# "Understanding the dual specificity of UBA6"

# "Einblick in die duale Spezifität von UBA6"

## Doctoral thesis

for a doctoral degree at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,

## Section Biomedicine

submitted by

## Ngoc Truongvan

from

## Ha Nam, Viet Nam

Würzburg 2021

**Submitted on:** ……………………………………………………………………….

## Members of the thesis committee:

**Chairperson:**

**Primary Supervisor: Prof. Dr. Hermann Schindelin**

**Supervisor (Second): Prof. Dr. Sonja Lorenz**

**Supervisor (Third): Prof. Dr. Christoph Sotriffer**

**Supervisor (Additional): Prof. Dr. Alexander Buchberger**

**Date of public Defence:** …………………………………………………....
**Date of Receipt of Certificates:** ……………………………………………….

# TABLE OF CONTENTS

# SUMMARY

Ubiquitylation is a protein post translational modification, in which ubiquitin is covalently attached to target protein substrates resulting in diverse cellular outcomes. Besides ubiquitin, various ubiquitin-like proteins including FAT10 exist, which are also conjugated to target proteins. The underlying modification mechanisms are conserved. In the initial step, ubiquitin or a ubiquitin-like protein is thioester-linked to a catalytic cysteine in the E1activating enzyme in an ATP-dependent manner. The respective protein modifier is then transferred to an E2 conjugating enzyme in a transthioesterification reaction. Finally, an E3 ubiquitin ligase E3 catalyzes the covalent attachment of the protein modifier to a substrate. In the case of ubiquitin, multiple ubiquitin molecules can be attached to a substrate in the form of either linear or branched polyubiquitin chains but also as single ubiquitin modifications. Depending on the nature of the ubiquitin chain, the substrates are destined to various cellular processes such as their targeted destruction by the proteasome but also non-degradative outcomes may occur.

As stated above FAT10 is a ubiquitin-like protein modifier which typically targets proteins for proteasomal degradation. It consists of two ubiquitin-like domains and is mainly expressed in cells of the human immune system. The reported involvement of FAT10 modifications in cancers and other diseases has caught the attention of the scientific community as an inhibition of the FAT10ylation process may provide avenues for novel therapeutic approaches. UBA6 is the E1 activating enzyme that resides at the apex of the FAT10 proteasomal degradation pathway. UBA6 not only recognizes FAT10 but can also activate ubiquitin as efficiently as the ubiquitin specific E1 UBA1. The dual specificity of UBA6 may complicate the inhibition FAT10ylation since targeting the active site of UBA6 will also inhibit the UBA6-catalyzed ubiquitin activation. Therefore, it is important to understand the underlying principles for the dual specificity of UBA6 prior to the development of compounds interfering with FAT10ylation.

In this thesis important novel insights into the structure and function of UBA6 were derived by X-ray crystallography and biochemical methods. The first crystal structure of UBA6 reveals the multidomain architecture of this enzyme in atomic detail. The enzyme is composed of a rigid core including its active and inactive adenylation domains as well as a 4 helix bundle. Overall, the molecule adopts a "Y" shape architecture with the core at the base and the first and second catalytic half domains forming one arm of the "Y" and the ubiquitin fold domain constituting the other arm. While UBA6 shares the same domain architecture as UBA1, substantial differences were revealed by the crystal structure. In particular, the first catalytic

half domain undergoes a significant shift to a position more distal from the core. This rigid body movement is assumed to generate room to accommodate the second ubiquitin-like domain of FAT10. Differences are also observed in a hydrophobic platform between the core and the first catalytic half domain and the adenylation active site in the core, which together from the binding sites for ubiquitin and FAT10. Site directed mutagenesis of key residues in these areas altered the UBA6-catalyzed activation of ubiquitin and FAT10. UBA6 variants were generated with the goal of trying to block the activation of FAT10 while still maintaining that of ubiquitin activation, in order to fully explain the dual specificity of UBA6. However, none of these mutations could block the activation of FAT10, while some of these UBA6 variants blocked ubiquitin activation. Preliminary inhibition assays with a group of E1 inhibitors belonging to the adenosyl sulfamate family demonstrated potent inhibition of FAT10ylation for two compounds. The dual specificity of UBA6 hence needs to be further examined by biochemical and structural methods. In particular, the structure of a complex between UBA6 and ubiquitin or FAT10 would provide key insights for further biochemical studies, ultimately allowing the targeted inhibition of the FAT10ylation machinery.

# ZUSAMMENFASSUNG

Der Prozess der Ubiquitinierung stellt eine posttranslationale Modifikation dar, bei der das kleine Protein Ubiquitin kovalent an ein Zielprotein angehängt wird, was zu verschiedenen zellulären Effekten führt. Neben Ubiquitin existieren verschiedene ubiquitinähnliche Proteine, wie z.B. FAT10, die an Zielproteine angehängt werden können. Die der Modifikation zugrunde liegenenden Mechanismen der Proteinmodifikation sind konserviert. Im ersten Schritt wird Ubiquitin oder das ubiquitinähnliche Protein in einer ATP-abhängigen Reaktion kovalent an das katalytische Cystein des aktivierenden Enzyms (E1) gebunden. Danach wird es durch Transthioestherifizierung an ein konjugierendes Enzym (E2) übertagen und schließlich durch eine Ligase (E3) kovalent an das Substrat gehängt. Ubiquitin kann entweder einzeln oder in Form linearer oder verzweigter Ketten an ein Substrat angehängt werden, was wiederum zu verschiedenen funktionalen Konsequenzen wie dem Abbau das Proteins durch das Proteasom führen kann.

Wie schon erwähnt ist FAT10 ein ubiquitinähnliches Protein, das üblicherweise Zielproteine für den Abbau durch das Proteasom markiert. Es besteht aus zwei ubiquitinähnlichen Domänen und wird im Menschen hauptsächlich in Zellen des Immunsystems exprimiert. Die Beteiligung von FAT10 an der Entstehung von Krebs and anderen Krankheiten hat die Aufmerksamkeit der wissenschaftlichen Gemeinschaft erregt, da Inhibition des ‚FAT10ylation' Prozesses einen neuen therapeutischen Ansatz zur Behandlung dieser Krankheiten darstellen könnte. UBA6 fungiert hierbei als E1 s Enzym, das am Anfang des FAT10-abhängigen proteasomalen Abbaus steht. UBA6 aktiviert neben FAT10 auch Ubiquitin mit ähnlicher Effizienz wie das ubiquitinspezifische E1 UBA1. Diese Bispezifität von UBA6 könnte die Inhibition der FAT10ylierung erschweren, da die Inhibition der katalytischen UBA6 Aktivität gleichzeitig UBA6-abhängige Ubiquitinaktivierung behindern würde. Daher ist für die zukünftige Entwicklung FAT10-spezifischer UBA6 Inhibitoren ein grundlegendes Verständnis der UBA6 Bispezifität unerlässlich.

In dieser Dissertation wurden wichtige, neue Einsichten in die Struktur und Funktion von UBA6 durch Röntgenkristallographie und biochemische Methoden gewonnen. Die erste Kristallstruktur von UBA6 zeigt die Multidomänenarchitektur des Enzyms bei atomarer Auflösung. Das Protein besteht aus einem starren Kern, der sowohl seine aktive als auch inaktive Adenylierungsdomäne sowie ein 4-Helix Bündel enthält. Das Molekül nimmt eine an ein Y erinnnernde Form ein, in der der Kern die Basis, die erste und zweite katalytischen Halbdomänen einen Arm und die ubiquitinähnliche gefaltete Domäne den zweiten Arm darstellen. Zwar ähneln sich der Domänenaufbau von UBA6 und UBA1, jedoch zeigte die

Kristallstruktur bedeutende Unterschiede zwischen den beiden auf. Speziell die erste katalytische Halbdomäne ist in UBA6 im Vergleich zu UBA1weiter vom Enzymkern entfernt. Diese ‚Bewegung' erlaubt wahrscheinlich die Platzierung der zweiten UBL-Domäne von FAT10. Weitere Unterschiede konnten auch in der hydrophoben Oberfläche zwischen Kern, erster katalytischer Halbdomäne und dem aktiven Zentrum für die Adenylierung im Kern beobachtet werden, die zusammen die Bindestelle für Ubiquitin und FAT10 bilden. Durch ortsgerichtete Mutagenese von Schlüsselpositionen in dieser Region kontte die UBA6-katalysierte Aktivierung von entweder Ubiquitin oder FAT10 unterbunden werden. Um die Bispezifität von UBA6 zu entschlüsseln wurden UBA6 Varianten mit dem Ziel erzeugt, die Aktivierung von FAT10 unter Aufrechterhaltung der von Ubiquitin zu blockieren. Obwohl keine dieser Mutationen die FAT10-Aktivierung unterband, verhinderten einige jedoch die Aktivierung von Ubiquitin. Vorläufige Inhibitionsexperimente mit E1-Inhibitoren aus der Adenosylsulfamat Klasse zeigten starke Inhibition der FAT10ylierung durch zwei Verbindungen. Die Bispezifität von UBA6 bedarf weiterer strukturbiologischer und biochemischer Untersuchungen. Vor allem Kristallstrukturen von UBA6 in FAT10 und Ubiquitin-gebundener Form würden wichtige Erkenntnisse für weiteregehende biochemische Untersuchungen und schließlich die gezielte Unterdrückung der FAT10ylierungsmaschinerie liefern.

# 1. MAIN INTRODUCTION
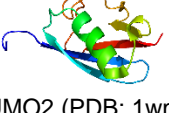
## 1.1. The process of ubiquitylation
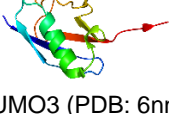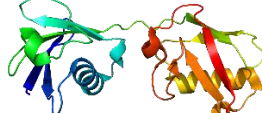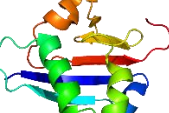
### 1.1.1. Ubiquitin and ubiquitylation
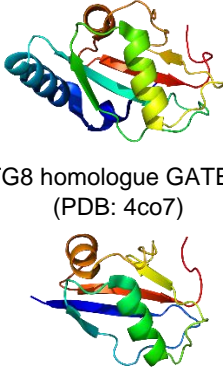
Ubiquitin is a 76 residue protein and is the founding member of a large family of proteins containing the β-grasp fold, which also includes FAT10, SUMO1/2/3/4, Nedd8, ISG15, Ufm1, Urm1, ATG8/12 (Table 1). Collectively, they are referred to as ubiquitin-like modifiers (UBLs). Ubiquitin is used as a posttranslational modification which alters the structure, function, and/or localization of target proteins (Wilkinson 2005). Ubiquitylation is the process in which ubiquitin molecules are covalently attached to substrate proteins by forming an isopeptide bond between the substrate and ubiquitin (Komander and Rape 2012, Scaglione, Basrur et al. 2013). A key feature of ubiquitin and UBLs is the presence of a conserved di-Gly motif at their C-termini which plays a critical role in the conjugation reaction. The carboxyl group of the C-terminal glycine residue is the site of attachment to substrates, while amino groups in the side chains of lysine residues are the most common target sites for coupling to ubiquitin within the substrate proteins. The conjugation of ubiquitin results in an isopeptide bond between ubiquitin and the substrates (Kerscher, Felberbaum et al. 2006). Non-canonical ubiquitylation events may involve the N-terminal amino group (Scaglione, Basrur et al. 2013) of a target protein or the side chains of cysteine, serine or threonine residues (Wang, Herr et al. 2009).

In the simplest case a single ubiquitin is attached to a protein in a process referred to as mono-ubiquitylation. Since ubiquitin has seven lysine residues and one N-terminal amino group, this modification can be extended, resulting in the formation of oligo-/poly-ubiquitin chains involving any of the eight amino groups of ubiquitin. This may lead to the formation of homogeneous Ub-chains in which the same linkage type is consistently utilized or mixed chains in which different linkages are employed and even branched chains (Komander and Rape 2012). Depending on the configuration of the attached ubiquitin chains, ubiquitylation can target the modified substrates to different biological pathways (Komander and Rape 2012, Huang and Dixit 2016). All possible linkages have been detected in cells (Peng, Schwartz et al. 2003, Xu, Duong et al. 2009) with the Lys48 and Lys63 linkages being the best characterized examples (Haglund and Dikic 2005). Lys48 linked ubiquitin chains with at least four ubiquitins can efficiently target a substrate to the subsequent degradation by the proteasome (Thrower, Hoffman et al. 2000). In

contrast, Lys63 linkages are involved in cellular processes such as DNA repair and signal transduction events (Chan and Hill 2001).

**Table 1. ubiquitin like proteins in human.**

| UBL | Structure | Sequence identity to Ubiquitin | Selected functions |
|---|---|---|---|
| Ubiquitin |  (PDB: 6dc6) | 100% | Proteasomal degradation, antigen processing, apoptosis, regulation of the cell cycle , DNA transcription and repair (Hasselgren and Fischer 1997) |
| FAT10 (ubiquitin D) |  (PDB :6gf1 and 6gf2) | 29% and 36% for N and C domain, respectively | Proteasomal degradation (Hipp, Kalveram et al. 2005), regulation of apoptosis (Ross, Wosnitzer et al. 2006), (Aichem and Groettrup 2020) |
| SUMO1/2/3/4 |  SUMO1 (PDB: 2uyz)  SUMO2 (PDB: 1wm2)  SUMO3 (PDB: 6nnq) SUMO4 (N/A) | 18%, 16%, 16% and 14% for SUMO1/2/3 /4, respectively | Cell cycle control, nuclear transport, DNA repair, (Johnson 2004, Meulmeester and Melchior 2008). |
| NEDD8 |  (PDB: 1ndd) | 58% | NF-κB signaling, cell growth, transcriptional control, genomic integrity and tumor suppression (Hori, Osaka et al. 1999, Xirodimas, Saville et al. 2004) |
| ISG15 |  (PDB: 1z2m) | 30% and 37% for N and C domain, respectively | Response to bacterial and viral infection (Hsiang, Zhao et al. 2009, Bogunovic, Byun et al. 2012), cytokine production (Swaim, Scott et al. 2017) |
| Ufm1 |  (PDB: 5iaa) | 21% | Brain development (Nahorski, Maddirevula et al. 2018), apoptosis, reticulophagy (Li, Zhang et al. 2017, Liang, Lingeman et al. 2020) |
| Urm1 |  | 25% | RNA modification (Schlieker, Van der Veen et al. 2008) |

8

| | Yeast homologue (PDB: 2pko) | | |
|---|---|---|---|
| ATG8/ATG12 |  ATG8 homologue GATE-16 (PDB: 4co7)  ATG12 (PDB: 4gdk) | 13% and 17% for GATE-16 and ATG12, respectively | Autophagy (Mizushima, Sugita et al. 1998) |

### 1.1.2. The ubiquitylation cascade in humans

Ubiquitylation is carried out by an enzymatic cascade including E1 activating enzymes, E2 conjugating enzymes and E3 ligating enzymes (Figure 1A). At first, an E1 enzyme binds ubiquitin and Mg-ATP, and catalyzes the C-terminal acyl-adenylation of ubiquitin which generates a ubiquitin-AMP adduct and releasing $PP_i$. In the second step, the catalytic cysteine of the E1 attacks the ubiquitin-AMP intermediate to form a thioester bond between the cysteine and the C-terminal glycine of ubiquitin resulting in the formation of an E1~ubiquitin/UBL adduct where the "~" designates a covalent linkage between the two proteins. In the second step, E2 enzymes bind to the E1~ubiquitin conjugate, triggering a transthioesterification reaction in which ubiquitin is transferred to the active site cysteine present in each E2. Finally, E3 enzymes catalyze the ligation of ubiquitin to the substrates. From a mechanistic perspective there are two groups of E3s. The first group, which includes the HECT (homologous to E6AP C-terminus) ligases (Rotin and Kumar 2009), RBR (RING-in between-RING) ligases (Deshaies and Joazeiro 2009), and the IpaH (invasion-plasmid antigen H) ligase (Li, Jiang et al. 2017), which are structurally and evolutionarily distinct E3 enzymes, bind to the E2~ubiquitin adduct and, in another transthioesterification reaction, accept the activated Ub on their active site cysteine. The E3s themselves then catalyze the aminolysis reaction in which a substrate amino group attacks the reactive thioester to finally form an isopeptide bond between Ub and the acceptor lysine in the substrate. The second group are the RING (really interesting new gene) E3 ubiquitin ligases, which constitute by far the largest E3 family with around 600 members (Metzger, Pruneda et al. 2014). They lack a catalytic cysteine and therefore function as scaffolds which bring the E2~ubiquitin complex and the substrate into close spatial proximity to facilitate the transfer of ubiquitin onto the substrate protein. Additionally, both classes of E3

enzymes contain a sub-family of ligases, in which the melanoma antigen (MAGE) proteins assemble with the E3s and function as adaptors mediating the recruitment of substrates to the ligase complex for ubiquitylation (Yang, Huang et al. 2020). MAGEs bind to specific single-subunit E3 ubiquitin ligases, both RING and HECT type ligases, and regulate substrate specificity, ligase activity, and subcellular localization (Florke Gee, Chen et al. 2020). In summary, the process of ubiquitylation consists of an initial activation step followed by sequential transthioesterification reactions that terminate in an aminolysis step (Schulman and Harper 2009, Bedford, Lowe et al. 2011, Keszei and Sicheri 2017, Wang, Argiles-Castillo et al. 2020). In humans and other mammals, two ubiquitin activating enzymes exist, which are referred to as UBA1 and UBA6, together with approximately forty E2 enzymes (Stewart, Ritterhoff et al. 2016) and 500-1000 E3 ligases (Nakayama and Nakayama 2006). In non-mammalian enzymes with the curious exception of the sea urchin, there is only a single E1 enzyme, corresponding to UBA1, and the number of E2 and E3 enzymes decreases from mammals to simpler organisms. The divergence of the cascade with one or two E1 enzymes at the apex and increasingly larger number of E2 and E3 enzymes generates the necessary substrate specificity, which elicits and controls the multitude of biological processes following ubiquitylation (Figure 1A).

The sequential action of E1, E2 and E3 enzymes is also a hallmark of the activation and transfer of all other ubiquitin-like protein modifiers. Nevertheless, the complexity of the ubiquitylation cascade is always highest as reflected in the largest number of participating enzymes. In certain cases, such as FAT10 and Urm1, the participating E2 (Urm1-conjugation) and E3 enzymes have not yet been identified and, at least in theory, may not even exist. The general existence of designated E1, E2 and E3 enzymes, with minor exceptions such as UBA6 recognizing ubiquitin and FAT10, in each pathway implies that ubiquitylation and related pathways are controlled independently.

(A)



(B)
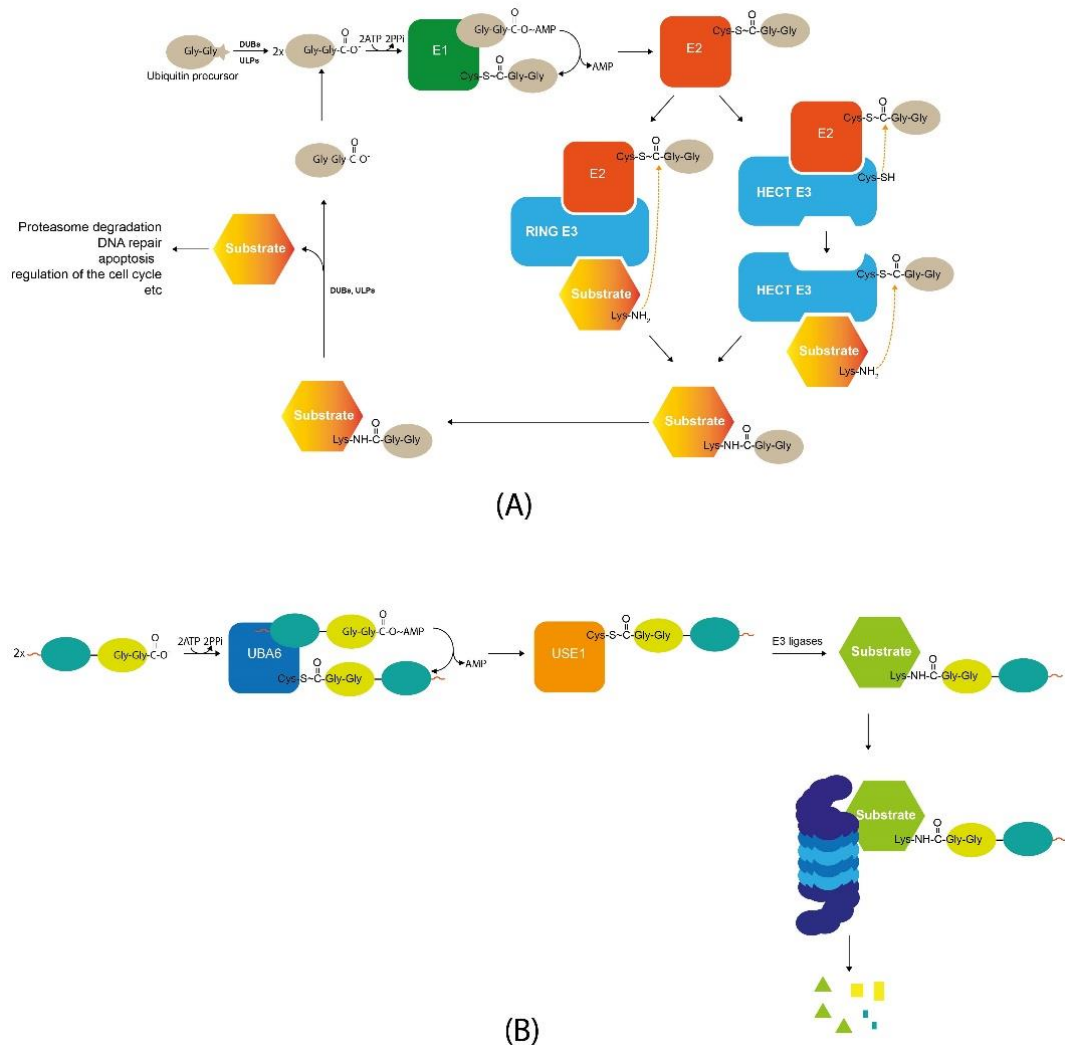
**Figure 1: Reaction cascade of ubiquitylation (A) and FAT10ylation (B).**

### 1.1.3. Ubiquitin activating (E1) enzymes

The E1 enzymes, as stated previously, reside at the apex of the ubiquitylation (or related) cascades. Up to now, eight E1 enzymes have been reported which activate ubiquitin and UBLs. These are: UBA1 and UBA6, which activate ubiquitin, with UBA6 also activating FAT10; the heterodimeric SAE1-SAE2 enzyme (also known as SAE1-UBA2), which activates SUMO; another heterodimeric enzyme, NAE1-UBA3, which activates NEDD8; UBA7, the activating enzyme for ISG15; UBA4, which targets Urm1; UBA5, the activating enzyme for Ufm1; and, finally, Atg7 which acts on both ATG12 and ATG8. E1 enzymes are divided into two groups, canonical and noncanonical, based on their domain architectures and enzymatic mechanisms, (Schulman and Harper 2009). With the exception of UBA6 and UBA7 the structures of all E1 enzymes are known (Figure 2 and Figure 3).

The *Escherichia coli* proteins MoeB and MoaD are involved in molybdenum cofactor biosynthesis (Rajagopalan 1997), which is an evolutionarily conserved pathway present in bacteria, archaea and eukaryotes, with the notable exception of *S. cerevisiae*. MoaD and UBLs display the same β-grasp fold including the C-terminal di-Gly motif (Rudolph, Wuebbens et al. 2001) (Figure 3D). The homodimeric MoeB enzyme binds two MoaD subunits and activates them in a conserved mechanism as the E1s activate ubiquitin or the UBLs up to the point of acyladenylate formation. This suggests that the UBLs and the E1s are derived from two ancestral genes closely related to MoaD and MoeB, respectively (Rudolph, Wuebbens et al. 2001). Following acyladenylate formation the pathways diverge leading to the formation of a thiocarboxylate on the MoaD C-terminus which acts as the S-donor during molybdenum cofactor biosynthesis instead of the standard thioester discussed in the context of the E1 enzymes.

The canonical group of E1 enzymes includes UBA1, SAE1-UBA2, NAE1-UBA3, UBA6 and UBA7. The structurally characterized enzymes in this group show a common hand-shaped structure with a conserved domain architecture (Figure 2). The rigid palm of the hand is formed by the adenylation domains, which are composed of two MoeB-like domains fused into a single polypeptide (Lee and Schindelin 2008). While MoeB as a homodimer harbors two active sites, the adenylation domains in canonical E1 enzymes have lost one active site and hence are named active and inactive adenylation subdomains (AAD and IAD, respectively). The AAD domain contains a nucleotide binding motif Gly-*X*-Gly-*X*-*X*-Gly, which binds to the triphosphate moiety of ATP and the two domains together catalyze the C-terminal adenylation of ubiquitin or the respectively UBL. The AD also provides a central platform that recognizes and binds ubiquitin or the cognate UBL. The IAD is catalytically inactive and does not bind to the UBLs mostly due to the presence of a 4-helix bundle that sterically blocks the interaction and the lack of most catalytic residues. While the IAD primarily provides structural stability, it nevertheless harbors a conserved arginine at its N-terminus which is essential for the adenylation reaction (Lee and Schindelin 2008, Lv, Williams et al. 2018).
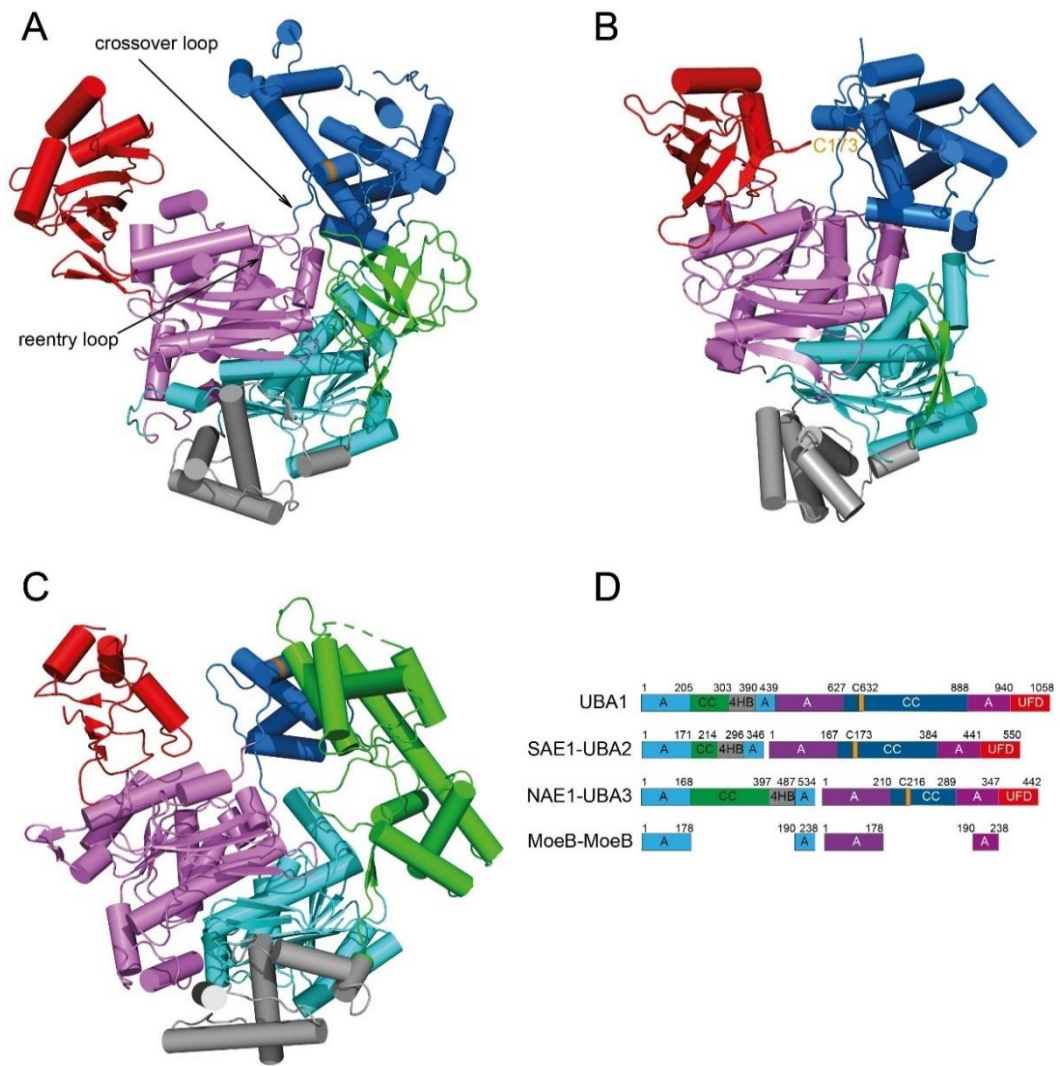
**Figure 2: Known structures of human canonical activating enzymes. (A) UBA1 (PDB entry: 6dc6), (B) the heterodimeric SAE1-UBA2 (PDB entry: 1y8q) and (C) NAE1-UBA3 (PDB entry: 1r4m) enzymes. (D) Domain architechtures of canonical E1 enzymes with inactive adenylation domain (IAD) in cyan, active adenylation domain (AAD) in violet, the first catalytic cysteine half (FCCH) domain in green, the second catalytic cysteine half (SCCH) domain in blue, the four helix bundle (4HB) in grey and the ubiquitin fold domain (UFD) in red. The position of the active site cyteine residue is shown in orange in (A) - (D).**

The first catalytic cysteine half (FCCH) domain, which represents the thumb of the hand, is connected to the IAD and 4HB domain via flexible linkers. The FCCH domain is involved in ubiquitin/UBL-binding and the catalytic mechanism of the E1 (Hann, Ji et al. 2019). The fingers of the hand are formed by the second catalytic cysteine half (SCCH) domain and the C-terminal ubiquitin folding domain (UFD).

The SCCH domain contains the active site cysteine residue, which attacks the adenylated intermediate to form the thioester bond with ubiquitin or the UBL, respectively, while the UFD primarily binds to cognate E2 enzymes to bring them, in particular, their active site cysteine residues, into close proximity to the thioester bond to allow the transthioesterification reaction, which transfers ubiquitin/UBL to the E2 enzymes. The crossover loop connecting the AAD and the SCCH domain orients the C-terminal tail of ubiquitin/UBL into the adenylation active site by engaging in critical interactions with ubiquitin/UBL (Lee and Schindelin 2008, Hann, Ji et al. 2019, Williams, Qie et al. 2019) (Figure 2A) and provides flexibility to allow for domain rearrangements during the catalytic cycle (see below). The reentry loop constitutes a second flexible connection between the AAD and the SCCH and, like the crossover loop, accommodates domain motions during the catalytic cycle.

In contrast to UBA1 and UBA6, the SUMO E1 and NEDD8 E1 are heterodimeric enzymes containing two subunits. The first subunit (also referred as UBA2 and UBA3, respectively) contains the AAD, the SCCH domain and the UFD. The second subunit (SAE1 and NAE1, respectively) is composed of the IAD, the FCCH domain and the 4-helix bundle (4HB) (Walden, Podgorski et al. 2003, Lois and Lima 2005). Initial structures of canonical E1s in complex with ubiquitin/UBL and ATP revealed a distance of 35 Å between the C-terminal glycine of ubiquitin/UBLs and the active site cysteine in the SCCH domain of the E1s. This immediately suggested that significant conformational changes are required to facilitate thioester bond formation (Walden, Podgorski et al. 2003, Lee and Schindelin 2008, Lv, Williams et al. 2018). Subsequent crystal structures of the SUMO E1 with an adenylate analog and a E1~SUMO tetrahedral intermediate demonstrated dramatic conformational changes involving an approximately 130° rotation of its SCCH domain (Olsen, Capili et al. 2010). The remodeling of two helices in this domain including the one that contains the catalytic cysteine residue and two helices in the adenylation active site, displacement of the N-terminal helices of the IAD and a rearrangement of the crossover and re-entry loops generates a structural framework for the second half reaction (Figure 4). These conformational changes result in the replacement of nearly half of the active site residues involved in Mg-ATP binding with residues from the SCCH domain which, in turn, are required for thioester bond formation (Olsen, Capili et al. 2010, Hann, Ji et al. 2019).

