UNIVERSITÀ
DI PAVIA

DIPARTIMENTO DI SCIENZE DEL FARMACO

Direttore Chiar.ma Prof.ssa Simona Collina

**LAUREA MAGISTRALE A CICLO UNICO IN
CHIMICA E TECNOLOGIA FARMACEUTICHE**

Development and characterisation of a peptide library for the investigation of SDE2 cleavage mechanism by USP5

Sviluppo e caratterizzazione di una libreria di peptidi per lo studio del meccanismo di proteolisi di SDE2 mediato da USP5

Relatore: *Prof.ssa Sofia Giorgetti*

Correlatore: *Dott.ssa Chiara Maniaci*

Tesi di Laurea Magistrale a Ciclo Unico di
Flavia Foggetti

Anno Accademico 2024/2025

Index

Abstract

Posttranslational modifications (PTMs) are chemical variations at the aminoacidic level that can modulate protein functions and regulate stress response pathways. Among these, the proteolytic removal of ubiquitin-like domains is a key regulatory step.

The protein Silencing Defective 2 (SDE2) is synthesised with an N-terminal ubiquitin-like domain ($SDE2_{UBL}$) that must be cleaved by the deubiquitinating enzyme USP5 (Ubiquitin Specific Protease 5) to generate the active C-terminal fragment ($SDE2_{CT}$), that participates in DNA repair, ribosome biogenesis and mRNA splicing. SDE2 processing occurs at lysine 78 (K78), downstream the conserved G76G77 motif. Preliminary data from an in vitro assay performed by incubating USP5 with 25-amino-acid peptide representing $SDE2_{66–90}$ sequence revealed two additional cleavage sites beyond the one previously reported. This finding raised the hypothesis that additional sequence or structural determinants might influence substrate recognition and processing.

In this thesis, the sequence determinants required for $SDE2_{UBL}$ processing were investigated, to elucidate the mechanism of SDE2 maturation and ultimately support the design of USP5 inhibitors or modulators.

To do so, a library of peptides has been designed, mutating the sequence of the original peptide, and tested in an in-vitro assay and analysed by MALDI-TOF.

The analysis revealed that cleavage occurred exclusively at lysine and arginine residues, regardless of the identity of amino acids in the sequence. However, the observed cleavage pattern was not replicated when a new batch of recombinant USP5 was employed in the assay, raising the possibility that a non-specific protease contaminant had been present in the initial preparation.

Subsequent proteomic analysis of both batches identified Oligopeptidase B, a contaminant from *E. coli*, present exclusively in the first batch, suggesting that this enzyme was responsible for the observed peptide cleavage activity.

Le modificazioni post-traduzionali (PTMs) sono variazioni chimiche a livello amminoacidico che possono modulare le funzioni proteiche e regolare i meccanismi di risposta a stress cellulari. Tra queste, la rimozione dei domini, incluso i domini *ubiquitin-like,* rappresenta un passaggio regolatorio chiave.

La proteina *Silencing Defective 2* (SDE2) viene sintetizzata con un dominio N-terminale *ubiquitin-like* ($SDE2_{UBL}$), che deve essere rimosso dall'enzima deubiquitinante USP5 (*Ubiquitin Specific Protease 5*) per generare il frammento C-terminale attivo ($SDE2_{CT}$), che partecipa alla riparazione del DNA, alla biogenesi dei ribosomi e allo splicing dell'mRNA. La maturazione di SDE2 avviene in corrispondenza della lisina 78 (K78), dove è localizzato il motivo G76G77. Dati preliminari provenienti da un saggio in vitro eseguito incubando USP5 con un peptide di 25 amminoacidi rappresentante la sequenza $SDE2_{66\text{-}90}$ hanno rivelato due siti di idrolisi aggiuntivi oltre a quello precedentemente riportato. Questo risultato ha sollevato l'ipotesi che ulteriori componenti della sequenza o struttura potessero influenzare il riconoscimento e la maturazione del substrato.

In questa tesi sono stati investigati i residui della sequenza necessari per il *processing* di $SDE2_{UBL}$, con l'obiettivo di chiarire il meccanismo di maturazione di SDE2 e, in ultima analisi, supportare la progettazione di inibitori o modulatori di USP5.

A tal fine, è stata progettata una libreria di peptidi con mutazioni nella sequenza del peptide originale, e i peptidi sono stati testati utilizzando lo stesso approccio.

L'analisi ha rivelato che l'idrolisi avveniva esclusivamente in corrispondenza di residui di lisina e arginina, indipendentemente dall'identità degli altri amminoacidi nella sequenza. Tuttavia, il *pattern* osservato non è stato replicato quando un nuovo lotto di USP5 ricombinante è stato utilizzato nel saggio, suggerendo la possibile presenza di una contaminazione da proteasi aspecifica nella preparazione iniziale.

La successiva analisi proteomica di entrambi i lotti ha identificato l'Oligopeptidasi B, un contaminante proveniente da *E. coli*, presente esclusivamente nel primo lotto, indicando che tale enzima fosse responsabile dell'attività osservata.

Introduction

The ubiquitin system

Living cells are constantly exposed to stressors, such as variations in temperature, pH, pathogens, oxidative species or misfolded proteins. Perturbations of cellular homeostasis, the process by which biological systems maintain a balanced internal environment despite external fluctuations[1], can either trigger adaptive stress-management and repair pathways or, when recovery fails, programmed cell death. To survive, organisms have evolved sophisticated surveillance and recovery systems that detect damage, limit its spread, and restore function[2], [3].

Post-translational modifications (PTMs) are enzymatic and non-enzymatic chemical variations that can occur to specific residues of a protein, enriching protein heterogeneity beyond the genome encoded sequence and governing cellular effects[4].

Ubiquitination exemplifies a PTM that is highly conserved amongst eukaryotes. Ubiquitin is a 76-amino-acid protein that adopts a β-grasp (also referred to as ubiquitin-like) fold, consisting of a five-strand antiparallel β-sheet wrapped around a central α-helix and a short C-terminal tail ending in the characteristic diglycine motif required for conjugation to substrates[5]. Ubiquitination consists of a three-step enzymatic cascade where the catalytic cysteine of an E1 ubiquitin-activating enzyme forms a thioester bond with Ub C-terminal glycine (G76) via an ATP-dependent step. The activated ubiquitin is then transferred to an E2 ubiquitin-conjugating enzyme trough a transesterification reaction, producing an E2-Ub thioester intermediate. As the last step, an E3 ubiquitin ligase binds the E2-Ub complex and the substrate protein, catalysing the formation of an isopeptide bond between Ub G76 and a nucleophilic residue of the target protein (Figure 1)[3]. This

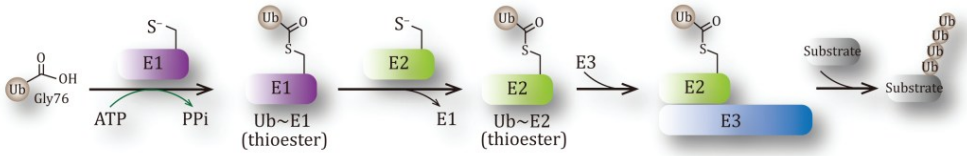


Figure 1: Schematic illustration of ubiquitination via the E1–E2–E3 enzyme cascade. Ubiquitin is first activated by the E1 enzyme in an ATP-dependent reaction, generating a high-energy thioester intermediate. It is then transferred to an E2 conjugating enzyme through a transesterification step. An E3 ligase subsequently recruits both the ubiquitin-charged E2 and the substrate, positioning them for efficient transfer. Successive rounds of this reaction lead to the assembly of a polyubiquitin chain on the substrate. [3]

hierarchical system confers substrate specificity, as a single E2 can partner with different E3s that recognize distinct substrates[6].

Ubiquitin can be linked to lysine, serine, threonine, or cysteine residues on target proteins, generating either a single ubiquitin attachment (monoubiquitination) or extended polymeric chains, producing different cellular responses[7].

While ubiquitination is well known for its proteolytic role, it also performs non-proteolytic functions, regulating protein-protein interactions, endocytosis, cell cycle advancement, and the activation or inhibition of specific substrates. It also participates in inflammatory signalling pathways, autophagic processes, DNA repair mechanisms and control of enzymatic activities (Figure 2) [8], [9], [10]. Therefore, disruption of the ubiquitination system could cause misregulation of signalling pathways, including those that drive cancer or control metabolism, and can interfere with the proper formation of protein complexes needed for inflammation regulation or DNA repair mechanisms. Moreover, faulty ubiquitination may lead to protein misplacement or buildup of misfolded proteins, contributing to neurodegenerative disorders[11].

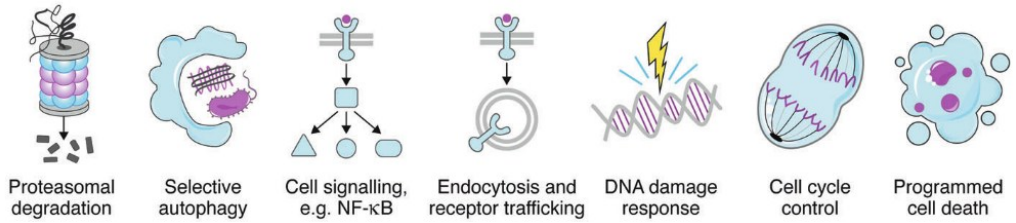


Figure 2: Representation of the diverse roles of ubiquitin in regulating protein fate and cell physiology. Ubiquitin conjugation can target proteins for proteasomal degradation, modulate their localization, influence signalling pathways, alter DNA damage responses, and control processes such as cell-cycle progression and autophag[10].

Ubiquitin-like-proteins (UBLs) and UBL domains

Ubiquitin-like proteins (UBLs) are a family of small modifiers that share with ubiquitin its characteristic fold and a C-terminal diglycine motif that enables enzymatic attachment to lysine residues on target proteins through an E1-E2-E3 cascade, thereby regulating processes such as transcription, cell-cycle control, DNA-damage responses, innate immunity, autophagy, and ribosome biogenesis[12]. The functional outcome of a UBL modification depends on the specific modifier: NEDD8 activates cullin-RING ligases[13], SUMO modulates nuclear protein interactions and chromatin dynamics[14].

Moreover, many cellular proteins contain a UBL domain within their primary sequence, that acts as regulator [12]. An example is FUBI, a UBL that is cleaved from the ribosomal protein eS30 during early 40S subunit biogenesis. This processing step is essential for proper maturation of the small ribosomal subunit and for maintaining proteostasis within the nucleolus[15].

Silencing defective 2 (SDE2)

SDE2 (Sde2 in lower eukaryotes) is a nuclear protein that has emerged in recent years as an important factor in several fundamental cellular pathways. Its name originates from genetic screenings in fission yeast *Schizosaccharomyces pombe*, where the deletion of Sde2 gave rise to defects in gene silencing and replication, hence Silencing Defective 2[16]. Homologs of Sde2 have been identified in other eukaryotes, including

mammalian cells. The *S. pombe* Sde2 protein comprises 263 amino acids, whereas its human homolog consists of 451 amino acids. Both proteins are synthesized as a precursor containing an N-terminal ubiquitin-like domain ($Sde2_{UBL}$, $SDE2_{UBL}$), a conserved GGKGG motif, and a C-terminal domain ($Sde2_{CT}$, $SDE2_{CT}$)[17]. The N-terminal UBL domain adopts the typical ubiquitin like β-grasp fold (Figure 3). Proteolytic processing at the conserved diglycine motif releases the active C-terminal fragment, a key activation step that enables SDE2 to participate in spliceosomal regulation and genome maintenance pathways[18], [19].

