylation. Structural studies showed that NEDDylation leads to striking conformational changes to enhance the CRL ubiquitination activity[55, 56, 57]. Therefore, NEDDylation control protein degradation indirectly.

Recent reports have concluded that NEDD8 has additional functions besides regulation of E3 ligases, such as transcriptional regulation[58]. In addition, an inhibitor that targets NEDD8 activation is being evaluated in clinical trials[59].

1.3.3 FUBI

Ribosomal protein FUBI-eS30/FAU is the precursor of FUBI, a ubiquitin-like protein that shares 36% identity with ubiquitin[60, 61]. Since FUBI bears a typical C-terminal diGly motif and ubiquitin-like core (Fig 1.2.G), it was speculated to function as a post-translational modification. Several substrates linked to immune response and apoptosis have been identified, including T cell antigen receptor-α[62], Bcl-G[63], endophilin II[64], and cytosolic 10-formyltetrahydrofolate dehydrogenase[65].

Despite of the known roles, FUBI is still a poorly studied modifier in many other aspects. Recently, a set of tools based-on recombinant FUBI protein enabled discovery of FUBI proteases through proteomics and quantitative biochemical assays[66]. However, how FUBI is attached to its substrates is still unknown.

1.3.4 Other modifiers

Apart from the aforementioned Ubls, more than 10 other ubiquitin-like modifiers have been discovered and studied. Representative roles of some Ubls have been revealed: FAT10 (F-adjacent transcript 10) is a regulator of immune responses[67]; the ATG family takes control of autophagy[68]; Ubiquitin-fold modifier 1 (UFM1) is the modifier to modulate ER-phagy[69]; Ubiquitin-related modifier 1 (URM1) acts as a carrier protein to transfer sulfur[70, 71].

Over the past decades, technological advances in structural biology, proteomics and molecular biology have increased the knowledge of the ubiquitin and Ubl system. Decoding the architecture and functions of ubiquitin and Ubls with proteomics and tailored chemical biology tools will accelerate the further understanding of their roles in physiological and pathological processes[72].

1.4 Conjugation of Ubiquitin and Ubiquitin-like Modifiers

The initiation of ubiquitin conjugation requires the activation of the terminal carboxylic acids of glycine residue, in which E1 activating enzyme together with ATP and magnesium cations transform ubiquitin into the active intermediate Ub~AMP (tilde (~) denotes covalent bond) with liberation of a pyrophosphate molecule (Fig 1.4.A). The adenylated ubiquitin is subsequently captured by the catalytic cysteine on the E1 to form a covalent complex Ub~AMP~E1 as a tetrahedral intermediate which is further transferred the catalytic cysteine of the E2 enzyme as a relatively stable E2~Ub thioester[73]. With the assistance of E3 ligases in a noncovalent manner or by forming an E3~Ub thioester, ubiquitin is conjugated to the target[74]. Multiple cycles can be repeated to polymerize the ubiquitin chain. In most cases, Ubl conjugations are also realized through the same E1-E2-E3 cascade with the involvement of different dedicated enzymes.

Although the E2~Ub complex is usually attacked by a lysine residue to form the conjugation through an isopeptide bond, increasing evidence has shown that other residues with nucleophilic substitution capabilities, such as threonine and serine, can also conjugate with ubiquitin through an ester bond. In addition, substrate scope has been expanded following the discovery of ubiquitinated lipids and lipopolysaccharide[75, 76, 77]. The discovery of non-protein ubiquitinated species has been actively executed which will deepen the insight into the substrate diversity.

1.4.1 E1 Activating Enzymes and E2 Conjugating Enzymes

In human genomes, eight E1 enzymes have been found to initiate Ub/Ubl conjugation[78]. UBA1 was the first identified E1 to activate ubiquitin and has been deemed as the sole E1 for ubiquitin activation for a long while. By using a homology motif search tool, UBA6 was characterized as a ubiquitin E1 which shares about 40% identity with UBA1[79]. Bioinformatics analysis and protein engineering study revealed that UBA1 and UBA6 are orthogonal E1s to be involved in modifications of distinctive substrates[80]. Advances in the structural biology have expanded the understanding of the cascade. With the development of specialized chemical biology tools, the snapshot of the adenylation and following attack by E1 enzyme was visualized in crystal structures[73, 81, 82]. In the E1 activation step, Ub~AMP~E1 intermediate exists as a high-energy state which can be mimicked by a vinyl sulfonamide moiety to give the suicide probe Ub~AVSN (Fig 1.4.B). The Michael acceptor captured the catalytic cysteine of E1 to form a stable complex Ub~AVSN~E1. The crystal structure of Ub~AVSN in complex with yeast Uba1 revealed the structural basis of the execution of E1 activation (Fig 1.4.C). So far, more than 24 E2 enzymes have been discovered, among which USE1 is the dedicated E2 for UBA6[83, 84] (Fig 1.4.D).

Besides ubiquitin conjugation, UBA6 in pair with USE1 has been reported to activate FAT10[84]. SUMO modifications are initiated by the heterodimeric complex of SUMO-activating enzymes 1 and 2 (SAE1::SAE2, double colon (::) denotes non-covalent interaction). The snapshot of SUMO activation was captured using a SUMO~AVSN probe [81, 82]. After activation, SUMO protein is transferred to E2 UBC9 to form UBC9~SUMO which can be hosted by dedicated E3s for conjugation. Thereby, isopeptide-linked RanGAP1~SUMO reconstituted by UBC9 has been a standard model substrate for biochemical research[85].

Similarly, NEDD8 activation also relies on a specialized heterodimeric complex formed by NAE1 (NEDD8-activating enzyme 1) and UBA3 (Fig 1.4.D). Two known E2s for NEDD8 are UBE2F and UBE2M which are specifically recognized by different NEDD8 E3 ligases to control the substrate specificity[86, 87]. NEDDylation has a pivotal role in regulating enzymatic activity of E3 ligases, among which cullins are the most predominant clients of NEDD8[55, 88].

Although the mature form of ISG15 has the same C-terminal residues as ubiquitin, the E1 enzyme UBA7 for ISG15 does not activate ubiquitin[89]. The downstream following UBA7 activation is controlled by E2 enzyme UBCH8 whose expression level is also controlled by interferon[90].

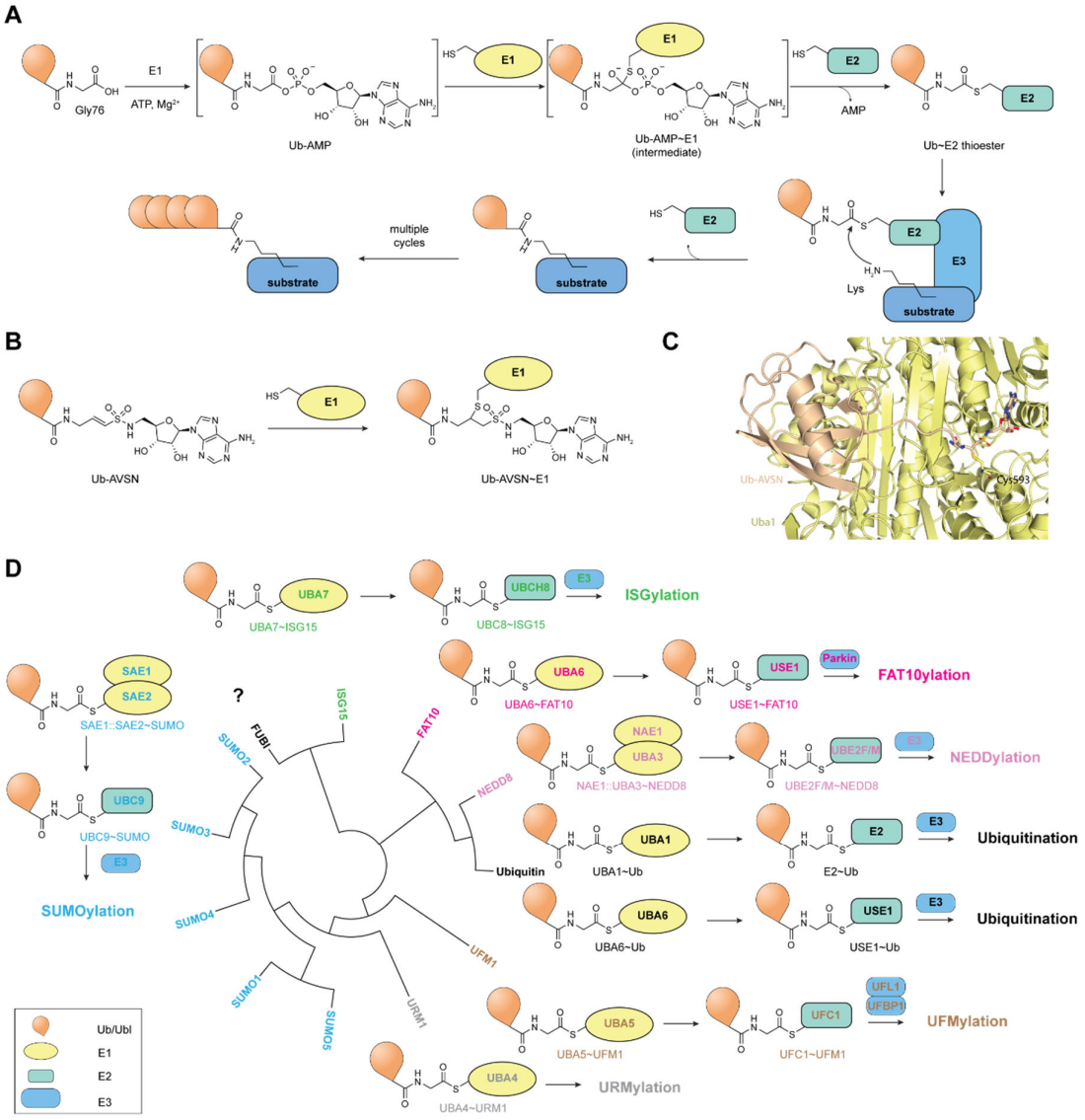


Fig 1.4. Conjugation of Ub/Ubl modifications. **A**. Overview of Ub conjugation. **B**. Structure of ubiquitin (PDB ID: 1UBQ). **C**. Structure of yeast Uba1 in covalent complex with Ub~AVSN (PDB ID: 6O83). **D**. Representative enzymes involved in Ub/Ubl conjugation, "~" denotes covalent complex, "::" denotes non-covalent interaction. This figure was adapted from references [26&78].

Apart from the aforementioned Ub/Ubl conjugation, URM1 and UFM1 have been reported to be activated by dedicated E1 enzymes UBA4 and UBA5 respectively. However, URM1 modifies substrates through a peroxidatic cysteine instead of the canonical E2/E3-dependent pathway[91]. The conjugation process of UFM1 requires E2 enzyme UFC1 and E3 complex of UFL1 and UFBP1[92, 93].

Unfortunately, the enzymes involved in FUBI conjugation have not been clarified yet. Further efforts are highly needed to fully characterize the FUBI modification since it is engaged in immune system[66].

1.4.2 E3 Ligases

E3 ligases are the key regulators for the specificity of protein ubiquitination. Human genome-wide annotation revealed that more than 600 putative E3s and substrate-recognition subunits of E3 complexes[94]. All E3 enzymes contain a domain for hosting Ub~E2 and are historically classified based on the structural characteristics and ubiquitin transfer mechanism into two classes: the RING (really interesting new gene) and the HECT (homologous to the E6AP carboxyl terminus) E3s[95]. With the investigation into the catalytic mechanisms of more E3s, a type of E3s acting like RING/HECT hybrids was discovered[96, 97]. This type of E3s has two RING domains (RING1 and RING2) which are separated by an in-between-RING domain (IBR) and thereby were named as RING-between-RING (RBR). Recently, the neuro-associated E3 ligase MYCBP2 has been depicted with a surprising ubiquitin esterification activity[98]. It has two essential catalytic cysteines to relay ubiquitin and thereby was named as RING-Cys-relay (RCR) E3 ligase.

The RING family members as the most abundant E3s feature a RING or U-box catalytic domain[95]. The RING domains contain a cysteine-rich motif to coordinate zinc ions to maintain the overall structures while U-box shares the similar structure without zinc coordination. RING E3s act as a scaffold protein to induce proximity of ubiquitin-charged E2 and substrate followed by a direct ubiquitin transfer[99]. RING E3 ligases are able to act in monomeric, homodimeric or heterodimeric manner. Some RING E3 contain multiple subunits which increase the diversity of E3 complexes.

In contrast to the one-step ubiquitination stimulated by RING E3s, HECT and RBR utilize an active cysteine to relay ubiquitin from E2 to substrate in two steps. HECT family is consisted of 28 members, of which the N-lobe is responsible for binding E2 and the C-lobe which contains an active cysteine for catalysis[8]. RBR family which have 14 members shares the features of RING and HECT E3s. The RING1 is the domain for recruiting ubiquitin-charged E2. The RING2 mediates the ubiquitination in a HECT-like mechanism[8]. Notably, some E3s catalyze linkage-specific ubiquitin chains. For example, UBE3C is a HECT E3 ligase which constructs Lys29 and Lys48-linked polyubiquitin[100, 101]. Apart from some mammalian E3s with linkage specificity, some bacterial E3s have also been reported to polymerize ubiquitin chains with defined linkage, such as NleL (non-Lee-encoded effector ligase) with the ability of assembling Lys6 and Lys48-linked ubiquitin chains[102]. The discovery and identification of specificity of E3s enable tool development for biochemical and mechanistic studies.

1.5 Deconjugation of Ubiquitin and Ubiquitin-like Modifiers

Ub/Ubl modifications are covalent attachments to substrates controlling plentiful cellular events. Reverse reactions are stringently required to maintain the proper modifications spatially and temporally. DUBs and evolutionarily related enzyme classes called ubiquitin-like proteases (ULPs) are the executors to protect protein homeostasis and prevent promiscuous and redundant Ub/Ubl modifications which could lead to a wide range of adverse effects and diseases[103].

DUBs and ULPs catalyze the maturation of Ub/Ubl precursors as well as cleave Ub/Ubl from substrates. DUBs and ULPs are usually assigned into two classes based on their catalytic mechanism. The majority of the DUBs/ULPs are cysteine proteases, which resemble the well-studied papain protease and rely on the active nucleophilic cysteine in corporation with adjacent histidine residue to hydrolyze the covalent bond between Ub/Ubl C-terminal glycine and their substrates. The other class is metalloproteases which utilize a zinc cation in the catalytic site in synergy with neighboring residues to active the water molecule to attack the covalent bond[104]. These enzymes are categorized into different families based on their primary sequences and features of their catalytic domains but the categories do not correlate with the specificity toward Ub/Ubls. Cross-reactivities in DUBs/ULPs have been observed frequently which contribute to the multiplexed functions of DUBs/ULPs in different cellular processes[103]. How many DUBs/ULPs have cross-reactivities, how those DUBs/ULPs achieve the cross-reactivities and how these enzymes behave in cellular settings are still poorly understood and remain systematic investigation.

Ubiquitin modifications often exist in the polymerized forms which can be composed of eight possible linkages. Most DUBs cleave ubiquitin chains indiscriminately while some DUBs show dedicated preference toward one or a small subset of linkages[105, 106]. Dissecting the mechanism of DUB linkage specificity would provide insights into linkage related biology and diseases. Despite of the accumulating structural information over the past years, *in silico* prediction of linkage and modifier specificity of DUBs/ULPs is still challenging since their specificities are not only determined by the C-terminal residues but also some other interaction sites, such as hydrophobic patches. Recent advances in development of biochemical assays, such as mass spectrometry-based DUB profiling using polyubiquitin with different linkages, and activity-based probes have enabled extensive characterization and study of specificity and revealed the structural basis of specificity[107, 108, 109].

1.6 Deubiquitinases

The human deubiquitinating enzyme family has around 100 members which are divided into seven subfamilies: USP (ubiquitin-specific proteases), UCH (ubiquitin carboxy-terminal hydrolases), MJD (Machado-Josephin domain-containing proteases), OTU (ovarian tumor proteases), MINDY (motif-interacting with ubiquitin-containing novel DUB family), ZUFSP (zinc finger with UFM1-specific peptidase domain protein) and JAMM (JAB1, MPN, MOV34 family) (Fig 1.5.A).

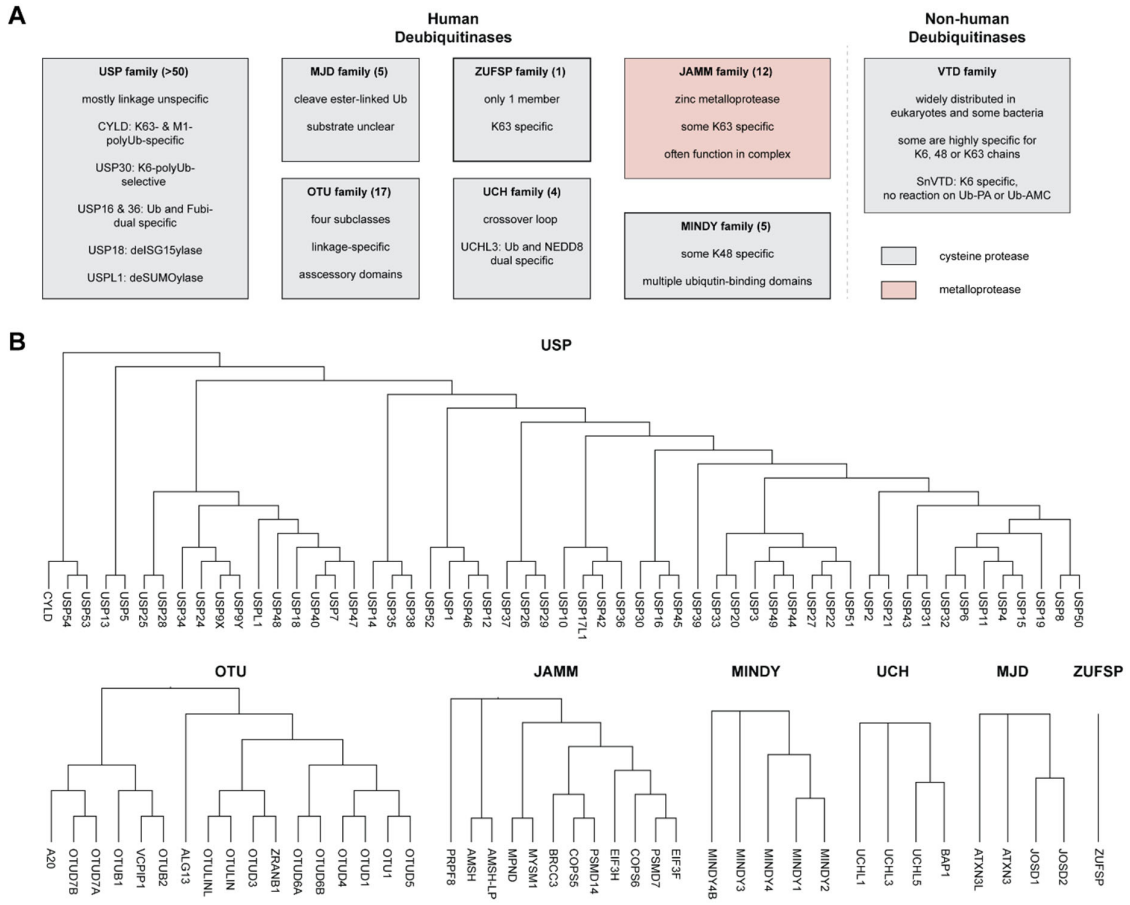


Fig 1.5. Overview of the classifications of human deubiquitinases. **A**. Families of human and non-human deubiquitinases depicted with features. Cysteine proteases are highlighted in grey and metalloproteases in salmon. **B**. Phylogenetic trees of human deubiquitinases based on their full-length sequences.

Recent bioinformatic exploration and biochemical validation have led to the discovery of a new class of DUBs as distant homologs of herpesviral large tegument proteins (VTD) in non-human species[110, 111]. Some of the VTDs display significant linkage preference. For example, SnVTD from *Simkania negevensis* is Lys6-linkage specific but has no activity toward ubiquitin-derived suicide substrate Ub-PA or fluorogenic substrate Ub-AMC[111]. Although the

classification is based on the protein sequence and conservation, DUBs within the same subfamily might possess distinctive enzymatic behaviors due to the subcellular localization and intrinsic specificity.

Among the subfamilies, USP is the largest group as represented by more than 50 validated or putative DUBs (Fig 1.5.B). The second largest DUB class is OTU family which has 17 members, of which linkage specificity attracts a lot of research interest[112]. JAMM family has 12 members while some of them are predicted inactive[113]. MINDY family contains 5 members, some of which show significant preference toward Lys48 linkage[114]. UCH and MJD have 4 members respectively while the newly uncovered ZUFSP forms a family on its own[115].

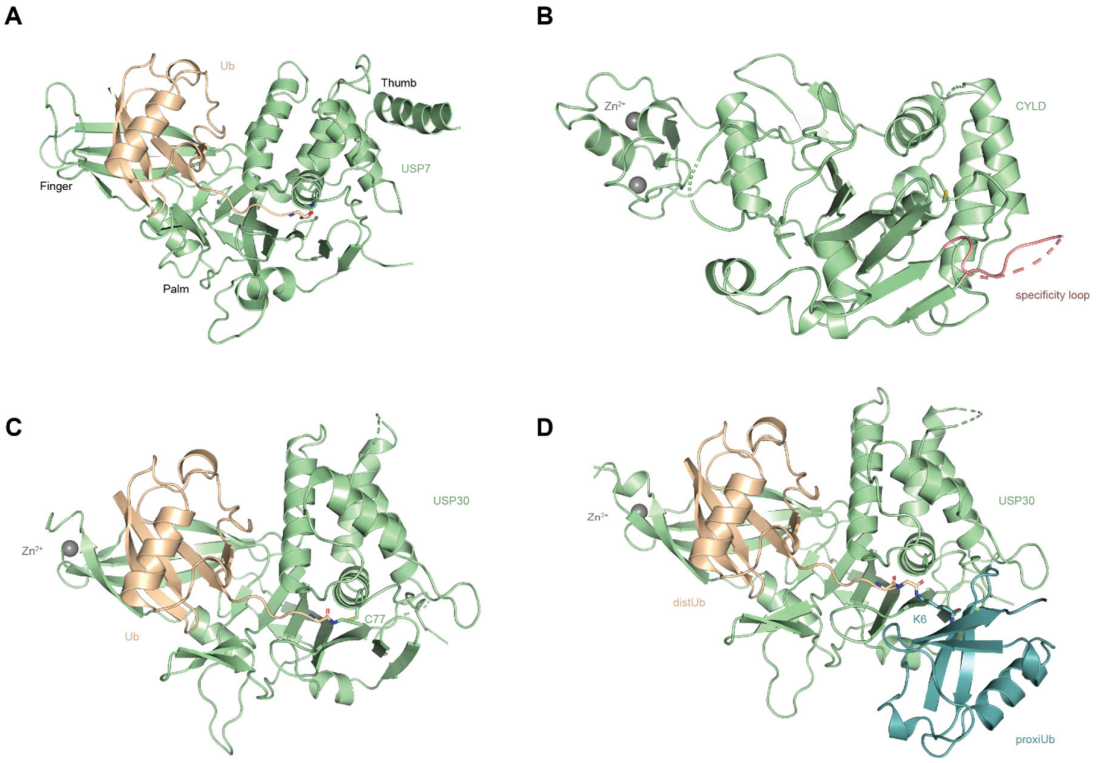


Fig 1.6. Representative structures of DUBs in apo form or in complex with ubiquitin. **A**. Structure of USP7~Ub-aldehyde (PDB ID: 1NBF). **B**. Crystal structure of apo CYLD (PDB ID: 2VHF). **C**. Crystal structure of USP30~Ub-PA covalent complex (PDB ID: 5OHK). **D**. Crystal structure of USP30 catalytically inactive mutant in complex with Lys6-linked diUb (PDB ID: 5OHP).

USP family members share the conserved architecture consisting of three domains, namely fingers, palm, and thumb which were annotated from the crystal structure of USP7 in apo form and in complex with ubiquitin, the first structure in USPs (Fig 1.6.A)[116]. Given that USP7 is associated with the stability of the tumor suppressor p53, medicinal chemistry efforts have led to a series of highly potent inhibitors[117]. Although most USPs cleave ubiquitin chains unselectively, a few outliers stand out with linkage specificity[105]. CYLD specifically cleaves

Lys63 and linear Met1 polyubiquitin chains to antagonize cytokine-mediated signaling pathway. Crystal structure and following biochemical study elucidated that the Lys63 specificity is dominated by an extended loop which shields the catalytic cysteine (Fig 1.6.B). This extended loop is remarkably distinctive from other previously reported USPs and confers the Lys63 specificity[118]. Another unique member is USP30 which has selectivity toward Lys6-linked polyubiquitin and functions as a negative modulator of mitophagy. The crystal structures of USP30 in complex respectively with mono-Ub and Lys6-diUb (Fig 1.6.C&D) revealed the structural basis of USP30 in regulating mitochondrial ubiquitination[119, 120].

Drug development targeting USPs has advanced into a new era due to increasing understanding of their functions and druggability. Recently, two inhibitors targeting USP30 have entered into clinical trials for treatment of acute kidney injury and Parkinson's disease, respectively[121]. Genome-wide screening also identified USP1 as a promising target inducing synthetic lethality[122]. Several inhibitors are being assessed in clinical studies (NCT05240898, NCT06065059).

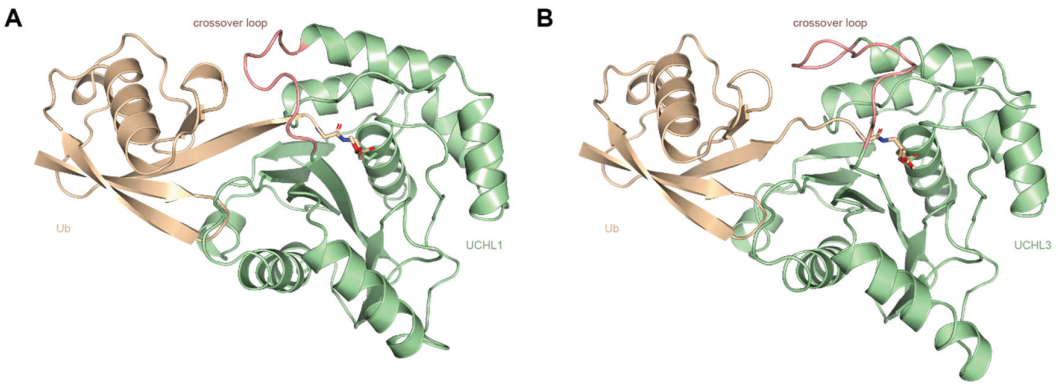


Fig 1.7. Representative structures of UCHs. **A**. Structure of covalent complex of UCHL1~Ub (PDB ID: 3KW5). **B**. Structure of covalent complex of UCHL3~Ub (PDB ID: 1XD3). The crossover loops are both highlighted in salmon color.

OTU DUBs are categorized in four subclasses: OTUDs, Otubains, OTULIN and A20-like OTUs. Systematic diUb cleavage assay elaborated the linkage specificity of all OTU members in four distinctive mechanisms[123]. OTUD1 shows highly activity and specificity on Lys63-linked chains. OTUB1 and A20 cleave Lys48-linked chains preferentially. Cezanne and Cezanne2 are both dedicated to Lys11-linked ubiquitin chains. Interestingly, OTULIN prefers to trim linear polyubiquitin chains and has been validated as a negative regulator of inflammation and autoimmunity[124]. Based on the specificity, some OTUs have been successfully employed for qualitatively identifying substrate linkage types in the Ubi-CREST assay[125]. Other

small modifications can also alter enzymatic acitivity. For instance, phosphorylation modifications act as a switch to enhance the deubiquitinating activity of OTUD5[126]. Strikingly, the Lys48-specific OTUD4 shifts its specificity to Lys63-linked chains upon phosphorylation by casein kinase II[127]. These results displayed the diverse modulation and interplay between post-translational modifications.

UCHs, the first discovered deubiquitinases, have four group members: UCHL1, UCHL3, UCHL5, and BAP1[128]. Structural studies showed that UCHL1 and UCHL3 both have a flexible crossover loop near the catalytic site acting as a substrate filter which prevents them from cleaving large substrates (Fig 1.7.A&B). Extension of the loop enabled both of them to cleave diUb substrates[129]. UCHL5 is bound to proteasome to debranch ubiquitin chains on substrates before degradation[130]. UCHL3 and UCHL5 have been reported to cross-react with NEDD8, while UCHL1 binds but does not cleave NEDD8[131, 132].

MJD family members have shown selective threonine and serine esterase activity[133]. However, the substrate scope of MJDs is still not very clear. MINDYs were discovered and characterized as Lys48-specific DUBs[134, 135]. MINDY1 and MINDY2 have five ubiquitin-binding domains (UBDs) to accommodate long ubiquitin chains whose length could affect the mode of cleavage[114]. Their cellular functions are still poorly studied and require further exploration.

The newly discovered ZUFSP is a Lys63-specific DUB whose specificity is realized through multiple tandem UBDs[136, 137, 138, 139]. Functionally, ZUFSP responses to DNA damage and maintains genome stability. Surprisingly, ZUFSP was initially annotated as an inactive UFM1 protease due to it was predicted to lack a histidine to form catalytic triad. Combined bioinformatic prediction and chemical biology tools led to the discovery and characterization of ZUFSP which highlight the importance of developing high-quality tools for studying DUBs and discovering unknown DUBs.

1.7 Ubiquitin-like Proteases

In analogy to DUBs, ULPs are the enzymes to activate the precursors of Ubls and cleave Ubls from their substrates. Cleavage of Ubls is not always executed by specific ULPs but by some DUBs due to their cross-reactivities. A better understanding of the cross-reactivities on the biochemical level would facilitate the decoding of the understudied cellular functions of DUBs/ULPs.

1.7.1 DeSUMOylases

SUMOs are one of the most abundant modifications among the Ubls. The already known SUMO proteases are from three distinctive families: sentrin-specific proteases (SENPs), de-SUMOylating isopeptidases (DESI) and USPL1 (ubiquitin-specific peptidase-like protein 1) (Fig 1.8.A). SENPs are consisted of six members: SENP1, SENP2, SENP3, SENP5, SEPN6, and SENP7. Although SENP8 was assigned into SENP family, it does not catalyze SUMO deconjugation, but plays a significant role in removal of NEDDylation[140, 141]. SENPs have specific subcellular localization which determine each characteristic functions[142]. Notably, members within SENP family also show preference toward SUMO paralogues[143]. DESI was identified as the second class of deSUMOylases from a yeast two-hybrid screening[144]. Cleavage assays showed that DESIs cleave isopeptide-linked SUMO1 but not precursor of SUMO1. DESI-1 was reported to trim polymerized SUMO2/3 as well[144]. Recently, a study investigated the deubiquitinating activity of DESI-2 and discovered ribosomal protein S7 as its substrate[145]. However, more biochemical data are needed for further validation.

USPL1 is included in the USP family but not equipped with any deubiquitinating activity. The development of SUMO-based covalent probe enabled the identification and validation of USPL1 as a SUMO protease. USPL1 shows preference toward SUMO2/3 over SUMO1[146].

1.7.2 DeNEDDylases

NEDDylation as the key regulator of E3 ligase enzymatic activity requires tight control by dedicated enzymes. UCHL3, but not UCHL1 was identified as a deNEDDylase from a yeast two-hybrid system[132]. Steady-state kinetic analysis revealed that UCHL3 has 1000-fold preference toward ubiquitin over NEDD8[141]. However, pull-down assay showed that UCHL3 possesses similar affinity toward ubiquitin and NEDD8[132]. These ambiguous data prevent further exploration of its functions in cellular context.

Another member in UCH family UCHL5 possesses capability of cleaving NEDD8 substrate[131]. COPS5 (also referred to as CSN5) from the JAMM family as a subunit of COP9 signalosome cleaves NEDD8 from cullin E3 ligase to suppress its ubiquitinating activity[147]. With the development of NEDD8 vinyl sulfone which is also an activity-based probe, SENP8 was identified as a deNEDDylase[141]. Increasing evidence showed that OTUB1 and OTUB2 both binds to NEDD8 and interestingly OTUB1 cleaves NEDD8[148, 149]. However, their functions related to NEDDylation remains unclear.

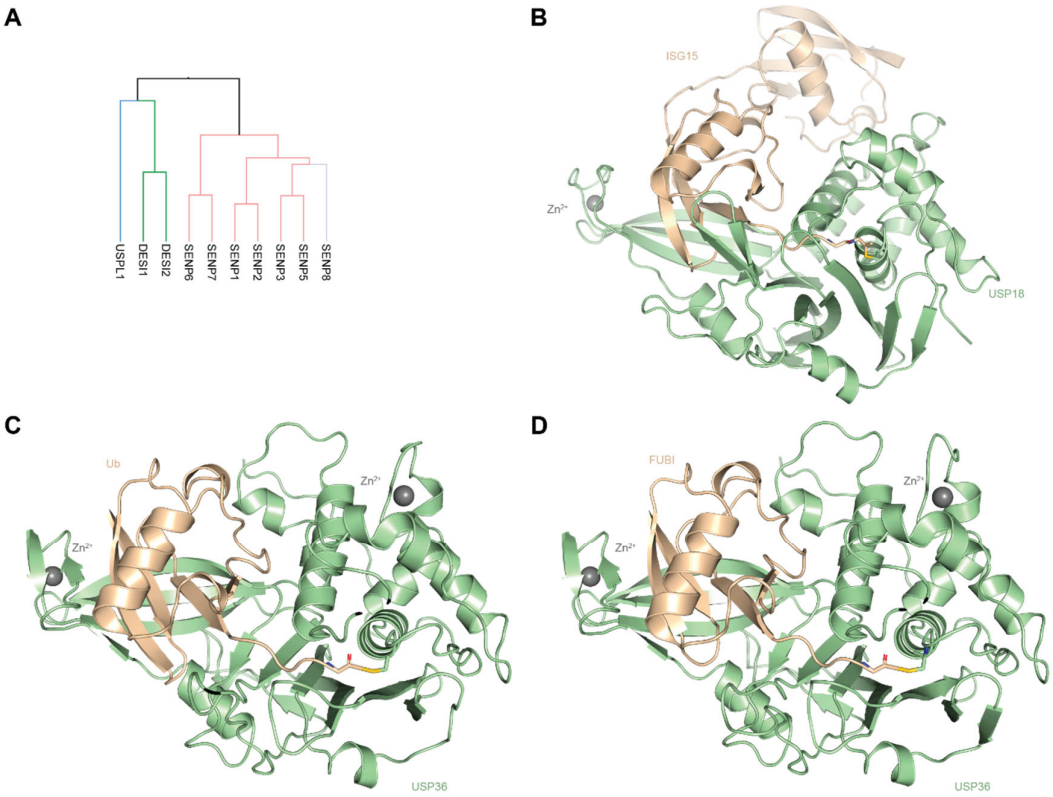


Fig 1.8. Overview of Ubl proteases. **A**. Phylogenetic trees of deSUMOylases which are highlighted in colors based on the subfamily classification. SENP8 is shown in purple since it acts on NEDD8 instead of SUMOs. **B**. Crystal structure of USP18~ISG15 (PDB ID: 5CHV). **C**. Crystal structure of USP36~Ub (PDB ID: 8BS9). **D**. Crystal structure of USP36~FUBI (PDB ID: 8BS3).

1.7.3 DeISGylases

ISG15 is associated with the immune response and thereby can be recognized and cleaved by proteases from viruses. The most outstanding example is the papain-like protease PLpro from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which preferentially cleaves ISG15 to regulate viral spread and attenuate innate immunity[150, 151]. In human genome, USP18 specifically deconjugates ISG15 but shows no any activity toward ubiquitin. The crystal structures of mouse USP18 in apo form as well as in covalent complex with ISG15 revealed the structural basis of its specificity toward ISG15 instead of ubiquitin (Fig 1.8.B). A hydrophobic interaction between USP18 and ISG15 contributes significantly to the structural basis of specificity.

Activity-based probes have enabled discovery of USP2, USP5, USP13 and USP14 as ISG15-reactive proteases[152, 153]. Isopeptide-bond-based substrates are needed to test these enzymes and rule out the false-positive candidates which will be beneficial for cellular studies of

deISGylases. Another Ubl that is actively involved in immune response is FAT10 whose conjugation machinery has been well-studied. However, the proteases responsible for deconjugating FAT10 are still unknown.

1.7.4 DeFUBIylases

USP36 was recently identified as a deFUBIylase for ribosomal subunit 40S maturation[61]. To search for proteases with FUBI specificity in genome-wide scale, activity-based covalent probe was designed and prepared. Through pull-down and proteomics, USP16 and USP36 were found to act on both ubiquitin and FUBI[66]. Crystal structures of USP36 in complex with FUBI and ubiquitin elaborated the substrate recognition mechanism and following mutagenesis assays explained the key residues involved in the cross-reactivities toward FUBI and ubiquitin (Fig 1.8.C&D). Structural information resulted in reprogramming the closely related USP42 into a FUBI protease by simply introducing an Arg-Tyr motif on the α-helix which is highly important in USP16 and USP36 for recognizing FUBI.

1.8 Tools for Deubiquitinases and Ubiquitin-like Proteases

Development of tools and assays in various formats over the past decades has tremendously facilitated the identification, validation, and qualitative and quantitative study of DUBs and ULPs[154]. Most DUBs and ULPs are cysteine proteases which are amenable to be targeted by activity-based probes (ABPs) to form covalent adducts[155]. However, the readout of ABPs reacting with DUBs/ULPs is usually based on gel electrophoresis which is time-consuming and unintuitively. To achieve kinetic and quantitative results, microplate-reader-based assays have been developed where Ub/Ubls derivatives can be turned over by DUBs/ULPs to release signals, such as fluorescence intensity, fluorescent polarization, and luminescence[156]. The combined uses of aforementioned tools provide a platform to discover new DUBs/ULPs and comprehensively understand the specificity of DUBs/ULPs.

1.8.1 Activity-based Probes

Ub/Ubl-based ABPs are normally consisted of three elements: (1) a reactive warhead which can covalently react with catalytic cysteine through an addition or substitution reaction; (2) a recognition element which is Ub/Ubl protein to be recognized and bound by DUBs/ULPs; (3)

a reporting tag which is used for orthogonal detection or enrichment but dispensable for structural biology research. Biotin, fluorophores or peptide epitopes have been installed at N-termini of Ub/Ubl to realize those purposes. Additionally, ubiquitination enzymes also sometimes employ catalytic cysteine to relay ubiquitin. ABPs thereby are suitable for broader applications in identifying and characterizing unknown enzymes involved in Ub/Ubl modifications. Ubiquitin ABPs can be classified into three groups based on the reaction types: direct addition, conjugate addition and nucleophilic substitution.

Ubiquitin-aldehyde (Ubal) was the first reported ABP in the Ub/Ubl research field in the occasion of investigating the catalytic mechanism of a DUB[157]. During the hydrolysis, the DUB was inactivated by borohydride which means the reaction intermediate was reduced. This discovery led to the development of first-generation ubiquitin covalent probe Ubal[158]. DUBs form a tetrahedron intermediate with Ubal by direct addition (Fig 1.9.A). Therefore, it served as a general DUB inhibitor for mechanistic studies[159]. Although the Ubal was developed in 1986, the first application in structural biology was in 1999 to give the crystal structure of yeast UCH, Yuh1 in complex with Ubal[160]. Following this successful application, the first complex structure from USP family, USP7~Ubal was reported which laid the foundation for drug discovery targeting USP7[116]. Ubal has been one of the most frequently used probes to capture DUBs and contributed to at least 12 complex structures[161]. However, the reversibility of hemithioacetal intermediate resulted to high heterogeneity as evidenced by the apo form of USP7 existing together with two USP7~Ubal complexes in the same asymmetric unit which increase the difficulty of crystallization[116]. Similarly, Ub-CN was also designed as a direct addition probe to form reversible complex with DUBs but not used widely[159]. Propargylamine was installed at C-termini of ubiquitin and SUMO to give the products which are referred to as Ub/Ubl-PA or Ub/Ubl-Prg. The alkyne handle of Ub/Ubl-PA was initially designed for click chemistry to forge model substrates for mechanistic studies[162, 163].

Strikingly, Ub-PA showed strong inhibitory activity against some DUBs. Subsequent studies proved its capability of trapping DUBs by forming a vinyl thioether which is irreversible and thereby suitable for structural biology studies[164]. So far, PA-based strategies have been implemented in other Ubls, including ISG15[165], SUMO[166, 167], and FUBI[66].

The first series of ubiquitin probes with Michael acceptors (Ub-VS (vinylmethyl sulfone), Ub-VSPh (vinylphenyl sulfone), Ub-VME (vinyl methylester), and Ub-VCN (vinylcyanide)) were assembled for functional proteomics leading to discover a series of DUBs[168] (Fig 1.9.B). Ub-OEtVS (vinylethoxy sulfone) was introduced shortly after as an alternative for comparison[169].

Among these probes, Ub-VS and variants for ISG15 and FUBI have been successfully used for discovering and characterizing previously unknown proteases[66, 152]. Ub-VME has been widely used for crystallography, especially for UCHL1[170], UCHL3[171] and USP46[172]. Recently, Ub-Dha of which Gly76 of ubiquitin was replaced with a dehydroalanine (Dha) was developed as a cascading ABP to capture a plethora of E1, E2, and E3 enzymes[173]. The same warhead was adapted to ISG15 for proteomic application[153].

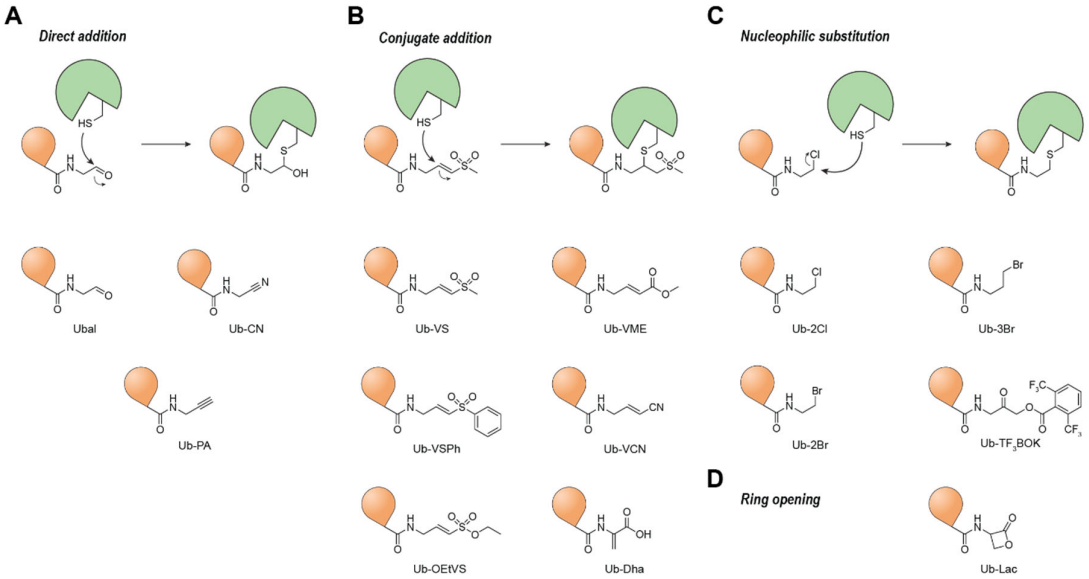


Fig 1.9. Reported mono-Ub-based covalent probes. **A**. Probes with direct addition mechanism. **B**. Probes with conjugate addition mechanism. **C**. Nucleophilic substitution-based covalent probes. **D**. Lactone-based ubiquitin probe with potential ring opening mechanism upon reacting with DUBs.

The third type of Ub/Ubl ABPs is featured by their nucleophilic substitution warheads which function as a leaving group upon attack by DUBs/ULPs (Fig 1.9.C). This type of ABPs has shown broad applications in structural biology, especially in understanding the specificity of viral DUBs toward Ub and ISG15[174, 175]. Moreover, a lactone-bearing ubiquitin probe was developed but barely applied[169].

Overall, various Ub/Ubl-based ABPs have been developed for DUBs with versatile functions. Nevertheless, there are still some challenges to overcome. Since all the ABPs are protein-based, cell permeability is a roadblock for cellular applications. Recently work showed that ABPs modified with cell-penetrating peptides are able to cross the cell membrane for intracellular profiling[176]. A photo-caged ubiquitin probe was further developed to enable temporal profiling of DUBs[177]. Another challenge is to develop probes for zinc-dependent DUBs. Zinc-

chelating moiety has been installed into ubiquitin for JAMM DUBs[178]. Although these probes have been successfully developed, more application scenarios should be tested to interrogate the functions of DUBs.

1.8.2 Fluorescence Intensity and Polarization Assays

Ub/Ubl-based ABPs are versatile tools for the characterization of DUBs as well as ubiquitinating enzymes. However, the intrinsic features of ABPs, such as low throughput, limit their applications. Plate-based assays are amenable to miniaturization which is suitable for high through-put screening. Derivatized from the cleavage sites in Ub/Ubls, a variety of substrates have been designed, among which fluorescence intensity and polarization assays are the most frequently used because the assays are easily set up and suitable for high-throughput screening (Fig 1.10.A). The fluorogenic substrates for DUBs/ULPs whose C-termini are modified through an aryl amide bond with a fluorophore have been widely employed for drug screening[179, 180]. The fluorophore is quenched but can be released upon activation of DUBs/ULPs to transmit signals. Ub/Ubl-RhoG is a typical fluorogenic substrate which contains a Rhodamine 110-glycine (RhoG). Previous efforts have led to multiple methods to prepare these substrates[106, 181, 182] (Fig 1.10.B). However, these methods utilized harsh conditions in either reaction or purification steps which could result in misfolding. Moreover, the acryl amide bond is significantly different from the natural isopeptide which might lead to misinterpretation of catalytic specificity of DUBs/ULPs.

Fluorescence polarization substrates (Ub/Ubl-KG-TAMRA) have natural isopeptide bonds to connect fluorophore TAMRA through the amine on the lysine side chain with Ub/Ubl (Fig 1.10.C). This assembly reveals the authentic linkage between Ub/Ubls and their substrates. Thereby fluorescence polarization substrates are considered as suitable DUBs/ULPs substrates for quantitative analysis of specificity as well as inhibitor evaluation[183].

However, the preparation of fluorescence polarization substrates requires native chemical ligation using a δ-mercaptolysine-containing peptide which requires nearly ten-step synthesis, free-radical-based desulfurization, purification by high-performance liquid chromatography (HPLC) and subsequent lyophilization. The resulting powder needs refolding in aqueous buffer which can lead to heterogeneity especially in the case of Ubls[184]. Previous studies showed that fluorescence polarization substrates cleaved by DUBs/ULPs cannot reach full conversion to the level of free KG-TAMRA which limits their broader applications[184, 185]. Overall, a mild, facile

and transferable method for preparing Ub/Ubl fluorescence polarization substrates is in high demand for this field not only for delineating the specificity of DUBs/ULPs but also for future drug discovery.

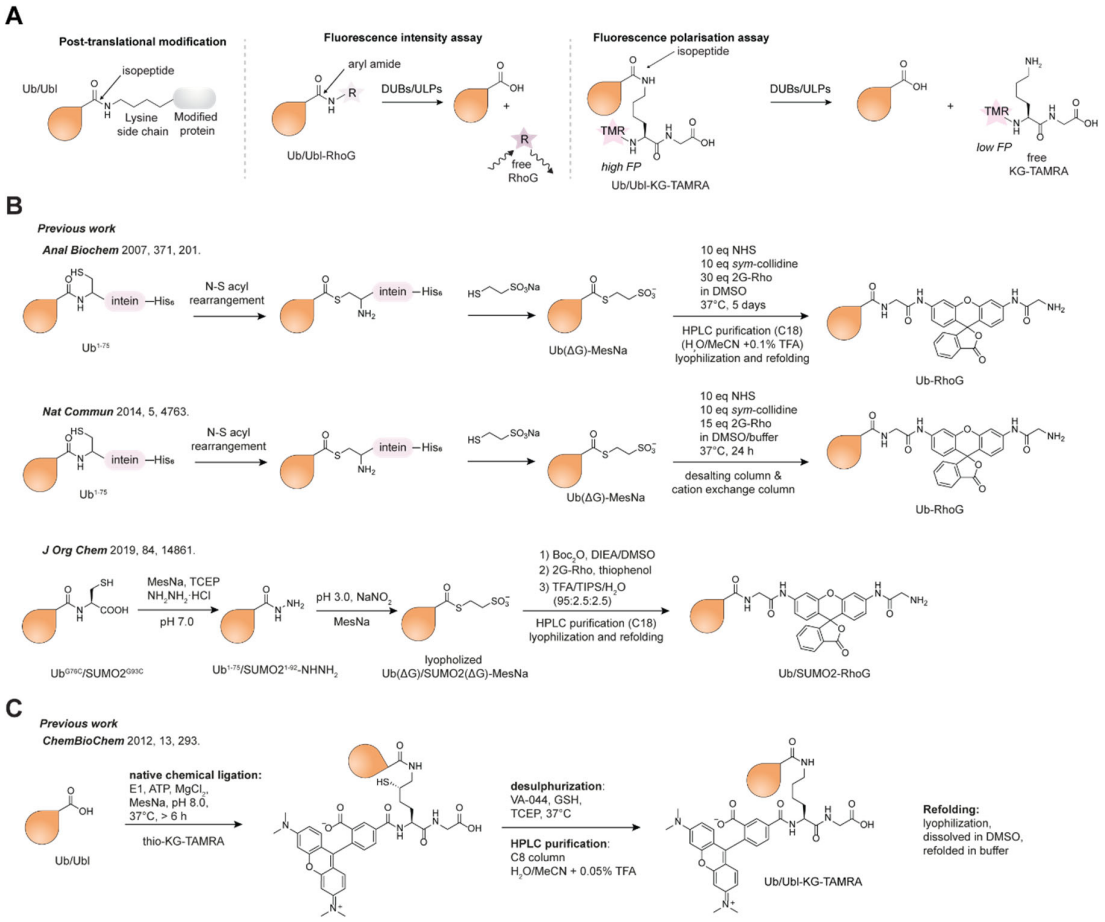


Fig 1.10. Overview of fluorescence-based Ub/Ubl substrates. **A**. The structures of native isopeptide bond, Ub/Ubl-RhoG, and Ub/Ubl-KG-TAMRA. **B**. Previous efforts to generate the fluorescence intensity assay platform. **C**. Previous work that describes the preparation of Ub/Ubl-KG-TAMTA. VA-044 is a desulfurization reagent.

1.9 Di-Ubiquitin Activity-based Probes: Design and Applications

The complexity of the ubiquitin modifications determines that DUBs must possess multiple layers of specificity from substrate specificity to ubiquitin chain linkage specificity. Linkage specificity of DUBs is usually characterized by polyubiquitin cleavage assay which is resolved on gel electrophoresis. To understand how the linkage specificity is achieved, di-ubiquitin (diUb) ABPs equipped with an internal covalent warhead are powerful tools because they react with the catalytic cysteine and interact with both distal (S1) and proximal (S1') sites which would provide answers to the specificity. Moreover, some HECT and RBR E3 ligases show

extraordinary preference toward forging specific linkage type which might be revealed by the diUb ABPs.[100, 186]. DiUb ABPs would provide a platform for dissecting the mechanism of linkage specificity of E3s since they might be capable of stabilizing the DUBs or E3 ligases which would be beneficial for the structural determination. Additionally, potential applications in proteomics might lead to discovery of new enzymes in the ubiquitin system.

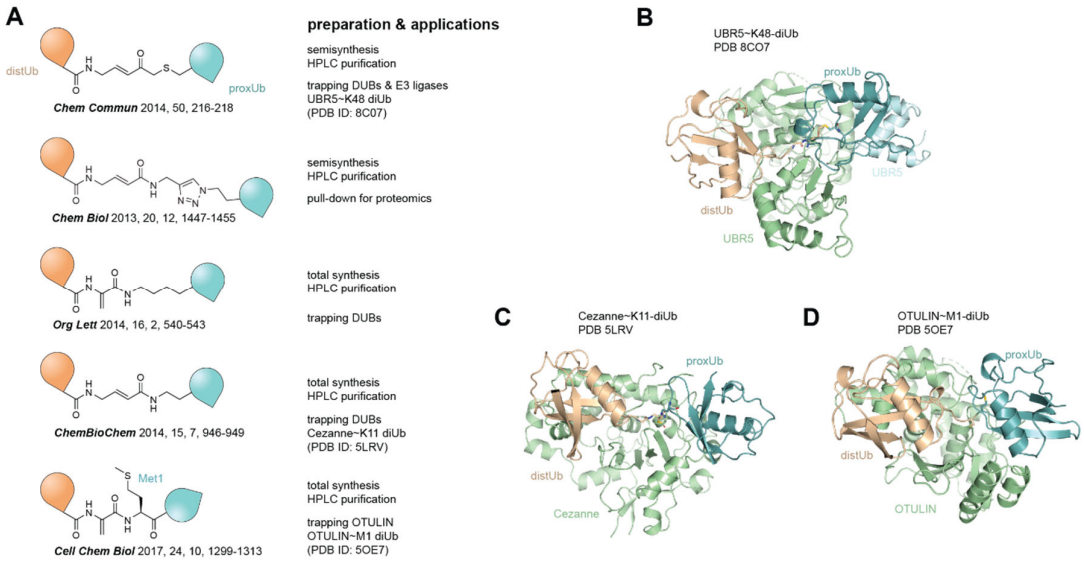


Fig 1.11. DiUb-based covalent probes with internal warheads for trapping DUBs. **A**. Representative structures of reported diUb covalent probes with Michael acceptors. **B**. The covalent structure of HECT E3 UBR5 with Lys48-linked diUb. **C**. Crystal structure of Cezanne~Lys11-linked diUb. **D**. Structure of OTULIN~Met1-linked diUb.

1.9.1 Previous Development of DiUb Probes

Mono-Ub-based ABPs enable the visualization of how DUBs bind and cleave ubiquitin, but cannot reveal how the proximal ubiquitin interacts with DUBs to achieve linkage specificity. To obtain the complex structure of DUBs and related linkage-specific diUb, the catalytic cysteine was usually mutated to alanine for crystallization. However, sample homogeneity might be an issue due to excessive addition of diUb chains[187, 188]. Therefore, developing diUb ABPs with internal covalent warheads to capture DUBs would enable structural biology studies to analyze the interaction between DUBs and distal and proximal ubiquitin molecules.

Early efforts in trapping DUBs employed a Michael acceptor at the end of ubiquitin coupled with some peptides surrounding the specific lysine[189]. However, this strategy cannot fully represent the diUb molecule since the proximal ubiquitin might form interaction with DUBs in different area[189]. Expanded from the mono-Ub ABPs with conjugation addition warheads, the

proximal ubiquitin was connected through various chemistry to form diUb ABPs with defined linkages (Fig 1.11.A). Some of the probes have been successfully employed for structural determination of Lys48-specific HECT E3 UBR5[190] (Fig 1.11.B), Cezanne in complex with Lys11-linked diUb[191] (Fig 1.11.C), and covalent complex of OTULIN with the Met1-linked linear diUb-Dha probe[192] (Fig 1.11.D).