After thioester bond formation and the release of AMP, the UBL linked SCCH domain rotates back to the open conformation leaving the empty adenylation site

for the binding of a second ubiquitin/UBL (Schafer, Kuhn et al. 2014). The presence of ubiquitin or the UBLs in the SCCH domain triggers outward rotations of the SCCH domain and the UFD, thereby generating more space between the two domains to accommodate the respective E2 and provide binding interfaces for this enzyme (Lois and Lima 2005, Olsen and Lima 2013, Williams, Qie et al. 2019). The crystal structures of complexes between the isolated UFD of selected E1 enzymes and E2 enzymes suggested that the latter bind to the UFD before being sandwiched between the UFD and SCCH domains. Available structures of Uba1 indicate a rotation of the UFD from a distal to a proximal position relative to the SCCH between the open and closed conformation (Williams, Qie et al. 2019). The UFD was also reported to utilize conserved residues to build a distinctive interaction network with the respective E2 (Lv, Rickman et al. 2017, Williams, Qie et al. 2019). The sandwiching binding involves closing rotations of the UFD and the SCCH domain. These conformational movements bring the E2 active site cysteine residue into close spatial proximity of active site cysteine in the E1, which is thioesterified to ubiquitin or the UBLs so that the transthioesterification reaction can proceed (Williams, Qie et al. 2019).
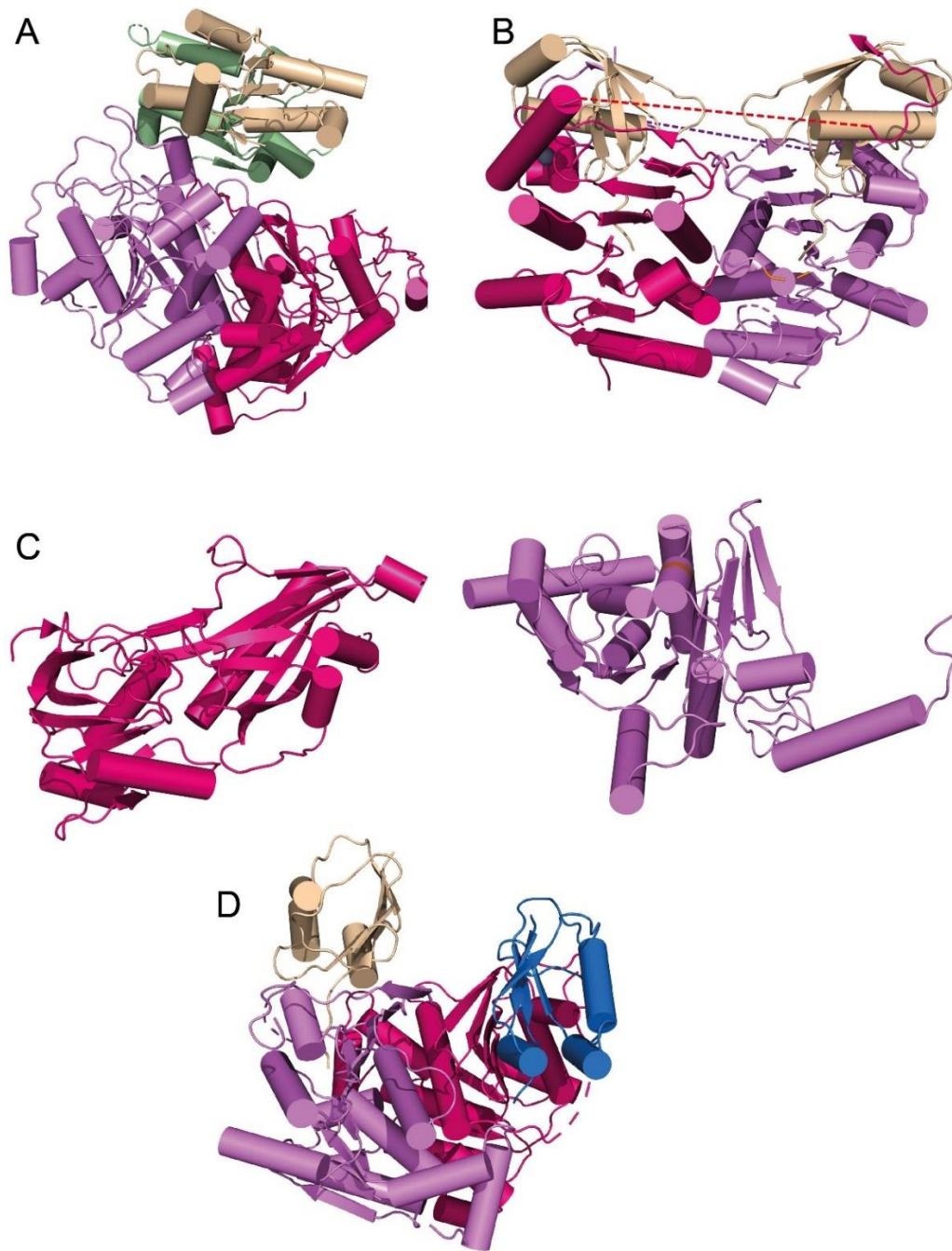
**Figure 3: Structures of noncanonical activating enzymes. (A)** Homodimeric UBA4 (PDB: 6yub) from *Chaetomium thermophilum* with the two identical adenylation subunits in violet and hot pink and their rhodanese-homology domains (RHD) depicted in green and wheat. **(B)** Homodimeric UBA5 in complex with Ufm1 (PDB: 5iaa), with the adenylation subunits in violet and hot pink and the two Ufm1 proteins in wheat. **(C)** Structures of the N-terminal (hot pink, PDB: 3ruj) and C-terminal domains (violet, PDB: 3rui) of ATG7 from *S. cerevisiae*. **(D)** Structure of the homodimeric MoeB (two monomers in violet and hot pink) in complex with MoaD (two monomers in blue and wheat) (PDB: 1jw9).

The noncanonical human E1 enzymes include UBA4 (also known as molybdenum cofactor synthesis protein 3, MOCS3), UBA5 and ATG7. While the human MOCS3 structure is still unknown, the structure and mechanism of its fungal UBA4 homologue has been reported. UBA4 is a two-domain protein containing an adenylation domain (AD) and a rhodanese-homology domain (RHD), which are essential for the dual functionality of the initial ubiquitin-like conjugation followed by a sulfur transfer reaction (Furukawa, Mizushima et al. 2000, Pabis, Termathe et al. 2020). The homodimeric structure of UBA4 harbors two nucleotide binding sites and, while each active site is primarily formed by one subunit, residues located near the N-terminus of the other subunit complement the active site. This is in analogy to MoeB and to the aforementioned canonical E1 enzymes, however, in the latter case only a single adenylation active site is formed. In the case of UBA5 this suggests that AD homodimerization is essential for the adenylation of its UBL, Urm1. Unlike canonical E1 enzymes, the active site cysteine residue of UBA4 resides within the AD in close distance to the Urm1 C-terminal glycine residue. This facilitates thioester bond formation without the need for the significant conformational changes occurring in the canonical E1 enzymes. No E2s for UBA4 have been reported so far (Pabis, Termathe et al. 2020).

ATG7 (also known as APG7L) activates two UBLs, ATG8 and ATG12, which are both involved in autophagy. The covalent modification of ATG12 and ATG8 and its substrate is essential for autophagy to occur (Mizushima, Noda et al. 1998, Ichimura, Kirisako et al. 2000). The crystal structure of ATG7 from yeast (Hong, Kim et al. 2011) was determined (Figure 2C), which revealed that this enzyme also consists of two domains. The N-terminal domain functions like the UFD domain in the canonical E1s since it recruits the E2 enzymes ATG3 and ATG10. The C-terminal domain contains the active site cysteine, an adenylation site and a module mediating homodimerization. Similar to UBA4, the active site cysteine of ATG7 is located close to the C terminus in the co-crystal structure with ATG8 (Hong, Kim et al. 2011). This implies a catalytically competent state appropriate for covalent bond formation between ATG7 and its UBLs. When ATG7 is in action, it binds to ATG3 and ATG8 to form a heterohexameric complex with 2:2:2 stoichiometry composed of the ATG7 homodimer, two ATG3 subunits with each binding to the N-terminal domain of ATG7 and, finally, two ATG8 subunits engaging in interactions with the C-terminal domains. The detailed mechanism of how ATG7 transfer ATG8 to ATG3 still remains unclear (Hong, Kim et al. 2011).

UBA5, with a molecular weight of only 45 kDa is the smallest E1. It contains a minimalistic structure with an adenylation domain and a long loop. As expected, UBA5 also forms a homodimer, however, it exhibits a unique mechanism for UBL adenylation and UBL transfer to the E2. The homodimerization of UBA5 is essential since Ufm1 binds to the UBA5 homodimer in a trans-binding mechanism. Ufm1 simultaneously binds to the Ufm1-interacting sequence in an extended loop of one UBA5 subunit and to the adenylation domain in the other subunit (Figure 2B). The active site cysteine is located in the crossover loop in close proximity to the C terminus of Ufm1. This allows for thioester bond formation with only minor conformational changes compared to canonical E1s. It is unclear whether Ufm1 adenylation takes place in only one UBA5 subunit. The trans-binding mechanism of UBA5 also applies to the binding of UFC1, the corresponding E2 enzyme, in that UFC1 binds to the adenylation domain of one UBA5 subunit and the extended loop of the other subunit. Based on these observations, a hypothesis has been proposed that one subunit holds the protein substrates, while the other subunit catalyzes the activation of Ufm1 and the transfer to UFC1 (Oweis, Padala et al. 2016).

In summary, the E1 activating enzymes evolved from the bacterial MoeB enzyme or its counterpart ThiF, which is involved in thiamin biosynthesis. While noncanonical E1s closely resemble MoeB which homodimerizes to catalyze the activation of the MoaD (Rudolph, Wuebbens et al. 2001), canonical E1s acquired additional domains to exert their functions beyond the initial adenylation step.
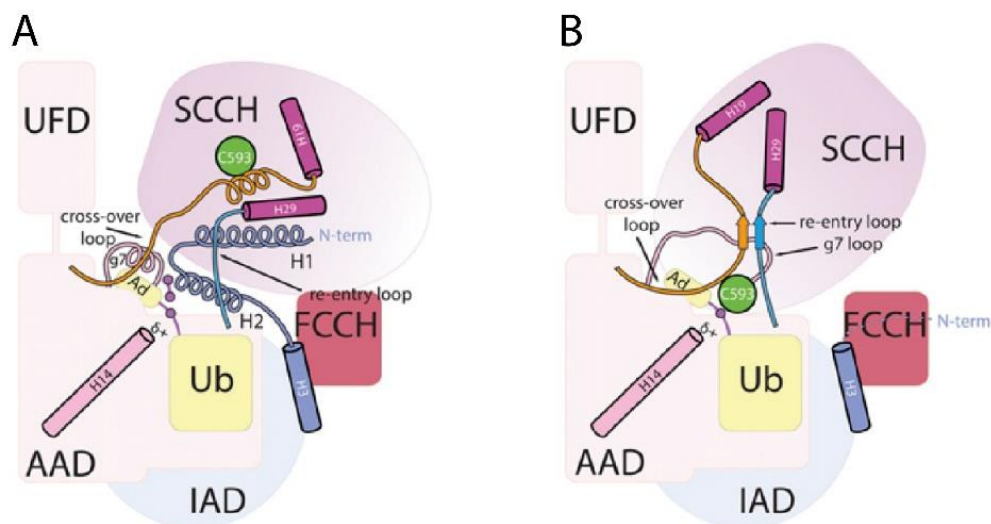


**Figure 4: Structures of UBA1/Ub complexes in the open (A) and closed (B) states (Hann, Ji et al. 2019).**

### 1.1.4. UBA6

The ubiquitin activating enzyme 1 (UBA1) was first discovered and characterized in yeast and plants in 1981 and 1982, respectively (Ciechanover, Heller et al. 1981, Ciechanover, Elias et al. 1982, Haas, Warms et al. 1982). Ten years later, the gene encoding HsUBA1 (HsUBA1) was cloned (Handley, Mueckler et al. 1991). For a long time UBA1 was thought to be the only E1 that activates ubiquitin, however, in 2007 three different groups discovered a second ubiquitin activating enzyme. While searching for E1 enzymes that recognize the UBL FAT10 among proteins that were known to bind ubiquitin or contain E1 consensus motifs, researchers identified a protein which shares 40% sequence identity with UBA1. This protein, now referred to as UBA6, was found to be present only in vertebrates and sea urchins. UBA6 features conserved E1 motifs such as the ATP-triphosphate-binding motif G*XG*XX*GCE and the putative active site cysteine within the consensus sequence P*XCT*XXX*P. The enzyme was found to activate ubiquitin and FAT10 both *in vitro* and *in vivo* and to also charge E2 enzymes with ubiquitin and FAT10. The simultaneous discovery gave rise initially to some confusion with the Pelzer group naming the enzyme UBE1L2 and the Chiu group E1-L2 (E1-like 2), while Jin and others chose the name UBA6 (Chiu, Sun et al. 2007, Jin, Li et al. 2007, Pelzer, Kassner et al. 2007). Among these suggestions, UBA6 represents the most systematic name and has been used widely by scientists since then, consequently it will also be used in this work.

Since the discovery of UBA6, one major question has been, why does ubiquitin need two activating enzymes? Several facts suggest that UBA6 does not simply represent a backup system. The first hint is that UBA1 and UBA6 display different selectivities towards various ubiquitin conjugating enzyme E2s. In a test of 29 E2 enzymes, Jin and others found that eight of them (Table 2) could be charged efficiently by both UBA1 and UBA6, six E2s (Table 2) were charged by UBA6 only poorly when compared to UBA1, eight other E2s (Table 2) were not charged by UBA6 at all and could only be loaded by UBA1, five E2s (Table 2) could neither be charged by UBA1 nor UBA6 and a single E2, referred to as either UBE2Z or USE1, could be specifically charged only by UBA6 (Jin, Li et al. 2007). These different E2 selectivities allow these two E1 enzymes to direct ubiquitin to distinct subsets of E3 enzymes and consequently substrates. The second observation was that UBA1 and UBA6 function in spatially distinct fashions. Although UBA6 is widely expressed in human tissues and cell lines, its expression is ten times lower compared to that of UBA1 in HeLa, U2OS, 3T3, HCT116, ts85 and ts20 cell lines. Furthermore, whereas

only 50% of UBA6 is activated with ubiquitin in proliferating 293T cells, UBA1 is fully activated (Jin, Li et al. 2007). This suggests that UBA1 is the major player in the ubiquitylation pathways (Wang and Zhao 2019). In contrast to the generally lower basal expression levels of UBA6, enzyme levels were found to be up-regulated in human epidermal keratinocyte cells under heat stress and in dendritic cells following treatment with lipopolysaccharide (Ebstein, Lange et al. 2009, Echchgadda, Roth et al. 2013). This suggests that UBA6 may exert specific functions or enhance the activity of UBA1 under certain conditions (Wang and Zhao 2019). Finally, UBA6 and USE1 are the only cascade for FAT10ylation as will be discussed in detail below (Jin, Li et al. 2007).

**Table 2. E2 charging activities of UBA1 and UBA6**

|  | Charging by | |
| --- | --- | --- |
|  | UBA1 | UBA6 |
| E2H | + | - |
| E2A/Rad6a | + | + (weak) |
| E2B/Rad6b | + | + (weak) |
| E2G2 | + | + |
| E2G1 | + | + (weak) |
| E2R2/Cdc34A | + | - |
| E2R1/Cdc34B | + | - |
| E2C/UbcH10 | + | + (weak) |
| E2S | + | + |
| E2D1/UbcH5A | + | + |
| E2D2/UbcH5B | + | + |
| E2D3/UbcH5C | + | + |
| E2D4/UbcH5D | + | + |
| E2E1 | + | + (weak) |
| E2E2 | + | + (weak) |
| E2K | + | - |
| E2T | + | + |
| E2E3 | + | + |

| | | |
|---|---|---|
| E2L3 | + | + |
| E2J1 | + | - |
| E2J2 | + | - |
| E2Q2 | + | - |
| EW2 | + | - |
| E2L6 | - | - |
| E2F | - | - |
| E2M/Ubc12 | - | - |
| E2U | - | - |
| E2I/UBC9 | - | - |
| USE1/UBE2Z | - | + |

## 1.1.5. FAT10

FAT10 (also known as ubiquitin D) is a member of the UBL family, and the gene encoding this protein is located on chromosome 6 in the human major histocompatibility complex class I locus (Fan, Cai et al. 1996). FAT10 is a two-domain protein with both domains belonging to the β-grasp fold family. FAT10 is expressed in mature dendritic cells and B cells, but it is also inducible by the proinflammatory cytokines γ-interferon and tumor necrosis factor α in cells derived from various tissues. While ubiquitin and other UBLs typically require limited proteolysis at their carboxy termini to liberate the C-terminal diglycine motif, FAT10 is synthesized as a mature protein with an exposed diglycine motif at its C-terminus and hence can be directly activated and conjugated to substrate proteins (Raasi, Schmidtke et al. 2001). FAT10 targets proteins for rapid proteasomal degradation. Interestingly, however, FAT10 is also degraded by the proteasome (Figure 1B). The intrinsically disordered heptapeptide with the sequence APNASC located at the N-terminus of FAT10 (residues 1-7) enables FAT10-conjugates to bypass the requirement of unfolding activity by the AAA+ ATPase VCP/p97 prior to degradation by the 26 S proteasome and this sequence motif, together with the poor stability of FAT10, enables a fast, direct and irreversible targeting of FAT10 along with its substrates to the proteasome (Aichem, Anders et al. 2018).

In attempts to solve the structure of the poorly soluble FAT10 full-length protein, it was separated into its two domains and all of its 5 cysteine residues were replaced with Ser to improve FAT10 solubility and stability. Due to these improvements, the

crystal structure of the FAT10 N-terminal domain and NMR structure of its C-terminal domain could be deciphered (Aichem, Anders et al. 2018). As expected, due to sequence identities of 29% and 36% to ubiquitin for the N and C terminal domain, respectively, both domains exhibit the β-grasp fold. The two domains are connected by a flexible linker (residues 82 to 86) and deletion of this linker abrogates FAT10 conjugation by UBA6. Superimposition of both domains with ubiquitin identified unique surface properties that enable them to interact with distinct binding partners (Aichem, Anders et al. 2018).

### 1.1.6. Limited specificity of UBA6, ATG7 and SAE1-UBA2

Among the E1 enzymes, UBA6, the SUMO activator SAE1-UBA2 and the autophagy E1 ATG7 are somewhat distinct from other activating enzymes as they are less specific and can activate more than one ubiquitin-like modifier. SAE1-UBA2 is the heterodimeric canonical E1 which recognizes and activates SUMO as described above. In humans, there are 4 members in the SUMO protein family (SUMO1-4), while there are still three family members in yeast, referred to as Smt3, Smt3p and K7-Smt3p. Human SUMO-1 displays approximately 50% sequence identity with the 96% identical SUMO-2 and SUMO-3 pair (Muller, Hoege et al. 2001). Finally, SUMO-4 is 87% identical to SUMO-2 and 42 % identical to SUMO-1 (Bohren, Nadkarni et al. 2004). SUMO family members are only distantly related to ubiquitin, e.g. SUMO-1 shares only 18% sequence identity with human ubiquitin. Nevertheless, all SUMO family members also adopt the β-grasp fold (Bayer, Arndt et al. 1998, Huang, Ko et al. 2004, Ding, Xu et al. 2005).

As stated above, the noncanonical E1 ATG7 is involved in autophagy and activates the distantly related (only 17% sequence identity) UBLs ATG8 and ATG12. The structures of yeast ATG8 (PDB: 2kq7) and yeast ATG12 (PDB: 3w1s) also revealed the β-grasp ubiquitin-like fold with minor variations. ATG8 features a short N-terminal extension folding into a single α-helix (Table 1) (Schwarten, Stoldt et al. 2010, Metlagel, Otomo et al. 2013, Noda, Fujioka et al. 2013).

Unlike the two aforementioned E1s, UBA6 recognizes two fundamentally different ubiquitin-like proteins, namely ubiquitin and FAT10 in the sense that FAT10 contains two ubiquitin-like domains and ubiquitin only a single domain. Consequently, FAT10 (165 residues) is more than double the size of ubiquitin (76 residues). Although it will appear as two ubiquitin molecules fused together, it is topologically different from most di-ubiquitin moieties formed during ubiquitin chain

formation with the exception of linear di-ubiquitin. FAT10 exhibits a higher affinity towards UBA6 compared to ubiquitin suggesting that additional contacts exist between the E1 and the N-terminal domain of FAT10 (Gavin, Chen et al. 2012). The linker between the two domains of FAT10 also plays an important role in its activation and conjugation. UBA6 does not activate a FAT10 variant lacking the linker, although it can activate the recombinant linear di-ubiquitin which topologically, most closely resembles FAT10 (Aichem, Catone et al. 2014). Moreover, mutations in this linker abrogate auto-FAT10ylation of USE1 (UBE2Z), the E2 solely conjugating FAT10 (Aichem, Anders et al. 2018). A chimeric construct of ubiquitin at the N-terminus with the C-terminal domain of FAT10 behaved just like wild type FAT10 while ubiquitin fused to the C-terminus of the N-terminal domain of FAT10 also abrogated auto-FAT10ylation of USE1. These data indicate that the C-terminal domain together with the linker are essential for the recognition of FAT10 by the E1 UBA6 and the E2 USE1, although the inhibition of the auto-FAT10ylation of USE1 does not generally affect its function and FAT10 conjugation (Aichem, Anders et al. 2018). Somewhat surprisingly, however, neither of the individual FAT10 domains alone nor a mixture of the two could efficiently compete with ubiquitin in its binding to UBA6 (Gavin, Chen et al. 2012). This observation suggests that the C-terminal domain of FAT10 does not bind as tightly to UBA6 as ubiquitin, however, due to avidity affects from its N-terminal domain FAT10 displays an overall enhanced binding to UBA6.

The C-terminal four residues (CYCI) prior to the di-glycine motif were reported to be important for the specificity of FAT10. When these residues were replaced with the corresponding residues [71]LRLR[74] at the C-terminal tail of ubiquitin, the resulting FAT10 variant behaved just like ubiquitin in that it could be activated by both UBA1 as well as UBA6, and UBA6 could transfer the FAT10 variant to USE1 and other E2 enzymes like UBE2D3 and BIRC6. Nevertheless, the ubiquitin variant with CYCI at its C-terminal tail did not behave like wild type FAT10, meaning that this ubiquitin variant was still activated by UBA1 (Schelpe, Monte et al. 2016). This is suggesting that other elements exist, which are involved in the selectivity of FAT10 towards the E1s, not excluding the FAT10 N-terminal domain, the linker and the E1 enzymes themselves. In summary, the following conclusions regarding different structural elements of FAT10 can be drawn: (1) Its N-terminal tail and the flexible two domain structure enable the fast, irreversible proteasomal degradation of FAT10; (2) The linker contributes to its tight binding and activation by UBA6 and the auto-FAT10ylation of USE1; (3) The C-terminal motif is responsible for its specificity

towards UBA6 and USE1. Nevertheless, a full understanding of the structural basis for these interactions and their selectivities are still missing.

Since ubiquitin and FAT10 share the activating enzyme UBA6, the question arises as to whether the two protein modifiers interfere with each other? As mentioned earlier, FAT10 exhibits a higher affinity to UBA6 compared to ubiquitin, however, the transfer rate of FAT10 to USE1 is 6-fold slower than that of ubiquitin (Gavin, Chen et al. 2012). This implies that FAT10 may out-compete ubiquitin during activation by UBA6 in the cell, especially when FAT10 expression is up-regulated. The auto-FAT10ylation of USE1 mentioned earlier leading to proteasomal degradation suggests a regulatory role of USE1-FAT10ylation which limits the conjugation of FAT10 after cytokine-mediated induction (Aichem, Catone et al. 2014). Moreover, FAT10 targets UBA1 for proteasomal degradation, although the FAT10-UBA1 conjugation does not interfere with the function of UBA1 in the ubiquitylation pathway and the amount of the FAT10-UBA1 conjugate was observed to be very low (Bialas, Groettrup et al. 2015). Conversely, FAT10 was found to be ubiquitylated with mono and polyubiquitin chains, however, FAT10 ubiquitylation is not required for its degradation by the proteasome (Hipp, Kalveram et al. 2005). This implies a cross regulation of FAT10 by ubiquitin. Interestingly, FAT10 also interferes with SUMOylation as it binds non-covalently to the adenylation domain in the UBA2 subunit of the SUMO E1, thereby blocking the interaction between SUMO1/2/3 and its E1. Subsequently, FAT10 is efficiently thioester-linked to the active site cysteine residue of SUMO E1 and transthioesterified to the SUMO E2 enzyme UBC9, however, FAT10 will not be conjugated to SUMO substrates, thus suggesting that SUMO E3 ligases do not recognize FAT10. Hence, the conjugation of FAT10 to UBC9, the SUMO-specific E2, allows for a specific inhibition of SUMOylation by FAT10, particularly upon proinflammatory-mediated FAT10 overexpression (Aichem, Sailer et al. 2019).

### 1.1.7. The higher selectivity of UBA6 during the FAT10ylation of E2 enzymes

The human genome encodes for more than 40 E2 enzymes which act as conjugating enzymes for ubiquitin and all UBLs, the majority of which are involved in ubiquitylation events. These enzymes primarily conduct two types of reactions, namely transthioesterification (transfer from a thioester to a thiol group) and aminolysis (transfer from a thioester to an amino group). All E2s have interactions with their specific E1 and several E2s may share a common E1 (Stewart, Ritterhoff et al. 2016), however, they can typically interact with multiple E3 enzymes. The E3

enzymes follow two distinct mechanisms as already mentioned above to catalyze the ligation of the UBLs to the substrates. In the case of E2 enzymes interacting with RING type E3s, the E2 directly engages a substrate protein and, therefore, plays a role in the determination of where and how a substrate is modified with ubiquitin or a UBL (Stewart, Ritterhoff et al. 2016). In special cases, the E2 enzymes even conjugate ubiquitin or a UBL to the substrate in the absence of the E3 ligase (Haldeman, Xia et al. 1997, Ryu, Choi et al. 2008). Ubiquitin or a UBL is transferred mostly to a lysine residue, or (less frequently) to the N-terminus (Scaglione, Basrur et al. 2013, Fletcher, Christensen et al. 2015), of either a substrate or another ubiquitin/UBL molecule to generate mono-, multi- or poly-ubiquitin/UBL chains. In some unusual cases, the UBLs are transferred to the thiol group of a cysteine residue (Wenzel, Lissounov et al. 2011) or hydroxyl group of serine/threonine residues (Wang, Herr et al. 2009) of the substrates. The configurations of the ubiquitin/UBLs chains specifies the fate and function of the modified substrate proteins (Metzger, Pruneda et al. 2014).

Ubiquitin can be efficiently transferred to many E2 enzmes by both UBA1 and UBA6. However, FAT10 exhibits a considerably higher selectivity since it is charged to USE1 only and USE1 also accepts this modifier only from UBA6 (Table 2) (Jin, Li et al. 2007). The structures of more than 32 human E2 proteins have been solved so far, revealing a highly conserved architecture in their catalytic domain, referred to as ubiquitin conjugating (UBC) domain (Stewart, Ritterhoff et al. 2016), which is typically composed of 4 α-helices and a 4 stranded β-sheet. This also applies to the FAT10-specific E2 USE1 (also known as UBE2Z), however, USE1, like a few other E2 enzymes, also harbors N-terminal and C-terminal extensions. The C-terminal extension folds into two α-helices which reside on top of the UBC domain (Figure 5A). This extension shares a large interface with the core domain and is stabilized by 6 electrostatic interactions and 8 hydrogen bonds. The N-terminal extension (residues 1-99) is a long intrinsically disordered region, which in its amino acid content is highly biased towards alanine and glycine residues. The sequence alignment of USE1 and its homologs UBE2D3 and BIRC6 revealed four insertions in USE1 which were named LA, LB, LC and LD loops. Among them, the LB loop (residues 194-197) is located at the entrance to the active site cleft (Figure 5A). The deletions of either the N-terminal extension or the LB loop leads to activation of USE1 by UBA1 (Schelpe, Monte et al. 2016). Ube2D3 is a member of Ube2D E2 family and represents a minimalistic E2 containing only the UBC domain. The binding interactions of Ube2D3 are well studied (Stewart, Ritterhoff et al. 2016)

(Figure 5A). A superimposition of Ube2D3 and USE1 reveals that the N-terminal extension of FAT10 is located near the E1 binding site, thus possibly suggesting its involvement in binding to UBA6. In contrast, the LB loop is well conserved among class IV E2 enzymes such as BIRC6 and UBE2O. However, BIRC6 and UBE2O do not accept the FAT10 from UBA6 (Schelpe, Monte et al. 2016). This may be due to the residue composition of the LB loop itself or the N-terminal extension of USE1. The LB loop is positioned close to the E2 catalytic cysteine which may explain its observed influence on the reactivity of USE1 (Schelpe, Monte et al. 2016). In fact, mutation of the aspartate which is positioned at the entrance of the active site cleft of Ube2D3 severely reduced the transfer of ubiquitin to lysine side chains (Plechanovova, Jaffray et al. 2012) (Fig 4A). Moreover, the higher selectivity of FAT10ylation towards USE1 is also reflected in the auto-FAT10ylation of USE1 in *cis*, i.e. thioesterified FAT10 on the active site cysteine of USE1 is transferred to a lysine residue within the same USE1 molecule resulting in the formation of an isopeptide bond (Aichem, Pelzer et al. 2010). This auto-FAT10ylation only happens when the FAT10 C-terminal domain is present as mentioned earlier (Aichem, Anders et al. 2018) (Figure 5B). So far, Parkin is the only E3 enzyme of FAT10 that was identified. Mutations in the Parkin-encoding gene PARK2 are associated with familial Parkinson's disease. Parkin is an RBR E3 ligase which binds to USE1, auto-FAT10ylates itself, and facilitates FAT10ylation of the Parkin substrate Mitofusin2 *in vitro* and in cells (Roverato, Sailer et al. 2021).
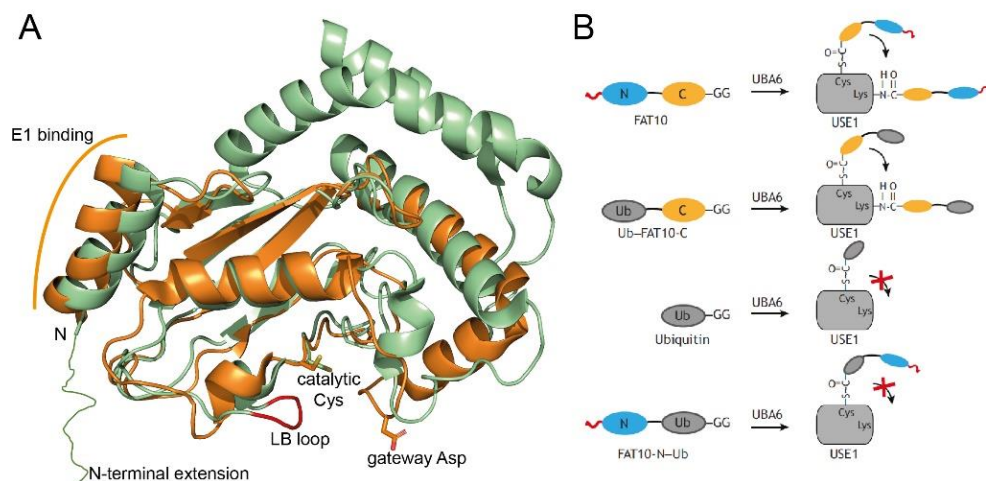


**Figure 5: USE1 structure and mechanism of auto-FAT10ylation. (A) Sumperimposition of USE1 (PDB entry: 5a4p in green) and UBE2D3 (PDB entry: 2fuh in orange). Highlighted are the LB loop of USE1 as well as its N-terminal extension and the E1 binding site of UBE2D3. (B) Mechanism of covalent auto-FAT10ylation by USE1.**

### 1.2. Targeting E1s for drug development

#### 1.2.1. UBA6 and its connections to various diseases

Since ubiquitylation regulates a variety of complex cellular processes, it is not surprising that malfunctions in one or more components of this system lead to variety of diseases (Popovic, Vucic et al. 2014). UBA6 has an important role in brain development as UBA6 knockout mice present not only with a general size reduction, but also morphological defects in the hippocampus and amygdala which play major roles in learning, memory and emotional responses. The mice also display social behavior abnormalities, defects in learning and memory and dendritic spine density reduction (Lee, Dodart et al. 2013). UBA6 knockout mice exhibited a markedly increased abundance of the HECT-type E3 ligase Ube3a and the post-synaptic density protein Shank3 in the amygdala (Lee, Dodart et al. 2013). Ube3a targets as one of its substrates the protein Arc (an activity-regulated cytoskeleton-associated protein), which is required for persistent forms of synaptic plasticity and memory in the mammalian brain (Day and Shepherd 2015). The loss of Uba6 in the amygdala correlates with a dramatic reduction in the steady-state levels of Arc. The proteasomal degradation of Ube3a was observed to be promoted by the UBA6-USE1 cascade in mouse embryonic fibroblasts. This cascade also ubiquitylated Ube3a *in vitro*. Interestingly, these activities occur in parallel with but spatially distinct from the UBA1-UbcH7 cascade (Lee, Dodart et al. 2013). These observations demonstrated a direct involvement of UBA6-mediated proteasomal degradation in brain-associated physiological and pathphysciological states. In patients with attention-deficit hyperactivity disorder and intellectual disabilities, the UBA6 gene was found to be duplicated (Matoso, Melo et al. 2013). In contrast, the gene was found to be deleted in patients with intellectual disability and behavioral disorders (Assawamakin, Wattanasirichaigoon et al. 2012, Matoso, Melo et al. 2013, Shimada, Okamoto et al. 2013, Utine, Haliloglu et al. 2014, Quintela, Barros et al. 2015). Finally, UBA6 was found to be overexpressed in human brains from patients with Alzheimer's disease (Andreev, Petyuk et al. 2012).

UBA6 was found to catalyze the polyubiquitylation and proteasomal degradation of the cytoskeleton linker protein ezrin. Ezrin expression in cells expressing anti-*UBA6* shRNA was localized more diffusely to both cytoplasm and nuclei, and about half of the spheroids, which are acini-like spherical structures, exhibited the mislocalization pattern of ezrin expression. Overexpression of Uba6 in cells expressing anti-*UBA6* shRNA abrogated the phenotypes such as enlarged spheroids and diffuse

subcellular localization of ezrin, indicating that the alterations in cellular morphology and ezrin expression resulted from Uba6 deficiency (Liu, Zhao et al. 2017). The change in ezrin localization correlated with poor prognosis in breast cancer patients (Sarrio, Rodriguez-Pinilla et al. 2006). This observation provides direct evidence of the role of UBA6-mediated proteasomal degradation in breast cancer.

The examples summarized above provide evidence for the involvement of the ubiquitin-UBA6-USE1 proteasomal degradation pathway in different pathophysiological states. One might wonder whether the FAT10-UBA6-USE1 pathway also plays any roles in diseases? Possible answers to this question can be derived by looking at covalent conjugation partners of FAT10. The proteomic analysis of endogenous FAT10 interactors after simulation with pro-inflammatory cytokines identified 571 interacting proteins. These partners are involved in different functional pathways such as autophagy, cell cycle regulation, apoptosis and cancer. Among them, 176 interactors were found to be covalently linked to FAT10 and were classified as putative substrates of the FAT10ylation pathway (Aichem, Kalveram et al. 2012). In another study an almost identical number (175 proteins) of FAT10ylated candidates, which are involved in a broad spectrum of cellular processes, was identified (Leng, Xu et al. 2014). The two proteomic studies demonstrated that a large number of FAT10 putative substrates is comprised of proteins involved in DNA and RNA processing. Among the FAT10-modified target proteins only a few (Table 3) were characterized with respect to the role FAT10ylation exerts on their biological function (Aichem and Groettrup 2020).