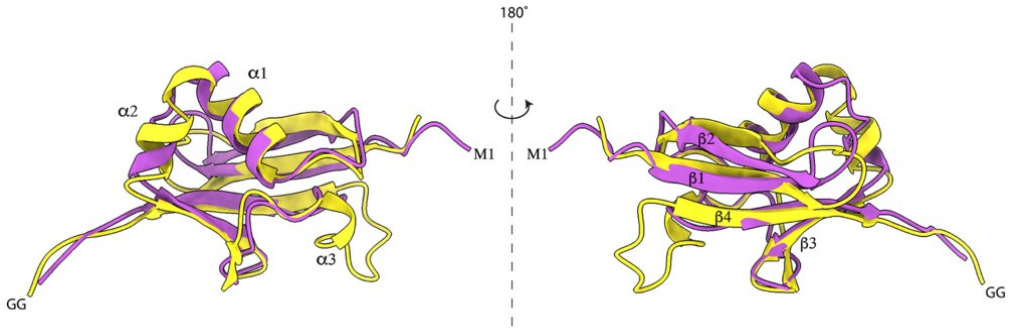


Figure 3: Structural overlay of the yeast (yellow) and human (purple) UBL domains. The N-terminal UBL domain of SDE2 adopts a conserved β-grasp fold characteristic of ubiquitin, composed of a central α-helix enclosed by a five-stranded β-sheet[34].

Intron specific mRNA splicing

SDE2 is a key regulator of intron-specific pre-mRNA splicing, facilitating the removal of introns with weak or non-canonical splicing signals. In *S. pombe*, $Sde2_{CT}$ integrates into the spliceosome and recruits Cactin/Cay1, promoting efficient splicing of a subset of challenging introns. Loss or improper processing of Sde2 leads to splicing defects, impaired telomeric silencing, genomic instability, and sensitivity to stress[18].

In mammalian cells, through interactions with SF3B1, U2AF1, and the auxiliary factor Cactin, SDE2 associates with the U2 snRNP complex, suggesting a role in 3′ splice-site recognition and late catalytic steps of splicing. Structural studies place SDE2 within the post-catalytic (P)

spliceosomal complex, where it acts after exon ligation. Loss of SDE2 leads to widespread alternative splicing alterations and cell death[20].

SDE2 role in ribosome biogenesis

Beyond its role in pre-mRNA splicing, SDE2 has been identified as a crucial factor in ribosome biogenesis. SDE2 directly associates with C/D box small nucleolar RNAs (snoRNAs) and is required for efficient processing of ribosome RNA (rRNA) precursors.

SDE2 exemplifies a non-canonical RNA-binding protein (RBP) that lacks classical RNA-binding motifs[21]. Instead, RNA recognition is mediated through an intrinsically disordered region (IDR) and a SAP domain[22].

SDE2 demonstrates selective binding to members of the C/D box snoRNA family, which are defined by two highly conserved C and D box motifs and two less conserved C′ and D′ boxes. The opposing arrangement of these motifs promotes the folding of snoRNAs into a characteristic hairpin structure, enabling assembly of the functional C/D box snoRNP complex. Biochemical analyses indicate that the SAP domain and/or IDR of SDE2 contact regions near the C box motif, facilitating snoRNA stabilization and maturation. Deficiency of SDE2 disrupts rRNA processing and nucleolar architecture, impairing ribosome biogenesis and reducing protein synthesis. The resulting nucleolar stress activates p53 pathway, leading to apoptotic cell death, highlighting SDE2's role as a pivotal link between RNA metabolism, splicing, and ribosome production that sustains cellular growth and homeostasis[20].

SDE2 Interaction with TIMELESS (TIM) and replication fork protection

The Fork Protection Complex (FPC), mainly consisting in TIMELESS (TIM) and TIPIN, engages with replication origins at the beginning of S phase, coupling helicase and polymerase activities[23]. SDE2 interacts with the C-terminal region of TIM, promoting its stability and localization at replication forks. Through its DNA-binding SAP domain, SDE2

anchors the FPC to chromatin, allowing it to move in coordination with the replisome. Knockdown of SDE2 or disruption of its interaction with TIM resembles TIM-deficient phenotypes, displaying extensive ssDNA accumulation, fork progression defects, and impaired recovery of stalled forks, underscoring SDE2's essential role in preserving replication fork stability and genome integrity[24].

Attenuation of replication stress via PCNA interaction

SDE2 plays a pivotal role in the human cellular response to replication stress through its interaction with Proliferating Cell Nuclear Antigen (PCNA), a central coordinator of DNA replication and repair. PCNA acts as a sliding clamp that recruits replication and repair factors, single-stranded (ss) DNA generated at stalled replication forks activates RAD18 ubiquitin E3 ligase to monoubiquitinate PCNA, creating a docking site for translesion synthesis (TLS) polymerases, enabling DNA synthesis across damaged templates[25][26]. SDE2 contains a conserved PCNA-interacting peptide (PIP) motif within its UBL domain, which mediates its recruitment to replication forks. Upon PCNA binding, SDE2 undergoes proteolytic cleavage, allowing $SDE2_{CT}$ to contribute to fork stabilization under normal conditions, by negatively regulates damage-inducible PCNA monoubiquitination. Under replication stress, however, the PCNA interaction facilitates SDE2 degradation via the $CRL4^{CDT2}$ E3 ubiquitin ligase complex. Removal of SDE2 enables the recruitment of TLS polymerases to monoubiquitinated PCNA, allowing lesion bypass and fork restart. Depletion of SDE2 leads to elevated PCNA-Ub levels, and hypersensitivity to agents that causes replication-associated damage and reduced cell survival. On the other hand, failure to cleave or degrade SDE2 impairs S-phase progression and cell proliferation[19].

Regulation through the N-end rule degradation pathway

The regulation of SDE2 cellular levels is mediated by the Arg/N-end rule pathway, a branch of the UPS that links a protein's half-life to the identity of its N-terminal residue[27].

Following cleavage of its UBL, the resulting $SDE2_{CT}$ fragment exposes an N-terminal lysine residue that acts as an N-degron, a degradation signal recognized by specialized E3 ubiquitin ligases called N-recognins. In mammalian cells the N-recognins of SDE2 are UBR1 and UBR2, which bind to the N-degron and facilitate polyubiquitination of the substrate, marking it for degradation by the proteasome[28].

Structural modelling shows that the K78 of $SDE2_{CT}$ forms hydrogen bonds with negatively charged residues (Asp118, Asp150 and Asp153) within the UBR box, that enables the N-recognins to identify positively charged N-degrons[29], [30]. Through this mechanism, the Arg/N-end rule pathway ensures rapid turnover of SDE2 once its functional role has been fulfilled, which is crucial for maintaining replication fork integrity and promoting cell survival under replication stress[28].

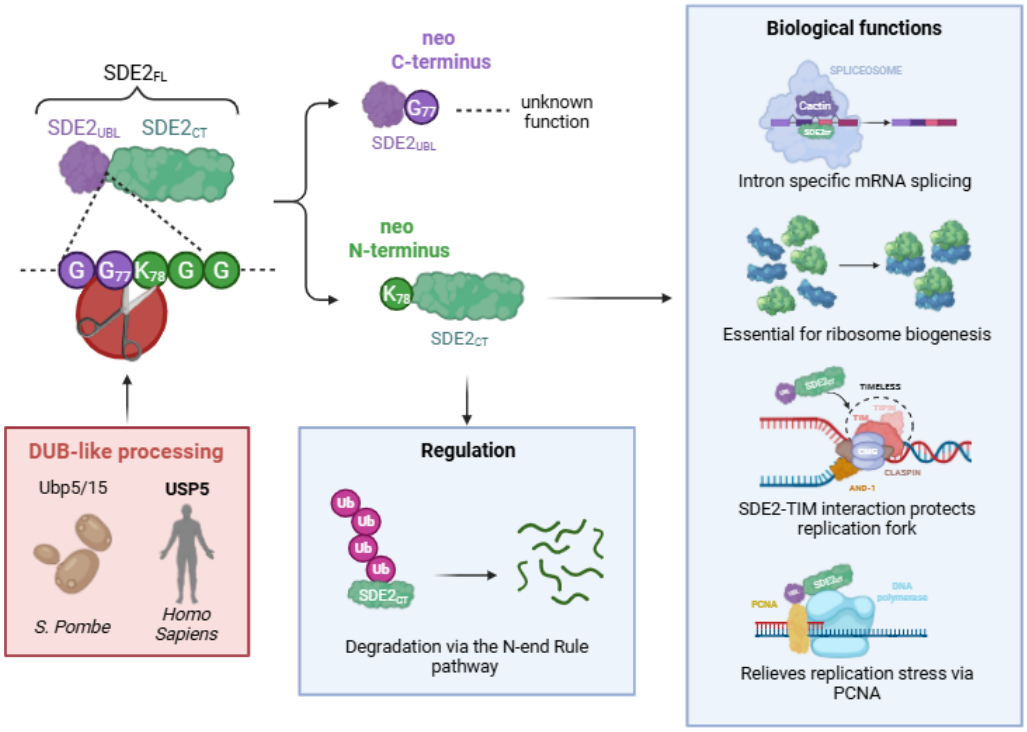


Figure 4: Schematic representation of SDE2 processing and biological functions. Cleavage after the conserved diglycine (Gly-Gly) motif generates two products: a neo C-terminus containing the UBL domain, and a neo N-terminus corresponding to $SDE2_{UBL}$. This processing is mediated by DUB enzymes, Ubp5/15 in S. pombe and USP5 in humans, and regulates SDE2 stability, with the C-terminal fragment targeted for degradation via the N-end rule pathway. On the right, major biological functions of SDE2 are illustrated, including roles in intron-specific mRNA splicing, ribosome biogenesis, replication-fork protection through interaction with TIM, and alleviation of replication stress via PCNA[34].

Deubiquitinating enzymes (DUBs)

Ubiquitination and ubiquitin-like modifications rely on reversible signalling, so cells require mechanisms both to attach these modifiers and to remove them when appropriate. This is carried out by a large and diverse group of enzymes known as deubiquitinating enzymes (DUBs). By cleaving ubiquitin from modified proteins, either by removing single units from the ends of ubiquitin chains or by releasing entire chains in a single step, DUBs modulate substrate fate and recycle ubiquitin for new rounds of conjugation, this way maintaining proteostasis and broader cellular homeostasis[31].

DUBs fall into two major enzyme classes: cysteine proteases and metalloproteases. The cysteine protease group comprises six families,

whose members share conserved sequence features and characteristic domains. They typically use a catalytic triad involving cysteine, histidine, and either asparagine or aspartate to cleave isopeptide bonds (Figure 5). Metalloprotease DUBs, consist of a zinc-dependent catalytic mechanism and a serine residue to hydrolyse ubiquitin-substrate linkages[32].

Among the DUB families, the USP (ubiquitin-specific protease) group, of the cysteine proteases family, is the largest. USPs are characterized by a distinctive architectural arrangement often compared to a hand, consisting of a thumb, a palm and fingers subdomains. The catalytic residues sit between the thumb and palm, while the fingers engage ubiquitin to stabilize substrate binding[33].