1.9.2 Challenges in DiUb Probes

Although only three complex structures empowered by the diUb ABPs have been reported so far, the snapshots of proximal ubiquitin binding to the enzymes have already provided deeper insight into the mechanism of specificity. Unfortunately, all the methods to prepare diUb ABPs cannot avoid either harsh reaction conditions or HPLC purification which leads to misfolding. Additional atoms in the warheads might hinder proper recognition by DUBs. All the aforementioned factors restrain the accessibility and applicability of the diUb probes.

Preparation of the diUb probe for capturing UBR5 firstly requires synthesis of acetal-protected warhead (Fig 1.12.A). Following the installation of caged warhead, trifluoroacetic acid (TFA) was mixed into the reaction to remove the protecting group and release the covalent warhead. The resulting intermediate was harshly treated with cold ether to form precipitate for purification[193]. The sophisticated synthesis and purification workflow restricted the broad application especially for structural biology studies which usually require a large quantity of probes.

For the purpose of proteomics studies, click chemistry was utilized to generate a series of diUb ABPs [194] (Fig 1.12.B). The C-terminus of proximal ubiquitin was functionalized by intein chemistry with a covalent warhead as well as an alkyne handle for the purpose of click reaction. The distal ubiquitin was incorporated by genetical expansion with an azidonorleucine. Although this method was straightforward in preparing the complex, the length of the linker between proximal and distal ubiquitin molecules is significantly longer than the natural isopeptide bond which might affect the recognition of some DUBs that have very sensitive binding mechanisms toward polyubiquitin chains.

To keep the length of the linker in consistency with the natural isopeptide bond, some other warhead design and synthetic methods were explored. Given that ubiquitin has no any cysteine residues, incorporation of cysteine into specific position followed by radical thiol elimination was a successful strategy to generate Dha-based diUb ABPs[195] (Fig 1.12.C). A linear diUb probe bearing a Dha warhead at the N-terminus of distal ubiquitin was also prepared using the

same strategy. In addition, the total synthesis of ubiquitin fragments requires intense synthetic efforts which might not be easily transferable between laboratories.

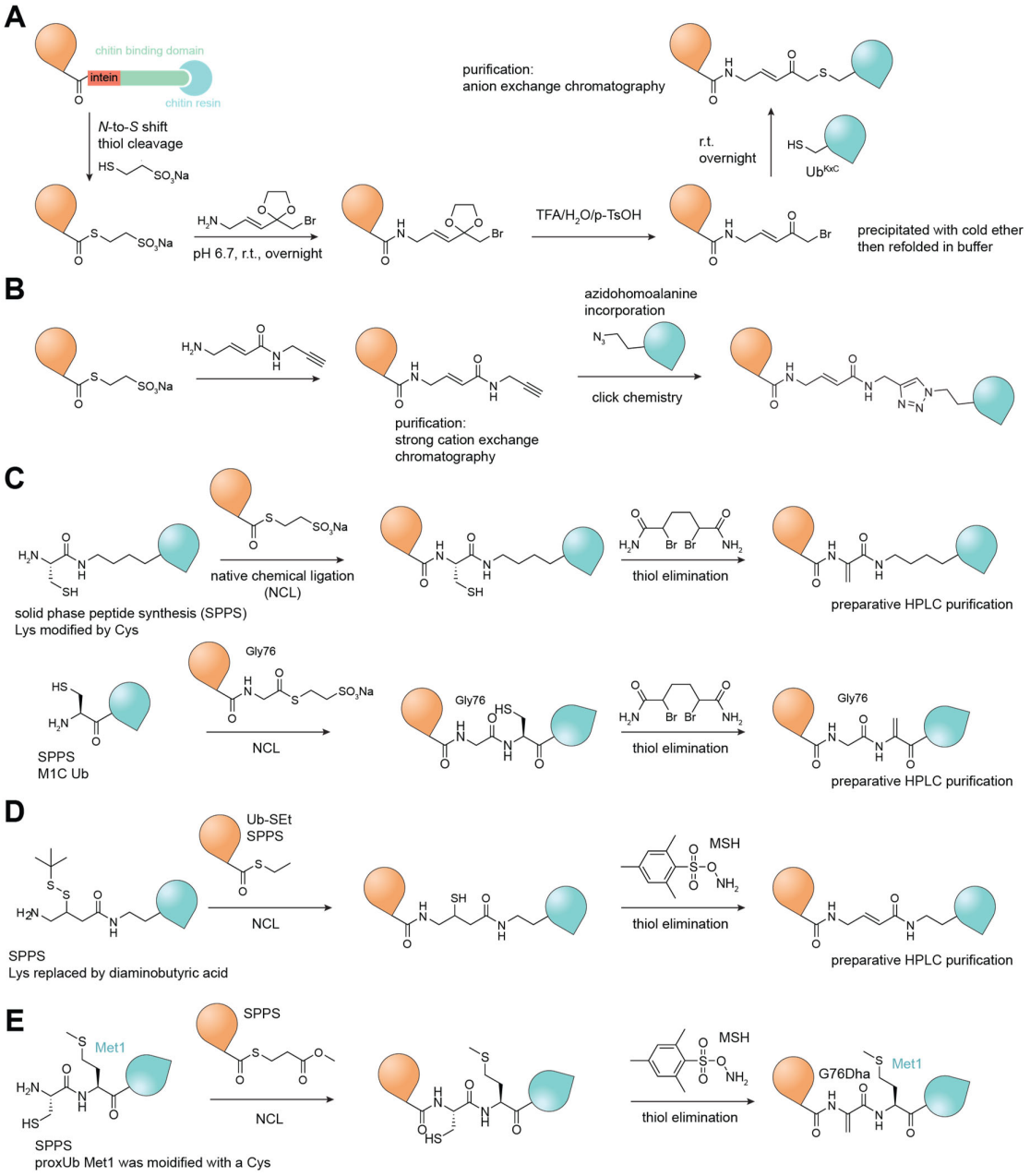


Fig 1.12. Previous methods in preparation of representative diUb probes with internal warheads. **A**. Incorporation of a Michael acceptor into diUb through an acetal-protected warhead. **B**. Click chemistry enabled the preparation of diUb probes. The proximal ubiquitin was functionalized with a warhead and a terminal alkyne. Azidonorleucine was incorporated into distal ubiquitin. **C**. Solid phase peptide synthesis and subsequent thiol elimination led to the Dha-containing diUb probes. **D**. Preparation of Michael acceptor-containing diUb probe using total synthesis method. **E**. Preparation of linear diUb-Dha probe using SPPS and thiol elimination at position 76 of the proximal ubiquitin.

Similarly, thiol elimination method was utilized for the preparation of another Michael acceptor-based diUb ABP which was successfully implemented for solving the crystal structure of Cezanne in covalent complex with the Lys11-linked diUb probe[191] (Fig 1.12.D). Incorporation of Dha into the first residue of proximal ubiquitin disabled the reactivity toward Met1-specific OTULIN due to the wrong positioning (Fig 1.12.C). However, another design that introduced Dha warhead into the Gly76 position of distal ubiquitin resulted in the structure determination of OTULIN in covalent complex with the linear diUb probe though the resolution was worse than the previously reported complex of catalytically inactive OTULIN with native linear diUb[196] (Fig 1.12.E).

It is obvious that difficulties in syntheses result in limited access to the diUb covalent probes. To complement previously reported probes, novel tools with differential covalent modification mechanism which can be prepared in a facile manner and in a large quantity could offer an alternative to study the linkage specificity of DUBs. Beyond the applications in investigating DUBs, new tool would offer additional options for studying those E3s which have a catalytic cysteine.

1.10 Aims

Extensive studies in the specificity mechanism have triggered the conceptualization of multiple layers of specificity that DUBs possess to modulate the ubiquitin code[105]. These layers are the molecular basis of regulating Ub/Ubl-related cellular events in multiple dimensions. Innovative chemical biology tools are the key to dissect the mechanism underlying the DUB specificities in many aspects. However, one of the biggest challenges in preparing Ub/Ubl-based tools is the refolding step which dominates the quality of the tools because harsh reaction conditions, HPLC purification in organic solvent and following lyophilization are able to completely destroy the correct conformations[197]. Although refolding from DMSO stock into buffer might help ubiquitin-based substrates refold to a certain degree, some Ubls are still difficult to obtain due to the refolding procedure. To aid preparing tools for quantitative studies on DUBs, it is worthwhile to explore mild reaction conditions and gentle purification protocols by fast protein liquid chromatography (FPLC) in aqueous buffer at mild pH to avoiding the lyophilization step. Hence, the aims of the thesis are to investigate two major layers of DUB specificities, namely (I) modifier specificity and (II) linkage specificity enabled by the innovation of chemical biology tools which are prepared in native states.

To achieve the aim I, a bespoke protocol for native semisynthesis of isopeptide-containing Ub/Ubl-KG-TAMRA is to be implemented for quantitative analysis of several DUBs/ULPs toward Ub/Ubls in fluorescence polarization assays. The deSUMOylase USPL1 will be extensively investigated with the aid of structural biology to elaborate how this USP evolutionarily achieves preference toward SUMO2/3 modifications instead of ubiquitin or other SUMOs. On the basis of the assays, a trial high-throughput screening campaign based on the fluorescence intensity assay will be launched to discover novel tool compounds targeting USPL1. To unleash the power of the fluorescence polarization assay platform, a variety of DUBs and ULPs will be examined to discover unknown cross-reactivities towards Ub/Ubls in a quantitative manner.

To achieve the second goal, novel diUb-based probes with internal covalent warheads will be designed and prepared in native states. The probes will be tested in different settings, including the gel-based binding assay and X-ray crystallography. Linkage-specific DUB SnVTD from bacteria will be selected as a case study to visualize how its binding sites are occupied by both proximal and distal ubiquitin molecules simultaneously. Beyond the applications in DUBs, HECT-type E3 ligases will be explored with the covalent diUb probes.

Collectively, this work has the potential to establish an assay platform of which substrates are linked with fluorophores via a natural isopeptide linkage for quantitative analysis of modifier specificity of different DUBs/ULPs. This effort to establish the assay platform would pave a new way for elucidating the functions of DUBs/ULPs and provide new insights into future mechanistic research. In addition to the modifier specificity, the design and development of novel diUb probes would offer a new set of tools to study the linkage specificity of DUBs as well as E3 ligases. These tools would provide new options to understand these two major layers of specificity of DUBs/ULPs.

2. MODIFIER SPECIFICITY

Dissecting the specificity of DUBs/ULPs toward Ub/Ubl modifiers on the biochemical level would provide the foundation for functional analysis of DUBs/ULPs in cells. To quantitatively investigate the catalytic specificity of DUBs/ULPs, a continuous readout of cleavage is preffered which not only provides high-throughput capabilities but also requires less protein consumption. To monitor the cleavage kinetics, a series of fluorescence intensity and polarization substrates have been introduced but most of them are prepared in harsh conditions which results in misleading data interpretation. Mild reaction and tailored purification tactics are the solutions to overcome the difficulties in preparation of high-quality assay substrates as well as provide convenient access to a good quantity.

2.1 Access to the Fluorescence Intensity Substrates

Early effort to obtain fluorescence intensity used trypsin, which is a protease that can cleave protein after arginine residues, to activate ubiquitin followed by the attack by the terminal amine of diGly-AMC (aminomethylcoumarin)[198]. Intein-based chemistry was utilized to optimize the production of the substrates[141]. AMC-based substrates enabled the qualitative validation of USPL1 specificity. However, there is a critical drawback of Ub/Ubl-AMC, especially when they are used for screening and evaluating inhibitors. AMC has a short excitation wavelength and low quantum yield. The intrinsic defects limit the optimization of assays[181]. Therefore, a better fluorophore Rhodamine 110 quenched by glycine residues was installed at the C-terminus of ubiquitin and purified by HPLC[181]. Another work reported a method where all the free amines are protected by tert-butyloxycarbonyl (Boc) group in DMSO, functionalized and purified by HPLC. These methods require a refolding step which reduce the quality of substrates. To retain the native state, FPLC purification is an option[199].

Intein-fused Ub/SUMO proteins were expressed in bacteria, followed by the on-bead cleavage by the thiol in sodium 2-mercaptoethane sulfonate (MesNa), to obtain active but stable intermediate Ub/SUMO-MesNa (Fig 2.1.A). To convert the thioesters to product, NHS was added to the reaction together bis-glycyl-Rhodamine 110 (see MATERIALS & METHODS for synthesis) to attack the thioesters. Following straightforward purification by cation exchange chromatography, these reagents were prepared in pure form (Fig 2.1.B). USPL1 catalytic domain was purified (Fig 2.1.C) and tested in the cleavage assay using Ub/SUMO-RhoG substrates

(Fig 2.1.D). Quantitative analysis showed that USPL1 cleaves SUMO2 preferentially with a 4-fold higher catalytic efficiency than SUMO1, but no action on ubiquitin substrate which is in line with previous result (Fig 2.1.E&F)[146].

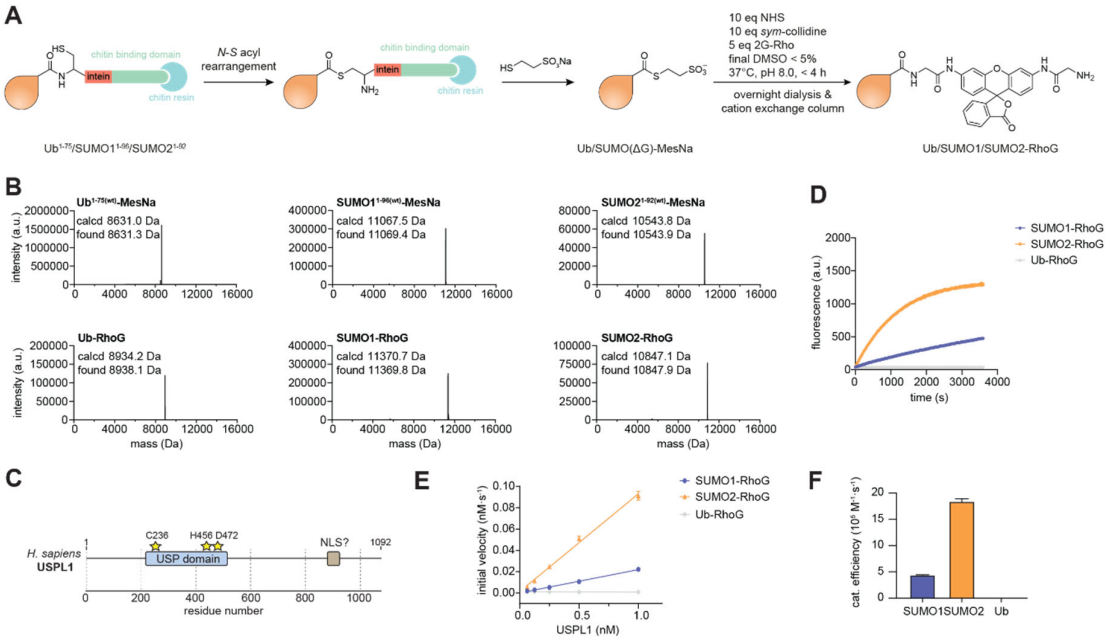


Fig 2.1. Preparation of fluorescence intensity substrates and their applications in USPL1. **A**. Workflow of native synthesis of Ub/SUMO-RhoG. 2G-Rho: bis-glycyl-Rhodamine 110, NHS: *N*-hydroxylsuccinimide. **B**. Deconvoluted mass spectra of Ub/SUMO-MesNa and Ub/SUMO-RhoG. **C**. Overview of USPL1 protein architecture. The catalytic USP domain and the nuclear localization sequence (NLS) are shown as boxes, and residues of the catalytic triad as stars. **D**. Representative fluorescence over time trace ([USPL1] = 0.4 nM, [Ub/Ubl-RhoG] = 50 nM). **E**. Linear regression of initial velocity of USPL1 on the substrates over the enzyme concentration. **F**. Catalytic efficiencies of USPL1 determined for indicated substrates as mean ± standard error.

The well-folded substrates prepared in large quantity enabled quantitative comparison but the interpretated data cannot reveal the authentic specificity due to the artificial linkage between Ub/SUMO and fluorophore which is different from the natural isopeptide linkage[200]. Therefore, fluorescence polarization substrates whose fluorophores are linked via an isopeptide bond are suitable for examining the specificity of USPL1[200]. Previous work utilizing solid phase peptide synthesis (SPPS) to forge the isopeptide bonds was performed and purified in organic solvents[184]. Instead of using SPPS, native semisynthesis starting from recombinant protein offers mild reaction condition. It has been reported that generation of acyl azides from ubiquitin hydrazides with nitrous acid followed by subsequent conversion into amides to afford ubiquitin-based tools (Fig 2.2.A)[201, 202]. Instead of using trypsin to activate ubiquitin, the stable thioester (Ub-MesNa) was quickly transformed to the hydrazide form which was highly concentrated (Fig 2.2.B). To functionalize the C-terminal glycine, Ub-hydrazide was kept at -10°C followed by the treatment of sodium nitrite and citric acid for a few minutes to give the active

ubiquitin acyl azide. KG-TAMRA was then immediately added into the mixture and incubated at 37°C for several minutes to substitute the azide moiety to forge a new isopeptide bond (Fig 2.2.A). Previous work to incorporate the KG-TAMRA required a thiol on the lysine side chain as an auxiliary group to facilitate the thioesterification which is the prerequisite of the formation of isopeptide bonds. However, the unnatural thiol-modified lysine required intense synthetic efforts. Free KG-TAMRA synthesized in solid phase method as well as an optimized method in solution were developed (synthetic procedures in MATERIALS & METHODS). In addition, the reaction and purification should be shielded from light to avoid any potential bleaching of the fluorophore.

Similarly, the reaction was performed on SUMO1, but side reaction occurred on the cysteine which was nitrosylated. However, this modification can be reversed by simply adding reducing agent tris(2-carboxyethyl)phosphine (TCEP). In general, this functionalization protocol enabled desired product in 20-40% yield, with various potential side products including dimers, hydrolyzed Ub/SUMO, as well as lactams. Because the installation of TAMRA which is tautomerized dependent on pH value affects charge states, purification of Ub/SUMO-KG-TAMRA was challenging and thereby tailored purification procedures were scouted and optimized (Fig 2.2.C). Ub-KG-TAMRA was purified by two-round cation exchange chromatography at pH 6 and 4.5 respectively. SUMO1-KG-TAMRA was firstly purified by size-exclusion chromatography to remove free KG-TAMRA and then refined by high-resolution anion exchange chromatography. SUMO2-based reagent was also purified in two steps. All the reagents were purified in pure form and validated by gel analysis and intact protein mass spectrometry (Fig 2.2.D&E).

To further validate if the functionalized proteins are still naturally folded, circular dichroism (CD) spectroscopy was employed to assess the quality since the CD property is the golden standard for the determination of protein secondary structures. Ub/SUMO2-KG-TAMRA both possess similar CD properties compared to their parental forms (Fig 2.2.F). The CD data thereby confirmed the natural confirmation of Ub/SUMO2-KG-TAMRA. Since deSUMOylase USPL1 is not active on ubiquitin substrates, quality check for Ub-KG-TAMRA was realized by deubiquitylating enzyme USP2. The fluorescence polarization assay and data analysis confirm the high quality of Ub-KG-TAMRA because the substrate was fully cleaved and reached the level of free KG-TAMRA (Fig 2.2.G). Since the assay was performed using a broad range of USP2 concentration, quantification of its catalytic efficiency on Ub-KG-TAMRA was possible (Fig 2.2.H).

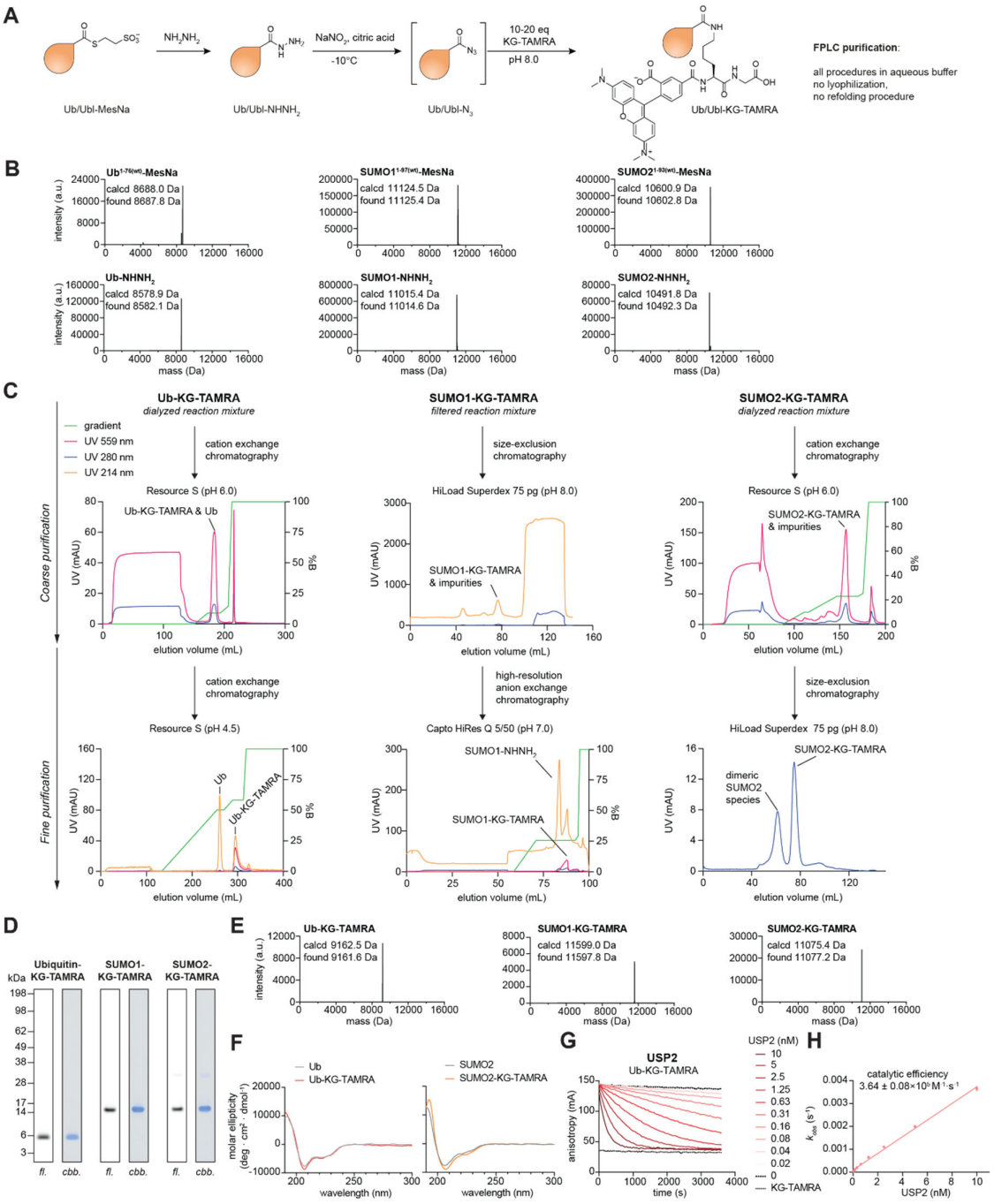


Fig 2.2. Preparation, characterization and quality assurance of fluorescence polarization substrates Ub-, SUMO1-, and SUMO2-KG-TAMRA. **A**. Workflow for synthesis of Ub/Ubl-KG-TAMRA. Ub/SUMO-MesNa thioesters were generated by the aforementioned intein chemistry. Active intermediates that cannot be purified is shown in brackets. **B**. Deconvoluted mass spectra of stable intermediates generated in **A**. All the Ub and SUMOs are wildtype. **C**. Chromatograms of two-step purification workflow for Ub/SUMO-KG-TAMRA by FPLC. B% indicates the percentage of elution buffer from inlet B on FPLC. All the chromatography columns are commercially available and used according to the instruction. **D**. Purity check by gel-based analysis of indicated substrates; *fl*, fluorescence; *cbb*, Coomassie brilliant blue-stained. **E**. Deconvoluted mass spectra of purified Ub/SUMO-KG-TAMRA. **F**. Circular dichroism spectra of Ub/SUMO2-KG-TAMRA and their parental proteins. **G**. Quality check of Ub-KG-TAMRA by DUB USP2 in the fluorescence polarization assay. **H**. Catalytic efficiency derived from the slope of k_{obs}/[USP2] plot is shown as insert.

2.2 Examining the Specificity of DeSUMOylase USPL1

The native semisynthesis protocol provides useful tools with high quality as evident from the CD spectra data for quantitative analysis and comparison of the substrate preference of USPL1 as well as SENP1 and yeast deSUMOylase ULP1 which both served as positive controls. USPL1 cleaved SUMO2 quickly but showed significantly less activity against SUMO1 and no any activity against ubiquitin-based substrate (Fig 2.3.A). SENP1 showed quite similar activity against SUMO1 and SUMO2 substrates and of course inactive on Ub-KG-TAMRA (Fig 2.3.B). Surprisingly, ULP1 showed totally different behaviors on the two SUMO-based reagents as ULP1 cleaved SUMO1 significantly faster than SUMO2 (Fig 2.3.C). Measurements at broader concentration ranges enabled the proper calculation of observed rate constants as well as the derived catalytic efficiencies of the three enzymes (Fig 2.3.D-F & Appendix Fig 1).

Since enzymes in the cells normally operate with limited concentration of substrates, catalytic efficiency can be used as a viable parameter for comparing how efficiently enzymes process their substrates. It is striking that USPL1 cleaved SUMO2-KG-TAMRA with an around 25-fold higher catalytic efficiency than SUMO1-KG-TAMRA which is far beyond the ratio (around 4-fold difference) reported in the fluorescence intensity assays (Fig 2.1.F). It is obvious that the specificity of USPL1 toward SUMO2/3 over SUMO1 was significantly underestimated based on the fluorescence intensity assays because of the artificial linkage in the substrates. These data firmly demonstrated the advantages of utilizing isopeptide-linked substrates for profiling the modifier specificity of DUBs and ULPs. As expected, SENP1 did not discriminate SUMO1 and SUMO2 significantly (Fig 2.3.E). Contrarily, ULP1 exhibited extreme preference toward SUMO1 over SUMO2 because SUMO1 shares higher similarity with yeast SUMO protein (Fig 2.3.F).

To make full use of the obtained fluorescence polarization substrates, catalytically inactive enzymes USPL1 and USP21 were purified and tested. USPL1 and USP21 both showed preferences toward SUMO2 and ubiquitin respectively (Fig 2.3.G). Catalytic mutant might have higher binding affinity for ubiquitin than wildtype[203] and thereby this assay can be used to estimate the binding affinity of DUBs and ubiquitin. Overall, the utilization of Ub/SUMO-KG-TAMRA reagents validated the quality which is from the mild reaction and purification methods. They can be used not only for distinguishing the modifier specificity, but also as fluorescent tracers to measure the binding affinity.

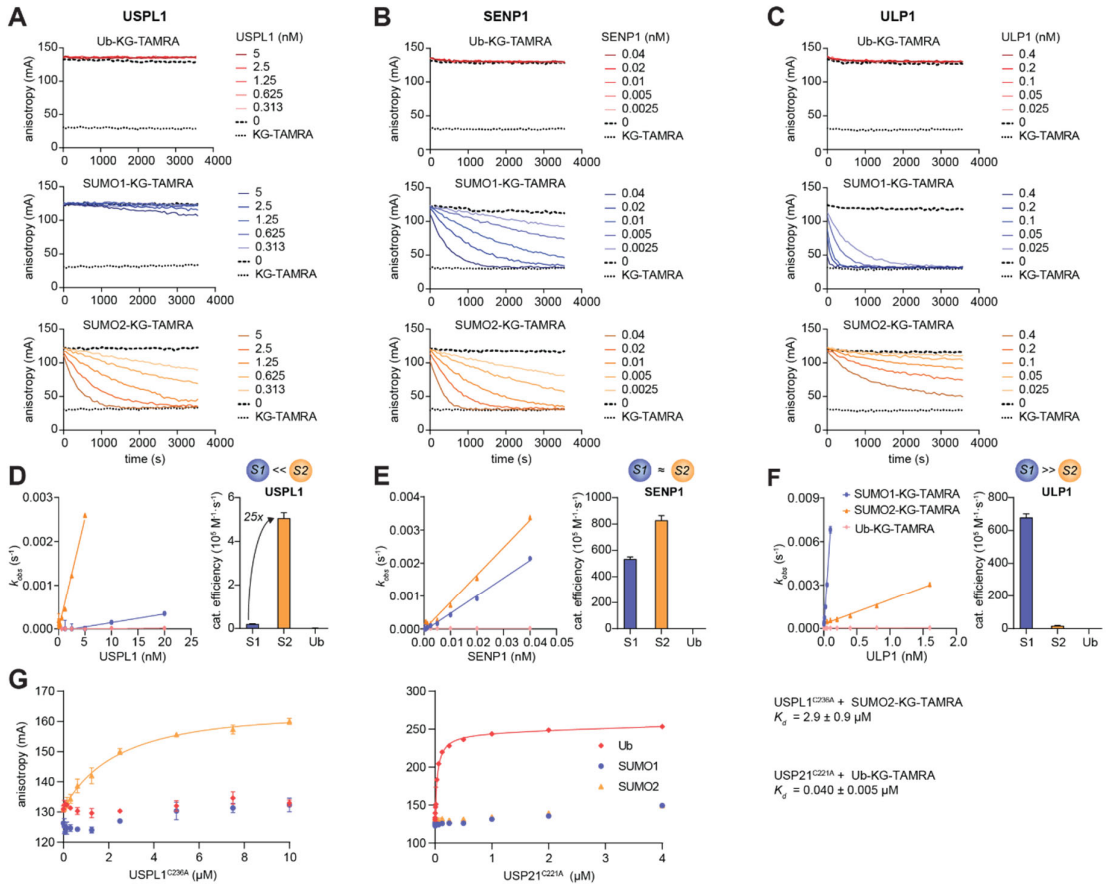


Fig 2.3. Assessment of the specificity toward SUMO paralogues. **A-C**. Human USPL1, human SENP1, yeast ULP1 were tested in fluorescence-polarization-based cleavage assays. Averages of technical triplicates are shown, which are representative of three independent experiments. **D-F**. Plots of observed rate constants over enzyme concentrations determined from assays shown in **A-C** and Appendix Fig 1. Catalytic efficiencies derived from the slopes of k_{obs}/[enzyme] plots (left) are shown as bar graphs as the mean ± standard error (right). G. Binding assay of inactive USPL1 and USP21 based on fluorescence polarization assays. The dissociation constant values were shown.

2.3 Structural Overview of USPL1 in Complex with SUMO2/3

The natural isopeptide-linked substrates exhibited superior applications so that provided quantitative insights into the specificity of USPL1 toward SUMO2/3 as well as an estimation of binding affinity of USPL1 with SUMO2. However, the reason leading to the uniqueness of USPL1 in USP family as well as the mechanism of modifier specificity remained unclear because the missing of structural information. Another unique member in the USP family, USP18 was captured and stabilized by its substrate ISG15-based covalent probe and then crystallized. According to the strategy of studying USP18 as well as the previous design of SUMO-based probes[204, 205], it would be informative to determine the complex structure of USPL1 with SUMO2/3.

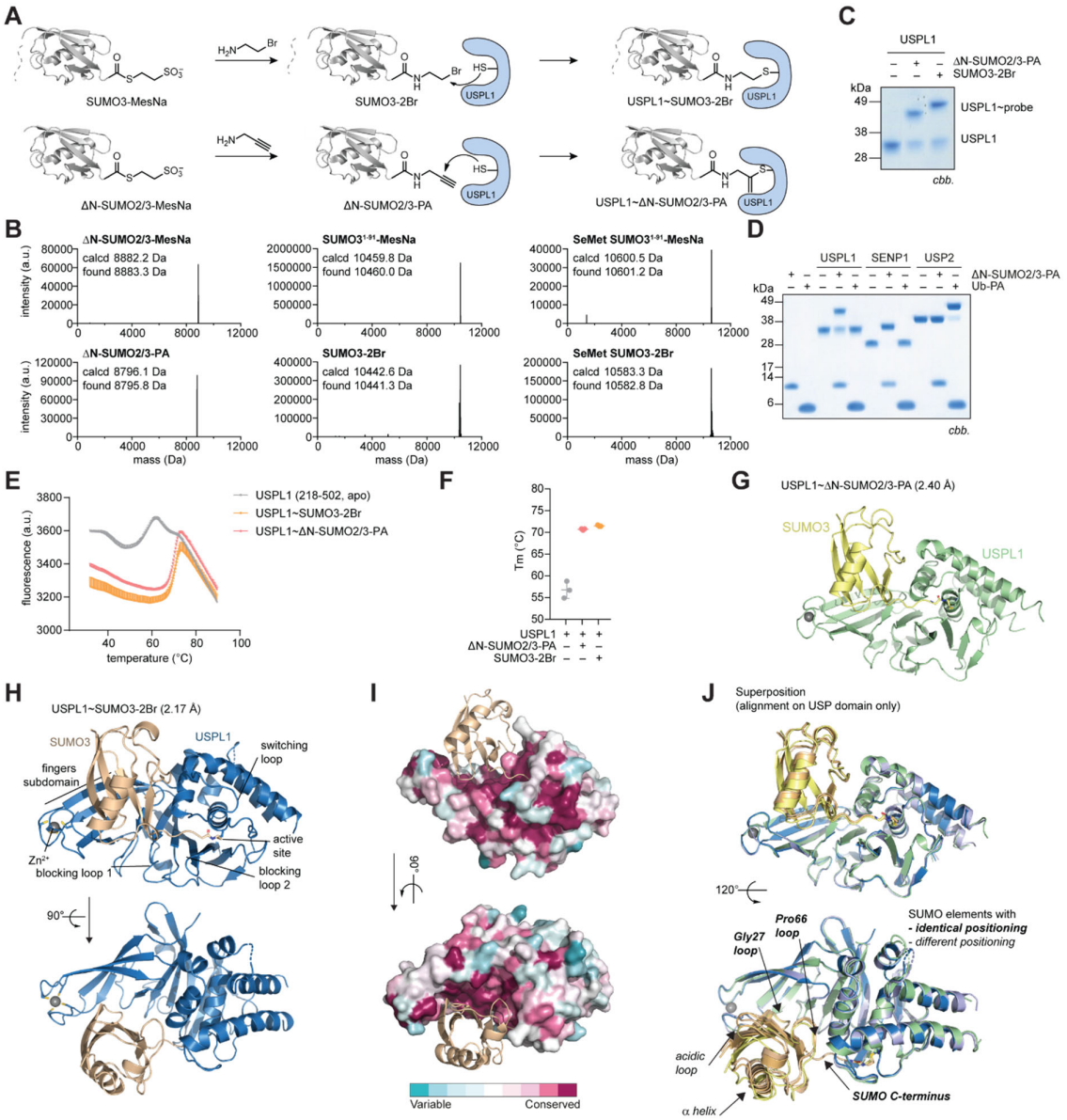


Fig 2.4. Structural determination of USPL1 in complex with substrates. **A**. Generation of substrate-trapped USPL1 complexes. The unstructured N-terminal sequence of SUMO is indicated by a dashed line. **B**. Characterization of SUMO thioesters and probes. **C**. USPL1 catalytic domain reacts with indicated probes. **D**. USPL1, SENP1 and USP2 reacted with SUMO2/3-PA or Ub-PA probes. **E-F**. Thermal shift assay of USPL1 and probes. **G**. Cartoon representation of USPL1~ΔN-SUMO2/3-PA. **H**. Cartoon representation of USPL1~SUMO3-2Br. **I**. Sequence conservation was mapped as a colored surface. **J**. Superposition of the independent geometries of USPL1~SUMO complexes with alignment on the USP domain. Regions of SUMO that interact with USPL1 and whose relative positioning toward USPL1 is consistent across crystal forms are highlighted in bold.

Since SUMO2 and SUMO3 share identical folds and differ only in the unstructured N-termini, a full-length SUMO3-2Br probe equipped with a 2-bromoethyl warhead and a truncated ΔN-SUMO2/3-PA probe armed with a propargylamine warhead were semisynthetically prepared using intein chemistry (Fig 2.4.A). To solve the phase problem in crystallography, selenomethionine-incorporated SUMO3 probe was obtained as well. All the thioesters and final products were validated by intact protein mass spectrometry (Fig 2.4.B). Unsurprisingly, both tactics

were successful as evident from a shift on gel due to the formation of covalent complexes (Fig 2.4.C). Intriguingly, ΔN-SUMO2/3-PA can only react with deSUMOylase instead of USP2 which implied its broader applications beyond structural biology as it is milder than the recently reported Dha-based probe (Fig 2.4.D)[206]. Both probes showed strong stabilization effect on USPL1 with almost 15°C increment on the melting temperature which was a positive indication for crystallization (Fig 2.4.E&F).

After several rounds of optimization, crystal structures of USPL1 in covalent complex with ΔN-SUMO2/3-PA to 2.4 Å resolution, as well as of USPL1 in covalent complex with SUMO3-2Br to 2.17 Å resolution in different crystal forms were solved by single-wavelength anomalous dispersion measurements with the assistance of a selenomethionine-containing SUMO3-2Br probe (Fig 2.4.G&H and Appendix Table 1). The structure of USPL1~SUMO3-2Br has two almost identical copies in the asymmetric unit while the other structure contains only one copy (Fig 2.4.J & Fig 2.5.A). Both structures displayed the typical USP-fold with well-defined electron density of all regions and thereby allow unambiguous analysis and interpretation of the geometric arrangement (Fig 2.5.B-E). In the catalytic center, all the three residues of catalytic triad (Cys263, His456, and Asp472) were explicitly observed with defined electron density (Fig 2.5.F). Mutagenesis studies based on SUMO2-RhoG cleavage assay as well as gel-based assay confirmed that mutations of these residues are able to abolish the SUMO2 cleavage (Fig 2.5.G&H). However, mutants H456A and D472A still showed very weak binding to SUMO3-2Br probe because the high reactivity of the nucleophilic-substitution-based mechanism (Fig 2.5.H).

USPL1 was phylogenetically assigned into the USP family since it owns a typical USP-fold structure. The conservation analysis based on the sequences of 183 orthologs which were annotated in the Ensembl database highlighted that the highly conserved regions appear not only in the surrounding residues of catalytic center, but also in the fingers domain (Fig 2.4.I) which might provide the decisive factor for the specific recognition of SUMO2/3 by USPL1 even though it has a USP scaffold. Structural superposition of the three copies from the two structures revealed consistent positioning of SUMO in these highly conserved areas including the SUMO C-terminus, the hydrophobic Pro66 loop, and the Gly27 loop interacting with the USPL1 fingers domain (Fig 2.4.J). They are two major interacting sites between USPL1 and SUMO2/3 besides the interaction formed by SUMO2/3 tail. How USPL1 interacts with the Gly27 loop using its finger domain which is highly conserved across 183 species, and with the Pro66 loop might contribute to the uniqueness of USPL1 in the USP family.

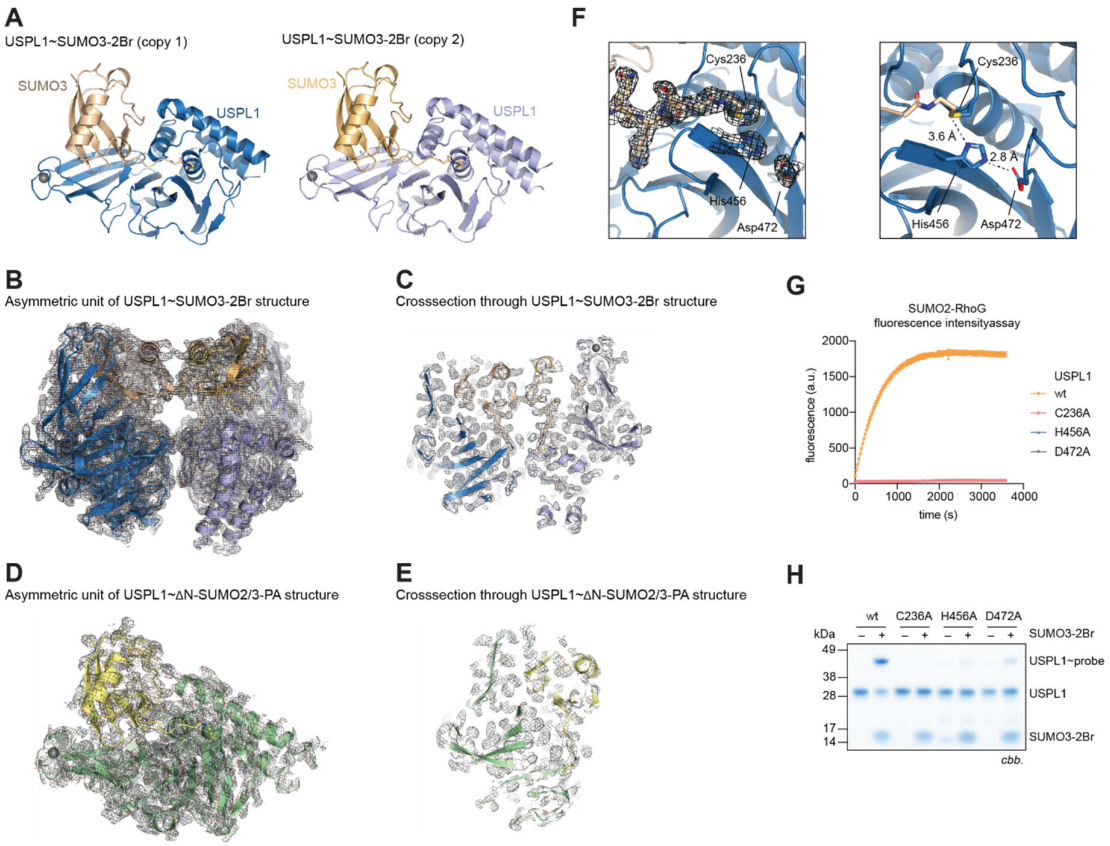


Fig 2.5. Crystal structures of USPL1 in covalent complexes with SUMO2/3. **A**. Cartoon representation of both copies in the asymmetric units of the USPL1~SUMO3-2Br structure. Zinc atoms are shown as grey spheres. **B**. Electron density map of the asymmetric unit of the USPL1~SUMO3-2Br structure shown as the weighted $2|F_O|-|F_C|$ density contoured at $\sigma = 1.0$. **C**. Cross-section through the density shown in **B**, centered on the SUMO C-terminal residues. **D**. Electron density map of the asymmetric unit of the USPL1~ΔN-SUMO2/3-PA structure. **E**. Cross-section through the density shown in **D**, centered on the SUMO C-terminal residues. **F**. Close-up view of the active site, showing alignment of the three residues of the catalytic triad. Hydrogen bonds are indicated with dashed lines, and atomic distances are measured. **G**. SUMO2-RhoG cleavage assay with wild-type USPL1 or indicated mutants of the catalytic triad. An average of three technical replicates is shown ([USPL1] = 0.25 nM, [SUMO2-RhoG] = 50 nM). **H**. Coomassie-stained SDS-PAGE gel of indicated USPL1 proteins, pre-incubated with SUMO3-2Br probe where shown.

2.4 Mechanism of SUMO2/3 Specificity in USPL1

To investigate the specificity mechanism of USPL1 based on the structural information, key residues can be mutated to confirm the importance using SUMO2-RhoG or gel-based cleavage assays. Sequence alignment of ubiquitin, NEDD8, ISG15 and SUMOs in combination with structural information rationalized the design of mutations (Fig 2.6.A&B). In general, key residues involved in USPL1 recognition of SUMO2/3 were mutated into the equivalent residues in SUMO1 or ubiquitin. The SUMO2 mutants were purified and assembled as isopeptide-linked conjugates with the model protein RanGAP1 using enzymatic method[146]. The RanGAP1~SUMO2 cleavage assay was optimized and visualized on gel (Fig 2.6.C). The wildtype

and mutants were tested at two time points (10 min for partial cleavage and 30 min for almost complete turnover). It is obvious that residues Pro66 and Asp71 both play significant roles in distinguishing SUMO2/3 and ubiquitin as the P66Q and D71R mutants were cleaved by USPL1 with a much lower velocity. However, mutations P66R and D71H that transform SUMO2 more SUMO1-like did not show negative effects on USPL1 cleavage as they were cleaved to the extent as the wildtype substrate (Fig 2.6.D).

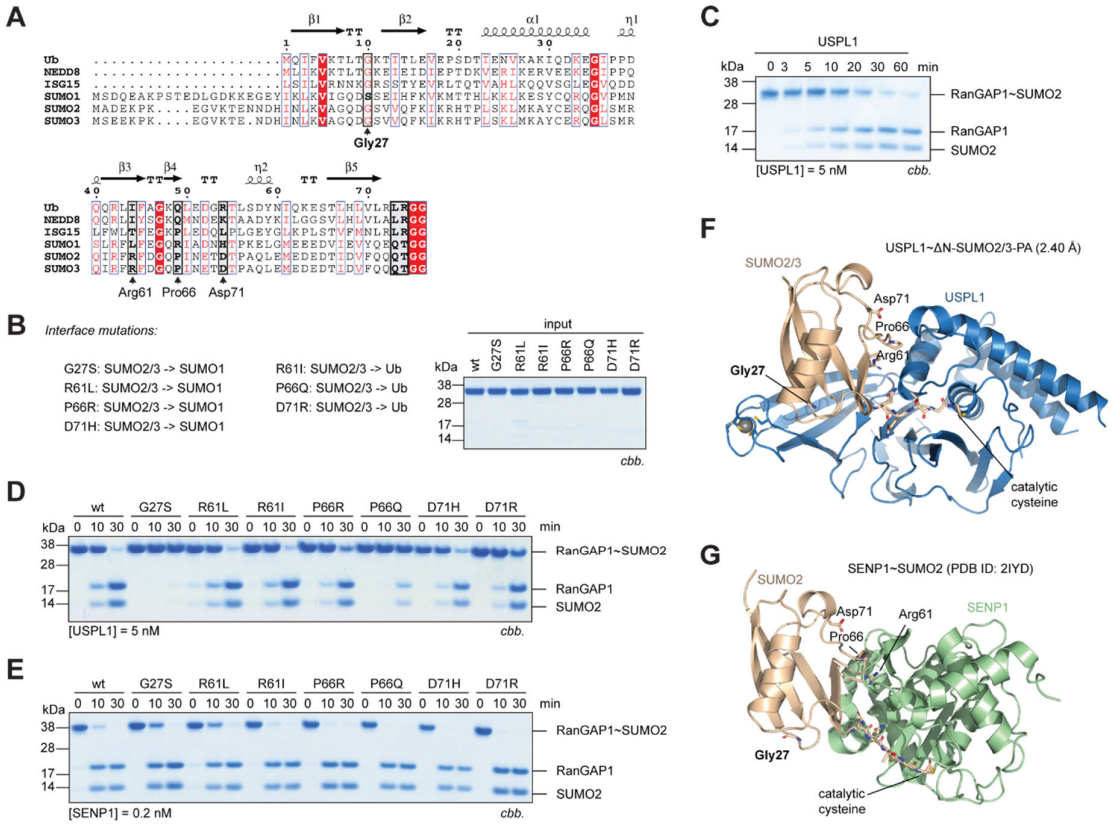


Fig 2.6. Investigation into the key residues of SUMO2/3 for USPL1 recognition. **A**. Sequence alignment of the human Ub/Ubl. SUMO2 and SUMO3 differ in the disordered N-terminal region; folded parts are identical yet differ by one in amino acid numbering. Residue numbering is based on the SUMO2 sequence throughout this work. **B**. Key residues of SUMO2/3 at the interface with USPL1 were selected for mutations into the corresponding amino acids in Ub or SUMO1. RanGAP1~SUMO2 was assembled with mutations on SUMO2 and visualized on gel. **C**. Coomassie-stained SDS-PAGE gel of the isopeptide-linked RanGAP1~SUMO2 cleavage assay with USPL1 for indicated time points for assay optimization. **D**. RanGAP1~SUMO2 cleavage assay with USPL1. **E**. RanGAP1~SUMO2 cleavage assay with SENP1. **F**. Structure of the USPL1~ΔN-SUMO2/3-PA. Key residues interacting with USPL1 are highlighted in sticks **G**. Structure of the SENP1~SUMO2 covalent complex (PDB ID: 2IYD). Key residues interacting with USPL1 are highlighted in sticks.

Interestingly, R61L and R61I mutants were cleaved as efficiently as the wildtype. Ile44 patch is a typical hydrophobic region in ubiquitin for recognition, but equivalent Arg61 in SUMO2 is not the determinant for specific recognition by USPL1. Strikingly, when the G27S mutant, which was supposed to mimic SUMO1, was incubated with USPL1, the substrate acquired resistance to USPL1. This firmly demonstrated that Gly27 loop in the SUMO2/3 is the

determinant part for the specificity of USPL1 and thus explained the catalytic efficiency of USPL1 on SUMO2 is 25-fold higher than SUMO1. In contrast to USPL1, SENP1 cleaved all the RanGAP1~SUMO2 substrates efficiently even at very low enzyme concentration (Fig 2.6.E). In-depth analysis and comparison of the structures of USPL1~ΔN-SUMO2/3-PA and SENP1~SUMO2 (Fig 2.6.F&G) revealed that the USP-fold of USPL1 possesses multiple interacting regions. SENP1 binds only to the SUMO2 C-terminus which is sufficient to cleave the isopeptide bonds indiscriminately and these data were in line with the SUMO1/2-KG-TAMRA cleavage assays (Fig 2.3.E).

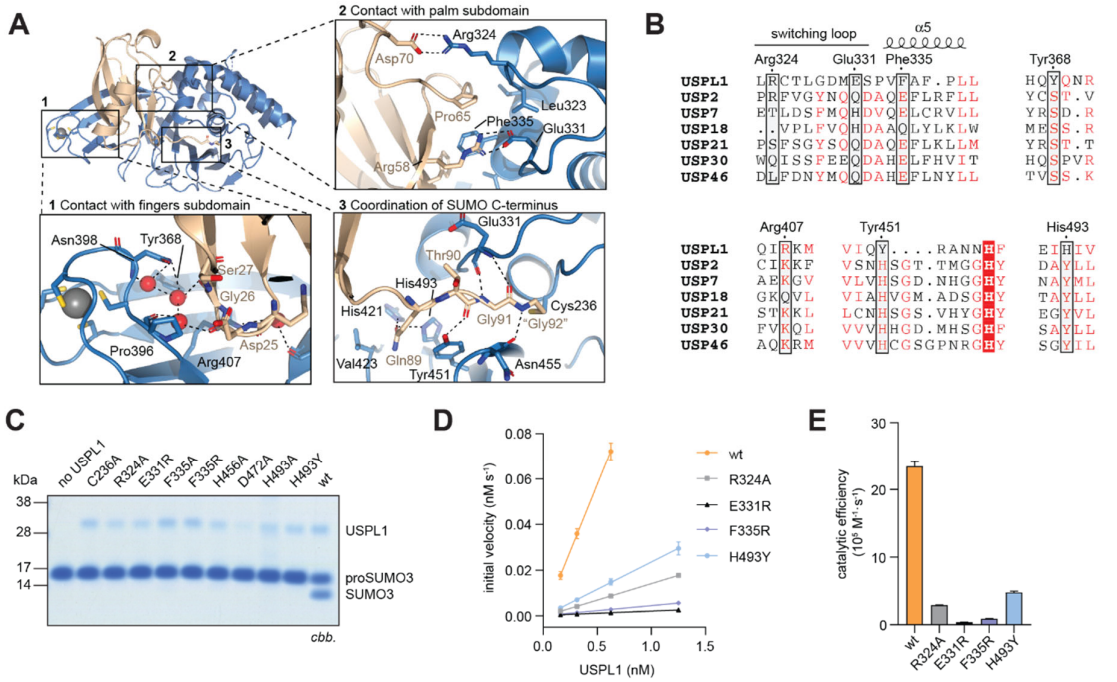


Fig 2.7. Intrinsic features of SUMO2/3 specificity of USPL1. **A**. Close-up view on three areas mediating SUMO2/3 recognition by USPL1, numbered according to small inserts, of the USPL1~SUMO3-2Br structure. Residues (sticks) and water molecules (red spheres) involved in the interaction between USPL1 and SUMO3 are highlighted. Hydrogen bonds are shown as dashed lines. Zinc atoms were shown in grey sphere. **B**. Protein sequence alignment of USPL1 with other representative human USPs. Key residues from areas important for SUMO interaction are highlighted in boxes. **C**. Coomassie-stained gel of proSUMO3 incubated with wild-type USPL1 or indicated mutants. **D**. SUMO2-RhoG cleavage assay using wild-type USPL1 or indicated mutants with different concentrations (N=3). **E**. Catalytic efficiencies calculated from **D** are shown as mean ± standard error.

Through introducing key mutations into SUMO2, the reason that USPL1 prefers SUMO2 instead of SUMO1 or ubiquitin has been clarified. However, the intrinsic features that determine the USPL1 specificity evolutionarily different from other USP members remained unclear. Closely examining the interface in the USPL1~SUMO3-2Br structure disclosed that some key interactions in three major contacting surface (Fig 2.7.A). In the contact with finger domain, multiple water molecules mediated several hydrogen bonds between the Gly27 (Gly26 in SUMO3) loop. In the palm domain, two salt bridges were formed which contributed strongly

to the binding affinity. Interestingly, a hydrophobic residue Phe335 formed hydrophobic interaction with Pro65 (Pro66 in SUMO2) which resembled a sandwich-like interaction together with the salt bridges. Moreover, several residues of USPL1 coordinated the backbone of C-terminus to enhance the binding during cleavage. Sequence alignment of USPL1 and other representative USPs uncovered several distinctive residues in USPL1 which were highlighted in boxes (Fig 2.7.B). For example, Arg324 which formed salt bridge with Asp70 of SUMO3 is not highly conserved while the residues correlated to Phe335 in other USPs are mostly hydrophilic. Systematic comparison of the alignment rationalized a series of mutants on USPL1 which were firstly tested by a proSUMO3 cleavage assay (Fig 2.7.C). It is obvious that all mutants were not able to cleave proSUMO3. In the following SUMO2-RhoG cleavage assay, E331R and F335R almost completely abolished the enzymatic activities (Fig 2.7.D&E). The introduction of hydrophilic residue into position 331 disrupted the hydrophobic interaction with Pro65 which is also a key residue to distinguish SUMO2/3 and ubiquitin. E331R might interfere with the SUMO C-terminus due to the altered charge to abolish the activity.