As described above FAT10 inhibits SUMOylation by either non-covalently binding to the SUMO E1 adenylation site or by forming a thioester link to the active site cysteine. Consequently, FAT10 downregulates the SUMOylation of target proteins. The transcription factor JunB is one of the SUMO substrates. JunB was also found to be FAT10ylated by UBA6 and this leads to its proteasomal degradation (Aichem, Sailer et al. 2019). The SUMOylation of JunB regulates its ability to induce the transcription of genes encoding cytokines such as IL-2, IL-4, and IL-10 (Garaude, Farras et al. 2008). In other words, FAT10 downregulates the expression of these cytokines. This may explain the high expression levels of FAT10 in cancer cell types where pro-inflammatory cytokines are also expressed (Lukasiak, Schiller et al. 2008). The promyelocytic leukemia (PML) protein is another SUMO substrate and PML SUMOylation was found to be down regulated by FAT10. PML bodies are important regulators of several cellular processes such as DNA repair and genome

maintenance. Therefore, downregulation of PML SUMOylation might favor the transforming capacities of cells expressing FAT10 and promote tumor formation (Aichem, Sailer et al. 2019).

The tumor suppressor protein p53 is another FAT10 target. Li and others showed that FAT10ylation greatly elevated the transcriptional activity of p53 (Li, Santockyte et al. 2011). Furthermore, FAT10 was observed to be overexpressed in many tumor tissues (Lukasiak, Schiller et al. 2008). It remains unclear why the overexpression of FAT10 does not induce apoptosis. Li *et al.* put forward the following arguments: (1) p53 is the most commonly mutated gene in human cancer (Stratton 1992) and FAT10 might fail to activate the mutated forms of p53. (2) FAT10 expression is negatively regulated by p53 (Zhang, Jeang et al. 2006). The overexpression of FAT10 in those tumor cells is likely due to the loss of function of mutated forms of p53. In contrast, in non-transformed cells, native p53 keeps FAT10 expression at a low level where it is hardly detectable except in spleen and thymus (Liu, Pan et al. 1999). In a contradictory report, Choi and others described that stimulation of FAT10 expression by the signal transducer and activator of transcription 3 (STAT3) and nuclear factor κB (NF-κB). They observed the repression of p53 transcriptional activity due to the overexpression of FAT10. Together with the downregulation of FAT10 expression of p53, as mentioned above, this FAT10-p53 double-negative regulation is critical in the control of tumorigenesis. This is in line with the overexpression of FAT10 in many cancer cell types as argued by Choi *et al.* (Choi, Kim et al. 2014). While Li observed that only the conjugation of active FAT10 (FAT10 with the double glycine at the C-terminal tail) enhanced the activity of p53 (Liu, Pan et al. 1999), Choi *et al.* observed that p53 downregulated both active FAT10 or inactive FAT10 could repress the transcriptional activity of p53. The FAT10-p53 cross regulation observed in the studies of Choi *et al.* and Li *et al.* suggest different mechanisms (non-covalent and covalent interactions of p53 and FAT10), however, additional studies will be required to completely unveil the underlying mechanisms.

The autophagosomal receptor p62/SQSTM1 was also reported as a FAT10 substrate and its FAT10ylation leads to proteasomal degradation. As expected, downregulation of either UBA6 or USE1 clearly inhibited the formation of p62-FAT10 conjugates, thus confirming that p62 FAT10ylation was carried out through the UBA6-USE1 cascade (Aichem, Kalveram et al. 2012). p62 plays a key role in signaling functions central to tumor initiation in the epithelium, and suppression of

tumor progression in the stroma (Moscat, Karin et al. 2016). This represents an indirect evidence of the involvement of FAT10ylation in cancer.

UBA1, the major activating enzyme in the ubiquitin proteasome pathway was found to be mono-FAT10ylated under endogenous conditions (Rani, Aichem et al. 2012). FAT10ylation of UBA1 leads to its degradation by the proteasome (Bialas, Groettrup et al. 2015), however, only a small portion of UBA1 is modified and this modification did not decrease overall levels of ubiquitin conjugates. UBA6 alone is sufficient to catalyze the FAT10ylation of UBA1 *in vitro*. The function of the UBA1 FAT10ylation remains somewhat enigmatic (Bialas, Groettrup et al. 2015). On the other side, FAT10 was found to be polyubiquitylated, however, the ubiquitylation of FAT10 does not affect FAT10 degradation (Hipp, Kalveram et al. 2005). This, still, raises questions on the function of ubiquitylated FAT10 and the cross regulation of ubiquitylation and FAT10ylation.

Phosphodiesterase 6 (PDE6), which is expressed in human retina is a pivotal effector enzyme for phototransduction and vision. The non-covalent interaction of PDE6 and FAT10 inhibits PDE6 activity while its covalent interaction leads to the proteasomal degradation of PDE6. The FAT10-dependent downregulation of PDE6 function could cause visual impairment during inflammation (Boehm, Bialas et al. 2020).

The deubiquitylating enzyme otubain 1 (OTUB1) belongs to the ovarian tumor family of cysteine protease (Komander, Clague et al. 2009). It removes polyubiquitin chains form the tumor suppressor p53, thus leading to p53 stabilization (Sun, Challagundla et al. 2012). OTUB1 is also involved in the immune response, DNA damage response and pathogen biology, though its catalytic activity is not always required. FAT10ylation of OTUB1 leads to its degradation by the proteasome (Bialas, Boehm et al. 2019). This provides indirect evidence that FAT10 downregulates p53 levels.

*In vivo* and *in vitro* evidence was provided that the Wnt-induced secreted protein-1 (WISP1) can suppress hepatocellular carcinoma (HCC) cell proliferation. In HCC cells, FAT10 is overexpressed and FAT10ylation facilitated WISP1 degradation by the proteasome, hence, promoted tumor progression (Yan, Lei et al. 2018). The proliferating cell nuclear antigen (PCNA) is associated with DNA replication, chromatin remodeling, DNA damage repair, sister chromatid cohesion and cell cycle control (Kannouche, Wing et al. 2004, de Medina-Redondo and Meraldi 2011).

Expression of both FAT10 and PCNA are upregulated in HCC where FAT10ylation of PCNA leads to its degradation by the proteasome. This suggests that FAT10 may be involved in the DNA damage response and therefore the progression of HCC (Chen, Zhang et al. 2018).

**Table 3. List of confirmed proteins which are covalently modified by FAT10**

| | Degraded by the proteasome | Influence on activity | Reference |
|---|---|---|---|
| P53 | Unknown | Activated | (Li, Santockyte et al. 2011) |
| P62/SQSTM1 | Degraded | Unknown | (Aichem, Kalveram et al. 2012) |
| USE1 | Degraded | No | (Aichem, Pelzer et al. 2010, Aichem, Catone et al. 2014) |
| LRRFIP2 | | Inactivated | (Buchsbaum, Bercovich et al. 2012) |
| UBA1 | Degraded | No | (Rani, Aichem et al. 2012, Bialas, Groettrup et al. 2015) |
| JunB | Degraded | Unknown | (Aichem, Boehm et al. 2019) |
| PCNA | Degraded | Unknown | (Chen, Zhang et al. 2018) |
| OTUB1 | Unknown | Activated | (Bialas, Boehm et al. 2019) |
| WISP1 | Degraded | Increased WISP1 mRNA expression by stabilization of β-catenin | (Yan, Lei et al. 2018) |

| AIPL1 | Unknown | Unknown | (Bett, Kanuga et al. 2012) |
|---|---|---|---|
| SUMO E1 SAE1-UBA2 | Thioester | Inhibited | (Aichem, Sailer et al. 2019) |

### 1.2.2. Targeting E1 enzymes with inhibitors

The ubiquitin proteasome systems (UPS) offers multiple opportunities for therapeutic interventions. The first successfully exploited target in this system was the proteasome itself. Bortezomib is a first-in-class covalent-reversible inhibitor which binds to the active sites of the proteasome, blocking the whole UPS (Adams, Behnke et al. 1998, Groll, Berkers et al. 2006). Proteasome inhibition by bortezomib inhibits the growth and induced apoptosis of multiple myeloma cell lines (Hideshima, Richardson et al. 2001). Another proteasome inhibitor is carfilzomib, which potently binds to one of its active sites and specifically inhibits the chymotrypsin-like activity of both the proteasome and immunoproteasome. The inhibition by carfilzomib is irreversible and it induces a dose- and time-dependent inhibition of proliferation, ultimately leading to the apoptosis of multiple myeloma cells (Kuhn, Chen et al. 2007). These inhibitors have been approved by the United States Food and Drug Association to treat leukemia, multiple myeloma, and some lymphomas (da Silva, Paiva et al. 2013). Additional targets in the UPS are E3 ligases and deubiquitylating (DUB) enzymes, which offer the following interesting properties. Due to the large number of E3 enzymes they would allow for a tailored therapeutic intervention by altering the levels of the relatively small number of substrate proteins targeted by these enzymes. DUB enzymes together with E3 ligases of the HECT and RBR family offer the advantage that their catalytic centers can be directly addressed, in contrast to the large family of RING E3 ligases where inhibitors would need to target sites of protein-protein-interactions. In contrast to the more tailored approach of E3 and DUB inhibitors, inhibition of the proteasome interferes of course with all ubiquitin-mediated degradative processes.

Similar to the proteasome inhibitors compounds targeting the ubiquitin E1 will globally interfere with protein ubiquitylation, a situation, which is less pronounced for other E1 enzymes. Nevertheless, many attempts were made to target the E1 family including UBA1 itself. The first compound reported compound was PYR-41 which inhibits the ubiquitin E1 enzyme and prevents ubiquitin-mediated proteasomal degradation. The mechanism likely involves a covalent link to the

active site cysteine of UBA1, hence blocking E1~ubiquitin thioester formation. PYR-41 led to an accumulation and transcriptional activation of the tumor suppressor p53. Interestingly, PYR-41 enhanced SUMOylation via an unknown mechanism (Yang, Kitagaki et al. 2007). This is in line with the UBA6-p53 negative cross regulation mentioned earlier. Similarly, PYZD-4409 inhibits UBA1 and blocks the degradation of p53. PYZD-4409 induced cell death in malignant cells and preferentially inhibited the clonogenic growth of primary acute myeloid leukemia cells (Xu, Ali et al. 2010). The PYZD-4409 structure is similar to that of PYR-41 with an exocyclic double bond and an N-aryl bond which are potential sites for the nucleophilic attack by the active site cysteine of UBA1. Hence, these two inhibitors might adopt the same inhibitory mechanism. Another inhibitor that inhibits E1-ubiquitin thioester formation is NSC624206 (Ungermannova, Parker et al. 2012), which block the formation of the ubiquitin-thioester in a similar mechanism like PYR-41. The group of adenosine sulfamate inhibitors inhibit the E1 enzymes with the same mechanism called substrate-assisted inhibition in which the respective E1 catalyzes the formation of a covalent UBL-inhibitor adduct which resembles the UBL-AMP intermediate (Soucy, Dick et al. 2010, Chen, Tsu et al. 2011, Nawrocki, Griffin et al. 2012).

MLN4924 was the first reported inhibitor in this class and it selectively inhibits the NEDD8 E1. This inhibitor interacts with the ATP binding site of NEDD8 E1 and forms a covalent adduct that mimics the catalytic NEDD8-AMP intermediate, thus blocking subsequent NEED8-E1 thioester formation (Soucy, Smith et al. 2009). As a consequence, the neddylation of the cullin RING-type E3 ligases is blocked, an essential requirement for full catalytic activity of these E3 enzymes, thus significantly impairing the ubiquitin transfer by these ligases (Duda, Borg et al. 2008). As a result, treatment with MLN4924 led to the accumulation of several oncoproteins and tumor suppressors such as NRF2, p27 and IκB, which are substrates of the culin RING-type E3 ligases (Emanuele, Elia et al. 2011). Indeed, MLN4924 treatment suppressed the growth of human tumor xenografts in mice. (Soucy, Smith et al. 2009). ABPA3 is an adenosyl sulfamate that is covalently linked to ubiquitin and NEDD8, thus, inhibiting ubiquitylation and neddylation. In studies with this compound an upregulation of p53 but also of the cyclin dependent kinase inhibitor p21, resulting in apoptosis of A549 lung cancer cells was observed. Similar to PYR-41, ABPA3 treatment elevated SUMO conjugate levels robustly, although the underlying mechanism still remained uncharacterized (An and Statsyuk 2015). The inhibitor TAS4464, another adenosyl sulfamate, also selectively targets the

33

NEDD8 E1 and inhibits the neddylation of the cullin RING-type E3 ligases. This subsequently increases the accumulation of the E3 substrates, resulting in apoptosis in human cancer cell lines (Yoshimura, Muraoka et al. 2019). To date, several other inhibitors targeting the neddylation of culin E3 ligases have been reported (Yu, Jiang et al. 2020), with the aforementioned MLN4924 (Swords, Erba et al. 2015, Bhatia, Pavlick et al. 2016, Sarantopoulos, Shapiro et al. 2016) and TAS4464 (Yamamoto, Shimizu et al. 2021) having entered clinical trials.

ML-792 is a potent and selective inhibitor of the heterodimeric SAE1-UBA2SUMO E1. As an adenosyl sulfamate, ML-792 inhibits SUMO E1 activity by forming an adduct with SUMO in an ATP-dependent mechanism catalyzed by the SAE1-UBA2 itself. Cell viability in cancer cell lines was decreased upon ML-792 treatment, however, as a side effect of the loss of SUMOylation a failure to complete mitosis and chromosome-segregation defects in the HCT116 human colon carcinoma cells were observed (He, Riceberg et al. 2017). In another study (Lv, Yuan et al. 2018), the compound COH000 was reported to selectively inhibit SUMO E1. The allosteric inhibitor COH000 covalently binds to a cysteine residue in a cryptic pocket at the AAD-IAD interface that is distinct from the active site of SUMO E1. The binding of the inhibitor induced significant conformational changes in the enzyme including the formation of suitable binding pocket. These changes include the disassembly of the ATP binding site and a $180^o$ rotation of the SCCH domain which contains the active site cysteine residue. As a consequence, the SUMO E1 completely lost its adenylation and thioester forming activities. In addition, several natural products with unknown mechanisms were reported which inhibit SUMOylation. These include ginkgolic acid (Fukuda, Ito et al. 2009), davidiin (Takemoto, Kawamura et al. 2014), tannic acid (Suzawa, Miranda et al. 2015), and kerriamycin B (Fukuda, Ito et al. 2009).
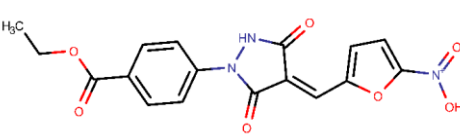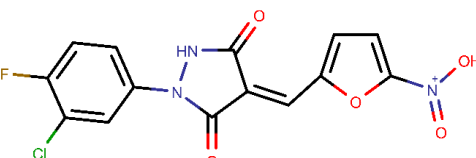
Another adenosine sulfamate, referred to as compound 1, was shown to selectively inhibit UBA6 by forming covalent adducts with both ubiquitin and FAT10. Interestingly compound 1 showed a potent inhibition of ubiquitin activation (half-maximal inhibitory concentration ($IC_{50}$) = 0.15 µmol/L) but only a poor inhibition of FAT10 activation ($IC_{50}$ = 100 µmol/L) (Gavin, Chen et al. 2012). MLN7243 (also known as TAK243) is another adenosyl sulfamate which targets both UBA6 and UBA1 (Misra, Kuhn et al. 2017, Hyer, Milhollen et al. 2018). MLN7243 treatment induced the death of HCT-116 colon cancer cells and demonstrated antitumor activity (Hyer, Milhollen et al. 2018). Since ABPA3 is an adenosine sulfamate
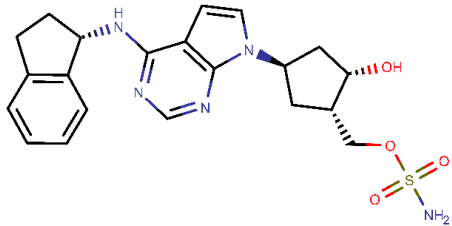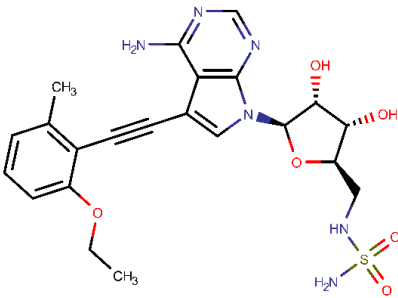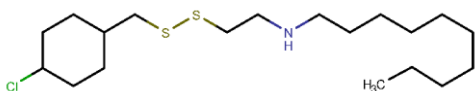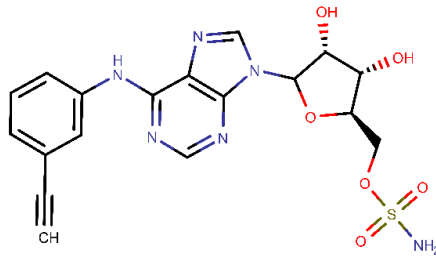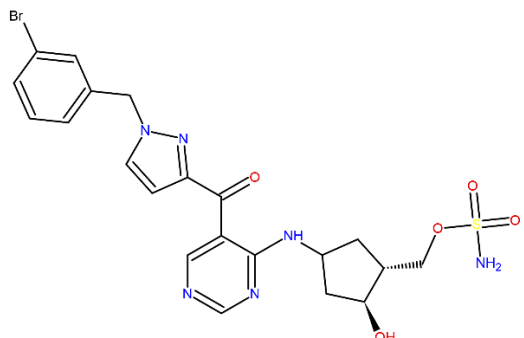
inhibitor and hence forms adducts with ubiquitin (An and Statsyuk 2015, Misra, Kuhn et al. 2017), there is a high probability that ABPA3 will also be covalently linked to FAT10. As stated above, FAT10 was shown to bind to the SUMO E1 and effectively inhibits SUMOylation (Aichem, Sailer et al. 2019). One possible explanation would be that the FAT10-ABPA3 adduct might disrupt this binding, thus, interfering with the SUMOylation inhibition of FAT10. This may explain the upregulation of SUMOylation in A549 lung cancer cell upon ABPA3 treatment (An and Statsyuk 2015).
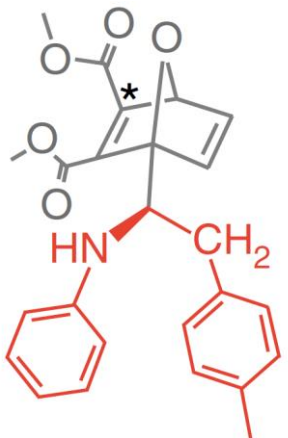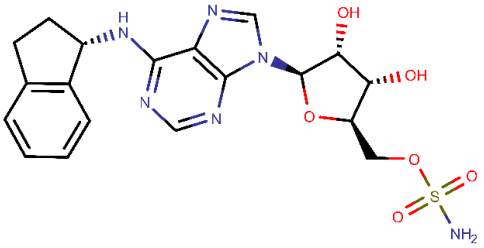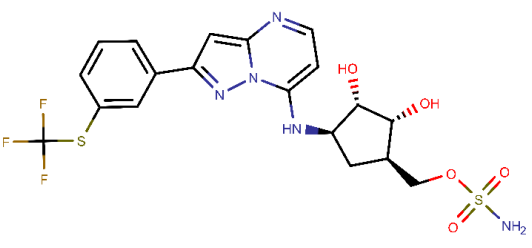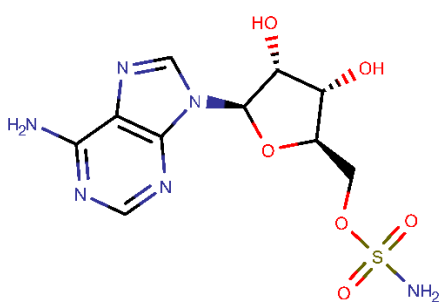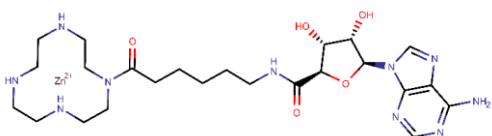
The noncanonical E1 UBA5 was observed to be inhibited by another adenosine sulfamate inhibitor, referred to as ADS, and consequently forms a covalent adduct with Ufm1 (Gavin, Hoar et al. 2014). The adduct tightly binds to the active site of UBA5, thus inhibiting UBA5-mediated cellular effects. Other UBL-ADS adducts were also detected in cells, suggesting a nonselective nature of E1 inhibition of ADS (Gavin, Hoar et al. 2014). Compound 8.5 is an organometallic inhibitor that incorporates adenosine and a zinc(II)-cyclen amide within its core scaffold. Compound 8.5 selectively and noncompetitively with respect to ATP inhibits UBA5 via an unknown mechanism (da Silva, Paiva et al. 2016).
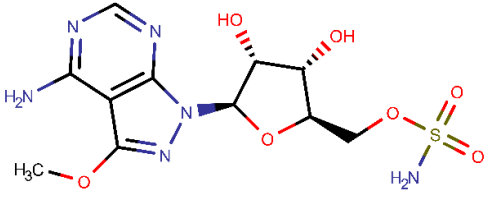
Finally, Huang and others designed and synthesized a pyrazolopyrimidine sulfamate compound that selectively inhibits ATG7 (Huang, Adhikari et al. 2020). Cellular levels of the autophagy markers, LC3B and NBR1, are downregulated following treatment with this compound.

**Table 4. List of selected E1 inhibitors**

| Inhibitor structure and name | Targeted E1 | Inhibition Mechanism | Reference |
|---|---|---|---|
| <br>PYR-41 | Ubiquitin E1 | Potentially covalent linked to active site cysteine | (Yang, Kitagaki et al. 2007). |
| <br>PYZD-4409 | Ubiquitin E1 | Potentially covalently linked to active site cysteine | (Xu, Ali et al. 2010) |

| | NEDD8 E1 | Forming covalent adduct with ubiquitin C-terminus | (Soucy, Smith et al. 2009) |
|---|---|---|---|
| MLN4924 | | | |
| TAS4464 | NEDD8 E1 | Forming covalent adduct with NEDD8 C-terminus | (Yoshimura, Muraoka et al. 2019) |
| NSC624206 | Various E1 enzymes | Covalently linked to active site cysteine | (Ungermannova, Parker et al. 2012) |
| ABPA3 | Various E1 enzymes | Forming covalent adduct with ubiquitin C-terminus | (An and Statsyuk 2015) |
| MLN-792 | SAE1-SAE2 | Forming covalent adduct with SUMO1/2/3 C-terminus | (He, Riceberg et al. 2017) |

| | | SAE1-SAE2 | Allosteric | (Lv, Yuan et al. 2018) |
|---|---|---|---|---|
|  COH000 | | | | |
|  Compound 1 | | UBA1 UBA6 | Forming covalent adduct with ubiquitin C-terminus | (Gavin, Chen et al. 2012) |
|  MLN7243 (TAK243) | | UBA1, UBA6 | Covalent adducts with ubiquitin and FAT10 C-terminus | (Hyer, Milhollen et al. 2018) |
|  ADS | | E1 enzymes | Covalent adduct with C-terminus of Ufm1 | (Gavin, Hoar et al. 2014) |
|  Compound 8.5 | | UBA5 | unknown | (da Silva, Paiva et al. 2016) |

| | ATG7 | Likely forming a covalent adduct with C-terminus of ATG12 | (Huang, Adhikari et al. 2020) |
|---|---|---|---|
| ATG7-Compound 1 | | | |

## 1.3. Aim of this work

Due to the involvement of ubiquitylation and FAT10ylation in various cellular processes in both physiological and pathophysiological setting, it is imperative to study the underlying processes in more detail. Ultimately, the targeting of UBA6 with specific inhibitors would not only enable scientists to study the consequence of blocking the entire FAT10ylation reactions and those ubiquitylation events catalyzed by UBA6, but to also develop novel therapies. In the work presented in this thesis and as a first step towards this long-term objective, this thesis aims to elucidate the mechanism and the determinants for the dual specificity of UBA6 by using X-ray crystallography and biochemical methods.

## 2. MATERIALS AND METHODS

### 2.1. DNA constructs

The human UBA6 gene encoding for residues 35-1052 was cloned into the pGEX-2TK expression vector with glutathione S-transferase (GST) tag at its N-terminus and a His$_6$-tag at its C-terminus using SLIC cloning. The GST tag was fused with UBA6 through the TEV cleavage site. A C-terminally His$_6$-tagged HsUBA1 gene encoding for residues 41-1058 was cloned into the pET23b vector. The FAT10 gene was N-terminally fused to a His$_6$-tagged SUMO gene in the pET28a vector. The gene encoding human ubiquitin was introduced into the pET30a vector without any affinity tag. As for the expression of USE1 (UBE2Z), the construct pET28a-USE1 with His$_6$ tag at the N-terminus was used. All of the genes encoded for the *Homo sapiens* variants of the proteins.https://www.sigmaaldrich.com/life-science/molecular-biology/molecular-biology-products.html?TablePage=111547123

### 2.2. Mutants

All mutants were generated using site-directed mutagenesis and sequence and ligation independent cloning (SLIC). The list of mutants is provided in Table 5.

**Table 5. List of mutants**

|   | Mutant | Vector | Remarks |
|---|--------|--------|---------|
| 1 | C625A | Δ34 HsUBA6-pGEX | Inactivated UBA6 |
| 2 | UBA6-FCCH | Δ34 HsUBA6-pGEX | HsUBA6's residues 203 to 298 were replaced by HsUBA1's residues 206 to 302 |
| 3 | UBA6-SCCH | Δ34 HsUBA6-pGEX | HsUBA6's residues 623 to 889 were replaced by HsUBA1's residues 631 to 889 |
| 4 | UBA6-UFD | Δ34 HsUBA6-pGEX | HsUBA6's residues 942 to 1052 were replaced by HsUBA1's residues 944 to 1058 |

39

| 5 | UBA6-AAD2 | Δ34 *Hs*UBA6-pGEX | *Hs*UBA6's residues 916 to 942 were replaced by *Hs*UBA1's residues 916 to 943 |
|---|---|---|---|
| 6 | UBA6-AAD4 | Δ34 *Hs*UBA6-pGEX | |
| 7 | UBA6-8CS | Δ34 *Hs*UBA6-pGEX | 8 Cysteines (C178, C311, C347, C414, C433, C682, C721, C770) were substituted by Serine |
| 8 | C162L | *Hs*FAT10-pETSUMO | |
| 9 | Cysteine free FAT10 (C0-FAT10) | Δ4 *HsFAT10*-pETM41 | Included the following substitutions: C7T, C9T, C134L, C162S, C160S |
| 10 | E601Q | Δ34 *Hs*UBA6-pGEX | |
| 11 | H614S | Δ34 *Hs*UBA6-pGEX | |
| 12 | H599N | Δ34 *Hs*UBA6-pGEX | |
| 13 | H599N/H614S | Δ34 *Hs*UBA6-pGEX | |
| 14 | M594L | Δ34 *Hs*UBA6-pGEX | |
| 15 | H599N/M594L | Δ34 *Hs*UBA6-pGEX | |
| 16 | Q608E | Δ40 *Hs*UBA1-pET23b | |
| 17 | S621H | Δ40 *Hs*UBA1-pET23b | |

## 2.3. DNA cloning

### 2.3.1. Site directed mutagenesis

The DNA mutations were generated using site-directed mutagenesis. At first, forward and reverse primers were designed to introduce the desired variations of the sequences. Next, 25µl of a mixture containing 0.2 µM of each primer, 1 ng/µl of the DNA template, 0.2 mM of deoxynucleotide triphosphates (dNTPs), 1x Q5 reaction buffer, 0.5 µl of Q5 DNA polymerase (New England Biolabs) and water was prepared. This mixture was then subjected to 25 rounds of the polymerase chain reaction (PCR) in a thermal cycler using the followed parameters: 98 °C denaturation for 10 seconds, 60 °C annealing for 30 seconds, and 72 °C elongation for 30 seconds per 1 kb of template sequence. Thereafter, 1x Cutsmart buffer and 1µl of DpnI digestion enzyme (New England biolabs) were added to the mixture

followed by incubation for 1 hour at 37 °C to destroy the template. The mixture was transformed into *E. coli* DH5α competent cells and plated onto LB agar plates supplemented with the appropriate antibiotic selection marker. In the next step, single and separated colonies were picked and inoculated in LB medium test tubes containing the appropriate antibiotic. Finally, the respective plasmid DNA was isolated using the Mini-prep kit (Fisher Scientific UK) and the DNA sequence was verified by automated DNA sequencing.

### 2.3.2. SLIC cloning

All chimeric UBA6 constructs were engineered using sequence and ligation independent cloning (SLIC). The procedure of SLIC was as follows: Two primers were designed to amplify the DNA fragment of interest, two other primers were used to linearized the plasmid using DNA PCR. The DNA fragment and the linearized plasmid were complementary at their ends. One unit of the DpnI restriction enzyme (New England BioLabs) was added to the PCR samples followed by incubation at 37 °C for 2 hours to destroy the templates. The samples were purified to remove dNTPs. 1000 ng of each the fragment encoding the gene of interest and linearized plasmid were then partially digested using an exonuclease, T4 DNA polymerase (New England Biolabs) to generate single-stranded DNA overhangs. In the next step, 8 µl of the fragment was mixed with 2 µl of the linearized plasmid. The mixture was kept on ice for at least 10 minutes and then transformed into *E. coli* DH5α competent cells. The transformed cells were plated onto LB agar plates containing the appropriate antibiotic. On the next day, several single colonies were picked and used to inoculate 5 ml of LB medium with the respective antibiotic. Finally, the DNA constructs were purified with the bacterial plasmid Miniprep kit (Fisher scientific UK) and DNA sequences were confirmed by automated sequencing.

### 2.4. FAT10 transfer in yeast.

To test the activation and transfer of FAT10 by ScUBA1 *in vivo*, we cloned FAT10 into two different yeast expression vectors, one with the constitutive, strong ADH1 promoter, and the other with the strong MET25 promoter which is induced by lack of methionine in the growth medium. Yeast cells were harvested from the log phase cultures and were transformed with these plasmids or empty vectors as negative controls. Anti-FAT10 Western blots of whole cell lysates were made to check for FAT10 expression and potential FAT10 conjugates. GAPDH was blotted as a loading control.

### 2.5. Protein purification

*Hs*UBA1, UBA6, USE1 and FAT10 proteins were expressed in *E. coli* BL21 (DE3) RIL, *E. coli* soluble BL21 prare, *E. coli* BL21 DE3 and *E. coli* BL21 rosetta II cells, respectively. The cells were grown in LB Broth medium (Carl ROTH) in a shaking incubator to an optical density measured at 600 nm ($OD_{600}$) of 0.6-1. The expression was induced by adding 0.2mM IPTG followed by overnight growth at 15°C (UBA6), 20°C (FAT10 and *Hs*UBA1) and 25 °C (USE1). Typically, 8 l cultures were employed to express the protein of interest. The amount of culture was increased to 32 l when larger amounts of the proteins were needed for crystallization trials. All of the following purification steps were carried out at 4 °C. The cells were harvested by centrifugation (15 min at 7500 x g) and were then resuspended in lysis buffer A (50 mM Tris-HCl pH 8, 300 (FAT10, USE1 and *Hs*UBA1)-500 mM NaCl (UBA6), 5 mM 2-mercaptoethanol, complete protease inhibitor cocktail (EDTA-free, Roche) (1 tablet for 100ml of the lysate) and DNaseI. The lysate was then passed twice through a microfluidizer (M-110P microfluidics) at a pressure of 1.5 bar to break the cell walls. The cell debris was removed by centrifugation for 1 hour at 36000 x g and the supernatant was loaded onto the first affinity column containing the appropriate matrix.

### 2.5.1. Purification of Uba6 and its variants

The resulting supernatants containing UBA6 and its variants, which contained a GST-tag at the N-terminus and a $His_6$-tag at the C-terminus, were passed twice through a 100 ml column containing 10 ml of glutathione agarose beads (Thermo Fisher). After washing the beads with lysis buffer, the proteins were eluted with lysis buffer supplemented with 50 mM reduced L-glutathione (Merk). 200 µl of TEV protease (7 mg/ml) was added to the 100 ml protein samples and the samples were dialyzed against buffer A (50 mM Tris-HCl pH 8, 150 mM NaCl and 5 mM 2-mercaptoethanol) overnight to cleave off the glutathione S-transferase tag. The following day, the protein samples were diluted 1:1 with dilution buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 50 mM imidazole, 5 mM 2-mercaptoethanol) and passed twice through a 100 ml column containing 10ml Ni-NTA agarose beads (Thermo Fisher). The beads were washed with buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 25 mM Imidazole and 5 mM 2-mercaptoethanol). The intact proteins were eluted with elution buffer (50 mM Tris-HCl pH8, 500 mM NaCl, 300 mM imidazole and 5 mM 2-mercaptoethanol). In the final step, the protein samples were concentrated to 5 ml and loaded onto a HiLoad Superdex 16/600 200 pg size exclusion

chromatography (SEC) column (GE Healthcare). The proteins were eluted with 1.2 column volumes in SEC buffer (25mM Tris-HCl pH 8, 500 mM NaCl, 1 mM TCEP (Tris (2-carboxyethyl) phosphin hydrochloric, ROTH)) and collected in 2 ml fractions. 15% SDS PAGE was used to analyze the purity of the elution fractions. Fractions of sufficient purity eluting around 70 ml were pooled and concentrated. The resulting proteins were divided into 40 µl aliquots in thin walled PCR tubes and flash frozen in liquid nitrogen followed by storage at -80 °C.