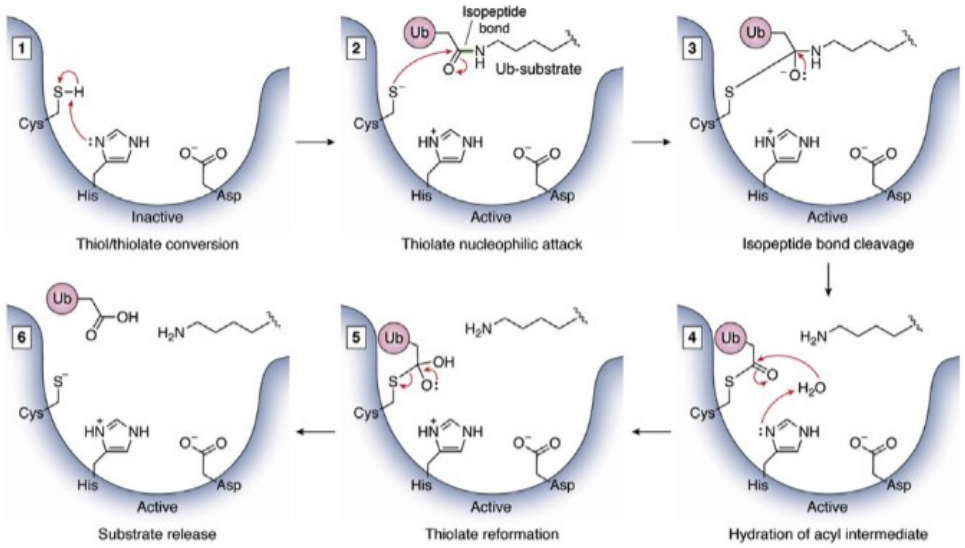


Figure 5: Schematic representation of the mechanism of cysteine protease DUBs: The catalytic core operates through a classic triadic arrangement of Cys, His and Asp residues. First, the Asp side chain stabilises and polarises the adjacent His, enabling His to abstract a proton from the Cys thiol and generate a highly nucleophilic thiolate anion. This thiolate then attacks the carbonyl carbon of the ubiquitin-linked isopeptide bond, producing a tetrameric intermediate in which the ubiquitin's acyl group is covalently linked to the enzyme's Cys. Collapse of this intermediate releases the substrate's amine, while the ubiquitin remains thioester-bound to the DUB. A water molecule subsequently hydrolyses the thioester, converting it into a carboxylate intermediate that resolves to free ubiquitin and regenerates the catalytic cysteine thiolate. Finally, both the liberated ubiquitin monomer and the deubiquitinated substrate dissociate from the enzyme, completing the catalytic cycle[33].

Sde2 processing in yeast by Ubp5 and Ubp15

Genetic and biochemical studies have identified two deubiquitinating enzymes in *S. pombe*, Ubp5 and Ubp15, as being jointly responsible for

processing Sde2 in yeast. Deletion of either Ubp15 or Ubp5 leads to reduced cleavage, but only the double mutant (ΔUbp5 ΔUbp15) fully abolishes Sde2 processing. Subcellular localization experiments further refined this model: by fusing Sde2's UBL to GFP modified with either a nuclear localization signal (NLS) or a nuclear export signal (NES), it was shown that efficient processing requires nuclear localization[17].

SDE2 processing in human cells by USP5

Until recently, the identity of the human protease responsible for this maturation step was unknown. A recent preprint, however, conducted Protein BLAST searches using Ubp5 and Ubp15 as queries to identify the closer human homolog to the yeast proteases. USP7 emerged as the closest human match, showing 34% overall identity and up to 50% identity within the catalytic domain, making it a plausible functional analogue. Despite this sequence match, USP7 failed to cleave recombinant full-length human SDE2 in vitro, even though it processed a di-ubiquitin control substrate. To understand this discrepancy, the authors compared yeast and human SDE2 and found that, while the proteins share 39% identity and a highly conserved GGKGGFGS cleavage motif, the yeast UBL domain contains an extra loop and α3 helix absent in human SDE2. Structural modelling confirmed that these divergent features lie away from the conserved cleavage region, suggesting that species-specific differences may affect protease recognition. Through a combined workflow, including in vitro assays, quantitative proteomics, mass spectrometry, and activity-based probes, the authors identified USP5 as the protease responsible for SDE2 processing in humans and demonstrated that USP5 directly cleaves SDE2 both in vitro and in cells[34]. This is unprecedented as USP5 was so far only known to specifically process unanchored linear ubiquitin chains via the recognition of the free C-terminal.

They further showed via biophysical binding studies that the SDE2 UBL domain binds USP5, though with lower affinity compared to ubiquitin.

USP5 (also known as isopeptidase T) is an 835-residue multidomain enzyme. Its domain architecture includes two ZnF-UBP (zinc-finger ubiquitin–binding) domains. The first being an N-terminal ZnF-UBP (residues 1–156) that is tightly associated with the catalytic core and is essential for catalytic activity. The second is a C-terminal ZnF-UBP (residues 163–291) that binds the free C-terminal diglycine motif of ubiquitin, helping recruit unanchored ubiquitin chains. A catalytic USP domain hosts the active-site essential for proteolysis. Two UBA (ubiquitin-associated) domains near the C-terminus (residues 634–684 and 698–749), contribute to ubiquitin binding[35].

The model proposed in the preprint suggests that the $SDE2_{UBL}$ mimics ubiquitin in binding the ZnF-UBP domain, presenting its diglycine motif in a way that allows the catalytic site to access and cleave the peptide bond. The lower binding affinity observed, compared to ubiquitin, may reflect subtle differences in the UBL fold[34].

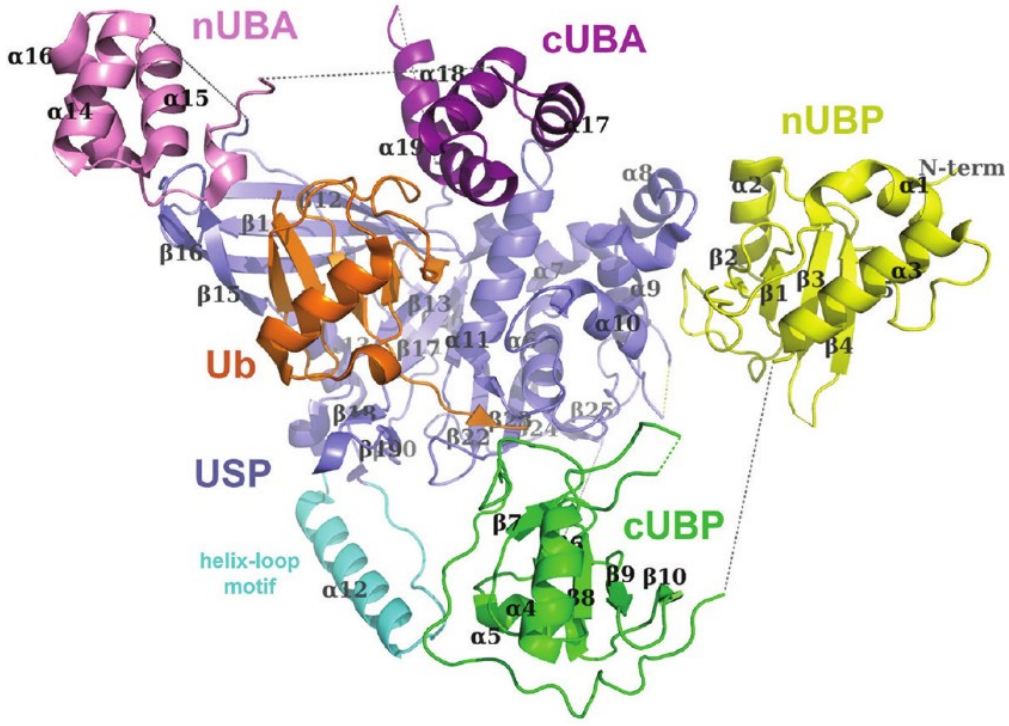


Figure 6: Ribbon representation of USP5 in complex with Ub (orange), nUBP in yellow, cUBP in green, nUBA in violet, and cUBA in magenta [35].

Aim of the work

Building on the identification of USP5 as the protease responsible for cleaving the N-terminal ubiquitin-like domain of SDE2, this thesis aims to investigate the sequence determinants required for $SDE2_{UBL}$ processing with the objective to understand the mechanism of SDE2 maturation in order to design inhibitors/modulators of USP5.

My thesis work concerned the study of the residues involved in SDE2 cleavage.

For this purpose, 25-amino-acid peptide, synthesised on solid phase, was tested in a recombinant cleavage assay monitored by MALDI-TOF mass spectrometry.

Then, a library of peptide variants was designed, introducing targeted mutations across the cleavage region to identify the residues essential for processing.

Methods

Reagents and solvents used were purchased from Sigma-Aldrich, Merck, Fluorochem or CEM.

Peptide synthesis

The synthesis was performed using a Liberty Blue™ 2.0 peptide synthesiser, on 90 mesh rink amide resin, N-terminally protected with Fmoc, with a loading capacity of 0,18 g/mol. The resin was swollen with DCM on a shaker at room temperature for 30 min. DCM was then drained on a vacuum manifold and the resin was washed with peptide grade DMF before transferring it into the reaction vessel. Synthesis cycle: 1) Fmoc deprotection with 4 mL 10% piperidine in DMF, 1 min at 90°C; 2) peptide grade DMF wash (4 × 4 mL); 3) 90°C 2 min amino acid coupling, coupling agents were DIC and Oxyma in peptide grade DMF; 4) peptide grade DMF wash (4 × 4 mL). Double coupling approach was used for arginine, leucine and amino acids right after. 0,01 mmol scale reaction: 0,056 g resin, 0,08 M amino acids, 0,1 M DIC, 0,1 M Oxyma. 0,1 mmol scale reaction: 0,560 g resin, 0,2 M amino acids, 1 M DIC, 1 M Oxyma. The resin was collected, rinsed with DCM and dried on vacuum. The resin was then incubated at r.t. for 3 h with cleavage mix (92.5% TFA, 5% water, 2.5% TIPS). The peptides were precipitated in cold diethyl ether and centrifuged at 4000 rpm, 4°C for 10 min. The precipitate was dissolved in 50:50 water/acetonitrile and freeze-dried overnight.

Peptide analysis by LC-MS

LCMS was performed on a Zorbax 300SB-C3 5 µm (150 × 2.1 mm) column coupled to an Agilent system composed of an autosampler, a 1260 Bin Pump VL and a vacuum degasser, a 1290 Thermostatted Column Compartment and a 6130 Quadropole LC/MS. Samples were prepared in 50:50 H2O/MeCN, injection volumes were between 1-5 µL, the mobile

phase was a gradient of 10-95% MeCN (0.05% TFA) in H2O (0.05% TFA) over 30 min. Data analysis was performed using Agilent ChemStation software.

Peptides cleavage assay

Working solutions of each peptide at a concentration of 40 µM were prepared by diluting a 10 mM stock solution (in DMSO) with buffer (Tris 50 mM, pH 7,5, 1 mM DTT). Working solutions of USP5WT at a concentration of 4 µM were prepared in the same buffer. 5 µL of peptide first and then enzyme working solutions were pipetted in 384 wells plates and incubated at 37°C for 1 h. The reaction was quenched with 2 µL of 12% TFA. Samples were mixed 1:1 with 2′,6′-Dihydroxyacetophenone (DHAP) MALDI matrix. The matrix was prepared combining 375 µL of a 133 mM 2,5-DHAP saturated solution in ethanol and 125 µL of a 110 mM solution of diammonium hydrogen citrate in water. 0,8 µL MALDI matrix mixed samples were then spotted onto a MTP 384 ground steel plate (Bruker). The samples were then analysed in reflective negative mode by MALDI-TOF on a UltrafleXtreme (Bruker. Daltonics).