Collectively, the specific recognition of SUMO2/3 by USPL1 was visualized through two high-resolution crystal structure of USPL1 in covalent complex with SUMO2/3 probes. Mutations that transformed SUMO2 into multiple Ub/SUMO1-like modifiers which were further incorporated to the isopeptide-linked conjugates. The time-dependent cleavage assays provided an intuitive comparison of the contribution of each residue to the USPL1 specificity. Gly27 loop as the anchor point for hosting SUMO2/3 and distinguishing SUMO1 was proved by the introduction of serine residue which interrupted the interaction of Gly27 loop with USPL1 finger domain which was analyzed as a conserved area (Fig 2.4.I). Mutation of Pro66 into a glutamine which exists in ubiquitin led to the significant resistance to USPL1 cleavage. Reciprocally, Pro66 interacts with Phe335 in USPL1 through a hydrophobic interaction. The mutation of Phe335 into a positively charged arginine residue almost completely attenuated the SUMO cleavaging activity. These data showed that Phe335 is a residue which is not only important for differentiating SUMO2/3 and ubiquitin evolutionarily (Fig 2.7.B) but also the fundamental for maintaining its catalytic activity.

Successful applications of the native semisynthesis and bespoke purification workflow resulted in three fluorescence polarization substrates for characterizing understudied USPL1 in a quantitative manner. In combination with structural biology approaches, the underlying specificity mechanism of USPL1 was elaborated.

2.5 Examining UCHL3 and USP18

Native semisynthesis and subsequent customized purification workflow led to the development of three fluorescence polarization substrates for quantitatively characterizing the catalytic specificity of deSUMOylase USPL1. To further expand the substrate panel for broadly studying other DUBs and ULPs, NEDD8 thioester and hydrazide were prepared using the aforementioned protocol and characterized by mass spectrometry (Fig 2.8.A). Since the C-terminal domain (CTD) of ISG15 is sufficient for recognition[204], ISG15(CTD) intermediates carrying a cysteine to serine mutation which was intended for simplicity were obtained and characterized (Fig 2.8.A).

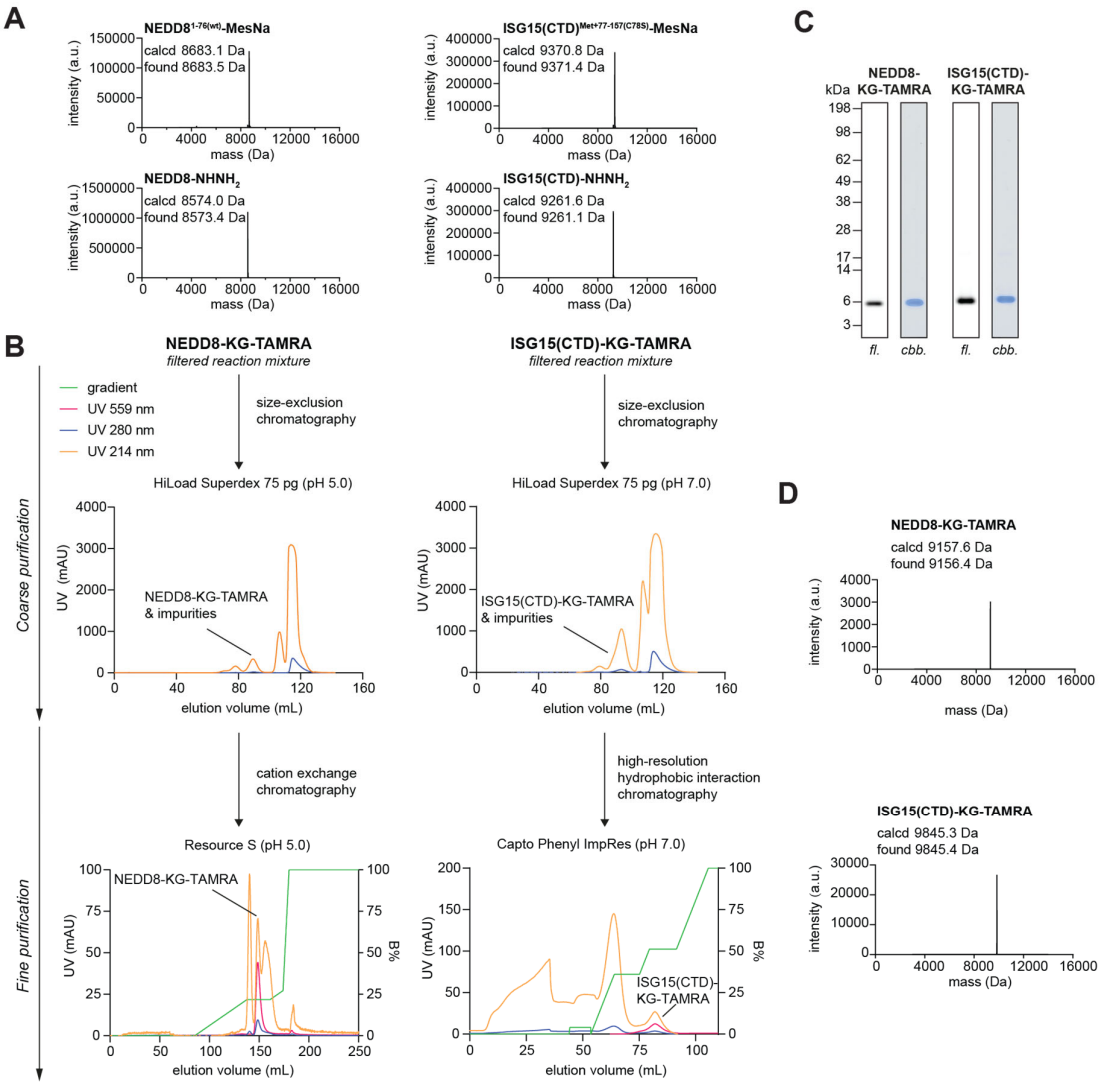


Fig 2.8. Preparation and characterization of NEDD8- and ISG15(CTD)-KG-TAMRA. **A**. Characterization of NEDD8 and ISG15(CTD) thioesters and hydrazides by intact mass spectrometry. **B**. Two-step purification workflow for NEDD8- and ISG15(CTD)-KG-TAMRA. **C**. Gel-analysis of NEDD8- and ISG15(CTD)-KG-TAMRA. **D**. Deconvoluted mass spectra of NEDD8- and ISG15(CTD)-KG-TAMRA.

The reaction mixture of NEDD8 was initially filtered and then subjected to a size-exclusion chromatography for an isocratic elution at pH 5.0 to separate protein from the excessive KG-TAMRA molecule (Fig 2.8.B). At the same pH, a cation exchange chromatography was applied to finely purify NEDD8-KG-TAMRA with a gradient elution. The reaction mixture of ISG15(CTD) was also filtered to remove any insoluble components and purified firstly by size-exclusion chromatography at pH 7.0. According to previous experience, the protein-containing mixture was finely purified on both anion and cation exchange column at various pH but the ISG(CTD)-KG-TAMRA cannot be obtained in pure form. Given that installation of TAMRA might increase the hydrophobicity of the whole molecule, a high-resolution hydrophobic interaction column (Capto Phenyl ImpRes) was tested in a gradient elution. Owing to the difference of hydrophobicity, pure ISG15(CTD)-KG-TAMRA was separated from other impurities (Fig 2.8.B). In-gel analysis (Fig 2.8.C) and intact protein mass spectrometry (Fig 2.8.D) both proved the purities of obtained NEDD8 and ISG15 reagents.

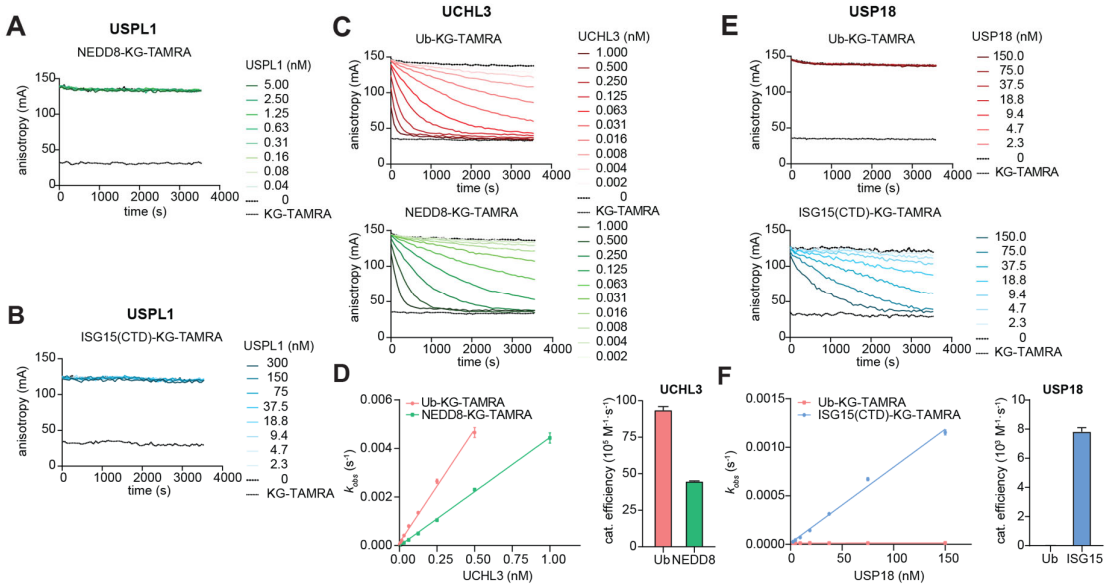


Fig 2.9. Application of NEDD8- and ISG15(CTD)-KG-TAMRA. **A**. Fluorescence polarization assay of USPL1 sing NEDD8-KG-TAMRA. **B**. Fluorescence polarization assay of USPL1 using ISG15(CTD)-KG-TAMRA. **C**. Fluorescence polarization assay of UCHL3 using Ub- and NEDD8-KG-TAMRA. **D**. Plots of observed rate constants over UCHL3 concentration to calculate the catalytic efficiencies which are shown as bar graphs. **E**. Fluorescence polarization assay of human USP18 using Ub- and ISG15(CTD)-KG-TAMRA. **F**. Plots of observed rate constants over USP18 concentration to calculate the catalytic efficiencies which are shown as bar graphs.

The NEDD8 and ISG15 fluorescence polarization substrates were respectively tested by USPL1 which was used as a negative control. As expected, USPL1 did not show any cleavage on NEDD8 and ISG15 at any concentration (Fig 2.9.A&B). NEDD8-KG-TAMRA was then used for revisiting the catalytic preference of UCHL3 because it was previously reported as a deNEDDylase with around 1000-fold lower activity compared to its deubiquitinating activity

in the fluorescence intensity assay[141]. Strikingly, UCHL3 only showed approximate 3-fold difference in the isopeptide-bond-based assay (Fig 2.9.C&D). This highlighted the superiority of using isopeptide-containing substrates for characterizing the modifier specificity since the obtained data is in consistent with the pull-down assay[132]. Mouse USP18 has been tested in the fluorescence polarization assay but human USP18 has not been investigated[204]. Human USP18 showed extremely good preference toward ISG15 (Fig 2.9.E). Further analysis of its catalytic efficiencies provided the preference of human USP18 quantitatively for the first time (Fig 2.9.F).

With the aid of fluorescence polarization substrates, modifier specificity of UCHL3 was revised which emphasized the previously underestimated NEDD8 cleavage function of UCHL3. ISG15-based substrate enabled the quantification of the deISGylase activity of human USP18. Since human USP18 exhibited limited ISG15 processing activity, its function as a scaffold protein in cellular environment should be further investigated with the input from the *in vitro* biochemical data[207].

2.6 Discovery of Triple Cross-reactivities in USP16 and USP36

Successful applications of the aforementioned five fluorescence polarization have provided broader understanding of the modifier specificity of several DUBs and ULPs. To further extend the available isopeptide-based toolbox for investigating the modifier specificity and discovering unprecedently known cross-reactivities in DUBs/ULPs, FUBI was selected as a case study since its proteases had been reported recently but their isopeptidase activities had not been quantified[66].

Mature FUBI protein has a cysteine residue in its position 57. To avoid the potential nitrosylation of Cys57 for easier reaction control, this residue was mutated to an alanine residue. FUBI thioester and hydrazide were sequentially prepared using the standard intein chemistry and then aminolysis with excessive hydrazine, and finally characterized by intact protein mass spectrometry (Fig 2.10.A). FUBI hydrazide was concentrated and functionalized by KG-TAMRA by the activation of nitrous acid on ice. The filtered reaction mixture was crudely purified on a size-exclusion chromatography. The pooled protein-containing fractions were purified by a high-resolution anion exchange column (Fig 2.10.B). The purified FUBI-KG-TAMRA was analyzed on gel (Fig 2.10.C) and characterized by intact protein mass spectrometry to confirm its purity (Fig 2.10.D).

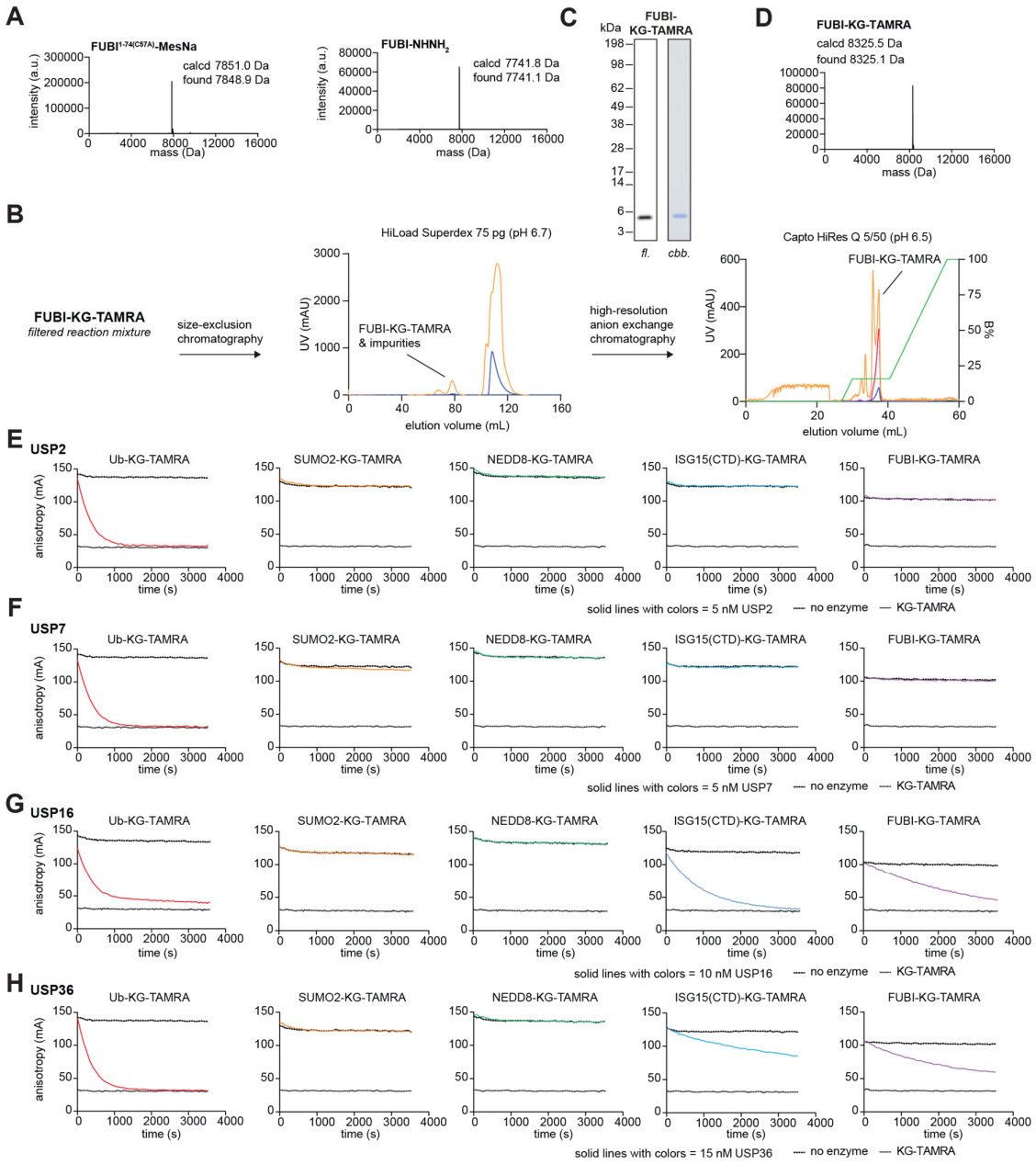


Fig 2.10. Preparation of FUBI-KG-TAMRA and its application for screening cross-reactivities of DUBs/ULPs. **A**. Intact mass spectra of FUBI thioester and hydrazide. **B**. The purification workflow tailored for FUBI-KG-TAMRA in two steps by Dr. Rachel O'Dea. **C**. Gel analysis of the purified FUBI-KG-TAMRA. **D**. Deconvoluted mass spectrum of purified FUBI-KG-TAMRA. **E**. Fluorescence polarization assays for USP2. **F**. Fluorescence polarization assays for USP7. **G**. Fluorescence polarization assays for USP16. **H**. Fluorescence polarization assays for USP36.

Following the purification of FUBI-KG-TAMRA, a panel of fluorescence polarization substrates was thereby assembled. SUMO1-KG-TAMRA was excluded for easily handing the assay and high similarity to SUMO2 reagent. This panel was firstly examined by USP2 at single concentration (Fig 2.10.E). USP2 showed extraordinary specificity toward ubiquitin instead of ISG15 which clarified the misleading results generated from hyperreactive probes[152].

Full-length USP7 was also selected as a case study and showed strong catalytic effect on ubiquitin (Fig 2.10.F). Interestingly, USP7 slightly cleaved SUMO2-KG-TAMRA which implied its potential role in regulating SUMO modifications. However, this required further exploration to provide solid evidence since the single-concentration assay cannot give a quantification of the catalytic activities.

The substrate panel was then employed to examine the cross-reactivities of USP16 and USP36 which are both reported to be capable of cleaving ubiquitin and FUBI. In the fluorescence polarization assay, USP16 and USP36 both cleaved ubiquitin and FUBI efficiently (Fig 2.10.G&H). Surprisingly, they can also efficiently cleave ISG15 which is evolutionarily close to FUBI because of the hydrophobic proline residues. The single-concentration assay led to the discovery of the first tri-specific isopeptidases targeting ubiquitin, ISG15, and FUBI. To perform quantitative comparison of the modifier specificity of USP16, the fluorescence polarization assays were carried out with a broad range of enzyme concentration (Fig 2.11.A). FUBI-KG-TAMRA was fully cleaved to reach the same level of free KG-TAMRA which confirmed its quality. Plots of the observed rate constants over enzyme concentration resulted in the quantification of catalytic efficiencies toward the tested substrates (Fig 2.11.B). Similarly, USP36 was step-wise diluted and evaluated in the fluorescence polarization assay (Fig 2.11.C). Quantification of catalytic efficiencies of USP36 revealed different preference toward ISG15 and FUBI in comparison to USP16 (Fig 2.11.D). USP16 preferred to cleave ISG15 over FUBI while USP36 displayed similar deubiquitinating activity and moderate activity on FUBI, but very weak activity on ISG15 substrate.

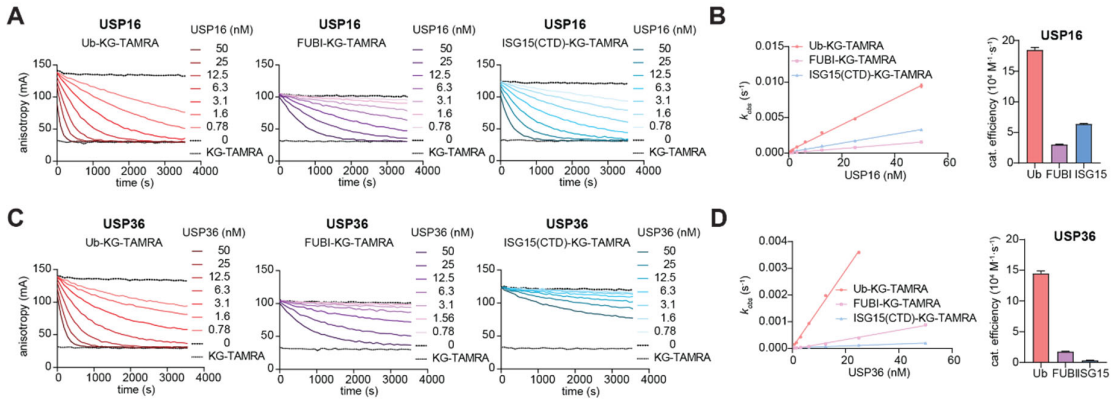


Fig 2.11. Quantitative analysis of the triple specificity of USP16 and USP36 toward ubiquitin, FUBI and ISG15. **A**. Fluorescence polarization assay of USP16 using Ub-, FUBI-, and ISG15(CTD)-KG-TAMRA. **B**. Quantification of the catalytic efficiencies of USP16 toward ubiquitin, FUBI, and ISG15. The catalytic efficiencies are shown as bar graphs. **C**. Fluorescence polarization assay of USP36 using Ub-, FUBI-, and ISG15(CTD)-KG-TAMRA. **D**. Quantification of the catalytic efficiencies of USP36 toward ubiquitin, FUBI, and ISG15. The catalytic efficiencies are shown as bar graphs.

Although USP16 and USP36 both possess tri-specific activities toward ubiquitin, FUBI and ISG15, the preference profile of USP16 and USP36 differ significantly which might contribute to a variety of cellular events since they are localized in different cellular compartment. The biochemical data unraveled their preference quantitatively which shed light on their multiple functions of processing modifiers and provided the hints for mechanistic studies of their cellular functions in the future. To elaborate explicitly their functions of processing FUBI, a two-tier model was proposed to explain how USP16 and USP36 coordinate the FUBI cleavage[66]. However, how they behave as ISG15 proteases remain unclear[208]. Since the deISGylase activities were discovered by biochemical assays, cellular data are needed to interrogate how USP16 and USP36 implement their tri-specific functions.

In summary, the native semisynthesis and customized purification methods enabled by FPLC maintained the natural conformation of Ub/Ubls after the functionalization of their C-termini with an isopeptide-linked fluorophore. The fully folded fluorescence polarization substrates established an assay platform for characterizing, revising and quantifying the cross-reactivities of DUBs/ULPs. The substrates can also be adapted to a binding assay instead of the format of cleavage-based kinetic assay if the catalytically inactive enzymes were incubated which provided the binding affinity.

The preparation of different substrates was carefully explored and eventually resulted in the assembly of a panel of substrates for studying cross-reactivities of DUBs/ULPs (Fig 2.12). The substrate panel quantified the catalytic specificity of USPL1 whose mechanism was uncovered by following structural studies and biochemical validation. The preference of UCHL3 toward ubiquitin and NEDD8 was revised to provide the biochemical evidence for future mechanistic studies. The specificity toward ISG15 of human USP18 was confirmed quantitatively for the first time. The panel was a powerful platform leading to the unprecedented discovery of the first tri-specific DUBs/ULPs USP16 and USP36.

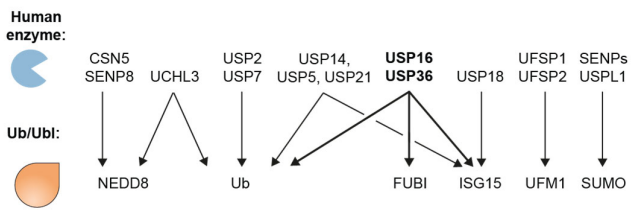


Fig 2.12. Cross-reactivities of human DUBs and ULPs. Modifier preference is indicated with arrows. Tri-specific isopeptidases USP16 and USP36 are highlighted in bold. Enzymes studied in this work with a panel of substrates include USP2, USP7, USP16, and USP36. UCHL3, USP18, SENP1, USP21, and USPL1 were studied using subsets of substrates.

2.7 High-throughput Screening Campaign for USPL1

Research on DUBs and ULPs over the past years has been advanced significantly but the lack of specific inhibitors of high-quality hinders the interrogation of the functions of these enzymes. Since most of DUBs and ULPs are cysteine-centric isopeptidases, many promiscuous compounds were discovered due to their high reactivities toward the catalytic cysteine residue. Moreover, the poor assay quality led to identify invalid compounds. A notorious example is the compound LDN-57444 which was discovered as a UCHL1 specific inhibitor in a Ub-AMC assay[209]. After it has been widely used for nearly two decades, a study shown that LDN-57444 cannot engage its target in cells and showed no activities against UCHL1 in various assays[210]. Moreover, in comparison to the coumarin-based substrate, Rhodamine possesses superior optical advantages to reduce the interference from the assayed compounds and thereby Ub/Ubl-RhoG is more suitable for setting up the high-throughput screen.

USPL1 as a unique member in the USP family requires specific tool compounds to distinguish its catalysis-related functions from scaffold-based functions for mechanistic studies[211]. Since a large quantity of SUMO2-RhoG was prepared in native conditions, a high-throughput screening campaign can be launched efficiently.

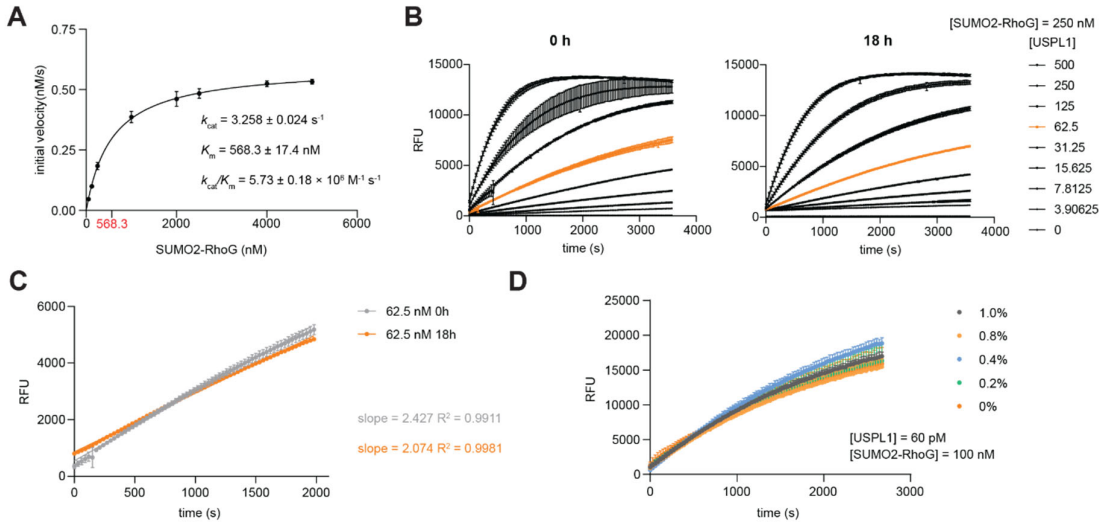


Fig 2.13. Assay development for high-throughput screening campaign to discover novel inhibitors targeting USPL1. **A**. Enzymatic characterization of USPL1 using SUMO2-RhoG as the substrate. The initial velocities at different substrate concentrations were plotted over the substrate concentrations, the USPL1 concentration was at 183 pM. The derived parameters were shown as inserts. **B**. Titration of a broad concentration range of USPL1 with a fixed concentration of SUMO2-RhoG at 250 nM. USPL1 was tested at 0 h as well as kept on ice for 18 h. Data were shown as mean ± standard error. **C**. The stability of USPL1 was compared at 62.5 pM from **B**. The slopes from linear regression were shown. **D**. DMSO tolerance of USPL1 was tested at various concentration. High-throughput screening campaign was executed by Dr. Sonja Sievers and Dr. Philipp Lampe.

In order to establish the fluorescence-intensity-based assay for discovering USPL1 inhibitors, the characterization of USPL1 is an essential step. Titration of SUMO2-RhoG was performed with a fixed concentration of USPL1 at 183 pM to give the Michaelis-Menten parameters of USPL1 (Fig 2.13.A). Given that substrate concentration will have different effects on discovery of inhibitors with different mechanisms, substrate concentration was set to 250 nM, which is equivalent approximately to half value of Michaelis-Menten constant (Fig 2.13.B). A wide range of USPL1 concentration were tested to optimize the assay. Furthermore, the stability of USPL1 was tested as well upon storage on ice for 18 hours. When USPL1 was used below 62.5 pM, the assay readout showed good linearity (Fig 2.13.B). The enzymatic activities of USPL1 can be retained even after 18 hours which was endorsed by the consistent results from the linear regression of the hydrolysis of SUMO2-RhoG by USPL1(62.5 pM) (Fig 2.13.C). Therefore, the enzyme which has good robustness can be stored on ice for overnight in case the high-throughput screening was interrupted. Since the compounds were stored using DMSO as solvent, up to 1% DMSO concentration was evaluated in the SUMO2-RhoG cleavage assay (Fig 2.13.D). USPL1 showed good tolerance of DMSO which indicated the assay was successfully established.

The high-throughput screening was performed by the team at the Compound Management and Screening Center (COMAS) of the Max Planck Institute of Molecular Physiology. Hit compounds that showed more than 50% inhibitory activities and were not active in the counter screen were tested in multiple concentrations to give the IC_{50} values. Two hit compounds (COMAS ID: 183051 and 130646) were identified for further evaluation (Fig 2.14.A). These two compounds share the same core consisting of an *N*-benzoyl pyrazole scaffold which might be attacked by the USPL1 catalytic cysteine. To test if these two compounds are covalent inhibitors, USPL1 was pre-incubated with these compounds for three different time points respectively. The IC_{50} values were shifted in proportion to the pre-incubation time (Fig 2.14.B). Compound 183051 showed stronger inhibition against USPL1 which might be due to the additional methyl group. The time-dependent inhibition of USPL1 implied that compound 183051 might covalently bind to USPL1 through an addition-elimination mechanism. The catalytic cysteine of USPL1 firstly attacked the carbonyl of compound 183051 to form a hemithioacetal intermediate which was further eliminated to give the covalent complex (Fig 2.14.C). To validate this hypothesis, intact protein mass spectra were recorded for indicated incubation time. Protein mass spectrometry data revealed that USPL1 forms a covalent thioester complex with the compound 183051 in a time-dependent manner (Fig 2.14.D). After incubation for 2 hours, USPL1

was completely labelled by the covalent inhibitor. Although the experimental data showed that compound 183051 inhibited USPL1 covalently, the structural information of how it binds to USPL1 is still missing. To test if the covalent complex can be set up for crystallization, thermal shift assay using different reducing agents was performed. Interestingly, the covalent inhibitor did not stabilize USPL1 (Fig 2.14.E). The mechanism should be studied more carefully.

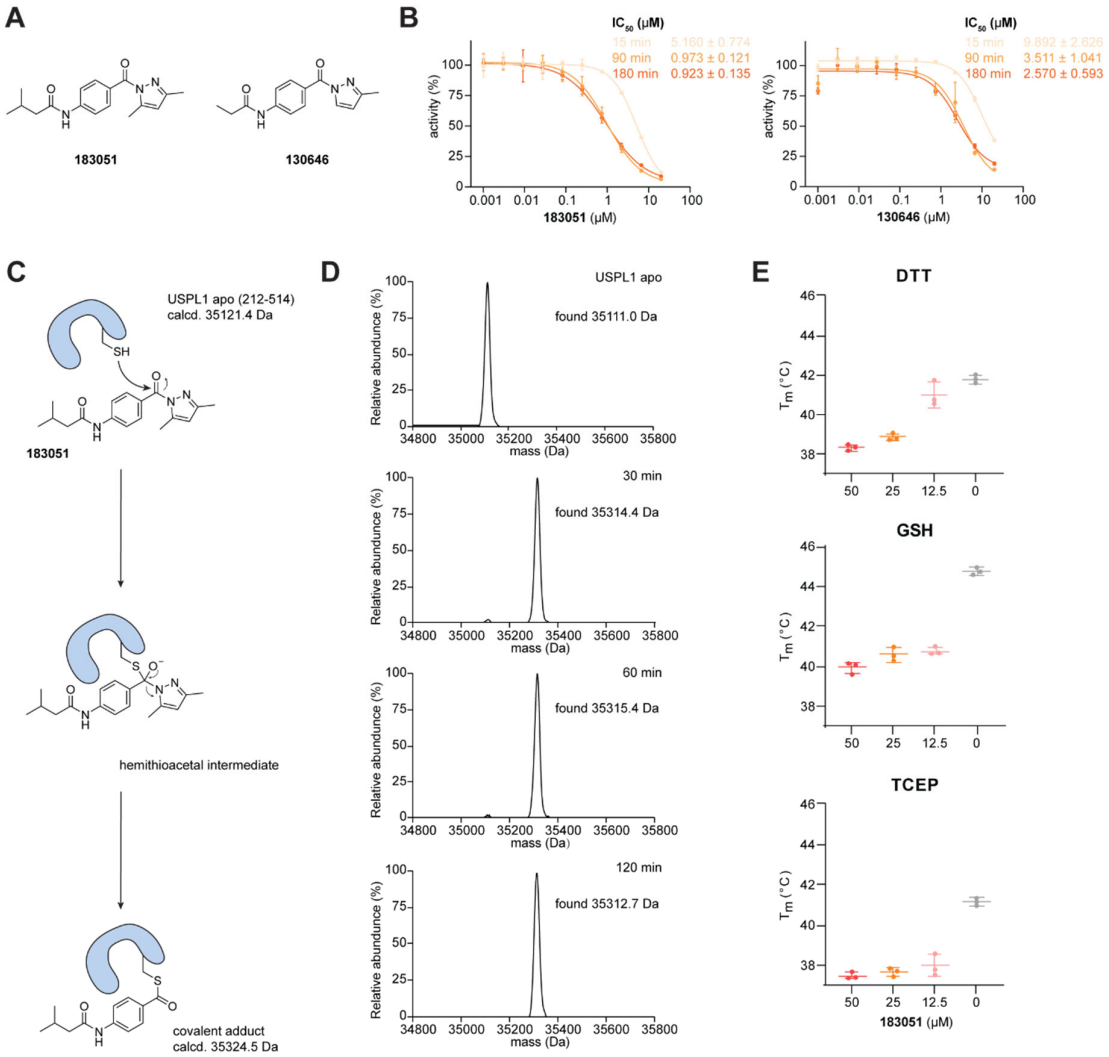


Fig 2.14. Preliminary evaluation of two hits targeting USPL1 discovered from the high-throughput screening. **A**. Chemical structures of the hit compounds 183051 and 130646. **B**. Inhibitory activities of hit compounds against USPL1 with indicated pre-incubation time using SUMO2-RhoG cleavage assay. **C**. The proposed mechanism of USPL1 inhibition by covalent modification. The calculated molecular weight of apo USPL1 as well as covalent adduct was shown. **D**. Time-dependent covalent modification of USPL1 by compound 183051 measured by intact protein mass spectrometry. **E**. Thermal shift assay results of compound 183051 in the same assay buffer with the exception of reducing agents (1 mM of each). The assay was performed in triplicate and the results were shown as mean ± standard error.

Previously, a series of *N*-benzoyl pyrazole derivatives were reported to inhibit neutrophil elastase[212]. Structure-activity relationship analysis revealed that *N*-benzoyl pyrazole derivatives can achieve relatively strong inhibition at nanomolar level against neutrophil elastase. Notably,

neutrophil elastase is a serine protease which attacks the carbonyl group with the same mechanism that USPL1 does to form a covalent complex. Apart from neutrophil elastase, fatty acid amide hydrolase (FAAH) which is also a serine enzyme hydrolyzes the endocannabinoid anandamide and related amidated signaling lipids. Many carbamate-based inhibitors were reported to react with FAAH through an addition-elimination process to achieve covalent inhibition[213].

Since there are many other cysteine or serine proteases in cellular environment, it is foreseeable that these compounds might have more off-target effects. The inhibitory activities rely on the reactivities of the carbonyl group instead of being recognized by USPL1 and subsequently forming covalent complex. The chemical space of hit compounds is considerably limited for further modifications. Therefore, additional structural optimization of the hit compounds was not considered.

2.8 Summary

In order to characterize the catalytic specificity of the deSUMOylase USPL1, fluorescence-intensity-based assays were firstly employed for quantitative comparison of modifier specificity. Ub, SUMO1, and SUMO2-RhoG substrates were semisynthesized by intein chemistry with minimal DMSO for dissolving the fluorophore. Ion exchange chromatography was applied to purify these three substrates. In the fluorescence intensity assay, USPL1 showed around 4-fold preference toward SUMO2 over SUMO1 and did not cleave Ub-RhoG at all. Given that the substrates with artificial linkages cannot represent the natural isopeptide bonds, a native semisynthetic method was designed to incorporate the fluorophore into the C-termini of ubiquitin, SUMO1 and SUMO2. The reaction was performed under mild condition in just a few minutes which maintain the native conformation as the parental modifiers. After customized two-step purification, the modifier specificity of USPL1 was examined in the fluorescence polarization assays. Quantification of the cleavage data showed that USPL1 has much more significant preference toward SUMO2 than previously anticipated. To explicitly dissect the mechanism of specificity of USPL1, two SUMO2/3-based suicide inhibitors were prepared to capture USPL1 covalently. The crystal structures of USPL1 in complex with SUMO2/3 revealed how they interact with each other. Bioinformatic-guided mutagenesis studies on both sides revealed that USPL1 utilizes its finger domain to host the Gly27 loop of SUMO2 which is unique compared with ubiquitin and SUMO1. This unique binding mode also provided structural basis for designing specific USPL1 inhibitors by targeting its finger subdomain. Another mutation at Pro66

to make SUMO2/3 more ubiquitin-like highlighted the importance of the hydrophobic interaction of SUMP2/3 with USPL1. Since a large quantity of SUMO2-RhoG was prepared, a high-throughput screening was performed at COMAS which led to discovery of two hit compounds.

The mild reaction conditions and native purification for preparing fluorescence polarization substrates were further explored to generate NEDD8 and ISG15-based substrates. Full-length ISG15 was truncated to a C-terminal domain for easier handling. After functionalization at the C-termini of NEDD8 and ISG15(CTD), the purification protocols were then optimized to obtain pure NEDD8 and ISG15(CTD)-KG-TAMRA (Fig 2.8.B). The installation of KG-TAMRA molecule changed the isoelectric point values of Ub/Ubls. The most difficult case was ISG15(CTD)-KG-TAMRA because it could not be purified by either anion or cation exchange chromatography. However, the changes in hydrophobicity upon modification by KG-TAMRA can be distinguished by hydrophobic interaction chromatography which was successfully applied to the purification of ISG15(CTD)-KG-TAMRA. In fact, hydrophobic interaction chromatography was also suitable for purifying SUMO1-KG-TAMRA which was finely purified by high-resolution anion exchange chromatography (data not shown). Further optimization of the purification protocols might be unified by combining coarse purification of size-exclusion chromatography to remove excessive KG-TAMRA molecules and fine purification enabled by hydrophobic interaction chromatography.

The successful purification of NEDD8-KG-TAMRA led to the revision of the preference of UCHL3 which was previously reported to be a dual ubiquitin and NEDD8 proteases. Quantitative data showed that its ubiquitin and NEDD8 processing capabilities should be similar instead of previously reported 1000-fold difference. Moreover, the catalytic efficiency of human USP18 was reported due to the availability of pure ISG15(CTD)-KG-TAMRA.

To apply the semisynthetic method more broadly, FUBI was selected as an example because two FUBI proteases USP16 and USP36 were characterized recently. The functionalization and purification of FUBI-KG-TAMRA were straightforward using standard methods. USP2, USP7, USP16, and USP36 were tested at single concentration by a panel of fluorescence polarization substrates consisting of ubiquitin, SUMO2, ISG15(CTD), NEDD8, and FUBI. As positive control enzymes, USP2 and USP7 both cleaved Ub-KG-TAMRA completely to the level that free KG-TAMRA reached. Unprecedently, USP16 and USP36 were both observed to cleave ubiquitin, ISG15, and FUBI. Concentration-dependent assay further confirmed their triple cross-reactivities. The substrate panel can be used to discover and validate more enzymes with cross-reactivities. The native preparation methods might be adapted to other Ubls to expand the panel.

In summary, the development of native semisynthesis of natural isopeptide-linked fluorescence polarization substrates enabled the characterization and discovery of the modifier specificity of DUBs/ULPs. The deeper understanding of the cross-reactivities on the biochemical level will provide more information for cellular investigation into the functions of DUBs/ULPs. A larger scale of preparation of these Ub/Ubl-KG-TAMRA substrates will accelerate the drug development targeting DUBs/ULPs.

3. LINKAGE SPECIFICITY

Polyubiquitin chains assembled with eight linkages are the root of sophisticated cellular signaling pathways. Besides the linear polyubiquitin forged by ordinary peptide bonds, the majority of polyubiquitin chains are based on isopeptide bonds formed by the amine moiety on the side chains of seven lysine residues and the carboxylic acid functional group of C-terminal glycine. Because each linkage type has specific functions in regulating corresponding cellular events, the modification of specific ubiquitin linkage requires precise control to maintain the homeostasis in cellular environment which is enabled by some DUBs in possession of dedicated linkage specificity[105] (Fig 3.1.A).

Biochemical characterization of linkage specificity of DUBs usually relies on a panel of polyubiquitin substrates with eight different linkages which are prepared either by enzymatic assembly or chemical synthesis. This gel-based assay format can inform the linkage specificity but not explain the underlying mechanism. In order to fully elaborate the linkage specificity of cysteine-based DUBs, visualization of the binding interface that proximal and distal sites occupied by ubiquitin molecules using structural biology is a key solution. A tactic that introducing mutation at the catalytic cysteine into alanine has been successfully implemented for several DUBs, such as CYLD in complex with Lys63-linked diUb[214], human and zebrafish USP30 with Lys6-linked diUb[120, 187], MINDY-1 or MINDY-2 in complex with Lys48-linked diUb[114], and LotA from *Legionella pneumophila* with Lys6-linked diUb[215]. This tactic leading to formation of non-covalent complex which might be sensitive to some crystallization conditions. To capture DUBs covalently, diUb-based covalent probes with various warheads were developed. By using a totally chemical synthesized diUb probe which has a conjugate addition warhead in the linker, the crystal structure of Lys11-specific DUB Cezanne in covalent with its substrate was solved[191]. With the invaluable structural information, the mechanism of the linkage specificity could be revealed and the proteases could be used as research tools for other applications.

The chemical synthesis of diUb probes might not be easily transferred between different laboratories and thereby constrain broader application in studying other linkage-specific DUBs. Therefore, a streamlined semisynthetic method based on recombinant ubiquitin should be designed for generating a large quantity of probes. Previous success in using mono-ubiquitin-based covalent probe to capture DUBs could be the foundations for the design of novel di-ubiquitin-based probes.

3.1 Design of DiUb Probes with Internal Alkyl Bromide Warheads

Previously reported diUb-based covalent probes all employed the warheads with addition-based mechanism to occupy the proximal and distal ubiquitin binding sites and capture DUBs. Alkyl-halide-based mono-ubiquitin covalent probes have been widely used for proteomic studies as well as structural characterization of DUBs which were captured by a nucleophilic substitution reaction to form an irreversible thioether bond.

Ub-2Br has a bromide leaving group which is connected by two carbon atoms at the C-terminus of ubiquitin (Fig 3.1.B). Ub-2Br has been successfully used for structural determination of USP46 and OTUB2[216, 217]. To design an alkyl-bromide-based diUb covalent probe, the carbon atom accommodating the bromide atom could be the potential exit vector to chain extension to form an isopeptide-like linker by additional five carbon atoms (Fig 3.1.C). Given that a cysteine mutation could be introduced into the proximal ubiquitin for forging the isopeptide-like linker, a sulfur atom was placed in the "lysine" side chain (Fig 3.1.D).

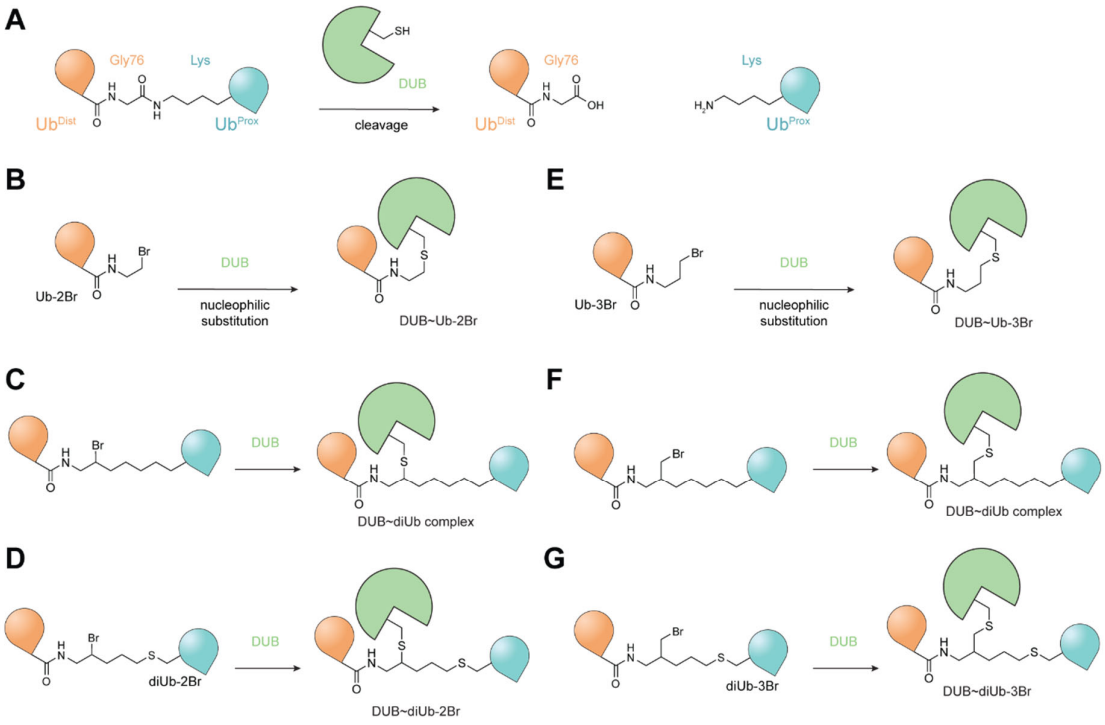


Fig 3.1. Design principle of covalent diUb probes with warheads utilizing nucleophilic substitution mechanism to capture DUBs. **A**. Schematic illustration of native isopeptide bond between two ubiquitin molecules and how DUB cleaves the bond. **B**. Ub-2Br probe covalently reacts with a DUB through a nucleophilic substitution. **C**. Design of a diUb-2Br probe in which the lysine side chain contains four carbon atoms. **D**. Design of a diUb-2Br probe in which the linker has a sulfur atom. **E**. Ub-3Br probe covalently reacts with a DUB through a nucleophilic substitution. **F**. Design of a diUb-3Br probe in which the lysine side chain contains four carbon atoms. **G**. Design of a diUb-3Br probe in which the linker has a sulfur atom.

Similarly, Ub-3Br which has one more carbon atom and was successfully applied for solving several crystal structures of viral DUBs also has the same mechanism to trap DUBs (Fig 3.1.E). Therefore, diUb-3Br probes could be designed in analogy to diUb-2Br probes (Fig 3.1.F&G). Since bromide warhead might be sensitive to harsh reaction conditions, orthogonal synthetic routes compatible with warheads are required to ensure the integrity and quality of the final products.

3.2 Retrosynthetic Analysis of Newly Designed DiUb Probes

Although four diUb probes with bromide-based warheads have been designed, the probes whose linkers consisting of merely carbon atoms might require synthesis performed in organic solvents which might damage the quality of final products. Therefore, the probes (Fig 3.1.C&F) were not chosen for further development. Instead, the other two probes were developed.

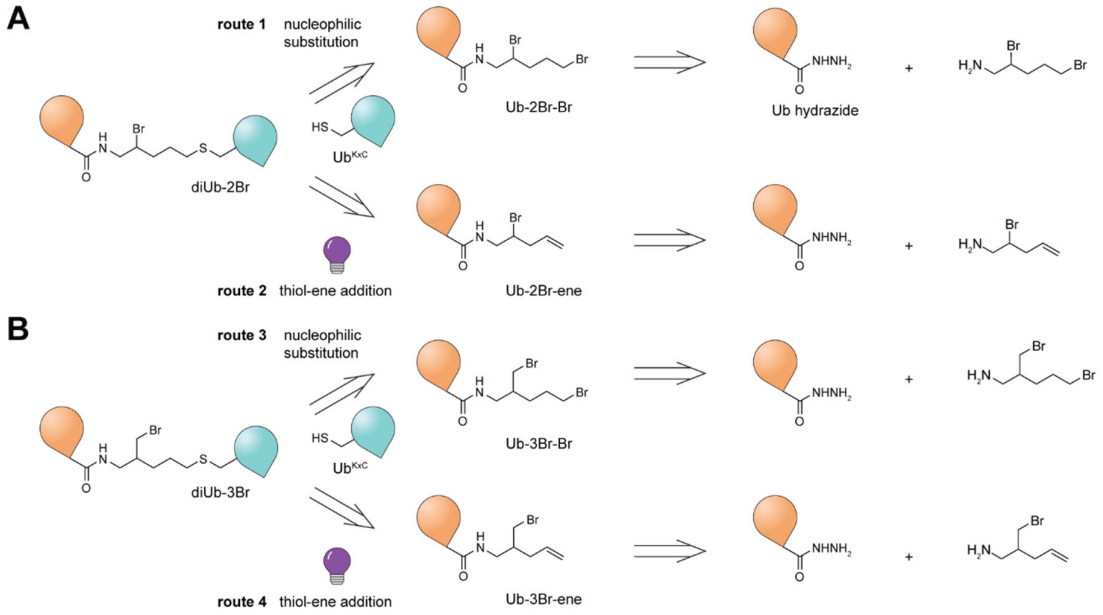


Fig 3.2. Retrosynthetic analysis of diUb-Br probes. **A**. Retrosynthetic routes for assembling diUb-2Br probe using either nucleophilic substitution (route 1) or thiol-ene chemistry (route 2). **B**. Retrosynthetic routes for assembling diUb-3Br probe using either nucleophilic substitution (route 3) or thiol-ene reaction (route 4). Required small molecules are shown on right.

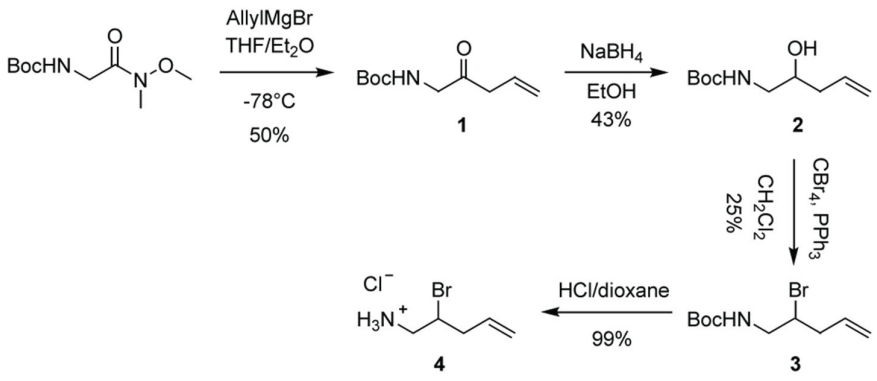
Since ubiquitin has no any cysteine residues, a cysteine residue can be introduced into the specific lysine position (Ub^{KxC}) for conjugation. The thiol moiety of the Ub^{KxC} is able to react with either nucleophilic substitution or thiol-ene addition. For nucleophilic substitution, the distal ubiquitin has to be functionalized with a long chain with a terminal bromide (Fig 3.2.A). However, the nucleophilic substitution usually requires alkaline condition which will lead to

side reaction since there are two bromide atoms in the linker (route 1). To overcome this drawback, thiol-ene reaction under UV light could be employed for ligation with its orthogonality[218] (route 2). The reaction is performed in acidic condition and thereby bromide atom will not affect the ligation between cysteine and alkene. Retrosynthetic analysis was applied to design diUb-3Br probe as well (Fig 3.2.B). Last but not least, syntheses of two small molecules are essential for generating the diUb-2Br and diUb-3Br probes.

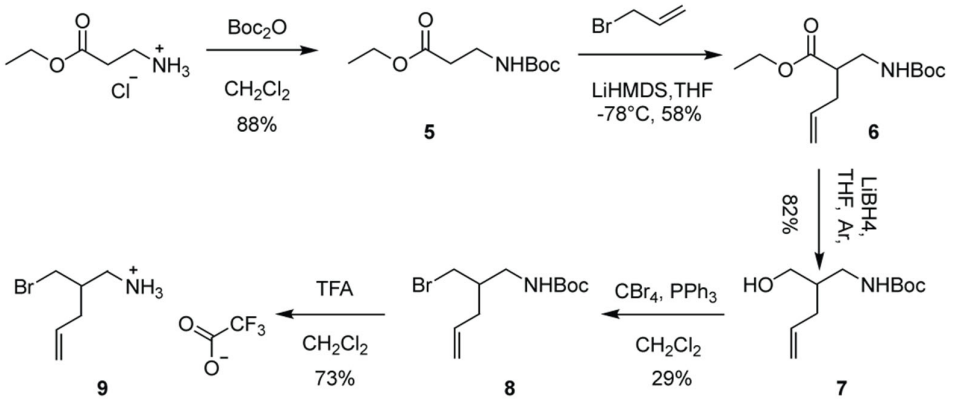
3.3 Chemical Synthesis of the Warheads for Assembling DiUb-Br Probes

Since the two warheads differ in the length of carbon atoms, the synthetic routes require customized design. Due to the amine and bromide might react intermolecularly, the small molecules should exist as salt forms to enable stability for longer shelf lives.

To synthesize the warhead for generating diUb-2Br, a commercially available Boc-protected glycine-based Weinreb amide was attacked by allylmagnesium bromide in a dry ice-acetone bath to extend the chain with a terminal alkene to obtain compound **1** with a moderate yield (Scheme 3.1). Compound **1** was then treated with sodium borohydride in ethanol solution to reduce the ketone to a hydroxyl group. It is speculated that some of the products might be lost during concentration on the rotary evaporator because they are volatile. Therefore, a bulb-to-bulb distillation might be used to improve the yield of these two steps. Followed by an Appel reaction, the hydroxyl group was replaced by a bromine atom. After purification by silica gel chromatography, compound **3** was deprotected to obtain the final compound **4** in a HCl salt with a quantitative yield.



Scheme 3.1. Synthesis of warhead for generating diUb-2Br. Reaction conditions and yield are shown. AllylMgBr: allylmagnesium bromide, $NaBH_4$: sodium borohydride, CBr_4: carbon tetrabromide, PPh_3, triphenyl phosphine.



Scheme 3.2. Synthesis of warhead for generating diUb-3Br. Reaction conditions and yields are shown around the arrows. Boc_2O: di-*tert*-butyl decarbonate, LiHMDS: lithium bis(trimethylsilyl)amide, $LiBH_4$: lithium borohydride, TFA: trifluoracetic acid.