### 2.5.2. Purification of Uba1 and its variants

HsUBA1 and its variants feature a C-terminal His$_6$-tail. Following cell lysis and centrifugation the resulting supernatants were chromatographed twice over a column containing Ni-NTA agarose beads (Thermo Fisher). Washing buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 25 mM imidazole and 5 mM 2-mercaptoethanol) was used to wash the beads. The proteins were eluted with elution buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 300 mM imidazole and 5 mM 2-mercaptoethanol). The resulting samples were diluted 1:2 with dilution buffer (50 mM Tris-HCl pH 8 and 5 mM 2-mercaptoethanol) resulting in a final NaCl concentration of 100 mM. Uba1 and variants were further purified by ion exchange chromatography using a Mono Q column 10/100 GL (GE Healthcare). A salt gradient from 100 mM to 1 M NaCl was used to elute the proteins. After analyzing the purity of the elution fractions by SDS-PAGE (15 % gels) the selected fractions were pooled and concentrated to 5 ml and loaded onto a HiLoad Superdex 16/600 200 pg SEC column equilibrated with SEC buffer (25 mM Tris-HCl pH 8, 150 mM NaCl and 1 mM TCEP) and stored at -80 °C after concentrating the proteins to 20 mg/ml.

### 2.5.3. Purification of Use1

Use1 proteins feature a His$_6$-tag at their N-terminus. The purification procedures were similar to that of hsUBA1 with the following modifications: The washing buffer was added 50 mM imidazole. The proteins were eluted from the Ni-NTA agarose beads with elution buffers containing elevated concentration of imidazole (100, 200 and 300 mM final concentration). The elution fractions were then evaluated in 15% SDS-PAGE gel. The fractions with less impurities were collected and purified further.

### 2.5.4. Purification of FAT10

The wild type and variants of FAT10 were N-terminally tagged with a His$_6$-SUMO tag. The cells were resuspended in lysis buffer containing 50 mM HEPES pH 7.5

adjusted with 1 M NaOH, 300 mM NaCl, 25 mM imidazole, 5 mM 2-mercaptoethanol, complete cocktail protease inhibitor (1 tablet per 100 ml of buffer and 10 mg of DNAse I per 100 ml of buffer. After breaking the cells and removing cell debris as described in Section 2.4, the supernatants were passed twice through 100 ml column containing 10 ml of Ni-NTA agarose beads, followed by washing with buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 25 mM imidazole, 5 mM 2-mercaptoethanol). The proteins were eluted with the elution buffer containing 300 mM instead of 25 mM imidazole. 500 ul of the SUMO protease ULP1 (1 mg/ml) was added to 100 ml samples which were then dialyzed against buffer (50 mM HEPES 7.5, 100 mM NaCl, 5 mM 2-mercaptoethanol) overnight at 4 °C. On the next day, the samples were filtered and loaded onto a Mono S 10/100 GL ion exchange column (GE health care) and the proteins were eluted using linear salt gradients from 100 mM to 1 M NaCl in 50 mM HEPES 7.5 and 2-mercaptoethanol. SDS-PAGE with 15 % gels was used to analyze the elution fractions. The pooled fractions were concentrated to 5 ml using Millipore Amicon (10 kDa MWCO, Merk) concentrators and loaded on a HiLoad Superdex 16/600 75 pg SEC column which was equilibrated with SEC buffer (25 mM Tris-Cl pH 7.5, 150 mM NaCl, 1 mM TCEP). The proteins were eluted using 150 ml of the buffer at a flowrate of 1 ml/min. The elution fractions which contained a protein band with a molecular weight of ~18.5 kDa proteins were pooled and concentrated to a final concentration of 7 mg/ml. The proteins were divided into 40 μl aliquots and were stored at -80 °C.

## 2.6. Circular dichroism spectroscopy

During the purification of UBA6 two different species eluted from SEC columns. Circular dichroism (CD) spectroscopy was used to check whether both samples were properly folded. 1 μM of each protein was exchanged into 0.1 M sodium phosphate buffer pH 7.5 and pipetted into a transparent quartz cuvette. The measurements were conducted with a JASCO J-810 spectropolarimeter at room temperature. The UV spectra from 190 nm to 250 nm were recorded with a scanning speed of 20 nm/min, a response time of 1 second and a bandwidth of 2 nm. Ten spectra were recorded for each sample and averaged to optimize the signal to noise ratio. The buffer spectrum was also recorded and subtracted from the protein spectra. Excel was used to plot the graphs for analysis.

## 2.7. Activity assays

### 2.7.1. E1 activity assay

To test the functionalities of purified UBA6 and *S. cerevisiae* Uba1 (ScUBA1), 3 µM of the respective E1 enzyme was mixed with either 20 µM of unlabeled FAT10 or 3 µM of ubiquitin, which was labeled with the 800CW infrared fluorescent dye (IRdye, LI-COR), in reaction buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 2.5 mM ATP and 5 mM $MgCl_2$). In the case of UBA1 the activity assay was only conducted with labeled ubiquitin. The reactions were conducted at room temperature for 2 hours. The reactions were stopped by adding protein loading buffer without reducing agent (50 mM Tris-HCl pH 6.8, 2% SDS, 6% glycerol, 0.01% Bromophenol blue). SDS PAGE with 4-20 % gradient gels was used to analyze the results. In the reactions with FAT10, the SDS gels were stained with Coommasie Brilliant Blu G250 in an aqueous solution containing 50% (v/v) ethanol, 10% (v/v) acetic acid, 1% (w/v) Coommasie G250) and destained in a solution containing 50% (v/v) ethanol and 10% (v/v) acetic acid. An additional band corresponding to the E1 thioesterified with FAT10 and exhibiting an ~20 kDa shift above the band corresponding to the unmodified E1 (as determined in the negative control where no ATP was present) indicated the functionality of the respective E1 enzyme (WT or variant). In the reactions with ubiquitin, the gels were scanned with an Odyssey (LICOR, Odyssey, USA) scanner. The protein bands of E1 enzymes which were thioesterified with IRdye labeled Ubiquitin emitted fluorescence at a wavelength of 800 nm.

### 2.7.2. E1-E2 transthioesterification assay

The transfer of activated FAT10 from UBA6 to the E2 enzyme USE1 was analyzed by mixing 9 µM of USE1 with the reaction mixtures of the E1 enzymes UBA6, containing either FAT10 or ubiquitin. The detection of USE1 thioesterified to either FAT10 or ubiquitin was based on Commassie G250 stained 15% SDS PAGE gels taking into account the expected mass difference to the band corresponding to uncharged USE1.

### 2.7.3. Inhibition assays

The reaction mixtures were prepared by mixing 3 µM *Hs*UBA1/UBA6 with 3 µM IRdye labeled ubiquitin/ 20 µM FAT10 and either 2.5 % (final concentration) DMSO or the inhibitor at the same concentration of DMSO in 25 mM Tris. HCl pH 7.5, 2.5 mM ATP, 5mM MgCl2 and 150 mM NaCl. The reactions were occurred at room temperature for one hour. The protein loading buffer without reducing agent and 35 mM EDTA was added to the samples to stop the reactions. The samples were analyzed on 4-20 % SDS-PAGE gels under non-reducing conditions. The bindings

of thioesterified ubiquitin to *Hs*UBA1/UBA6 and thioesterified FAT10 to UBA6 were detected like described in the section 2.6.1.

## 2.8. Crystallization and data collection

Protein samples were prepared by incubating the inactive UBA6$^{C625A}$ variant with either FAT10 or ubiquitin in a 1:1.5 molecular ratio in incubation buffer (25 mM Tris-HCl pH 8, 250 mM NaCl, 1 mM TCEP, 1 mM ATP and 2 mM MgCl$_2$) overnight. After mixing the samples were concentrated to 4 mg/ml and centrifuged at 30000 *x g* for 1 hour at 4 °C. Protein crystals were obtained by vapor diffusion experiments by mixing 400 nl of protein sample with 200 nl of crystallization buffer either in 96-well sitting drop plates using an NT8 crystallization robot (Formulatrix, City, USA) or manually by mixing 1 μl of protein solution and 0.5 μl of reservoir solution in 24-well hanging drop plates. The plates were incubated at 4 °C for up to 4 weeks.

The UBA6-ubiquitin complex was crystallized by incubating the protein mixture against a reservoir solution containing 0.12-0.16 M Ca- Acetate, 0.08 M Na-Cacodylate pH 6.5,13.2-14.4 % w/v PEG 8000, 16-20% v/v glycerol). The crystals were then fished in mother liquor supplement with 30% glycerol as cryo-protectant and flash cooled in liquid nitrogen for subsequent data collection. Initial crystals were tested at beamline 14.1 at the BESSY II synchtrotron (HZB, Berlin, Germany) or beamline P13 at the EMBL outstation in Hamburg at the PETRA III storage ring of DESY (Hamburg, Germany). Two datasets were collected at beamline P13 with crystals obtained under the same conditions.

## 2.9. Structure determination and refinement

The structure was solved by molecular replacement (MR) with Phaser (A.J. McCoy, et al., J. Appl. Cryst. 40, 658-674, 2007) using the yeast Uba1 structure ((Lee and Schindelin 2008); PDB entry 3cmm) as a search model in a sequential domain by domain approach against a dataset belonging to space group P2$_1$22$_1$ collected at beamline P13 at a wavelength of 0.9672 Å. Based on a packing analysis two copies of Uba6 were predicted to be present in the asymmetric unit. In the first round the core of Uba1 consisting of its AAD, IAD and FCCH was used as search model and the resulting structure was refined with Refmac. Subsequently, another round of MR in which the two copies of the correctly placed AAD-IAD-FCCH domains were searched again, followed by a search for two copies of the SCCH domain was conducted and the resulting model was refined again with Refmac. Next, another round of MR was carried out to locate the two copies of the UFD, followed by

refinement with Refmac, which resulted in a model of the UBA6-ubiquitin complex which was complete with the exception of ubiquitin. Based on the crystallization setups ubiquitin was expected to be part of the complex, however, no electron density features corresponding to it could be detected at this stage. After further refinement with Refmac and manual rebuilding with Coot (Ref.) including modeling of the bound ATP, which was clearly visible in the electron density maps, the phases were improved by twofold averaging with DM. The model was extensively rebuilt and refined with Refmac and Phenix (Ref.) in multiple rounds.

At this point the C2 dataset became available and, due to superior diffraction, refinement was continued and completed against this dataset. This dataset was also collected at beamline P14 at a wavelength of 0.9762 Å. Anisotropy of the diffraction data was corrected with an I/$\sigma$I cutoff of 0.8 with the Staraniso server (https:/staraniso.globalphasing.org/cgi-bin./staraniso.cgi) from Global Phasing Limited. The structure was solved by molecular replacement using one copy of the preliminarily refined UBA6 structure from the $P2_122_1$ dataset as search model with Phaser. The $C_2$ dataset was again supposed to also contain ubiquitin, however, there was no indication of its presence in the electron density maps. The structure was finally refined with Buster (G. Bricogne et al., Buster Version 2.10.3) incorporating ncs restraints and TLS refinement where each chain was treated as a single TLS body.

## 2.10.    Structure analysis and visualization

The protein structures were analyzed using The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. The secondary structure depiction was generated using ESPript 3.0 (Robert and Gouet 2014).

## 3. RESULTS AND DISCUSSION

### 3.1. FAT10 purification

FAT10 was reported to be a not highly soluble protein (Aichem, Anders et al. 2018). Therefore, several *E. coli* strains were tested to optimize expression of FAT10 in this organism. The *E. coli* BL21 (DE3) strain contains the λDE3 prophage that encodes the gene for T7 RNA polymerase under control of a lac promoter, thus allowing the expression of T7 RNA polymerase to be induced with IPTG. At the same time, this strain is deficient in the genes encoding the Lon and OmpT proteases. This alteration is supposed to reduce the degradation of the heterologously expressed protein in the bacterial host. *E. coli* BL21 (DE3) and its derivatives were used to test the expression of FAT10. The derivative Rosetta II is designed to enhance the expression of eukaryotic proteins that contain codon, which are rarely used in *E. coli*. This strain supply tRNAs for 7 rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) on a plasmid carrying a chloramphenicol-resistance gene. The tRNA genes are driven by their native promoters. The pLysS strain expresses T7 lysozyme, which suppresses the basal expression of T7 RNA polymerase prior to induction, thus avoiding the leaky expression of the recombinant target proteins that might affect cell growth and viability. ArcticExpress cells encode for additional chaperonins which confer improved protein folding at lower temperatures, thus potentially increasing the yield of active and soluble recombinant proteins, especially when induction is carried out at a reduced temperature. The SoluBL21 pRare competent *E. coli* strain is a significantly improved BL21 host strain for the expression of soluble mammalian proteins.

The expression of human FAT10 was induced by the addition of 0.2 mM IPTG at 20 °C and overnight incubation. Cells were lysed by sonification, and samples of the cell lysate before and after IPTG induction of the *E. coli* host strains were collected and analyzed by SDS-PAGE (Figure 6). Extra bands corresponding to the N-terminally His$_6$ tagged fusion protein SUMO-FAT10 (calculated molecular weight of 34 kDa) were observed in the induced samples of BL21, Rosetta II and ArcticExpress host strains but not in the SoluBL21 and pLysS strains. Rosetta II were chosen for further studies since this host strain is known to support human protein expression and resulted in a higher yield and less impurities than the ArcticExpress and BL21 strains.

For large scale purification, FAT10 expression was induced by the addition of 0.2 mM IPTG, followed by growth at 20 °C for 20 h. Cells were harvested by centrifugation at 4 °C and 5000 g for 15 min. The resulting cell pellet was resuspended in lysis buffer. After lysing the cells by passing them twice through a Microfluidizer and removing the cell debris by centrifugation at 4 °C and 38000 g for 15 min, the cleared lysate was applied to Ni-NTA agarose beads. Ni-NTA (refers to $Ni^{2+}$ ios that have been coupled to nitrilotriacetic acid) agarose is an affinity chromatography matrix for purifying recombinant proteins carrying a His tag. The histidine residues in the $His_6$ tag bind to the vacant positions in the coordination sphere of the immobilized nickel ions with high specificity and affinity. Hence, the His-tagged proteins are retained while other proteins pass through the beads. After washing, His-tagged proteins were eluted with a buffer containing imidazole which competes with the His-tag for binding to the nickel ions. After elution of the fusion protein, the SUMO domain together with the His-tag was cleaved off by adding the SUMO specific protease ULP1 and incubation at 4 °C for 15 hours (Figure 6C). Some aggregation of cleaved FAT10 was observed as the resulting solution was slightly turbid. This is not uncommon since the addition of the highly soluble SUMO protein is known to increase the solubility of the target protein. Following centrifugation to remove insoluble FAT10, the FAT10 protein left in the supernatant was further purified by cation exchange chromatography on a Mono S 10/100 GL column (GE healthcare) since its calculated isoelectric point is 9. FAT10 was finally purified by size exclusion chromatography on a HiLoad 16/60 superdex 75 prep grade (GE healthcare) column. The protein was shown to be clean and homogeneous (Figure 6E and Figure 6G) as judged by a single band on a 15% SDS PAGE and a near-symmetrically shaped peak in the SEC. The final yield of the protein was 2.3 mg per liter of culture medium.
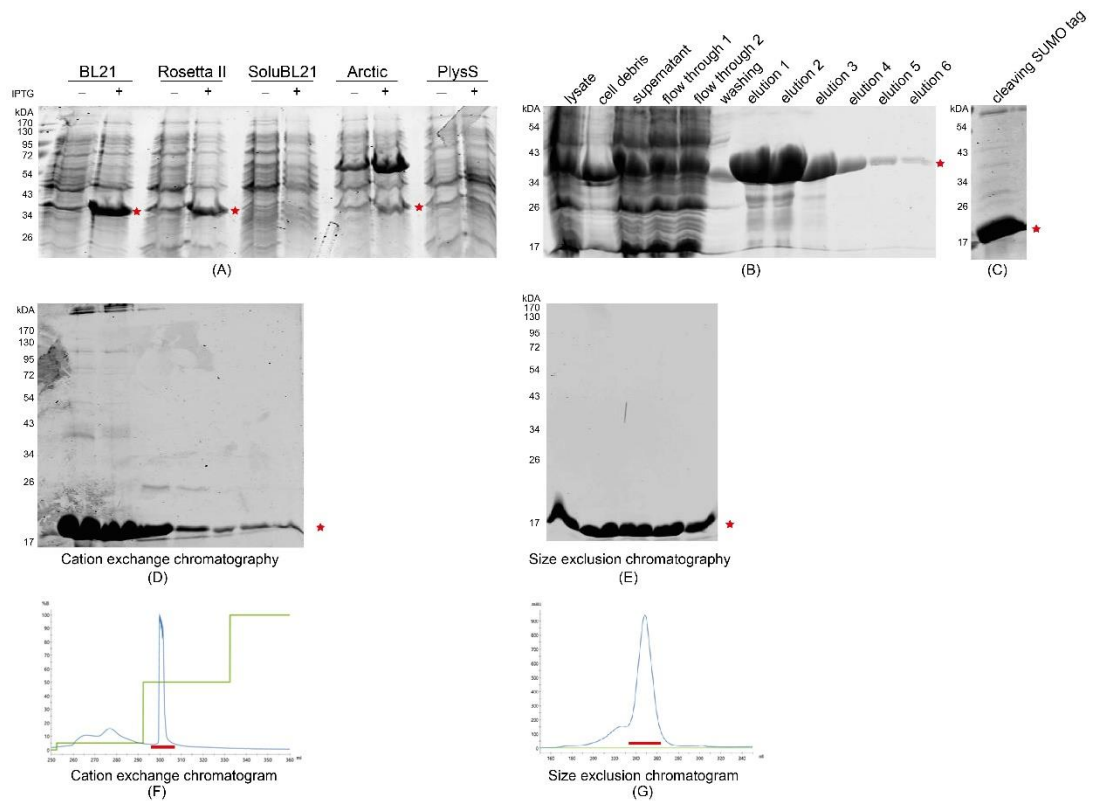
**Figure 6: FAT10 purification. (A)** *E. coli* **expression host strains tested for overexpression before (-) and after (+) addition of IPTG. (B) Fractions from the Ni-NTA affinity chromatography step. (C) SDS PAGE of FAT10 after cleavage of the SUMO tag by TEV protease. The SUMO tag is not visible because it overlapped with FAT10. (D and E) SDS-PAGE analyses of the cation exchange and size exclusion chromatography steps with (F and G) the corresponding chromatograms of the cation exchange and size exclusion chromatography steps. The position of molecular weight marker proteins is marked on the left for all gels (panels A-E).**

## 3.2. USE1 purification

Full-length human USE1 was purified following an established protocol (Schelpe, Monte et al. 2016) using *E. coli* BL21 DE3 cells as expression host. Expression of N-terminally His$_6$ tagged USE1 was induced by the addition of 0.2 mM IPTG at 25 °C (Figure 7A) followed by overnight growth. Cells were harvested by centrifugation at 5000 g for 15 min at 4 °C and lysed by passing them twice through a microfluidizer. Another centrifugation step (38000 g, 1 h at 4 °C) eliminated unbroken cells and cells debris. USE1 purification was initiated by Ni-NTA affinity chromatography (Figure 7B). After flowing the supernatant from the last centrifugation step through the Ni-NTA agarose beads, the beads were washed

thoroughly with increasing NaCl concentrations (100 mM to 400 mM) to get rid of unspecifically bound proteins including a possible *E. coli* chaperon with a molecular weight of ~70 kDa. The His$_6$-USE1 protein was eluted with buffers containing 100 mM, 200 mM and 300 mM imidazole. The samples from the 200 mM and 300 mM imidazole elutions were collected, combined and further purified by anion exchange chromatography on a Mono Q 10/100 GL column (GE healthcare) and size exclusion chromatography on a HiLoad 16/60 superdex 200 prep grade column (GE healthcare). The resulting protein migrated as a single band on an SDS PAGE. The final yield of full-length USE1 was 3 mg of protein per liter of culture medium. An identical purification procedure was used to purify a truncated version of USE1 (Δ99USE1) in which 99 residues of an N-terminal extension containing residues, which are predicted to be unstructured, were not present. In this case the yield was slightly higher with 3.4 mg of protein per liter of culture. The Δ99USE1 was generated to be used in future structural studies of the UBA6-USE1 complex.
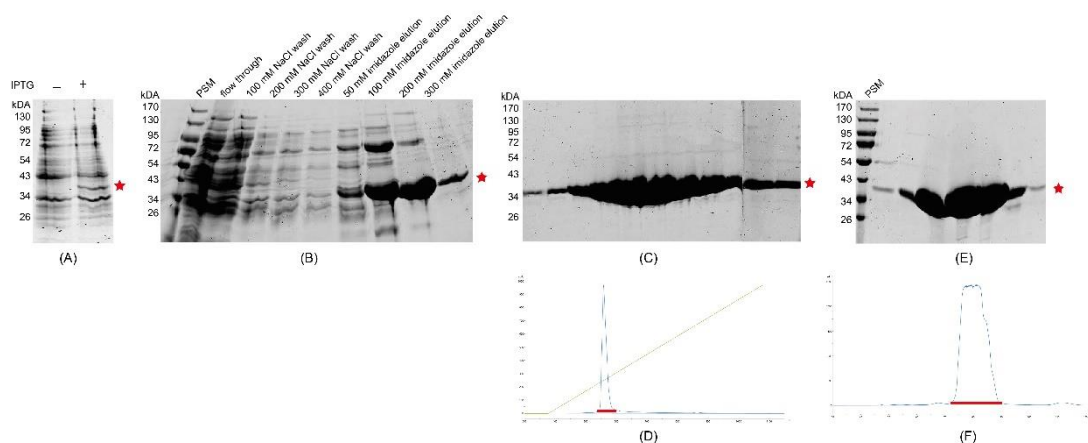


**Figure 7: USE1 purification. (A) Induction of full-length USE1 in *E. coli* BL21 DE3 cells after IPTG addition. (B) SDS-PAGE analysis of the Ni-NTA affinity chromatography step. (C and D) SDS-PAGE analysis and chromatogram of the anion exchange chromatography step. (E and F) SDS-PAGE analysis and chromatogram of the size exclusion chromatography step. The USE1 band is marked by the red star on each gel. Positions of marker proteins are indicated on the left in (A-C), while the actual bands of the pre-stained protein marker (PSM) mix is shown in (E).**

## 3.3. UBA6 purification

The human *uba6* gene was cloned into two expression vectors: (1) A pBADM11 vector with an araBAD promoter where protein expression can be induced by the addition of arabinose. The full-length UBA6 gene was cloned into this vector with a

TEV cleavage site and a His$_6$ tag at the N-terminus. (2) A pGEX-2TK vector with an IPTG-inducible lac promoter containing a glutathione S-transferase (GST) tag fused to the N-terminus of UBA6 including a TEV cleavage site in the connecting linker. In this case the first 34 residues (Δ34UBA6), which, based on secondary structure predictions and in analogy with other ubiquitin activating enzymes, were presumed to be disordered, were omitted.

### 3.3.1. *E. coli* expressions tests

UBA6 is a rather large protein with a molecular weight of 118 kDa for the full-length protein or 114 kDa for the N-terminally truncated version. Forcing high level expression of UBA6 in bacteria might hence cause misfolding and instability of the target protein. The two UBA6 expression vectors were transformed into several *E. coli* host strains to analyze expression levels. While pBADM11 vector did not express UBA6 following arabinose induction in all five *E. coli* strains tested, the pGEX vector expressed GST-fused UBA6 in ArticExpress, SoluBL21 pRare, Rosetta II and pLysS host strains (Figure 8), with the Arctic and Rosetta II strains displaying higher expressions levels.
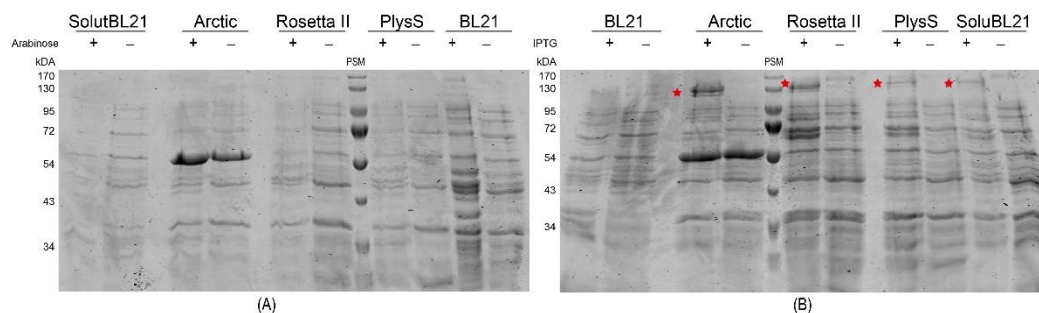


**Figure 8: UBA6 expression tests. (A) Expression of UBA6 from the pBADM11 vector. (B) Expression of Δ34UBA6 from the pGEXvector. Red asterisk indicates the bands corresponding to UBA6.**

### 3.3.2. UBA6 expression in the BL21 Rosetta II strain

UBA6 was initially purified from the Rosetta II strain. Protein expression was induced with 0.2 mM IPTG at 20 °C in LB medium. The cells were harvested by centrifugation at 4 °C for 15 min. The resulting cell pellet was resuspended in lysis buffer (50 mM Tris. HCl pH 8, 300 mM NaCl, 5 mM 2-mercaptoethanol). After lysing the cells by passing them twice through a Microfluidizer and removing the cell debris by centrifugation at 38000 g for 60 min at 4 °C, the cleared lysate was applied to a glutathione agarose column. The bound GST-UBA6 fusion protein was eluted using a buffer containing reduced glutathione (GSH). Analysis of the elution fractions by

SDS PAGE revealed very high amounts of degradation products with molecular weights of ~72 kDa, ~54 kDa and ~30 kDa in the elution fractions (Figure 9) with a substantially weaker band of the full-length protein at ~130 kDa. Following cleavage of the GST tag with TEV protease (Figure 9) the full-length protein as well as the degradation bands displayed the expected reduction in size. UBA6 was further purified by anion exchange chromatography with a Q Sepharose fast flow column (GE healthcare) and size exclusion chromatography with a HiLoad 16/60 superdex 200 prep grade column (GE healthcare). Due to the high levels of degradation products and the correspondingly low amount of the full-length protein only insufficient quantities of UBA6 in a highly impure state could be obtained (Figure 9C and Figure 9D).
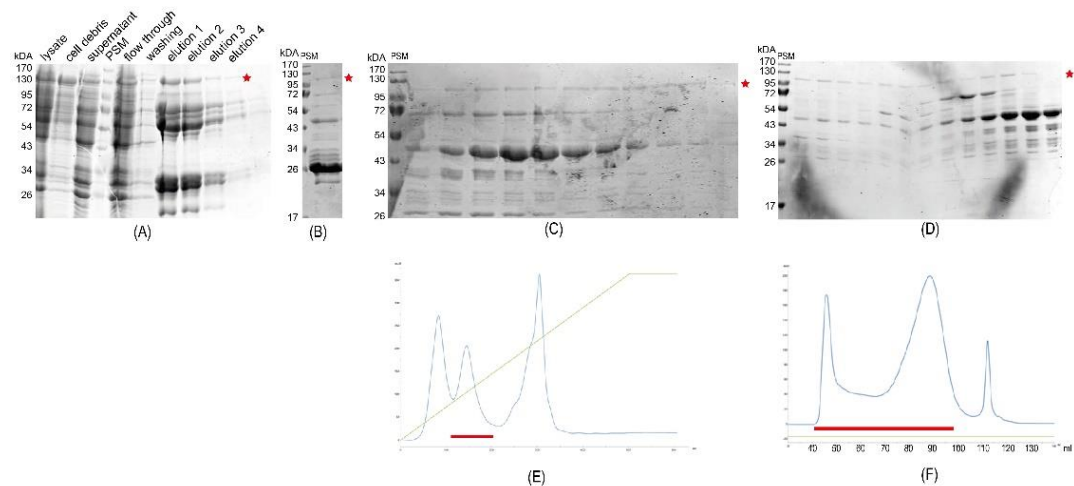


**Figure 9: UBA6 purification with protein expressed in *E. coli* BL21 Rosetta II cells. (A) SDS PAGE analysis of fractions from the glutathione affinity chromatography step. (B) Cleavage of the GST tag using TEV protease. (C and E) SDS-PAGE analysis and chromatogram of the ion exchange chromatography step. (D and F) SDS-PAGE analysis and chromatogram of the size exclusion chromatography step.**

### 3.3.3. UBA6 expression in the *E. coli* BL21 ArcticExpress strain

The ArcticExpress cells were expected to support the proper folding of UBA6 at the lower temperature typically used for the expression of recombinant proteins in *E. coli*. Cell growth took place initially at 30 °C until an $OD_{600}$ of 0.8~1 was reached. The temperature was then decreased to 13 °C and 0.2 mM IPTG was added to the culture to induce the expression of UBA6. Under these conditions UBA6 clearly demonstrated enhanced stability resulting in fewer degradation products in the glutathione affinity chromatography step (Figure 10A) compared to the purification

from the Rosetta II host strain. The subsequent removal of the GST-tag with TEV protease displayed the expected decrease in mass with the appearance of a prominent band corresponding to GST. Unfortunately, most of UBA6 remained bound to the Q Sepharose fast flow anion exchange chromatography column, which could only be eluted at a very high salt concentration of 2 M NaCl, while only a small portion of the protein eluted at a salt concentration of 400 mM within the standard salt gradient (Figure 10D). This minor fraction was concentrated and the protein applied to an HiLoad 16/60 superdex 200 prep grade size exclusion chromatography column, however, no UBA6 could be detected in the elution fractions (Figure 10E).
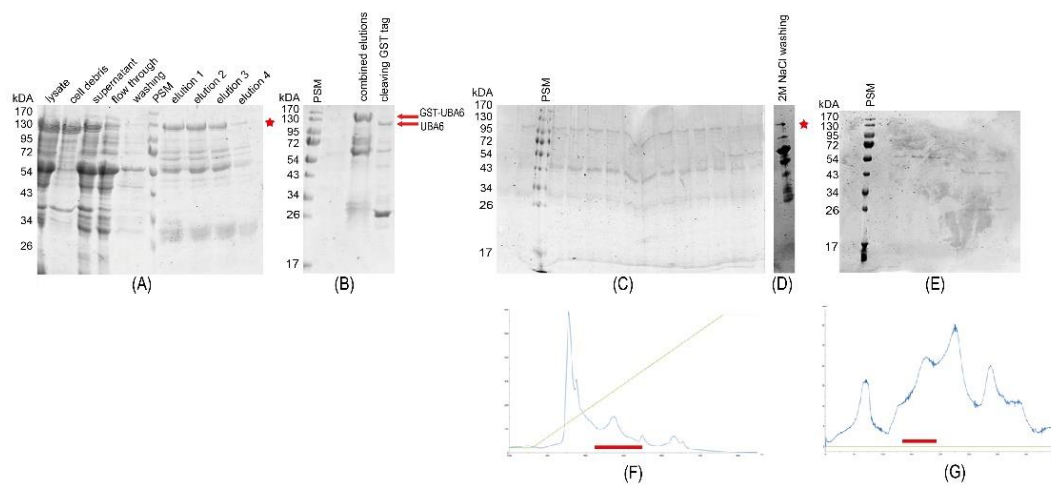


**Figure 10: UBA6 purification from *E. coli* BL21 ArcticExpress cells. (A) SDS PAGE analysis of fractions from the glutathione affinity chromatography. (B) Cleavage of the GST tag using TEV protease. (C and F) SDS-PAGE analysis and chromatogram of the ion exchange chromatography step. (D) Eluting proteins following a wash step with 2 M NaCl. (E and G) SDS-PAGE analysis and chromatogram of the size exclusion chromatography.**

### 3.3.4. UBA6 expression in the *E. coli* SoluBL21 pRare strain

Due to the difficulties with expression and/or purification in the aforementioned strains, the SoluBL21 pRare strain was studied in more detail. UBA6 expression was induced with 0.2 mM IPTG at 20 °C in LB medium at an $OD_{600}$ of 0.8-1. Cells were harvested and lysed and then protein purified as described before. The GST affinity chromatography showed a good amount of GST tagged UBA6, nevertheless substantial amounts of different degradation products with approximate molecular weights of 72 kDa , 54 kDa and 26 kDa remained (Figure 11A). During anion exchange chromatography the protein eluted in a major peak at a salt concentration

of 300 mM and a minor peak at a salt concentration of approximately 500 mM (Figure 11D and Figure 11F), while no protein remained bound to the column as the high salt wash fraction was devoid of UBA6. Following concentration of the protein to 7 mg/ml by ultrafiltration with a 50 kDa membrane cutoff, size exclusion chromatography with a HiLoad Superdex 200 16/60 prep grade (GE healthcare) revealed three bands at elution volumes of 65 ml, 75 ml and 90 ml, respecitvely. The first peak (highlighted in green) eluted right after the column void volume of 40 ml possibly suggesting that it represents an aggregation product. The species in the second peak (highlighted in orange) also contained Uba6 with minor impurities, possibly corresponding to UBA6 monomers and the third peak only contained impurities and/or degradation products which showed smaller molecular weighst (Figure 11D and Figure 11F). The respective yield for the second peak was quite low with ~0.2 mg per liter of culture.
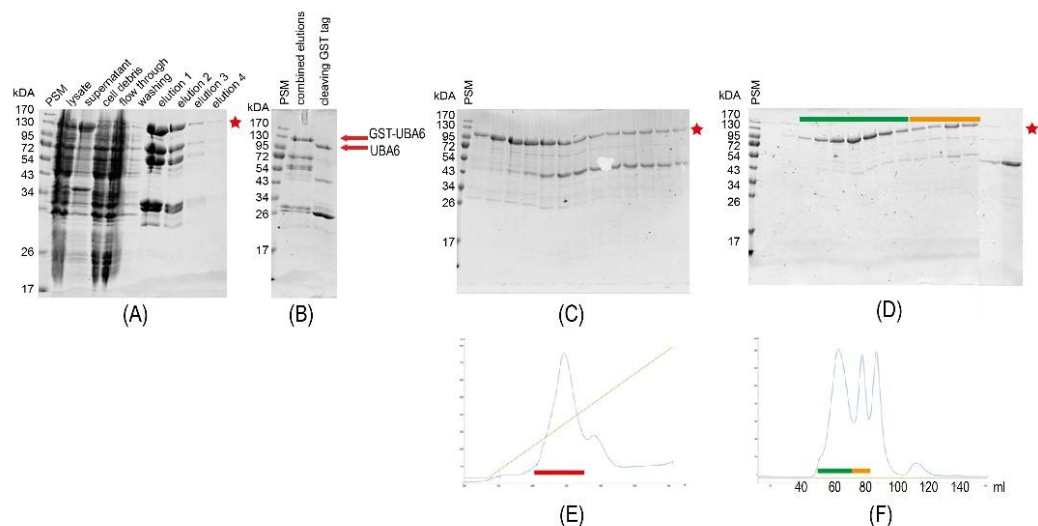