His-USP5^{WT} expression

Competent Escherichia coli LOBSTR cells were incubated on ice for 30 min with the respective plasmid (purchased from MRC PPU reagents and services), heat-shocked at 42°C in water bath for 45 sec followed by 2 min incubation on ice. SOC media (950 µL) was added to the cells, and the resulting suspension incubated at 37°C, 300 rpm for 1 h. after this time, 250 µL of the cultures were plated onto ampicillin LB agar plates and incubated at 37°C overnight. One colony from both plates was picked and used to inoculate starter cultures in TB media (160 mL, supplemented with ampicillin to final concentration of 100 µg/mL), which were incubated in a shaking incubator at 37°C, 180 rpm for 18 h. 20 mL of each starter culture were used to inoculate TB media (1 L supplemented with

ampicillin at 100 µg/mL), and incubated at 37°C, 180 rpm until an OD600 of 0.8. Protein expression was then induced with addition of IPTG (100 µM final concentration). Cultures were incubated for 24 h at 18°C, 240 rpm. The cultures were centrifuged at 4°C, 4000 rpm for 30 min. Resulting pellets were collected and stored at -80°C.

His-USP5^{WT} purification

The pellets were resuspended in lysis buffer (50 mM HEPES pH 7,5, 500 mM NaCl, 0.5 mM AEBSF, 5 nM leupeptidin hemisulfate) and sonicated on ice with an ultrasonic processor for 10 min (10 sec on/ 15 sec off, 65% amplitude). The lysates were centrifuged at 4°C, 20000 rpm for 35 min. The supernatants were collected and filtered through a 0,22 µm filter. The lysates were first purified by affinity chromatography on Ni-NTA agarose 5ml prepacked column using a ÄKTA go™ system. Column was washed with 5 column volumes of Buffer A (50 mM HEPES, pH 7,5, 500 mM NaCl, 5 mM imidazole) and protein eluted by imidazole gradient elution (15%-45% buffer A/buffer B), where buffer A (50 mM HEPES, pH 7,5, 500 mM NaCl, 5 mM imidazole) and buffer B (buffer 50 mM HEPES, pH 7,5 ,500 mM NaCl, 500 mM imidazole) over 10 column volumes. Elution was collected in fractions and analysed by SDS-PAGE gel. Fractions containing desired protein were combined and concentrated to a volume of 3 mL using Amicon®Ultra Centrifugal Filter. The solutions were then further purified by size exclusion chromatography on ÄKTA go™ system using a HiLoad 16/600 Superdex 200 pg column (Cytiva) and eluted with buffer (50 mM HEPES, 500 mM NaCl, 5% glycerol, pH 7.5, 1 mM DTT). Collected fractions were analysed by SDS-PAGE, concentrated using Amicon®Ultra Centrifugal Filter, aliquoted and stored at -80°C.

Cleavage assay gel

Solutions of USP5^{WT} and USP5^{C335A} at 2 µM were prepared in buffer (50 mM Tris pH 7,5 1 mM DTT), mixed in a 1:1 ratio with di-ubiquitin and

SDE2, at final concentration of 3 µM. The resulting mixtures were incubated for 30 min at 37°C. The samples were quenched with 4×LDS/DTT, boiled at 95°C for 5 min and loaded into a 4–12% Bis-Tris Nupage™ Midi gel. The SDS-PAGE was run with MOPS buffer at 150 V for 70 min. The gel was then stained with Coomassie blue and imaged on ChemiDoc (Biorad).

Mass spectrometry

Samples (38 µg of purified protein in 25 µL) were reduced with 5 mM TCEP solution at 55°C for 15 min, followed by alkylation with iodoacetamide, final concentration of 40 mM, for 30 min at r. t., and acidification (final concentration 2.5% phosphoric acid). 165 µL of binding buffer was added to the samples (100 mM TEAB in 90% methanol) and the resulting solutions were loaded onto S-trap micro columns (ProtiFi) and centrifuged at 4,000 g for 30 sec. The columns were washed 3 times with 150 µL of binding buffer and centrifuged at 4,000 g for 30 sec between each washing step. To fully remove residues of the binding buffer, the columns were centrifuged at 4000 g for 1 min. The proteins were digested with 20 µL of digestion buffer (50 mM TEAB, 1 µg of trypsin) at 47°C for 2 h. The resulting peptides were eluted with 40 µL of buffer 1 (50 mM TEAB), 40 µL of buffer 2 (50 mM TEAB, 0,2% formic acid), and 40 µL of buffer 3 (50% MeCN in water), each elution step was carried out by centrifuging at 4000 g for 1 min. The elutions were combined and dried in a SpeedVac at 45°C for 2 h. The samples were analysed by the Mass Spectrometry facility of the MRC PPU of the University of Dundee.

Results

Design and synthesis of the peptide library

The investigation started with the generation of an originator peptide spanning residues **66–90** of SDE2. This peptide contains the conserved cleavage motif (GGKGG) flanked by 10 amino acid residues on either side. The canonical cleavage happens between residues G77 and K78, therefore, this peptide was selected as resembles the wild-type SDE2 sequence and an ideal starting point for establishing an in vitro cleavage assay. The peptide was synthesised using solid phase peptide chemistry as described in the Methods section.

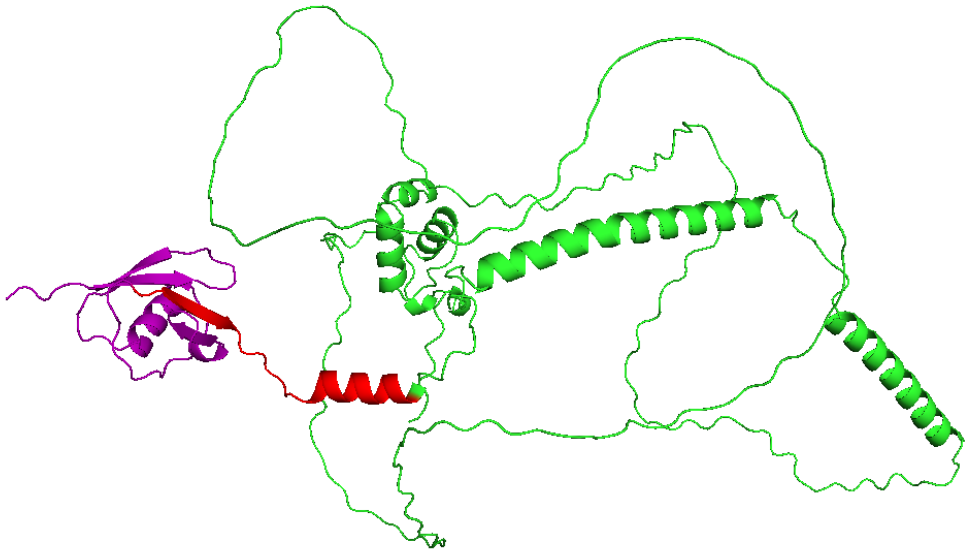


Figure 7: AlphaFold-predicted 3D structure of SDE2. The UBL domain is shown in purple, residues 66–90 (corresponding to the sequence used for peptide design) are highlighted in red, and the C-terminal domain is depicted in green. The AlphaFold prediction of the the C-terminal domain shows a highly disordered structure, however, there is limited experimental evidence to confirm the 3D structure.

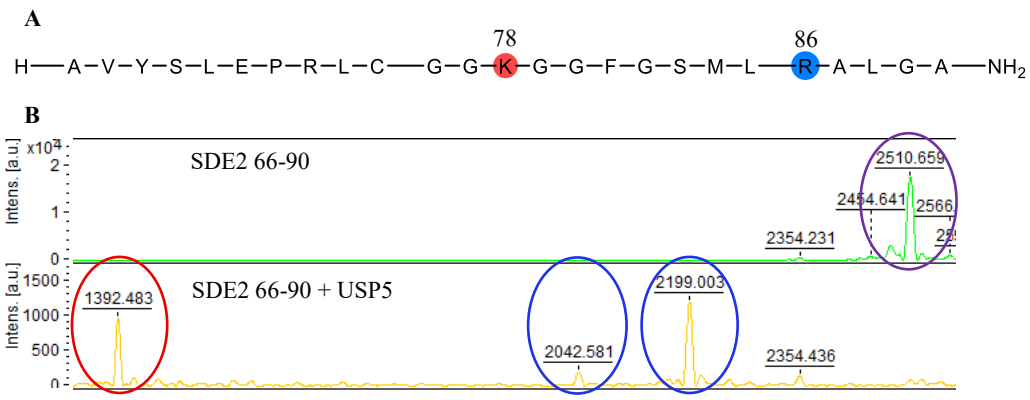


Figure 8: A. Sequence of the originator peptide (SDE2 66–90, MW 2510 Da). The two cleavage sites are highlighted: lysine 78 in red and arginine 86 in blue. **B**. MALDI-TOF spectrum of the cleavage assay, performed as described in the Methods section. The upper panel shows the intact full-length peptide (SDE2 66–90), with the corresponding peak circled in purple. In the lower panel, the full-length peptide peak is absent due to processing by USP5 during the 1 h incubation. The resulting cleavage products are circled, with fragments corresponding to cleavage at K78 shown in red and those corresponding to cleavage at R86 shown in blue.

As next step, a series of *in vitro* cleavage assays were performed to assess the processing of the synthetic peptide, as described in the Methods chapter. The reactions were carried out in Tris buffer adjusted to pH 7.5, in presence of dithiothreitol (DTT) to maintain the catalytic cysteine in its reduced, active state by preventing oxidation. Diluted peptides in assay buffer were incubated with recombinantly expressed USP5, respective final concentration of 20 µL and 2 µL, at 37 °C for 1 h to allow for the cleavage reaction to occur. The reaction was subsequently quenched by the addition of 2 µL of 12% TFA, ensuring immediate termination of enzymatic activity. For the initial assay (Figure 8), di-ubiquitin served as the experimental control, as it is a well-characterized substrate for USP5 that reliably reports enzymatic activity. In subsequent assays, the originator peptide, SDE2 66-90, was used as control to ensure consistency across experiments. The samples were subsequently analysed by MALDI-TOF for evidence of processing.

Surprisingly, the MALDI revealed peaks at mass1392 Da, consistent with a processing at position K78-G79, which is incongruent with the processing of the full-length protein, which is cleaved at positions G77-K78. Furthermore, two additional less intense peaks were present compatible with a cleavage at sites R86-A87 and, to a lesser extent, L85-R86. This surprising deviation from the full-length SDE2 processing pattern suggests that additional sequence or structural determinants might influence substrate recognition and processing, prompting further investigation. Consequently, a series of mutated peptides was designed

and synthesized, targeting the residues at which cleavage occurred as well as the GG motif essential for full-length SDE2 processing [17].

Lysine 78 mutations

In order to investigate the contribution of lysine residues to peptide processing, a panel of peptides comprising residues **66–90** of SDE2 that included point mutations at K78 was generated. Lysine-to-alanine (K→A) substitution was introduced to investigate the specific role of the functionalised side chain. Lysine-to-serine (K→S) mutation was included as this would maintain polarity while removing the positive charge by introducing a small, uncharged hydroxyl-containing residue. Lysine-to-arginine (K→R) substitutions preserved a positively charged side chain at pH 7.5 while incorporating a bulkier guanidinium group. The introduction of a negatively charged residue, enabled by lysine-to-glutamic acid (K→E) mutations, allowed the assessment of charge reversal. Lysine-to-phenylalanine (K→F) substitutions were used to replace the charged side chain with a bulky, apolar aromatic group, in order to examine the effects of hydrophobicity and steric bulk. Lysine-to-proline (K→P) mutations incorporated a rigid cyclic residue known to restrict backbone flexibility and alter local peptide conformation. Moreover, double point mutants at both K78 and R86 sites were also generated to investigate potential interactions between the two sites, including simultaneous substitution of K78 and R86 to alanine, as well as a reciprocal swap (K78R and R86K).