The effort to obtain the small molecule for generating diUb-3Br started from commercially available ethyl 3-aminopropanoate hydrochloride. The amine functional group was firstly masked with a Boc protecting group to give compound **5**. In a dry ice-acetone bath, compound **5** was deprotonated by the non-nucleophilic base lithium bis(trimethylsilyl)amide (LiHMDS), allyl bromide was then added into the mixture to afford the intermediate compound **6**. The ester moiety of compound **6** was reduced to an alcohol group by lithium borohydride in THF solution under the protection of argon to give compound **7**. Followed by an Appel reaction, bromine was introduced to the hydroxyl position to give compound **8**. After deprotection by trifluoroacetic acid (TFA), final compound **9** was obtained in a salt form. Since both of the final products exist as salt forms, they will thereby not react with themselves. Due to the formation of the salt, the final products can be easily dissolved in aqueous buffer without the assistance of any other organic solvents which is good to the following reaction on ubiquitin. To achieve better physical properties for easier handling, lyophilization was applied to both final products. Finally, all the intermediates and final products obtained in the aforementioned synthetic routes were fully characterized with NMR spectroscopy and high-resolution mass spectrometry. The final products were stored under the protection in an inert atmosphere.

In general, the two synthetic routes were performed successfully to obtain a large quantity of small molecules for the purpose of functionalization of distal ubiquitin as well as subsequent photo-triggered thiol-ene ligation with proximal ubiquitin which bears a lysine-to-cysteine mutation. The synthetic routes only contain four or five steps which do not require special equipment. Notably, the synthetic routes both utilize commercially available and inexpensive starting materials.

3.4 Preparation of DiUb-2Br and DiUb-3Br Probes

Since the warhead-containing small molecules already mimic the length of the terminal glycine residue as well as part of the lysine side chain, a truncated ubiquitin which lacks the C-terminal glycine residue was cloned into pTXB1 vector. The expressed ubiquitin was in fusion at C-terminal part with an intein domain and a chitin-binding domain (Fig 3.3.A). Based on the *N*-to-*S* shift mechanism, excessive thiol molecule was added into the reaction to induce the thiol cleavage to obtain intermediate Ub-MesNa which was used without purification.

Ub-MesNa as a relatively active thioester can undergo efficient aminolysis upon addition of excessive amount of hydrazine which usually lifts the pH value up to around 9. The obtained intermediate Ub-hydrazide (also referred to as Ub-$NHNH_2$) was then concentrated to around 2 mM for further reaction or long-term storage at -80°C (Fig 3.3.B). To functionalize Ub-hydrazide, the same reaction which was applied to the semisynthesis of Ub/Ubl-KG-TAMRA was employed on truncated ubiquitin again. Owing to the last residue in the truncated ubiquitin is still a glycine which has no chiral center, there is no any concern about the epimerization of the last residue. Ub-hydrazide was cooled in a sodium chloride-ice bath to minimize the occurrence of unwanted side reactions due to higher reaction temperature, such as Curtius rearrangement. This reaction usually leads to the generation of isocyanate intermediate by releasing nitrogen gas. The isocyanate is considerably active and thereby might further react with a series of nucleophilic moieties, such as hydroxyl groups from amino acids, water, and amine groups from ubiquitin or small molecules to form unfavored carbamate, free amine or urea derivative at the terminal glycine residue[219].

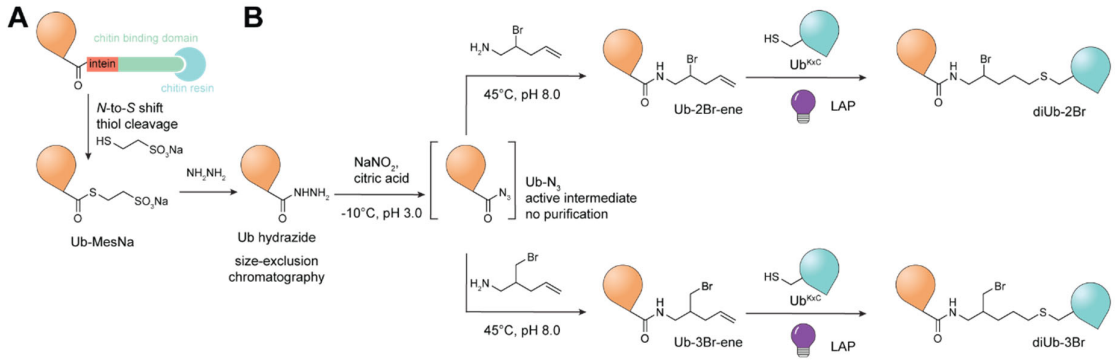


Fig 3.3. Semisynthetic routes of newly designed diUb-2Br and diUb-3Br probes. **A**. Preparation of ubiquitin-based intermediate using intein chemistry from recombinant ubiquitin. **B**. Semisynthetic routes for diUb-2Br and diUb-3Br from intermediate Ub-hydrazide which was activated by nitrous acid functionalized by respective small molecules. The obtained Ub-2Br-ene and Ub-3Br-ene were conjugated with Ub^{KxC} through a UV-initiated free radical reaction using the free radical initiator lithium phenyl-2,4,6-trimethylbenzoylphosphinate (LAP).

Highly concentrated Ub-hydrazide was activated by mixing sodium nitrite, followed by the addition of citric acid to decrease pH to around 3.0 and thereby resulted in the formation of oxidative nitrous acid. Ub-hydrazide was oxidized into a highly reactive intermediate Ub-N_3 of which purification is not necessary. A solution of small molecules in pH 8.0 buffer (1.5 M HEPES to maintain the pH value) was immediately added into Ub-N_3. The reaction was transferred directly to a water bath at 45°C for several minutes to generate Ub-2Br-ene or Ub-3Br-ene which are both able to be purified by size-exclusion chromatography at pH around 5 which is suitable for next step. Since the two small molecules are in salt forms, up to 40 equivalencies can be used in the reaction to achieve good conversion but also save the small molecules from be wasted. The purified Ub-2Br-ene or Ub-3Br-ene proteins were highly concentrated to around 2 mM. Mass spectra were recorded to confirm the quality and purity of these two intermediates. However, side products were observed in the functionalization reaction. The purification was performed on a size-exclusion chromatography which is not able to separate desired products from some other side products with similar hydrodynamic radius. In the purified Ub-2Br-ene, a small portion of Ub-2OH-ene in which hydroxyl replaced the bromide atom by hydrolysis was observed (Fig 3.4.A). Additionally, free ubiquitin with a deletion of Gly76 molecules were also observed. There were also some non-covalent adducts potentially formed by sodium ion or acetonitrile from the buffer or mobile phase. Similarly, the deconvoluted mass spectrum of Ub-3Br-ene also showed a small amount of hydrolyzed Ub-3OH-ene and free ubiquitin but the undesired side products are obviously much lower than the Ub-2Br-ene sample which might be due to its primary carbon with higher stability (Fig 3.4.B). Given that next step is merely dependent on the UV-triggered free radical reaction, the hydrolyzed side product and free ubiquitin will not affect the thiol-ene addition, the protein samples were used without fine purification.

In parallel, three different Ub mutants (K6C, K48C, and K63C) were purified and concentrated to around 2 mM as well using acid precipitation and subsequent cation exchange chromatography. The buffer was exchange to the same buffer for storing intermediates Ub-2Br-ene or Ub-3Br-ene. All the Ub mutants were analyzed by intact protein mass spectrometry. The deconvoluted mass data are consistent with the theoretical values (Fig 3.4.C&D&E).

The thiol-ene reaction was performed by mixing 50 µL Ub-2Br-ene or Ub-3Br-ene, 50 µL Ub mutants, and 1 µL LAP with final concentration around 0.5 mM which was dissolved in advance in the same buffer on ice. Although previous work indicated a specific model of UV light for initiating the reaction[218], UV light (365 nm, 8 Watt) was sufficient for the reaction.

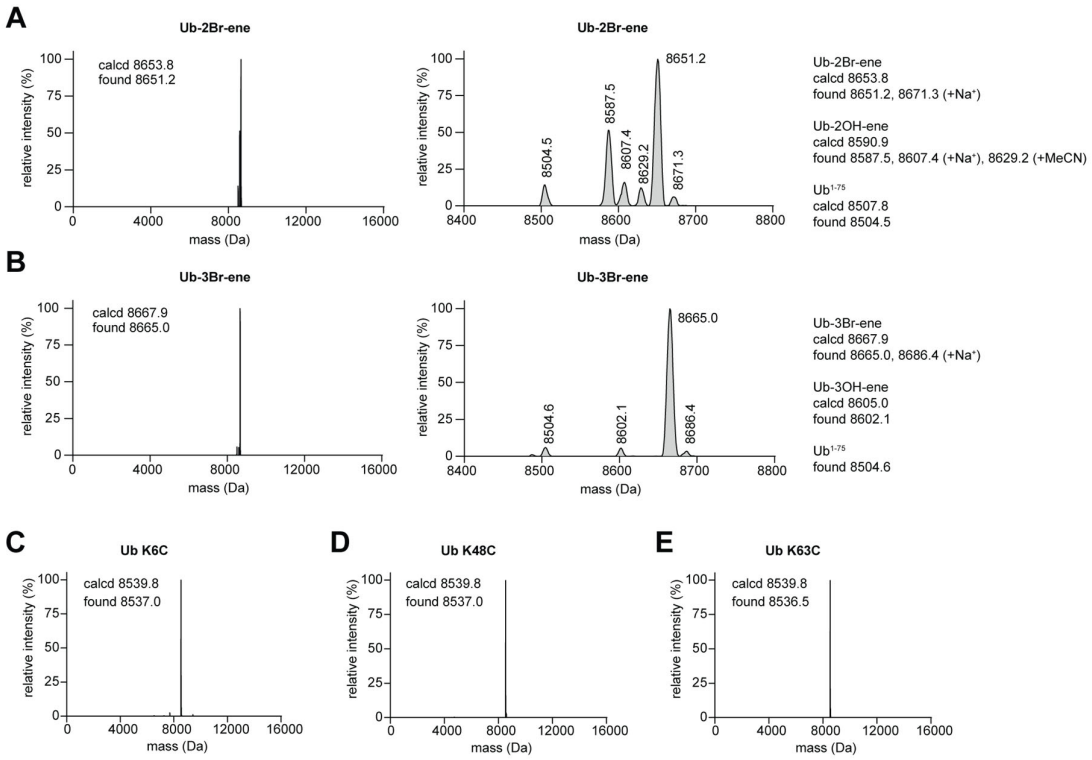


Fig 3.4. Characterization of intermediate proteins for generating diUb-2Br and diUb-3Br probes. **A**. Deconvoluted mass spectrum of intermediate Ub-2Br-ene. Close-up view is shown as well where desired product is shown horizontally. Detailed analysis of side products is listed on right. **B**. Deconvoluted mass spectrum of intermediate Ub-3Br-ene. Close-up view is shown in the middle. Detailed analysis of side products is listed on right. **C**. Deconvoluted mass spectrum of Ub K6C. **D**. Deconvoluted mass spectrum of Ub K48C. **E**. Deconvoluted mass spectrum of Ub K63C. Calculated and observed mass data shown as inserts.

The reaction was set up in a considerably small volume to ensure sufficient penetration of UV light to initiate the free radical reaction (see Appendix Fig 2). With the irradiation of UV light, the reaction progress was monitored by intact protein mass spectrometry. Raw mass data were deconvoluted for analysis in detail. Once the starting materials were consumed almost completely, the reaction in several tubes were pooled together. For easier operation, the reaction mixture was subjected to a size-exclusion chromatography for isocratic elution. Alternatively, cation exchange chromatography might be also employed for fine purification since the di-ubiquitin shows different chromatographic behavior from mono-Ub, such as free ubiquitin and starting materials. Given that the diUb-2Br and diUb-3Br covalent probes will be used in the scenario of capturing DUBs irreversibly and the resulting complex will be further finely purified, trace impurities of mono ubiquitin impurities might not affect the qualitative applications. Due to the existence of hydrolyzed impurities, namely Ub-2OH-ene and Ub-3OH-ene, diUb-2OH and diUb-3OH might be obtained after the thiol-ene ligation. Since hydroxyl group is inert toward cysteine, there is no any concern if the final products contained such impurities.

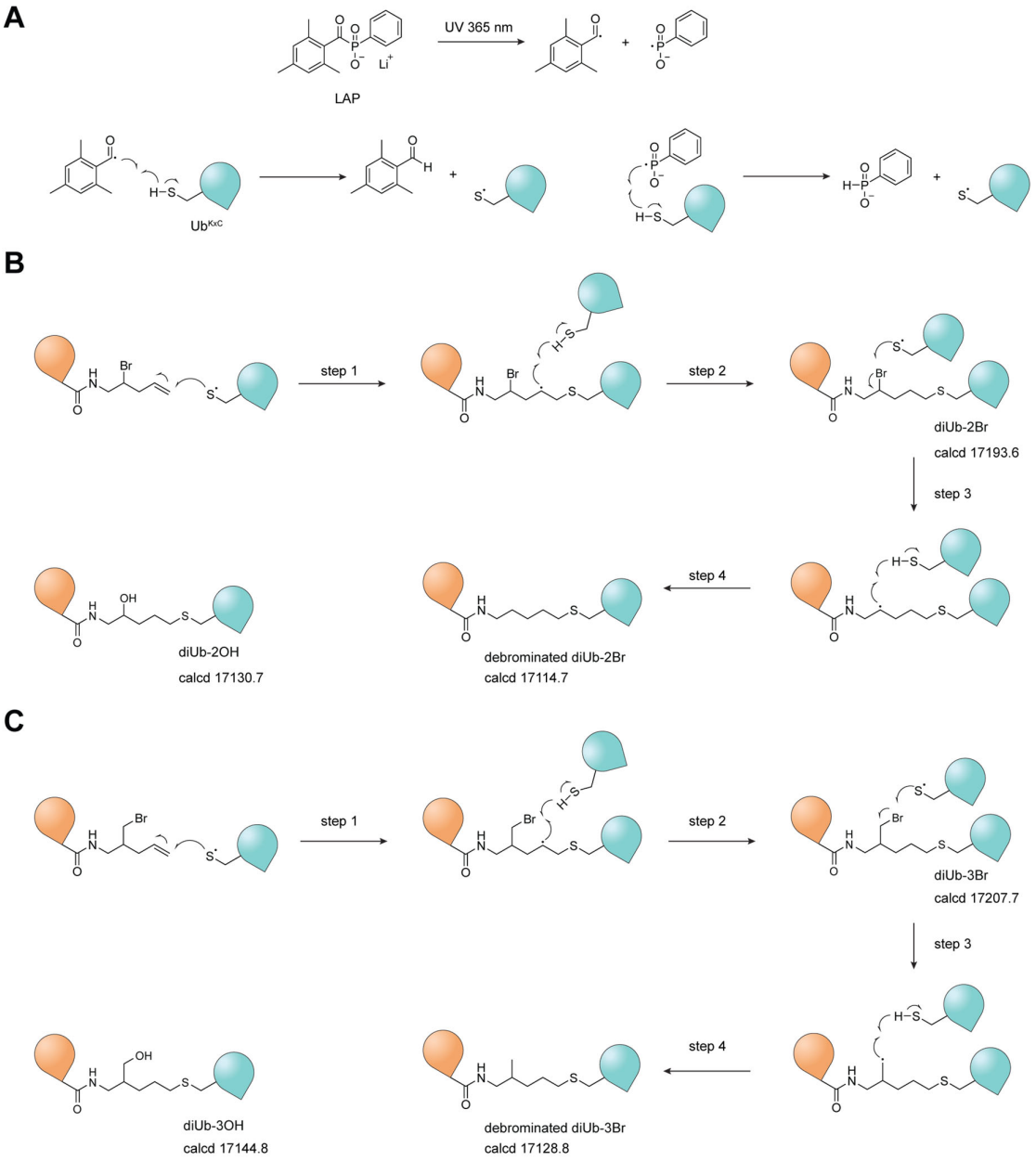


Fig 3.5. Overview of the mechanism of photo-initiated free radical reaction to generate probes and mechanism leading to potential side products. **A**. Homolytic cleavage by UV light 365 nm to form free radicals. Both free radicals can induce propagation of free radical formation on the side chain of cysteine residues. **B**. Proposed mechanism to explain the side products observed in the diUb-2Br assembly reaction. **C**. Proposed mechanism underlying the generation of side products during the preparation of diUb-3Br probe. Calculated mass data are shown under the correct final products as well as potential side products.

Although these potential impurities would theoretically not have a negative effect on trapping DUBs covalently, it would be beneficial to know the underlying mechanism of how side products are formed to provide more information for accurate data interpretation. Moreover, a better quality of these probes might broaden their applications, such as for proteomic studies to discover new linkage-specific DUBs or E3 ligases.

To ignite the free radical reaction upon UV light exposure, lithium phenyl-2,4,6-trimethylbenzoylphosphinate (LAP) was selected as the photoinitiator since it had been successfully introduced into protein conjugation[218]. LAP is a well-characterized and commercially available photoinitiator which has been widely used in free radical polymerization of hydrogels and other materials[220].

In the reaction mixture, photoinitiator LAP was instigated by UV light to undergo homolytic cleavage which led to the production of free radicals for reaction initiation (Fig 3.5.A). One molecule of LAP disassociated into two free radical species which were both able to the react with thiol functional group in the ubiquitin bearing a cysteine mutation to produce thiyl free radicals. In the case of preparation of diUb-2Br, thiyl free radicals were coupled in succession to the terminal alkene moiety in the Ub-2Br-ene (Fig 3.5.B). The first step led to the formation of isopeptide-like bond with an unpaired free radical which was further coupled with cysteine to obtain the desired product Ub-2Br-ene after the second step. However, the thiyl radicals might continue to propagate which could result in the homolytic cleavage of carbon-bromine bond to give a secondary free radical in the linker between proximal and distal ubiquitin molecules. This active secondary free radical would proceed to react with cysteine residues in the ubiquitin. Eventually, a side product which lost a bromine atom was formed without capability of trapping DUBs covalently. Since there was already some hydrolyzed side product Ub-2OH-ene in the starting material (Fig 3.4.A), diUb-2OH can also be observed after the thiol-ene addition. The side product diUb-2OH theoretically should be inert because of the hydroxyl group should not react with the catalytic cysteine of DUBs. Therefore, these probes can be directly used without optimization of the purification protocols.

Three different isopeptide-like linkages, including Lys6, Lys48, and Lys63, were forged using this thiol-ene reaction. After purification by a size-exclusion chromatography, all the diUb-2Br probes were analyzed by intact protein mass spectrometry. In the K6C-linked diUb-2Br sample, desired product diUb-2Br was observed (Fig 3.6.A). Hydrolyzed diUb-2OH also existed in the sample. For K48C-linked diUb-2Br, desired product was found but with some diUb-2OH as well as a debrominated side product (Fig 3.6.B). Similar results were observed as well in the K63C-linked diUb-2Br sample but desired product was the dominant component (Fig 3.6.C). The variations in the desired product diUb-2Br and undesired impurities from the different samples might originate from the positions of cysteine residue in either disordered loop or rigid secondary structures which could enhance the possibility of carbon-bromine bond cleavage in a free radical manner (Fig 3.5.B&C).

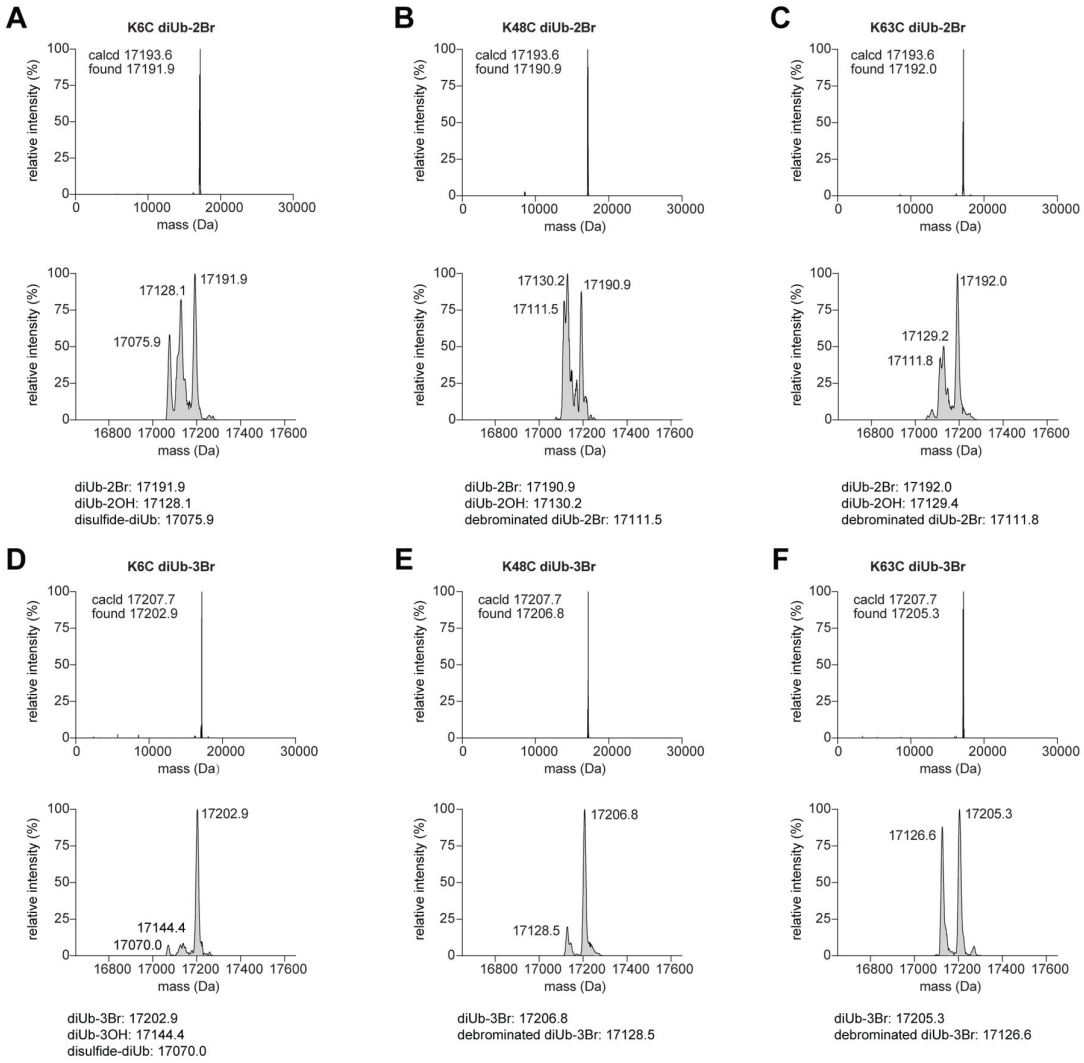


Fig 3.6. Deconvoluted intact protein mass spectra of diUb-2Br and diUb-3Br probes with Lys6, Lys48, and Lys63 linkages respectively. **A**. Deconvoluted mass spectrum of K6C diUb-2Br and close-up view (bottom). **B**. Deconvoluted mass spectrum of K48C diUb-2Br and close-up view (bottom). **C**. Deconvoluted mass spectrum of K63C diUb-2Br and close-up view (bottom). **D**. Deconvoluted mass spectrum of K6C diUb-3Br and close-up view (bottom). **E**. Deconvoluted mass spectrum of K48C diUb-3Br and close-up view (bottom). **F**. Deconvoluted mass spectrum of K63C diUb-3Br and close-up view (bottom). Detailed analysis of potential impurities is shown and annotated under each figure.

The diUb-3Br probes were also analyzed by intact protein mass spectrometry. In the K6C-linked diUb sample, expected product diUb-3Br was the most abundant component (Fig 3.6.D). Small portions of impurities were observed and assigned as diUb-3OH. Additionally, disulfide-linked diUb which might be from incomplete reduction. In K48C-linked diUb sample, desired diUb-3Br was the major component (Fig 3.6.E). Interestingly, K63C-linked diUb-3Br sample contained more diUb-3OH which might be due to the structural feature in this position. Since primary free radicals are less stable than the secondary ones, debrominated species were significantly less than in diUb-2Br probes (Fig 3.5.B&C).

3.5 Validation of the Reactivity of DiUb-2Br and DiUb-3Br Probes

In order to minimize the complexity in the protein purification steps, size-exclusion chromatography was favored and employed for easier purification in the two-step semisyntheses. All the probes were kept in buffer with a pH value below 7 and stored for long term at -80°C to reduce any potential hydrolysis of the nucleophilic-substitution-oriented warheads which might be labile if stored in alkaline environment. Characterization of the obtained probes with intact protein mass spectrometry assured the quality. Although some side products were observed, they were fully assigned based on the mechanism of thiol-ene free radical reaction. The undesired impurities are considered as neutral components which might slightly interfere the binding by DUBs to decelerate the reaction but would not affect the formation of complex of since those impurities are not armed with covalent warheads.

Previous development of covalent diUb probes with addition-based mechanisms displayed the art and beauty of chemical synthesis of proteins. However, the preparation of desired products requires cutting-edge peptide synthesizer and related instruments which may limit the broad distribution of these probes. For example, the Dha-based probes were synthesized chemically with a cysteine residue incorporated to replace the terminal glycine of the distal ubiquitin. The elimination reaction was performed to remove the thiol group which led to the formation of the Dha warhead[195]. Two types of linkage were forged, namely Lys48 and Lys63-linked probes. Linkage-specific DUBs, such as CYLD and OTUB1 were tested since they possess Lys63 and Lys48 chain specificity. They both reacted specifically toward corresponding Dha-based diUb probes. Non-specific DUBs, such as USP2, reacted directly with both probes. OTULIN, a Met1-specific DUB, showed considerably weak off-target effects after incubation with probes. These data showed promising results of the application of Dha-based probes which are selective and mildly reactive.

Although several DUBs were tested with the Dha-based probes with isopeptide bonds, these probes were not further employed for structural determination of DUBs. Ovaa and co-workers reported a novel thiol handle which was not only used for native chemical ligation to forge isopeptide-linke bonds, but also served as a masked group for late-stage functionalization in the linker between proximal and distal ubiquitin molecules[221]. Followed by elimination of thiol group, a Michael acceptor warhead was introduced. This method was applied to the all seven linkages and the resulting probes were validated by the promiscuous DUB USP7 on a gel-based assay.

Owing to the development of these probes, the covalent complex structure of Cezanne and Lys11 diUb was successfully solved[222]. Given that the synthesis is considerably difficult, this type of probes was only applied to the study of Cezanne. Another Michael-acceptor-based diUb probe was devised by the Zhuang group through a mild semisynthetic workflow[223]. After around a decade since its first disclosure, a Lys48-linked probe was successfully employed to study a HECT E3 UBR5 which prefers to forging Lys48-linked polyubiquitin chains[190]. Initial test of these probes was performed on promiscuous USP2 and USP21 which both showed significant molecular weight shift on the gel-based assay. UCHL1 was further tested but showed only weak off-target effect on Lys48-based diUb probe. OTUB1 reacted only with Lys48-based probe as it is a Lys48-linkage specific DUB.

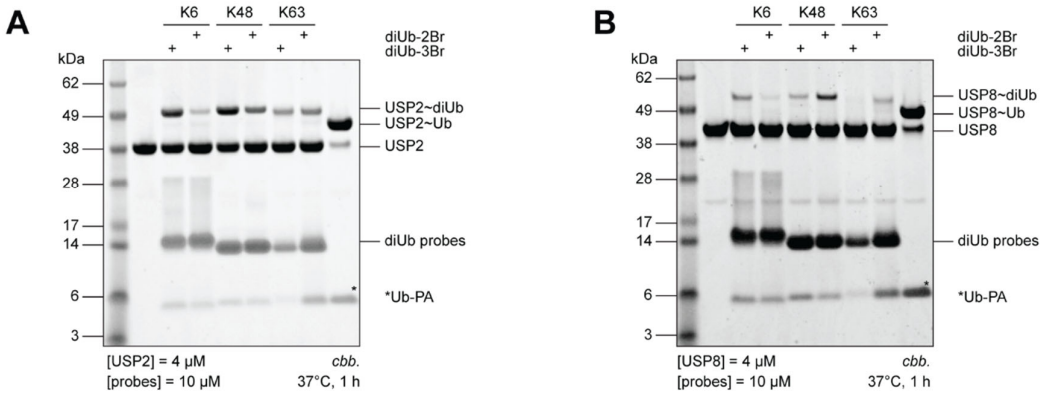


Fig 3.7. Test reactions to validate the diUb-2Br and diUb-3Br probes with two promiscuous DUBs without linkage specificity. **A**. USP2 was incubated with the panel of diUb-based covalent probes. Incubation time and temperature are listed below the figure. **B**. USP8 was tested with probes in the aforementioned condition. Proteins are annotated on the right side. Mono-Ub-based covalent probe Ub-PA was used as a positive control to confirm the deubiquitinating activities of USP2 and USP8 enzymes. Ub-PA was highlighted with an asterisk.

In analogy to the previous development of diUb-based covalent probes, the newly designed alkyl-bromide-based diUb probes were assembled as an assay panel. Promiscuous DUBs USP2 and USP8 catalytic domains were selected to validate the reactivities of diUb-2Br and diUb-3Br probes. After incubation with excessive probes at 37°C for 1 hour, USP2 broadly reacted with these probes with clear band shifts on the gel (Fig 3.7.A). Mono-Ub-based covalent probe Ub-PA was incubated with USP2 as well to confirm the deubiquitinating activity. The USP2~Ub complex showed less shift on the gel compared to the USP2~diUb species. It is interesting that diUb-2Br and diUb-3Br behaved quite differently. In Lys6-linked probes, diUb-3Br formed more covalent complex with USP2 while diUb-2Br showed less. This effect might be due to the more impurities observed in diUb-2Br sample. In the case of USP8 (Fig 3.7.B), Lys48-linked diUb-2Br behaved better which implied that the selection of probes should be customized to achieve an optimal readout of the gel-based assays.

3.6 Scouting for Linkage-specific DUBs with DiUb Probes

Since the reactivity of alkyl-bromide-based diUb probes have been validated by two promiscuous DUBs USP2 and USP8, further exploration on understudied DUBs with intrinsic linkage specificity should be performed. Although USP has the most abundant members in the DUB family, most USPs do not possess linkage preference. One of the outliers in the USP family is USP30 which prefers to cleave Lys6-linked polyubiquitin chains to a certain degree. However, the complex of USP30 with the catalytically inactive mutation with Lys6-linked diUb substrate was solved and in-depth studied[120, 187]. Therefore, USP30 was not chosen for further test with alkyl-bromide-based diUb probes. Interestingly, a recent work with a focus on USP20 reported that USP20 selectively cleaves Lys48 and Lys63 polyubiquitin chains[224]. However, the data were contradictory to previous work that using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry-based deubiquitinating assay to characterize a series of DUBs where USP20 showed no selectivity toward different eight linkages[107]. In the output of MALDI-TOF-based assays, most USPs were confirmed as non-selective DUBs toward specific linkages while USP9Y might be an exception which requires additional biochemical validation.

In search of linkage-specific DUBs, a ground-breaking work provided informative resources on the OTU family through systematic biochemical validation[123]. Besides some OTUs with linkage specificity which have been extensively studied, such as OTUD5 and OTUB1, OTUD4 stands out as an attractive enzyme since phosphorylation at a specific site could switch its specificity from Lys48 linkage to Lys63-linked chains[127]. OTUD4 purified from bacterial and mammalian cells both confirmed the specificity reorientation upon phosphorylation. To express and purify OTUD4, a plasmid construct based on pOPIN-B was prepared which encoding OTUD4 (residues 1-300). However, after several rounds of optimization, no intact OTUD4 (residues 1-300) could be successfully induced and purified from bacteria which was contradictory to what previously reported[127]. Therefore, the newly designed alkyl-bromide-based probes were not tested on OTUD4.

To include more types of DUBs with differential background and specificities, four enzymes with three different linkage preferences (Lys6, Lys48, and Lys63) from human (ZUFSP), yeast (Mug105) and bacteria (WcVTD and SnVTD) were selected and studied using gel-based assays. All these enzymes were kindly provided by the Kay Hofmann lab at the University of Cologne.

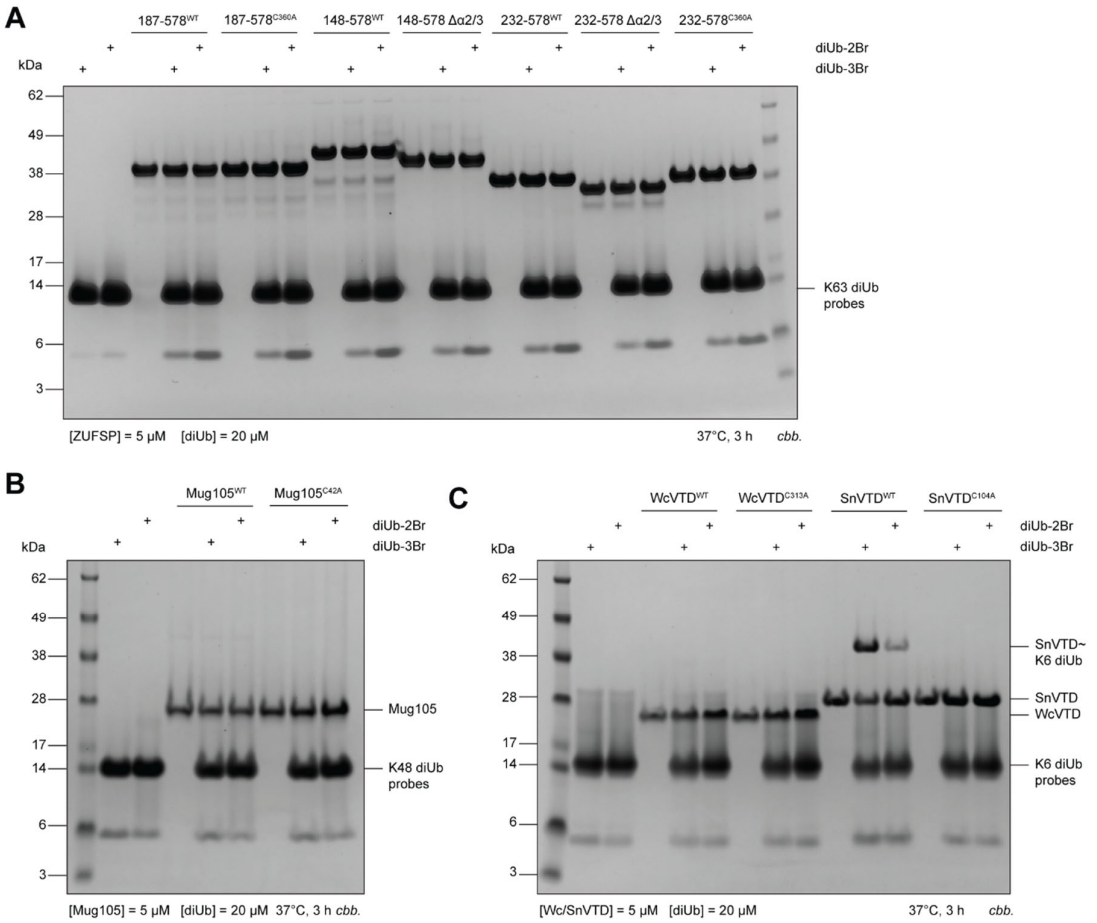


Fig 3.8. In search of potential DUBs with linkage specificity that can react with the alkyl-bromide-based covalent diUb probes. Representative DUBs from human, yeast and bacterial were selected. **A**. Several truncated versions of Lys63-linkage specific ZUFSP were tested with the Lys63-linked probes. Δ indicates deletion of specific regions. **B**. Lys48-linkage specific Mug105 from fission yeast *Schizosaccharomyces pombe* was tested with Lys48-linked diUb probes. **C**. Lys6-linkage specific VTD DUBs WcVTD from *Waddlia chondrophila* and SnVTD from *Simkania negevensis* were tested with Lys6-linked diUb probes. All the DUBs shown in this figure are kind gifts from the Hofmann lab, University of Cologne.

ZUFSP is a Lys63-linkage specific DUB which was recently identified using various approaches, such as bioinformatic analysis and proteomic strategies. The covalent complex structures of ZUFSP with probe Ub-PA have been reported as well[137]. However, how it achieves Lys63 linkage specificity at the proximal ubiquitin binding site is still unclear. As ZUFSP itself constitutes a completely new subfamily within DUBs in human genome, structural characterization of its cleavage preference would enable deeper understanding of its cellular functions. Unfortunately, even after incubation at 37°C for three hours, none of the ZUFSP enzymes with different length showed band shift on the gel (Fig 3.8.A). Given that diUb probes were used in four equivalency, ZUFSP enzymes should be saturated with probes. Although this experiment included some truncated versions of ZUFSP, such as Δα2/3 which was putative proximal ubiquitin binding site, no reaction was observed.

Since Mug105 was discovered as a Lys48 specific DUB from fission yeast as the compact homolog of ZUFSP, Lys48-linked diUb probes were incubated with Mug105 and the reactions were resolved on gel. Similar to what had been observed in ZUFSP, Mug105 did not form any covalent complex with alkyl-bromide-based diUb probes (Fig 3.8.B). This might be due to the low activity of enzymes which cannot initiate effective attack against the warheads. Another possible explanation is that bromine atom is too bulky for enzymes to accommodate due to the narrow catalytic sites.

Given that linkage-specific enzymes from human and yeast did not show promising results, it would be interesting to test some newly discovered and identified DUBs from bacteria[111]. WcVTD1 is a recent characterized DUB from *Waddlia chondrophila*[110]. Gel-based assay showed that WcVTD1 can react with Ub-PA and form covalent complex. WcVTD1 was biochemically profiled with a panel of eight diUb substrates in which it showed preference toward Lys6-linked substrate. After incubated with diUb-based covalent probes, faint band of covalent complexes can be observed which indicate that longer incubation time might be beneficial for the complex formation (Fig 3.8.C). However, the trace complexes were not enough for purification, subsequent characterization by mass spectrometry or biophysical methods and crystallization studies.

Another DUB candidate for testing from bacteria is SnVTD which was bioinformatically discovered and biochemically identified from *Simkania negevensis*, an intracellular Chlamydia-like pathogen of the respiratory tract[111]. SnVTD shares considerably similar biochemical profile with WcVTD1 due to they are both specific to Lys6-linked polyubiquitin chains. Interestingly, SnVTD cannot be captured by Ub-PA probe to form covalent complex which is different from the behavior of WcVTD1. Therefore, structural characterization of SnVTD is hard due to the apo form is usually much less stable than the substate-bound DUB.

Gratifyingly, SnVTD reacted efficiently with both diUb-2Br and diUb-3Br probes and form intense band shifts on gel (Fig 3.8.C). The catalytic inactive SnVTD (C104A) cannot form any complex with probes due to the loss of catalytic reactivity. Probes with same linkage but different length of warheads showed different reactivity in the case of USP2 and USP8 (Fig 3.8.A&B) which cannot be predicted due to the intrinsic enzymatic activities or quality of probes. In the case of SnVTD, diUb-3Br probe showed superior complex formation effect in comparison with diUb-2Br according to the intensity of complexes (Fig 3.8.C). Since the reaction condition was 37°C, it would be beneficial if lower temperature and shorter reaction course could be tested to prevent complexes from potential damage, such as aggregation.

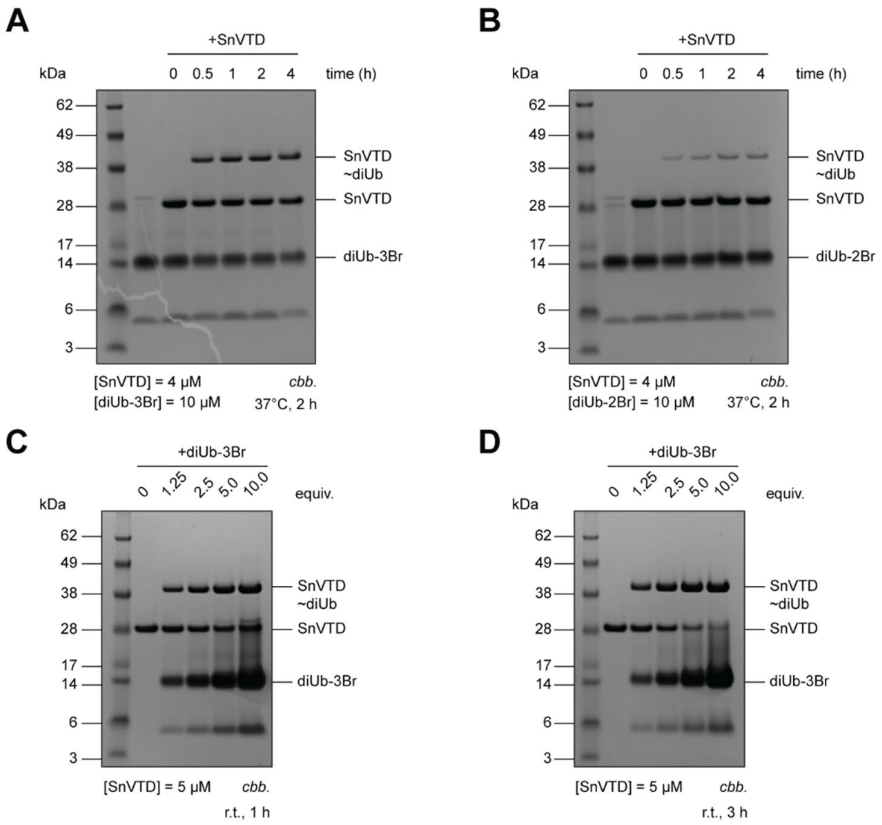


Fig 3.9. Optimization of gel-based assay for SnVTD with covalent probes. **A**. Time-dependent assay for the reaction between SnVTD with Lys6-linked diUb-3Br. **B**. Time-dependent assay for the reaction between SnVTD with Lys6-linked diUb-2Br. **C**. Excessive diUb-3Br mixed with SnVTD to increase the conversion of SnVTD into complex. The reactions were performed for 1 hour at room temperature. **D**. Excessive diUb-3Br incubated with SnVTD for 3 hours at room temperature. Reaction conditions are shown below each gel.

To test the reaction time, 2.5-fold equivalent of diUb-3Br probe was firstly incubated with SnVTD at 37°C for two hours (Fig 3.9.A). The probe reacted with SnVTD in a time-dependent manner. Strikingly, incubation for longer period, such as four hours led to significantly more complex product in comparison to 30 minutes. Time-dependent assay was performed as well for diUb-2Br probe using the same reaction condition and stoichiometry (Fig 3.9.B). Although diUb-2Br reacted with SnVTD in a time-dependent manner, the product formed by diUb-2Br was significantly less than diUb-3Br probe.

To further optimize the reaction condition in milder temperature to minimize potential protein misfolding and aggregation, various stoichiometry of diUb-3Br probe was mixed with 5 μM SnVTD. Gel analysis showed that ten equivalent of diUb probe led to the most conversion of SnVTD into covalent complex (Fig 3.9.C). To further boost the conversion, longer incubation time was tested (Fig 3.9.D). Gel analysis of the reaction of diUb-3Br with SnVTD showed increasing incubation time with higher stoichiometry could lead to almost complete conversion.

In general, the reaction was optimized by varying several parameters, such as reaction temperature, stoichiometry as well as incubation period. Gel-based readouts suggested that stoichiometry plays the most important role in achieving ideal conversion of apo SnVTD into covalent complex even the reaction was performed at room temperature (Fig 3.9.D). High stoichiometry might not be feasible since excessive probes might bind to the DUB in a non-covalent manner to interfere with the purification. Moreover, these probes contained a chiral center which explained that some probes could not form covalent complexes with SnVTD even after incubation for longer period.

3.7 Capturing SnVTD with DiUb-3Br

The effort for the scouting for DUBs with dedicated linkage specificity using newly designed probes with novel alkyl bromide warheads led to the visualization of Lys6 linkage preference of SnVTD in the gel-based assay format. These probes showed mild reactivity with SnVTD, but they were inactive against ZUFSP and its yeast homolog Mug105. Since SnVTD was successfully trapped with the diUb-2Br and diUb-3Br probes, structural understanding of its linkage preference would be possible.

VTD-type DUBs as the most recently discovered DUB family from a bioinformatic screen are broadly distributed in eukaryotic and bacterial organisms[110] (Fig 3.10.A). VTD DUBs are classified into four subclasses according to their origins. Some of VTDs possess preference toward linkage types, among which Lys6 specificity is a unique activity since DUBs encoded in human genome do not have such dedicated linkage specificity. In eukaryotes, OTUD3 has been validated with capability of cleaving Lys6 chain as well as Lys11 linkage[123]. USP30 also cleaves Lys6 chains more selectively than other chain types but not as specific as WcVTD and SnVTD[119]. Therefore, elaboration of their unique Lys6-linkage specificity might stimulate further investigation into relevant infection biology.

Among bacterial DUBs, the bacterial effector LotA with two OTU domains from *Legionella pneumophila* was characterized as an exclusively Lys6-specific DUB[225]. Structural determination of LotA in apo form as well as catalytically inactive LotA in complex with Lys6-linked diUb substrate revealed a substrate-assisted catalysis mechanism to cleave Lys6-linked polyubiquitin chains specifically[215]. The characterization of LotA provided not only a comprehensive insight into its functions during pathogen infection, but also a highly specific tool for biochemical research.

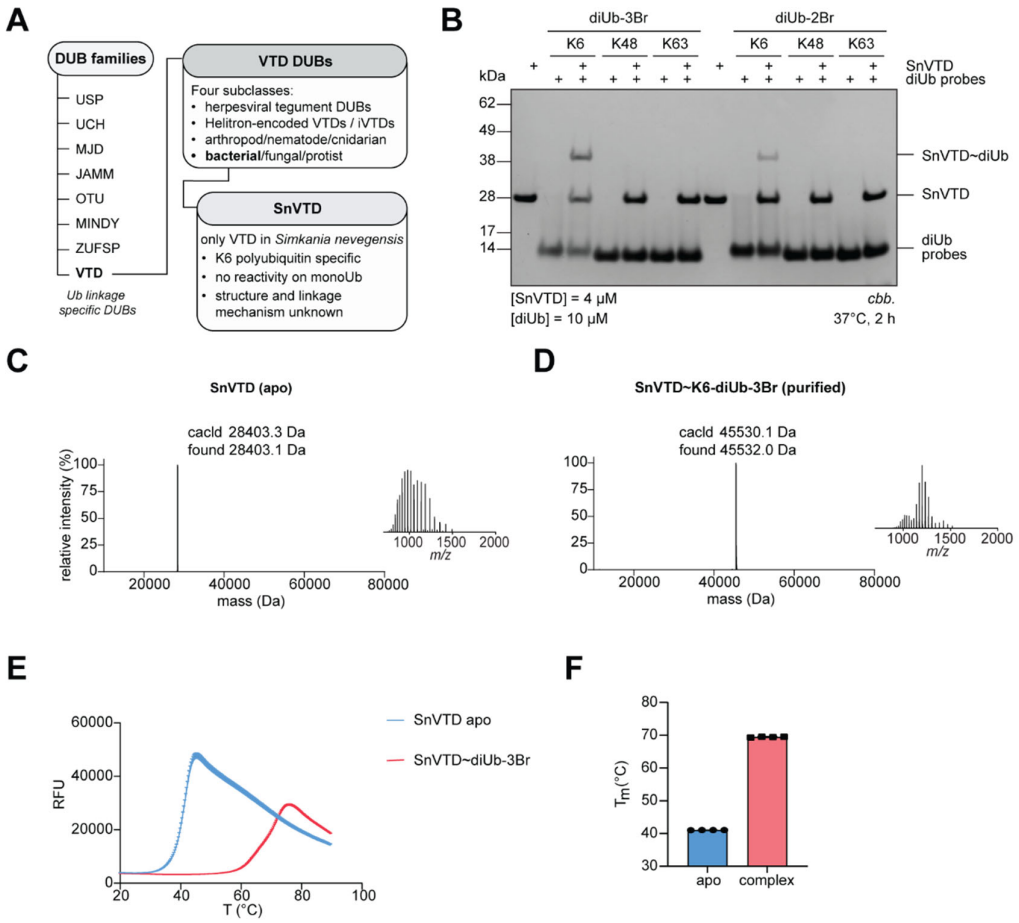


Fig 3.10. Characterization of the covalent complex of SnVTD with diUb-3Br probe. **A**. Overview of DUB families and the features of newly identified VTD DUBs. **B**. SnVTD was tested with the panel of alkyl-bromide-based covalent diUb probes to examine its linkage specificity. Reaction details are showed under the gel. **C**. Deconvoluted mass spectrum of apo SnVTD. Theoretical and observed mass are shown. **D**. Deconvoluted mass spectrum of purified SnVTD in covalent complex with Lys6-linked diUb-3Br. Raw *m/z* data are shown as inserts. **E**. Melting curves of thermal shift assay for apo SnVTD and purified SnVTD in covalent complex with Lys6-linked diUb-3Br. **F**. Melting temperature of apo SnVTD and purified SnVTD~diUb-3Br complex.

It is hypothesized that Lys6 linkage specificity of VTD DUBs might contribute to the defense systems which still require more evidence[110]. Since SnVTD showed decent reactivity toward the alkyl-bromide-based probes, structural characterization of SnVTD in covalent complex could provide new insights and understandings into the linkage-related cellular functions. Previous cleavage assay using a full panel of diUb substrates with eight linkages validated the specificity[111]. It was not clear that how the alkyl-bromide-based probes behave after long-term incubation with SnVTD. To test if SnVTD specifically reacts with Lys6-linked probes, SnVTD was incubated individually with each probe at 37°C for two hours (Fig 3.10.B). Gel analysis validated that SnVTD only reacted with Lys6-linked probes and also confirmed the mild reactivities of alkyl-bromide-based probes.

Subsequently, intact protein mass spectrometry was employed to record the mass of apo SnVTD where the observed value was consistent with the theoretical one (Fig 3.10.C). The apo SnVTD was incubated with excessive diUb-3Br probe to form covalent complex. After several rounds of fine purification using gel filtration and ion exchange chromatography, the complex of SnVTD with diUb-3Br was successfully purified. Intact protein mass spectrometry validated the complex with correct mass as well as the purity (Fig 3.10.D). Both raw data and deconvoluted spectra confirmed the purity of the covalent complex which was a positive indicator for crystallization because of the sample homogeneity.

Given that a stabilized complex can be more easily crystallized, thermal shift assay based on the folding as indicated by the fluorescence intensity changes was employed to evaluate the protein stabilization effect in comparison to apo SnVTD (Fig 3.10.E). Melting curve of SnVTD~diUb-3Br showed significantly right-shifted effect as it was more stable than the apo form. Quantitative analysis of the melting curves showed striking stabilization effect of diUb-3Br binding to SnVTD (Fig 3.10.F). The melting temperature of the covalent complex showed nearly thirty degrees increasement compared to the apo SnVTD which is beneficial for protein crystallization.

3.8 Structural Basis for the Lys6 Specificity of SnVTD

In order to fully reveal the mechanism underlying the catalytic specificity of SnVTD toward Lys6-linked polyubiquitin chains, purified covalent complex of SnVTD with diUb-3Br in large quantity was subjected to sitting-drop crystallization in a 96-well format using twelve commercially available screening kits. Theoretically, the complex of SnVTD with diUb-2Br resembles the reaction intermediate during chain cleavage since diUb-3Br contains an extra carbon atom. However, preparation of diUb-3Br is more convenient due to less hydrolysis of the covalent warhead during the reaction. Moreover, aforementioned gel-based assay showed that diUb-3Br reacts with SnVTD more efficiently than diUb-2Br. Therefore, diUb-3Br instead of diUb-2Br was selected to generate the covalent complex.

The purification after the reaction was initially performed on size-exclusion chromatography. Unfortunately, this procedure led to significant loss of sample. In addition, free SnVTD can also bind to diUb impurities to form non-covalent complexes which cannot be separated from the covalent complex. Fortunately, anion exchange chromatography was successfully used to purify the covalent complex with a slowly increased elution gradient.

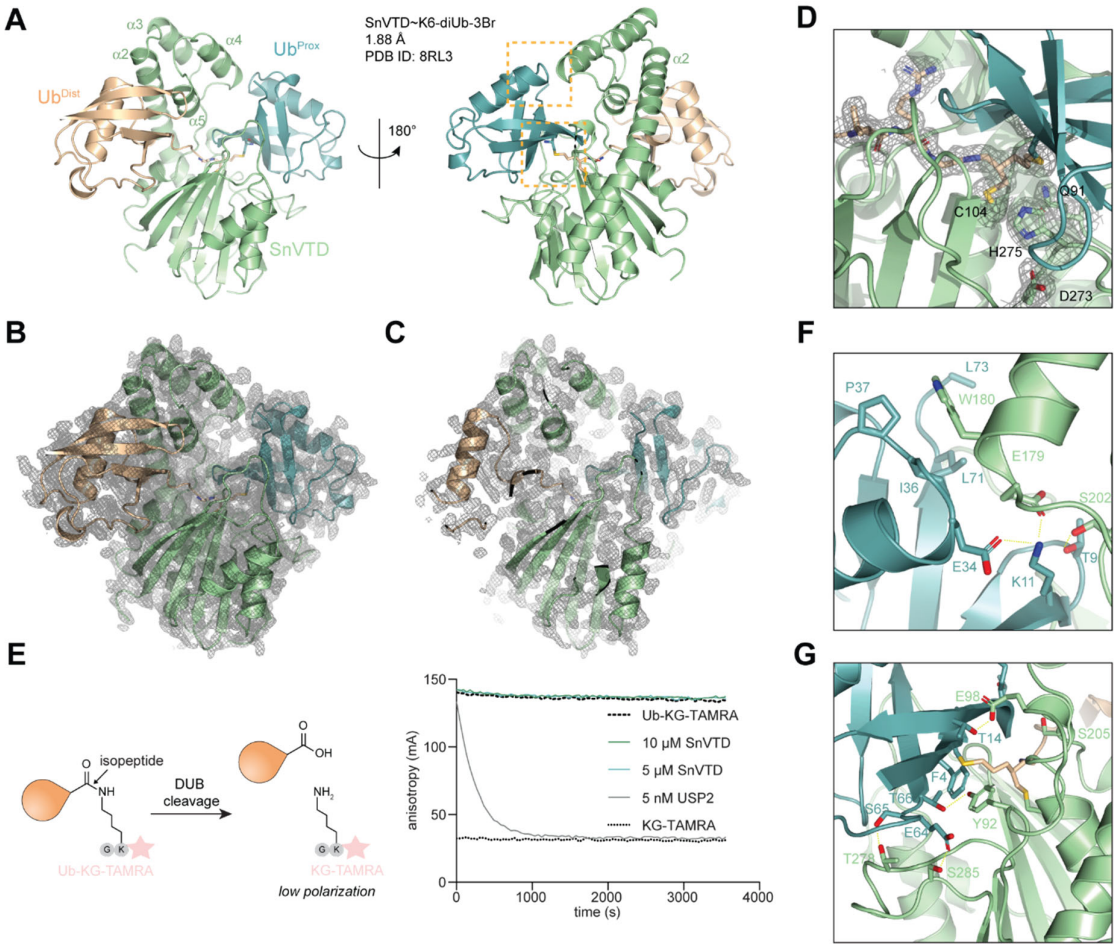


Fig 3.11. High-resolution crystal structure of SnVTD in covalent complex with Lys6-linked diUb-3Br probe (PDB ID: 8RL3). **A**. Overview of the complex structure. SnVTD is shown in light green, distal ubiquitin in wheat and proximal ubiquitin in cyan. Some secondary structures of SnVTD are annotated according to the sequence. Interacting sites between SnVTD and proximal ubiquitin are highlighted in orange boxes. **B**. $2mF_O$-dF_C map of overall complex structure. **C**. 2mFo-DFc map of the cross section of the complex. **D**. Electron density map of the catalytic triad residues of SnVTD, the oxyanion hole residue and C-terminal residues of distal ubiquitin. All the electron density maps are contoured at 1.0 σ. **E**. Schematic illustration of fluorescence polarization assay. USP2 was used as positive control. The final concentration of Ub-KG-TAMRA was 100 nM. **F**. Close-up view of interaction between SnVTD and the TEK-box and I36 hydrophobic patch of proximal ubiquitin. Hydrogen bonds are shown as yellow dashed lines. **G**. Close-up view of interaction around the π-π stacking formed by the Y92 residue in SnVTD and F4 residue in the proximal ubiquitin.