**Figure 11: UBA6 purification after expression in the *E. coli* SoluBL21 pRare strains. (A) Glutathione affinity chromatography. (B) Cleavage of the GST tag using TEV protease. (C and E) SDS-PAGE gel analysis and chromatogram of the ion exchange chromatography. (D and F) SDS-PAGE gel analysis and chromatogram of the size exclusion chromatography.**

Up to this point, all UBA6 expressions were induced with 0.2 mM IPTG. To investigate the influence of IPTG concentrations on the expression level, increasing as well as decreasing concentrations ranging from 0.05 mM to 1 mM IPTG were examined, however, no significant change in the amounts of GST-tagged UBA6 could be detected (Figure 12A). Interestingly, changing the amount of IPTG

negatively affected the amount of cell pellet at higher IPTG concentrations and possibly also at the lowest concentration tested (Figure 12B). Next the expression levels of UBA6 in LB and TB medium were compared. From the same culture volume, TB medium produced nearly the double amount of cell pellet, however, the quantity of UBA6 seemed less when compared with that obtained with LB medium (Figure 12C). Finally, different expression temperatures were analyzed which revealed maximal UBA6 levels at 15 °C (Figure 12D) with 20 °C resulting in the second highest yield. This is consistent with the general concept of the ArcticExpress strain, which at least in this case cannot be further exploited by going to even lower temperatures. In summary, after optimization the temperature after induction was changed from 20 °C to 15 °C.
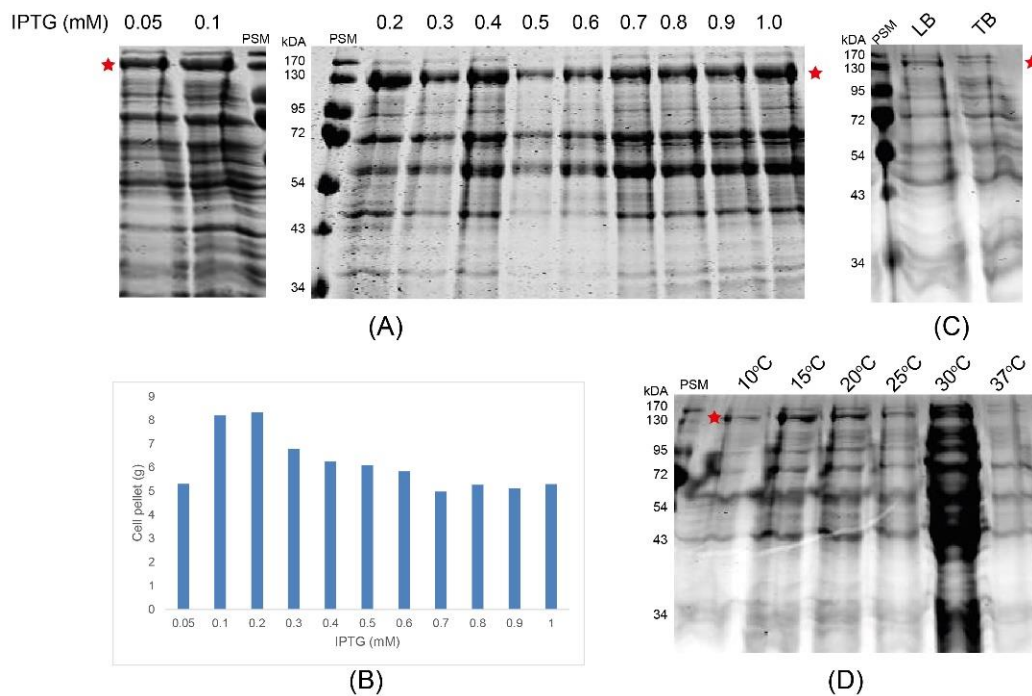


**Figure 12: UBA6 expression optimization in the *E. coli* ArcticExpress strain. (A) Optimization of IPTG concentration. (B) The amount of cell pellet as a function of IPTG optimization. (C) Comparison of LB and TB media. (D) Optimization of induction temperature.**

Examining the shifts of protein bands on SDS PAGE corresponding to full-length UBA6 and its degradation products following removal of the GST tag (Figure 10B), it is evident that not only the full-length protein but also its degradation products were shortened by removal of the GST tag. This suggested that addition of a second affinity tag at the C-terminus should allow the elimination of the degradation

56

products as they would not contain the second tag and hence would be unable to bind to the affinity. Hence, a new UBA6 construct was engineered with contained a $His_6$ tag at the C-terminus of UBA6, resulting in the GST-UBA6- $His_6$ construct. By using tandem affinity chromatography to sequentially trap the GST and $His_6$ tags, the purification of UBA6 could be substantially improved. In order to optimize the purification procedure, two protocols were tested. In the first (Figure 13B), the cell lysate was first passed through the Ni-NTA agarose beads and the bound GST-UBA6-$His_6$ was then eluted using buffer containing imidazole. Subsequently, the protein sample was applied to glutathione agarose beads and the bound proteins were eluted using buffer containing reduced GSH and, finally, TEV protease was added to the sample to cleave the GST tag (Figure 13B). In the second approach (Figure 13A), GST-UBA6-$His_6$ was first purified by glutathione affinity chromatography, followed by removal of the GST-tag with TEV protease. Next the resulting UBA6-$His_6$ construct was further purified by Ni-affinity chromatography. As the second method yielded significantly reduced levels of impurities (compare the last lanes of the SDS-PAGE gels in Figure 13A and Figure 13B), it was chosen as the standard protocol. After tandem affinity chromatography the protein was further purified by ion exchange and size exclusion chromatography (Figure 13C-F). SDS PAGE analysis clearly documented a substantial improvement in protein quality (Figure 13E). While the elution profile from the anion exchange chromatography column differed significantly with UBA6 eluting as a clearly separate peak at a salt concentration of 250 mM, UBA6 still eluted in more than a single peak from the size exclusion chromatography column. The first peak still appeared right after the void volume (Figure 13F), thus again presumably corresponding to aggregated UBA6. The remainder of UBA6 eluted as a second major peak at a volume of 68 ml with a shoulder at a smaller elution volume of 63 ml. Protein fractions across the entire second peak were pooled resulting in a yield of 0.2 mg per liter of culture, which was unchanged compared to the previous protocol, however, the protein quality was significantly improved (compare fractions highlighted in orange on Figure 11D and Figure 13E).
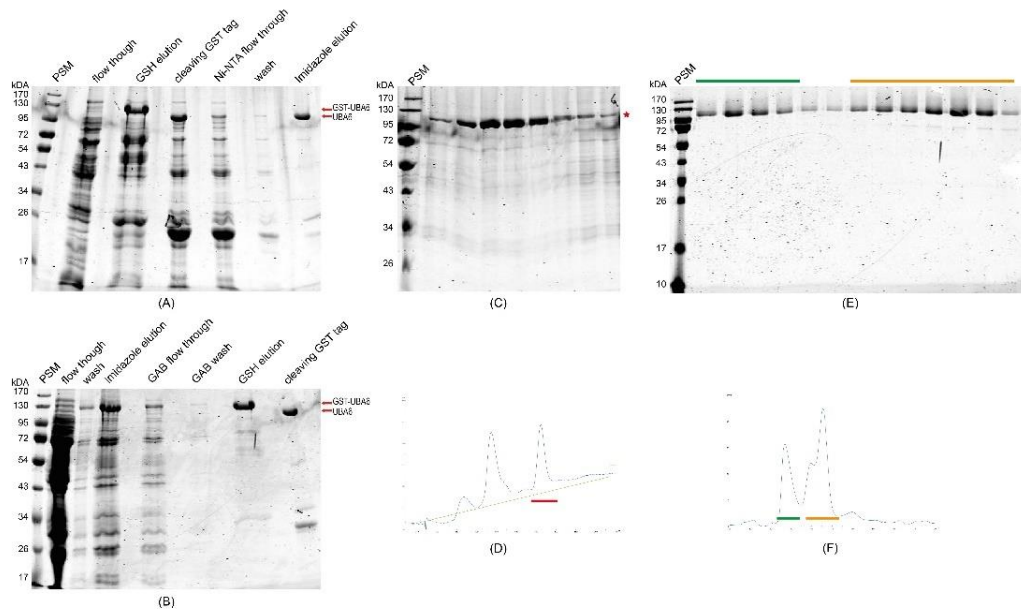
**Figure 13: Purification of the GST-UBA6-His$_6$ construct. (A) UBA6 tandem affinity chromatography on glutathione agarose and Ni-NTA agarose beads. (B) UBA6 tandem affinity chromatography on Ni-NTA agarose and glutathione agarose beads. (C and D) SDS-PAGE analysis of the fractions after and chromatogram of the ion exchange chromatography step. (E and F) SDS-PAGE analysis of the fractions after and chromatogram of the size exclusion chromatography step.**

### 3.3.5. Biophysical and enzymatic characterization of purified UBA6

To generate enough UBA6 for further characterization and crystallization, protein expression was scaled up from 16 liter to 48 liters of LB medium. At the same time, the ion exchange chromatography step was eliminated since it did not have a significant impact on protein purity (Figure 14). With this modification the size exclusion chromatography still showed two distinct peaks containing UBA6 (Figure 14A and Figure 14B), however, the second peak eluting at 67 ml presumably corresponding to soluble UBA6 no longer showed a shoulder at lower elution volumes, while the first peak, presumably still corresponding to aggregated UBA6, which eluted at 45 ml, now exhibited a shoulder at higher elution volumes (Figure 14B). With this final modification the yield of UBA6 could be improved to 0.3 mg per liter of cell culture.

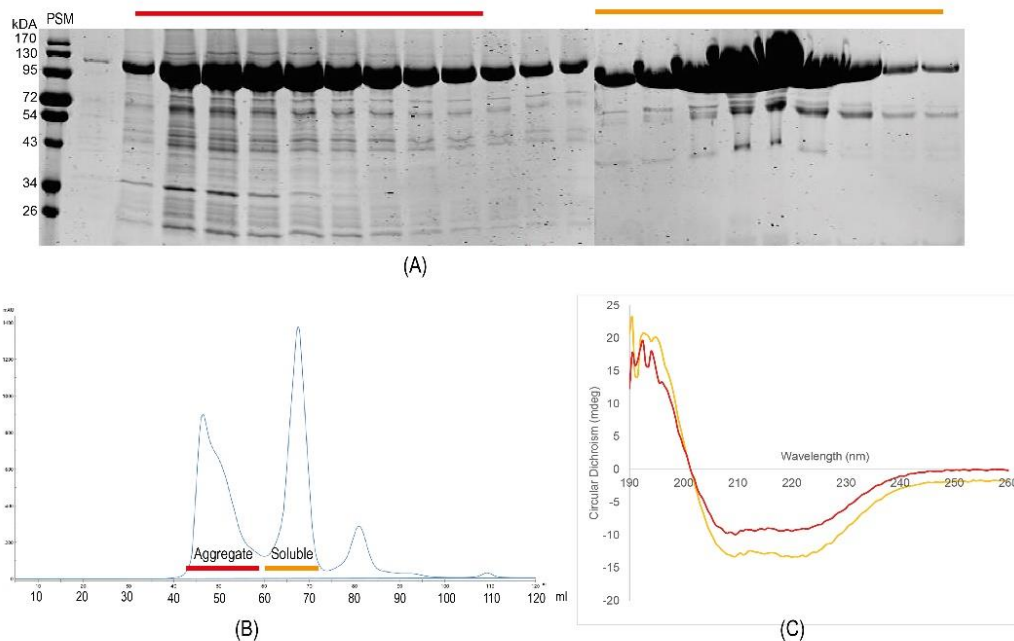**Figure 14: (A) SDS-PAGE analysis of UBA6 after size exclusion chromatography. (B) Chromatogram of the size exclusion chromatography indicating the fractions analyzed in (A). (C) Circular dichroism spectra of two protein species after size exclusion chromatography.**

To investigate possible differences between the two UBA6 fractions (aggregated and soluble protein) the two samples were analyzed by circular dichroism (CD) spectroscopy. CD spectroscopy is a technique which evaluates the secondary structure and folding properties of proteins. Following the exchange of UBA6 into buffer containing 133 mM $Na_2HPO_4$ and 133 mM $KH_2PO_4$ CD spectra were recorded in the range from 190 nm to 260 nm. The two UBA6 samples showed very similar profiles with the differences in the intensity of the double minimum at 210 nm and 225 nm presumably resulting from differences in protein concentration. The observed CD spectra are typical for proteins with an α-helical content of ~50%. Furthermore, the CD spectra indicate that both UBA6 samples are properly folded (Figure 14C).

As the CD analysis did not reveal obvious differences between the two UBA6 samples, the protein batches were further analyzed by a combination of size exclusion chromatography and multi-angle light scattering (SEC-MALS). In SEC-MALS, the SEC column serves to separate the protein molecules by size, more precisely by hydrodynamic radius. After exiting the column, the molecules pass through a MALS detector and are probed by a laser beam. The MALS signals, along

59

with the protein concentration, which is measured by UV absorbance and/or the differential refractive index (dRI) signal, are analyzed to measure the absolute molar mass of the molecules. The SEC column was a Superose 6 Increase 10/300 GL column (GE healthcare), which can separate molecules ranging in size from 5 kDa to 5000 kDa. This column thus covers a broader size range than the Superdex 200 column employed during protein purification, which is suited for the separation of molecules in the range from 10 kDa to 600 kDa. The aggregated UBA6 sample exited the SEC column very early, starting at an elution volume of 8 ml (Figure 15A) volume while the soluble UBA6 sample eluted from the column beginning at a volume of 14 ml leading into a small shoulder at 15 ml and a single major peak at 16 ml (Figure 15B). While it was not possible to deduce a molar mass for aggregated UBA6, thus suggesting considerable sample heterogeneity (Figure 15A), the soluble UBA6 species yielded a molar mass signal at 117 kDA across the entire main peak corresponding very accurately the calculated molecular weight of 117 kDa for the slightly truncated version of UBA6 studied (Figure 15B). A higher molecular weight of 230 kDa was obtained for the shoulder, best visible in Figure 15D, which most likely corresponds to a UBA6 dimer.

Based on these initial results SEC-MALS was used to optimize the buffer conditions to increase the fraction of soluble UBA6. Increasing the concentration of the reducing agent TCEP showed hardly any impact. In buffer containing 5 mM TCEP the shoulder was reduced (Figure 15C), however, the elution peak became less symmetric. Somewhat surprisingly, at 10 mM TCEP the shoulder reappeared, even resulting in a minor peak, while the major peak again appeared very symmetrical (Figure 15D). Decreasing the pH to either 6.5, which also necessitated the use of Bis-Tris or Hepes instead of Tris-HCl buffer, or increasing the pH to 9 while maintaining Tris as buffer complement or to pH 10 with glycine as buffer substance was detrimental to protein quality. Especially at acidic pH, irrespective of the buffer substance, the shoulder became much more prominent. While this trend was less pronounced at pH 9 and 10, the chromatograms were clearly inferior to those obtained at pH 8.5 (Figure 15E-H). In contrast, increasing the salt concentration favorably affected protein homogeneity (Figure 15I and Figure 15J), since in the presence of 500 mM NaCl in the buffer the shoulder was reduced while still retaining an almost perfectly symmetric peak with a molar mass signal corresponding to a UBA6 monomer (Figure 15J). Consequently, 500 mM NaCl was added to all buffers during UBA6 purification. Under these conditions preparative size exclusion chromatography runs showed an improved resolution and the quantity of soluble

UBA6 increased at the expense of the aggregated form of UBA6 (Figure 16A and Figure 16B). All subsequent UBA6 purifications were conducted with this final protocol. Under these conditions the protein yield was significantly increased, from 0.3 mg to 1.2 mg per liter culture.
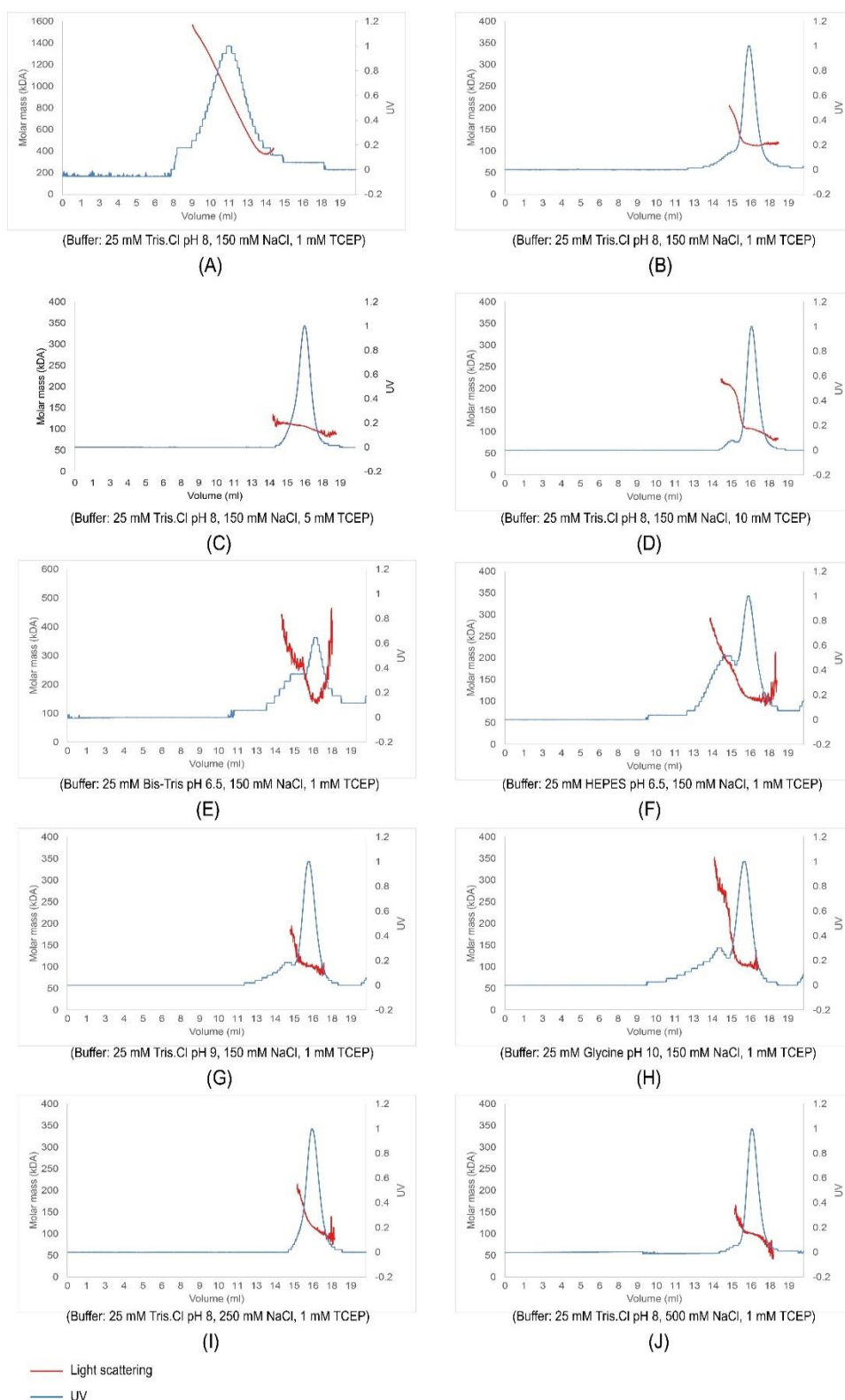
**Figure 15: SEC-MALS analyses of UBA6 in different buffer conditions. Buffer compositions are indicated below each panel.**
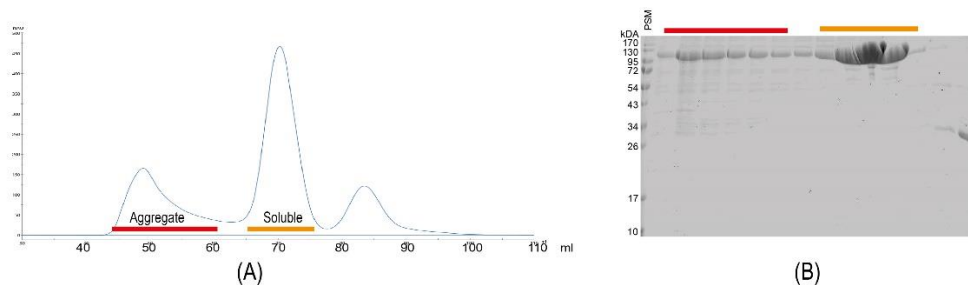
**Figure 16: Final purification step of UBA6 after buffer optimization. (A) Chromatogram of a size exclusion chromatography with a HiLoad 16/60 superdex 200 prep grade column. (B) SDS-PAGE analysis of the fractions from the HiLoad 16/60 superdex 200 prep grade column.**

Besides the biophysical characterization of UBA6 and the ensuing steps which led to a further increase in its yield and quantity it was also essential to ensure that the purified protein retained catalytic activity. Hence activity assays in which the charging of UBA6 with either ubiquitin or FAT10 and the subsequent transfer of the modifier to an E2 enzymes was analyzed. In this assay either ubiquitin or FAT10 will be first adenylated at their respective C-terminus followed by the attack of the active site cysteine (Cys625) on the C-terminal glycine of the ubiquitin/FAT10 adenylate and its covalent attachment to the catalytic cysteine via a thioester bond. This covalent intermediate can be visualized by SDS PAGE provided the sample buffer does not contain reducing agents and due to the addition of either 8.5 kDa in case of ubiquitin or 17.5 kDa in the case of FAT10 an upward shift of the UBA6 band will be visible in the gel. The charging assay can be modified to a transthioesterification assay by adding a cognate E2 enzyme, which will accept the activated protein modifier (ubiquitin or FAT10) on its active site cysteine. Again, a shift in the migration pattern of the E2 enzyme will indicate the formation of the covalent E2~ubiquitin/FAT10 adduct. In the case of the ubiquitin assays the resulting E1~ubiquitin and E2~ubiquitin complexes can be easily visualized with a commercially available, fluorescently labeled ubiquitin derivative in which the near-infrared fluorescence IRDye 800CW NHS dye (LI-COR) is randomly attached to the side chain $NH_2$-groups of lysine residues. In the case of FAT10 such a dye-modified variant is not available and detection was carried out by staining with Coomassie Brilliant Blue G250, which, of course not only labels the desired E1~FAT10 and E2~FAT10 adducts but also the non-modified E1 and E2 enzymes.

As can be seen (Figure 17, right panel) in the presence of UBA6, ubiquitin and Mg-ATP the E1-ubiquitin adduct is readily formed (lane 2) while this does not occur in

the absence of Mg-ATP (lane 1). Addition of either full-length USE1 (lane3) or the N-terminally truncated Δ99USE1 results in the respective E2~ubiquitin adduct, which, due to their smaller size migrates farther into the gel. Carrying out the assay with unlabeled FAT10 (Figure 17, right panel) likewise results in the formation of the UBA6~FAT10 adduct (lane 2), which does not occur in the absence of Mg-ATP. As before, addition of full-length USE1 or Δ99USE1 results in the formation of the respective E2~FAT10 adducts.
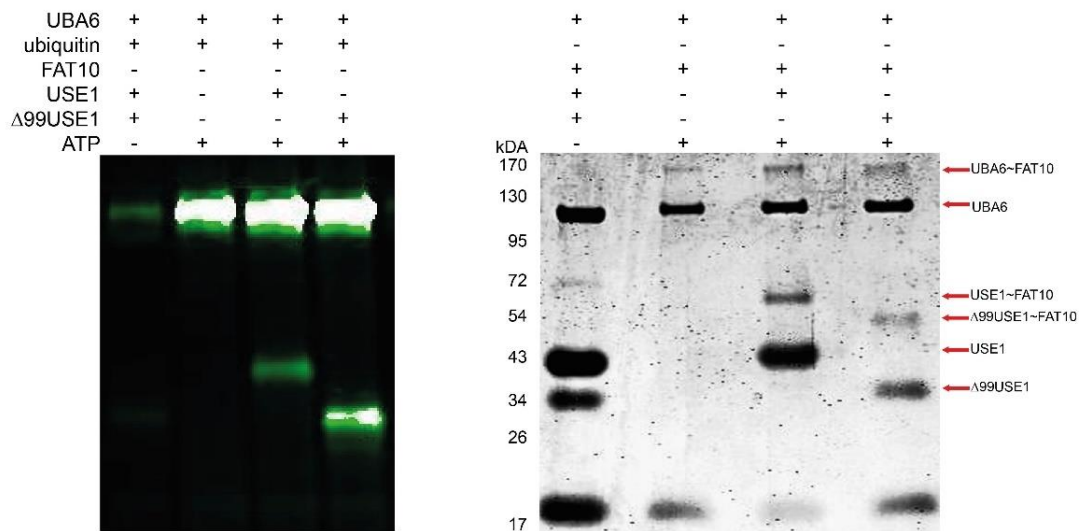


**Figure 17: Purified UBA6 retains catalytic active. UBA6 charging assay with ubiqutin/FAT10 and UBA6-USE1 transthioesterification assays with either full-length USE1 or N-terminally truncated Δ99USE1. (Left panel) Assays with dye-labeled ubiquitin and visualization of ubiquitin-containing bands after illumination with a laser at a wavelength of 800 nm. (Right panel) Assays with FAT10 and visualization of all proteins (E1, E1~FAT10, E2, E2~FAT10 and FAT10) by Coomassie staining.**

## 3.4. Protein crystallization as well as structure determination and refinement

Once the catalytic activity had been confirmed, crystallization setups of UBA6 were performed. The primary objective was to crystallize Uba6 either in complex with ubiquitin or FAT10 and ATP. To prevent the formation of either the UBA6~ubiquitin or UBA6~FAT10 adducts the active site Cys625Ala variant of UBA6 was used in which the reaction will be stopped after the formation of the respective ubiquitin/FAT10 adenylate as the attacking nucleophile for thioester bond formation, the thiol group of Cys625, is missing. The respective non-covalent complexes of UBA6 and the acyladenylated protein modifiers were formed in solution (Figure 18). While there is a significant shift in the elution volume in case of FAT10 in the

presence of ATP (panel C), no significant change in elution volume can be detected with ubiquitin. Nevertheless, SDS PAGE analysis employing silver staining of the elution fractions revealed the presence of ubiquitin in the early elution fractions corresponding to the fractions were UBA6 would elute (Figure 18C). In the case of FAT10 its co-elution withUBA6 can be already be visualized during the seize exclusion chromatography step and can be further confirmed by Coomassie staining.
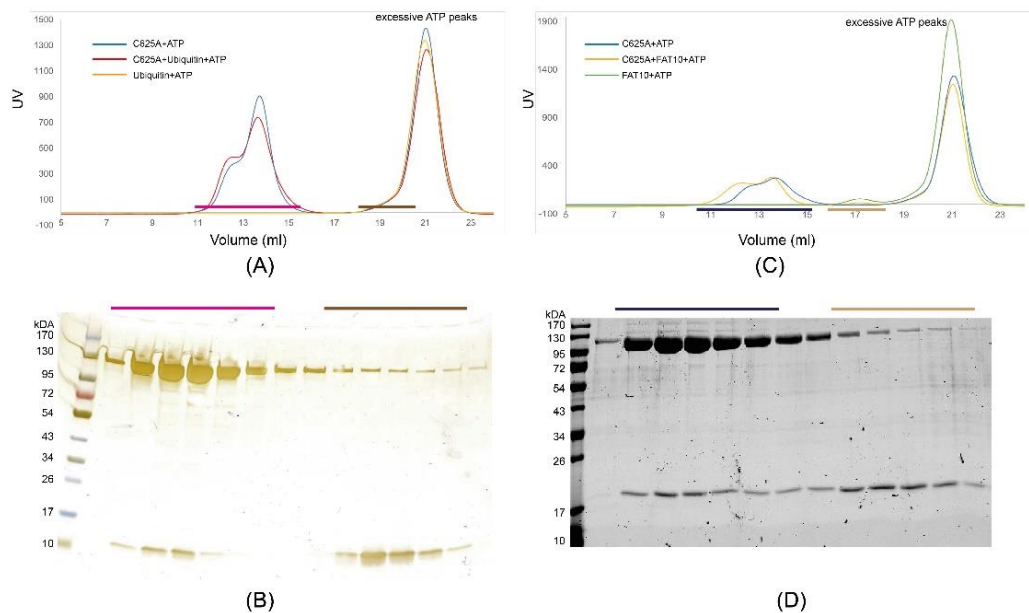


**Figure 18: Formation of UBA6 in complexes with either ubiquitin of FAT10 in the presence of ATP. (A and B) Size exclusion chromatography and SDS-PAGE analysis of the complex between UBA6 and the ubiquitin-acyladenylate. The gel image demonstrates that ubiquitin co-eluted with UBA6. (C and D) The size exclusion chromatography profiles and SDS-PAGE analysis of UBA6 and FAT10 showrf that FAT10 vo-eluted with UBA6.**

Once the formation of the non-covalent UBA6-ubiquitin/FAT10-acyladenylate complexes had been confirmed, crystallization trials in the sitting drop format were set up with a Honeybee crystallization robot (Cartesian) at a concentration of 4 mg/ml with a 2:1 (V:V) ratio of protein solution and mother liquor. Crystals were obtained with either protein modifier but only when additional ATP was added to the protein complex. Crystal growth was rapid with fully grown crystals appearing overnight but only at 4°C. The resulting crystals appeared as either stacked plates for the Uba6-ubiquitin complex or irregularly shaped prisms for the Uba6-FAT10 complex. The former crystals grew with a mother liquor containing 0.16 M calcium

acetate, 0.1 M sodium cacodylate pH 6.5, 14-15 % polyethylene glycol 8000 and 16 % glycerol, while the latter formed in the presence of 0.1 M LiCl, 0.1 M Sodium citrate pH 5.5, 18% polyethylene glycol 4000 and 10% iso-propanol. Initial diffraction experiments with crystals of either morphology revealed poor diffraction with maximal resolutions of 12 Å for the plates and 6 Å for the prism-shaped crystals. The poor diffraction of the crystals was attributed to their rapid formation. In order to slow down crystal growth, the hanging drop method was modified by adding a layer of paraffin oil on top of the reservoir solution to slow down the vapor diffusion process (Figure 19B). As expected, crystal formation was slowed down with crystals being formed only after three to four days. While the diffraction properties of the prism-shaped crystals could not be improved (data not shown), the diffraction properties of the stacked plate-shaped crystals (Figure 19A) of the complex of UBA6 and ubiquitin could be significantly improved (Figure 19C) resulting in diffraction to just beyond 3 Å resolution (Figure 19D). This particular crystal, which was utilized for high resolution refinement, formed in buffer containing 0.16 M calcium acetate, 0.1 M sodium cacodylate pH 6.5, 14.4% PEG 8000 and 16% glycerol. Prior to conducting diffraction experiments the crystal were flash cooled in liquid nitrogen after they had been transferred into artificial mother liquor containing 25% glycerol for both crystal morphologies.
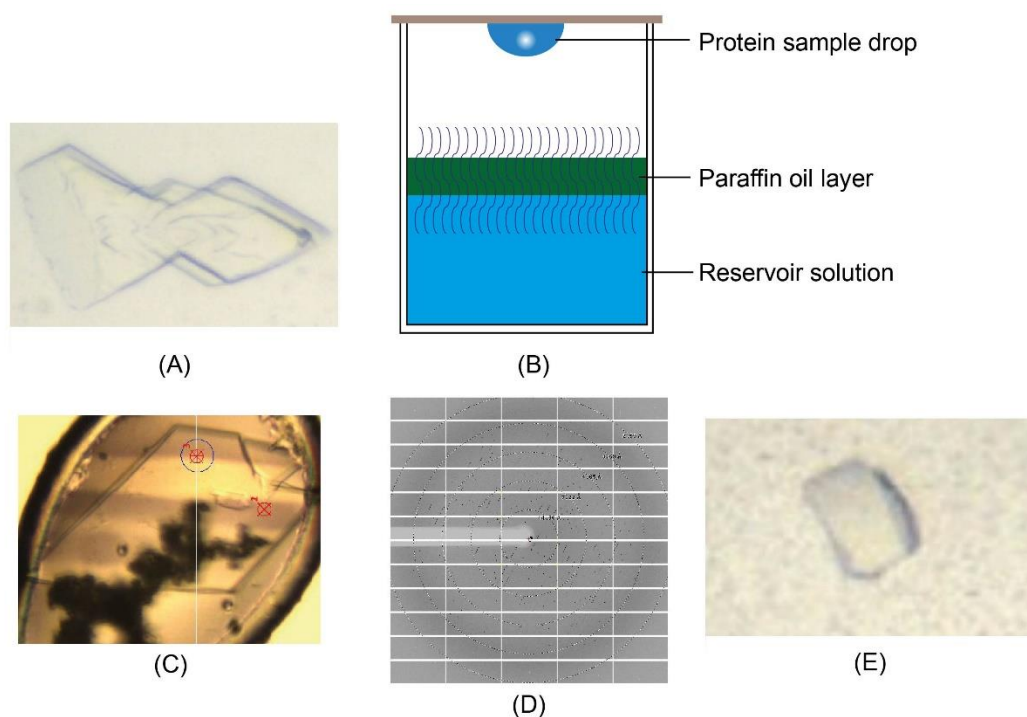


(A)          (B)

(C)          (D)          (E)

**Figure 19: Crystallization of UBA6. (A) Initial crystals of the UBA6-ubiquitin-acyladenylate complex with typical dimensions of 30 x 100 x 2 $\mu$m. (B)**

**Schematic drawing of the optimized hanging drop method. (C) Optimized crystal of UBA6-ubiquitin-acyladenylate complex mounted in a nylon loop. As can be seen, the crystal still appears to consist of at least two stacked plates. The dark shadow represents ice crystals which either formed during the vitrification process or attached onto the crystal during crystal storage or transport. The blue circle indicates the beam size (diameter of 6 μm). (D) X-ray diffraction image of the optimized crystal of UBA6-ubiquitin-acyladenylate complex. (E) Initial crystals of the UBA6-FAT10-acyladenylate complex with typical dimensions of 20 x 50 x 5 μm.**

Data collection at beamline P14 of the EMBL outstation in Hamburg at the Deutsche Elektronensynchrotron (DESY) revealed that crystals of the putative UBA6-ubiquitin-acyladenylate complex crystallized in two space groups, $P2_122_1$ and C2, under identical conditions. The $P2_122_1$ crystals diffracted to a resolution of 3.3 Å (Table 6). Packing calculations assuming a molecular weight of 117 kDa indicated the presence of 2 molecules in the asymmetric unit for either crystal form corresponding to Matthew's coefficients of 2.59 $Å^3$/Da and 2.75 $Å^3$/Da and corresponding solvent contents of 52.5% and 55.3 %.