GG motif mutations

To determine the function of the G76G77 and G79G80motifs, which is conserved across UBL and UBL-like domains [12], and it is known to play a key role in the recognition and processing of substrates by USP5, the Gs were replaced with A. Alanine substitutions were introduced as single point mutations to assess the contribution of each residue individually. Furthermore, double mutants were generated, targeting either G76 and

G77 or G79 and G80, as well as a peptide where all four Gs were mutated, in order to evaluate the combined effect of disrupting the motif.

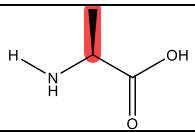
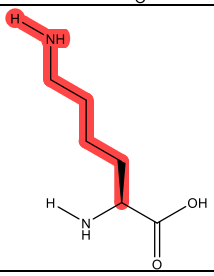
Arginine 86 mutations

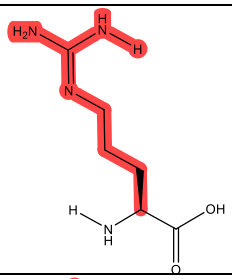
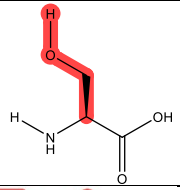
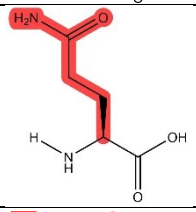
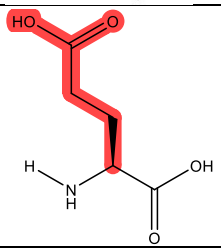
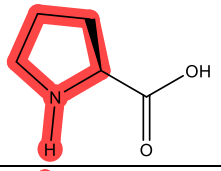
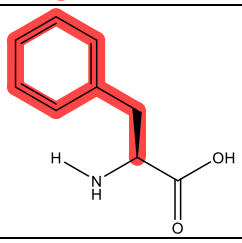
The same mutational strategy applied to residue K78 was used to examine the additional cleavage site, targeting the R86 residue.

Replacement with alanine (R→A) provided a minimalistic side chain, enabling evaluation of how strongly the native guanidinium group influences processing. The serine (R→S) and glutamine (R→Q) variants were generated to retain a polar side chain while removing the positive charge, with serine offering a compact hydroxyl-containing group and glutamine providing an uncharged amide with a length more comparable to the native residue. Substitution with lysine (R→K) preserved a positively charged residue but introduced a shorter aliphatic side chain. Incorporation of phenylalanine (R→F) allowed assessment of the effect of replacing the charged side chain with a bulky, hydrophobic aromatic moiety, whereas mutation to glutamic acid (R→E) inverted the local charge environment.

The chemical structures of the amino acids used for substitutions are shown in Table 1.

Table 1: Summary of the amino acid mutations introduced to investigate the contribution of individual residues to SDE2 cleavage by USP5. Side chain structures are provided to illustrate the chemical changes introduced by each substitution.

Amino acid	Structure	Notes
Ala (A)		Minimal side chain, used for K78, R86 and GG motif
Lys (K)		Positively charged, used for R86K

Arg (R)		Positively charged, bulky, used for K78R, R86K
Ser (S)		Polar uncharged, used for K78 and R86
Gln (Q)		Polar uncharged, used for R86
Glu (E)		Negatively charged, used for K78 and R86
Pro (P)		Conformationally rigid, used for K78
Phe (F)		Bulky hydrophobic, used for K78 and R86

Synthesis strategy and optimisation

To reduce synthesis time and minimize the use of organic solvents, a parent 10-residue peptide comprising SDE2 residues 81-90, which was common among the whole library of peptides, was synthesised on a 0.1 mmol scale, omitting the final Fmoc deprotection step. The resin was then collected, dried, and divided into 10 batches, which were subsequently used for the synthesis of the K- and G-substituted mutant peptides.

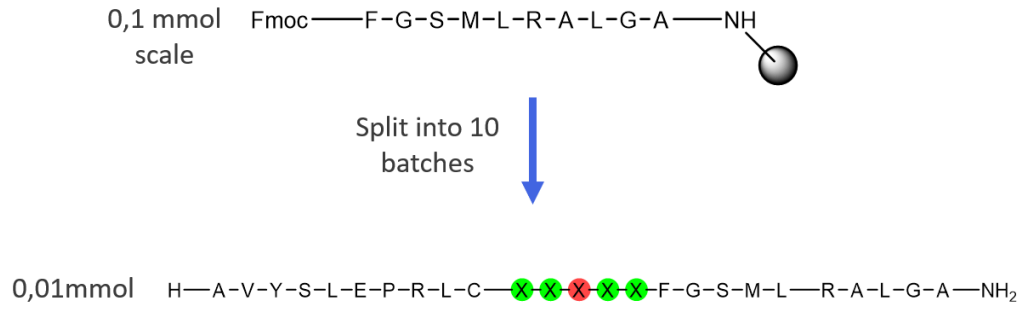


Figure 9: Schematic representation of the peptide synthesis strategy used to optimize time and solvent consumption. The parent 10-residue peptide (SDE2 81-90) was synthesized on a 0.1 mmol scale with the final Fmoc deprotection step omitted. The resin-bound peptide was then collected, dried, and divided into 10 smaller batches (0.01 mmol scale), each used for the synthesis of peptides containing K78 and G mutations at specific positions (highlighted in green and red).

As illustrated in the Methods chapter, peptides were synthesised using Fmoc-based solid-phase peptide synthesis on Rink amide resin with an automated Liberty Blue™ 2.0 synthesiser. The N-terminal Fmoc group served as a protective group throughout the assembly, and its removal was monitored by UV absorbance to ensure efficient deprotection at each step, as shown in Figure 11, using peptide SDE2 66-90 G77A as an example. After chain assembly, peptides were cleaved from the resin using a trifluoroacetic acid-based cleavage cocktail, precipitated in cold diethyl ether, and lyophilised. The identity and purity of the final peptides were confirmed by LC-MS analysis, as shown in Figures 12 and 13 for the peptide SDE2 66-90 G77A.

The synthesis of the peptide panel presented several challenges due to both sequence composition and peptide length.

Arginine and leucine residues, as well as amino acids immediately following them, often exhibited reduced coupling efficiency, necessitating the use of double-coupling steps to ensure complete incorporation. In addition, the peptides consisted in 25 residues, which made solid-phase synthesis more challenging because longer chains can impose steric strain on the resin and reduce accessibility during coupling. To address this, two Rink amide resins were evaluated: a 100–200 mesh resin with a loading capacity of 0.35 mmol/g and a 90-mesh resin with a lower loading capacity of 0.18 mmol/g. The lower-loading 90 mesh resin produced superior results, as coupling efficiency began to decline after approximately 15 residues when using the higher-loading resin. These optimizations were essential to obtain peptides of sufficient purity and yield for the subsequent cleavage assays.

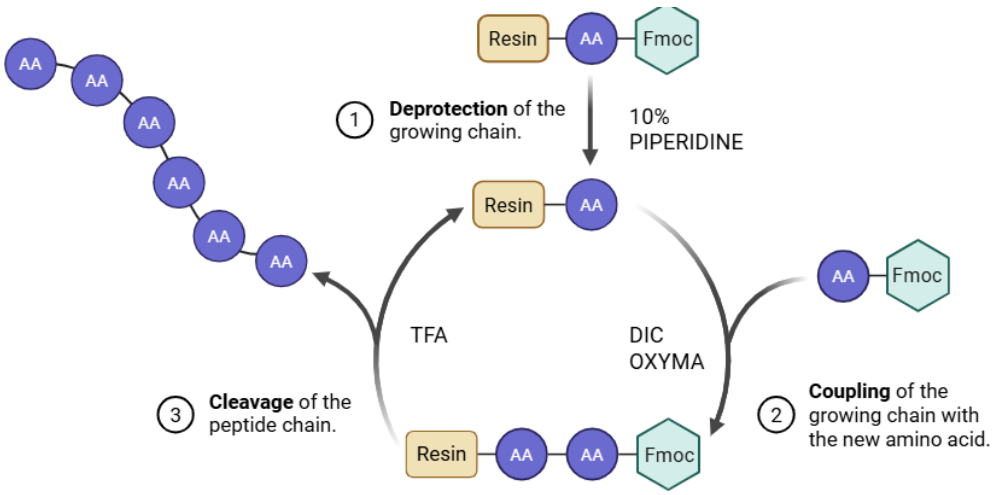


Figure 10: Scheme illustrating the solid-phase peptide synthesis (SPPS) cycle using Fmoc chemistry. (1) Deprotection of the growing peptide chain with 10% piperidine removes the Fmoc group. (2) Coupling of the next Fmoc-protected amino acid is facilitated by DIC and Oxyma, extending the peptide chain on the resin. (3) Final cleavage of the fully assembled peptide from the resin is achieved using TFA.

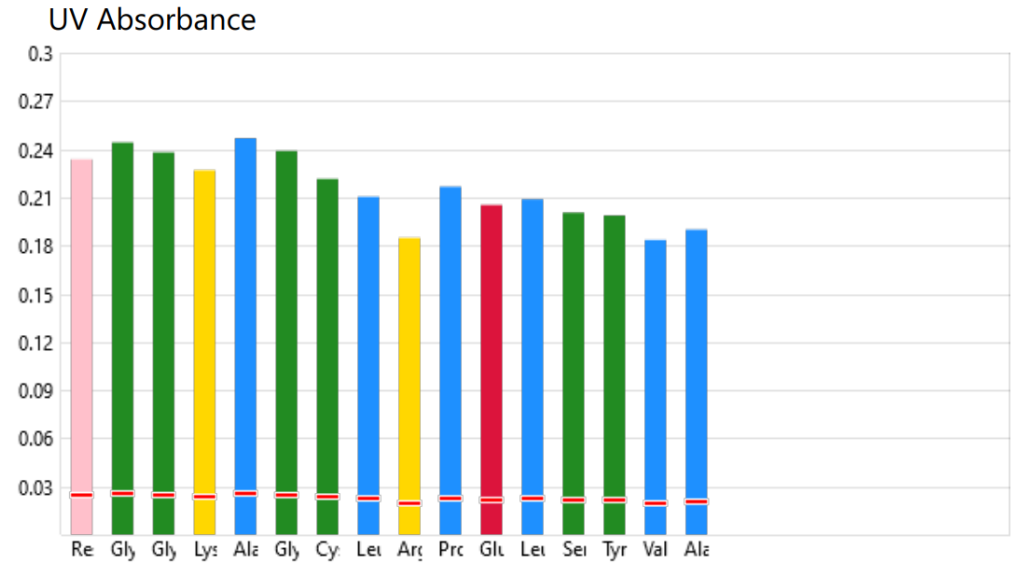


Figure 11: UV absorbance spectrum of the peptide SDE2 66-90 G77A, synthesized by continuing the peptide elongation from the parent resin-bound peptide SDE2 81-90.