The twelve screening plates provided multiple crystals from various buffer conditions which have precipitants including PEG300, PEG400, PEG2000 and PEG Monomethyl Ether 550. Notably, four conditions contain PEG400. Crystals from the buffer containing PEG Monomethyl Ether 550 were significantly larger than other crystals. However, all crystals found in the initial screening were either sea urchin-like or thin plates intergrown with each other. Fine screening was then set up based on the conditions giving crystals. Unfortunately, none of the crystals from fine screening gave good diffraction results. Further analysis of the obtained datasets informed crystal twinning results.

Given that the resolution was low and the search model from the AlphaFold was not accurate enough, further processing of the data would be very challenging. Therefore, the initial screening plates were manually examined again to check if there were new crystals which grew slowly. After around seven weeks, some new crystals were found from the Morpheus screening plate[226]. These crystals were collected for X-ray diffraction. Although the crystals were fished directly from the initial screen without optimization, two of them diffracted well. A search model from AlphaFold with manual chimerization based on the structure of previously reported structure of herpesvirus-encoded cysteine protease M48 (PDB ID: 2J7Q) was prepared and used for phasing[227, 228]. The best diffracted crystal was solved and refined to achieve optimal statistics (Appendix Table 2).

Finally, the crystal structure of SnVTD in covalent complex with diUb-3Br probe was determined at a resolution of 1.88 Å (Fig 3.11.A). The secondary structural elements were analyzed and annotated for clarity (Appendix Fig 3.A&B). The C-terminal part of SnVTD which contains a transmembrane domain was not included due to the difficulty in expression. Residues resolved in the crystal structure were highlighted in black (Appendix Fig 3.C). The structure of SnVTD contains multiple α helices, among which α2 helix is the longest one consisting of 38 amino acids (residues 122-159) as the core backbone of the enzyme. SnVTD is significantly different from the general features of USPs whose structures usually resemble a right hand. Followed by α2 helix, a cluster of three short α helices form a cap-like structure. In the complex structure, SnVTD was found to interact intensively with the proximal ubiquitin as highlighted in the orange boxes (Fig 3.11.A). The bottom part of SnVTD consists a core structure formed by four β sheet elements. Since the structure has high resolution, the electron density map of the overall structure showed good coverage (Fig 3.11.B). Moreover, the electron density map of the cross section also showed good quality which further confirmed the quality of the complex structure (Fig 3.11.C). Close-up view of the catalytic center of SnVTD showed that the catalytic residue Cys104 formed a covalent thioether bond with the probe by substituting the bromide warhead (Fig 3.11.D). Adjacent residues which formed the catalytic site were highlighted. The side chain of Gln91 is shown which is involved in forming an oxyanion hole for catalytic cleavage. According to the electron density map, all the catalytic residues as well as the C-terminal tail in the distal ubiquitin were well defined. Notably, the probe should contain a mixture of enantiomers due to the chiral center in the linkage. In the refinement process, the *S*-enantiomer was fitted into the structure empirically and covalently linked with SnVTD and two ubiquitin molecules based on the orientation and continuity of the density map.

SnVTD was previously characterized as a DUB by cleaving isopeptide-linked diUb substrate but not active against suicide probe Ub-PA and fluorogenic substrate Ub-AMC[111]. Given that neither Ub-PA or Ub-AMC is functionalized through a natural isopeptide bond, it is speculated that bond type might contribute to the resistance of Ub-PA and Ub-AMC. To investigate if the bond types of substrates could affect the activity of SnVTD, isopeptide-linked fluorescence polarization substrate Ub-KG-TAMRA was employed. USP2 was selected as positive control enzyme which cleaved Ub-KG-TAMRA completely and led to the full conversion to reach the anisotropy level of free KG-TAMRA (Fig 3.11.E). Strikingly, SnVTD at either 10 or 5 μM concentration could not cleave any Ub-KG-TAMRA substrate in 1 hour incubation. This assay proved that bond types are not the dominant factor to affect the deubiquitinating activity of SnVTD. Occupation of the distal ubiquitin binding site in SnVTD could not activate the deubiquitinating activity. There should be a mechanism in the proximal ubiquitin binding site in SnVTD to initiate the cleavage.

Analysis of the interaction of SnVTD with proximal ubiquitin has led to the discovery of two major binding sites as shown in the orange boxes (Fig 3.11.A). The close-up view of the upper orange box led to identification of an important hydrophobic binding site which is formed by the well-known Ile36 hydrophobic patch with structurally bulky Trp180 in SnVTD (Fig 3.11.F). In addition, three other hydrophobic residues in the proximal ubiquitin, Pro37, Leu71 and Leu73 also contribute to form a large hydrophobic pocket in coordination with Ile36. Besides the hydrophobic interaction for molecular recognition, adjacent residues Glu179 and Ser202 are actively involved in hydrogen bonding with the TEK-box of ubiquitin (Fig 3.11.F). Glu179 forms a hydrogen bond with the amine group on the side chain of Lys11 and Ser202 interacts with the hydroxyl group of Thr9 which could boost the binding affinity. The Gln34 in the TEK-box also contributes to the binding affinity through forming a hydrogen bond with Lys11 to stabilized the long side chain of Lys11 to boost the hydrogen bond with Glu179.

In the lower orange box (Fig 3.11.A), the core interaction is a hydrophobic π-π stacking formed by the Phe4 in ubiquitin and Tyr92 in SnVTD (Fig 3.11.G). To further enhance the π-π stacking, the hydroxyl side chain of Thr66 in ubiquitin interacts with the phenolic hydroxyl through a hydrogen bond. As a typical hydrophobic patch for interaction, Phe4 patch usually consists of Thr14 residue. In the complex structure, Thr14 contributes to the interaction through a hydrogen bond with Glu98 in SnVTD. Additionally, Glu64 in ubiquitin which is also close to Tyr92 forms a hydrogen bond with Ser285. Ser65 in ubiquitin also interacts with Thr278 through a hydrogen bond.

Fluorescence polarization assays proved that bond type is not the reason for the resistance of mono-Ub-based reagents toward SnVTD deubiquitinating activity. In-depth analysis of the high-resolution crystal structure of SnVTD with Lys6-linked diUb suicide substrate revealed several key interaction sites that might be responsible for the explanation why SnVTD cleaves Lys6-linked chains specifically. Two major hydrophobic interaction sites could be the determinant factors that induce SnVTD to bind to Lys6-linked polyubiquitin chains and execute cleavage. The Ile36 and Phe4 patches, which are both well-characterized hydrophobic patches in ubiquitin structure, play important roles in the molecular recognition of proximal ubiquitin by SnVTD in the solvated state. The classic Ile36 patch is usually composed of three residues, namely Ile36, Leu71 and Leu73 as the canonical cluster[229]. However, structural analysis of SnVTD and proximal ubiquitin identified an additional hydrophobic residue Pro37 might be also involved in the hydrophobic pocket to accommodate Trp180 in SnVTD. In some cases, Leu8 also participates in the Ile36 patch to create an expanded hydrophobic pocket. For instance, Leu8 acts as one of the recognition element for USP36 to interact with ubiquitin[66]. However, this residue was not involved in the accommodation of SnVTD. This giant hydrophobic interaction is enhanced by adjacent hydrogen bonds contributed by the TEK-box consisting of residues Lys6, Lys11, Thr12, Thr14 and Glu34. Since Lys6 was used for forging isopeptide bond, it thereby could not contribute to the interaction. Nevertheless, Lys11, Thr14 and Glu34 formed multiple hydrogen bonds both intermolecularly and intramolecularly to enhance the interaction. Interestingly, Phe4 patch employed a mechanism different from the completely hydrophobic interaction orchestrated by four hydrophobic residues in the Ile36 patch. Phe4 and Tyr92 interact with each other through a π-π stacking which is also strengthened by a hydrogen bond. Apart from this hydrophobic stacking, the rest of interaction nearby existed as hydrogen bonds. In the canonical Phe4 patch, Gln2 is an important component which however was not engaged in binding to SnVTD.

In general, the proximal ubiquitin utilized two classic hydrophobic patches as well as some adjacent hydrogen bonds for enhancing the binding affinity to accommodate SnVTD specifically. Induced by the proximal ubiquitin in the specific orientation, SnVTD would be able to precisely bind to Lys6-linked chains instead of other seven linkages and then implement its enzymatic cleavage. This behavior for SnVTD to recognize Lys6-linked polyubiquitin explains why SnVTD showed no any activity toward mono-Ub-based reagents. Small molecule modifications at the C-terminal of ubiquitin could not either form hydrophobic interaction and π-π stacking or induce conformation change to initiate cleavage.

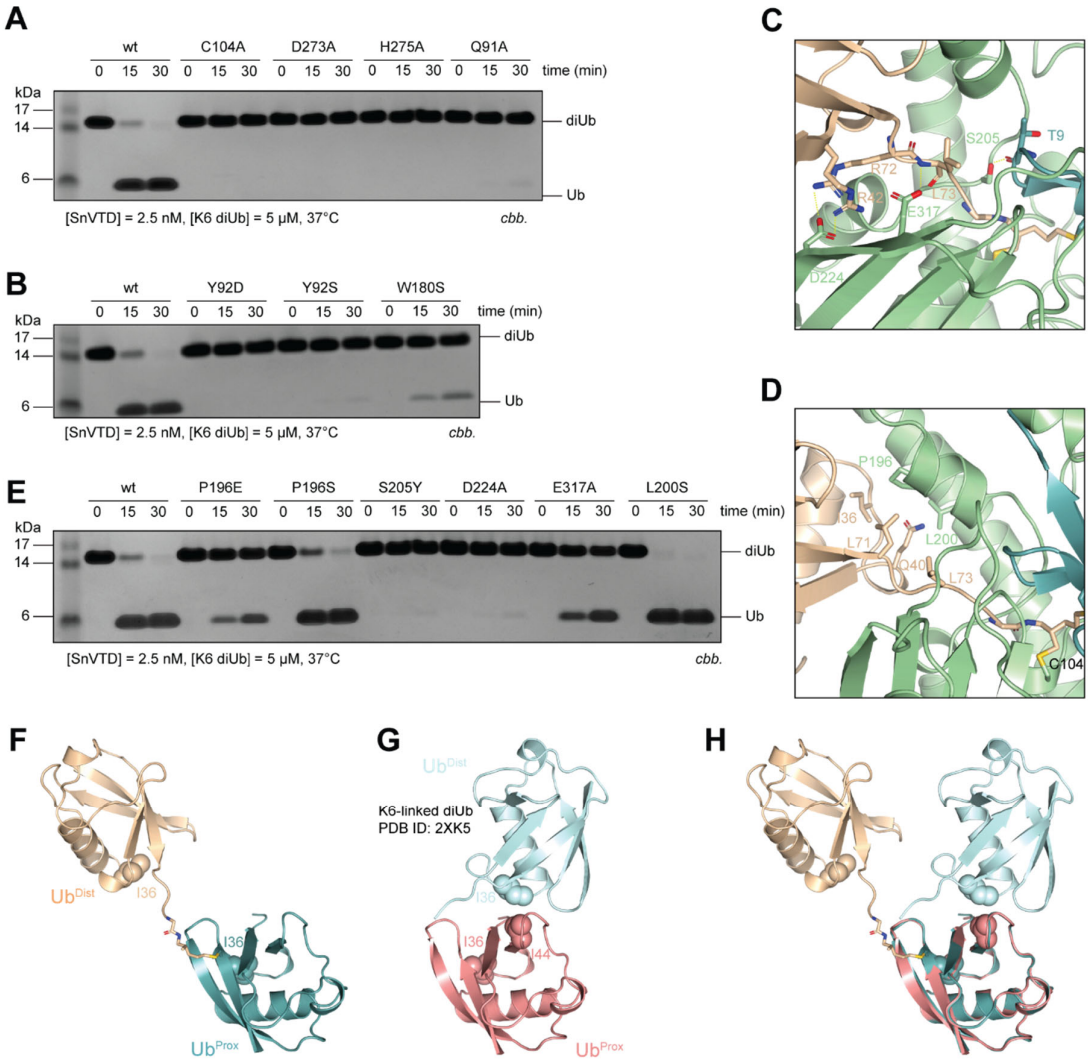


Fig 3.12. Mutagenesis studies to elaborate the Lys6 linkage specificity of SnVTD using time-dependent gel-based cleavage assay with Lys6-linked diUb as the substrate. **A**. Mutations in the catalytic triad and oxyanion hole residues. **B**. Mutations in the residues for binding to proximal ubiquitin. **C**. Close-up view of the interaction of SnVTD and the C-terminal tail of distal ubiquitin. **D**. Close-up view of the hydrophobic interaction formed by SnVTD and I36 patch in distal ubiquitin. **E**. Mutations in the residues which are responsible for the interaction of I36 patch and hydrogen bonding in distal ubiquitin. **F**. Structure of diUb isolated from SnVTD~diUb-3Br complex. **G**. Crystal structure of native isopeptide-linked Lys6 diUb. **H**. Superposition of structures shown in **F** and **G**. Key residues involved in hydrophobic interactions are shown in spheres and annotated.

Structural analysis of the catalytic sites as well as proximal ubiquitin binding site revealed some key residues might be involved in determining the catalytic activity. Although VTD DUBs from various origins showed different behavior and preference toward ubiquitin chains, sequence alignment of SnVTD with M48 and other previously characterized VTDs is consistent with the structural analysis of the catalytic triad residues (Appendix Fig 4). Moreover, Gln91 is also highly conserved in the bioinformatic analysis which is speculated to act as an oxyanion hole to facilitate the hydrolysis of ubiquitin chains.

To test the hypothesis, plate-based readouts, such as Ub-RhoG and Ub-KG-TAMRA, are not suitable because SnVTD does not cleave them and thereby will not give any signal for interpretation. To overcome this drawback but also achieve continuous readout, fluorescence resonance energy transfer (FRET) assay using Lys6-linked diUb substrate of which proximal and distal ubiquitin molecules are labeled with a pair of FRET donor and acceptor fluorophores respectively might offer another option for quantitative analysis[230]. However, such FRET substrates include two additional fluorophores which might be interfered with enzymatic cleavage to hinder proper data interpretation.

The cleavage assay using natural Lys6-linked diUb as substrate was selected. Previous effort has led to the assembly of Lys6-linked diUb protein by E3 ligase NleL[102]. The gel-based assay could be used for validating the key residues in SnVTD by introducing corresponding mutations. Although the gel-based readout is not as continuous as the plate-based assays, it still provides straightforward comparison of the effects on the cleavage activity between different mutants.

According to the aforementioned structural analysis as well as sequence alignment of various VTDs, all three putative catalytic triads as well as Gln91 were mutated to alanine. Lys6-linked diUb substrate was mixed with wildtype and mutant SnVTDs individually at 37°C. The reactions were then analyzed at three time points. Coomassie stains of the gel showed that wildtype SnVTD cleaved diUb efficiently in a time-dependent manner while all the four mutants could not cleave substrate any more since their catalytic sites were destroyed (Fig 3.12.A). The experimental data were consistent with the structural analysis and bioinformatic alignment.

After successful validation of the residues in the catalytic site, hydrophobic interaction between Trp180 residue in SnVTD and Ile36 patch of the proximal ubiquitin was studied by replacing tryptophan with serine which is not only significantly less bulky than tryptophan, but also contains a hydrophilic hydroxyl to abolish the hydrophobic interaction. Apart from the hydrophobic Trp180 residue, Tyr92 which serves as a recognition element for specific binding to Lys6-linked diUb through π-π stacking was substituted with a hydrophilic aspartate or milder serine residue. In the same reaction condition for studying catalytic mutants, mutant SnVTDs bearing Y92D, Y92S or W180S individually were incubated with diUb substrate. Strikingly, single mutation on either position 92 or 180 can abolish the deubiquitinating activity of SnVTD which validated the importance of these two hydrophobic residues for recognizing and cleaving Lys6-linked polyubiquitin chains specifically in line with the aforementioned structural analysis (Fig 3.12.B). Although the roles of Lys6 linkage specificity of SnVTD in infection biology remain

poorly understood, the key residues Tyr92 and Trp180 are the hotspots which might be specifically targeted by potential small molecule inhibitors. These small molecule inhibitors would act as versatile tools since genetic interruption might not be feasible in an infection model.

After structural and experimental validation of two key residues involved in molecular recognition at the proximal ubiquitin binding site, the distal ubiquitin was analyzed as well to find other key interactions which might promote the linkage preference of SnVTD. The distal ubiquitin interacts intensively with SnVTD through its long C-terminal tail which enters into the catalytic center for cleavage upon recognition of the proximal ubiquitin by SnVTD (Fig 3.12.C). Surrounding the C-terminal tail, two acidic residues in SnVTD are involved in strong interaction with ubiquitin through hydrogen bonds. The carboxylic acid moiety on the Asp224 side chain coordinates with two positively charged Arg42 and Arg72 residues through two hydrogen bonds (Fig 3.12.C). The other acidic residue Glu317 does not form hydrogen bond with any moieties on side chain, but with the backbone amide moiety of Leu73 to enhance the binding affinity. Interestingly, Ser205 which is spatially close to the C-terminal glycine interacts with the main chain of Thr9 in the proximal ubiquitin through a hydrogen bond (Fig 3.12.C). This hydrogen bond reorients the hydroxyl group in Ser205 which might spare more space for accommodating ubiquitin C-termini.

Besides hydrogen bonds at the interface of SnVTD and distal ubiquitin, the canonical patch formed by Ile36, Leu71 and Leu73 residues in ubiquitin orchestrates a large hydrophobic pocket which hosts the hydrophobic residue Pro196 in SnVTD (Fig 3.12.D). Another hydrophobic residue Leu200 which is adjacent to Pro196 is not involved in the hydrophobic interaction because it is hindered by the hydrophilic side chain on Gln40 in the distal ubiquitin (Fig 3.12.D).

The hydrophobic interaction between SnVTD and distal ubiquitin is weaker than what has been observed in the proximal ubiquitin binding site. In the distal ubiquitin binding site, only one proline interacts with the Ile36 patch while in the proximal ubiquitin binding site, planar and rigid tryptophan residue plays a key role in the hydrophobic interaction with the Ile36 patch. Moreover, Tyr92 as another planar and rigid residue also makes contribution to the specific recognition of Lys6-linked chains through a hydrophobic π-π stacking. Therefore, it is speculated that the hydrophobic interactions in the proximal ubiquitin binding sites are responsible for the specific recognition while hydrophobic interaction in the distal site affects the catalytic cleavage.

On the basis of the structural analysis of the interface between SnVTD and the distal ubiquitin, multiple mutations were introduced into SnVTD. Followed by the established Lys6-linked diUb cleavage assay, deubiquitinating activities of these mutant SnVTD enzymes were investigated (Fig 3.12.E). Firstly, the hydrophobic interaction site was inspected. Pro196 in SnVTD was substituted with a glutamic acid residue of which the hydrophilic carboxylic acid side chain was proposed to disrupt the hydrophobic pocket. In comparison to wildtype, P196E mutant lost its cleaving activity almost completely. A milder mutation was introduced at the same position, namely P196S. Strikingly, this mutation did not lower the enzymatic activity significantly which indicates that this mutation is tolerable for SnVTD. The same serine residue was introduced to Leu200 which is slightly far away from the Ile36 patch. Interestingly, L200S mutant cleaved diUb substrate efficiently which might be attributed to the new hydrogen bonding formed between the newly introduced serine and Gln40 in the distal ubiquitin. This new hydrogen bond could boost the binding affinity and thereby maintain the cleaving activity. Two acidic residues Asp224 and Glu317 were replaced with an alanine respectively. Consequently, the deubiquitinating activity of D224A mutant was completely abolished. while E317A mutant was nearly inactive. Interestingly, SnVTD completely lost cleaving activity upon mutagenesis occurred to Ser205 with a bulky tyrosine residue. There might be two major reason to explain why SnVTD lost activity with the installation of S205Y mutation. Introduction of the tyrosine residue could lead to the disassociation of the hydrogen bond between Ser205 and Thr9, though tyrosine has a phenolic hydroxyl but the bulky phenyl ring might conflict with the Thr9 loop. The second explanation is that bulky side chain could prevent the Lys6-linked isopeptide bond from entering into the catalytic center. The proximity between isopeptide bond and catalytic cysteine was hindered by the introduction of a bulky residue which led to the fact that SnVTD would lose activity completely.

The interface between SnVTD and diUb was studied carefully with the combination of structural analysis and subsequent mutagenesis-based biochemical assays. Besides the close-up view of how SnVTD interacts with Lys6-linked diUb with dedicated specificity, the overall structural features of the conformation changes of diUb were also depicted (Fig 3.12.F). The Ile36 patches in each ubiquitin were both highlighted in spheres which showed a stretched-off conformation compared to Lys6-linked diUb which was crystallized alone[102] (Fig 3.12.G). Since SnVTD requires the proximal ubiquitin for activation to initiate the cleavage, proximal ubiquitin molecules were fixed and these two structures were aligned based on the proximal ubiquitin molecules (Fig 3.12.H). The superposition featured a substantial rotation of the distal

ubiquitin upon binding to SnVTD. However, in the crystal structure of native Lys6-linked diUb, Ile44 patch in the proximal ubiquitin and Ile36 patch in the distal ubiquitin interact with each other through the exclusive hydrophobic interaction in the solvated state. On the basis of this hydrophobic interaction, Lys6-linked diUb displayed a compact conformation in its crystal structure (Fig 3.12.G).

In contradictory to the compact diUb, diUb from the complex structure with SnVTD showed an open conformation with a stark structural shift. The hydrophobic Ile36 patch in the distal ubiquitin was not used for interacting with the proximal ubiquitin, but for accommodating the hydrophobic Pro196 in SnVTD. When the open conformation was achieved, the isopeptide bond could be finally accessible by the catalytic center for further processing. Such huge conformational changes in the diUb chains have been previously observed in other DUBs, such LotA and USP30[119, 215]. Crystal structure of LotA bound to Lys6-linked diUb showed that Phe4 patch in the proximal ubiquitin was involved in a hydrophobic interaction but not through a hydrophobic π-π stacking. Moreover, USP30 interacts with Phe4 patch through several hydrophobic residues, including a planar and rigid tryptophan, but does not form a π-π stacking. The feature that Tyr92 in SnVTD forms π-π stacking interaction with Phe4 patch in ubiquitin might be the fundamental reason to explain why SnVTD favors Lys6-lnked polyubiquitin chains but shows no any activity toward Ub-PA, Ub-AMC and Ub-KG-TAMRA due to lack of driven force by Phe4 patch. In comparison to LotA and USP30, SnVTD requires stringent π-π stacking through a planar residue to ignite its chain-cleaving activity. Unlike the P196S mutant which still retained the activity of SnVTD, milder mutation with a serine on Tyr92 has a fatal effect on the activity of SnVTD. Overall, the two hydrophobic interaction sites at the proximal ubiquitin binding interface constitute of dual criteria for substrate selection to assure the specificity.

Bioinformatic analysis of VTD sequences from various species can offer accurate prediction of residues assembling the catalytic center, but detailed information for explaining their enzymatic activation mechanism and substrate specificity is still not available from the primary sequences. Structural determination of VTD DUBs in complex with their substrates, either solely distal ubiquitin, or both distal and proximal ubiquitin molecules would provide panoramic analysis of interaction for subsequent experimental validation. Although AlphaFold provided predicted models, the accuracy of models remains challenging for solving phasing in structural determination of some VTDs. Therefore, initial efforts to solve the structures of VTDs utilized selenomethionine incorporation[110].

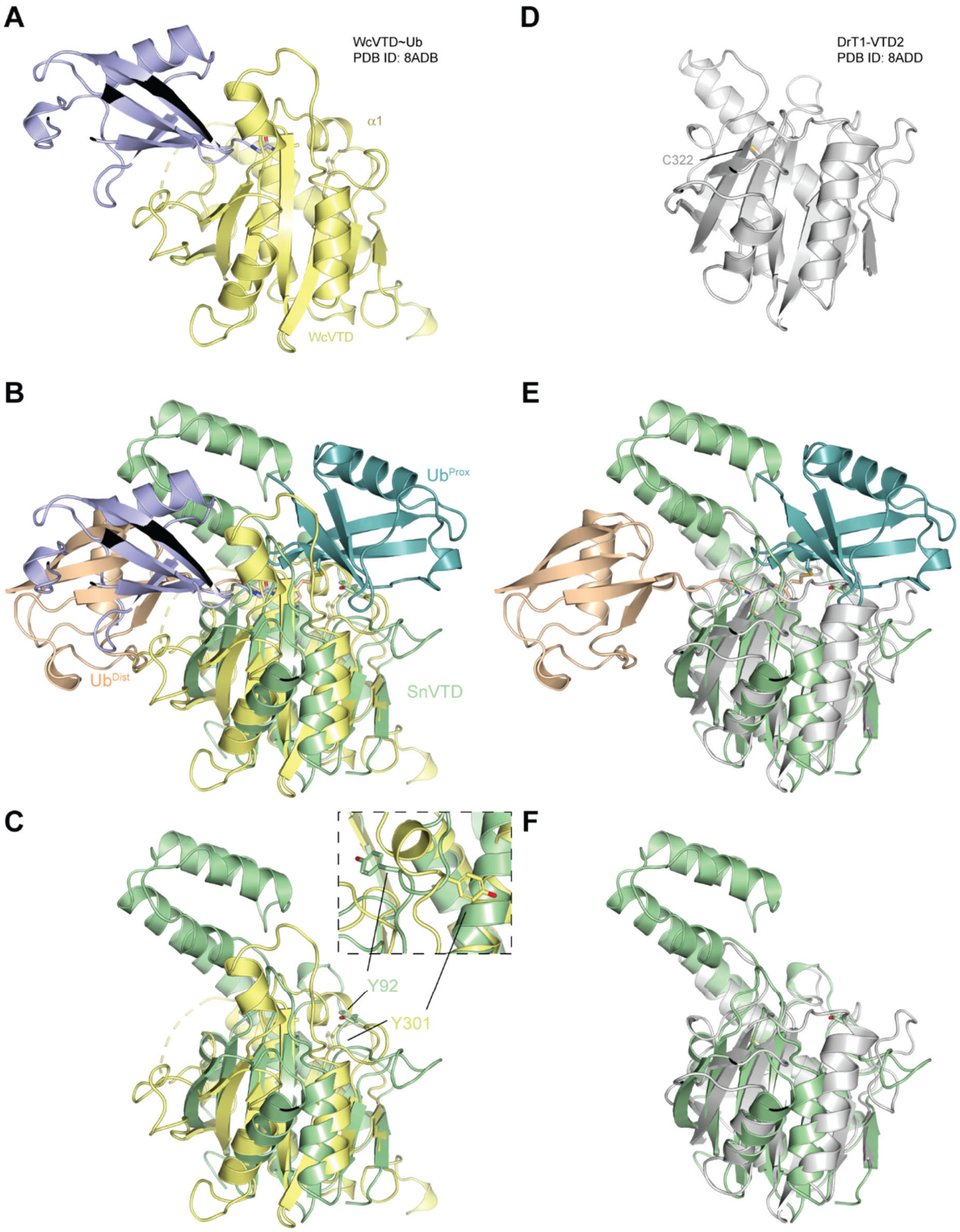


Fig 3.13. Structural alignment of SnVTD and previously reported WcVTD in covalent complex with Ub-PA and DrT1-VTD2 in apo form. **A**. Structural overview of WcVTD~Ub (PDB ID: 8ADB), of which WcVTD is shown in yellow. α1 helix is annotated. **B**. Superposition of SnVTD~diUb and WcVTD~Ub. **C**. Superposition of SnVTD~diUb and WcVTD~Ub where ubiquitin molecules are omitted for clarity. Y92 in SnVTD and equivalent Y301 in WcVTD are highlighted. **D**. Crystal structure of apo DrT1-VTD2 (PDB ID: 8ADD) which is shown in grey. The catalytic cysteine C322 is shown as sticks. **E**. Structural alignment of SnVTD~diUb and DrT1-VTD2. **F**. Superposition of SnVTD and DrT1-VTD2 without diUb.

So far, only two VTDs have been structurally characterized, namely WcVTD (UniProt ID: D6YWY5) and DrT1-VTD2 (UniProt ID: A0A8M9QCC7) in apo form. WcVTD is specific to Lys6 linkage but possesses slightly weak activity toward Lys11-linked chains[110]. DrT1-VTD2 is a recently identified DUB found in *Danio rerio*. As a eukaryotic DUB, DrT1-VTD2 cleaves polyubiquitin chains with preference toward Lys11, Ly48 and Lys63 linkages[110].

The apo and ubiquitin-bound structures of WcVTD have been both solved which provided the structural basis for understanding how it interacts with distal ubiquitin. Although the structure of diUb-bound WcVTD is not available, structural comparison of apo and Ub-PA-bound WcVTD could provide some hints. It was speculated that α1 helix in WcVTD might determine the specificity since it shifts significantly upon binding of the distal ubiquitin (Fig 3.13.A). Structural alignment of WcVTD and SnVTD revealed explicitly that the long α2 helix of SnVTD coordinate a cluster of several helices while WcVTD lacks this feature (Fig 3.13.B). Since the existence of the helices, distal ubiquitin in the SnVTD structure has a stark rotation compared to the ubiquitin in WcVTD. Sequence alignment implied that Tyr92 in SnVTD is highly conserved in several VTDs. Structural overlay of SnVTD and WcVTD revealed that Tyr92-equivalent residue Tyr301 in WcVTD sits next to α1 helix. Tyr92 is in compact formation due to the interaction with proximal ubiquitin while Tyr301 is in a flexible position and points outward which might reorient to interact with Phe4 patch upon binding with Lys6 polyubiquitin chains (Fig 3.13.C). Moreover, sequence analysis also revealed a long insertion which is only owned by SnVTD and contains the hydrophobic residue Trp180 for interacting with the Ile36 patch. Due to the dual hydrophobic interactions in the proximal ubiquitin binding site, the linkage specificity of SnVTD was ensured at the expense of losing activity toward mono-Ub-based reagents. However, why WcVTD has the capability of reacting with Ub-PA but no activity against Ub-AMC substrate remains unclear. Further efforts into solving the structure of WcVTD in complex with diUb would provide convincing explanation.

The structure of DrT1-VTD2 was presented as apo form (Fig 3.13.D). DrT1-VTD2 does not have a long backbone helix from the structural overlay with SnVTD (Fig 3.13.E). When the diUb was omitted, SnVTD showed good alignment with DrT1-VTD2 besides the featured helical cluster in SnVTD (Fig 3.13.F). However, the residue equivalent to Tyr92 in DrT1-VTD2 is a serine which could not strongly interact with Phe4 and thereby lead to the loss of linkage specificity. As the first structure of VTD DUB in complex with diUb substrate, SnVTD~diUb-3Br enabled explicit analysis of the proximal and distal ubiquitin binding sites. SnVTD might be a powerful tool for studying Lys6 linkage, such as in the Ubi-CREST assay.

3.9 Revisiting VTD DUBs and Exploring E2s

The decent reactivity of alkyl-bromide-based probes has been successfully validated by SnVTD biochemically and structurally with a moderate linkage specificity. To further explore the applications of these newly designed probes, other members from VTD family could be tested though early effort to covalently trap WcVTD with alkyl-bromide-based probes failed. Given that structural variations in VTDs are highly unpredictable, gel-based assay readout could give comprehensive results of how these enzymes are interrogated by the probes. Moreover, E2 conjugating enzymes also use their catalytic cysteines to transfer activated ubiquitin from E1 activating enzymes to E3 ligases for modifying specific substrates. If these probes were active toward E2 enzymes, there will be broader applications for deep understanding of the process of ubiquitin transfer.

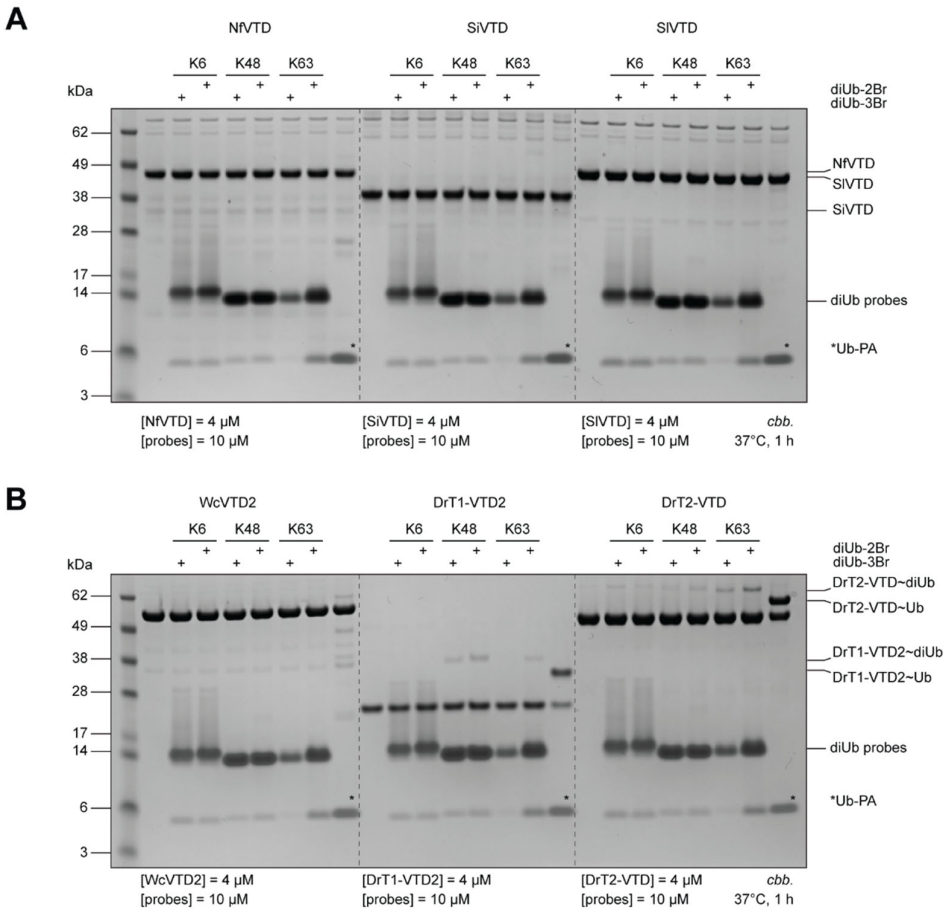


Fig 3.14. Exploration of alkyl-bromide-based probes for other VTD DUBs. **A**. Gel analysis of NfVTD, SiVTD and SlVTD incubated with the panel of diUb probes with three different linkages. **B**. Gel analysis of WcVTD2, DrT1-VTD2 and DrT2-VTD incubated with diUb probes. VTDs in covalent complex with probes are annotated on the right side. Reaction conditions are listed below the gels.

The probe panel of diUb-2Br and diUb-3Br consisting of Lys6, Lys48 and Lys63 linkages was prepared for the assay with six previously reported VTD DUBs[110]. All the VTDs are generous gifts from the Hofmann lab in Cologne. Detailed protocols for protein expression and related purification were reported previously[110].

NfVTD (UniProt ID: A0A6A5BFE9) is a protozoan DUB from *Naegleria fowleri*, which is also commonly known as brain-eating amoeba. Biochemical characterization showed that NfVTD can cleave three types of ubiquitin chains, including Lys6, Lys11 and Lys48, of which Lys6 is the most favored linkage. However, one hour incubation with probes did not show any covalent adduct formation (Fig 3.14.A). Previous bioinformatic study corroborated that NfVTD shares similarity with other two fungal VTDs, SiVTD (UniProt ID: G4TM62) from *Serendipita indica*, and SlVTD from *Serpula lacrymans* (UniProt ID: F8PCL5) since they are all Lys6-oriented VTDs[110]. In comparison with NfVTD, both SiVTD and SlVTD show even more specific preference toward Lys6-linked chains. Unfortunately, neither SiVTD or SlVTD could be covalently trapped by Lys6-linked diUb probes (Fig 3.15.A). Notably, none of NfVTD, SiVTD and SlVTD could react with Ub-PA which implied another activation mechanism involved in this cluster of VTDs different from SnVTD. This observation implied the complexity of catalytic mechanism of VTDs are not synchronized evolutionarily.

Another set of VTDs were also tested by the probes. WcVTD2 (UniProt ID: D6YRI4) was discovered from *Waddlia chondrophila* which also encodes WcVTD. WcVTD2 is biochemically similar to WcVTD with activity toward Ub-PA probe. In contrast to the Lsy6 preference of WcVTD, WcVTD2 preferentially cleaves Lys48-linked chains, with weak activity against Lys11 linkage[110]. Lys63-linked diUb probes did not form covalent adducts with WcVTD2 (Fig 3.14.B). DrT1-VTD2 (UniProt ID: A0A8M9PPM7) was discovered from zebrafish Helitron DrT1 as the second VTD domain. DrT2-VTD (UniProt ID: A0A8M9QC81) was identified from zebrafish Helitron DrT2. As Helitron-encoded VTDs, DrT1-VTD2 and DrT2-VTD not only reacted with Ub-PA and NEDD8-PA, but also cleaved Ub-AMC and NEDD8-AMC[110]. They do not have strong linkage specificity. Therefore, gel-based analysis revealed that DrT1-VTD2 can react with some probes including Lys48 and Lys63 probes but not Lys6 probes which is consistent with reported diUb cleavage results. Interestingly, DrT2-VTD can be trapped with all probes but formed small quantity of products (Fig 3.14.B). This exploration expanded the application of alkyl-bromide-based probes for characterizing the specificity of VTDs. It is foreseeable that this panel of probes could be expanded to cover all seven linkages for characterizing newly discovered DUBs.

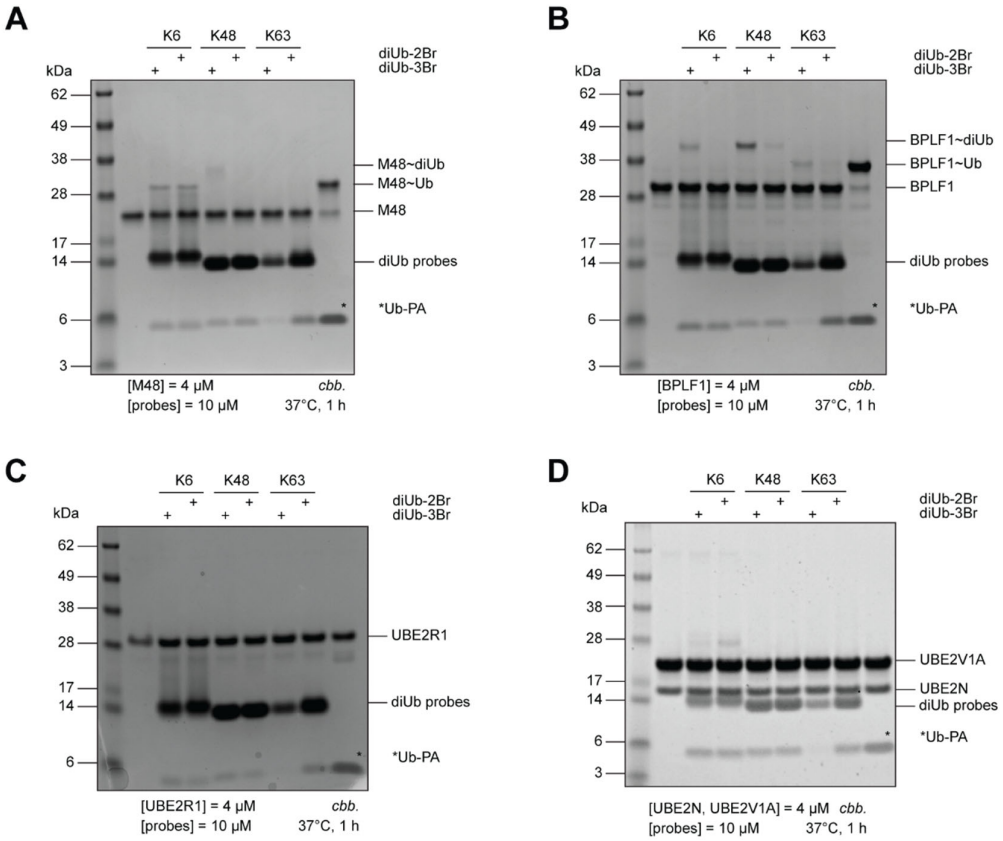


Fig 3.15. Exploring new applications of alkyl-bromide-based diUb probes on other DUBs and E2 conjugating enzymes. **A**. M48 DUB incubated with probes and analyzed on gel. Covalent adducts are annotated. **B**. BPLF1 was tested with probes. Coomassie stains visualized the results as annotated on the right side. **C**. E2 enzyme UBE2R1 was assayed with probes. **D**. UBE2V1A and UBE2N complex was tested with the covalent probes and the reactions were resolved on gel. Ub-PA as positive control is annotated with an asterisk.

Followed by the test on VTDs, DUBs from viruses were assayed with the panel of substrates using the same reaction condition. M48 is from herpesvirus and adopts a papain-like fold. Only Lys48-linked diUb-3Br showed weak activity with M48 (Fig 3.15.A). However, M48 can cleave at least Lys48 and Lys63 linkages[228]. BPLF1 from Epstein-Barr virus was incubated with the diUb panel and analyzed (Fig 3.15.B). BPLF1 was prone to form covalent adducts with Lys6 and Lys48-linked diUb-3Br probes. This phenomenon indicated that the length of linker might change the behaviors of probes and thereby both diUb-2Br and diUb-3Br should be tested to achieve optimal labelling results.

Given that E2 enzymes all have a catalytic cysteine for ubiquitin transfer, UBE2B1 was tested but did not show any covalent complex (Fig 3.15.C). The heterodimer of UBE2V1A/UBE2N was assayed as well. No adducts were observed after one hour incubation (Fig 3.15.D). The mild reactivity of the bromide-based warhead was not suitable for covalently capturing E2 enzymes. Optimization of the warhead with a better leaving group might be an option.

3.10 Beyond DUBs

The development of alkyl-bromide-based probes has led to the successful investigation into the newly identified SnVTD, which belongs to the eighth class of DUBs. So far, genome mining is still limited in several species. Deeper and more systematic genome mining might lead to discovery of previously unknown VTDs, for example from plants. The SnVTD~diUb-3Br structure together with WcVTD, WcVTD~Ub-PA and DrT1-VTD2 would act as potential search models or provide new training datasets for AlphaFold to generate more highly accurate models for solving structures of other VTDs. Apart from focusing on the investigation into cysteine-based DUBs, diUb probes might be employed to study some HECT E3s which usually have a catatlytic cysteine for transferring activated ubiquitin. Previous reported Michael-acceptor-based diUb probe has been successfully applied to study HECT E3 UBR5 which prefers to construct Lys48-linked polyubiquitin chains[190].

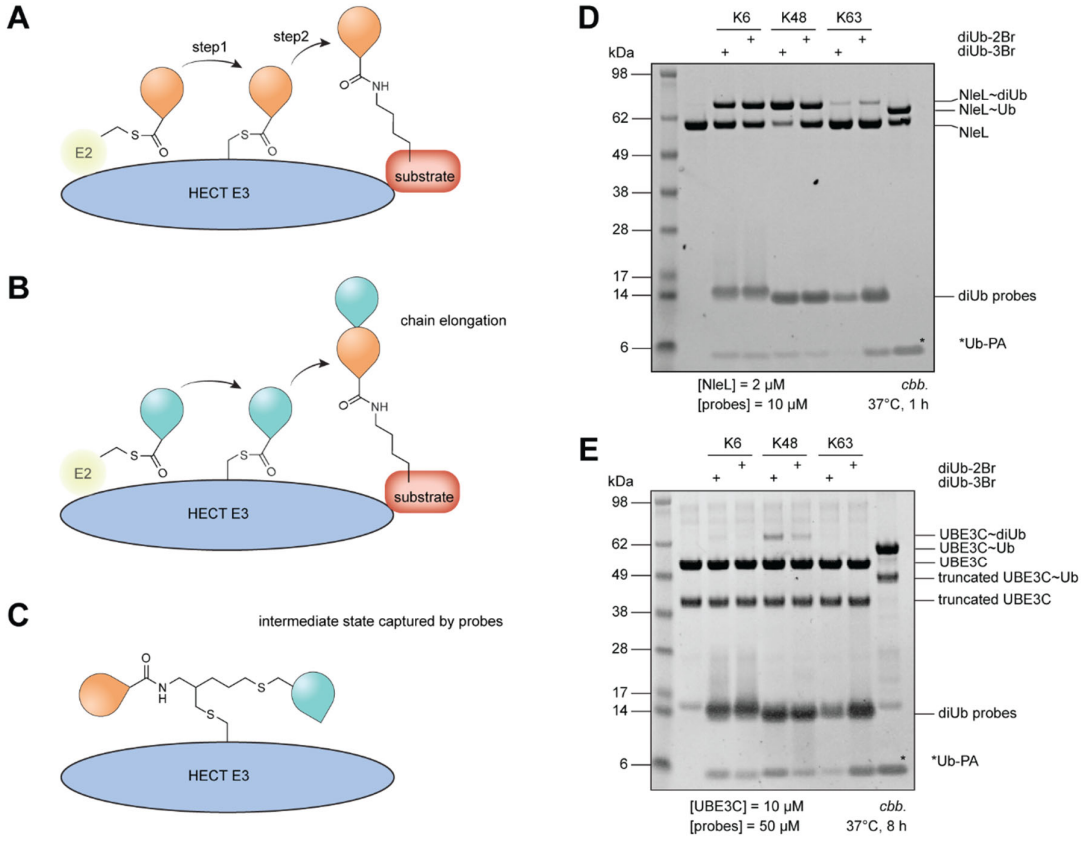


Fig 3.16. Application of alkyl-bromide-based diUb probes for capturing HECT E3s which have a catalytic cysteine. **A**. Schematic illustration of ubiquitination process orchestrated by HECT E3 ligases. **B**. Schematic illustration of polymerization of ubiquitin chains by HECT E3. **C**. The intermediate state of ubiquitination by HECT E3 can be captured by alkyl-bromide-based probes. The crosslinked product of HECT E3 and diUb-3Br is shown. **D**. Gel-based analysis of bacterial E3 NleL incubated with various diUb probes as well as Ub-PA and visualized by Coomassie stains. **E**. Reactions of UBE3C with probes were resolved on gel. The reaction conditions are shown.

With the aid of covalent diUb probe, UBR5 was modified with Lys48-linked diUb. The complex structure was then visualized by Cryo-EM. This application implied that alkyl-bromide-based probes might also be capable of trapping HECT E3s. The overall mechanism of HECT E3s is based on their catalytic cysteine which delivers the E2-charged ubiquitin by forming an active thioester as the first step. In parallel, HECT E3s can host a substrate of which a lysine residue can attack the thioester intermediate to forge an isopeptide bond (Fig 3.16.A). Besides isopeptide bonds formed by lysine residues, the N-terminal methionine can be also modified by activated ubiquitin. This process can be iteratively performed for several rounds to generate polyubiquitin chains (Fig 3.16.B). In some E3s, specific linkages can be achieved to form homogeneous polyubiquitin chains. This mechanism requires more structural information for clarification. Besides HECT E3s, another type of E3s, RBRs also have a catalytic cysteine using a similar mechanism to modify specific substrates and then polymerize ubiquitin chains. Recent work focusing on the RBR-type E3 RNF216 has revealed its Lys63 linkage specificity by structural determination of RNF216 in non-covalent complex with its diUb substrate[186]. The obtained structure informed how RNF216 achieve linkage specificity and guided further validation experiments.

According to the successful application of Michael-acceptor-based diUb probe in UBR5, it is hypothesized that alkyl-bromide diUb probes could also be reactive toward the catalytic cysteine of HECT E3 to form a thioether-based complex (Fig 3.16.C). For HECT E3s with specific linkage preference, HECT E3s might form some specific interaction to host proximal and distal ubiquitin molecules to achieve desired linkages. To validate the hypothesis of alkyl-bromide-based diUb probes for capturing HECT E3s, two in-house available E3s NleL and UBE3C were submitted for the gel-based assay.

NleL is a bacterial E3 with capability of constructing Lys6 and Lys48 linkages[102]. NleL was treated with excessive probes individually at 37°C for an hour. Gel analysis showed that Lys6 and Lys48-linked diUb probes reacted with NleL efficiently with clear band shifts over NleL~Ub-PA complex. Slight off-target effects were also observed in the lanes of Lys63 diUb probes (Fig 3.16.D). The second E3 was UBE3C which is responsible for Lys29 and Lys48 chain types. Since this enzyme is poorly active and prone to aggregate, buffer supplemented with glycerol was used for the reactions. These reactions were performed for longer period because of the low enzymatic activity. Gel analysis explicitly proved the selectivity of Lys48-based probes (Fig 3.16.E). Moreover, diUb probes with other linkages can also be made on demand for tailored study of other HECT or RBR E3s of interest.

3.11 Summary

Polyubiquitin linkage-related cellular events are tightly regulated by dedicated E3 ligases and DUBs. Structural determination of complexes between DUBs and their ubiquitin substrate is a straightforward approach to visualize how the linkage specificities are achieved and modulated. Advances in X-ray crystallography and Cryo-EM provides complementary options for structural determination of DUBs in complex with substrates. However, successful structural determination often requires covalent probes to capture DUBs to form a highly stable complex which is usually beneficial for capturing the snapshot of catalysis. To clearly elaborate the catalytic mechanism of linkage preferences in many DUBs, diUb-based probes armed with an internal warhead provided valuable tools for trapping DUBs. Moreover, these probes could be use as well for trapping E3s which use a catalytic cysteine for ubiquitin transfer from E2 to their specific substrates.

In this chapter, diUb-2Br and diUb-3Br probes were designed based on the inspirations from earlier design of mono-Ub-based probes[168]. Through the integration of organic synthesis and protein chemistry, a panel of alkyl-bromide-based probes was assembled. These probes shown good performance toward promiscuous USP2 and USP8. Clear band shifts can be observed in almost every probe. On the hunt for linkage-specific DUBs with the probes, several DUBs with different linkage preference and origins are selected. ZUFSP was one of the selected enzymes. It did not form covalent adducts with Lys63-linked diUb probes. Previous diUb cleavage data implied that diUb might be a good substrate for ZUFSP which could explain why no covalent complex was observed. Hence, further efforts to forge a tri or tetra-Ub probe armed with a covalent warhead might be a solution to trap ZUFSP covalently.

Gel-based assays informed the selection of SnVTD. To fully unleash the potential of these probes, SnVTD, a Lys6-specific DUB from the newly identified VTD family, was captured to form a covalent complex. After successfully determining a high-resolution crystal structure of SnVTD~diUb-3Br, the interface between SnVTD and diUb was explicitly visualized. Through a detailed analysis of the binding interface between SnVTD and diUb, several potentially important residues in SnVTD which might be involved in specific recognition was mutated. The obtained mutants were tested by a gel-based Lys6-linked diUb cleavage assay. The assay results suggested that SnVTD uses unique dual hydrophobic “locks” for its activation and recognition of Lys6-linked polyubiquitin chains. This intriguing mechanism also highlighted the importance of the development of novel diUb probes with internal warheads.

Beyond DUBs, these probes were employed to capture HECT E3s NleL and UBE3C in a linkage-specific manner. Since RBR E3s also have a catalytic cysteine for transferring ubiquitin, continued exploration on these E3s with alkyl-bromide-based probes would expand the understanding of their linkage specificity. For those E3s with high molecular weight, Cryo-EM would be an option for structural analysis of their linkage specificity.

In summary, the ambition to understand linkage specificity in DUBs has led to the development of a panel of diUb-based probes with internal alkyl-bromide-based warheads. Three linkages were prepared using an expedited method based on the photoinitiated thiol-ene addition. Due to the facile preparation workflow, variants with an epitope tag could be installed for pull-down assays and proteomics studies Applications of these probes will facilitate the understanding of linkage specificity of DUBs and E3 ligases.

4. DISCUSSION & OUTLOOK

The discovery of ubiquitin paved a new avenue for investigating the fate of proteins in cells. The complexity of the ubiquitin system has dramatically expanded with the increasing discovery of other ubiquitin-like modifiers as well as eight different linkages in the polyubiquitin chains. The modifications of Ub/Ubls are reversibly controlled by a large class of enzymes DUBs and ULPs. Similarly, the regulation of specific linkage is controlled by some unique DUBs to modulate related downstream effects which are often involved in physiology and pathology. This thesis focused on the tool development for studying two layers of specificities in DUBs/ULPs. The first layer of specificity was the modifier specificity in DUBs/ULPs, of which various DUBs/ULPs were investigated with a panel of natively prepared substrates in a quantitative manner. The second layer of specificity was the linkage specificity which was studied with newly designed diUb probes with internal covalent warheads. The general goals of this thesis are to investigate the modifier specificities of DUBs/ULPs quantitatively as well understand the mechanism linkage specificity with novel diUb covalent probes for structural determination. The integrated platform would facilitate drug development and provide structural insights into mechanistic studies of DUBs/ULPs.

4.1 Modifier Specificity from a Biochemical Perspective

Over the past four decades, studies focusing on DUBs/ULPs are heavily dependent on Ub/Ubl-based substrates, especially fluorogenic substrates which are quenched and conjugated to Ub/Ubls. Triggered by DUBs/ULPs, substrates were cleaved over time and the released fluorophores can be excited by light and then emit fluorescence which can be detected and analyzed quantitatively. The major drawback in this type of substrates is the linkage between C-termini of Ub/Ubls and fluorophores, which is an artificial amide bond instead of a natural isopeptide bond. The artificial amide bond can be forged by intein chemistry after optimization of the reaction conditions. However, forging the isopeptide bond is considerably difficult and usually requires an auxiliary thiol group on the side chain of lysine. After functionalization, the thiol group has to be removed by harsh condition, such as radical desulfurization which could damage the natural conformation of proteins, especially some Ubls. Chemical synthesis in organic solvents enabled the preparation of Ub/Ubl-KG-TAMRA as fluorescence polarization substrates with the benefit of isopeptide linkage but could also lead to heterogeneity.