**Table 6. Data collection statistics and preliminary refinement parameters for space group $P2_122_1$**

| Data collection | |
|---|---|
| Space group | $P2_122_1$ |
| Cell dimensions | |
| $a, b, c$ (Å) | 104.8, 114.3, 206.7 |
| $\alpha, \beta, \gamma$ (°) | 90, 90, 90 |
| Resolution limits (Å) | 50.17 - 3.32 (3.47 - 3.32) |
| $^aR_{merge}$ | 0.376 (4.031) |
| $^bR_{pim}$ | 0.118 (1.360) |
| $CC_{1/2}$ | 0.996 (0.408) |
| $^c<I / \sigma I>$ | 5.2 (0.6) |
| Completeness (%) | 99.3 (94.7) |
| Redundancy | 10.8 (9.3) |
| Wilson B-factor ($Å^2$) | 97.4 |

Crystals belonging to space group $P2_122_1$ were available first and hence structure determination by molecular replacement was carried out with these crystals, followed by preliminary refinement. The molecular replacement approach with the program Phaser involved the sequential placement of the domains of UBA6 by using the corresponding domains of ScUBA1 (PDB entry: 3cmm) as search models. A native Patterson map indicated the presence of a non-crystallographic symmetry translation with coordinates of x=0.5, y=0.5 and z=0.1462 with a height of 47% of the origin peak. Subsequent molecular replacement in Phaser took this translational non-crystallographic symmetry into account, hence only one orientation had to be identified in the rotation function. Initial calculations in which the core of ScUBA1 consisting of its AAD-IAD core was employed revealed one clear solution corresponding to two molecules due to the non-crystallogrphic translational symmetry, which exhibited a translation function (TF) Z-score of 13 and and a log likelihood gain of 302. Initial refinement of this partial molecular replacement solution with Refmac yielded an R-factor of ~40% with an $R_{free}$ of 48%. Assuming that these domains were correctly placed the subsequent search with the SCCH domain of ScUBA1 again produced two clear solution with TF Z-scores of 4.5 as well as an overall LLG (including the AAD-IAD core) of 623. This model was again refined with Refmac resulting in improved R-factors of 37% ($R_{free}$ of 47%). A third round of molecular replacement searching for the UFD again produced two solutions with TF Z-scores of 5.7 and a combined LLG of 378. Preliminary refinement in Refmac reduced the R-factors to 33% and 44% ($R_{free}$). Based on the R-factors and the quality of the electron density maps the molecular replacement solutions were assumed to be correct. Iterative rounds of refinement and model building with Refmac reduced the R-factors to 30.7% and 41.9%. The electron density maps at this point indicated where the FCCH domain should be located and this domain was positioned accordingly in Coot in both UBA6 molecules by hand. At this stage refinement was continued with Phenix and after multiple rounds of rebuilding and refinement a preliminary model with R-factors of 23.8% and 28.5% ($R_{free}$) was obtained. Throughout the entire structure solution process no density for ubiquitin could be observed, hence it was concluded that the complex must have disintegrated during crystallization. At the same time clear density of Mg-ATP was already visible in the early stages and hence Mg-ATP was included in the model. At this stage the higher resolution dataset derived from crystals belonging to the monoclinic space group C2 was obtained.

**Table 7. R-factors at different stages of molecular replacement and refinement against the P2$_1$22$_1$ dataset**

| Stage | R/R(free) |
|---|---|
| MR AAD/IAD/FCCH (2 copies) | 0.5012/0.5185 |
| After one round of refinement in Refmac | 0.3984/0.4795 |
| Known AAD/IAD/FCCH and MR of SCCH (2 copies) | 0.5068/0.5213 |
| After one round of refinement in Refmac | 0.3709/0.4734 |
| Known AAD/IAD/FCCH/SCCH and MR of UFD (2 copies) | 0.5130/0.5153 |
| After one round of refinement in Refmac | 0.3309/0.4416 |
| After model building and extensive refinement in Refmac | 0.3073/0.4187 |
| After model building and extensive refinement in Phenix | 0.2382/0.2852 |

Refinement in space group C2 was initiated by molecular replacement using a single UBA6 subunit from the preliminarily refined P2$_1$22$_1$ structure. Two solutions with TF Z-scores of 9.0 and 10.6 and an overall LLG of 302 were readily identified. In this space group the two molecules are related by a non-crystallographic rotation and hence two independent orientations had to be determined. As already a visual inspection of the diffraction images revealed significant anisotropy in the diffraction limits of the C2 crystals, which was confirmed during data processing of these crystals, the data were anisotropically truncated with the Staraniso server from Global Phasing (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi) revealing significant anisotropy with maximal resolution limits of 2.7/3.8 Å in the best/worst direction. Refinement was initiated against anisotropically corrected data in Phenix and was completed with Buster. The final model exhibited R-factors of 21.2% and 23.5% (R$_{free}$) with satisfactory model quality as documented by low deviations from ideality for the stereochemical restraints and a Ramachandran diagram with approximately 95% of the residues in the most favored regions and 0.5% outliers (Figure 20G).

**Table 8. Data collection statistics and refinement parameters for the UBA6 structure in space group C2.**

| Data collection | |
|---|---|
| Space group | C2 |
| Cell dimensions | |
| $a$, $b$, $c$ (Å) | 123.9, 113.7, 183.5 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 96.5, 90 |
| Resolution limits (Å, number for high resolution limit in the best/worst diffraction) | 48.25 – 2.71 / 3.83 |
| [a]$R_{merge}$ | 0.294 (2.570) |
| [b]$R_{pim}$ | 0.123 (1.158) |
| $CC_{1/2}$ | 0.9847 (0.1695) |
| [c]$<I / \sigma I>$ | 6.0 (0.92) |
| Completeness (%, spherical/elliptical) | 0.725 / 0.948 (0.218 / 0.774) |
| Redundancy | 6.66 (5.56) |
| Wilson B-factor (Å$^2$) | 70.3 |
| Refinement | |
| No. reflections | 49,253 (986) |
| [d]$R_{work}$ / [e]$R_{free}$ | 0.212 / 0.235 (0.233 / 0.274) |
| No. of atoms | |
| Protein | 15,654 |
| Water | 107 |
| ATP | 299 |
| $B$-factors (Å$^2$) | |
| Protein | |
| Water | |
| ATP | |
| [f]Ramachandran statistics (%) | 95.0 / 4.5 / 0.5 |
| Overall B-factor (Å$^2$) | 74.7 |
| Data precision index (Å) | 0.39 |
| RMS deviations in | |

| Bond lengths (Å) | 0.01 |
|---|---|
| Bond angles (°) | 1.14 |
| Torsion angles (°) | 3.39 |
| Planar groups (Å) | 0.012 |

[a]$R_{sym} = \Sigma_{hkl}\Sigma_i \mid I_i - <I> \mid / \Sigma_{hkl}\Sigma_i I_i$ where $I_i$ is the $i^{th}$ measurement and $<I>$ is the weighted mean of all measurements of *I*.

[b]$R_{pim} = \Sigma_{hkl}\overline{1/ (N\text{-}1)} \; ^{1/2} \Sigma_i \mid I_i(hkl) - I (hkl) \mid / \Sigma_{hkl}\Sigma_i I(hkl)$, where N is the redundancy of the data and $I (hkl)$ the average intensity.

[c]$<I / \sigma I>$ indicates the average of the intensity divided by its standard deviation.

[d]$R_{work} = \Sigma_{hkl} \mid\mid F_o\mid - \mid F_c\mid\mid / \Sigma_{hkl}\mid F_o\mid$ where $F_o$ and $F_c$ are the observed and calculated structure factor amplitudes.

[e]$R_{free}$ same as R for 5% of the data randomly omitted from the refinement. The number of reflections includes the $R_{free}$ subset.

[f]Ramachandran statistics were calculated with MolProbity in PHENIX.

Numbers in parentheses refer to the respective highest resolution data shell in the dataset.

## 3.5. Analysis of the UBA6 structure

### 3.5.1. Overall structure of Uba6

As mentioned in the introduction, UBA6 belongs to the canonical class of E1 enzymes and hence shares their conserved multidomain architecture consisting of a core composed of the pseudo-symmetrically arranged AAD and IAD domains together with the 4-helix-bundle, which is decorated with the FCCH and SCCH domains on one side and the UFD on the other (Figure 20). The conserved domain architecture was, of course, also the reason why the structure could be solved by molecular replacement. As can be seen in Figure 20E, the overall level of sequence identity between UBA6 and ScUBA1 is slightly above 40%, which is well within the range of successful molecular replacement structure solutions. Nevertheless, in such a multi-domain architecture the movements of individual domains may complicate structure determination by molecular replacement with the full-length template as search model. Such a situation also occurred during the UBA6 molecular replacement approach and hence a divide and conquer approach was utilized starting with the AAD-IAD-4HB core and the subsequent addition of the peripheral domains.

Before the structure of UBA6 will be described in more detail, the structural variability of UBA6 displayed by the two molecules present in the asymmetric unit of the C2 crystals will be analyzed. The IAD, AAD and 4HB domains are packed

together to form a rigid body of the UBA6 structure. This body in the two UBA6 molecules in the asymmetric unit exhibited almost identical conformations with a root-mean-square deviation (rmsd) of 0.44 Å over 481 aligned Cα atoms (Figure 20F). Taking into account the average coordinate error as reflected in the data precision index with 0.39 Å (Table 3), the structure of this core region is virtually invariant. Around this conserved core the accessory components, the FCCH and SCCH domains as well as the UFD display enhanced variability. In particular, the UFD adopts slightly different orientations as already observed and described in great detail for Uba1 (Lee and Schindelin 2008, Lv, Williams et al. 2018). Individually, these domains can be superimposed with comparable rmsd values as the core, however, these domains undergo rigid body motions. Consequently, a superimposition of the two full-length UBA6 molecules results in a rmsd of 2.13 Å for 986 (out of a total of 1017) aligned Cα atoms. This larger rmsd is due to the aforementioned motions of the FCCH, SCCH and UFD domains, which are linked to the rigid body with the flexible loops (Figure 20A-C).

The FCCH domain is tethered to the IAD and 4HB domains by two loops called the β7 and β15 loops, respectively, (Figure 20A and Figure 21A). Likewise the SCCH domain is linked to the AAD by two loops which are generally referred to as crossover loop and reentry loop (Figure 20A-C) with the first leading into the SCCH domain and the latter connecting this domain back to the core. Finally, the UFD domain is tethered to the AAD by a single loop (Figure 20A-C). The crossover loop traverses from one side of the molecule to the other and was reported to play an important role in directing the ubiquitin/UBL C-terminal tail to the adenylation site in the respective E1 (Lee and Schindelin 2008). Overall, human UBA6 adopts a 'Y' shape with the IAD, AAD and 4HB domains forming the base of the letter 'Y' and the UFD and SCCH domain as the two arms. Consequently, these two domains are situated across from each other forming a large canyon in between that serves to accommodate the E2 enzymes during the transthioesterification reaction that transfers ubiquitin/UBL from the active site cysteine residue in the E1 to its counterpart in the E2. The multidomain structure and plasticity of the canonical E1s were reported to (1) differentiate the UBLs for activation which involves many specific interactions and requires large conformational changes (Olsen, Capili et al. 2010, Olsen and Lima 2013) and (2) selectively recruit the E2s and subsequently catalyze transfer of the UBLs from E1s to E2s (Olsen and Lima 2013, Lv, Rickman et al. 2017).
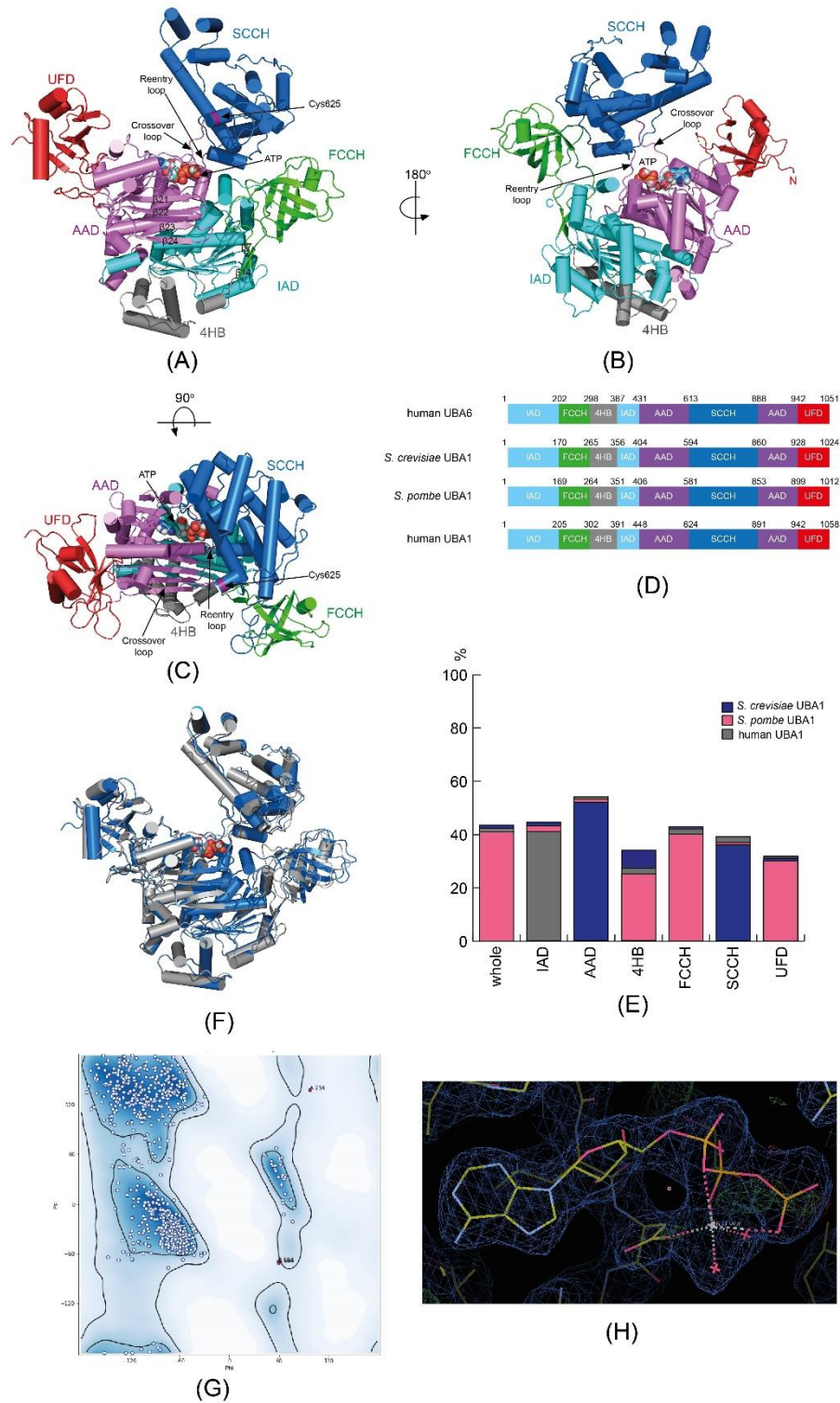
**Figure 20: Overall structure of human UBA6 in front (A), top (C) and back (B) views with the inactive adenylation domain (IAD) in cyan, the active adenylation domain (AAD) in purple, the first catalytic cysteine (FCCH) domain in green, the second catalytic cysteine (SCCH) domain in blue, the 4-**

helix-bundle (4HB) in grey and the ubiquitin fold domain (UFD) in red. (D) Domain architectures of UBA6 and the three, structurally characterized UBA1 orthologs. (E) Sequence identities (overall and for the isolated domains) between UBA6 and these UBA1 proteins. (F) Superimposition of the A-chain (blue) and B-chain (grey) corresponding to the two UBA6 monomers present in the asymetric unit. (G) Ramachandran diagram of the refined UBA6 structure with outliers 664 in both chains and 734 in the A-chain. Preferred regions of the Ramachandran diagram are indicated by intensifying saturation of the blue color. (H) Electron density map of the ATP molecule in the A-chain with the coordinated $Mg^{2+}$-ion in the UBA6 adenylation site.



(A)
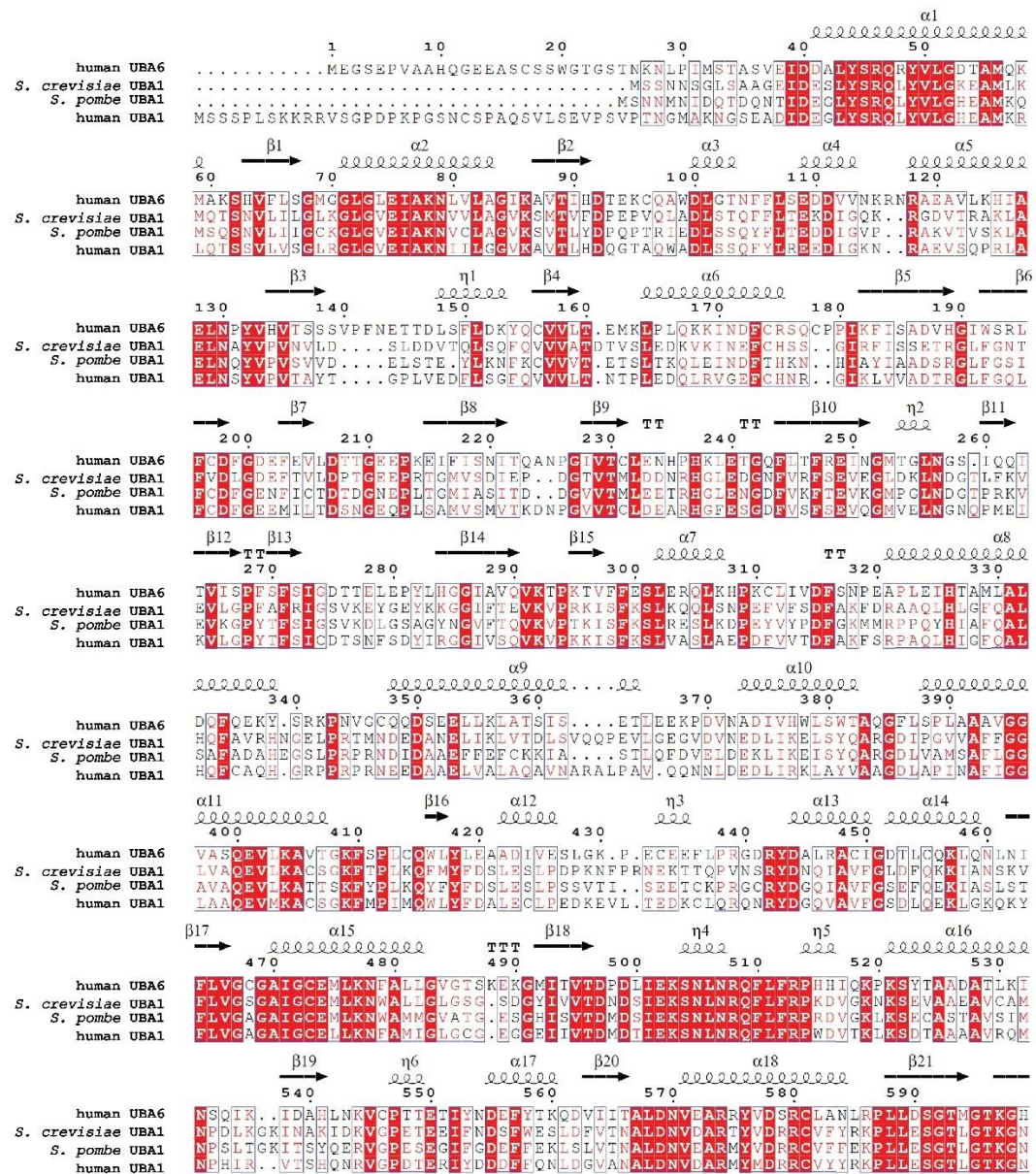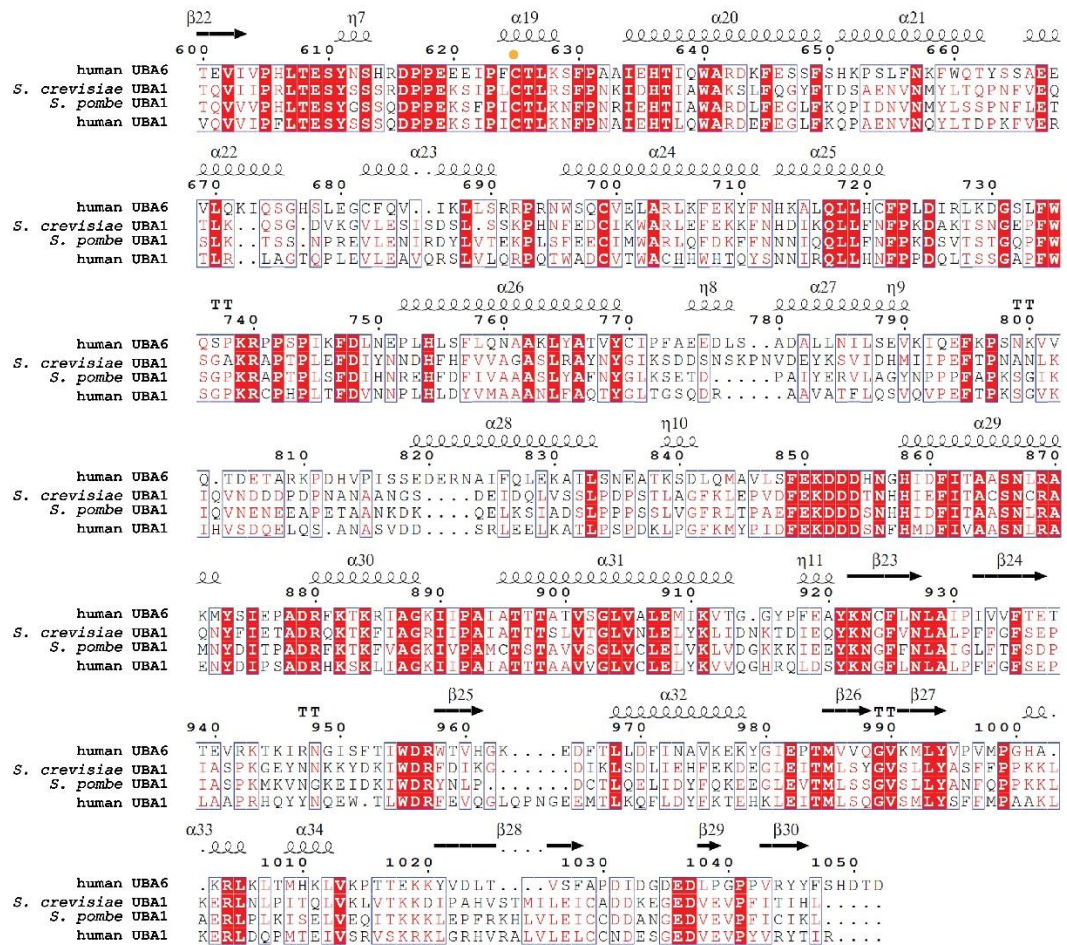
**Figure 21: (A) and (B) Multiple sequence alignment of human UBA6 and the three structurally characterized UBA1 proteins. Above the alignment, the secondary structure elelments present in UBA6 are depicted with helices as spirals (designated with α for α-helices and η for 3₁₀-helices) and β-strands of arrows. Sequences were aligned using Clustal Omega and the figure was generated with ESPript. Red shades indicate conserved residues, while red letters in white boxes indicate similar residues. The catalytically active cysteine (Cys625 of UBA6) is indicated by an orange dot.**

As outlined above the sample set up for crystallization actually contained UBA6 and ubiquitin, the latter presumably in its acyladenylated state plus a 166-fold excess of Mg-ATP. One possible explanation for the absence of the acyladenylated ubiquitin in the structure and the presence instead of Mg-ATP could be an exchange reaction in which the nucleotide displaced the AMP-modified ubiquitin. While the equilibrium would presumably reside on the side of the UBA6-ubiquitin-acyladenylate complex, crystals packing forces and the high stoichiometry excess of Mg-ATP might have

shifted the equilibrium towards a UBA6-ATP complex. In fact, at the regions where ubiquitin would be expected to bind crystallographic neighbors can be found. Consequently, the binding of Mg-ATP could be visualized in this structure. Mg-ATP is well defined in the electron density (Figure 20H) and resides in its expected position in the adenylation active site of the AAD domain (Figure 20). While it is difficult to model the $Mg^{2+}$-cation satisfactorily at this resolution, it appears as if the metal is coordinated by 3 phosphate oxygens, 2 $H_2O$ molecules and and one oxygen from the side chain of Asp569 with roughly octahedral geometry. The residues known to be involved in the adenylation reaction including Arg46 and Arg508 coordinating the trisphosphate moiety and Asp497 (Lake, Wuebbens et al. 2001) are conserved in UBA6 and interact with the nucleotide in an analogous fashion.

### 3.5.2. Structural comparisons of UBA6 with UBA1

As stated above the overall architecture and domain organization is conserved between UBA6 and UBA1. Superimposition of UBA6 and ScUBA1 in complex with ATP (Misra et al., unpublished data) revealed a rmsd of 2.2 Å in the positions of the respective C$\alpha$-atoms for 835 aligned residues (out of a total of 998 residues). It should be stated that although HsUBA1 was cocrystallized with Mg-ATP only the triphosphate moiety could be visualized in the map. Figure 22F displays the results of the corresponding superimposition, which reveals a maximal agreement for the core region composed of the AAD-IAD-4HB domains and larger deviations in the peripheral domains, in particular the FCCH domain, as described in more detail in the next section. For the SCCH domain and the UFD domain movements of up to 20.5 Å and rotations of up to 21° were detected. The structure of HsUBA1 (Lv, Williams et al. 2018)presumably also represents a complex with Mg-ATP, however, only the triphosphate moiety of Mg-ATP could be visualized.

### 3.5.3. The FCCH domain

Superimposition of UBA6 and HsUBA1 (PDB entry: 6dc6), yeast UBA1 (PDB entry: 3cmm) and *S. pombe* UBA1 (PDB entry: 4ii2) revealed a significant shift of the FCCH domain in UBA6 (Figure 22B). Compared to HsUBA1, the FCCH domain in UBA6 exhibits a movement of 7.5 Å away from the AAD-IAD core and hence away from where the ubiquitin molecule binds in the AAD domain as well as a 29° rotation (Figure 22A). This change consequently introduces a broader cleft between the FCCH domain and the core of UBA6. One attractive hypothesis derived from this

structural change would predict that the extra space is required to accommodate the second ubiquitin-like domain of FAT10.

Upon closer examination of the resulting cleft, a conserved cluster of hydrophobic residues including Phe203, Val205, Val297 in the β7 and β14 loops and Phe199, Phe299 and Pro412 in UBA6 tightly connects the FCCH domain to the rigid body at the bottom of the cleft (Figure 22C, lower panel). Further up in the cleft are two ionic interaction networks which further connect the FCCH to the UBA6 core (Figure 22C upper panel). The first network involves Lys381 of the IAD and Asp207 and Lys292 of the FCCH. The second ionic network is formed between Lys597 of the AAD and Glu211 of the FCCH. Although the latter two charged residues are conserved in the sequence of UBA6 (Figure 22E), the distances between them (Lys597 and Glu211) are larger as a consequence of the outward movement of its FCCH domain. The interface between the FCCH and the SCCH in HsUBA1 (Figure 22D, grey residues) includes a hydrophobic cluster (referred as FCCH-SCCH hydrophobic cluster) which involves Met220, Leu235 and Pro745 and the ionic interaction featuring Lys747 and Glu252 (Figure 22D, grey residues). These interactions are conserved among the UBA1 orthologs structures and sequences (Figure 22D and Figure 22E). Although the residues which are involved in the FCCH-SCCH hydrophobic cluster of UBA1 are type-conserved in UBA6 based on the sequence alignment (Figure 22E), the interactions are unlikely to form due to the significantly increased distance (Figure 22D), which is a consequence of the FCCH domain shift. Instead, Glu249 in the FCCH of UBA6 is likely involved in an ionic interaction network with Lys248 and Glu876 in the SCCH (Figure 22D). This shift in the ionic interactions could at least partially stabilize the FCCH domain of UBA6 in the distal position observed in the UBA6 structure.
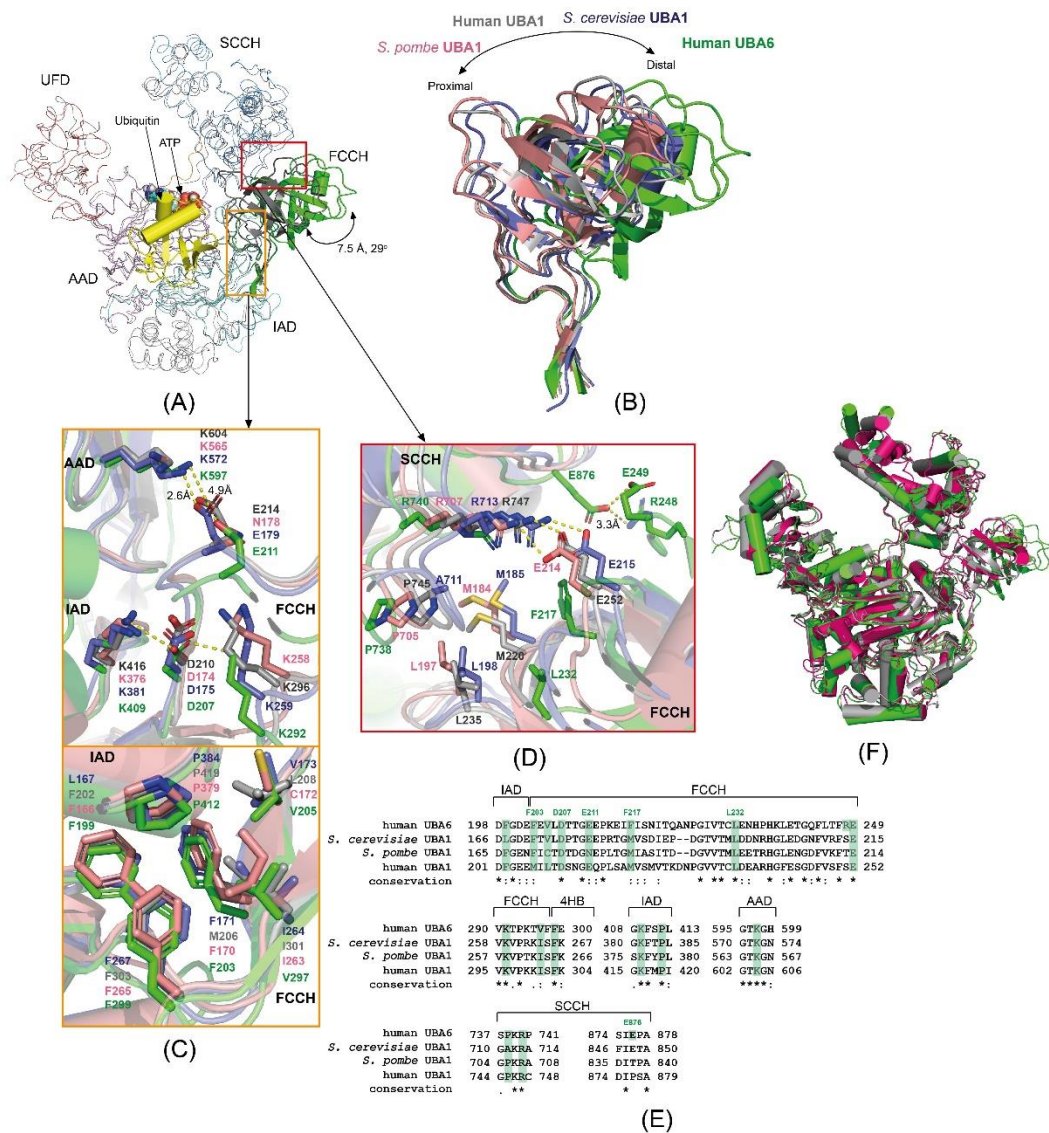
## Figure 22

(A) UFD, SCCH, Ubiquitin, ATP, AAD, IAD, FCCH, 7.5 Å, 29°

(B) Human UBA1, *S. cerevisiae* UBA1, *S. pombe* UBA1, Human UBA6, Proximal, Distal

(C)
AAD — K604, K565, K572, K597, 2.6Å, 4.9Å, E214, N178, E179, E211
IAD — K416, K376, K381, K409, D210, D174, D175, D207
FCCH — F258, K296, K259, K292
IAD — L167, F202, P384, P419, P379, P412, V173, L208, C172, V205, F199
F171, M206, F170, F203, F267, F303, F265, F299, I264, I301, I263, V297, FCCH

(D) SCCH, E876, E249, R740, R707, R713, R747, R248, 3.3Å, P745, A711, M185, E214, E215, M184, E252, P705, P738, F217, L197, L198, M220, L232, L235, FCCH