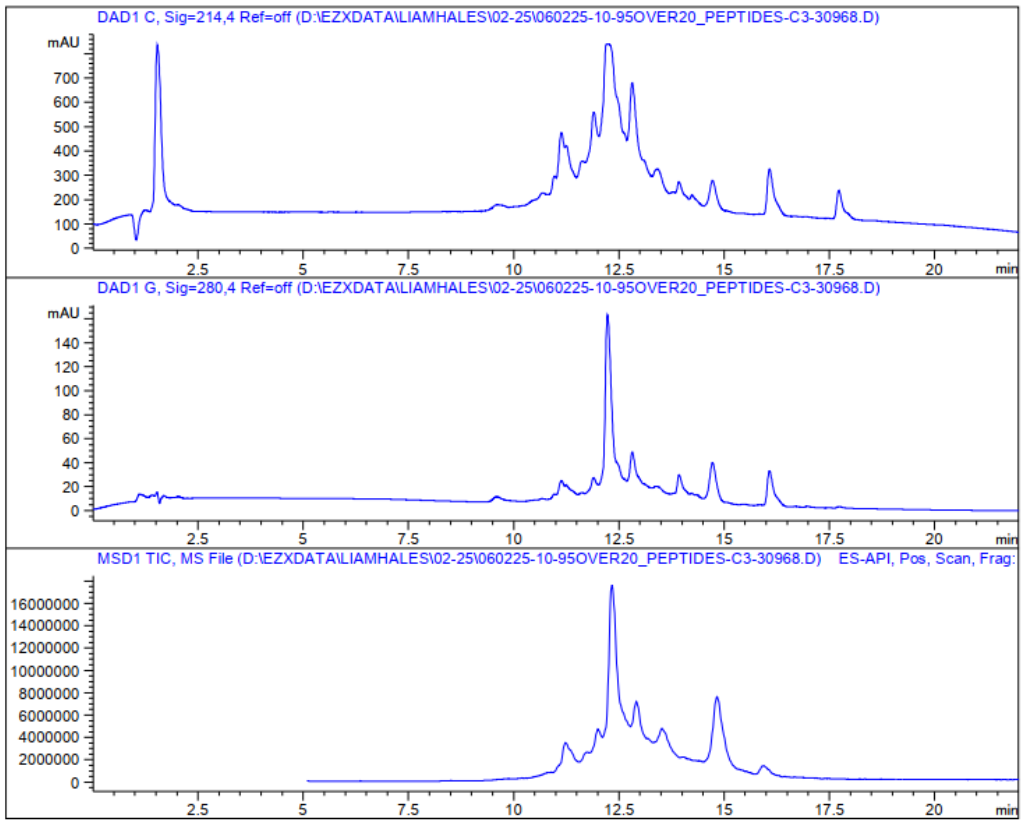


Figure 12: LC-MS analysis of the peptide SDE2 66-90 G77A. The top chromatogram shows the UV absorbance at 214 nm, which primarily detects the peptide bond backbone. The middle chromatogram displays the UV absorbance at 280 nm, highlighting the presence of aromatic residues, phenylalanine and tyrosine. The bottom trace represents the total ion count (TIC) from mass spectrometry, confirming the peptide's elution profile and ionization efficiency.

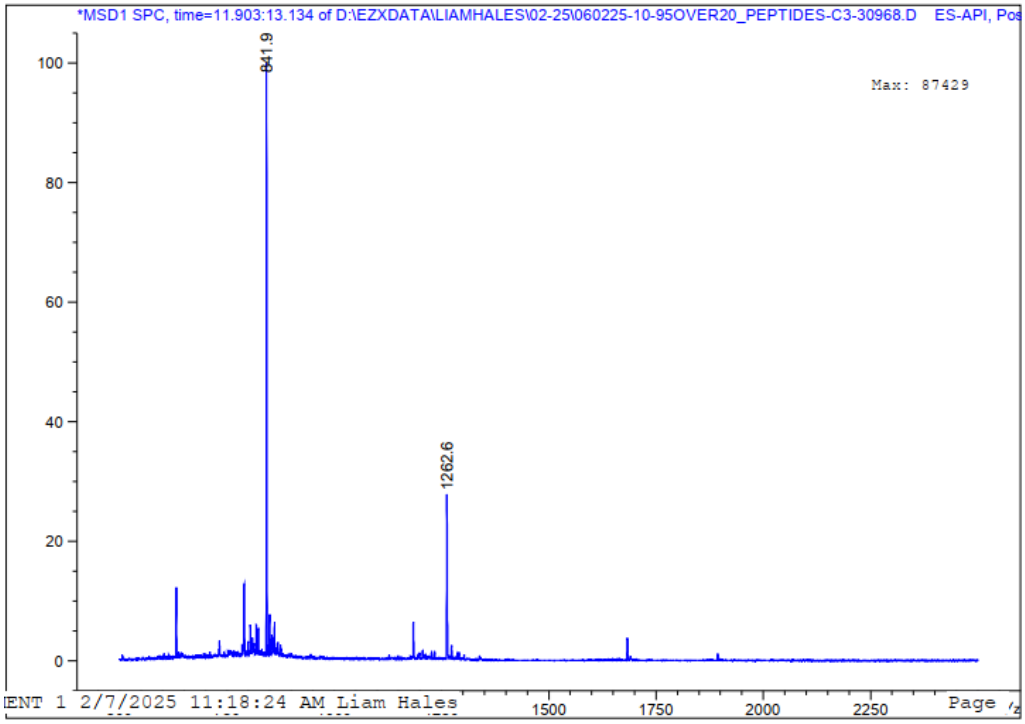


Figure 13: LC-MS mass spectrum of the SDE2(66–90) G77A peptide acquired at a retention time of 11.903 min. The two dominant peaks correspond to the multiply charged molecular ions, with the left peak representing the 3+ charge state (m/z=841,9) and the right peak representing the 2+ charge state (m/z=1262,6).

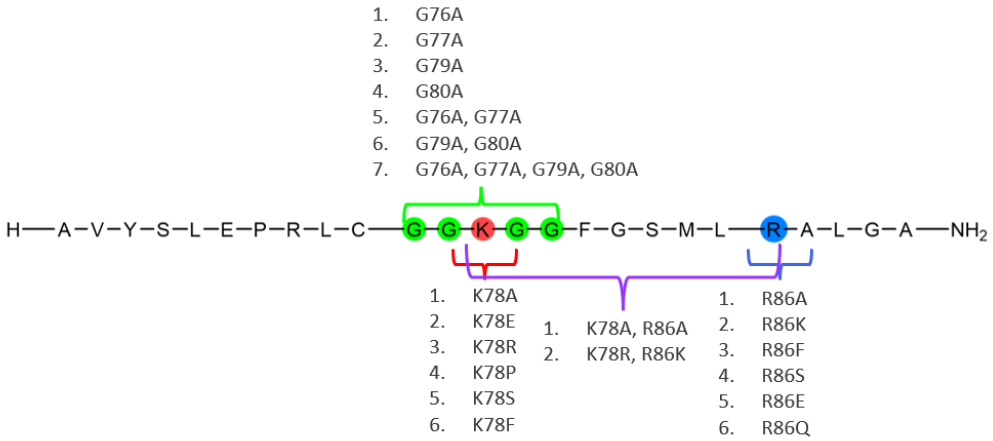


Figure 14: Schematic representation of the originator peptide sequence and the positions targeted for mutagenesis. The peptide sequence (SDE2 66-90) is shown with individual amino acids labelled. Residues G76–G80 (highlighted in green), K78 (highlighted in red), and R86 (highlighted in blue) were selected for systematic substitution. The lists indicate the specific single, double, and quadruple mutations introduced at each site, including glycine-to-alanine variants (G76A–G80A), lysine substitutions at position 78 (K78A/E/R/P/S/F), and arginine substitutions at position 86 (R86A/K/F/S/E/Q). These variants were generated to assess the structural and functional contribution of each residue to peptide behaviour.

Across all synthesized variants, the average peptide yield was around 65%, reflecting an overall efficient and consistent production process. Together, these optimizations ensured that the peptide library (Figure 14) was generated with sufficient yield for all downstream analyses. With the library successfully obtained, the following chapter focuses on assessing their processing through a series of cleavage assays.

Cleavage assay results

The cleavage assays were carried out as previously described for the first peptide, SDE2 66-90, that was used here as positive control.

The quantification of the cleavage was done by integrating the area under the curve (AUC) of the MALDI traces, which reflects the relative abundance of each ion species in the sample. The AUC of the full-length peptide was compared to the AUCs of its fragment ions to estimate the percentage of peptide that was processed and to identify the preferred cleavage sites.

K78 mutated peptides

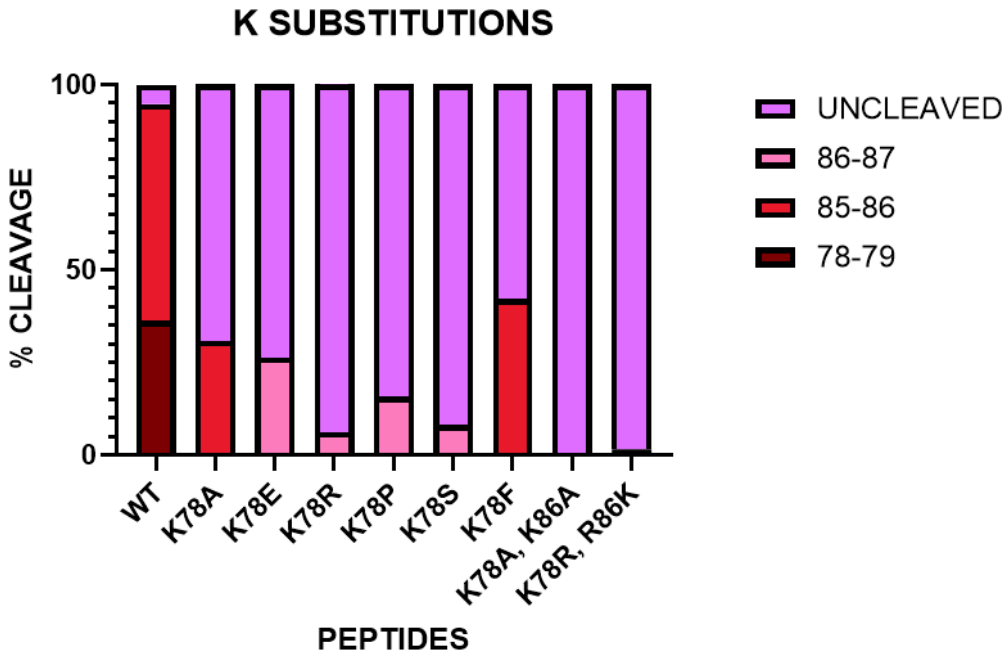


Figure 15: Effects of K78 substitutions on peptide cleavage patterns. Stacked bar plots show the distribution of cleavage products for each peptide variant containing substitutions at position K78, alone or in combination with R86 mutations. Cleavage at positions 78–79, 85–86, and 86–87 is represented by dark red, red, and pink bars, respectively, while uncleaved peptide is shown in purple. Compared to the wild-type peptide, most K78 variants exhibit markedly reduced processing, with several substitutions resulting in near-complete resistance to cleavage.

The cleavage assay revealed that the wild-type peptide was processed at both the 78–79 and 85–86 positions, with only a small fraction remaining uncleaved. Substitution of K78 with alanine (K78A) completely abolished cleavage at the 78–79 site and markedly reduced overall processing,

indicating that the lysine side chain is required for efficient recognition. The charge-reversal mutation K78E and the conservative substitution K78R similarly resulted in a strong reduction in cleavage, with most of the peptide remaining intact. Mutations to bulkier or conformationally restrictive residues (K78P and K78F) also impaired cleavage, although K78F retained a detectable fraction of processing at position 85–86. The polar, uncharged substitution K78S produced only minimal cleavage, further supporting the importance of a positively charged side chain at this position. Double mutants involving both K78 and R86 (K78A/K86A and K78R/R86K) were almost completely resistant to cleavage, indicating that disruption of both basic residues abrogates all detectable processing. Overall, these results show that cleavage at residue 78 is highly sensitive to substitutions that alter charge, size, or conformation of the lysine side chain, and that the presence of either lysine or arginine is essential for efficient processing.