To avoid chemical synthesis and subsequent preparative HPLC purification workflow, the C-terminus of ubiquitin was functionalized by intein chemistry to form ubiquitin hydrazide as a stable intermediate. In aqueous buffer, ubiquitin hydrazide was treated subsequently by sodium nitrite and citric acid. The latter two formed nitrous acid which is able to oxidize the C-terminal tail of ubiquitin hydrazide to unstable intermediate ubiquitin acyl azide. KG-TAMRA was then added into the *in situ* generated ubiquitin acyl azide to form an isopeptide bond. This protocol was applied to other Ubls, including SUMO1, SUMO2, FUBI, ISG15(CTD) and NEDD8. For each substrate, two-step purification methods were customized to give pure final products. It is worth further exploring the purification to generalize the protocol. Coarse purification by size-exclusion chromatography can be performed first to remove excessive small molecule, especially free KG-TAMRA which is the major impurity during following ion exchange purification. Secondly, fine purification could be achieved by hydrophobic interaction chromatography since the installation of KG-TAMRA significantly changed the hydrophobicity of the entire molecule.

The fluorescence polarization substrates were used to characterize the substrate specificity of USPL1, an atypical member in USP family. Quantitative analysis of USPL1 with fluorescence polarization substrates revealed 25-fold selectivity of SUMO2 over SUMO1 while fluorogenic substrates reported a selectivity ratio of 5 folds. To fully elaborate the mechanism of its modifier specificity, two crystal structures of USPL1 in covalent complex with SUMO2/3 were solved. Mutagenesis studies based on the structural analysis and bioinformatic mining could provide convincing data for explaining the fundamental mechanism of modifier specificity in USPL1. Notably, a similar work which reported a complex structure of USPL1 with SUMO2 was published earlier which was enabled by the release of AlphaFold[231]. Further translation of these biochemical data into cellular context would facilitate the understanding of the functions of USPL1 toward SUMO2/3. For comparison, deSUMOlyases SENP1 and yeast ULP1 were also studied with the substrates quantitatively which both confirmed the homogeneity of fluorescence polarization substrates.

After understanding the catalytic specificity of USPL1, an obsolete DUB UCHL3 was tested. UCHL3 was known for its dual specificity toward ubiquitin and NEDD8. By comparing the data from fluorescence polarization assay and previously reported data based on Ub/NEDD8-AMC fluorogenic substrates, a huge difference was observed. This comparison implied that deNEDDylase activity of UCHL3 was underestimated. The dual roles of UCHL3 needs to be clarified in cellular setting since NEDD8 regulates protein degradation indirectly.

Human USP18 was investigated with Ub/ISG15(CTD)-KG-TAMRA substrates for the first time. Interestingly, human USP18 showed lower enzymatic activity than mouse USP18 which was investigated previously[204]. This result implied the function of human USP18 might be realized as a scaffold protein to coordinate protein-protein interactions[232]. Given that ISG15 is actively involved in immune response, cellular functions of USP18 as a deISGlyase or an adaptor protein remain further elucidation.

Sequence alignment of Ub/Ubls revealed that ISG15 and FUBI are closely relevant (Fig 1.2.A). Recent work using a FUBI-based covalent probe in combination with proteomics identified two DUBs USP16 and USP36 with profound activity toward FUBI[66]. With the aid of the panel of fluorescence polarization substrates, USP16 and USP36 were validated as deISGylases which means they are the first DUBs with triple specificities. This expanded the complexity of their enzymatic functions in cellular environment. How they cleave different substrates upon different stimulation needs to be deeply investigated. Notably, USP16 was also recently discovered as a deISGylase through proteomics[208]. USP5 was also reported as a deISGylase. However, the profile of modifier specificity of USP5 remains unclear which could be further tested with the panel of fluorescence polarization substrates.

The method for preparing fluorescence polarization substrates natively provided homogenous assays to distinguish the modifier specificities of DUBs and ULPs. As isopeptide-bond-based reagents, they are structurally close to the natural substrates. Several DUBs/ULPs were selected for quantitative studies. Given that DUBs/ULPs have more than 100 members, there is still a lot of space for exploration using isopeptide-bond-linked substrates. Moreover, the substrates can be prepared in a larger quantity for high-throughput screening campaign to discover novel modulators targeting DUBs/ULPs.

4.2 Further Exploration of Other Modifiers?

This work presented only six Ub/Ubl functionalized with KG-TAMRA fluorophore at their C-termini. There are still more than 10 Ubls for exploration using Ubl hydrazide as starting materials for C-terminal functionalization. Installation of a covalent warhead at C-termini of Ubls would enable discovery of corresponding proteases by proteomics. Since some Ubls are very fragile and sensitive to oxidation, C-terminal functionalization is a challenging task. The classic intein-based functionalization usually takes several hours to days which could cause heavy oxidation and subsequent aggregation of Ubls[66].

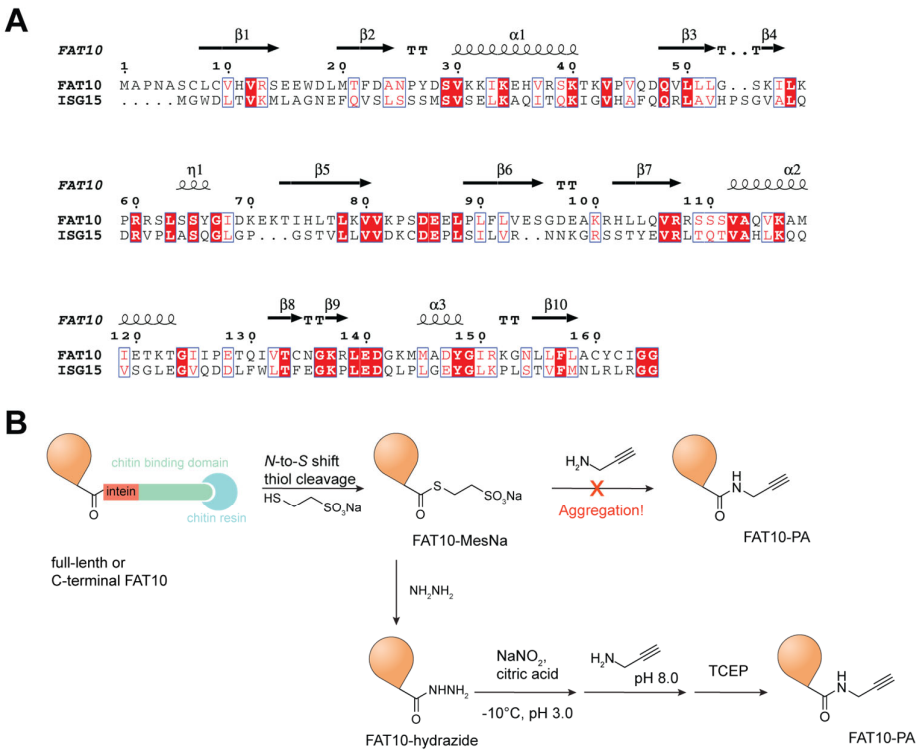


Fig 4.1. Proposed route for functionalization of FAT10. **A**. Sequence alignment of mature ISG15 and FAT10. Secondary structure of FAT10 is annotated according to a structure of FAT10 (PDB ID: 7PYV). **B**. Proposed synthetic route for preparing covalent probe FAT10-PA which might reduce protein precipitation.

FAT10 is a notorious protein which is prone to form aggregation due to its C-terminal cysteine residues[233]. So far, how FAT10 is attached to its substrates has been intensively studied which led to the identification of corresponding conjugating enzymes[234]. However, which proteases are responsible for cleaving FAT10 remains understudied. FAT10, similar to ISG15, has two Ubl domains (Fig 4.1.A). FAT10 has two cysteine residues adjacent to its terminal diGly motif which could form oxidation and aggregation. If FAT10 is prepared from a fusion protein with intein and chitin-biding domain, FAT10-MesNa could be obtained. Functionalization with a PA warhead is usually time-consuming which might not be compatible with FAT10 stability. Alternatively, FAT10-MesNa could be immediately transformed into FAT10 hydrazide with excessive hydrazine which will be further functionalized by nitrous acid (Fig 4.1.B). Since the cysteine residues will be nitrosylated as well, excessive TCEP could reverse the nitrosylation. The anticipated route would provide opportunity to obtain FAT10-based probes for discovering related proteases in cells. Alternatively, FAT10-KG-TAMRA might be prepared. Previously reported deISGylases can be tested with this substrate. Moreover, this method could be further optimized and adapted for other understudied Ubls to characterize other previously unknown ULPs.

4.3 Unmet Needs for Studying Linkage Specificity

The second layer of ubiquitin specificity investigated in this thesis was linkage specificity which was enabled by designing novel diUb-based covalent probes. Previously reported diUb probes with various internal covalent warheads have not been widely applied for studying DUBs because of difficulties in synthesis. Recent discovery of VTD DUBs showed that some of them possess high linkage specificity. To understand the specificity mechanism of VTDs, novel diUb probes with differential trapping mechanism are necessary for structural biology work. In this thesis, two types of novel diUb-based probes were designed. Notably, preparation of these probes does not require either special equipment or extensive training in solid phase peptide synthesis.

Alkyl-bromide-based diUb probes were designed with the inspiration from Ub-2Br and Ub-3Br probes which are based on nucleophilic substitution to capture catalytic cysteine of DUBs. Therefore, the newly designed diUb probes were named as diUb-2Br and diUb-3Br respectively. The probes were prepared semisynthetically in two steps. The first step is to introduce warhead into distal ubiquitin. The resulting intermediate has a terminal alkene which can be conjugated with the mutated cysteine in proximal ubiquitin to form an isopeptide-like linker by photoinitiated thiol-ene addition. Since the probes are used for covalent trapping which does not require high purity as long as there is no interference with the covalent reaction with catalytic cysteine by side products, size-exclusion chromatography was used for purification. Although there were still some impurities in the final products, they are inert as lacking in covalent warheads. Three different linkages were prepared for screening of DUBs. Initial screening included Lys63-specific ZUFSP, Lys48-specific Mug105, Lys6-specific WcVTD and SnVTD. SnVTD formed covalent adducts with the probes. Therefore, the purified adduct of SnVTD with diUb-3Br was further characterized and crystallized. With the assistance of diUb binding, SnVTD~diUb-3Br showed significantly increased stability in the thermal shift assay. The crystal structure was solved and analyzed. Biochemical validation found that SnVTD utilized two hydrophobic interactions to recognize Lys6-linked polyubiquitin and initiate the cleavage. These unique hydrophobic interactions explained that SnVTD cannot cleave mono-Ub-based substrates since they cannot occupy the proximal binding site in SnVTD.

Although diUb cleavage assay provided solid evidence to analyze which residues are actively involved in the binding and cleavage, mutations on ubiquitin would provide more evidence reciprocally. Moreover, FRET assay would be another choice for monitoring the cleavage by

SnVTD variants in a quantitative manner. SnVTD as a Lys6 specific DUB can be used as a powerful tool for UbiCrest assay to analyze ubiquitin chains. The alkyl-bromide-based probes were able to trap HECT E3s as well. Further test on RBR E3s is necessary to broaden their applications.

Apart from polyubiquitin linkage specificity, ubiquitination on histone attracts a lot of research efforts. Recent studies using Cryo-EM elaborated the structural basis of H2AK119 specificity of BAP1[235, 236]. Moreover, Cryo-EM was also used to investigate how USP16 cleave H2AK119 ubiquitin[237]. In theory, the established method for preparing diUb probes can be optimized to generate ubiquitinated histone which bears a covalent warhead in the linker. Given that cysteine is not present in most histone H2A or H2B, site specific incorporation of a ubiquitin armed with an alkyl bromide warhead is feasible. The resulting probes can be used to covalently trap DUBs that cleave ubiquitin on histone.

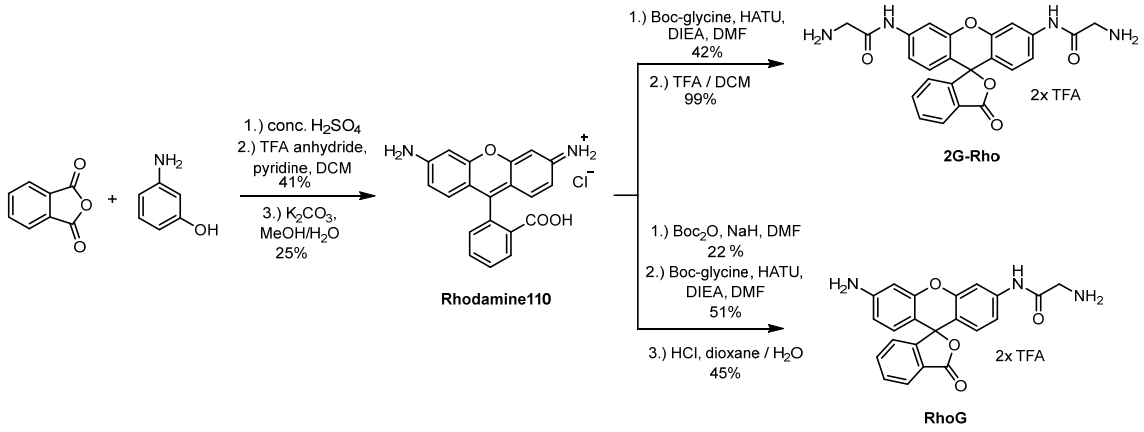
Collectively, this thesis established two assay platforms for studying modifier specificity of DUBs/ULPs and linkage specificity of DUBs. The semisynthetical methods used in this thesis can be further employed for modifying other protein of interest. The combined platforms would facilitate the deeper understanding of DUBs/ULPs biochemically and structurally which will be beneficial for translational research in cell biology as well as drug research and development.

5. MATERIALS & METHODS

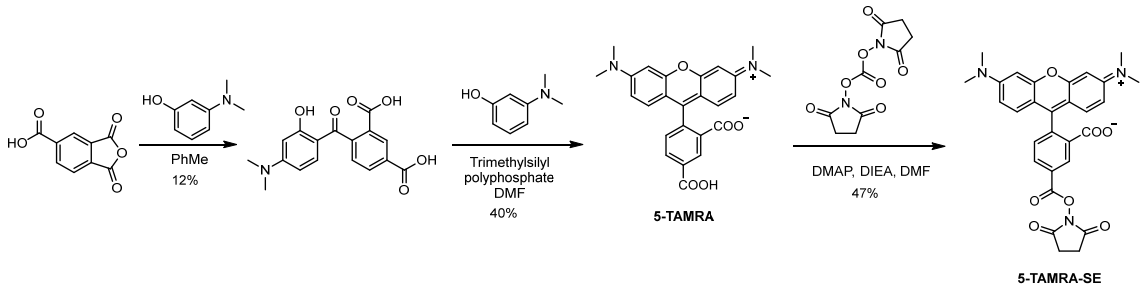
5.1 Chemistry: General Methods

Reagents and solvents were purchased from commercial resources and used without further purification. LC-MS was measured on an Agilent 1260 UPLC coupled to an MSD mass analyzer (single quadrupole) equipped with an electrospray ionization (ESI) source. ^{1}H and ^{13}C NMR spectra were recorded on Bruker AV 500 Avance III HD, AV 600 Avance III HD, Avance NEO 600 or AV 700 Avance III HD spectrometers. The chemical shifts of all spectra are specified in ppm and were referenced to the respective solvent peaks where possible. Thin-layer chromatography was carried out using silica gel aluminum plates (silica gel 60 GF254, Merck). Column chromatography was performed on silica gel (200-300 mesh). Preparative reverse phase (RPLC) purification was performed using an 1260/1290 Infinity II series system (Agilent, for high pressure purification) or a Pure C-850 system (Büchi, for medium pressure purification) equipped with a VP125/21 Nucleodur C18 Gravity column (5 µm, Macherey Nagel) and solvents as specified (mobile phase A: H_2O supplemented with 0.1% TFA, mobile phase B: acetonitrile supplemented with 0.1% TFA). High resolution mass spectrometry data were measured on an LTQ Orbitrap (Thermo Fisher).

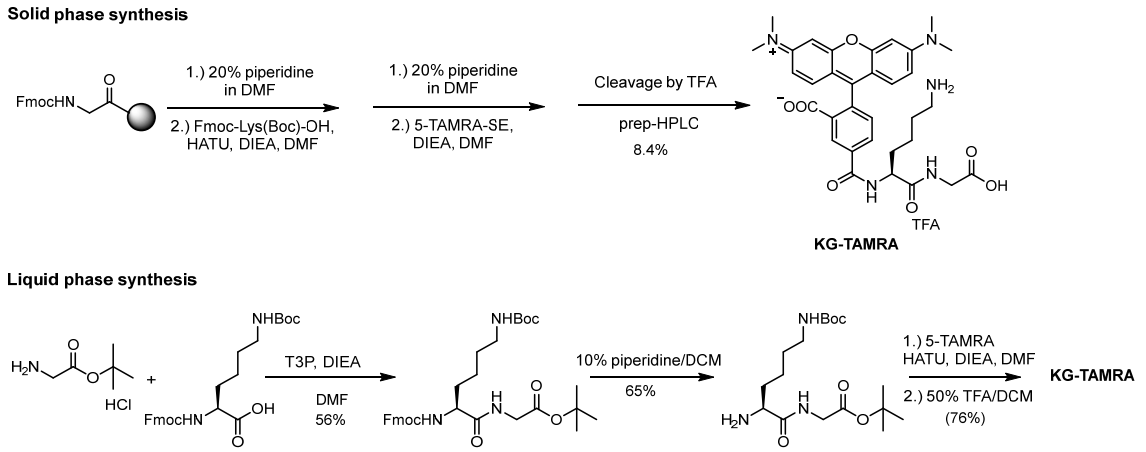
5.2 Chemical Synthesis of Fluorophores



Scheme 5.1. Rhodamine110 was prepared as shown, and then converted into 2G-Rho for fluorogenic substrate generation (upper path) and into RhoG for signal normalization (lower path).

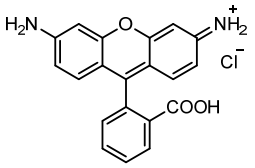


Scheme 5.2. Synthesis of 5-TAMRA and 5-TAMRA-SE.



Scheme 5.3. Preparation of free KG-TAMRA for fluorescence polarization substrate generation. While the compound (referred to as TAMRA-Lys-Gly) has been referenced[184] and used widely[204, 238, 239], a procedure for its synthesis as well as characterization data have not been reported. Two synthesis routes were explored: Through solid phase peptide synthesis (upper) as well as a solution phase route (lower).

Rhodamine 110

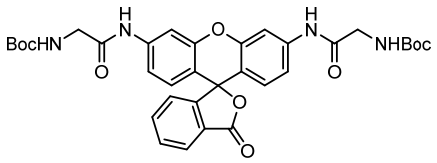


Rhodamine 110

Synthesis of free Rhodamine 110 was performed according to a previously reported procedure[240] with some modifications kindly suggested by Luke Lavis (personal communication). In a round-bottom flask, phthalic anhydride (15.0 g, 101.2 mmol, 1.0 eq) and 3-aminophenol (18.8 g, 172.1 mmol, 1.7 eq) were dissolved in 75 mL concentrated sulfuric acid and heated at 180 °C for 5 h. Then the reaction mixture was poured into crushed ice, and subsequently diluted with water and concentrated ammonia solution. The mixture was kept at 4°C overnight and

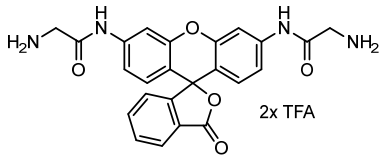
filtered. The filtrate was refluxed in 1 M HCl (400 mL) for 20 min and then hot filtered. The crude solution was allowed to stand at 4°C for 40 h. The solid was filtered and dried in the oven to afford crude Rhodamine 110 as a black solid (27 g). To further purify Rhodamine 110, trifluoroacetic anhydride was used to protect the amines. To a solution of crude Rhodamine 110 (330 mg, 1.0 mmol, 1.0 eq) in dichloromethane (CH_2Cl_2), trifluoroacetic anhydride (0.524 mL, 3.8 mmol, 3.8 eq) and pyridine (0.3 mL) were added at room temperature, and it was stirred for 3 h. The reaction mixture was extracted with CH_2Cl_2, dried over $MgSO_4$, and purified by silica column using petroleum ether (PE)/ethyl acetate (EtOAc) = 6:1 as eluant to give the product as an orange solid (200 mg, 41% yield). To a solution of TFA-protected Rhodamine 110 (110.8 mg, 0.21 mmol, 1.0 eq) in methanol (MeOH), potassium carbonate (785.8 mg, 5.69 mmol, 27 eq) in 2 mL water was added and stirred overnight at room temperature. The reaction was then quenched with concentrated HCl and filtered. The filter cake was washed with cold water. The solid was dried at 60°C overnight to give a purple solid (20.8 mg, 25.4% yield). ESI-MS: *m/z* for $C_{20}H_{15}N_2O_3^+$ $[M+H]^+$ calcd: 331.1, found: 331.1; ^{1}H NMR (700 MHz, DMSO-*d*$_6$) δ 13.11 (s, 1H), 8.20 (d, *J* = 7.8 Hz, 1H), 8.06 (s, 4H), 7.85 (td, *J* = 7.5, 1.1 Hz, 1H), 7.78 (t, *J* = 8.2 Hz, 1H), 7.43 (d, *J* = 7.4 Hz, 1H), 7.04 - 6.91 (m, 2H), 6.90 - 6.73 (m, 4H).

Bis-BocGly-Rhodamine 110



To a solution of Boc-glycine (524.4 mg, 3.0 mmol, 6.0 eq) were added HATU (1.14 g, 3.0 mmol, 6.0 eq) and DIEA (1.56 mL, 9.0 mmol, 9.0 eq) in DMF (5 mL), and it was stirred for 30 min. Then Rhodamine110 (183.0 mg, 0.5 mmol, 1.0 eq) was added and it was stirred overnight. The reaction mixture was diluted with 1 M HCl, extracted with EtOAc and the combined organic phases were dried over $MgSO_4$. The filtrate was concentrated and purified on silica column (CH_2Cl_2/MeOH = 20:1) to give an orange solid (133.9 mg, 41.5% yield). ESI-MS: *m/z* for $C_{34}H_{37}N_4O_9^+$ $[M+H]^+$ calcd: 645.3, found: 645.2; ^{1}H NMR (500 MHz, DMSO-*d*$_6$) δ 10.26 (s, 2H), 8.03 (d, *J* = 7.6 Hz, 1H), 7.83 (d, *J* = 1.8 Hz, 2H), 7.82 - 7.77 (m, 1H), 7.73 (t, *J* = 7.4 Hz, 1H), 7.29 (d, *J* = 7.6 Hz, 1H), 7.17 (dd, *J* = 8.7, 2.0 Hz, 2H), 7.10 (t, *J* = 6.1 Hz, 2H), 6.75 (d, *J* = 8.7 Hz, 2H), 3.73 (d, *J* = 5.9 Hz, 4H), 1.39 (s, 18H).

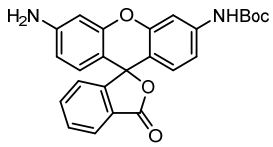
Bis-Gly-Rhodamine 110 (2G-Rho)



2G-Rho

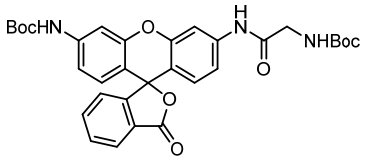
To a solution of bis-BocGly-Rhodamine 110 (506.5 mg, 0.79 mmol) in THF (10 mL), 50% TFA in CH_2Cl_2 (15 mL) was added slowly at 0°C. The reaction was moved to room temperature and stirred overnight. The solvent was removed under reduce pressure to give the final product as an orange solid (660.5 mg, 99.2% yield). ESI-MS: *m/z* for $C_{24}H_{21}N_4O_5^+$ $[M+H]^+$ calcd: 445.2, found: 445.2; ^{1}H NMR (500 MHz, D_2O) δ 8.09 - 8.01 (m, 1H), 7.79 - 7.69 (m, 2H), 7.40 (d, *J* = 8.2 Hz, 2H), 7.17 (d, *J* = 6.5 Hz, 1H), 7.08 - 6.99 (m, 2H), 6.77 (dd, *J* = 10.4, 4.3 Hz, 2H), 3.94 (s, 4H).

mono Boc-protected Rhodamine 110



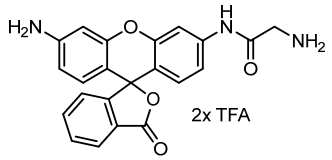
To a solution of Rhodamine 110 (200 mg, 0.55 mmol, 1.0 eq) in anhydrous DMF (10 mL) was added NaH (60% in mineral oil, 43.6 mg, 1.1 mmol, 2.0 eq) slowly over a period of 5 min. The resulting mixture was stirred for 1 h, then di-*tert*-butyl dicarbonate (Boc_2O, 101.2 mg, 0.46 mmol, 0.84 eq) was added and the mixture stirred at ambient temperature for 2 days. The reaction was quenched with acetic acid (1 mL) and water (5 mL). The mixture was diluted with EtOAc (50 mL) and washed with water and brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo*. The resulting crude product was purified on silica column chromatography (PE/ CH_2Cl_2/EtOAc = 4:4:2) to give the title compound as an orange solid (52.7 mg, 22.3% yield). ESI-MS: *m/z* for $C_{25}H_{23}N_2O_5^+$ $[M+H]^+$ calcd: 431.2, found: 431.2; ^{1}H NMR (700 MHz, DMSO-d_6) δ 9.64 (s, 1H), 7.97 (d, *J* = 7.7 Hz, 1H), 7.78 (td, *J* = 7.5, 1.0 Hz, 1H), 7.73 - 7.68 (m, 1H), 7.51 (s, 1H), 7.23 (d, *J* = 7.7 Hz, 1H), 7.08 (dd, *J* = 8.7, 2.0 Hz, 1H), 6.58 (d, *J* = 8.7 Hz, 1H), 6.43 (d, *J* = 2.0 Hz, 1H), 6.36 (d, *J* = 8.6 Hz, 1H), 6.32 (dd, *J* = 8.6, 2.0 Hz, 1H), 5.65 (s, 2H), 1.48 (s, 9H).

Dual-Boc-protected Gly-Rhodamine 110



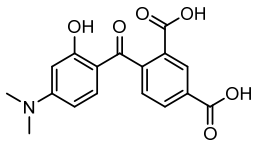
To a solution of Boc-glycine (36.6 mg, 0.21 mmol, 1.0 eq) in DMF were added HATU (106.0 mg, 0.28 mmol, 1.3 eq) and DIEA (0.49 mL, 0.28 mmol, 1.3 eq) and it was stirred for 30 min. Then mono Boc-protected Rhodamine 110 (30.0 mg, 0.70 mmol, 3.3 eq) was added and stirred overnight. The reaction was extracted with EtOAc and dried over $MgSO_4$. The filtrate was concentrated and purified on silica column (PE/ EtOAc = 1:1) to give an orange solid (21.0 mg, 51.2% yield). HRMS: *m/z* for $C_{32}H_{34}N_3O_8^+$ $[M+H]^+$ calcd: 588.2340, found: 588.2335; ^1H NMR (700 MHz, $CDCl_3$) δ 8.36 (s, 1H), 8.05 - 8.00 (m, 1H), 7.68 - 7.63 (m, 2H), 7.63 - 7.58 (m, 1H), 7.50 (s, 1H), 7.12 (d, *J* = 7.5 Hz, 1H), 7.04 (d, *J* = 8.4 Hz, 1H), 6.93 - 6.87 (m, 1H), 6.73 (d, *J* = 8.6 Hz, 1H), 6.69 (d, *J* = 8.6 Hz, 1H), 6.62 (s, 1H), 5.23 (t, *J* = 6.2 Hz, 1H), 3.93 (d, *J* = 6.1 Hz, 2H), 1.53 (s, 9H), 1.48 (s, 9H). ^{13}C NMR (176 MHz, $CDCl_3$) δ 169.68, 153.43, 152.49, 151.97, 151.89, 140.70, 135.27, 129.93, 128.80, 128.71, 126.47, 125.24, 124.05, 115.50, 114.35, 113.20, 108.00, 106.28, 82.69, 55.90, 29.85, 28.43.

Gly-Rhodamine 110 (RhoG)



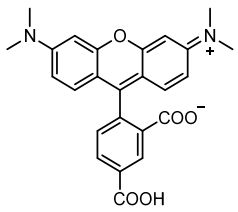
To a solution of dual-Boc-protected Gly-Rhodamine 110 (17.0 mg, 0.29 mmol) in a round-bottom flask in 1,4-dioxane (1 mL), 3 mL HCl (4 M in dioxane/H_2O) was added and stirred overnight with protection from light. The solvent was evaporated *in vacuo*. The crude mixture was purified by reverse phase HPLC. Pure fractions were pooled and lyophilized to give RhoG as an orange solid (6.0 mg, 45.0% yield). HRMS: *m/z* for $C_{22}H_{18}N_3O_4^+$ $[M+H]^+$ calcd: 388.1292, found: 388.1296; ^1H NMR (600 MHz, D_2O) δ 8.19 - 8.15 (m, 1H), 8.15 - 8.10 (m, 1H), 7.82 - 7.78 (m, 2H), 7.45 (d, *J* = 8.9 Hz, 1H), 7.43 - 7.39 (m, 3H), 7.00 (dd, *J* = 9.3, 2.0 Hz, 1H), 6.92 (d, *J* = 1.9 Hz, 1H), 4.09 (s, 2H). ^{13}C NMR (151 MHz, D_2O) δ 172.33, 166.11, 163.12, 162.88, 161.25, 159.62, 154.65, 144.21, 133.21, 133.04, 131.54, 130.53, 130.44, 129.61, 129.57, 119.65, 118.06, 117.78, 117.28, 117.24, 115.35, 106.94, 97.82, 41.41.

4-(4-(Dimethylamino)-2-hydroxybenzoyl)isophthalic acid



Synthesis of 5-TAMRA was based on the previous literature[241]. To a solution of 3-dimethylaminophenol (6.86 g, 50.0 mmol, 1.0 eq) in toluene, 1,2,4-benzenetricarboxylic anhydride (11.53 g, 60.0 mmol, 1.2 eq) was added and refluxed overnight. Product formation was observed via TLC ($CHCl_3$/MeOH/TEA = 85:10:5, R_f = 0.36) consistent to literature[241]. The reaction was then cooled down and filtered. The filter cake was washed with toluene (50 mL ×2), then dissolved in 200 mL MeOH and refluxed for 10 min. 50 mL acetic acid was added into the solution and evaporated under vacuum. The residue was dissolved in 100 mL MeOH and refluxed for 2 h and then moved to 4°C overnight. The reaction was filtered, and the filter cake was washed with cold MeOH to give the final product as a brown solid (1.94 g, 12% yield). The product was used without further purification. ESI-MS: *m/z* for $C_{17}H_{16}NO_6^+$ $[M+H]^+$, calcd: 330.1, found: 330.0. ^{1}H NMR (500 MHz, DMSO-d_6) δ 12.38 (s, 1H), 8.48 (d, J = 1.6 Hz, 1H), 8.22 - 8.14 (m, 1H), 7.50 (d, J = 7.8 Hz, 1H), 6.80 (d, J = 9.1 Hz, 1H), 6.21 (dd, J = 9.2, 2.5 Hz, 1H), 6.11 (d, J = 2.4 Hz, 1H), 3.01 (s, 6H).

5-TAMRA

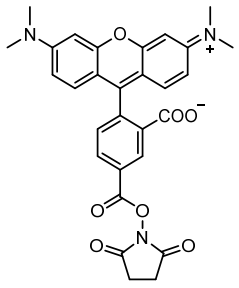


5-TAMRA

To a round-bottom flask, 4-(4-(dimethylamino)-2-hydroxybenzoyl)isophthalic acid (1.94 g, 5.9 mmol, 1.0 eq) and 3-dimethylaminophenol (1.05 g, 7.7 mmol, 1.3 eq) were dissolved in DMF (20 mL). Trimethylsilylpolyphosphate solution in chloroform (6 mL) was added and the mixture was refluxed for 3 h. Solvents were removed *in vacuo* and the residue was dissolved in 5% NaOH (20 mL) and stirred at room temperature overnight. The mixture was diluted with water (75 mL) and the product was precipitated with concentrated HCl (5 mL). The solid was filtered and washed with cold water (50 mL) and dried at 60°C overnight to give the 5-TAMRA as a purple solid, 1.0 g, 39.5% yield. The product was then used directly for next steps without

further purification. ESI-MS: *m/z* for $C_{25}H_{23}N_2O_5^+$ $[M+H]^+$, calcd: 431.2, found: 431.0. ^1H NMR (700 MHz, DMSO-d_6) δ 8.63 - 8.36 (m, 1H), 8.30 (d, J = 8.1 Hz, 1H), 7.40 (s, 1H), 6.61 (m, 6H), 3.01 (s, 12H).

5-TAMRA-SE



5-TAMRA-SE

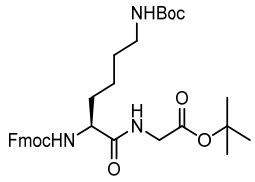
To a solution of 5-TAMRA (86 mg, 0.2 mmol, 1.0 eq), DIEA (66 µL, 0.4 mmol, 2.0 eq) and DMAP (49 mg, 0.4 mmol, 2.0 eq) in DMF (2 mL), disuccinimidyl carbonate (51 mg, 0.2 mmol, 1.0 eq) were added. The reaction was stirred vigorously for 6 h at room temperature and analyzed on TLC ($CHCl_3$/MeOH/TEA = 85:10:5, product: R_f = 0.3). Diethyl ether (20 mL) was added to precipitate the product. The solid was filtered, washed with EtOAc, CH_2Cl_2, and a small volume of cold MeOH and dried at 80°C to give the 5-TAMRA-SE as dark purple solid, 49 mg, 46.5% yield. 5-TAMRA-SE was used without further purification. ESI-MS: *m/z* for $C_{29}H_{26}N_3O_7^+$ $[M+H]^+$, calcd: 528.2, found: 528.2.

Solid phase synthesis of KG-TAMRA

Fmoc-Gly Wang resin (Novabiochem, 100-200 mesh, 36 mg, 0.7 mmol/g) in a fritted syringe was swollen with 5 mL DMF for 20 min. Then, 4 mL 20% piperidine in DMF was added and the resin was agitated for 20 min. The resin was then washed with DMF, MeOH, CH_2Cl_2 and DMF (3x 5 mL of each) sequentially. A mixture of Fmoc-Lys(Boc)-OH (60 mg, 0.13 mol, 5.0 eq) and DIEA (30 µL, 0.26 mmol, 10.0 eq) in 3 mL DMF was added into the syringe. HATU (50 mg, 0.13 mmol, 5.0 eq) was added and the resin agitated for 1 h. The resin was then washed as above and Fmoc was cleaved by 20% piperidine in DMF. The resin was then mixed with 5-TAMRA-SE (39.8 mg, 0.08 mol, 3.0 eq) and DIEA (30 µL, 0.26 mmol, 10.0 eq) for 4 h in DMF. The resin was then washed with DMF until the flow-through was colorless. The product was cleaved off the resin by a mixed solution consisting of TFA/H_2O/TIPS (95:2.5:2.5) for 30

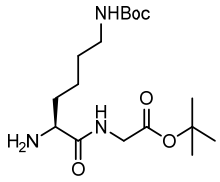
min. The product was finally purified on a preparative HPLC (Agilent) and lyophilized to give the title compound as a dark-purple solid (1.5 mg, 8.4% yield). HRMS: *m/z* for $C_{33}H_{38}N_5O_7^+$ $[M+H]^+$, calcd: 616.2766, found: 616.2762. ^{1}H NMR (700 MHz, D_2O) δ 8.52 (d, *J* = 1.8 Hz, 1H), 8.19 (dd, *J* = 7.8, 1.8 Hz, 1H), 7.64 (d, *J* = 7.8 Hz, 1H), 7.21 (d, *J* = 9.5 Hz, 1H), 7.20 (d, *J* = 9.5 Hz, 1H), 6.88 (dd, *J* = 9.5, 2.3 Hz, 2H), 6.45 (s, 2H), 4.68 (dd, *J* = 8.5, 6.2 Hz, 1H), 4.12 - 4.04 (m, 2H), 3.14 (s, 12H), 3.07 (t, *J* = 6.54 Hz, 2H), 2.09 - 1.97 (m, 2H), 1.84 - 1.76 (m, 2H), 1.68 - 1.60 (m, 2H). ^{13}C NMR (176 MHz, D_2O) δ 174.36, 173.47, 170.58, 169.37, 163.08, 162.88, 157.26, 156.97, 156.92, 135.88, 135.05, 134.78, 130.79, 130.65, 130.28, 128.83, 117.14, 115.48, 113.98, 112.84, 96.05, 56.22, 54.42, 41.37, 39.95, 39.22, 30.51, 26.34, 22.21.

Tert-butyl N^2-(((9H-fluoren-9-yl)methoxy)carbonyl)-N^6-(_tert_-butoxycarbonyl)-_L_-lysyl-glycinate



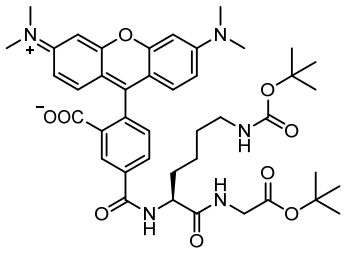
To a 100 mL round-bottom flask, Fmoc-Lys(Boc)-OH (1.406 g, 3.0 mmol, 1.0 eq), glycine *tert*-butyl ester HCl (528 mg, 3.15 mmol, 1.05 eq), T3P (50% in DMF, 2.75 mL, 3.9 mmol, 1.3 eq) and DIEA (1.05 mL, 6.0 mmol, 2.0 eq) were dissolved in anhydrous DMF (30 mL). The reaction was stirred at room temperature for 3 h and quenched by 20 mL H_2O. The reaction was extracted with EtOAc (100 mL ×3). The combined organic phase was washed with brine (100 mL ×3), dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The residue was further purified on silica column (CH_2Cl_2/MeOH = 30:1) to give the title compound as a white solid (970 mg, 55.6% yield). ESI-MS: *m/z* for $C_{32}H_{43}N_3O_7Na^+$ $[M+Na]^+$, calcd: 604.3, found: 604.2. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.76 (d, *J* = 7.5 Hz, 2H), 7.60 (d, *J* = 6.8 Hz, 2H), 7.42 - 7.37 (m, 2H), 7.34 - 7.29 (m, 2H), 6.47 (s, 1H), 5.47 (s, 1H), 4.64 (s, 1H), 4.48 - 4.34 (m, 2H), 4.27 - 4.14 (m, 2H), 4.00 - 3.85 (m, 2H), 3.19 - 3.01 (m, 2H), 1.96 - 1.48 (m, 4H), 1.46 (s, 9H), 1.43 (s, 9H), 1.41 - 1.23 (m, 2H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 171.77, 168.77, 156.32, 143.96, 143.92, 141.45, 127.86, 127.24, 125.25, 120.12, 82.58, 79.33, 67.20, 54.88, 47.31, 42.15, 40.02, 32.23, 29.77, 28.58, 28.19, 22.54.

***Tert*-butyl N^6-(*tert*-butoxycarbonyl)-*L*-lysylglycinate**



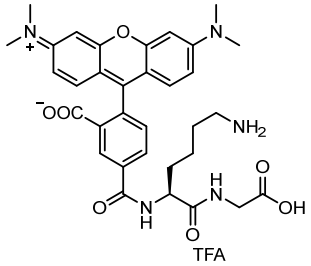
To a 50 mL round-bottom flask, *Tert*-butyl N^2-(((9H-fluoren-9-yl)methoxy)carbonyl)-N^6-(*tert*-butoxycarbonyl)-*L*-lysylglycinate (734 mg, 1.26 mmol, 1.0 eq) was mixed with 10% piperidine in CH_2Cl_2 (8 mL). The reaction was stirred for 1 h. 10 mL PE/diethyl ether (1:1) was added. The resulting mixture was filtered, and the organic phase was concentrated and purified on silica column (CH_2Cl_2/MeOH = 20:1) to give the title compound as a white sticky solid (296 mg, 65.4% yield). ESI-MS: *m/z* for $C_{17}H_{34}N_3O_5^+$ $[M+H]^+$, calcd: 360.2, found: 360.0. ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.72 (s, 1H), 4.59 (s, 1H), 3.98 - 3.88 (m, 2H), 3.42 (dd, *J* = 7.9, 4.5 Hz, 1H), 3.16 - 3.04 (m, 2H), 1.89 - 1.82 (m, 1H), 1.77 (s, 2H), 1.60 - 1.53 (m, 1H), 1.52 - 1.48 (m, 2H), 1.47 (s, 9H), 1.43 (s, 9H), 1.42 - 1.34 (m, 2H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 175.17, 169.32, 156.21, 82.27, 79.23, 55.08, 41.79, 40.31, 34.62, 30.02, 28.57, 28.21, 22.96.

(*S*)-2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-5-((2,2,16,16-tetramethyl-4,11,14-trioxo-3,15-dioxa-5,12-diazaheptadecan-10-yl)carbamoyl)benzoate



5-TAMRA (50 mg, 0.12 mmol, 1.0 eq) was dissolved in 4 mL anhydrous DMF, then *Tert*-butyl N^6-(*tert*-butoxycarbonyl)-*L*-lysylglycinate (50.1 mg, 0.14 mmol, 1.2 eq) and DIPEA (60.7 µL, 0.35 mmol, 3.0 eq) were added. After dissolving all components, HATU (48.6 mg, 0.13 mmol, 1.1 eq) was added and the mixture was stirred for 17 hours at room temperature. After full conversion of 5-TAMRA, the solvent was removed under reduced pressure and the product was purified by preparative medium pressure liquid chromatography eluting at 40-45 % solvent B to give the crude product which was directly used for the next step. ESI-MS: *m/z* for $C_{42}H_{54}N_5O_9^+$ $[M+H]^+$, calcd: 772.4, found: 772.4.

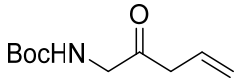
KG-TAMRA



(*S*)-2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-5-((2,2,16,16-tetramethyl-4,11,14-trioxo-3,15-dioxa-5,12-diazaheptadecan-10-yl)carbamoyl)benzoate was dissolved in 1.5 mL CH_2Cl_2 followed by addition of 1.5 mL TFA. The reaction mixture was stirred at room temperature until complete deprotection. After 90 minutes the reaction was diluted with CH_2Cl_2 and the solvent was removed under reduced pressure. KG-TAMRA was obtained as a purple powder (64.9 mg, 76.0% yield over two steps) after purification by preparative HPLC and lyophilization with virtually identical characteristics as specified above. ESI-MS: *m/z* for $C_{33}H_{38}N_5O_7^+$ $[M+H]^+$, calcd: 616.3, found: 616.0. ^{1}H NMR (700 MHz, D_2O) δ 8.40 (d, *J* = 1.8 Hz, 1H), 8.13 (d, *J* = 7.9 Hz, 1H), 7.67 - 7.58 (m, 1H), 7.23 (d, J = 9.4 Hz, 1H), 7.21 (d, *J* = 9.4 Hz, 1H), 6.90 (dd, *J* = 9.5, 1.9 Hz, 2H), 6.59 (s, 2H), 4.67 (dd, *J* = 8.6, 6.1 Hz, 1H), 4.07 - 3.98 (m, 2H), 3.17 (s, 12H), 3.07 (t, *J* = 7.5 Hz, 2H), 2.11 - 1.96 (m, 2H), 1.84 - 1.74 (m, 2H), 1.69 - 1.57 (m, 2H).

5.3 Chemical Synthesis of the Warhead for DiUb-2Br

***Tert*-butyl (2-oxopent-4-en-1-yl)carbamate (1)**

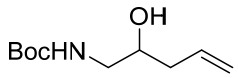


1

To a solution of *tert*-butyl (2-(methoxy(methyl)amino)-2-oxoethyl)carbamate (806.0 mg, 3.69 mmol, 1.0 eq) in anhydrous THF (15 mL) at -78°C under the protection of Ar, allylmagnesium bromide (1 M, 5.54 mL, 1.5 eq) was added dropwise over 10 min. The reaction was quenched by saturated NH_4Cl solution (10 mL) and the mixture was extracted by EtOAc (50 mL ×3). The combined organic phases were dried over anhydrous $MgSO_4$, filtered and concentrated. The resulting residue was further purified on silica gel column chromatography (PE/EtOAc = 5:1) to give the title compound as a colorless oil (368 mg, 1.85 mmol, 50% yield). R_f = 0.34,

(PE/EtOAc = 5:1). ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.90 (ddt, *J* = 17.2, 10.2, 7.0 Hz, 1H), 5.26 - 5.16 (m, 2H), 4.06 (d, *J* = 4.4 Hz, 2H), 3.20 (d, *J* = 7.0 Hz, 2H), 1.44 (s, 9H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 203.76, 155.74, 129.41, 120.05, 80.07, 50.04, 45.23, 28.44. HRMS: *m/z* for $C_{10}H_8NO_3^+$ $[M+H]^+$ calcd: 200.1281, found: 200.1283.

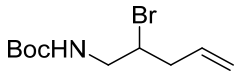
***Tert*-butyl (2-hydroxypent-4-en-1-yl)carbamate (2)**



2

To a solution of compound **1** (330 mg, 1.66 mmol, 1.0 eq) in absolute EtOH (20 mL) at -78°C, sodium borohydride (125 mg, 3.31 mmol, 2.0 eq) was added slowly. The reaction was stirred at -78°C for 2 h and quenched by saturated NH_4Cl solution (10 mL). The resulting mixture was extracted with EtOAc (50 mL ×3). The combined organic phase was dried over anhydrous $MgSO_4$, filtered and concentrated. The resulting residue was further purified on silica gel column chromatography (PE/EtOAc = 2:1) to give the title compound as a colorless oil (144 mg, 0.72 mmol, 43% yield). R_f = 0.43, (PE/EtOAc = 2:1). ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.86 - 5.76 (m, 1H), 5.18 - 5.14 (m, 1H), 5.13 (t, *J* = 1.0 Hz, 1H), 4.92 (s, 1H), 3.75 (tdd, *J* = 7.7, 5.1, 3.2 Hz, 1H), 3.33 (d, *J* = 13.6 Hz, 1H), 3.04 (dd, *J* = 14.1, 7.5 Hz, 1H), 2.33 - 2.15 (m, 2H), 1.45 (s, 9H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 134.00, 118.54, 79.74, 70.51, 39.35, 28.39, 28.00. HRMS: *m/z* for $C_{10}H_{20}NO_3^+$ $[M+H]^+$ calcd: 202.1438, found: 202.1430.

***Tert*-butyl (2-bromopent-4-en-1-yl)carbamate (3)**

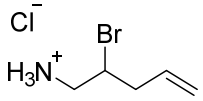


3

To a solution of compound **2** (101 mg, 0.5 mmol, 1.0 eq) and carbon tetrabromide (182 mg, 0.55 mmol, 1.1 eq) in CH_2Cl_2 (4 mL), triphenyl phosphine (144 mg, 0.55 mmol, 1.1 eq) was added slowly at 0°C. The reaction was then moved to room temperature until the TLC shows completion. The reaction was concentrated *in vacuo*. The residue was added to hexane and filtered. The filtrate was concentrated and further purified on silica gel column chromatography (PE/EtOAc = 8:1) to give the title compound as a colorless oil (32.7 mg, 0.12 mmol, 25% yield). R_f = 0.38, (PE/EtOAc = 8:1). ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.83 (ddt, 1H), 5.21 - 5.12 (m, 2H), 4.97 (s, 1H), 4.11 (dt, *J* = 12.2, 7.8 Hz, 1H), 3.64 (ddd, *J* = 14.2, 6.4, 3.8 Hz, 1H), 3.33 (ddd, *J* = 14.2, 7.8, 5.5 Hz, 1H), 2.69 - 2.53 (m, 2H), 1.45 (s, 9H). ^{13}C NMR (126 MHz, $CDCl_3$)

δ 155.82, 134.08, 118.65, 79.96, 55.35, 47.07, 40.58, 28.48. HRMS: *m/z* for $C_6H_{10}NO_2^+$ [M-Br-butyl+H]$^+$ calcd: 128.0707, found: 128.0705.

1-bromobut-3-en-1-amine hydrochloride (4)

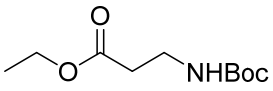


4

In a 25 mL round-bottom flask, compound **3** (16 mg, 0.61 mmol) was mixed with HCl solution in 1,4-dioxane (4 M, 2 mL) at 0°C. The reaction was then moved to room temperature and stirred vigorously for 2 h. The solvent was removed under reduced pressure to give the title compound as a white solid (12 mg, 0.60 mmol, 99% yield). ^{1}H NMR (600 MHz, DMSO-*d*$_6$) δ 8.24 (s, 3H), 5.80 (ddt, *J* = 17.0, 10.3, 6.8 Hz, 1H), 5.23 - 5.15 (m, 2H), 4.35 (tt, *J* = 8.8, 4.5 Hz, 1H), 3.14 (dd, *J* = 13.8, 8.9 Hz, 1H), 2.75 (m, 1H), 2.64 - 2.54 (m, 1H), 2.53 - 2.51 (m, 1H). ^{13}C NMR (151 MHz, DMSO-*d*$_6$) δ 133.96, 118.82, 51.24, 44.75, 40.06. HRMS: *m/z* for $C_5H_{11}{}^{79}BrN^+$, $C_5H_{11}{}^{81}BrN^+$ [M+H]$^+$ calcd: 164.0070, 166.0049, found: 164.0072, 166.0049.

5.4 Chemical Synthesis of the Warhead for DiUb-3Br

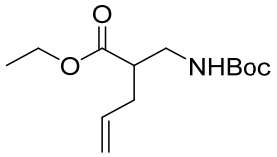
Ethyl 3-((tert-butoxycarbonyl)amino)propanoate (5)



5

To a solution of ethyl 3-aminopropanoate hydrochloride (1.54 g, 10.0 mmol, 1.0 eq) and triethylamine (2.8 mL, 20.0 mmol, 2.0 eq) in CH_2Cl_2 (30 mL), di-*tert*-butyl dicarbonate (2.62 g, 12.0 mmol,1.2 eq) was added dropwise at 0°C. The reaction was then moved to room temperature and stirred overnight. The reaction was diluted and washed with brine (50 mL ×3) and the organic phases were combined and dried over anhydrous $MgSO_4$. The filtrate was concentrated *in vacuo* to give the title compound as a colorless oil (1.92 g, 8.8 mmol, 88% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.02 (s, 1H), 4.14 (q, *J* = 7.1 Hz, 2H), 3.38 (q, *J* = 5.8 Hz, 2H), 2.50 (t, *J* = 6.1 Hz, 2H), 1.42 (s, 9H), 1.25 (t, *J* = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 172.64, 155.92, 79.45, 60.76, 36.22, 34.76, 28.50, 14.31. ESI-MS: *m/z* for $C_{10}H_{19}NO_4Na^+$ [M+Na]$^+$ calcd: 240.1, found: 240.2.

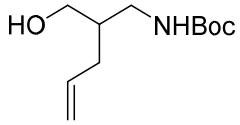
Ethyl 2-(((tert-butoxycarbonyl)amino)methyl)pent-4-enoate (6)



6

To a solution of compound **5** (660 mg, 3.0 mmol, 1.0 eq) in anhydrous THF (5 mL) at -78°C, lithium bis(trimethylsilyl)amide (1 M in THF, 6.1 mL, 2.0 eq) was added dropwise over 10 min. The mixture was stirred for 2 h followed by addition of allyl bromide (441 mg, 3.7 mmol, 1.2 eq) in anhydrous THF (5 mL) slowly. The reaction was stirred for 5 h at -78°C and monitored by TLC and LC-MS. The reaction was quenched by saturated NH_4Cl solution and extracted with EtOAc (30 mL × 3). The combined organic phases were dried over $MgSO_4$, filtered and concentrated. The residue was further purified on silica gel column chromatography (PE/EtOAc = 4:1) to give the title compound as a colorless oil (450 mg, 1.75 mmol, 58% yield). R_f = 0.31 (PE/EtOAc = 4:1). ^{1}H NMR (500 MHz, $CDCl_3$) δ 5.75 (ddt, *J* = 17.1, 10.1, 7.0 Hz, 1H), 5.13 - 5.03 (m, 2H), 4.87 (s, 1H), 4.16 (qd, *J* = 7.1, 2.9 Hz, 2H), 3.44 - 3.34 (m, 1H), 3.30 - 3.19 (m, 1H), 2.73 - 2.61 (m, 1H), 2.43 - 2.23 (m, 2H), 1.43 (s, 9H), 1.26 (t, *J* = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 174.46, 155.97, 134.66, 117.59, 79.50, 60.80, 45.35, 41.37, 34.17, 28.52, 14.41. HRMS: *m/z* for $C_{13}H_{24}NO_4^+$ $[M+H]^+$ calcd: 258.1700, found: 258.1693.

***Tert*-butyl (2-(hydroxymethyl)pent-4-en-1-yl)carbamate (7)**

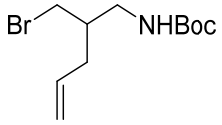


7

To a solution of compound **6** (257 mg, 1.0 mmol, 1.0 eq) in anhydrous THF (5 mL) under the protection of Ar, $LiBH_4$ (2 M in THF, 2.5 mL, 5.0 mmol, 5.0 eq) was added dropwise at 0°C over 5 min. The reaction was then moved to room temperature and stirred overnight. The reaction was monitored by TLC and quenched by saturated NH_4Cl solution (10 mL). The mixture was extracted with Et_2O (50 mL × 3). The combined organic phases were washed with brine (100 mL) and dried over anhydrous $MgSO_4$, filtered and concentrated. The residue was purified by silica gel column chromatography (PE/EtOAc = 4:1) to the title compound as a colorless oil (176 mg, 0.82 mmol, 82% yield). R_f = 0.17 (PE/EtOAc = 4:1). ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.77 (ddt, *J* = 17.1, 10.1, 7.1 Hz, 1H), 5.11 - 5.01 (m, 2H), 3.59 (dd, *J* = 11.7, 4.1 Hz, 1H), 3.41 (dd, *J* = 11.7, 6.8 Hz, 1H), 3.32 (dd, *J* = 14.5, 4.0 Hz, 1H), 3.10 (dd, *J* = 14.5, 6.8 Hz, 1H), 2.12

- 1.92 (m, 2H), 1.74 - 1.65 (m, 1H), 1.45 (s, 9H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 157.64, 136.48, 116.78, 80.13, 62.53, 41.46, 40.49, 33.59, 28.49. HRMS: *m/z* for $C_{11}H_{22}NO_3^+$ $[M+H]^+$ calcd: 216.1594, found: 216.1587.