(E) Sequence alignment

```
                IAD                      FCCH
          F203 D207 E211   F217                          L232
human UBA6      198 DFGDEFEVLDTTGEEPKEIFISNITQANPGIVTCLENHPHKLETGQFLTFRE 249
S. cerevisiae UBA1 166 DLGDEFTVLDPTGEEPRTGMVSDIEP--DGTVTMLDDNRHGLEDGNFVRFSE 215
S. pombe UBA1     165 DFGENFICTDTDGNEPLTGMIASITD--DGVVTMLEETRHGLENGDFVKFTE 214
human UBA1        201 DFGEEMILTDSNGEQPLSAMVSMVTKDNPGVVTCLDEARHGFESGDFVSFSE 252
conservation          *:*::::   * *::*     :::  :   * ** *:: *  :* *;*: * *

                FCCH  4HB          IAD            AAD
human UBA6      290 VKTPKTVFFE 300  408 GKFSPL 413  595 GTKGH 599
S. cerevisiae UBA1 258 VKVPRKISFK 267  380 GKFTPL 385  570 GTKGN 574
S. pombe UBA1     257 VKVPTKISFK 266  375 SKFYPL 380  563 GTKGN 567
human UBA1        295 VKVPKKISFK 304  415 GKFMPI 420  602 GTKGN 606
conservation         **.* .: *:       .** *:        ****:

                SCCH
                                  E876
human UBA6      737 SPKRP 741   874 SIEPA 878
S. cerevisiae UBA1 710 GAKRA 714   846 FIETA 850
S. pombe UBA1     704 GPKRA 708   835 DITPA 840
human UBA1        744 GPKRC 748   874 DIPSA 879
conservation         .  **         *  *
```

**Figure 22: Analysis of the FCCH domain movement. (A)** Ribbon diagram of a superimposition of UBA6 (mostly in ribbon representation with the domains colored as in Figure 20) and the complex of HsUBA1 and ubiquitin (PDB: 6dc6, mostly in grey ribbon representation). The FCCH domains (cartoon representation) of UBA6 is shown in green, that of UBA1 in grey and ubiquitin in yellow. **(B)** Relative positions of the FCCH domains of UBA6 (green), ScUBA1 (blue), HsUBA1 (grey) and *S. pombe* UBA1 (salmon) after superimposition of the respective core domains. **(C and D)** Graphical representation of the interactions between the FCCH domain and their neighboring domains in the aforementioned UBA1 enzymes. **(E)** Sequence alignment showing residues of these ubiquitin activating enzymes and UBA6, which are involved in interactions with the respective FCCH domains. **(F)**

**Superposition of UBA6 (green), HsUBA1 (grey) and ScUBA1 (hot pink) in complex with Mg-ATP (unpublished data).**

### 3.5.4. Modeling the UBA6-ubiquitin interactions and probing the predicted complex

By superimposing UBA6 and HsUBA1 in complex with ubiquitin, a model of the complex between UBA6 and ubiquitin was generated (Figure 23A). According to this model and in analogy with the first Uba1-ubiquitin structure (Lee and Schindelin 2008), the interactions of ubiquitin and UBA6 can be divided into three major interfaces (Figure 23A). Interface I would include contacts of the FCCH domain and one side of the globular β-grasp domain of ubiquitin (Figure 23B). The closest contact in this interface involves the side chain of Glu211 in the FCCH and the backbone amine of Thr9 of ubiquitin, however, with a distance of 6.7 Å it is beyond the limit of what would be considered a meaningful interaction (Figure 23B). Although ubiquitin has contacts with the FCCH domains of HsUBA1 and ScUBA1 (Lv, Williams et al. 2018), the significant shift of the FCCH of UBA6 likely abolished these interactions. Interface II includes the interactions of the residues from the 'bottom' of the globular β-grasp domain of ubiquitin as viewed in Figure 23A. Included in this interface is the 'Ile44 hydrophobic patch' of ubiquitin, which, besides Ile44, features Leu8, His68, and Val70. All four residues engage in contacts with a hydrophobic platform of HsUBA1 (Phe926, Phe933, Ser937 and Phe320) (Lv, Williams et al. 2018) (Figure 23C). This conserved hydrophobic platform is located on the β23 and β24 strands of UBA6 and includes Phe316, Phe925, Ile932 and Thr939 (Figure 23C). In addition, Val314 and Val934, which are unique to UBA6, are predicted to further strengthen the hydrophobic interactions between ubiquitin and UBA6, specifically around His68 of ubiquitin (Figure 23C). Ile44 of ubiquitin is surrounded by three phenylalanine residues in HsUBA1 whereas a 180° rotation of the Phe925 side chain compared to Phe926 of UBA1 and a shift by 4.2 Å of Phe316 of UBA6 compared to Phe320 of UBA1 are observed (Figure 23C). All of the changes in UBA6 hydrophobic platform likely enhance the hydrophobic interaction strength in the vicinity of His68 of ubiquitin, while the hydrophobic platform of HsUBA1 surrounding Ile44 of ubiquitin seems not as extensive. Interface III involves the interactions between the ubiquitin C-terminal tail and the adenylation site of UBA1/UBA6. The crossover loop which traverses the ubiquitin tail, directs it towards the adenylation site and positions the C-terminal Gly76 of ubiquitin for catalysis (Figure 23D). The presence of His599 and His614 instead of N606 and S621 in UBA1 renders the adenylation site of UBA6 more positively charged. This could

potentially create a repulsive force for Arg72 of ubiquitin (Figure 23D). The negatively charged residue Glu601 may balance the overall positive charge in this site by forming an ionic interaction with Arg72 of ubiquitin (Figure 23D), thus stabilizing the binding of the ubiquitin tail. Arg74 of ubiquitin not only forms an ionic interaction with Glu626 of HsUBA1 but also 1 or 2 hydrogen bonds. The flexibility of the crossover loop could potentially enable ionic interactions of Arg74 with Glu619 in the crossover loop of UBA6 (Figure 23D). Leu73 of ubiquitin is predicted to engage in hydrophobic interaction with Met594 of UBA6 and the side chain of Arg574 of UBA6 is predicted to form a hydrogen bond with the backbone of Gly75 of ubiquitin. These interactions would further stabilize the ubiquitin Gly-Gly tail in close proximity of the terminal phosphate of the ATP molecule (Figure 23D). This would be in complete analogy to ubiquitin activating enzymes where these interactions are completely conserved (Figure 23D and Figure 23E). In summary, although UBA6 and UBA1 are predicted to share common interactions in ubiquitin binding, UBA6 possesses unique features, which may also allow the recognition of FAT10.

The proposed UBA6-ubiquitin interface was probed by site directed mutagenesis. The UBA6 and HsUBA1 variants in which residues in their adenylation sites were exchanged for each other were generated and the corresponding proteins were purified as described in Section 3.3.4. The proteins behaved similar to the wild-type resulting in comparable yields, hence arguing against major structural impairments in their fold. The substitution of Glu601 of UBA6 to the corresponding residue Gln608 of HsUBA1 was predicted to ruin the charge balance and ionic interaction with Arg72 of ubiquitin. Indeed, this substitution completely eliminated the ability of UBA6 to activate ubiquitin (Figure 23F). While the UBA1 variant in which Gln608 was substituted to Glu did not exhibit any deleterious effect probably due to the hydrogen bonds with Ser621 or Asn606 and the ionic interaction of Glu626 and Arg74 of ubiquitin (Figure 23D) and the ionic interaction with Arg42 of ubiquitin (Figure 23G). In contrast, neither the individual substitutions of His599 to asparagine or His614 to serine nor the corresponding double mutation blocked the activation of ubiquitin by UBA6 (Figure 23F). The introduction of 2 His residues into the adenylation site of HsUBA1 neither blocked the activation to ubiquitin.
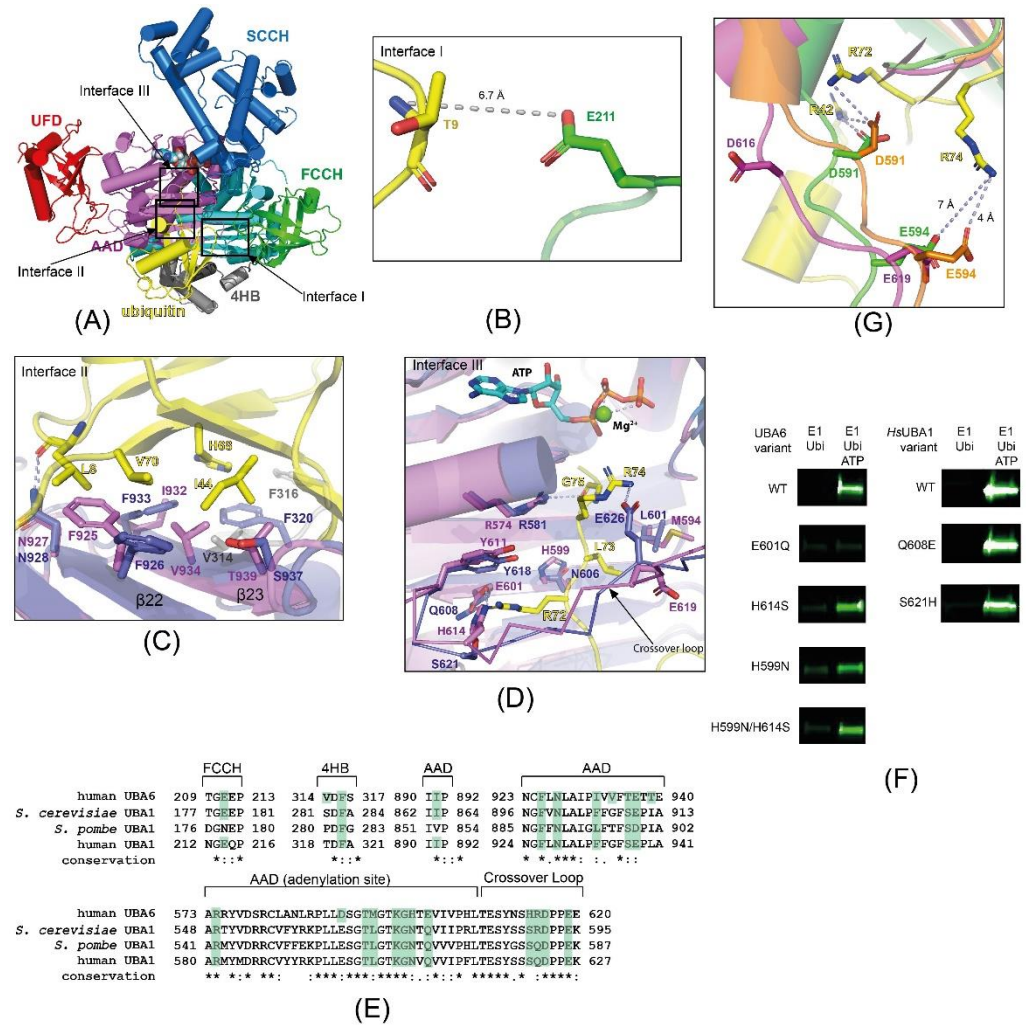
**Figure 23: Structural elements of the predicted UBA6-ubiquitin complex. (A)** Overall representation of the complex with UBA6 colored according to its domain architecture and ubiquitin in yellow. **(B)** Lack of interactions involving ubiquitin and the FCCH domain of UBA6. **(C and D)** Superimposition of HsUBA1 (blue) and UBA6 (magenta) highlighting the interactions with the hydrophobic patch of ubiquitin (C) and its C-terminal tail (D). **(E)** Sequence alignment of UBA6 and UBA1 orthologs showing the level of conservation for the residues interacting with ubiquitin with conserved residues being shaded in green. **(F)** Enzymatc assay investigating ubiquitin activation by UBA6 variants (left panel) and HsUBA1 variants (right panel). **(G)** Superimposition of UBA6 (purple), ScUBA1 in complex with Mg-ATP (green) or ubiquitin (orange). The binding of ubiquitin triggers the formation of ionic interactions between Glu594 of ScUBA1 and Arg74 of ubiquitin, since the distance between the two residues decreased. An additional ionic interaction is present between Asp591 of ScUBA1 and Arg42 of ubiquitin while Asp616 of

**UBA6 is oriented outward in the apo-form. The two ionic networks are predicted to be conserved in the complex of HsUBA1 and ubiquitin.**

### 3.1.1. UBA6 and FAT10 interactions

As described in section 1.1.6 FAT10 is composed of two ubiquitin-like domains and structural studies of the individual domains clearly confirmed this prediction (Aichem, Anders et al. 2018). The individual domains can be superimposed with ubiquitin resulting in rmsd values of 1 Å and 0.8 Å for the superimposition with the N-terminal and C-terminal domain, respectively. The superimposition suggested that the two domains have unique E1 binding surface properties (Aichem, Anders et al. 2018). In analogy to ubiquitin, the C-terminal di-Gly motif of FAT10 will have to reach the ATP binding site of UBA6 for the ensuing activation following ATP-hydrolysis. Therefore, the C-terminal ubiquitin-like domain of FAT10 (C-FAT10) is supposedly bound to UBA6 in the same position and orientation as to where ubiquitin was proposed to bind as described in the previous section. Consequently, the C-terminal domain of FAT10 (C-FAT10) was superimposed with ubiquitin in the model of the UBA6-ubiquitin complex (see Section 3.5.4) to generate a model of how the C-terminal domain of FAT10 (C-FAT10) might interact with UBA6 (Figure 24A). Since the C-FAT10 is an NMR structure featuring an ensemble of 10 models, the model was selected in which the C-terminal tail superimposed best with the corresponding residues in ubiquitin.
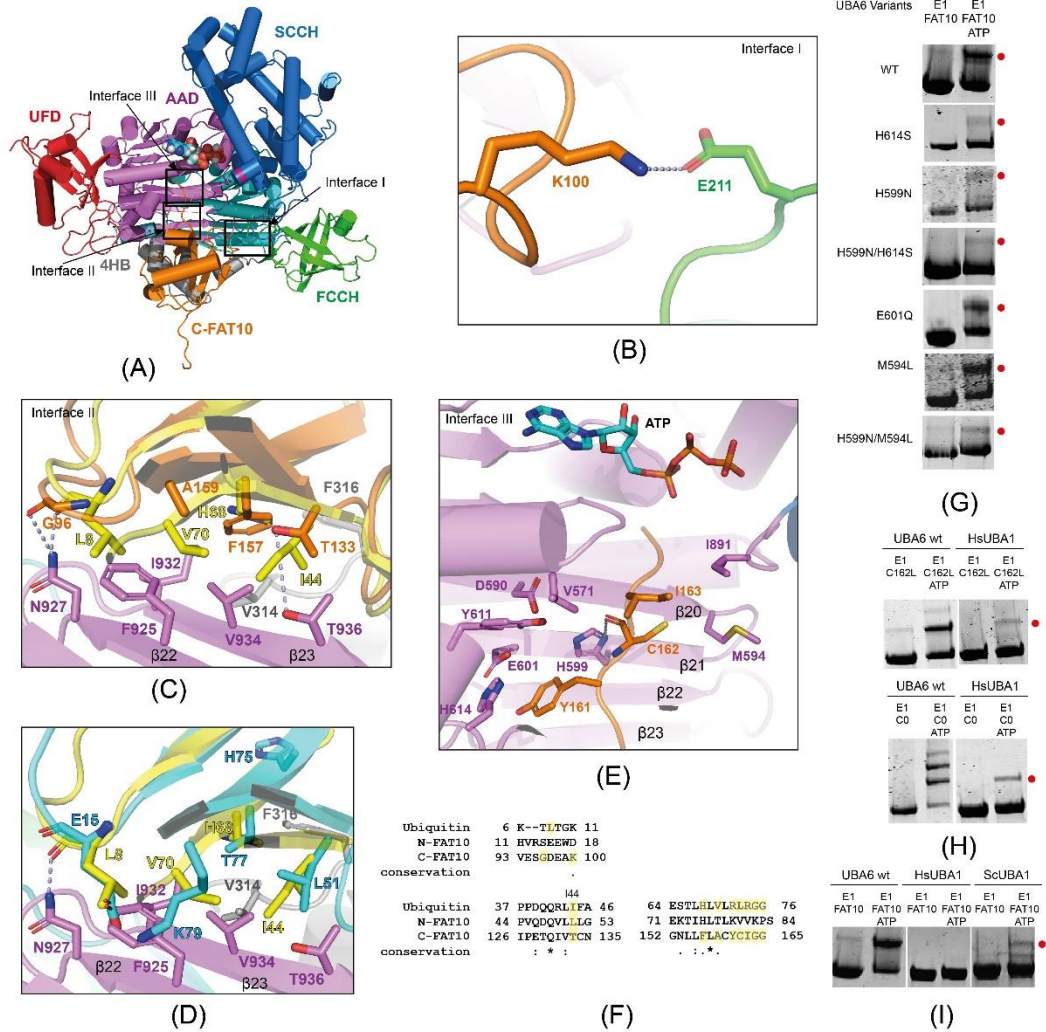
**Figure 24: Modeling the interactions between UBA6 and the C-terminal domain of FAT10 (C-FAT10). (A)** Ribbon diagram of UBA6 (colored according to its domain architecture as described in Figure 20) in complex with C-FAT10 (orange). **(B)** Predicted interations in interface I involving the FCCH domain. **(C)** Comparison of the predicted interactions involving interface II for C-FAT10 (orange) and ubiquitin (yellow). **(D)** Alternative model for interface II in which the N-terminal domain of FAT10 (N-FAT10) (cyan) is superimposed with ubiquitin (yellow). **(E)** Predicted interactions between UBA6 and the C-terminal tail of FAT10 (orange). **(F)** Sequence alignment of ubiquitin and the two domains of FAT10 with residues, which are predicted to be involved in interactions with UBA6, shaded in yellow. **(G)** Biochemical data showing the activation of FAT10 by the wild type and variants forms of UBA6. **(H)** Activity assays with UBA6 and HsUBA1 analyzing cysteine variants of FAT10, C162L and cysteine free FAT10 (C0). **(I)** FAT10 activation catalyzed by UBA6, HsUBA1 (HsUBA1) and *ScUBA1*.

Similar to the UBA6-ubiquitin complex, the interactions between C-FAT10 and UBA6 can be grouped into three interfaces (Figure 24A). Interface I shows an ionic interaction involving Lys100 of C-FAT10 and Glu211 of the FCCH domain in UBA6. The lysine residue is conserved in ubiquitin (Figure 24F), however, it is involved in an intramolecular ionic interaction with residue Glu34 of ubiquitin as seen in the crystal structure. Interface II again features the contacts at the 'bottom' of C-FAT10 and the hydrophobic platform of UBA6 (Figure 24C). A unique 'Phe157 hydrophobic patch' where Phe157 of C-FAT10 replaced His68 of ubiquitin is situated on the top of the UBA6 hydrophobic platform, which includes Val314, Phe316, Ile932, Val934 and Phe925 (Figure 24C). The central position and stronger hydrophobicity of Phe157 compared to His68 could potentially compensate for the weaker hydrophobicity of Gly96, Thr133 and Ala159 in FAT10, which are predicted to be present at the positions of Leu8, Ile44 and Val70 of ubiquitin, respectively (Figure 24C). The backbone of Gly96 of C-FAT10 is predicted to engage in a conserved hydrogen bond with Asn927 of UBA6 (Figure 24C). Interface III contains the predicted binding interactions involving the C-terminal tail of C-FAT10 and the adenylation site of UBA6 (Figure 24E). The presence of the residues His599 and His614 and Tyr611 of UBA6 may favor binding of the more hydrophobic tail of C-FAT10 which includes Tyr161, Cys162 and Ile163 (Figure 24E). Tyr161 is potentially sandwiched between His614 and Tyr611 of UBA6. While Tyr611 of UBA6 is a conserved residue in ubiquitin E1 enzymes, His614 is replaced with serine (Figure 24E). Replacements of the non-conserved residues His614 and His599 of UBA6 with the corresponding residues serine and asparagine of HsUBA1 were generated. Somewhat surprisingly, neither single substitution nor the double mutant completely inactivated UBA6 towards FAT10 (Figure 24G), however, a significant reduction was observed. A hydrogen bond is predicted to form between Tyr161 of C-FAT10 and Glu601 of UBA6 (Figure 24E), however, replacing Glu601 with Gln did not impair FAT10 activation by UBA6 (Figure 24G). Based on the model, Cys162 of C-FAT10 may form a hydrogen bond with His599 and engage in hydrophobic interactions with Met594 of UBA6 (Figure 24E). While the M594L variant revealed no effects on FAT10 activation, the M594L/H599N double mutant displayed a reduced activity (Figure 24G). The conserved residues Val571 and Ile891 of UBA6 engage in hydrophobic interactions with Ile163 of C-FAT10, further locking the FAT10 tail in position to facilitate catalysis (Figure 24E).

As stated above, the N-terminal domain of FAT10 (N-FAT10) also adopts the ubiquitin-like fold. In an alternative scenario it was assumed that N-FAT10 might

bind to UBA6 in the same orientation modeled for ubiquitin and C-FAT10 and the corresponding model was generated by superimposing N-FAT10 with ubiquitin (Figure 24D). While Ile44, the central residue of the hydrophobic patch of ubiquitin is replaced by the type-conserved Leu51 of N-FAT10 and hence seems to support this hypothesis, other predicted interactions clearly argue against this model. The residue Thr77 of N-FAT10 is in the position of His68 which is situated at the top of UBA6 hydrophobic platform (Figure 24D) and may reduce the hydrophobic interactions intensity which would impair the binding of N-FAT10. In particular, the hydrophobic residues Leu8 and Val70 of ubiquitin are replaced by the charged residues Glu15 and Lys79 of N-FAT10 (Figure 24D), and these alterations are predicted to severely interrupt the binding of N-FAT10 to UBA6. In addition, it remains totally unclear how the C-terminal tail of FAT10 could reach the adenylation active site in UBA6 if N-FAT10 would bind where ubiquitin is predicted to bind as it follows directly after the core of C-FAT10 and there would simply be not enough flexibility to reach the active site. Hence, N-FAT10 is unlikely to bind to the hydrophobic patch of UBA6 and adopt the predicted position of ubiquitin.

Immediately prior to the di-Gly motif at its C-terminal tail, ubiquitin harbors a Leu-Arg-Leu-Arg motif while FAT10 contains a more hydrophobic Cys-Tyr-Cys-Ile motif (Figure 24F). By swapping these two motifs ubiquitin-CYCI and FAT10-LRLR variants were generated. Studies showed that FAT10-LRLR behaved just like the ubiquitin wild-type as it can be activated by both HsUBA1 and UBA6. FAT10-LRLR could even be transferred to ubiquitin-specific E2 enzymes such as UBE2D3, BIRC6 by UBA6 (Schelpe, Monte et al. 2016). Substitutions in the C-terminal tail were investigated further, which revealed that the leucine substitution of Cys162 (C162L) of FAT10 was enough to alter the selectivity of FAT10 (Figure 24H, upper right panel). A FAT10 variant in which Cys7, Cys9 residues were substituted with Thr, Cys134 was replaced with Leu and Cys160 and Cys162 were replaced with Ser (FAT10 C0) was generated to increase the stability and solubility during structural studies with FAT10 (Aichem, Anders et al. 2018). Interestingly, FAT10 C0 showed a robust activation by UBA6, which was even stronger than FAT10 wildtype, possibly due to its increased solubility. Moreover, FAT10 C0 can also be activated by HsUBA1 (Figure 24H). However, the ubiquitin-CYCI variant, unexpectedly, was still found to be activated by HsUBA1, however, with decreased efficiency (Schelpe, Monte et al. 2016). These data raise several points: (1) The cysteine residues, specifically Cys162, of FAT10 are important for its specificity; (2) The adenylation sites of UBA6 and UBA1 may harbor important elements for their selectivity towards

FAT10; (3) The cysteine residues of FAT10 restrain its activation, since FAT10 C0 increased its activation; (4) Without Cys162, FAT10 can bind to UBA1 despite the differences in its surface properties and that of ubiquitin as well as the differences between HsUBA1 and UBA6 structures mentioned above (see Section 3.5.3 and 3.5.4); (5) Cys162 of FAT10 and the interactions involving the FAT10 C-terminal tail with the UBA6 adenylation site are not the only elements that define FAT10 specificity since ubiquitin-CYCI was still found to be activated by HsUBA1 (Schelpe, Monte et al. 2016). The sequence similarity between UBA6 and the UBA1 orthologs in the adenylation site and the crossover loop (Figure 24E) further supports the argument that FAT10 specificity determinants are not only located within the adenylation site.

In a control experiment the surprising discovery was made that ScUBA1 can activate FAT10 *in vitro* (Figure 24I), although at a reduced level compared to UBA6. This possibly suggests that a UBA1 ancestor had the ability to activate ubiquitin and FAT10, while HsUBA1 during the course of evolution lost its ability to activate FAT10. We have tested the transfer of FAT10 by ScUBA1 *in vivo* as described in section 2.4. FAT10 (18.5 kDa) was well expressed in yeast (Figure 25: lane 1 and 3). However, non of the FAT10 conjugates was detected. This suggested that FAT10 was expressed in yeast, but neither conjugated to potential target proteins nor to ScUBA1 itself. ScUBA1 shares 52% and 42% sequence identity with HsUBA1 and UBA6, respectively. The heterodimeric SUMO E1 which shares only 29% sequence identity with UBA6 and ScUBA1 was reported to activate FAT10 (Aichem, Sailer et al. 2019). These very different E1s can both activate FAT10. These observations add additional restraints to the FAT10 activation mechanism. Multiple sequence alignments of E1 enzymes were generated and analyzed to identify residues in HsUBA1 which are different from UBA6 and ScUBA1 as possible determinants for FAT10 activation. However, the introduction of these residues into UBA6 could neither block the activation of FAT10 nor that of ubiquitin (data not shown).
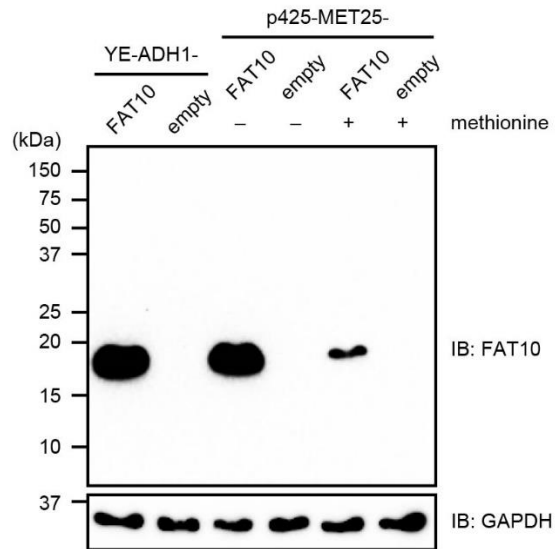
**Figure 25: Western blots of FAT10 expression in yeast. FAT10 was cloned into the two yeast expression vectors ADH1 and p425-MET25. MET25 promoter was induced by lack of methionine in the growth medium. The empty vectors were transformed as negative controls. GAPDH was blotted as a loading control.**

### 3.5.5. The interactions of UBA6 and USE1

Following the activation of either ubiquitin or FAT10, UBA6 would transfer the activated protein modifier to an E2 enzyme in a transthioesterification reaction. As mentioned earlier in the case of FAT10 the E2 enzyme USE1 plays a special role as it is the only family member to be modified with FAT10, hence the following section will focus on the predicted interactions between UBA6 and USE1.
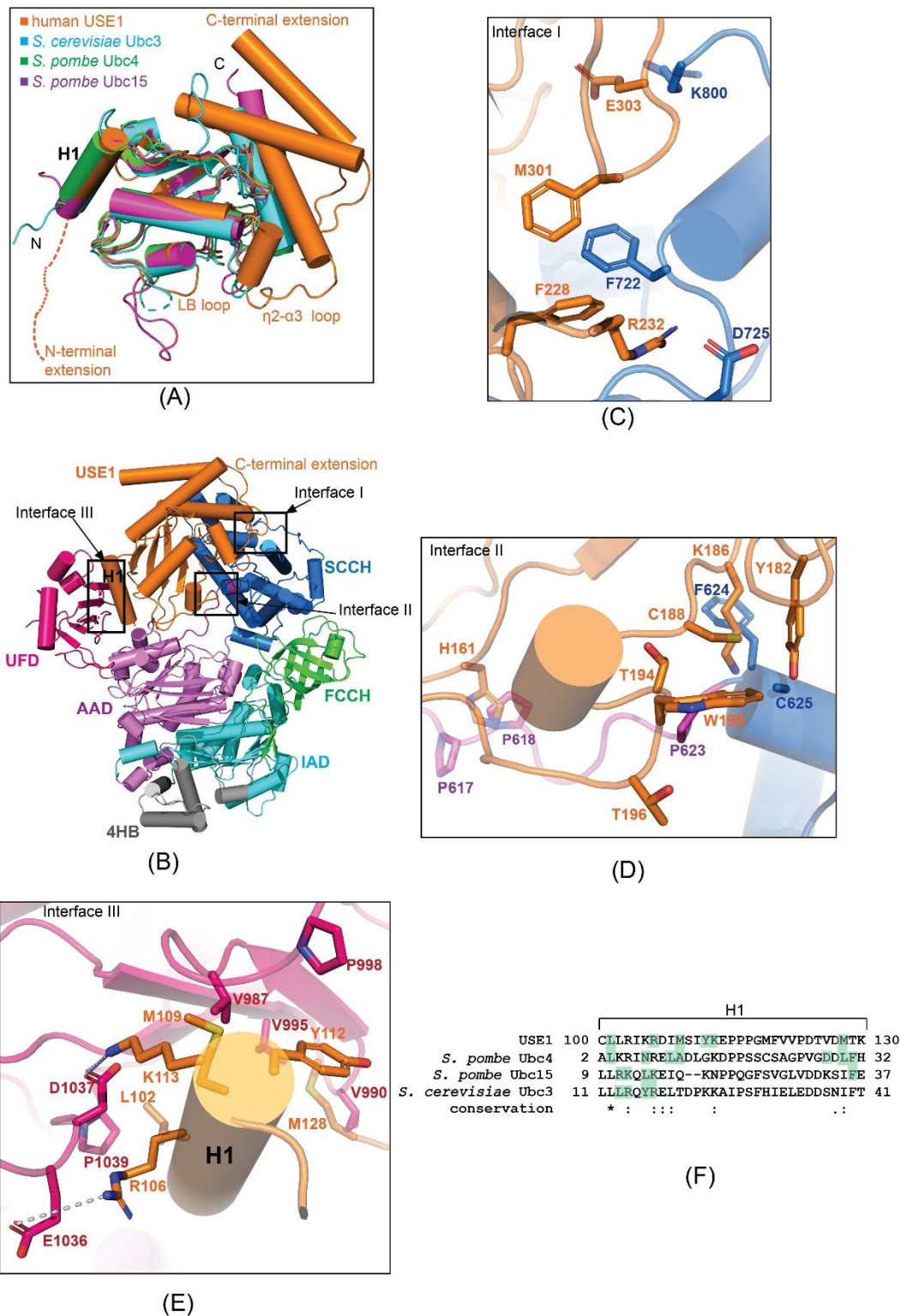
**Figure 26: Modeling the UBA6-USE1 interaction. (A) Superimposition of USE1 (orange) and selected yeast E2 enzymes for which structures in complex with Uba1 are known. *S. cerevisiae* Ubc3 is shown in cyan, *S. pombe* Ubc4 in green and *S. pombe* Ubc15 in magenta. (B) Model of USE1 (orange) in complex with UBA6, which is color coded according to its domain architecture. The three**

**binding interfaces are indicated by black rectangles. (C) Zoom into interface I which shows the interactions between USE1 and the SCCH domain of UBA6. (D) Enlarged view of interface II containing the interactions around the respective active site. (E) Close-up of interface III where helix H1 of USE1 contacts the UFD of UBA6. (F) Sequence alignment of helix H1 of the four E2 enzymes. Residues which are involved in interactions with UBA6 are shaded in green while the sole conserved residue is highlighted with "*" and type-conserved residues with ":" or ".".**

Among the available structures of the E2 enzymes visualized in complexes with either the *S. cerevisiae* or *S. pombe* UBA1 orthologs, Ubc4 from *S. pombe* (PDB: 4ii2) exhibits the highest sequence identity to USE1 with 38%. Hence, the model of the UBA6-USE1 complex (Figure 26B) was generated by superimposing UBA6 onto Uba1 and USE1 onto Ubc4 in the crystal structure of the UBA1-Ubc4 complex (PDB: 4ii2). Independent superimpositions of USE1 and three aforementioned E2 enzymes highlighted its unique features such as its N-terminal extension (not present in the structure) and $\alpha$-helical C-terminal extension as well as the loop connecting helices η2 and α3 (referred as η2-α3 loop) (Figure 26A). The interactions between UBA6 and USE1 again involve three separate interfaces, which will be analyzed separately (Figure 26B). Interface I involves contacts between USE1 and the SCCH domain of UBA6. These include ionic interactions between Asp725 and Lys800 of UBA6 and Arg232 and Glu303, respectively, of USE1 (Figure 26C). Phe722 of UBA6 is predicted to be sandwiched between the side chains of Phe228 and Met301 of USE1. Notably, Phe228 and Arg232 belong to the η2-α3 loop of USE1 while Met301 and Glu303 belong to the C-terminal extensions of USE1 (Figure 26C), thus suggesting that these interactions are likely to be unique in the UBA6-USE1 complex. Interface II includes the close interaction of the USE1 catalytic Cys188 and the catalytic Cys625 of UBA6 (Figure 26D) and residues Thr194, Trp195 and Thr196 of the LB loop. This loop was reported to be important for the selectivity of USE1 since, after its deletion, USE1 was recognized by HsUBA1 (Schelpe, Monte et al. 2016). The hydrophobicity of the LB loop in combination with Tyr182 and the catalytic Cys188 of USE1 may facilitate the interactions with a hydrophobic area in UBA6 featuring Pro623, Phe624 and Cys625 (Figure 26D). In addition, His161 of USE1 engages in hydrophobic interactions with Pro617 and Pro618 in the crossover loop of UBA6, thus possibly further strengthening the complex. Interface III harbors the interaction of helix H1 of USE1 and the UFD of UBA6 (Figure 26E). On one face of helix H1 ionic interactions

involving Lys113 and Arg106 of USE1 are predicted to take place with Asp1037 and Glu1036 of UBA6. On the other face of helix H1 hydrophobic interactions are predicted involving Leu102, Met109, Tyr112 and Met128 of USE1 with Pro1039, Val987, Val995, Pro998 and Val990 (Figure 26E). Specific interactions of residues located in helix H1 and the UFD are known to define the selectivities of the E2s toward the E1s (Lee and Schindelin 2008, Lv, Rickman et al. 2017, Lv, Williams et al. 2018) (Olsen and Lima 2013). Therefore, the residues which are involved in these interactions vary among the E2s (Figure 26F). Consequently, residues of the UFD which engage in interactions with helix H1 are not highly conserved among different E1 enzymes (Figure 27C). Complicating the analysis is the finding than an E1 may adopt different interactions with different E2s (Williams, Qie et al. 2019).