G mutated peptides

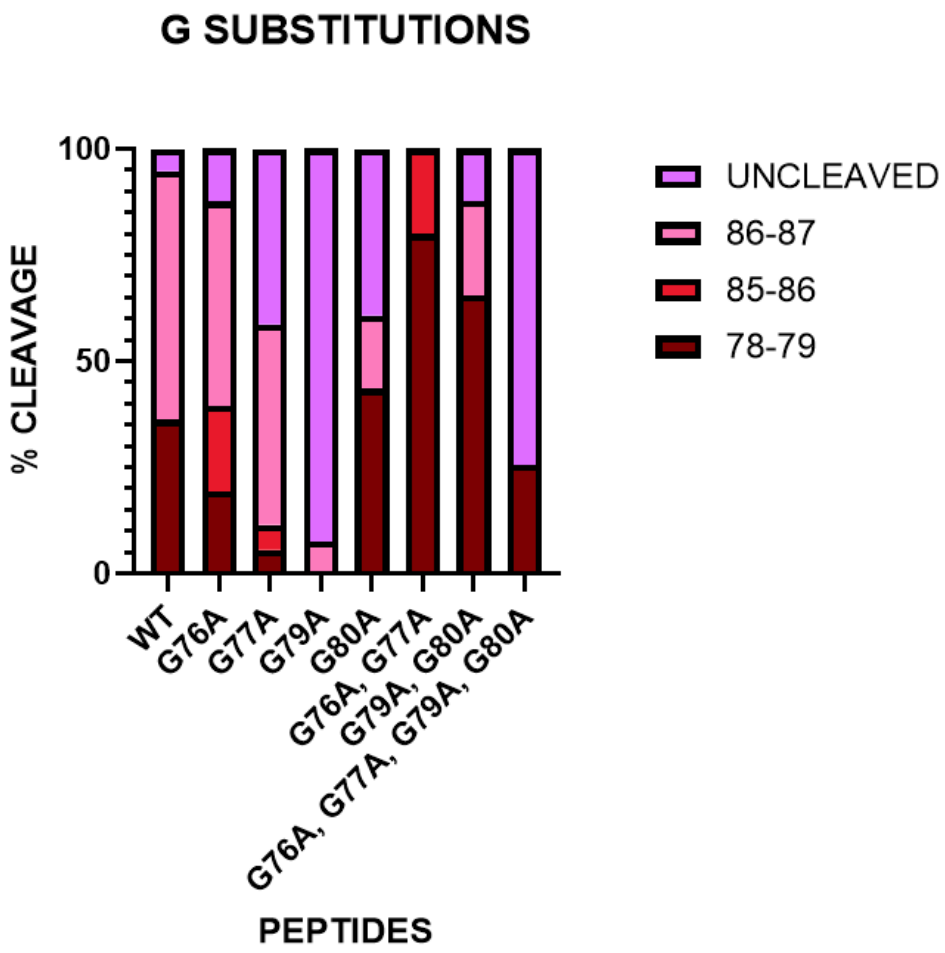


Figure 16: Effects of glycine (G76–G80) substitutions on peptide cleavage patterns. Stacked bar plots show the proportion of cleavage products generated from peptides carrying single, double or quadruple glycine-to-alanine substitutions within the G76–G80 motif. Cleavage at positions 78–79, 85–86, and 86–87 is represented in dark red, red, and pink, respectively, while uncleaved peptide is shown in purple. Compared with the wild-type peptide, glycine substitutions alter the distribution of cleavage events in a residue-dependent manner, with multi-site substitutions generally shifting processing toward the 78–79 site and reducing overall susceptibility to cleavage at downstream positions.

The glycine-to-alanine substitutions within the GG motif produced marked changes in the cleavage pattern compared with the wild-type peptide. While the WT substrate was cleaved at both 78–79 and 85–86, single alanine substitutions affected cleavage efficiency and site preference to varying degrees. Mutating G76 or G77 shifted cleavage predominantly toward the 78–79 site and reduced overall processing at 85–86, whereas substitutions at G79 or G80 strongly impaired cleavage, with most of the peptide remaining uncleaved. Double mutants further

amplified these effects: variants combining G76A/G77A or G79A/G80A displayed near-complete loss of cleavage at both canonical sites and showed an increased proportion of uncleaved peptide. The quadruple mutant (G76A/G77A/G79A/G80A) was almost entirely resistant to processing. Together, these results indicate that the integrity of the GG motif is critical for efficient USP5-mediated recognition, and even minor perturbations within this highly flexible region drastically reduce cleavage efficiency.

R86 mutated peptides

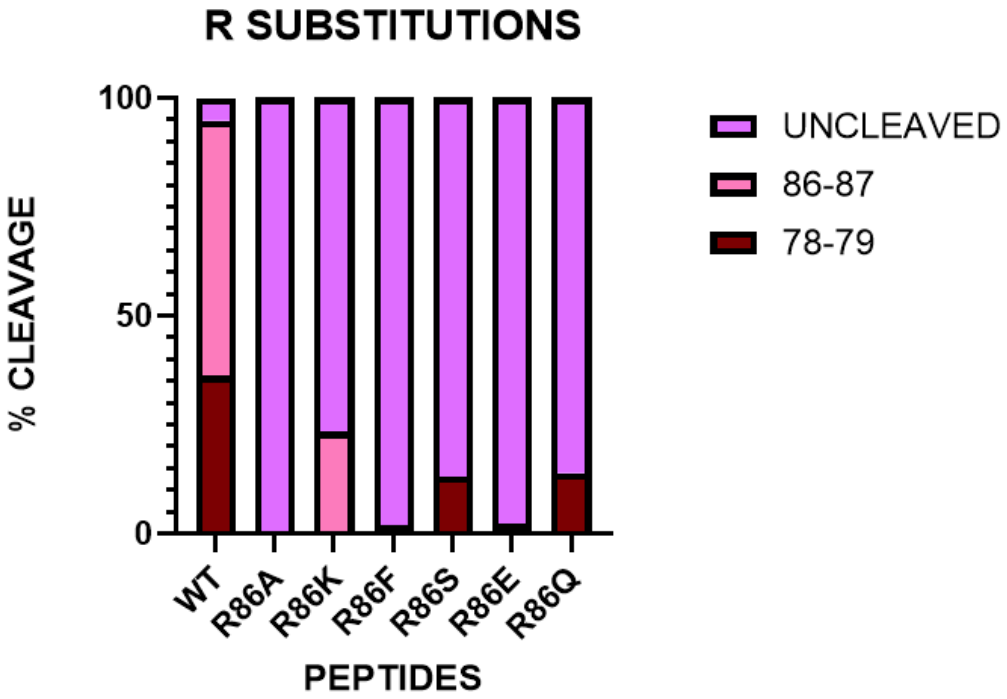


Figure 17: Effects of R86 substitutions on peptide cleavage patterns. Stacked bar plots show the distribution of cleavage products generated from peptides containing single amino-acid substitutions at position R86. Cleavage at positions 78–79, 85–86, and 86–87 is shown in dark red, red, and pink, respectively, while uncleaved peptide is shown in purple. Relative to the wild-type peptide, all substitutions at R86 markedly reduce downstream cleavage efficiency and shift processing toward the 78–79 site. Most variants exhibit near-complete loss of cleavage at 85–86 and 86–87, indicating that the native arginine at this position is essential for efficient recognition and processing at these sites.

Substitutions at R86 had a strong inhibitory effect on peptide processing. While the wild-type peptide was efficiently cleaved at both 78–79 and 86–87, replacing R86 with alanine (R86A), phenylalanine (R86F), serine (R86S), glutamic acid (R86E), or glutamine (R86Q) abolished cleavage at

the 86–87 site and resulted in almost complete loss of processing, with the majority of each variant remaining uncleaved. The conservative substitution R86K, which maintains a positive charge, produced only minimal cleavage at 86–87, indicating that the guanidinium group of arginine is specifically required for efficient processing at this position. These results demonstrate that R86 is essential for recognition and that both charge and side-chain geometry at this position are critical for supporting catalysis.

Processing of the SDE2-derived peptide is highly dependent on the basic residues K78 and R86, as substitutions that alter their charge, size, or geometry almost completely abolish cleavage at their respective sites. Integrity of the adjacent GG motif is equally critical, as even single glycine-to-alanine substitutions markedly reduce cleavage efficiency and shift site preference, while double or quadruple mutations nearly eliminate processing. Overall, efficient recognition by USP5 requires both the presence of lysine/arginine at key positions and the structural flexibility provided by the conserved GG motif.

Investigation of the Loss of Peptide-Cleavage Activity

A new batch of recombinant enzyme was expressed and purified following the established protocol described in the Methods section.

Upon re-evaluating the peptide panel with this freshly prepared enzyme, no detectable processing of any peptide variant was observed, which stood in clear contrast to the cleavage profiles obtained with the previous enzyme preparation. The absence of cleavage could reflect loss of enzymatic activity. To address this, a cleavage assay (Figure 18) was performed using two well-characterised substrates, full-length SDE2 and di-ubiquitin, both of which are reliably processed by the wild-type enzyme. In these control assay, the wild-type enzyme efficiently cleaved both substrates, demonstrating that the new batch retained full catalytic

activity. As expected, the catalytically inactive mutant enzyme, where the catalytic cysteine is mutated to an alanine ($USP5^{C335A}$), did not cleave either SDE2 or di-ubiquitin, thereby providing a negative control and confirming the validity of the assay conditions. Together, these results demonstrate that the recombinant enzyme was fully active, indicating that the absence of peptide processing in the new assays reflects a change in the experimental outcome rather than enzyme inactivation.

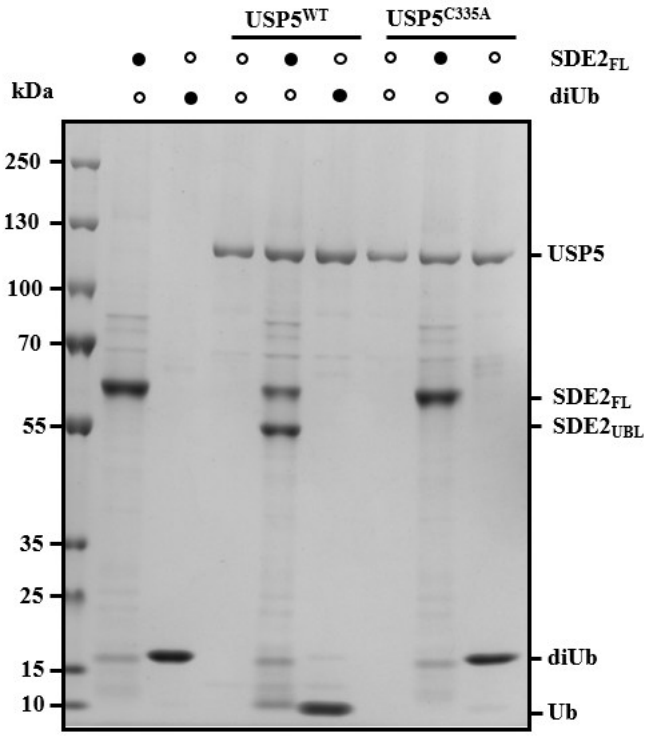


Figure 18: Assessment of enzymatic activity of newly prepared USP5 using established protein substrates. A freshly expressed and purified batch of recombinant USP5 was evaluated for catalytic activity following the unexpected absence of peptide cleavage in re-screening assays. SDS–PAGE analysis shows reactions containing wild-type USP5 ($USP5^{WT}$) or the catalytic mutant ($USP5^{C335A}$) incubated with either full-length SDE2 ($SDE2_{FL}$) or di-ubiquitin (diUb). $USP5^{WT}$ efficiently cleaved both $SDE2_{FL}$ and diUb, generating the expected $SDE2_{UBL}$ and mono-Ub products, confirming that the new enzyme preparation retained full catalytic activity. In contrast, $USP5^{C335A}$ showed no detectable processing of either substrate, validating the assay as a negative control.

Comparative proteomics of USP5 preparations

The unexpected absence of peptide processing by the newly purified USP5 enzyme raised a critical question: was USP5 truly responsible for the cleavage activity observed in the initial assays, or could it have been influenced by a contaminant present in the first enzyme preparation?