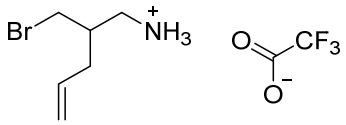
***Tert*-butyl (2-(bromomethyl)pent-4-en-1-yl)carbamate (8)**



8

To a solution of compound **7** (176 mg, 0.82 mmol, 1.0 eq) in CH_2Cl_2 (20 mL), tetracarbon bromide (298 mg, 0.9 mmol, 1.1 eq) and triphenyl phosphine (429 mg, 1.64 mmol, 2.0 eq) were added. The reaction was stirred at room temperature until the starting material was fully consumed. After evaporation of the solvent, silica gel column chromatography (PE/EtOAc = 5:1) was applied for purification to give the title compound as a colorless oil (65 mg, 0.23 mmol, 29% yield). R_f = 0.43 (PE/EtOAc = 5:1). ^{1}H NMR (600 MHz, $CDCl_3$) δ 5.74 (ddt, *J* = 17.2, 10.1, 7.2 Hz, 1H), 5.19 - 5.06 (m, 2H), 4.69 (s, 1H), 3.47 (m, 2H), 3.33 - 3.02 (m, 2H), 2.14 (t, *J* = 7.0 Hz, 2H), 2.00 (s, 1H), 1.44 (s, 9H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 156.15, 135.12, 117.84, 42.90, 39.63, 36.54, 34.88, 28.53. HRMS: *m/z* for $C_{11}H_{20}{}^{79}BrNO_2^+$, $C_{11}H_{20}{}^{81}BrNO_2^+$ $[M+H]^+$ calcd: 278.0750, 280.0730, found: 278.0746, 280.0723.

2-(bromomethyl)pent-4-en-1-amine trifluoroacetate (9)



9

To a solution of compound **8** (60 mg, 0.22 mmol) in CH_2Cl_2 (4 mL), trifluoroacetic acid (2 mL) was added. The reaction was stirred at room temperature for 2 h and evaporated *in vacuo* to give the title compound as a white solid (46 mg, 0.16 mmol, 73% yield). ^{1}H NMR (700 MHz, D_2O) δ 5.82 (ddt, *J* = 17.3, 10.1, 7.0 Hz, 1H), 5.23 (d, *J* = 17.1 Hz, 1H), 5.20 - 5.17 (m, 1H), 3.65 - 3.55 (m, 2H), 3.12 (ddd, *J* = 55.5, 13.2, 6.4 Hz, 2H), 2.31 - 2.18 (m, 3H). ^{13}C NMR (176 MHz, D_2O) δ 163.00 (d, *J* = 118.3 Hz), 134.23, 118.31, 116.31 (q, *J* = 291.7 Hz), 41.45, 37.11, 34.75, 34.15. HRMS: *m/z* for $C_6H_{13}{}^{79}BrN^+$, $C_6H_{13}{}^{81}BrN^+$ $[M+H]^+$ calcd: 178.02260, 180.0205, found: 178.0228, 180.0204.

5.6 Cloning, Protein Expression and Purification

Human USPL1 catalytic domain constructs (residues 212-514 and 218-502) were cloned from a cDNA template (coding sequence corresponding to UniProt: Q5W0Q7) into the pOPIN-K vector using the In-Fusion HD Cloning Kit (Takara Clonetech). Human SENP1 (residues 415-644, UniProt: Q9P0U3) and USP21 (residues 196-565, UniProt: Q9UK80) were subcloned into the pOPIN-B vector. Human UCHL3 (residues 1-229 in pOPIN-K, UniProt: P15374)[242], USP36 (residues 81-461 in pOPIN-K, UniProt: Q9P275)[243], USP2 (residues 258-605 in pOPIN-B, UniProt: O75604)[243], USP18 (residues 16-372 in pOPIN-B, UniProt: Q9UMW8)[243], and USP7 (residues 1-1102 in pFL, UniProt: Q93009)[244] were expressed and purified as described previously. Human ubiquitin (residues 1-75 and 1-76, UniProt: P0CG47) was subcloned into pTXB1 and pET17b. *Simkania negevensis* SnVTD (residues 74-326, Uni-Prot: F8L5E7) in pOPIN-K vector was generated from the Hofmann lab (University of Cologne)[245]. Ubiquitin (1-76) containing Lys to Cys mutations (Ub KxC) were generated based on the wildtype Ubiquitin pET17b construct. Site-directed mutagenesis was performed using a QuikChange protocol with Phusion Polymerase (M0530L, New England BioLabs).

For bacterial expression of USPL1, *Escherichia coli* Rosetta 2(DE3)pLacI cells (Merck) were chemically transformed and grown in 2x TY medium with antibiotics at 37°C. When the optical density at 600 nm (OD_{600}) reached ~0.8, the culture was cooled down to 18°C for 1 h and induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested after 12-16 h growth at 18°C by centrifugation at 4000 *g* for 10 min and purified immediately or stored at -80°C for further purification. For purification of USPL1, cell pellets were thawed and resuspended in binding buffer (50 mM sodium orthophosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0, 4 mM β-mercaptoethanol, supplemented with DNase I and lysozyme) and lysed using ultrasonication. Cell debris was discarded by centrifugation and the pooled lysate was filtered through a 0.45 µm filter. The clarified lysate was applied to a HisTrap Ni-NTA column (GE Healthcare) and then eluted by elution buffer (50 mM sodium orthophosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0, 4 mM β-mercaptoethanol). Protein-containing fractions were pooled, supplemented with His-3C protease, and dialyzed into binding buffer at 7°C overnight. The sample was then passed through a Ni-NTA column. The flow-through was collected and dialyzed into anion exchange low salt buffer (25 mM Tris, 50 mM NaCl, 4 mM DTT, pH 8.5) at 7°C overnight. The sample was applied to a Resource Q column (GE Healthcare) and purified with gradient elution from 50 to 500 mM NaCl. Pure fractions were collected, concentrated to ~2-3 mL and further purified by size-exclusion chromatography

(HiLoad 16/600 Superdex 75 pg, GE Healthcare) with SEC buffer (20 mM Tris, 100 mM NaCl, 4 mM DTT, pH 8.0). Fractions were analyzed on SDS-PAGE protein gels (NuPAGE 4-12%, Invitrogen) which were stained by Coomassie stains (ab119211, Abcam). Final protein samples were concentrated using 10K MWCO Amicon Ultra Filters (Merck Millipore), flash frozen in liquid nitrogen and stored at -80°C.

SENP1 was purified using a similar protocol as for USPL1 with the exception that binding, and elution buffers were adjusted to pH 7.6. The pooled fractions were digested by His-3C protease and dialyzed into binding buffer (pH 7.6). The sample was passed through a Ni-NTA column and the flow-through was collected, concentrated, and subjected to size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) with buffer (20 mM Tris, 100 mM NaCl, 4 mM DTT, pH 7.6). The final protein was concentrated, flash frozen in liquid nitrogen and stored at -80°C. USP21 was purified using the same protocol for USPL1 except for the SEC buffer which was supplemented with 5% glycerol.

Yeast ULP1 (residues 403-621 in pET19, UniProt: Q02724) was a kind gift from the Protein Chemistry Facility (Max Planck Institute of Molecular Physiology Dortmund).

For purification of human SUMO1 (1-97, UniProt: P63165), SUMO2 (1-93, UniProt: P61956), SUMO3 (1-92, UniProt: P55854), and SUMO3 pro-form (1-103), genes were subcloned into the pOPIN-B vector and expressed in Rosetta 2(DE3)pLacI cells as above. The proteins were purified on a Ni-NTA column as above. After His-3C cleavage, the proteins were passed through a Ni-NTA column again. The flow-through was collected, analyzed on a gel, and used without further purification.

To generate C-terminal thioesters (Ub/Ubl-MesNa), human Ub (1-75), Ub (1-76), SUMO1 (1-96), SUMO1 (1-97), SUMO2 (1-92), SUMO2 (1-93), ΔN-SUMO2 (18-92), SUMO3 (1-91), NEDD8 (1-76, UniProt: Q15843), C-terminal ISG15 (ISG15(CTD), 77-157, C78S, UniProt: P05161) and FUBI (1-74, C57A, UniProt: P62861) were subcloned into the pTXB1 vector and expressed in Rosetta 2(DE3)pLacI cells as described as above. Cells were resuspended in buffer A (20 mM HEPES, 50 mM sodium acetate, 75 mM NaCl, pH 6.5) supplemented with DNase I and lysed using ultrasonication (all buffer supplemented with 5% glycerol for Fubi purifications). The suspension was cleared by centrifugation and filtered through a 0.45 µm filter. The clarified lysate was loaded onto pre-equilibrated chitin resin (S6651L, New England BioLabs) and incubated with over-head rotation at 7°C overnight. The resin was washed with 1 L high salt buffer (20 mM HEPES, 50 mM sodium acetate, 500 mM NaCl, pH 6.5) and then

washed with 100 mL buffer A. Elution was initiated by addition of 50 mL sodium 2-mercaptoethanesulfonate (MesNa, 100 mM in buffer A). After incubation at 7°C for 24-36 h the eluate was collected and concentrated to less than 5 mL. Size-exclusion chromatography (HiLoad 26/600 Superdex 75 pg, GE Healthcare) was used to purify the protein into buffer A. The fractions were analyzed on a gel, concentrated, and stored at -80°C. Protein mass spectrometry revealed full cleavage of the N-terminal methionine for both SUMO2 (1-92) and SUMO3 (1-91) proteins. FUBI-MesNa was collected after eluted from chitin column, concentrated, flash frozen in liquid nitrogen and stored at -80°C for further use.

Selenomethionine (SeMet)-containing protein expression was conducted by the methionine pathway inhibition method in *E. coli* B834(DE3) cells (Novagen)[246]. The pTXB1 plasmid encoding SUMO3 (1-91) was freshly transformed into cells, and a culture in 2xTY medium with antibiotics was inoculated in the morning. After 10 h, a pre-culture in M9 minimal medium was started with a 1:1000 inoculum. On the second day, a large culture in M9 minimal medium was started with a 1:100 inoculum. When the OD_{600} reached 0.6, feed-back inhibition amino acids including 1.0 g each of lysine, threonine and phenylalanine, 0.5 g each of leucine, isoleucine and valine, and 0.5 g of selenomethionine for 10 L of culture were added. After shaking for 15 min, the culture was cooled down to 18°C for 1 h and then induced by 0.5 mM IPTG for 16 h. Protein purification was performed as above. Complete selenomethionine incorporation was validated by protein LC-MS analysis (mass calculated for SeMet-SUMO3-MesNa: 10600.4, found: 10601.2).

RanGAP1 (399-587) was cloned from a synthesized gene into pOPIN-B and expressed in *E. coli* Rosetta 2(DE3)pLacI cells. Purification was carried as described above by a Ni-NTA column, cleavage by His-3C protease, reverse Ni-NTA purification and anion exchange chromatography, similar to what was previously reported[247].

Human SAE1/SAE2 (from a pETDuet1 plasmid encoding GST-SAE1 and His6-UBA2), and UBE2I (from a pGEX-6P1 plasmid) were expressed in *E. coli* Rosetta 2(DE3) pLacI cells. Purification of SAE1/SAE2[248] and UBE2I[247] were carried out similarly to what was described previously.

For bacterial expression of SnVTD wildtype, *E. coli* Rosetta 2(DE3)pLacI cells (Merck) were chemically transformed and grown in 2xYT medium supplemented with chloramphenicol (34 µg/mL) and kanamycin (50 µg/mL) at 37°C. When the OD_{600} reached 0.6-0.8, the culture was cooled down to 18°C for 1 h and induced with 0.5 mM IPTG. Cells were harvested after 12-16

h growth at 18°C by centrifugation at 4000 *g* for 10 min and purified immediately or stored at -80°C for further purification. For purification of SnVTD, cell pellets were resuspended in binding buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0, 4 mM β-mercaptoethanol) and supplemented with DNase I, lysozyme and a tablet of EDTA-free protease inhibitor cocktail (Roche), and lysed in an ice bath using ultrasonication. Cell debris was discarded by centrifugation at 60000 *g* for 30 min and the combined supernatant was filtered through a 0.45 μm filter. The clarified lysate was applied to a HisTrap Ni-NTA column (GE Healthcare) and then eluted by elution buffer (50 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0, 4 mM 2-mercaptoethanol) using an ÄKTA pure chromatography system. Fractions showed UV 280 nm signal were pooled, mixed with His-3C protease, and dialyzed into binding buffer at 7°C overnight. The sample was then passed through a Ni-NTA column using a peristaltic pump. The flow-through was collected, concentrated, and further purified by size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) with buffer (20 mM Tris, 100 mM NaCl, 4 mM DTT, pH 8.0). Fractions were analyzed on SDS-PAGE protein gels which were visualized by Coomassie stains. Pure fractions were pooled and concentrated using 10K MWCO Amicon Ultra Filters (Merck Millipore), flash frozen in liquid nitrogen and stored at -80°C. For SnVTD mutants, pure proteins were obtained after two rounds of reverse Ni-NTA chromatography. The mutants were concentrated, flash frozen in liquid nitrogen and stored at -80°C.

To purify Ub KxC, cell pellets were thawed, resuspended in the buffer (50 mM Tris, pH 7.4, 1 mM EDTA) supplemented with lysozyme and lysed using ultrasonication. The cell debris was discarded by centrifugation at 60000 *g* for 30 min. The supernatant was transferred to a glass beaker on ice and stirred vigorously. The pH was adjusted to around 2.5 by adding 70% perchloric acid dropwise. The mixture was further stirred on ice for 10 min and spun down at 60,000 *g* for 30 min. The supernatant was pooled and dialyzed into buffer (50 mM NaOAc, pH 4.5) overnight. The sample was further clarified by a 0.22 μm filter and subjected to a cation exchange column (HiPrep SP FF 16/10, Cytiva). The protein was eluted with a gradient of 0-1 M NaCl. The fractions were analyzed on SDS-PAGE protein gels (NuPAGE 4-12%, Invitrogen) which were stained by Coomassie stains. Pure fractions were pooled and concentrated, and further purified by size-exclusion chromatography (HiLoad 16/600 Superdex 75pg, GE Healthcare) with buffer (250 mM NaOAc, 0.5 mM TCEP, pH 5.1). Pure fractions were pooled and concentrated using 3K MWCO Amicon Ultra Filters (Merck Millipore), flash frozen in liquid nitrogen and stored at -80°C.

Protein concentrations were determined by absorption at 280 nm unless noted otherwise. Intact protein mass spectrometry was recorded on an Agilent 1260 II Infinity system (Agilent, with Openlab software) and analyzed as reported previously[242]. The spectra were deconvoluted by ProMass 3.0.12 (Novatia, LLC). Exported data were plotted using GraphPad Prism 9.

5.7 Preparation and Purification of Substrates

Preparation and purification of Ub/SUMO1/2-RhoG Substrates

Reaction conditions were optimized from a previously reported method for Ubiquitin-RhoG[199]. To a solution of SUMO2(1-92)-MesNa (511 μM, ~10 mL in buffer A) were added 10 eq. *N*-hydroxysuccinimide (1 M in buffer A, pH adjusted to 7.5), 10 eq. *sym*-collidine and 5 eq. 2G-Rho (50 mM in DMSO, the final DMSO concentration is below 5%, see above for the synthesis of 2G-Rho), the pH which was measured by a pH meter (Mettler Toledo) was adjusted to 8.0 using 4 N NaOH and then the reaction mixture was heated to 37°C. The reaction was monitored by LC-MS until most of the starting material was converted to product or hydrolyzed. The reaction was then diluted and dialyzed against low salt cation exchange buffer (20 mM MES, pH 6.0, 50 mM NaCl, 1 mM TCEP) overnight. The sample was filtered, the filtrate applied to a Resource S column, and the product eluted with a gradient of 50-1000 mM NaCl. The fractions were analyzed on a gel and evaluated by intact protein mass spectrometry. Pure fractions were pooled and dialyzed into buffer (20 mM MES, pH 6.0, 100 mM NaCl). The final product was filtered and concentrated to ~1 mg/mL. SUMO1-RhoG was prepared and purified using the same method. Ub-RhoG was purified using the buffer with the same components, but the pH was adjusted to 6.5. The concentration was measured based on a calibration curve made from serial dilution of RhoG solution in aqueous buffer. The final products were validated by LC-MS. SUMO1-RhoG, calcd: 11370.7, found: 11369.8; SUMO2-RhoG, calcd: 10847.1, found: 10847.9; Ub-RhoG, calcd: 8934.2, found: 8938.1.

Preparation of Ub/Ubl-$NHNH_2$

Ub/Ubl-MesNa were treated with hydrazine monohydrate (500 μL added into 100 mL eluate). After 30 min at ambient temperature, the reaction was concentrated, filtered, and purified by size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) using buffer (20 mM MES, 50 mM NaCl, pH 6.7). HEPES salt cannot be used since it has a free hydroxyl

which might affect next step reaction by nitrous acid. Ub/Ubl-$NHNH_2$ proteins were validated by LC-MS: Ub-$NHNH_2$ calcd: 8578.9, found: 8582.1; SUMO1-$NHNH_2$ calcd: 11015.4, found: 11014.6; SUMO2-$NHNH_2$ calcd: 10491.8, found: 10492.3; NEDD8-$NHNH_2$ calcd: 8574.0, found: 8573.4; ISG15(CTD)-$NHNH_2$ calcd: 9261.6, found: 9261.1; Fubi-$NHNH_2$ calcd: 7741.8, found: 7741.1.

Preparation and purification of Ub-KG-TAMRA

To a solution of $Ub^{1\text{-}76(wt)}$-$NHNH_2$ (2 mM, 100 µL) at -10°C were added sodium nitrite (1 M, 50 µL) and citric acid (200 mM, 50 µL). After 2 min, KG-TAMRA (10 mM in 1.5 M HEPES pH 8.0, 200 µL) was added and the reaction was incubated at 30°C for 3 min. The reaction was analyzed by LC-MS, diluted to 40 mL, and dialyzed into buffer (20 mM MES, pH 6.0) overnight. The mixture was purified by cation exchange chromatography (Resource S, GE Healthcare) in the same buffer with gradient elution from 0 to 300 mM NaCl. Protein-containing fractions were pooled and dialyzed into low pH buffer (20 mM NaOAc, pH 4.5). These were further purified by cation exchange chromatography (Resource S, GE Healthcare) in low pH buffer with gradient elution from 0 to 300 mM NaCl. Ub-KG-TAMRA was concentrated, flash frozen in liquid nitrogen, stored at -80°C and validated by LC-MS.

Preparation and purification of SUMO1-KG-TAMRA

To a solution of $SUMO1^{1\text{-}76(wt)}$-$NHNH_2$ (1.8 mM, 100 µL) at -10°C were added sodium nitrite (1 M, 50 µL) and citric acid (200 mM, 50 µL). After 2 min, KG-TAMRA (20 mM in 1.5 M HEPES, pH 8.0, 200 µL) was added and the reaction was incubated at 30°C for 3 min. The reaction was analyzed by LC-MS upon which nitrosylated species were found. Nitrosylations were reversed by addition of TCEP solution (1 M in 20 mM MES, pH 7.0, 1 mL). The reaction mixture was then passed through a 0.45 µm filter and purified by size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) with isocratic elution in buffer (20 mM Tris, 50 mM NaCl, 1 mM DTT, pH 8.0). Protein-containing fractions were pooled, diluted with 20 mM Tris buffer (pH 7.0) and purified on a high-resolution anion exchange column (Capto HiRes Q 5/50, Cytiva) with gradient elution from 0 to 500 mM NaCl. SUMO1-KG-TAMRA-containing fractions were identified by LC-MS, concentrated, flash frozen in liquid nitrogen, stored at -80°C.

Preparation and purification of SUMO2-KG-TAMRA

To a solution of SUMO2$^{1\text{-}93(wt)}$-$NHNH_2$ (7.6 mM, 80 µL) at -10°C were added sodium nitrite (1 M, 40 µL) and citric acid (200 mM, 40 µL). After 2 min, KG-TAMRA (20 mM in 1.5 M HEPES, pH 8.0, 150 µL) was added and the reaction was incubated at 30°C for 3 min. The reaction was analyzed by LC-MS, diluted, and dialyzed into buffer (20 mM MES, pH 6.0) overnight which was shielded from light. The mixture was purified by cation exchange chromatography (Resource S, GE Healthcare) with gradient elution from 0 to 300 mM NaCl. Protein-containing fractions were pooled, dialyzed into buffer (20 mM HEPES, pH 8.0) and purified by size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) in the same buffer. SUMO2-KG-TAMRA was concentrated, flash frozen in liquid nitrogen, stored at -80°C and validated by LC-MS.

Preparation and purification of NEDD8-KG-TAMRA

To a solution of NEDD8$^{1\text{-}76(wt)}$-$NHNH_2$ (1.77 mM, 100 µL) at -10°C were added sodium nitrite (1 M, 50 µL) and citric acid (200 mM, 50 µL). After 3 min, KG-TAMRA (5 mM in 1.5 M HEPES, pH 8.0, 200 µL) was added and the reaction was incubated at 30°C for 5 min. The reaction mixture was analyzed by LC-MS, passed through a 0.45 µm filter and purified by size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) with isocratic elution in buffer (20 mM NaOAc, 50 mM NaCl, pH 5.0). Protein-containing fractions were pooled and further purified via cation exchange chromatography (Resource S, GE Healthcare) with gradient elution from 50 to 500 mM NaCl. Fractions were analyzed using LC-MS. The product NEDD8-KG-TAMRA was concentrated, flash frozen in liquid nitrogen, stored at -80°C and validated by LC-MS.

Preparation and purification of ISG15(CTD)-KG-TAMRA

To a solution of ISG15$^{Met+77\text{-}157(C78S)}$-$NHNH_2$ (1.77 mM, 150 µL) at -10°C were added sodium nitrite (1 M, 75 µL) and citric acid (200 mM, 75 µL). After 3 min, KG-TAMRA (10 mM in 1.5 M HEPES, pH 8.0, 200 µL) was added and the reaction was incubated at 30°C for 3 min. The reaction mixture was then passed through a 0.45 µm filter and purified by size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) with isocratic elution in buffer (20 mM Tris, 50 mM NaCl, pH 7.0). Protein-containing fractions were pooled and dialyzed into a high salt buffer (50 mM NaH_2PO_4, 1.5 M $(NH_4)_2SO_4$, pH 7.0) overnight. The

sample was further purified using hydrophobic interaction chromatography (HiTrap Capto Phenyl ImpRes, Cytiva) with gradient elution from 1.5 M to 0 M $(NH_4)_2SO_4$. Fractions were analyzed using LC-MS. The product ISG15(CTD)-KG-TAMRA was concentrated and buffer-exchanged into low salt buffer (50 mM NaH_2PO_4, pH 7.0), flash frozen in liquid nitrogen, stored at -80°C and validated by LC-MS.

Preparation and purification of FUBI-KG-TAMRA

To a solution of $FUBI^{1-74(C57A)}$-$NHNH_2$ (100 µL) at -10°C was added sodium nitrite (1 M, 50 µL) and the mixture was incubated for 2 min. Citric acid (200 mM, 50 µL) was then added. After 3 min, KG-TAMRA (20 mM in 1.5 M HEPES, pH 8.0, 200 µL) was added and the reaction was incubated at 30°C for 30 min. The reaction was diluted, and the buffer was exchanged to 50 mM NaCl, 20 mM MES, 5% glycerol, pH 6.5 using size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare). Protein-containing fractions were further purified by high-resolution anion exchange chromatography (Capto HiRes Q 5/50, GE Healthcare) with gradient elution from 50 to 500 mM NaCl. Purity of the fractions was assessed by LC-MS. Pooled FUBI-KG-TAMRA was concentrated, flash frozen in liquid nitrogen, stored at -80°C and validated by LC-MS.

Preparation and purification of RanGAP1~SUMO2

RanGAP1~SUMO2 wildtype and mutant conjugates were all assembled by the enzymatic catalysis of SAE1/SAE2 and UBE2I as described previously[249]. In general, 5 mL reactions in transport buffer (20 mM HEPES, 110 mM NaOAc, 2 mM $Mg(OAc)_2$, 1 mM DTT, pH 7.3 using 1 N KOH) containing 3.4 µM RanGAP1, 50 nM SAE1/SAE2, 30 nM UBE2I and 12 µM SUMO2 wildtype or mutants were initiated by addition of 2.5 mM ATP and incubated at 37°C for 2-3 h. The mixture was subjected to size-exclusion chromatography (HiLoad 26/600 Superdex 75 pg, GE Healthcare) with isocratic elution into transport buffer. The fractions were analyzed on a gel. The fractions containing pure fractions were pooled, flash frozen in liquid nitrogen and store at -80°C for further use.

5.8 Preparation and Purification of Probes

Preparation of SUMO3-2Br and ΔN-SUMO2-PA probes

To a solution of SUMO3(1-91)-MesNa (958.7 µM, 3.5 mL), 2-bromoethylamine hydrobromide (1.6 g in ~5 mL of buffer A) was added and the pH was adjusted by addition of 4 M NaOH to 8.0-8.5 at room temperature. The reaction was monitored by LC-MS. After 1-2 h, SUMO3(1-91)-MesNa was fully consumed, and the mixture was subjected to size-exclusion chromatography (HiLoad 26/600 Superdex 75 pg, GE Healthcare) in buffer A. The protein-containing fractions were pooled, concentrated, and stored at -80°C by flash freezing for further use. ΔN-SUMO2-PA was generated using a similar procedure by mixing ΔN-SUMO2-MesNa (767.5 µM, 1.5 mL) and propargylamine hydrocholoride (600 mg in ~3 mL of buffer A).

Preparation of diUb-2Brand diUb-3Br probes

In a 1.5 mL centrifuge tube, Ub-$NHNH_2$ (2 mM, 100 µL) was mixed with sodium nitrite (1 M, 50 µL) and citric acid (200 mM, 50 µL) at -10°C for 3 min. Then a solution of compound **4** or **9** (200 µL, 40-80 mM in 1.5 M HEPES, pH 8.0) was added and the resulting mixture was incubated at 45°C for 3 min. The reaction was monitored by LC-MS and purified by size-exclusion chromatography (HiLoad 16/600 Superdex 75pg, GE Healthcare) with buffer (250 mM NaOAc, 0.5 mM TCEP, pH 5.1). The protein-containing fractions were pooled and concentrated using 3K MWCO Amicon Ultra Filters to around 2 mM. The products Ub-2Br-ene and Ub-3Br-ene were analyzed by LC-MS. To generate diUb-2Br or diUb-3Br probes, Ub-2Br-ene or Ub-3Br-ene (50 µL) was mixed with Ub (KxC, 2 mM, 50 µL) in a PCR tube. LAP (CAS no. 85073-19-4, 1 µL, 50 mM stock solution in 250 mM NaOAc, 0.5 mM TCEP, pH 5.1) was added and the reaction was exposed to UV 365 nm (Herolab, Cat. No. H468.1, 8 W) on ice for 10-60 min which was monitored by LC-MS. The reaction was subjected to size-exclusion chromatography (HiLoad 16/600 Superdex 75pg, GE Healthcare) with buffer (20 mM MES, 100 mM NaCl, pH 6.5). Fractions were analyzed on SDS-PAGE protein gels and visualized by Coomassie stains. The diUb probes were concentrated, characterized by LC-MS, flash frozen in liquid nitrogen and stored at -80°C for further use.

5.9 Crystallography

USPL1 in complex with SUMO2/3

Covalent complexes of USPL1 (218-502) with SUMO3 probes were prepared by mixing USPL1 (1.0 eq, 0.19 mg/mL, after reverse Ni-NTA purification step in Ni-NTA binding buffer)

and SUMO3 probes (4.0 eq, 531 µM, fully purified) at room temperature for 4 h. The reaction mixture was subjected to purification on a Resource Q column as described above. Pure fractions containing the covalent complex were then concentrated, and buffer exchanged into SEC buffer (20 mM Tris, 100 mM NaCl, 4 mM DTT, pH 8.0). Since USPL1 could not be produced with SeMet labeling in high enough yields, crystals for the SAD dataset were prepared from unlabeled USPL1 and SeMet-containing SUMO3-2Br probe (3 Se sites per 378 residues).

Crystallization experiments were carried out at 18°C in 96-well sitting drop vapor diffusion plates (MRC format, Molecular Dimensions). Plates were set up using a Mosquito HTS robot (TTP Labtech) and fine screen plates were prepared with a Dragonfly robot (TTP Labtech). Coarse screening was carried out by mixing 200 nL of protein solution with 200 nL of reservoir. Fine screens were assembled in 600-900 µL drops with ratios as described below. USPL1~SUMO3-2Br (7.0 mg/mL) was crystallized in 100 mM Tris pH 8.8, 19% PEG 6000, 200 mM $CaCl_2$, from mixing protein : reservoir in a 1:2 ratio. USPL1~SeMet-SUMO3-2Br (5.4 mg/mL) was crystallized in 100 mM Tris pH 8.3, 21% PEG 4000, 200 mM $CaCl_2$, from mixing protein : reservoir in a 2:1 ratio. Crystals were collected with the protection of mother liquor containing 25% glycerol and vitrified in liquid nitrogen. USPL1~ΔN-SUMO2-PA (8.6 mg/mL) was crystallized in 100 mM CHES pH 8.9, 34% PEG 600, from mixing protein : reservoir in a 2:1 ratio. Crystals were cryo-protected in 100 mM CHES pH 9.2, 45% PEG 600, and vitrified in liquid nitrogen.

Diffraction data were collected at 100 K at the Swiss Light Source (SLS, Paul Scherrer Institute, Villingen, Switzerland) on beamlines PX2 and PX3. Images were integrated using either XDS[250] (for SeMet data from four 360° sweeps with 5° increments using the PriGo goniometer) or DIALS[251] (for native datasets), and scaled using Aimless[252]. Owing to the low sequence homology between the USPL1 catalytic domain and other previously crystallized members of the USP family (highest identity: USP7 at 16%), various molecular replacement efforts from the native datasets were not successful. The structure of USPL1~SUMO3-2Br was solved experimentally through the CRANK2 pipeline by a SAD protocol as implemented in the CCP4 suite of programs from the SeMet dataset involving substructure detection by SHELXD[253], density modification by Parrot[254] and model building/refinement by Buccaneer/Refmac[255]. The resulting initial model was then used to obtain phases for the native dataset, and the final model was obtained through several rounds of manual building in Coot[256] and refinement by Phenix.Refine[257]. The USPL1~ΔN-SUMO2-PA structure was solved through molecular replacement with Phaser[258] with one copy of the USPL1-SUMO3-2Br complex as search model.

See Table S1 for final statistics. Protein structures were deposited in the protein data bank with accession codes 7ZJU and 7ZJV.

SnVTD in complex with diUb-3Br

To obtain the covalent complex of SnVTD~diUb-3Br, SnVTD (12.0 μM) was incubated with K6-linked diUb-3Br (14.6 μM) in the reaction buffer (50 mM HEPES, 100 mM NaCl, 1 mM TCEP, pH 8.0) at room temperature for 3 h and the reaction was moved to 7°C overnight. The reaction was concentrated to 5 mL and purified on size-exclusion chromatography (HiLoad 16/600 Superdex 75 pg, GE Healthcare) with buffer (25 mM Tris, 4 mM DTT, pH 8.5). The pooled fractions which contained SnVTD~diUb was purified by anion exchange chromatography (Resource Q, GE Healthcare) with a gradient from 0 to 500 mM NaCl at pH 8.5. The fractions were analyzed on an SDS-PAGE gel. The fractions containing complex were pooled, concentrated, and purified on size-exclusion chromatography with buffer (25 mM Tris, 50 mM NaCl, 4 mM DTT, pH 8.0). The fractions were analyzed on an SDS-PAGE gel and complex-containing fractions were collected and purified by another round of anion exchange chromatography with a gradient from 0 to 300 mM NaCl at pH 8.0. The fractions were analyzed on an SDS-PAGE gel and the pure fractions were collected and exchanged into buffer (25 mM Tris, 100 mM NaCl, 4 mM DTT, pH 7.5) using a 10K MWCO Amicon Ultra Filter and concentrated to 6.94 mg/mL. The protein was mixed with reservoir solution in a ratio of 1:1 in SWISSCI MRC 3-Well sitting drop plates and grew for 7 weeks. Crystals were found in the well H9 of the Morpheus screening plate, collected in mother liquor (20 mM DL-glutamic acid monohydrate, 20 mM DL-alanine, 20 mM glycine, 20 mM DL-lysine monohydrochloride, 20 mM DL-serine, 100 mM Tris and Bicine, pH 8.5, 20% v/v PEG 500 MME, 10% PEG 20000) supplemented with 10% glycerol, and vitrified in liquid nitrogen.

Diffraction data were collected at beamline P14 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). Diffraction images were integrated by DIALS[259], and anisotropic data scaling was performed by the STARANISO server (Global Phasing Ltd)[260]. The structure was solved by molecular replacement using a hybrid search model based on the AlphaFold predicted SnVTD structure[261, 262] and a reported M48 USP structure(PDB ID: 2J7Q)[228]. The model was automatically built by Buccaneer in CCP4 v8.0[263]. The structure was further refined by Phenix[264] and manually inspected in Coot[265] iteratively. The final structure was deposited in protein data bank with the accession code 8RL3.

5.10 Biochemical and Biophysical Assays

Fluorescence intensity assay

Enzymes and Ub/SUMO-RhoG substrates were diluted into assay buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM DTT, 0.01 mg/mL BSA). Substrates were used at 50 nM of final concentration unless noted otherwise. Assays were carried out in black 384 well low volume non-binding plates (Greiner, 784900) with 20 µL total volume. After addition of enzyme, fluorescence (ex/em = 485/535 nm) was recorded for one hour on a Tecan Spark plate reader at ambient temperature.

For evaluation of USPL1 inhibitors, buffer supplemented with 0.01% Tween-20 and indicated reducing agents. USPL1 was used at a final concentration of 60 pM. USPL1 (5 µL) was first incubated with inhibitors (5 µL) for indicated time at ambient temperature, then SUMO2-RhoG (10 µL, 200 nM) was added into each well. The fluorescence was recorded according to aforementioned methods in triplicate. Initial velocity was converted into inhibitory rates based on the positive and negative controls. Inhibitory rates were plotted against inhibitor concentrations which were analyzed by GraphPad Prism 9 using four parameters dose-response analysis to calculate the IC_{50} values.

Fluorescence polarization assay

Enzymes and substrates were diluted into assay buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM DTT, 0.01 mg/mL BSA). Assays were performed in black 384 well low volume non-binding plates (Greiner 784900) with 20 µL total volume. For USP16, USP36, USP7, USP2 (single concentration with the five different substrates, and all USP36 experiments) and SnVTD, assays were performed in buffer (20 mM Tris, pH 8.0, 0.01% Triton X-100, 0.1 mg/mL BSA, 1 mM TCEP) in 384-well low volume black round bottom plates (Corning, 4514) with 20 µL total volume.

For binding assays, KG-TAMRA and Ub-KG-TAMRA were used at 3 nM of final concentration and SUMO1-KG-TAMRA and SUMO2-KG-TAMRA were used at 1 nM of final concentration unless noted otherwise. These assays were performed at least twice independently.

For cleavage assays, substrates were used at 100 nM of final concentration unless noted otherwise. After addition of enzyme, fluorescence polarization (ex/em = 535/590 nm) was recorded

on a Tecan Spark plate reader for one hour at ambient temperature. These assays were performed at least twice independently.

Data analysis for cleavage assays, anisotropy curves were subjected to nonlinear curve fitting to obtain the exponential decay in fluorescence (fit: plateau followed by one-phase exponential decay). The observed rate constants were then plotted over the enzyme concentration to give catalytic efficiencies as the slopes. For binding assays, anisotropies of replicate wells were averaged and plotted in GraphPad Prism 9 for nonlinear curve fitting to determine the affinity constant, K_d (fit: one-site binding - total).

Gel-based RanGAP1~SUMO2 cleavage assay

USPL1 (5 nM) and RanGAP1~SUMO2 (5 μM) were mixed at different time points and incubated at 37°C in assay buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM DTT, 0.01 mg/mL BSA). The reaction was quenched by adding lithium dodecyl sulfate (LDS) sample loading buffer (Thermo Fisher Scientific, NP0008) and analyzed on NuPAGE gels with Coomassie stains.

Gel-based pro-SUMO3 cleavage assay

USPL1 wildtype and mutants (500 nM) were mixed with full-length SUMO3 (10 μM) and incubated at 37°C for 2 h in assay buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM DTT, 0.01 mg/mL BSA). The reaction was quenched by adding LDS sample loading buffer and analyzed on NuPAGE gel with Coomassie stains.

Gel-based diUb cleavage assay

SnVTD wildtype and mutants and Lys6-linked diUb were diluted in assay buffer (50 mM HEPES, 100 mM NaCl, 1 mM TCEP, pH 8.0). Enzymes and diUb were mixed in a volume ratio of 1:1 and incubated at 37°C. Samples were taken from the reaction mixture and quenched with LDS sample loading buffer. The reactions were analyzed on SDS-PAGE gels which were visualized by Coomassie stains.

Gel-based diUb-2Br and diUb-3Br binding assay

Probes and enzymes were diluted in reaction buffer (50 mM HEPES, 100 mM NaCl, 1 mM TCEP, pH 8.0), mixed in a volume ratio of 1:1 and incubated at room temperature or 37°C for indicated time. The reactions were quenched with LDS sample loading buffer, resolved on an SDS-PAGE gel and visualized by Coomassie stains. For UBE3C, the reaction buffer was supplemented with 10% glycerol to avoid protein aggregation.

Thermal shift assay

For USPL1, assay was performed in triplicate in phosphate-buffered saline (PBS) containing 4 mM DTT. 5 µL USPL1 (10 µM) and 25 µL SUMO3 (19.2 µM) probe were mixed and incubated at room temperature for 30 min, then 10× Sypro-Orange (Invitrogen, S6650) (10 µL) was added.

For SnVTD, assay was performed in PBS buffer supplemented with 4 mM DTT. SnVTD apo (8 µM) or purified SnVTD~diUb-3Br covalent complex (8 µM) was mixed with the same volume of 10× SYPRO Orange. The assays for SnVTD were performed at the volume of 20 µL in quadruplicate in a white 96-well PCR plate (MLL9651, Bio-Rad) and the melting curves were recorded in the FRET mode on a BioRad CFX96 qPCR instrument. All the assays were repeated independently at least twice.

Circular dichroism spectroscopy

Protein samples were buffer exchanged into 10 mM sodium phosphate pH 7.0 and diluted to 0.1 - 0.5 mg/mL. CD spectra were measured on a Jasco J-815 spectrophotometer at 20 °C with 100 nm/min scan speed. Data were recording from 190 to 350 nm with a 1 nm pitch and as averages of three technical replicates. Protein concentrations of reference and sample proteins were adjusted by Coomassie-stained SDS PAGE gels and densitometry.

6. APPENDIX

Appendix Table 1. Data collection, phasing and refinement statistics of USPL1 crystals.

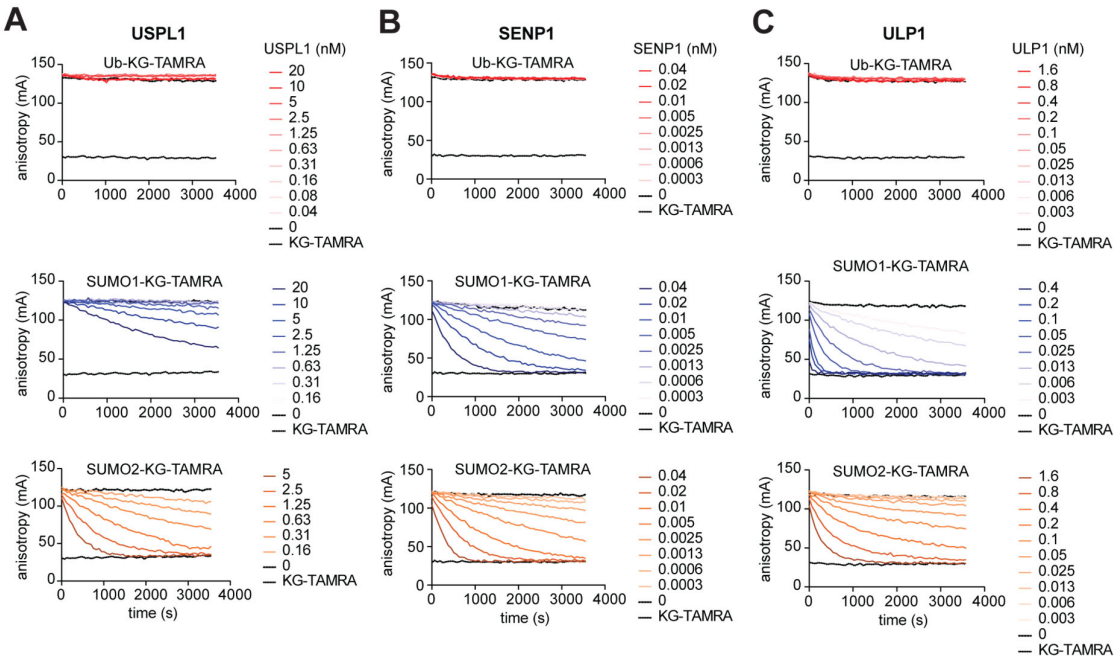
	USPL1~SUMO3-2Br (SeMet)	USPL1~SUMO3-2Br (PDB code: 7ZJU)	USPL1~ΔN-SUMO2-PA (PDB code: 7ZJV)
Data collection			
Beamline	SLS – PX3	SLS – PX2	SLS – PX2
Wavelength	0.9792 Å	1.000 Å	1.000 Å
Space group	$P\,2_1$	$P\,2_1$	$P\,4_1\,2_1\,2$
Cell dimensions			
a, *b*, *c* (Å)	66.84, 84.70, 71.58	66.06, 84.83, 71.11	95.71, 95.71, 82.88
α, *β*, *γ* (°)	90.0, 92.6, 90.0	90.0, 92.3, 90.0	90, 90, 90
Observed reflections	654,392	237,428	141,889
Unique reflections	24,690	41,570	15,670
Resolution (Å)	47.72 – 2.60 (2.72 – 2.60)	54.47 – 2.17 (2.25 – 2.17)	42.80 – 2.40 (2.49 – 2.40)
R_{merge}	0.221 (2.095)	0.112 (0.640)	0.053 (0.858)
R_{meas}	0.226 (2.149)	0.123 (0.711)	0.057 (0.933)
$I/\sigma(I)$	15.8 (1.7)	8.5 (2.4)	18.2 (1.8)
$CC_{1/2}$	0.999 (0.764)	0.996 (0.887)	0.999 (0.718)
Completeness (%)	100 (100)	99.9 (99.6)	100 (99.9)
Redundancy	26.5 (20.3)	5.7 (5.0)	9.1 (6.9)
Wilson *B* (Å^2)	42.3	33.8	79.3
Phasing			
Method	SAD	MR	MR
Resolution	2.6 Å		
Anom. completeness (%)	99.9 (99.8)		
Anom. multiplicity	13.5 (10.2)		
<FOM>	0.216		
Refinement			
Resolution		2.17 Å	2.40 Å
No. reflections		41549	15583
R_{work} / R_{free} (%)		19.2 / 23.8	23.0 / 25.8
No. atoms		5,527	2,659
Protein		5,166	2,621
Ligands		30	10
Water		331	28
B factors (Å^2)		45.3	110
Protein (Å^2)		45.0	110
Ligands (Å^2)		55.1	112
Water (Å^2)		49.5	103
R.m.s.d.			
Bond lengths (Å)		0.003	0.005
Bond angles (°)		0.64	1.13
Ramachandran (favored / allowed / outlier) (%)		97.2 / 2.8 / 0	97.0 / 3.0 / 0
Clashscore		3.5	6.0
Rotamer outliers (%)		0.6	1.5
Copies / a.s.u.		2	1

Each dataset was collected from a single crystal. Values in parentheses are for highest-resolution shells. Anom, anomalous; FOM, figure of merit; MR, molecular replacement; r.m.s.d., root mean square deviations; SAD, single-wavelength anomalous dispersion.

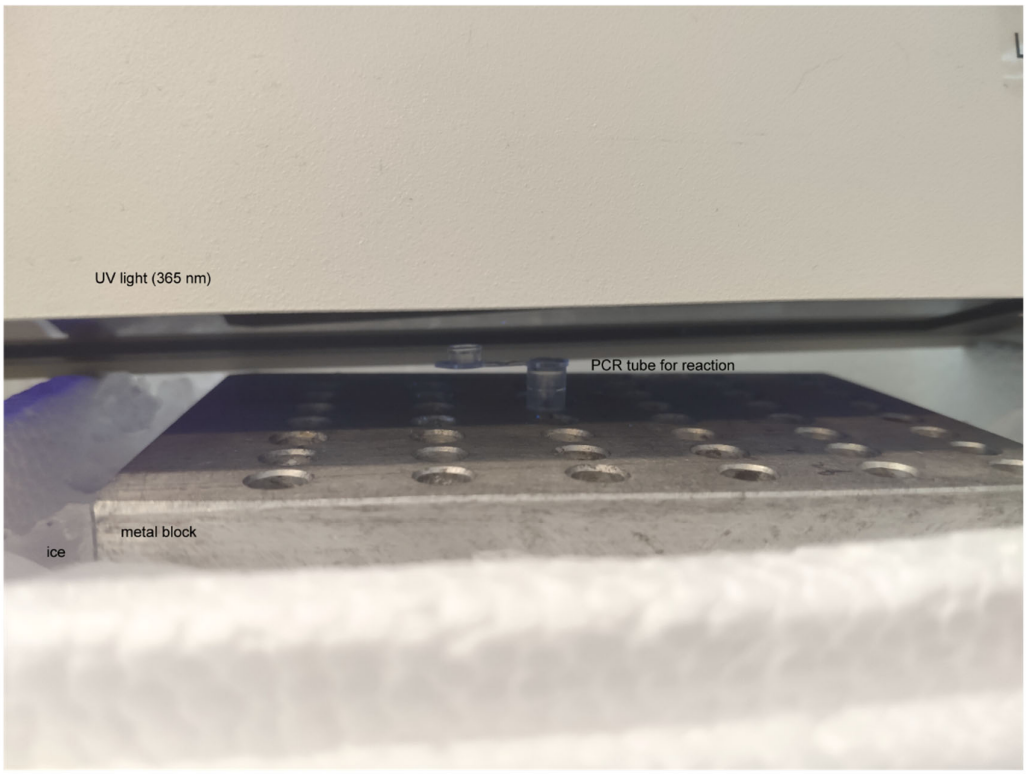
Appendix Table 2. Data collection and refinement statistics of SnVTD structure.

	SnVTD~K6-diUb-3Br (PDB code: 8RL3)
Data collection	
Beamline	PETRA III-P14
Wavelength (Å)	0.97625
Space group	$P\,2_1\,2_1\,2_1$
Cell dimensions	
a, *b*, *c* (Å)	54.25, 94.01, 95.06
α, *β*, *γ* (°)	90, 90, 90
Anisotropy correction	yes
Observed reflections	164,612 (7,655)
Unique reflections	30,022 (1,486)
Resolution (Å)	47.13 – 1.88 (2.02 – 1.88)
Ellipsoidal resolution limits (Å) [direction]	1.75 [a*] 2.10 [b*] 2.20 [c*]
R_{merge}	0.107 (0.615)
R_{meas}	0.119 (0.683)
$I/\sigma(I)$	8.9 (2.4)
$CC_{1/2}$	0.996 (0.794)
Spherical completeness (%)	74.2 (18.5)
Ellipsoidal completeness (%)	91.9 (61.0)
Redundancy	5.5 (5.2)
Wilson *B* (Å^2) [direction]	17.8 [a*] 26.8 [b*] 37.8 [c*]
Refinement	
Resolution (Å)	1.88 Å
No. reflections	30,017
R_{work} / R_{free} (%)	16.5 / 20.7
No. atoms	3,402
Protein	3,029
Ligand	9
Water	364
B factors (Å^2)	28.3
Protein (Å^2)	27.5
Ligand (Å^2)	18.7
Water (Å^2)	35.3
R.m.s.d.	
Bond lengths (Å)	0.010
Bond angles (°)	1.11
Ramachandran (favored / allowed / outlier) (%)	99.0 / 1.0 / 0.0
Clashscore	0.7
Rotamer outliers (%)	0.0
Copies / a.s.u.	1

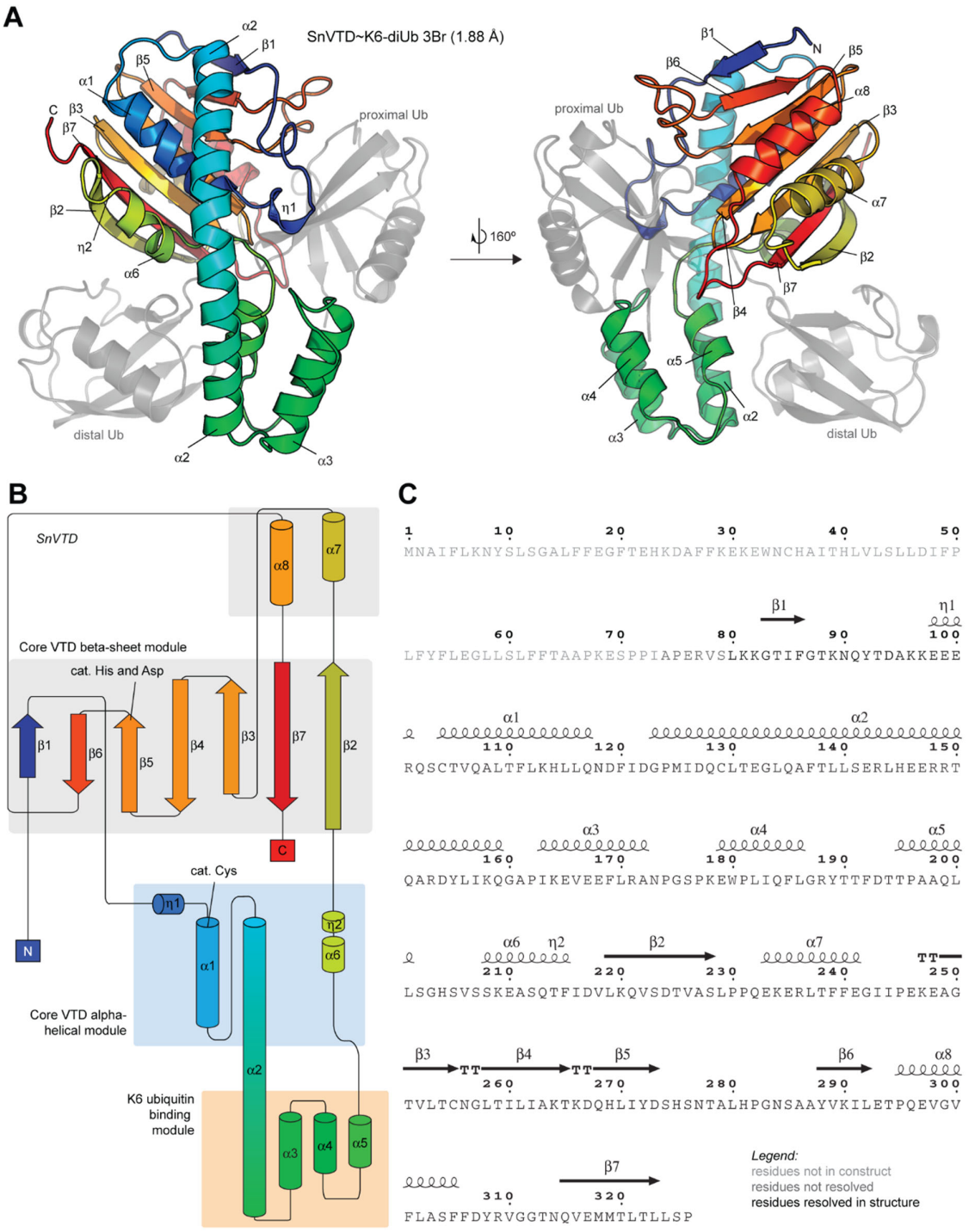
The dataset was collected from a single crystal. Values in parentheses are for highest-resolution shell. a.s.u., asymmetric unit. R.m.s.d., root mean square deviations.



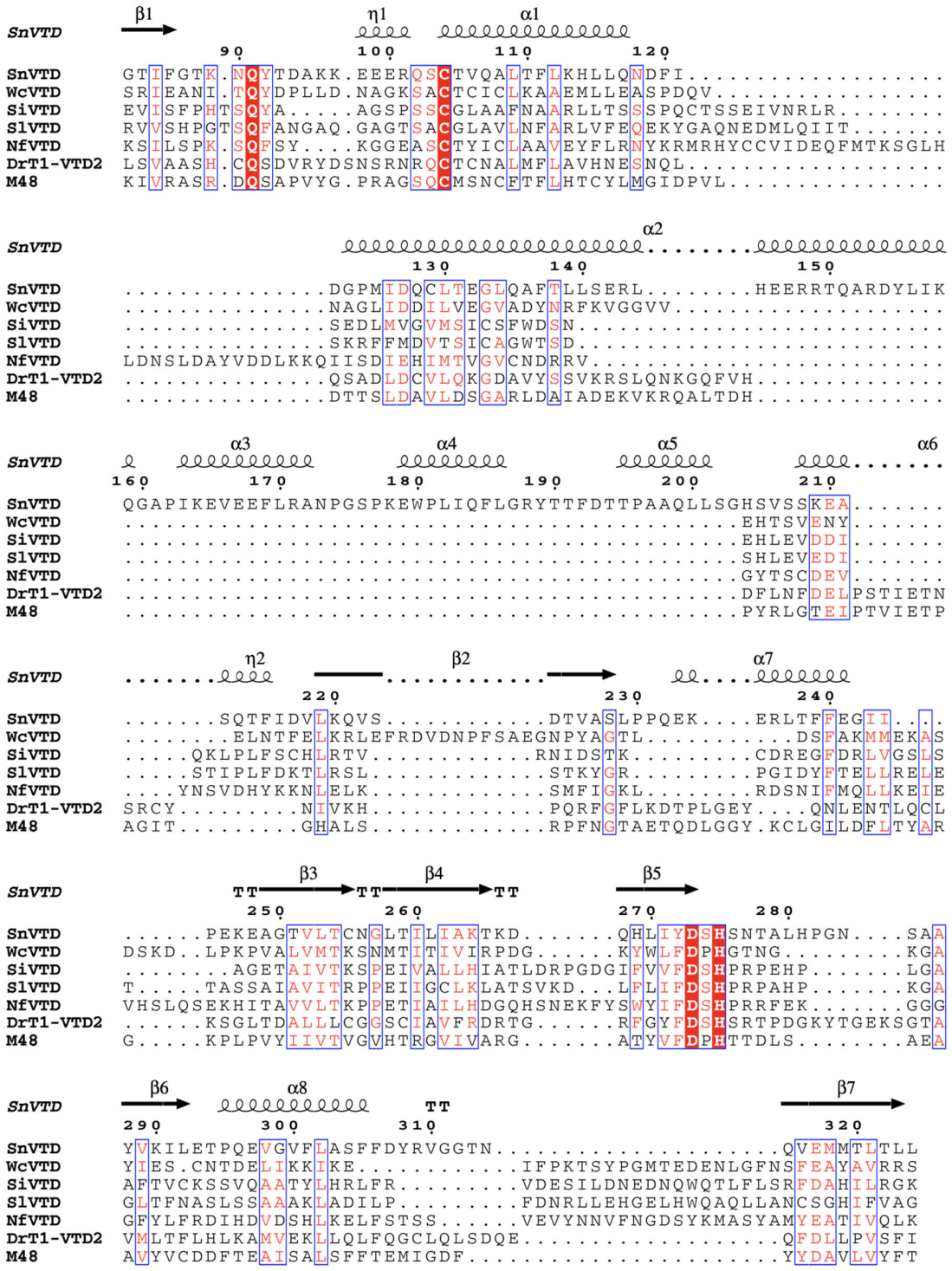
Appendix Fig 1. Fluorescence polarization kinetic assays for **A**. USPL1, **B**. SENP1 and **C**. ULP1.