To address this possibility, a comparative proteomic analysis of the two enzyme batches was undertaken. Both preparations were subjected to tryptic digestion, and the resulting peptides were captured and purified using S-Trap columns, according to the protocol described in the Methods section. Subsequent mass spectrometry analysis revealed a notable difference between the batches: the original preparation contained 14 proteins derived from *Escherichia coli* that were completely absent in the freshly prepared enzyme. To identify potential contributors to the observed peptide cleavage, a focused literature review was conducted on these contaminating proteins. Among them, Oligopeptidase B emerged as a strong candidate, as it is a serine protease known to cleave peptides shorter than 30 residues, with a preference for lysine (K) and arginine (R) residues[36]. The substrate specificity of Oligopeptidase B aligns closely with the cleavage patterns previously observed in the peptide panel, suggesting that the activity initially attributed to USP5 may have been due to this contaminant.

Conclusions

This work aimed to define the sequence determinants required for SDE2 processing by USP5, in order to better understand the mechanism of SDE2 maturation and support the design of USP5 inhibitors or modulators. An initial synthetic 25-amino-acid peptide, derived from the region surrounding the canonical K78 cleavage site of SDE2, was studied using a recombinant *in vitro* cleavage assay monitored by MALDI-TOF, revealing a shift in the expected processing site as well as two additional cleavage events not observed in the full-length protein. To investigate the basis of these unprecedented cleavage events, a library of peptide variants was generated via solid-phase synthesis, targeting the canonical cleavage site K78, the surrounding GG motif and R86 (the additional site detected in the initial assay), and analysed using the same experimental approach. Mutational analysis demonstrated that efficient processing requires both key basic residues, particularly K78 and R86, and the structural flexibility provided by the conserved GG motif, as alterations to charge or local flexibility substantially reduced or abolished cleavage.

Further analysis revealed that the unexpected cleavage sites, and the processing activity observed in the assay, were not due to USP5 itself, but resulted from contamination of the initial recombinant enzyme preparation by *E. coli* Oligopeptidase B. This finding highlights a major limitation of *in vitro* assays relying on recombinant proteins: their susceptibility to trace protease impurities introduced during bacterial expression and purification. To mitigate such issues, stringent controls, such as testing multiple protein batches, including appropriate negative controls and validating enzyme purity by mass spectrometry, are essential to ensure that observed substrate processing reflects the true specificity of the enzyme under study.

Figure index

Bibliography

[1] G. E. Billman, 'Homeostasis: The Underappreciated and Far Too Often Ignored Central Organizing Principle of Physiology', Mar. 10, 2020, *Frontiers Media S.A.* doi: 10.3389/fphys.2020.00200.

[2] L. Galluzzi, T. Yamazaki, and G. Kroemer, 'Linking cellular stress responses to systemic homeostasis', Nov. 01, 2018, *Nature Publishing Group*. doi: 10.1038/s41580-018-0068-0.

[3] X. Sheng, Z. Xia, H. Yang, and R. Hu, 'The ubiquitin codes in cellular stress responses', Mar. 01, 2024, *Oxford University Press*. doi: 10.1093/procel/pwad045.

[4] A. S. Venne, L. Kollipara, and R. P. Zahedi, 'The next level of complexity: Crosstalk of posttranslational modifications', Mar. 01, 2014, *Wiley-VCH Verlag*. doi: 10.1002/pmic.201300344.

[5] A. M. Burroughs, S. Balaji, L. M. Iyer, and L. Aravind, 'Small but versatile: The extraordinary functional and structural diversity of the β-grasp fold', *Biol Direct*, vol. 2, Jul. 2007, doi: 10.1186/1745-6150-2-18.

[6] L. Buetow and D. T. Huang, 'Structural insights into the catalysis and regulation of E3 ubiquitin ligases', Oct. 01, 2016, *Nature Publishing Group*. doi: 10.1038/nrm.2016.91.

[7] D. Komander and M. Rape, 'The ubiquitin code', *Annu Rev Biochem*, vol. 81, pp. 203–229, Jul. 2012, doi: 10.1146/annurev-biochem-060310-170328.

[8] M. A. Mansour, 'Ubiquitination: Friend and foe in cancer', Aug. 01, 2018, *Elsevier Ltd*. doi: 10.1016/j.biocel.2018.06.001.

[9] Y. Liao, I. Sumara, and E. Pangou, 'Non-proteolytic ubiquitylation in cellular signaling and human disease', Dec. 01, 2022, *Nature Research*. doi: 10.1038/s42003-022-03060-1.

[10] R. B. Damgaard, 'The ubiquitin system: from cell signalling to disease biology and new therapeutic opportunities', Feb. 01, 2021, *Springer Nature*. doi: 10.1038/s41418-020-00703-w.

[11] D. Popovic, D. Vucic, and I. Dikic, 'Ubiquitination in disease pathogenesis and treatment', Nov. 01, 2014, *Nature Publishing Group*. doi: 10.1038/nm.3739.

[12] L. Cappadocia and C. D. Lima, ‘Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism’, Feb. 14, 2018, *American Chemical Society*. doi: 10.1021/acs.chemrev.6b00737.

[13] K. Baek *et al.*, ‘NEDD8 nucleates a multivalent cullin–RING–UBE2D ubiquitin ligation assembly’, *Nature*, vol. 578, no. 7795, pp. 461–466, Feb. 2020, doi: 10.1038/s41586-020-2000-y.

[14] J. Keiten-Schmitz, K. Schunck, and S. Müller, ‘SUMO Chains Rule on Chromatin Occupancy’, Jan. 10, 2020, *Frontiers Media S.A.* doi: 10.3389/fcell.2019.00343.

[15] J. van den Heuvel *et al.*, ‘Processing of the ribosomal ubiquitin-like fusion protein FUBI-eS30/FAU is required for 40S maturation and depends on USP36’, *Elife*, vol. 10, Jul. 2021, doi: 10.7554/eLife.70560.

[16] R. Sugioka-Sugiyama and T. Sugiyama, ‘Sde2: A novel nuclear protein essential for telomeric silencing and genomic stability in Schizosaccharomyces pombe’, *Biochem Biophys Res Commun*, vol. 406, no. 3, pp. 444–448, Mar. 2011, doi: 10.1016/j.bbrc.2011.02.068.

[17] P. Thakran, P. A. Pandit, S. Datta, K. K. Kolathur, J. A. Pleiss, and S. K. Mishra, ‘ Sde2 is an intron-specific pre- mRNA splicing regulator activated by ubiquitin-like processing ’, *EMBO J*, vol. 37, no. 1, pp. 89–101, Jan. 2018, doi: 10.15252/embj.201796751.

[18] S. Chanarat and S. K. Mishra, ‘Emerging Roles of Ubiquitin-like Proteins in Pre-mRNA Splicing’, Nov. 01, 2018, *Elsevier Ltd*. doi: 10.1016/j.tibs.2018.09.001.

[19] U. Jo, W. Cai, J. Wang, Y. Kwon, A. D. D’Andrea, and H. Kim, ‘PCNA-Dependent Cleavage and Degradation of SDE2 Regulates Response to Replication Stress’, *PLoS Genet*, vol. 12, no. 12, Dec. 2016, doi: 10.1371/journal.pgen.1006465.

[20] J. Floro *et al.*, ‘SDE2 is an essential gene required for ribosome biogenesis and the regulation of alternative splicing’, *Nucleic Acids Res*, vol. 49, no. 16, pp. 9424–9443, 2021, doi: 10.1093/nar/gkab647.

[21] A. Castello *et al.*, ‘Comprehensive Identification of RNA-Binding Domains in Human Cells’, *Mol Cell*, vol. 63, no. 4, pp. 696–710, Aug. 2016, doi: 10.1016/j.molcel.2016.06.029.

[22] ‘PII: S0968-0004(99)01537-6’, 2000. [Online]. Available: www.chem.qmw.ac.uk/iubmb/enzyme/ATPases.html.

[23] A. R. Leman, C. Noguchi, C. Y. Lee, and E. Noguchi, ‘Human Timeless and Tipin stabilize replication forks and facilitate sister-chromatid cohesion’, *J Cell Sci*, vol. 123, no. 5, pp. 660–670, Mar. 2010, doi: 10.1242/jcs.057984.

[24] J. Rageul *et al.*, ‘SDE2 integrates into the TIMELESS-TIPIN complex to protect stalled replication forks’, *Nat Commun*, vol. 11, no. 1, Dec. 2020, doi: 10.1038/s41467-020-19162-5.

[25] A. R. Lehmann *et al.*, ‘Translesion synthesis: Y-family polymerases and the polymerase switch’, *DNA Repair (Amst)*, vol. 6, no. 7, pp. 891–899, Jul. 2007, doi: 10.1016/j.dnarep.2007.02.003.

[26] A. A. Davies, D. Huttner, Y. Daigaku, S. Chen, and H. D. Ulrich, ‘Activation of Ubiquitin-Dependent DNA Damage Bypass Is Mediated by Replication Protein A’, *Mol Cell*, vol. 29, no. 5, pp. 625–636, Mar. 2008, doi: 10.1016/j.molcel.2007.12.016.

[27] A. Varshavsky, ‘The N-end rule pathway and regulation by proteolysis’, Aug. 2011. doi: 10.1002/pro.666.

[28] J. Rageul, J. J. Park, U. Jo, A. S. Weinheimer, T. T. M. Vu, and H. Kim, ‘Conditional degradation of SDE2 by the Arg/N-End rule pathway regulates stress response at replication forks’, *Nucleic Acids Res*, vol. 47, no. 8, pp. 3996–4010, May 2019, doi: 10.1093/nar/gkz054.

[29] W. S. Choi *et al.*, ‘Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases’, *Nat Struct Mol Biol*, vol. 17, no. 10, pp. 1175–1181, Oct. 2010, doi: 10.1038/nsmb.1907.

[30] T. Tasaki, A. Zakrzewska, D. D. Dudgeon, Y. Jiang, J. S. Lazo, and Y. T. Kwon, ‘The substrate recognition domains of the N-end rule pathway’, *Journal of Biological Chemistry*, vol. 284, no. 3, pp. 1884–1895, Jan. 2009, doi: 10.1074/jbc.M803641200.

[31] F. E. Reyes-Turcu, K. H. Ventii, and K. D. Wilkinson, ‘Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes’, 2009. doi: 10.1146/annurev.biochem.78.082307.091526.

[32] Y. Li and D. Reverter, ‘Molecular mechanisms of dubs regulation in signaling and disease’, Feb. 01, 2021, *MDPI AG*. doi: 10.3390/ijms22030986.

[33] J. A. Harrigan, X. Jacq, N. M. Martin, and S. P. Jackson, ‘Deubiquitylating enzymes and drug discovery: Emerging opportunities’, Jan. 01, 2018, *Nature Publishing Group*. doi: 10.1038/nrd.2017.152.

[34] L. T. Hales *et al.*, ‘A Proteolytic Switch: USP5 controls SDE2 function via UBL-directed cleavage’, May 24, 2025. doi: 10.1101/2025.05.23.655772.

[35] G. V. Avvakumov *et al.*, ‘Two ZnF-UBP domains in isopeptidase T (USP5)’, *Biochemistry*, vol. 51, no. 6, pp. 1188–1198, Feb. 2012, doi: 10.1021/bi200854q.

[36] T. H. T. Coetzer, J. P. D. Goldring, and L. E. J. Huson, ‘Oligopeptidase B: A processing peptidase involved in pathogenesis’, *Biochimie*, vol. 90, no. 2, pp. 336–344, Feb. 2008, doi: 10.1016/j.biochi.2007.10.011.