Appendix Fig 2. Reaction setup for thiol-ene reaction.



Appendix Fig 3. Structural annotation of SnVTD. **A**. Annotation of secondary structures in SnVTD. Ubiquitin molecules are shown in grey. **B**. Illustration of secondary structures of SnVTD. **C**. Sequence of SnVTD catalytic domain.



Appendix Fig 4. Sequence alignment of SnVTD and other VTDs. Catalytic center residues are highlighted in red. Numbering is according to the sequence of SnVTD. Secondary structures are annotated according to the SnVTD structure. This alignment was performed by Prof. Kay Hofmann at the University of Cologne.

7. REFERENCES

[1] D. Komander, M. Rape, *Annual review of biochemistry* **2012**, *81*, 203.
[2] K. N. Swatek, D. Komander, *Cell Research* **2016**, *26*, 399.
[3] G. Goldstein, M. Scheid, U. Hammerling, D. H. Schlesinger, H. D. Niall, E. A. Boyse, *Proceedings of the National Academy of Sciences* **1975**, *72*, 11.
[4] A. Ciehanover, Y. Hod, A. Hershko, *Biochemical and Biophysical Research Communications* **1978**, *81*, 1100.
[5] A. Hershko, A. Ciechanover, H. Heller, A. L. Haas, I. A. Rose, *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1980**, *77*, 1783.
[6] K. D. Wilkinson, M. K. Urban, A. L. Haas, *Journal of Biological Chemistry* **1980**, *255*, 7529.
[7] A. Rousseau, A. Bertolotti, *Nature Reviews Molecular Cell Biology* **2018**, *19*, 697.
[8] L. Buetow, D. T. Huang, *Nature Reviews Molecular Cell Biology* **2016**, *17*, 626.
[9] J. A. Harrigan, X. Jacq, N. M. Martin, S. P. Jackson, *Nat Rev Drug Discov* **2018**, *17*, 57.
[10] S. Vijay-Kumar, C. E. Bugg, W. J. Cook, *Journal of molecular biology* **1987**, *194*, 531.
[11] F. Ohtake, Y. Saeki, K. Sakamoto, K. Ohtake, H. Nishikawa, H. Tsuchiya, T. Ohta, K. Tanaka, J. Kanno, *EMBO reports* **2015**, *16*, 192.
[12] T. Wauer, K. N. Swatek, J. L. Wagstaff, C. Gladkova, J. N. Pruneda, M. A. Michel, M. Gersch, C. M. Johnson, S. M. V. Freund, D. Komander, *Embo Journal* **2015**, *34*, 307.
[13] V. Chau, J. W. Tobias, A. Bachmair, D. Marriott, D. J. Ecker, D. K. Gonda, A. Varshavsky, *Science* **1989**, *243*, 1576.
[14] E. Oh, D. Akopian, M. Rape, *Annual Review of Cell and Developmental Biology, Vol 34* **2018**, *34*, 137.
[15] K. K. Deol, S. J. Eyles, E. R. Strider, *Journal of the American Society for Mass Spectrometry* **2020**, *31*, 1132.
[16] P. Xu, J. M. Peng, *Analytical Chemistry* **2008**, *80*, 3438.
[17] K. N. Swatek, J. L. Usher, A. F. Kueck, C. Gladkova, T. E. T. Mevissen, J. N. Pruneda, T. Skern, D. Komander, *Nature* **2019**, *572*, 533.
[18] M. K. Hospenthal, S. M. V. Freund, D. Komander, *Nature structural & molecular biology* **2013**, *20*, 555.
[19] K. E. Sloper-Mould, J. C. Jemc, C. M. Pickart, L. Hicke, *Journal of Biological Chemistry* **2001**, *276*, 30483.
[20] R. Beal, Q. Deveraux, G. Xia, M. Rechsteiner, C. Pickart, *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 861.
[21] C. T. Archer, T. Kodadek, *Nucleic Acids Research* **2010**, *38*, 789.
[22] L. Jin, A. Williamson, S. Banerjee, I. Philipp, M. Rape, *Cell* **2008**, *133*, 653.
[23] K. E. Wickliffe, S. Lorenz, D. E. Wemmer, J. Kuriyan, M. Rape, *Cell* **2011**, *144*, 769.
[24] A. M. Taherbhoy, B. A. Schulman, S. E. Kaiser, *Essays in Biochemistry* **2012**, *52*, 51.
[25] M. Hochstrasser, *Nature* **2009**, *458*, 422.
[26] L. Cappadocia, C. D. Lima, *Chemical Reviews* **2018**, *118*, 889.
[27] O. Kerscher, R. Felberbaum, M. Hochstrasser, *Annual Review of Cell and Developmental Biology* **2006**, *22*, 159.
[28] Y. Ichimura, Y. Imamura, K. Emoto, M. Umeda, T. Noda, Y. Ohsumi, *J Biol Chem* **2004**, *279*, 40584.
[29] H. Nakatogawa, Y. Ichimura, Y. Ohsumi, *Cell* **2007**, *130*, 165.
[30] P. J. Farrell, R. J. Broeze, P. Lengyel, *Nature* **1979**, *279*, 523.
[31] D. C. Blomstrom, D. Fahey, R. Kutny, B. D. Korant, E. Knight, *Journal of Biological Chemistry* **1986**, *261*, 8811.
[32] A. L. Haas, P. Ahrens, P. M. Bright, H. Ankel, *Journal of Biological Chemistry* **1987**, *262*, 11315.
[33] Y.-G. Chang, X.-Z. Yan, Y.-Y. Xie, X.-C. Gao, A.-X. Song, D.-E. Zhang, H.-Y. Hu, *Journal of Biological Chemistry* **2008**, *283*, 13370.

[34] S. D. Desai, A. L. Haas, L. M. Wood, Y.-C. Tsai, S. Pestka, E. H. Rubin, A. Saleem, A. Nur-E-Kamal, L. F. Liu, *Cancer Research* **2006**, *66*, 921.
[35] X. Zhang, D. Bogunovic, B. Payelle-Brogard, V. Francois-Newton, S. D. Speer, C. Yuan, S. Volpi, Z. Li, O. Sanal, D. Mansouri, I. Tezcan, G. I. Rice, C. Chen, N. Mansouri, S. A. Mahdaviani, Y. Itan, B. Boisson, S. Okada, L. Zeng, X. Wang, H. Jiang, W. Liu, T. Han, D. Liu, T. Ma, B. Wang, M. Liu, J.-Y. Liu, Q. K. Wang, D. Yalnizoglu, L. Radoshevich, G. Uzé, P. Gros, F. Rozenberg, S.-Y. Zhang, E. Jouanguy, J. Bustamante, A. García-Sastre, L. Abel, P. Lebon, L. D. Notarangelo, Y. J. Crow, S. Boisson-Dupuis, J.-L. Casanova, S. Pellegrini, *Nature* **2015**, *517*, 89.
[36] Y.-C. Perng, D. J. Lenschow, *Nature Reviews Microbiology* **2018**, *16*, 423.
[37] A. C. O. Vertegaal, *Nat Rev Mol Cell Biol* **2022**, *23*, 715.
[38] R. Mahajan, C. Delphin, T. Guan, L. Gerace, F. Melchior, *Cell* **1997**, *88*, 97.
[39] M. J. Matunis, E. Coutavas, G. Blobel, *Journal of Cell Biology* **1996**, *135*, 1457.
[40] R. Mahajan, L. Gerace, F. Melchior, *Journal of Cell Biology* **1998**, *140*, 259.
[41] A. Pichler, F. Melchior, *Traffic* **2002**, *3*, 381.
[42] D. Wen, J. Wu, L. Wang, Z. Fu, *Cell Reports* **2017**, *21*, 2147.
[43] A. C. Bellail, J. J. Olson, C. Hao, *Nature Communications* **2014**, *5*, 4234.
[44] A. Chen, H. Mannen, S. S.-L. Li, *IUBMB life* **1998**, *46*, 1161.
[45] J. T. Crowl, D. B. Stetson, *Proceedings of the National Academy of Sciences* **2018**, *115*, 6798.
[46] A. C. O. Vertegaal, *Nature Reviews Molecular Cell Biology* **2022**, *23*, 715.
[47] L. Wang, C. Wansleeben, S. Zhao, P. Miao, W. Paschen, W. Yang, *EMBO reports* **2014**, *15*, 878.
[48] K. M. Bohren, V. Nadkarni, J. H. Song, K. H. Gabbay, D. Owerbach, *Journal of Biological Chemistry* **2004**, *279*, 27233.
[49] D. Guo, M. Li, Y. Zhang, P. Yang, S. Eckenrode, D. Hopkins, W. Zheng, S. Purohit, R. H. Podolsky, A. Muir, J. Wang, Z. Dong, T. Brusko, M. Atkinson, P. Pozzilli, A. Zeidler, L. J. Raffel, C. O. Jacob, Y. Park, M. Serrano-Rios, M. T. M. Larrad, Z. Zhang, H.-J. Garchon, J.-F. Bach, J. I. Rotter, J.-X. She, C.-Y. Wang, *Nature Genetics* **2004**, *36*, 837.
[50] D. Guo, J. Han, B.-L. Adam, N. H. Colburn, M.-H. Wang, Z. Dong, D. L. Eizirik, J.-X. She, C.-Y. Wang, *Biochemical and Biophysical Research Communications* **2005**, *337*, 1308.
[51] Y.-C. Liang, C.-C. Lee, Y.-L. Yao, C.-C. Lai, M. L. Schmitz, W.-M. Yang, *Scientific Reports* **2016**, *6*, 26509.
[52] S. Kumar, Y. Yoshida, M. Noda, *Biochemical and Biophysical Research Communications* **1993**, *195*, 393.
[53] L. Zhou, W. Zhang, Y. Sun, L. Jia, *Cellular Signalling* **2018**, *44*, 92.
[54] M. D. Petroski, R. J. Deshaies, *Nature Reviews Molecular Cell Biology* **2005**, *6*, 9.
[55] D. M. Duda, L. A. Borg, D. C. Scott, H. W. Hunt, M. Hammel, B. A. Schulman, *Cell* **2008**, *134*, 995.
[56] K. Baek, D. T. Krist, J. R. Prabu, S. Hill, M. Klügel, L.-M. Neumaier, S. von Gronau, G. Kleiger, B. A. Schulman, *Nature* **2020**, *578*, 461.
[57] L. T. Henneberg, J. Singh, D. M. Duda, K. Baek, D. Yanishevski, P. J. Murray, M. Mann, S. S. Sidhu, B. A. Schulman, *Nature Chemical Biology* **2023**, *19*, 1513.
[58] R. I. Enchev, B. A. Schulman, M. Peter, *Nature Reviews Molecular Cell Biology* **2015**, *16*, 30.
[59] N. Yamamoto, T. Shimizu, K. Yonemori, S. Kitano, S. Kondo, S. Iwasa, T. Koyama, K. Sudo, J. Sato, K. Tamura, J. Tomomatsu, M. Ono, N. Fukuda, S. Takahashi, *Investigational New Drugs* **2021**, *39*, 1036.
[60] K. Kas, L. Michiels, J. Merregaert, *Biochemical and Biophysical Research Communications* **1992**, *187*, 927.
[61] J. van den Heuvel, C. Ashiono, L. C. Gillet, K. Dörner, E. Wyler, I. Zemp, U. Kutay, *eLife* **2021**, *10*, e70560.
[62] M. Nakamura, T. Tsunematsu, Y. Tanigawa, *Immunology* **1998**, *94*, 142.
[63] M. Nakamura, Y. Tanigawa, *European Journal of Biochemistry* **2003**, *270*, 4052.
[64] M. Nakamura, S. Shimosaki, *The FEBS Journal* **2009**, *276*, 6355.

[65] M. Nakamura, N. Watanabe, K. Notsu, *Biochemical and Biophysical Research Communications* **2015**, *464*, 1096.
[66] R. O'Dea, N. Kazi, A. Hoffmann-Benito, Z. Zhao, S. Recknagel, K. Wendrich, P. Janning, M. Gersch, *Nature Chemical Biology* **2023**, *19*, 1394.
[67] A. Aichem, M. Groettrup, *Journal of Cell Science* **2020**, *133*.
[68] H. Yamamoto, S. Zhang, N. Mizushima, *Nature Reviews Genetics* **2023**, *24*, 382.
[69] R. Ishimura, A. H. El-Gowily, D. Noshiro, S. Komatsu-Hirota, Y. Ono, M. Shindo, T. Hatta, M. Abe, T. Uemura, H.-C. Lee-Okada, T. M. Mohamed, T. Yokomizo, T. Ueno, K. Sakimura, T. Natsume, H. Sorimachi, T. Inada, S. Waguri, N. N. Noda, M. Komatsu, *Nature Communications* **2022**, *13*, 7857.
[70] C. D. Schlieker, A. G. Van der Veen, J. R. Damon, E. Spooner, H. L. Ploegh, *Proceedings of the National Academy of Sciences* **2008**, *105*, 18255.
[71] A. G. Van der Veen, K. Schorpp, C. Schlieker, L. Buti, J. R. Damon, E. Spooner, H. L. Ploegh, S. Jentsch, *Proceedings of the National Academy of Sciences* **2011**, *108*, 1763.
[72] C. Li, T. G. Nelson, A. C. O. Vertegaal, P. Thibault, *Nature Reviews Methods Primers* **2021**, *1*, 53.
[73] Z. S. Hann, C. Ji, S. K. Olsen, X. Q. Lu, M. C. Lux, D. S. Tan, C. D. Lima, *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 15475.
[74] J. M. Huibregtse, M. Scheffner, S. Beaudenon, P. M. Howley, *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 5249.
[75] E. G. Otten, E. Werner, A. Crespillo-Casado, K. B. Boyle, V. Dharamdasani, C. Pathe, B. Santhanam, F. Randow, *Nature* **2021**, *594*, 111.
[76] J.-i. Sakamaki, K. L. Ode, Y. Kurikawa, H. R. Ueda, H. Yamamoto, N. Mizushima, *Molecular cell* **2022**, *82*, 3677.
[77] J.-i. Sakamaki, N. Mizushima, *Trends in Cell Biology* **2023**, *33*, 991.
[78] B. A. Schulman, J. W. Harper, *Nature Reviews Molecular Cell Biology* **2009**, *10*, 319.
[79] J. Jin, X. Li, S. P. Gygi, J. W. Harper, *Nature* **2007**, *447*, 1135.
[80] X. Liu, B. Zhao, L. Sun, K. Bhuripanyo, Y. Wang, Y. Bi, R. V. Davuluri, D. M. Duong, D. Nanavati, J. Yin, H. Kiyokawa, *Nature Communications* **2017**, *8*, 14286.
[81] X. Lu, S. K. Olsen, A. D. Capili, J. S. Cisar, C. D. Lima, D. S. Tan, *Journal of the American Chemical Society* **2010**, *132*, 1748.
[82] S. K. Olsen, A. D. Capili, X. Lu, D. S. Tan, C. D. Lima, *Nature* **2010**, *463*, 906.
[83] B. A. Schulman, J. Wade Harper, *Nature Reviews Molecular Cell Biology* **2009**, *10*, 319.
[84] A. Aichem, C. Pelzer, S. Lukasiak, B. Kalveram, P. W. Sheppard, N. Rani, G. Schmidtke, M. Groettrup, *Nature Communications* **2010**, *1*, 13.
[85] A. Werner, A. Flotho, F. Melchior, *Molecular cell* **2012**, *46*, 287.
[86] D. T. Huang, O. Ayrault, H. W. Hunt, A. M. Taherbhoy, D. M. Duda, D. C. Scott, L. A. Borg, G. Neale, P. J. Murray, M. F. Roussel, B. A. Schulman, *Molecular cell* **2009**, *33*, 483.
[87] D. T. Huang, D. W. Miller, R. Mathew, R. Cassell, J. M. Holton, M. F. Roussel, B. A. Schulman, *Nature structural & molecular biology* **2004**, *11*, 927.
[88] A. Saha, R. J. Deshaies, *Molecular cell* **2008**, *32*, 21.
[89] W. Yuan, R. M. Krug, *The EMBO Journal* **2001**, *20*, 362.
[90] C. Zhao, S. L. Beaudenon, M. L. Kelley, M. B. Waddell, W. Yuan, B. A. Schulman, J. M. Huibregtse, R. M. Krug, *Proceedings of the National Academy of Sciences* **2004**, *101*, 7578.
[91] K. E. Ravichandran, L. Kaduhr, B. Skupien-Rabian, E. Shvetsova, M. Sokołowski, R. c. Krutyhołowa, D. Kwasna, C. Brachmann, S. Lin, S. Guzman Perez, P. Wilk, M. Kösters, P. Grudnik, U. Jankowska, S. A. Leidel, R. Schaffrath, S. Glatt, *The EMBO Journal* **2022**, *41*, e111318.
[92] J. J. Peter, H. M. Magnussen, P. A. DaRosa, D. Millrine, S. P. Matthews, F. Lamoliatte, R. Sundaramoorthy, R. R. Kopito, Y. Kulathu, *The EMBO Journal* **2022**, *41*, e111015.
[93] M. Kumar, P. Padala, J. Fahoum, F. Hassouna, T. Tsaban, G. Zoltsman, S. Banerjee, E. Cohen-Kfir, M. Dessau, R. Rosenzweig, M. N. Isupov, O. Schueler-Furman, R. Wiener, *Nature Communications* **2021**, *12*, 5708.
[94] W. Li, M. H. Bengtson, A. Ulbrich, A. Matsuda, V. A. Reddy, A. Orth, S. K. Chanda, S. Batalov, C. A. P. Joazeiro, *PloS one* **2008**, *3*, e1487.

[95] N. Zheng, N. Shabek, *Annual review of biochemistry* **2017**, *86*, 129.
[96] A. D. Capili, E. L. Edghill, K. Wu, K. L. B. Borden, *Journal of molecular biology* **2004**, *340*, 1117.
[97] D. M. Wenzel, A. Lissounov, P. S. Brzovic, R. E. Klevit, *Nature* **2011**, *474*, 105.
[98] K.-C. Pao, N. T. Wood, A. Knebel, K. Rafie, M. Stanley, P. D. Mabbitt, R. Sundaramoorthy, K. Hofmann, D. M. F. van Aalten, S. Virdee, *Nature* **2018**, *556*, 381.
[99] C. E. Berndsen, C. Wolberger, *Nature structural & molecular biology* **2014**, *21*, 301.
[100] Martin A. Michel, Paul R. Elliott, Kirby N. Swatek, M. Simicek, Jonathan N. Pruneda, Jane L. Wagstaff, Stefan M. V. Freund, D. Komander, *Molecular cell* **2015**, *58*, 95.
[101] Yosua A. Kristariyanto, Syed A. Abdul Rehman, David G. Campbell, Nicholas A. Morrice, C. Johnson, R. Toth, Y. Kulathu, *Molecular cell* **2015**, *58*, 83.
[102] M. K. Hospenthal, S. M. V. Freund, D. Komander, *Nature structural & molecular biology* **2013**, *20*, 555.
[103] J. A. Ronau, J. F. Beckmann, M. Hochstrasser, *Cell Research* **2016**, *26*, 441.
[104] M. J. Clague, I. Barsukov, J. M. Coulson, H. Liu, D. J. Rigden, S. Urbé, *Physiological Reviews* **2013**, *93*, 1289.
[105] T. E. T. Mevissen, D. Komander, *Annual review of biochemistry* **2017**, *86*, 159.
[106] M. S. Ritorto, R. Ewan, A. B. Perez-Oliva, A. Knebel, S. J. Buhrlage, M. Wightman, S. M. Kelly, N. T. Wood, S. Virdee, N. S. Gray, N. A. Morrice, D. R. Alessi, M. Trost, *Nature Communications* **2014**, *5*, 4763.
[107] V. De Cesare, J. Moran, R. Traynor, A. Knebel, M. S. Ritorto, M. Trost, H. McLauchlan, C. J. Hastie, P. Davies, *Nature Protocols* **2020**, *15*, 4034.
[108] P. Gong, G. A. Davidson, W. Gui, K. Yang, W. P. Bozza, Z. Zhuang, *Chem Sci* **2018**, *9*, 7859.
[109] X. Sui, Y. Wang, Y.-X. Du, L.-J. Liang, Q. Zheng, Y.-M. Li, L. Liu, *Chemical Science* **2020**, *11*, 12633.
[110] I. Erven, E. Abraham, T. Hermanns, U. Baumann, K. Hofmann, *Nature Communications* **2022**, *13*, 7643.
[111] V. Boll, T. Hermanns, M. Uthoff, I. Erven, E.-M. Hörner, V. Kozjak-Pavlovic, U. Baumann, K. Hofmann, *Nature Communications* **2023**, *14*, 7335.
[112] T. E. Mevissen, M. K. Hospenthal, P. P. Geurink, P. R. Elliott, M. Akutsu, N. Arnaudo, R. Ekkebus, Y. Kulathu, T. Wauer, F. El Oualid, S. M. Freund, H. Ovaa, D. Komander, *Cell* **2013**, *154*, 169.
[113] X. I. Ambroggio, D. C. Rees, R. J. Deshaies, *PLOS Biology* **2003**, *2*, e2.
[114] S. A. Abdul Rehman, L. A. Armstrong, S. M. Lange, Y. A. Kristariyanto, T. W. Gräwert, A. Knebel, D. I. Svergun, Y. Kulathu, *Molecular cell* **2021**, *81*, 4176.
[115] M. J. Clague, S. Urbé, D. Komander, *Nature Reviews Molecular Cell Biology* **2019**, *20*, 338.
[116] M. Hu, P. Li, M. Li, W. Li, T. Yao, J.-W. Wu, W. Gu, R. E. Cohen, Y. Shi, *Cell* **2002**, *111*, 1041.
[117] N. J. Schauer, R. S. Magin, X. Liu, L. M. Doherty, S. J. Buhrlage, *Journal of medicinal chemistry* **2020**, *63*, 2731.
[118] D. Komander, C. J. Lord, H. Scheel, S. Swift, K. Hofmann, A. Ashworth, D. Barford, *Molecular cell* **2008**, *29*, 451.
[119] M. Gersch, C. Gladkova, A. F. Schubert, M. A. Michel, S. Maslen, D. Komander, *Nature structural & molecular biology* **2017**, *24*, 920.
[120] Y. Sato, K. Okatsu, Y. Saeki, K. Yamano, N. Matsuda, A. Kaiho, A. Yamagata, S. Goto-Ito, M. Ishikawa, Y. Hashimoto, K. Tanaka, S. Fukai, *Nature structural & molecular biology* **2017**, *24*, 911.
[121] https://missiontherapeutics.com/media-centre/press-releases/.
[122] A. Simoneau, J. L. Engel, M. Bandi, K. Lazarides, S. Liu, S. R. Meier, A. H. Choi, H. Zhang, B. Shen, L. Martires, D. Gotur, T. V. Pham, F. Li, L. Gu, S. Gong, M. Zhang, E. Wilker, X. Pan, D. A. Whittington, S. Throner, J. P. Maxwell, Y. Chen, Y. Yu, A. Huang, J. N. Andersen, T. Feng, *Molecular Cancer Therapeutics* **2023**, *22*, 215.

[123] Tycho E. T. Mevissen, Manuela K. Hospenthal, Paul P. Geurink, Paul R. Elliott, M. Akutsu, N. Arnaudo, R. Ekkebus, Y. Kulathu, T. Wauer, F. El Oualid, Stefan M. V. Freund, H. Ovaa, D. Komander, *Cell* **2013**, *154*, 169.
[124] R. B. Damgaard, J. A. Walker, P. Marco-Casanova, N. V. Morgan, H. L. Titheradge, P. R. Elliott, D. McHale, E. R. Maher, A. N. J. McKenzie, D. Komander, *Cell* **2016**, *166*, 1215.
[125] M. K. Hospenthal, T. E. T. Mevissen, D. Komander, *Nature Protocols* **2015**, *10*, 349.
[126] O. W. Huang, X. Ma, J. Yin, J. Flinders, T. Maurer, N. Kayagaki, Q. Phung, I. Bosanac, D. Arnott, V. M. Dixit, S. G. Hymowitz, M. A. Starovasnik, A. G. Cochran, *Nature structural & molecular biology* **2012**, *19*, 171.
[127] Y. Zhao, M. C. Mudge, J. M. Soll, R. B. Rodrigues, A. K. Byrum, E. A. Schwarzkopf, T. R. Bradstreet, S. P. Gygi, B. T. Edelson, N. Mosammaparast, *Molecular cell* **2018**, *69*, 505.
[128] K. D. Wilkinson, K. Lee, S. Deshpande, P. Duerksen-Hughes, J. M. Boss, J. Pohl, *Science* **1989**, *246*, 670.
[129] Z.-R. Zhou, Y.-H. Zhang, S. Liu, A.-X. Song, H.-Y. Hu, *Biochemical Journal* **2011**, *441*, 143.
[130] K. K. Deol, S. O. Crowe, J. Du, H. A. Bisbee, R. G. Guenette, E. R. Strieter, *Molecular cell* **2020**, *80*, 796.
[131] M. Karpiyevich, S. Adjalley, M. Mol, D. B. Ascher, B. Mason, G. J. van der Heden van Noort, H. Laman, H. Ovaa, M. C. S. Lee, K. Artavanis-Tsakonas, *PLOS Pathogens* **2019**, *15*, e1008086.
[132] H. Wada, K. Kito, L. S. Caskey, E. T. H. Yeh, T. Kamitani, *Biochemical and Biophysical Research Communications* **1998**, *251*, 688.
[133] V. De Cesare, D. Carbajo Lopez, P. D. Mabbitt, A. J. Fletcher, M. Soetens, O. Antico, N. T. Wood, S. Virdee, *Proceedings of the National Academy of Sciences* **2021**, *118*, e2006947118.
[134] Syed A. Abdul Rehman, Yosua A. Kristariyanto, S.-Y. Choi, P. J. Nkosi, S. Weidlich, K. Labib, K. Hofmann, Y. Kulathu, *Molecular cell* **2016**, *63*, 146.
[135] Y. A. Kristariyanto, S. A. Abdul Rehman, S. Weidlich, A. Knebel, Y. Kulathu, *EMBO reports* **2017**, *18*, 392.
[136] T. Hermanns, C. Pichlo, I. Woiwode, K. Klopffleisch, K. F. Witting, H. Ovaa, U. Baumann, K. Hofmann, *Nature Communications* **2018**, *9*, 799.
[137] D. Kwasna, S. A. Abdul Rehman, J. Natarajan, S. Matthews, R. Madden, V. De Cesare, S. Weidlich, S. Virdee, I. Ahel, I. Gibbs-Seymour, Y. Kulathu, *Molecular cell* **2018**, *70*, 150.
[138] P. Haahr, N. Borgermann, X. Guo, D. Typas, D. Achuthankutty, S. Hoffmann, R. Shearer, T. K. Sixma, N. Mailand, *Molecular cell* **2018**, *70*, 165.
[139] D. S. Hewings, J. Heideker, T. P. Ma, A. P. AhYoung, F. El Oualid, A. Amore, G. T. Costakes, D. Kirchhofer, B. Brasher, T. Pillow, N. Popovych, T. Maurer, C. Schwerdtfeger, W. F. Forrest, K. Yu, J. Flygare, M. Bogyo, I. E. Wertz, *Nature Communications* **2018**, *9*, 1162.
[140] H. M. Mendoza, L.-n. Shen, C. Botting, A. Lewis, J. Chen, B. Ink, R. T. Hay, *Journal of Biological Chemistry* **2003**, *278*, 25637.
[141] T. Gan-Erdene, K. Nagamalleswari, L. Yin, K. Wu, Z.-Q. Pan, K. D. Wilkinson, *Journal of Biological Chemistry* **2003**, *278*, 28892.
[142] A. Nayak, S. Müller, *Genome Biology* **2014**, *15*, 422.
[143] J. Mikolajczyk, M. Drag, M. Békés, J. T. Cao, Z. e. Ronai, G. S. Salvesen, *Journal of Biological Chemistry* **2007**, *282*, 26217.
[144] E. J. Shin, H. M. Shin, E. Nam, W. S. Kim, J. H. Kim, B. H. Oh, Y. Yun, *EMBO reports* **2012**, *13*, 339.
[145] X. Xie, X. Wang, D. Jiang, J. Wang, R. Fei, X. Cong, L. Wei, Y. Wang, H. Chen, *Biochemical and Biophysical Research Communications* **2017**, *488*, 291.
[146] S. Schulz, G. Chachami, L. Kozaczkiewicz, U. Winter, N. Stankovic-Valentin, P. Haas, K. Hofmann, H. Urlaub, H. Ovaa, J. Wittbrodt, E. Meulmeester, F. Melchior, *EMBO Rep* **2012**, *13*, 930.
[147] G. A. Cope, G. S. B. Suh, L. Aravind, S. E. Schwarz, S. L. Zipursky, E. V. Koonin, R. J. Deshaies, *Science* **2002**, *298*, 608.
[148] Mariola J. Edelmann, A. Iphöfer, M. Akutsu, M. Altun, K. di Gleria, Holger B. Kramer, E. Fiebiger, S. Dhe-Paganon, Benedikt M. Kessler, *Biochemical Journal* **2009**, *418*, 379.

[149] K. Stuber, T. Schneider, J. Werner, M. Kovermann, A. Marx, M. Scheffner, *Nature Communications* **2021**, *12*, 5939.
[150] D. Shin, R. Mukherjee, D. Grewe, D. Bojkova, K. Baek, A. Bhattacharya, L. Schulz, M. Widera, A. R. Mehdipour, G. Tascher, P. P. Geurink, A. Wilhelm, G. J. van der Heden van Noort, H. Ovaa, S. Müller, K.-P. Knobeloch, K. Rajalingam, B. A. Schulman, J. Cinatl, G. Hummer, S. Ciesek, I. Dikic, *Nature* **2020**, *587*, 657.
[151] T. Klemm, G. Ebert, D. J. Calleja, C. C. Allison, L. W. Richardson, J. P. Bernardini, B. G. Lu, N. W. Kuchel, C. Grohmann, Y. Shibata, Z. Y. Gan, J. P. Cooney, M. Doerflinger, A. E. Au, T. R. Blackmore, G. J. van der Heden van Noort, P. P. Geurink, H. Ovaa, J. Newman, A. Riboldi-Tunnicliffe, P. E. Czabotar, J. P. Mitchell, R. Feltham, B. C. Lechtenberg, K. N. Lowes, G. Dewson, M. Pellegrini, G. Lessene, D. Komander, *The EMBO Journal* **2020**, *39*, e106275.
[152] A. Catic, E. Fiebiger, G. A. Korbel, D. Blom, P. J. Galardy, H. L. Ploegh, *PloS one* **2007**, *2*, e679.
[153] C. Li, T. Wang, L. Liang, G. Chu, J. Zhang, W. He, L. Liu, J. Li, *Science China Chemistry* **2023**, *66*, 837.
[154] X. Sui, Y. Wang, Y. X. Du, L. J. Liang, Q. Zheng, Y. M. Li, L. Liu, *Chem Sci* **2020**, *11*, 12633.
[155] D. S. Hewings, J. A. Flygare, M. Bogyo, I. E. Wertz, *The FEBS Journal* **2017**, *284*, 1555.
[156] Y. Jia, L. A. Claessens, A. C. O. Vertegaal, H. Ovaa, *ACS Chemical Biology* **2019**, *14*, 2389.
[157] C. M. Pickart, I. A. Rose, *Journal of Biological Chemistry* **1986**, *261*, 10210.
[158] A. Hershko, I. A. Rose, *Proceedings of the National Academy of Sciences* **1987**, *84*, 1829.
[159] Y. A. Lam, W. Xu, G. N. DeMartino, R. E. Cohen, *Nature* **1997**, *385*, 737.
[160] S. C. Johnston, S. M. Riddle, R. E. Cohen, C. P. Hill, *The EMBO Journal* **1999**, *18*, 3877.
[161] M. Gorka, H. M. Magnussen, Y. Kulathu, *Seminars in Cell & Developmental Biology* **2022**, *132*, 86.
[162] N. D. Weikart, S. Sommer, H. D. Mootz, *Chemical Communications* **2012**, *48*, 296.
[163] S. Sommer, N. D. Weikart, A. Brockmeyer, P. Janning, H. D. Mootz, *Angewandte Chemie-International Edition* **2011**, *50*, 9888.
[164] R. Ekkebus, S. I. van Kasteren, Y. Kulathu, A. Scholten, I. Berlin, P. P. Geurink, A. de Jong, S. Goerdayal, J. Neefjes, A. J. R. Heck, D. Komander, H. Ovaa, *Journal of the American Chemical Society* **2013**, *135*, 2867.
[165] A. Basters, P. P. Geurink, A. Röcker, K. F. Witting, R. Tadayon, S. Hess, M. S. Semrau, P. Storici, H. Ovaa, K.-P. Knobeloch, G. Fritz, *Nature structural & molecular biology* **2017**, *24*, 270.
[166] Z. Zhao, R. O'Dea, K. Wendrich, N. Kazi, M. Gersch, *Journal of the American Chemical Society* **2023**, *145*, 20801.
[167] S. Sommer, N. D. Weikart, U. Linne, H. D. Mootz, *Bioorganic & Medicinal Chemistry* **2013**, *21*, 2511.
[168] A. Borodovsky, H. Ovaa, N. Kolli, T. Gan-Erdene, K. D. Wilkinson, H. L. Ploegh, B. M. Kessler, *Chemistry & Biology* **2002**, *9*, 1149.
[169] K. R. Love, R. K. Pandya, E. Spooner, H. L. Ploegh, *ACS Chemical Biology* **2009**, *4*, 275.
[170] D. A. Boudreaux, T. K. Maiti, C. W. Davies, C. Das, *Proceedings of the National Academy of Sciences* **2010**, *107*, 9117.
[171] S. Misaghi, P. J. Galardy, W. J. N. Meester, H. Ovaa, H. L. Ploegh, R. Gaudet, *Journal of Biological Chemistry* **2005**, *280*, 1512.
[172] S. Dharadhar, M. Clerici, W. J. van Dijk, A. Fish, T. K. Sixma, *Journal of Structural Biology* **2016**, *196*, 437.
[173] M. P. C. Mulder, K. Witting, I. Berlin, J. N. Pruneda, K.-P. Wu, J.-G. Chang, R. Merkx, J. Bialas, M. Groettrup, A. C. O. Vertegaal, B. A. Schulman, D. Komander, J. Neefjes, F. El Oualid, H. Ovaa, *Nature Chemical Biology* **2016**, *12*, 523.
[174] G. C. Capodagli, M. K. Deaton, E. A. Baker, R. J. Lumpkin, S. D. Pegan, *Journal of Virology* **2013**, *87*, 3815.

[175] T. W. James, N. Frias-Staheli, J.-P. Bacik, J. M. Levingston Macleod, M. Khajehpour, A. García-Sastre, B. L. Mark, *Proceedings of the National Academy of Sciences* **2011**, *108*, 2222.
[176] W. Gui, C. A. Ott, K. Yang, J. S. Chung, S. Shen, Z. Zhuang, *Journal of the American Chemical Society* **2018**, *140*, 12424.
[177] W. Gui, S. Shen, Z. Zhuang, *Journal of the American Chemical Society* **2020**, *142*, 19493.
[178] D. S. Hameed, A. Sapmaz, L. Burggraaff, A. Amore, C. J. Slingerland, G. J. P. van Westen, H. Ovaa, *Angewandte Chemie International Edition* **2019**, *58*, 14477.
[179] A. C. Varca, D. Casalena, D. Auld, S. J. Buhrlage, *Star Protocols* **2021**, *2*.
[180] A. C. Varca, D. Casalena, W. C. Chan, B. Hu, R. S. Magin, R. M. Roberts, X. Liu, H. Zhu, H.-S. Seo, S. Dhe-Paganon, J. A. Marto, D. Auld, S. J. Buhrlage, *Cell Chemical Biology* **2021**, *28*, 1758.
[181] U. Hassiepen, U. Eidhoff, G. Meder, J.-F. Bulber, A. Hein, U. Bodendorf, E. Lorthiois, B. Martoglio, *Analytical Biochemistry* **2007**, *371*, 201.
[182] J. Fan, Y. Ye, G. Chu, Z. Zhang, Y. Fu, Y.-M. Li, J. Shi, *The Journal of Organic Chemistry* **2019**, *84*, 14861.
[183] J. A. Ward, L. McLellan, M. Stockley, K. R. Gibson, G. A. Whitlock, C. Knights, J. A. Harrigan, X. Jacq, E. W. Tate, *ACS Chemical Biology* **2016**, *11*, 3268.
[184] P. P. Geurink, F. El Oualid, A. Jonker, D. S. Hameed, H. Ovaa, *Chembiochem* **2012**, *13*, 293.
[185] J. N. Pruneda, C. H. Durkin, P. P. Geurink, H. Ovaa, B. Santhanam, D. W. Holden, D. Komander, *Molecular cell* **2016**, *63*, 261.
[186] T. R. Cotton, S. A. Cobbold, J. P. Bernardini, L. W. Richardson, X. S. Wang, B. C. Lechtenberg, *Molecular cell* **2022**, *82*, 598.
[187] M. Gersch, C. Gladkova, A. F. Schubert, M. A. Michel, S. Maslen, D. Komander, *Nat Struct Mol Biol* **2017**, *24*, 920.
[188] K. Keusekotten, Paul R. Elliott, L. Glockner, Berthe K. Fiil, Rune B. Damgaard, Y. Kulathu, T. Wauer, Manuela K. Hospenthal, M. Gyrd-Hansen, D. Krappmann, K. Hofmann, D. Komander, *Cell* **2013**, *153*, 1312.
[189] A. Iphöfer, A. Kummer, M. Nimtz, A. Ritter, T. Arnold, R. Frank, J. van den Heuvel, B. M. Kessler, L. Jänsch, R. Franke, *Chembiochem* **2012**, *13*, 1416.
[190] L. A. Hehl, D. Horn-Ghetko, J. R. Prabu, R. Vollrath, D. T. Vu, D. A. Pérez Berrocal, M. P. C. Mulder, G. J. van der Heden van Noort, B. A. Schulman, *Nature Chemical Biology* **2024**, *20*, 190.
[191] T. E. T. Mevissen, Y. Kulathu, M. P. C. Mulder, P. P. Geurink, S. L. Maslen, M. Gersch, P. R. Elliott, J. E. Burke, B. D. M. van Tol, M. Akutsu, F. E. Oualid, M. Kawasaki, S. M. V. Freund, H. Ovaa, D. Komander, *Nature* **2016**, *538*, 402.
[192] A. Weber, P. R. Elliott, A. Pinto-Fernandez, S. Bonham, B. M. Kessler, D. Komander, F. El Oualid, D. Krappmann, *Cell Chem Biol* **2017**, *24*, 1299.
[193] G. Li, Q. Liang, P. Gong, A. H. Tencer, Z. Zhuang, *Chemical Communications* **2014**, *50*, 216.
[194] Joanna F. McGouran, Selina R. Gaertner, M. Altun, Holger B. Kramer, Benedikt M. Kessler, *Chemistry & Biology* **2013**, *20*, 1447.
[195] N. Haj-Yahya, H. P. Hemantha, R. Meledin, S. Bondalapati, M. Seenaiah, A. Brik, *Organic Letters* **2014**, *16*, 540.
[196] A. Weber, P. R. Elliott, A. Pinto-Fernandez, S. Bonham, B. M. Kessler, D. Komander, F. El Oualid, D. Krappmann, *Cell Chemical Biology* **2017**, *24*, 1299.
[197] M. P. C. Mulder, R. Merkx, K. F. Witting, D. S. Hameed, D. El Atmioui, L. Lelieveld, F. Liebelt, J. Neefjes, I. Berlin, A. C. O. Vertegaal, H. Ovaa, *Angewandte Chemie International Edition* **2018**, *57*, 8958.
[198] L. C. Dang, F. D. Melandri, R. L. Stein, *Biochemistry* **1998**, *37*, 1868.
[199] M. S. Ritorto, R. Ewan, A. B. Perez-Oliva, A. Knebel, S. J. Buhrlage, M. Wightman, S. M. Kelly, N. T. Wood, S. Virdee, N. S. Gray, N. A. Morrice, D. R. Alessi, M. Trost, *Nat Commun* **2014**, *5*, 4763.
[200] A. Tirat, A. Schilb, V. Riou, L. Leder, B. Gerhartz, J. Zimmermann, S. Worpenberg, U. Eidhoff, F. Freuler, T. Stettler, L. Mayr, J. Ottl, B. Leuenberger, I. Filipuzzi, *Analytical Biochemistry* **2005**, *343*, 244.

[201] K. D. Wilkinson, S. E. Smith, L. O'Connor, E. Sternberg, J. J. Taggart, D. A. Berges, T. Butt, *Biochemistry* **1990**, *29*, 7373.
[202] Z. S. Hann, C. Ji, S. K. Olsen, X. Lu, M. C. Lux, D. S. Tan, C. D. Lima, *Proceedings of the National Academy of Sciences* **2019**, *116*, 15475.
[203] M. E. Morrow, M. T. Morgan, M. Clerici, K. Growkova, M. Yan, D. Komander, T. K. Sixma, M. Simicek, C. Wolberger, *EMBO reports* **2018**, *19*, e45680.
[204] A. Basters, P. P. Geurink, A. Rocker, K. F. Witting, R. Tadayon, S. Hess, M. S. Semrau, P. Storici, H. Ovaa, K. P. Knobeloch, G. Fritz, *Nat Struct Mol Biol* **2017**, *24*, 270.
[205] S. Sommer, N. D. Weikart, U. Linne, H. D. Mootz, *Bioorg Med Chem* **2013**, *21*, 2511.
[206] Y. Li, N. Varejão, D. Reverter, *Nature Communications* **2022**, *13*, 1819.
[207] V. Jové, H. Wheeler, C. W. Lee, D. R. Healy, K. Levine, E. C. Ralph, M. Yamaguchi, Z. K. Jiang, E. Cabral, Y. Xu, J. Stock, B. Yang, A. Giddabasappa, P. Loria, A. Casimiro-Garcia, B. M. Kessler, A. Pinto-Fernández, V. Frattini, P. D. Wes, F. Wang, *iScience* **2024**, *27*, 109593.
[208] J. Gan, A. Pinto-Fernández, D. Flierman, J. J. L. L. Akkermans, D. P. O'Brien, H. Greenwood, H. C. Scott, G. Fritz, K.-P. Knobeloch, J. Neefjes, H. van Dam, H. Ovaa, H. L. Ploegh, B. M. Kessler, P. P. Geurink, A. Sapmaz, *Proceedings of the National Academy of Sciences* **2023**, *120*, e2315163120.
[209] Y. Liu, H. A. Lashuel, S. Choi, X. Xing, A. Case, J. Ni, L.-A. Yeh, G. D. Cuny, R. L. Stein, P. T. Lansbury, *Chemistry & Biology* **2003**, *10*, 837.
[210] N. Panyain, A. Godinat, T. Lanyon-Hogg, S. Lachiondo-Ortega, E. J. Will, C. Soudy, M. Mondal, K. Mason, S. Elkhalifa, L. M. Smith, J. A. Harrigan, E. W. Tate, *Journal of the American Chemical Society* **2020**, *142*, 12020.
[211] L. Bragado, M. Magalnik, P. Mammi, A. Romero, N. Gaioli, B. Pozzi, A. Srebrow, *Nucleic Acids Research* **2022**, *50*, 12444.
[212] A. I. Khlebnikov, I. A. Schepetkin, M. T. Quinn, *Bioorganic & Medicinal Chemistry* **2008**, *16*, 2791.
[213] R. K. P. Tripathi, *European Journal of Medicinal Chemistry* **2020**, *188*, 111953.
[214] Y. Sato, E. Goto, Y. Shibata, Y. Kubota, A. Yamagata, S. Goto-Ito, K. Kubota, J.-i. Inoue, M. Takekawa, F. Tokunaga, S. Fukai, *Nature structural & molecular biology* **2015**, *22*, 222.
[215] G. D. Warren, T. Kitao, T. G. Franklin, J. V. Nguyen, P. P. Geurink, T. Kubori, H. Nagai, J. N. Pruneda, *Molecular cell* **2023**, *83*, 105.
[216] J. Yin, Allyn J. Schoeffler, K. Wickliffe, K. Newton, Melissa A. Starovasnik, Erin C. Dueber, Seth F. Harris, *Structure* **2015**, *23*, 2043.
[217] M. Altun, T. S. Walter, H. B. Kramer, P. Herr, A. Iphöfer, J. Boström, Y. David, A. Komsany, N. Ternette, A. Navon, D. I. Stuart, J. Ren, B. M. Kessler, *PloS one* **2015**, *10*, e0115344.
[218] E. M. Valkevich, R. G. Guenette, N. A. Sanchez, Y. C. Chen, Y. Ge, E. R. Strieter, *J Am Chem Soc* **2012**, *134*, 6916.
[219] E. F. V. Scriven, K. Turnbull, *Chemical Reviews* **1988**, *88*, 297.
[220] B. D. Fairbanks, M. P. Schwartz, C. N. Bowman, K. S. Anseth, *Biomaterials* **2009**, *30*, 6702.
[221] M. P. C. Mulder, F. El Oualid, J. ter Beek, H. Ovaa, *Chembiochem* **2014**, *15*, 946.
[222] T. E. T. Mevissen, Y. Kulathu, M. P. C. Mulder, P. P. Geurink, S. L. Maslen, M. Gersch, P. R. Elliott, J. E. Burke, B. D. M. van Tol, M. Akutsu, F. El Oualid, M. Kawasaki, S. M. V. Freund, H. Ovaa, D. Komander, *Nature* **2016**, *538*, 402.
[223] G. Li, Q. Liang, P. Gong, A. H. Tencer, Z. Zhuang, *Chem Commun (Camb)* **2014**, *50*, 216.
[224] X.-Y. Lu, X.-J. Shi, A. Hu, J.-Q. Wang, Y. Ding, W. Jiang, M. Sun, X. Zhao, J. Luo, W. Qi, B.-L. Song, *Nature* **2020**, *588*, 479.
[225] T. Kubori, T. Kitao, H. Ando, H. Nagai, *Cellular Microbiology* **2018**, *20*, e12840.
[226] F. Gorrec, *Drug Discovery Today* **2016**, *21*, 819.
[227] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583.

[228] C. Schlieker, W. A. Weihofen, E. Frijns, L. M. Kattenhorn, R. Gaudet, H. L. Ploegh, *Molecular cell* **2007**, *25*, 677.
[229] D. Komander, M. Rape, *Annu Rev Biochem* **2012**, *81*, 203.
[230] P. P. Geurink, B. D. M. van Tol, D. van Dalen, P. J. G. Brundel, T. E. T. Mevissen, J. N. Pruneda, P. R. Elliott, G. B. A. van Tilburg, D. Komander, H. Ovaa, *Chembiochem* **2016**, *17*, 816.
[231] Y. Li, N. Varejao, D. Reverter, *Nature Communications* **2022**, *13*.
[232] V. Jové, H. Wheeler, C. W. Lee, D. R. Healy, K. Levine, E. C. Ralph, M. Yamaguchi, Z. K. Jiang, E. Cabral, Y. Xu, J. Stock, B. Yang, A. Giddabasappa, P. Loria, A. Casimiro-Garcia, B. M. Kessler, A. Pinto-Fernández, V. Frattini, P. D. Wes, F. Wang, *iScience* **2024**, *27*.
[233] A. Aichem, S. Anders, N. Catone, P. Rößler, S. Stotz, A. Berg, R. Schwab, S. Scheuermann, J. Bialas, M. C. Schütz-Stoffregen, G. Schmidtke, C. Peter, M. Groettrup, S. Wiesner, *Nature Communications* **2018**, *9*, 3321.
[234] N. D. Roverato, C. Sailer, N. Catone, A. Aichem, F. Stengel, M. Groettrup, *Cell Reports* **2021**, *34*, 108857.
[235] W. Ge, C. Yu, J. Li, Z. Yu, X. Li, Y. Zhang, C.-P. Liu, Y. Li, C. Tian, X. Zhang, G. Li, B. Zhu, R.-M. Xu, *Nature* **2023**, *616*, 176.
[236] J. F. Thomas, M. I. Valencia-Sánchez, S. Tamburri, S. L. Gloor, S. Rustichelli, V. Godínez-López, P. De Ioannes, R. Lee, S. Abini-Agbomson, K. Gretarsson, J. M. Burg, A. R. Hickman, L. Sun, S. Gopinath, H. F. Taylor, Z.-W. Sun, R. J. Ezell, A. Vaidya, M. J. Meiners, M. A. Cheek, W. J. Rice, V. Svetlov, E. Nudler, C. Lu, M.-C. Keogh, D. Pasini, K.-J. Armache, *Science Advances* **2023**, *9*, eadg9832.
[237] H. Ai, Z. He, Z. Deng, G.-C. Chu, Q. Shi, Z. Tong, J.-B. Li, M. Pan, L. Liu, *Nature structural & molecular biology* **2024**.
[238] T. Klemm, G. Ebert, D. J. Calleja, C. C. Allison, L. W. Richardson, J. P. Bernardini, B. G. Lu, N. W. Kuchel, C. Grohmann, Y. Shibata, Z. Y. Gan, J. P. Cooney, M. Doerflinger, A. E. Au, T. R. Blackmore, G. J. van der Heden van Noort, P. P. Geurink, H. Ovaa, J. Newman, A. Riboldi-Tunnicliffe, P. E. Czabotar, J. P. Mitchell, R. Feltham, B. C. Lechtenberg, K. N. Lowes, G. Dewson, M. Pellegrini, G. Lessene, D. Komander, *Embo J* **2020**, *39*, e106275.
[239] J. N. Pruneda, C. H. Durkin, P. P. Geurink, H. Ovaa, B. Santhanam, D. W. Holden, D. Komander, *Mol Cell* **2016**, *63*, 261.
[240] D. Hammler, A. Marx, A. Zumbusch, *Chem Eur J* **2018**, *24*, 15329.
[241] M. V. Kvach, I. A. Stepanova, I. A. Prokhorenko, A. P. Stupak, D. A. Bolibrukh, V. A. Korshun, V. V. Shmanai, *Bioconjugate Chem* **2009**, *20*, 1673.
[242] C. Grethe, M. Schmidt, G. M. Kipka, R. O'Dea, K. Gallant, P. Janning, M. Gersch, *Nat Commun* **2022**, *13*, 5950.
[243] R. O'Dea, N. Kazi, A. Hoffmann-Benito, Z. Zhao, S. Recknagel, K. Wendrich, P. Janning, M. Gersch, *Nat Chem Biol* **2023**, *in revision*.
[244] A. Friese, S. Kapoor, T. Schneidewind, S. R. Vidadala, J. Sardana, A. Brause, T. Forster, M. Bischoff, J. Wagner, P. Janning, S. Ziegler, H. Waldmann, *Angew Chem Int Ed Engl* **2019**, *58*, 13009.
[245] V. Boll, T. Hermanns, M. Uthoff, I. Erven, E. M. Horner, V. Kozjak-Pavlovic, U. Baumann, K. Hofmann, *Nat Commun* **2023**, *14*, 7335.
[246] G. D. Van Duyne, R. F. Standaert, P. A. Karplus, S. L. Schreiber, J. Clardy, *J Mol Biol* **1993**, *229*, 105.
[247] V. Bernier-Villamor, D. A. Sampson, M. J. Matunis, C. D. Lima, *Cell* **2002**, *108*, 345.
[248] L. M. Lois, C. D. Lima, *EMBO J* **2005**, *24*, 439.
[249] A. Werner, A. Flotho, F. Melchior, *Mol Cell* **2012**, *46*, 287.
[250] W. Kabsch, *Acta Cryst* **2010**, *D66*, 125.
[251] J. Beilsten-Edmands, G. Winter, R. Gildea, J. Parkhurst, D. Waterman, G. Evans, *Acta Cryst* **2020**, *D76*, 385.
[252] P. R. Evans, G. N. Murshudov, *Acta Cryst* **2013**, *D69*, 1204.
[253] G. M. Sheldrick, *Acta Cryst* **2010**, *D66*, 479.
[254] K. Cowtan, *Acta Cryst* **2010**, *D66*, 470.

[255] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Cryst* **2011**, *D67*, 355.
[256] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Cryst* **2010**, *D66*, 486.
[257] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta Cryst* **2012**, *D68*, 352.
[258] A. J. McCoy, *Acta Cryst* **2007**, *D63*, 32.
[259] G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter, G. Evans, *Acta Crystallographica Section D-Structural Biology* **2018**, *74*, 85.
[260] I. J. Tickle, Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G., **2016**.
[261] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583.
[262] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Zídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, *Nucleic Acids Research* **2022**, *50*, D439.
[263] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallographica Section D-Structural Biology* **2011**, *67*, 235.
[264] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Crystallographica Section D-Structural Biology* **2019**, *75*, 861.
[265] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66*, 486.

ACKNOWLEDGEMENT

I never thought that my journey to pursue my Doctor rerum naturalium would take five years but I firmly believe that good science always takes time.

I would like to sincerely thank my supervisor Dr. Malte Gersch for giving me this wonderful opportunity to dive into the DUB ocean though I had no any experience in protein biochemistry as well as always being supportive. I really appreciate that we always focus on the quality of our scientific research instead of the quantity of publications.

I am immensely grateful to Prof. Herbert Waldmann for being my thesis advisory member and establishing the Chemical Genomics Centre III. I am so fortune to learn so much knowledge from this great chemical biologist. His generosity for sharing all the equipment in department IV gave me opportunities to try all the possibilities.

I would like to thank Prof. Daniel Summerer for being my thesis advisory member and sharing valuable plasmids.

I would like to thank Prof. Kay Hofmann and his team for generously sharing experimental resources and providing insightful suggestions. It was a great collaboration and I am always amazed by his sequence alignment.

I would like to thank all my colleagues in the Gersch lab for sharing reagents, ideas and knowledge. I would like to thank Dr. R. O'Dea, K. Wendrich and N. Kazi for helping me purify proteins and reagents for my projects and Dr. M. Schmidt for sharing resources. I would also like to thank G. Kipka for correcting the abstract.

I would like to thank Dr. L. Sironi and C. Hornemann for offering kind help and IMPRS for generous financial support for my studies.

I would like to thank Dr. S. Sievers and her team at the COMAS for high-throughput screen.

I would like to thank the support from local Chinese students and scholars, particularly Drs. L. Wang, F. Huang, J. Xie, G. Xue, X. Cheng, W. Ma, and H. Guo. I would also like to thank all the peers, especially Y. Liu for always being patient and supportive.

I would like to thank my friends Dr. Z. Li and T. Sun for their kind support.

Last but not least, I would like to thank my parents for their constant and unwavering support.