The structural and sequence alignments of the UFDs of different E1 enzymes suggest unique features of the UBA6 UFD (Figure 27). The deletion of two residues between Ala1002 and Lys1003 and 4 residues between Leu1024 and Thr1025 of UBA6 are reflected in the structure of its UFD (Figure 27C). The β27-β28 and H43-β27 loops in the UFD of UBA6 are dramatically shifted by 11 Å and 10 Å, respectively, compared to their counterparts in HsUBA1 (Figure 27A and Figure 27B). The aforementioned deletions are compensated at the C-terminus of UBA6 by the presence of 4 extra residues, leading to a shift of the UBA6 C-terminal tail (Figure 27B). HsUBA1 features an insertion in its UFD, which is referred to as the β27-H31 insertion loop, which is not conserved among the ubiquitin E1 enzymes from *S. cerevisiae* and *S. pombe* (Lv, Williams et al. 2018) but is present in vertebrate UBA1 orthologs. The differences in the UFD domains of UBA6 and the UBA1 orthologs may explain their selectivities towards specific E2 enzymes.

The model of UBA6 and USE1 complex suggests conserved interactions as well as unique contacts between the two enzymes (Figure 26). Helix H1 at the N-terminal end of USE1 is certainly involved in interactions with the UFD domain of UBA6 (Figure 26E). This helix was also reported to hold key elements that define the selectivities of the other E1 enzymes (Lee and Schindelin 2008, Olsen and Lima 2013, Lv, Rickman et al. 2017, Lv, Williams et al. 2018, Williams, Qie et al. 2019). In addition, USE1 has a long N-terminal extension that was reported to be important for its selectivity towards HsUBA1 (Schelpe, Monte et al. 2016). Hence the question arises to what extent this helix H1 plays a role in the selectivity of USE1. A similar question would apply for the interactions of the η2-α3 loop and the C-terminal extension of USE1 in contact with the SCCH domain of UBA6 when the deletion of

the LB loop of USE1 alone already altered the selectivity of USE1 (Schelpe, Monte et al. 2016) (Figure 26D). Moreover, the mechanisms of how the N-terminal extension and the LB loops could define the specificity of USE1 are still unknown. It is hence highly relevant to further investigate the interactions between UBA6 and USE1 using biochemical and structural methods.
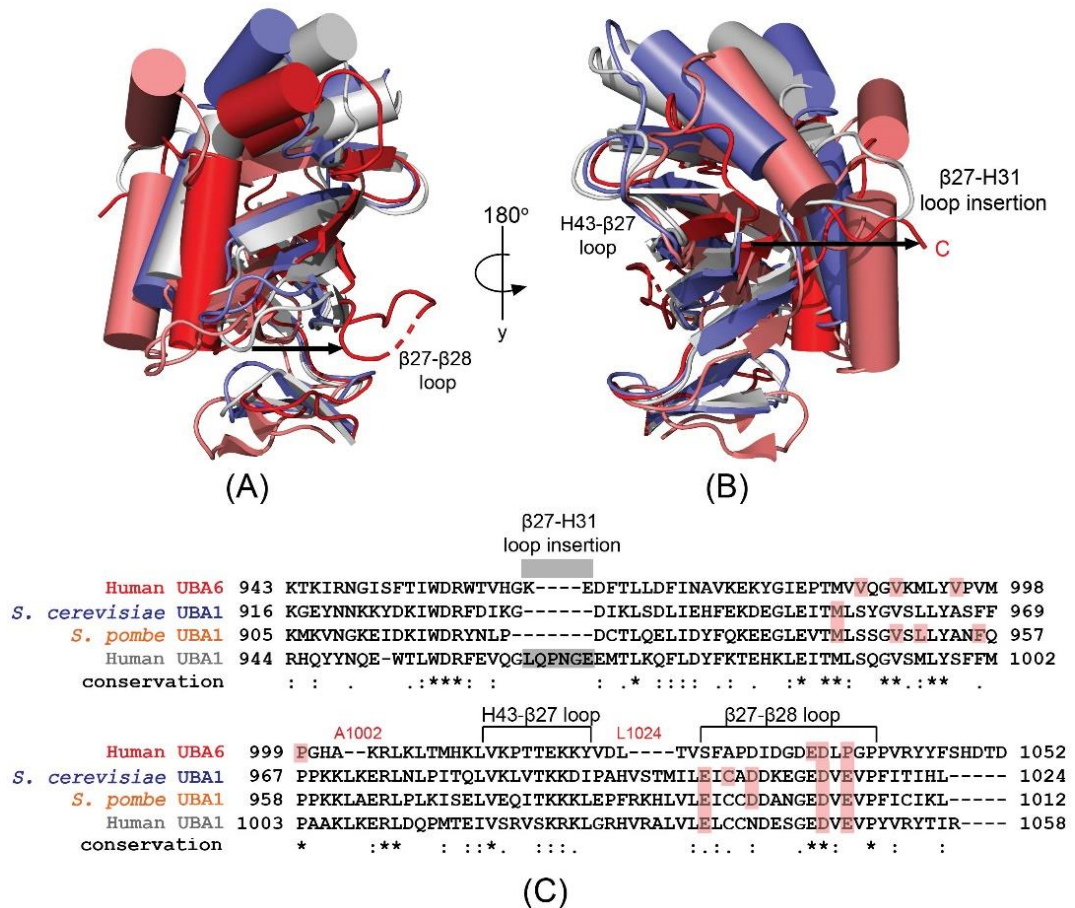


**Figure 27: Comparison of the UFD of UBA6 and the ubiquitin-specific E1 enzymes. (A) Front side of the superimposed UFDs of UBA6 (red), HsUBA1 (grey) (PDB: 6dc6), ScUBA1 (blue) (PDB: 3cmm) and *S. pombe* UBA1 (pink) (PDB: 4ii2). The black arrow indicates the shift of the β27-β28 loop. (B) Back side of the superposition. In this case the black arrow indicates the shift of the C-terminus of UBA6 and the white arrow indicates the shift of the H43-β27 loop (C) Sequence alignment of the UFDs being discussed in this section.**

## 3.5.6. Attempts to swap the selectivity for FAT10/ubiquitin between UBA6 and UBA1

The structure of UBA6 and preliminary biochemical studies suggested the binding mode and the specific interactions of ubiquitin with UBA6. While the 'Ile44 hydrophobic patch' of ubiquitin is predicted to engage in interactions with the unique hydrophobic platform on the β23 and β24 strands of UBA6, Arg72 presumably engages in an ionic interaction with Glu601 located on the β23-strand in the adenylation site of UBA6 (Figure 23C and Figure 23D). Substitutions of either of these two elements with the corresponding parts in HsUBA1 abolished the activation of ubiquitin by UBA6 (Figure 23F and Figure 28G), even though HsUBA1 is also a ubiquitin activating enzyme. This suggested that an intact hydrophobic platform and adenylation site of UBA6 is indispensable for ubiquitin activation. Unlike ubiquitin, FAT10 activation by UBA6 cannot be easily suppressed. Despite many efforts, it was not possible to block the activation FAT10. As can be seen from Figure 28A-E almost the entire front side of UBA6 including the entire adenylation site was replaced with the corresponding parts of HsUBA1. None of these UBA6 chimeras could inactivate UBA6 towards the activation of FAT10 (Figure 28G upper panel). Perhaps, replacing a single domain is not enough to block the activation of FAT10. Hence, an alternative would be to try combinations of the UBA6 chimeric constructs. Ultimately, the complete front side of UBA6 could be swapped with the front side of HsUBA1 as shown for a hypothetical construct in Figure 28F. Since all the UBA6 chimeras that were generated so far remained active, it is reasonable to expect that the UBA6-front side chimera would still be folded properly. However, it would be meaningless if blocking FAT10 activation would also prevent the activation of ubiquitin, hence one may not interfere with the adenylation site and the hydrophobic platform of UBA6.
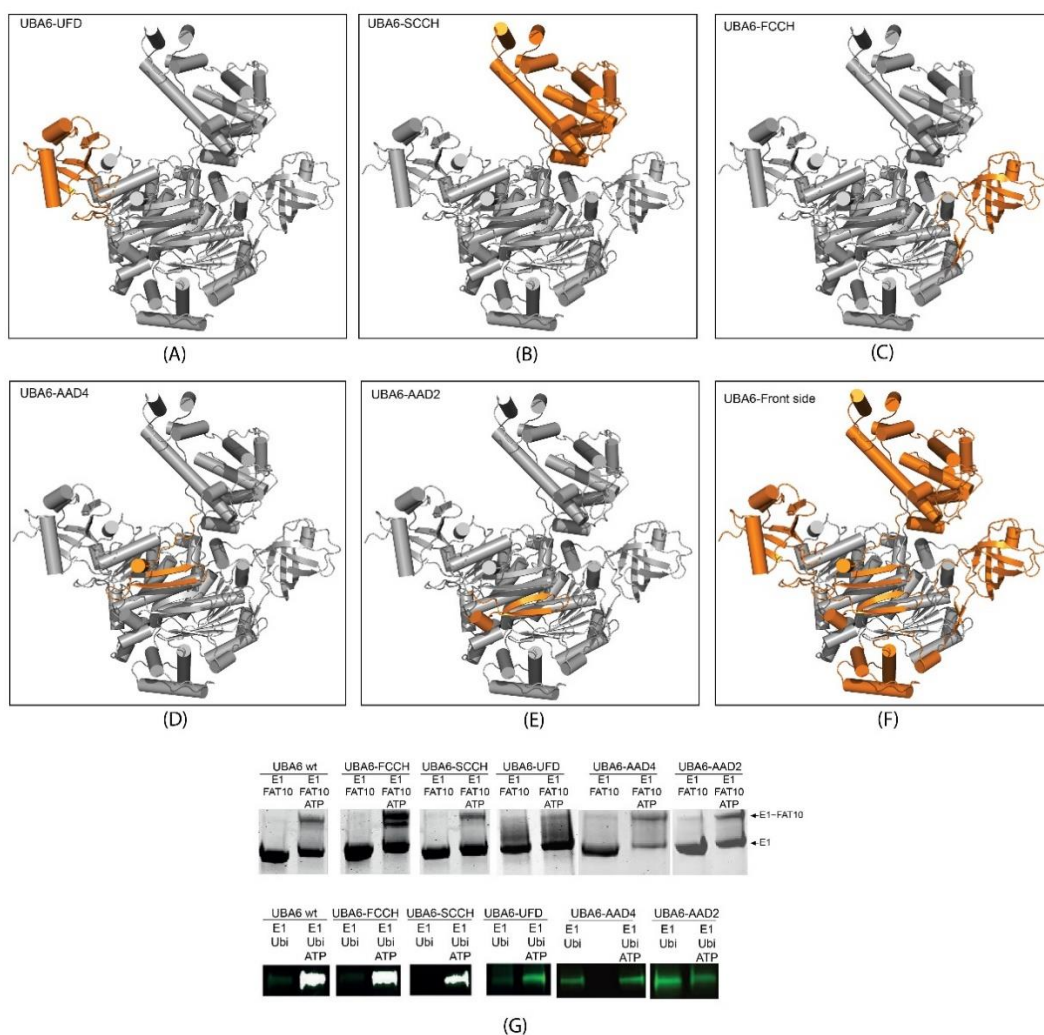
**Figure 28: Characterization of chimeric UBA6 constructs. (A-E) UBA6 structures in which individual domains which were replaced with the corresponding regions of HsUBA1 are highlighted in orange. (F) A proposed chimeric protein in which all individual domain swaps are combined. (G) Activity assays monitoring formation of the E1~FAT10 (top) and E1~ubiquitin (bottom) complexes.**

### 3.5.7. Inhibition of UBA6 by adenosyl sulfamates

As mentioned in Section 1.2.2, the adenosyl sulfamates ABPA3, MLN4924 and MLN7243 are potent inhibitors of different E1 enzymes. The potency of these inhibitors on UBA6 was analyzed in a qualitative manner by studying how these compounds inhibited the formation of either the UBA6~FAT10 or UBA6~ubiquitin intermediates. These assays revealed that MLN4924 is a really weak inhibitor of UBA6 for both FAT10 and ubiquitin activation showing hardly any inhibitory effect at concentrations as high as 0.7 mM. In contrast, ABPA3 and MLN7243 are much

more potent inhibitors with MLN7243 being slightly more effective than ABPA3 (Figure 29). In the case of UBA6~ubiquitin formation both compounds show inhibitory effects already in the high nanomolar range. In comparison, they seem to be less effective in blocking UBA6~FAT10 formation by roughly one order of magnitude, however, due to differences in the visualization of the respective adduct these differences should not be overinterpreted at this point. These data would be contradictory to those reported for another adenosine sulfamate referred to as Compound 1, which displayed a potent inhibitory effect on ubiquitin but not FAT10 activation by UBA6 (Gavin, Chen et al. 2012). This was attributed to the weaker binding of the FAT10-compound I adduct compared to the ubiquitin-compound I adduct, which is contradictory to the reported strong affinity of FAT10 to UBA6.
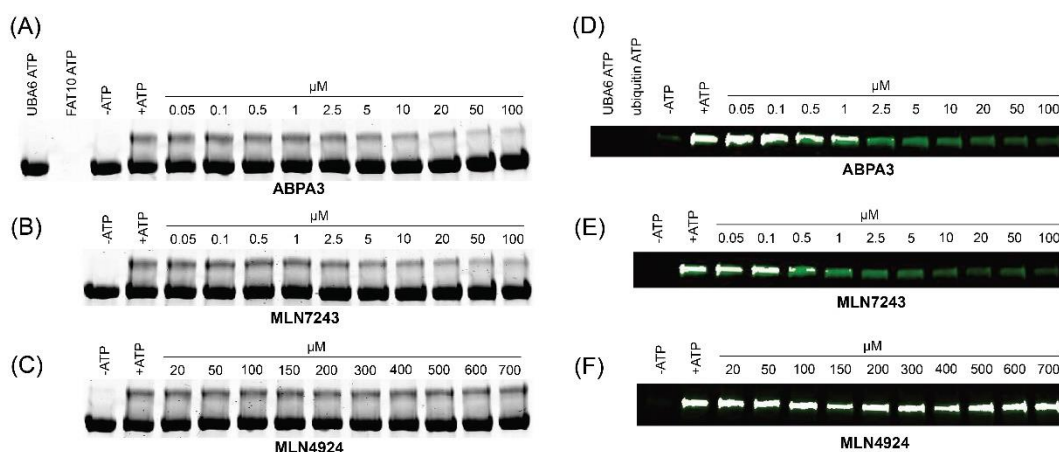


**Figure 29: Inhibition of UBA6 by adenosylsulfamates. Inhibition of FAT10 activation by ABPA3 (A), MLN7243 (B), MLN4924 (C) and ubiquitin activation by ABPA3 (D), MLN7243 (E), MLN4924 (F).**

## 4. OUTLOOK

As mentioned in section 1.2.1. FAT10ylation can be inhibited by mutating the di-Gly tail of FAT10, however, FAT10 is much more than a proteasome-targeting signal since it noncovalently interacts with other partners which are involved in signaling pathways or cancer development (Aichem and Groettrup 2020). FAT10 was also reported to bind noncovalently to the SUMO E1, thus inhibiting SUMOylation. FAT10 was found to be even activated by SUMO E1 and then transferred to the SUMO E2 but to be conjugated to substrates (Aichem, Sailer et al. 2019). Therefore, mutations of the di-Gly tail of FAT10 may also affect its function as a noncovalent binding partner or a covalent inhibitor of SUMOylation. Ideally, to stop the FAT10ylation, one could target UBA6. However, UBA6 exhibits dual specificity and global inhibition of UBA6 will therefore display unwanted side effects. Since FAT10ylation and ubiquitylation are involved in many important cell processes as well as severe diseases (Aichem and Groettrup 2016, Aichem and Groettrup 2020), attempts have been made to selectively inhibit FAT10ylation.

So far, it is still unknown where N-FAT10 binds to in UBA6 and whether it plays any role in the stable activation of FAT10 and could be responsible for the poor inhibitory effect of compound I. Studying the determinants of the dual specificity of UBA6 would make it possible to inhibit FAT10ylation without affecting its ubiquitylation, thus avoiding the side effects. This would ultimately allow researchers to study the consequences of blocking the entire FAT10ylation pathway in cellular processes and different diseases.

## 5. LIST OF PUBLICATIONS

**Peer reviewed publications**

Green, J.L., et al., Ubiquitin activation is essential for schizont maturation in Plasmodium falciparum blood-stage development. PLoS Pathog, 2020. **16**(6): p. e1008640.

# 6. ACKNOWLEDGMENT

# 7. ABBREVIATION

| | |
|---|---|
| o | Degree |
| $^oC$ | Degree Celsius |
| Å | Ångström |
| $OD_{600}$ | Optical density at 600 nm |
| ATP | Adenosine triphosphate |
| AMP | Adenosine monophosphate |
| BESSY | Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung m. b. H. |
| PDB | Protein data bank |
| kDa | KiloDaltons |
| DUB | Deubiquitinases |
| DNA | Deoxyribonucleic acid |
| DNase | Deoxyribonuclease |
| *E. coli* | *Escherichia coli* |
| *S. cerevisiae* | *Saccharomyces cerevisiae* |
| *S. Pombe* | *Schizosaccharomyces Pombe* |
| EDTA | Ethylenediaminetetraacetic acid |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |

| | |
|---|---|
| IPTG | Isopropyl β- d-1-thiogalactopyranoside |
| IR | Infrared |
| LB | Lysogen broth |
| M | Molar |
| min | Minute |
| mM | Milimolar |
| mL | Mililiter |
| nm | Nanometer |
| MWCO | Molecular weight cut off |
| NF-κB | Nuclear Factor Kappa B |
| SDS-PAGE | Sodium dodecyl sulphate–polyacrylamide gel electrophoresis |
| PEG | Poly ethylene glycol |
| PSM | Pre-stained protein marker |
| rmsd | Root mean square deviation |
| TB | Terrific Broth |
| TCEP | Tris(2-carboxyethyl)phosphine |
| TEV | Tobacco Etch Virus |
| UBL | Ubiquitin-like proteins |
| WT | Wild type |

| µl | Microliter |
| µM | Micromolar |

## 8. AFFIDAVIT

I here by confirm that my thesis entitled "Understanding the dual specificity of UBA6" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

**Würzburg,**

**Place, Date**

**Signature**

### Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation „Einblick in die duale Spezifität von UBA6" eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

**Würzburg,**

**Ort, Datum**

**Signature**

# REFERENCES

Adams, J., M. Behnke, S. Chen, A. A. Cruickshank, L. R. Dick, L. Grenier, J. M. Klunder, Y. T. Ma, L. Plamondon and R. L. Stein (1998). "Potent and selective inhibitors of the proteasome: dipeptidyl boronic acids." Bioorg Med Chem Lett **8**(4): 333-338.

Aichem, A., S. Anders, N. Catone, P. Rossler, S. Stotz, A. Berg, R. Schwab, S. Scheuermann, J. Bialas, M. C. Schutz-Stoffregen, G. Schmidtke, C. Peter, M. Groettrup and S. Wiesner (2018). "The structure of the ubiquitin-like modifier FAT10 reveals an alternative targeting mechanism for proteasomal degradation." Nat Commun **9**(1): 3321.

Aichem, A., A. N. Boehm, N. Catone, G. Schmidtke and M. Groettrup (2019). "Analysis of modification and proteolytic targeting by the ubiquitin-like modifier FAT10." Methods Enzymol **618**: 229-256.

Aichem, A., N. Catone and M. Groettrup (2014). "Investigations into the auto-FAT10ylation of the bispecific E2 conjugating enzyme UBA6-specific E2 enzyme 1." FEBS J **281**(7): 1848-1859.

Aichem, A. and M. Groettrup (2016). "The ubiquitin-like modifier FAT10 in cancer development." Int J Biochem Cell Biol **79**: 451-461.

Aichem, A. and M. Groettrup (2020). "The ubiquitin-like modifier FAT10 - much more than a proteasome-targeting signal." J Cell Sci **133**(14).

Aichem, A., B. Kalveram, V. Spinnenhirn, K. Kluge, N. Catone, T. Johansen and M. Groettrup (2012). "The proteomic analysis of endogenous FAT10 substrates identifies p62/SQSTM1 as a substrate of FAT10ylation." J Cell Sci **125**(Pt 19): 4576-4585.

Aichem, A., C. Pelzer, S. Lukasiak, B. Kalveram, P. W. Sheppard, N. Rani, G. Schmidtke and M. Groettrup (2010). "USE1 is a bispecific conjugating enzyme for ubiquitin and FAT10, which FAT10ylates itself in cis." Nat Commun **1**: 13.

Aichem, A., C. Sailer, S. Ryu, N. Catone, N. Stankovic-Valentin, G. Schmidtke, F. Melchior, F. Stengel and M. Groettrup (2019). "The ubiquitin-like modifier FAT10 interferes with SUMO activation." Nat Commun **10**(1): 4452.

An, H. and A. V. Statsyuk (2015). "An inhibitor of ubiquitin conjugation and aggresome formation." Chem Sci **6**(9): 5235-5245.

Andreev, V. P., V. A. Petyuk, H. M. Brewer, Y. V. Karpievitch, F. Xie, J. Clarke, D. Camp, R. D. Smith, A. P. Lieberman, R. L. Albin, Z. Nawaz, J. El Hokayem and A. J. Myers (2012). "Label-free quantitative LC-MS proteomics of Alzheimer's disease and normally aged human brains." J Proteome Res **11**(6): 3053-3067.

Assawamakin, A., D. Wattanasirichaigoon, C. Tocharoentanaphol, S. Waeteekul, M. Tansatit, W. Thongnoppakhun and C. Limwongse (2012). "A novel maternally-derived insertional translocation resulting in partial trisomy 4q13.2-q22.1 with complex translocation t(8;20) in a family with intellectual disability." Am J Med Genet A **158A**(4): 901-908.

Bayer, P., A. Arndt, S. Metzger, R. Mahajan, F. Melchior, R. Jaenicke and J. Becker (1998). "Structure determination of the small ubiquitin-related modifier SUMO-1." J Mol Biol **280**(2): 275-286.

Bedford, L., J. Lowe, L. R. Dick, R. J. Mayer and J. E. Brownell (2011). "Ubiquitin-like protein conjugation and the ubiquitin-proteasome system as drug targets." Nat Rev Drug Discov **10**(1): 29-46.

Bett, J. S., N. Kanuga, E. Richet, G. Schmidtke, M. Groettrup, M. E. Cheetham and J. van der Spuy (2012). "The inherited blindness protein AIPL1 regulates the ubiquitin-like FAT10 pathway." PLoS One **7**(2): e30866.

Bhatia, S., A. C. Pavlick, P. Boasberg, J. A. Thompson, G. Mulligan, M. D. Pickard, H. Faessel, B. J. Dezube and O. Hamid (2016). "A phase I study of the investigational NEDD8-activating enzyme inhibitor pevonedistat (TAK-924/MLN4924) in patients with metastatic melanoma." Invest New Drugs **34**(4): 439-449.

Bialas, J., A. N. Boehm, N. Catone, A. Aichem and M. Groettrup (2019). "The ubiquitin-like modifier FAT10 stimulates the activity of deubiquitylating enzyme OTUB1." J Biol Chem **294**(12): 4315-4330.

Bialas, J., M. Groettrup and A. Aichem (2015). "Conjugation of the ubiquitin activating enzyme UBE1 with the ubiquitin-like modifier FAT10 targets it for proteasomal degradation." PLoS One **10**(3): e0120329.

Boehm, A. N., J. Bialas, N. Catone, A. Sacristan-Reviriego, J. van der Spuy, M. Groettrup and A. Aichem (2020). "The ubiquitin-like modifier FAT10 inhibits retinal PDE6 activity and mediates its proteasomal degradation." J Biol Chem **295**(42): 14402-14418.

Bogunovic, D., M. Byun, L. A. Durfee, A. Abhyankar, O. Sanal, D. Mansouri, S. Salem, I. Radovanovic, A. V. Grant, P. Adimi, N. Mansouri, S. Okada, V. L. Bryant, X. F. Kong, A. Kreins, M. M. Velez, B. Boisson, S. Khalilzadeh, U. Ozcelik, I. A. Darazam, J. W. Schoggins, C. M. Rice, S. Al-Muhsen, M. Behr, G. Vogt, A. Puel, J. Bustamante, P. Gros, J. M. Huibregtse, L. Abel, S. Boisson-Dupuis and J. L. Casanova (2012). "Mycobacterial disease and impaired IFN-gamma immunity in humans with inherited ISG15 deficiency." Science **337**(6102): 1684-1688.

Bohren, K. M., V. Nadkarni, J. H. Song, K. H. Gabbay and D. Owerbach (2004). "A M55V polymorphism in a novel SUMO gene (SUMO-4) differentially activates heat shock transcription factors and is associated with susceptibility to type I diabetes mellitus." J Biol Chem **279**(26): 27233-27238.

Buchsbaum, S., B. Bercovich, T. Ziv and A. Ciechanover (2012). "Modification of the inflammatory mediator LRRFIP2 by the ubiquitin-like protein FAT10 inhibits its activity during cellular response to LPS." Biochem Biophys Res Commun **428**(1): 11-16.

Chan, N. L. and C. P. Hill (2001). "Defining polyubiquitin chain topology." Nat Struct Biol **8**(8): 650-652.

Chen, J. J., C. A. Tsu, J. M. Gavin, M. A. Milhollen, F. J. Bruzzese, W. D. Mallender, M. D. Sintchak, N. J. Bump, X. Yang, J. Ma, H. K. Loke, Q. Xu, P. Li, N. F. Bence, J. E. Brownell and L. R. Dick (2011). "Mechanistic studies of substrate-assisted inhibition of ubiquitin-activating enzyme by adenosine sulfamate analogues." J Biol Chem **286**(47): 40867-40877.

Chen, Z., W. Zhang, Z. Yun, X. Zhang, F. Gong, Y. Wang, S. Ji and L. Leng (2018). "Ubiquitinlike protein FAT10 regulates DNA damage repair via modification of proliferating cell nuclear antigen." Mol Med Rep **17**(6): 7487-7496.

Chen, Z., W. Zhang, Z. Yun, X. Zhang, F. Gong, Y. Wang, S. Ji and L. Leng (2018). "Ubiquitinlike protein FAT10 regulates DNA damage repair via modification of proliferating cell nuclear antigen." Mol Med Rep.

Chiu, Y. H., Q. Sun and Z. J. Chen (2007). "E1-L2 activates both ubiquitin and FAT10." Mol Cell **27**(6): 1014-1023.

Choi, Y., J. K. Kim and J. Y. Yoo (2014). "NFkappaB and STAT3 synergistically activate the expression of FAT10, a gene counteracting the tumor suppressor p53." Mol Oncol **8**(3): 642-655.

Ciechanover, A., S. Elias, H. Heller and A. Hershko (1982). ""Covalent affinity" purification of ubiquitin-activating enzyme." J Biol Chem **257**(5): 2537-2542.

Ciechanover, A., H. Heller, R. Katz-Etzion and A. Hershko (1981). "Activation of the heat-stable polypeptide of the ATP-dependent proteolytic system." Proc Natl Acad Sci U S A **78**(2): 761-765.

da Silva, S. R., S. L. Paiva, M. Bancerz, M. Geletu, A. M. Lewis, J. Chen, Y. Cai, J. L. Lukkarila, H. Li and P. T. Gunning (2016). "A selective inhibitor of the UFM1-activating enzyme, UBA5." Bioorg Med Chem Lett **26**(18): 4542-4547.

da Silva, S. R., S. L. Paiva, J. L. Lukkarila and P. T. Gunning (2013). "Exploring a new frontier in cancer treatment: targeting the ubiquitin and ubiquitin-like activating enzymes." J Med Chem **56**(6): 2165-2177.

Day, C. and J. D. Shepherd (2015). "Arc: building a bridge from viruses to memory." Biochem J **469**(1): e1-3.

de Medina-Redondo, M. and P. Meraldi (2011). "The spindle assembly checkpoint: clock or domino?" Results Probl Cell Differ **53**: 75-91.

Deshaies, R. J. and C. A. Joazeiro (2009). "RING domain E3 ubiquitin ligases." Annu Rev Biochem **78**: 399-434.

Ding, H., Y. Xu, Q. Chen, H. Dai, Y. Tang, J. Wu and Y. Shi (2005). "Solution structure of human SUMO-3 C47S and its binding surface for Ubc9." Biochemistry **44**(8): 2790-2799.

Duda, D. M., L. A. Borg, D. C. Scott, H. W. Hunt, M. Hammel and B. A. Schulman (2008). "Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation." Cell **134**(6): 995-1006.

Ebstein, F., N. Lange, S. Urban, U. Seifert, E. Kruger and P. M. Kloetzel (2009). "Maturation of human dendritic cells is accompanied by functional remodelling of the ubiquitin-proteasome system." Int J Biochem Cell Biol **41**(5): 1205-1215.

Echchgadda, I., C. C. Roth, C. Z. Cerna and G. J. Wilmink (2013). "Temporal gene expression kinetics for human keratinocytes exposed to hyperthermic stress." Cells **2**(2): 224-243.

Emanuele, M. J., A. E. Elia, Q. Xu, C. R. Thoma, L. Izhar, Y. Leng, A. Guo, Y. N. Chen, J. Rush, P. W. Hsu, H. C. Yen and S. J. Elledge (2011). "Global identification of modular cullin-RING ligase substrates." Cell **147**(2): 459-474.

Fan, W., W. Cai, S. Parimoo, D. C. Schwarz, G. G. Lennon and S. M. Weissman (1996). "Identification of seven new human MHC class I region genes around the HLA-F locus." Immunogenetics **44**(2): 97-103.

Fletcher, A. J., D. E. Christensen, C. Nelson, C. P. Tan, T. Schaller, P. J. Lehner, W. I. Sundquist and G. J. Towers (2015). "TRIM5alpha requires Ube2W to anchor Lys63-linked ubiquitin chains and restrict reverse transcription." EMBO J **34**(15): 2078-2095.

Florke Gee, R. R., H. Chen, A. K. Lee, C. A. Daly, B. A. Wilander, K. Fon Tacer and P. R. Potts (2020). "Emerging roles of the MAGE protein family in stress response pathways." J Biol Chem **295**(47): 16121-16155.

Fukuda, I., A. Ito, G. Hirai, S. Nishimura, H. Kawasaki, H. Saitoh, K. Kimura, M. Sodeoka and M. Yoshida (2009). "Ginkgolic acid inhibits protein SUMOylation by blocking formation of the E1-SUMO intermediate." Chem Biol **16**(2): 133-140.

Fukuda, I., A. Ito, M. Uramoto, H. Saitoh, H. Kawasaki, H. Osada and M. Yoshida (2009). "Kerriamycin B inhibits protein SUMOylation." J Antibiot (Tokyo) **62**(4): 221-224.

Furukawa, K., N. Mizushima, T. Noda and Y. Ohsumi (2000). "A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes." J Biol Chem **275**(11): 7462-7465.

Garaude, J., R. Farras, G. Bossis, S. Charni, M. Piechaczyk, R. A. Hipskind and M. Villalba (2008). "SUMOylation regulates the transcriptional activity of JunB in T lymphocytes." J Immunol **180**(9): 5983-5990.

Gavin, J. M., J. J. Chen, H. Liao, N. Rollins, X. Yang, Q. Xu, J. Ma, H. K. Loke, T. Lingaraj, J. E. Brownell, W. D. Mallender, A. E. Gould, B. S. Amidon and L. R. Dick (2012). "Mechanistic studies on activation of ubiquitin and di-ubiquitin-like protein, FAT10, by ubiquitin-like modifier activating enzyme 6, Uba6." J Biol Chem **287**(19): 15512-15522.

Gavin, J. M., K. Hoar, Q. Xu, J. Ma, Y. Lin, J. Chen, W. Chen, F. J. Bruzzese, S. Harrison, W. D. Mallender, N. J. Bump, M. D. Sintchak, N. F. Bence, P. Li, L. R. Dick, A. E. Gould and J. J. Chen (2014). "Mechanistic study of Uba5 enzyme and the Ufm1 conjugation pathway." J Biol Chem **289**(33): 22648-22658.

Groll, M., C. R. Berkers, H. L. Ploegh and H. Ovaa (2006). "Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20S proteasome." Structure **14**(3): 451-456.

Haas, A. L., J. V. Warms, A. Hershko and I. A. Rose (1982). "Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation." J Biol Chem **257**(5): 2543-2548.

Haglund, K. and I. Dikic (2005). "Ubiquitylation and cell signaling." EMBO J **24**(19): 3353-3359.

Haldeman, M. T., G. Xia, E. M. Kasperek and C. M. Pickart (1997). "Structure and function of ubiquitin conjugating enzyme E2-25K: the tail is a core-dependent activity element." Biochemistry **36**(34): 10526-10537.

Handley, P. M., M. Mueckler, N. R. Siegel, A. Ciechanover and A. L. Schwartz (1991). "Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1." Proc Natl Acad Sci U S A **88**(1): 258-262.

Hann, Z. S., C. Ji, S. K. Olsen, X. Lu, M. C. Lux, D. S. Tan and C. D. Lima (2019). "Structural basis for adenylation and thioester bond formation in the ubiquitin E1." Proc Natl Acad Sci U S A **116**(31): 15475-15484.

Hasselgren, P. O. and J. E. Fischer (1997). "The ubiquitin-proteasome pathway: review of a novel intracellular mechanism of muscle protein breakdown during sepsis and other catabolic conditions." Ann Surg **225**(3): 307-316.

He, X., J. Riceberg, T. Soucy, E. Koenig, J. Minissale, M. Gallery, H. Bernard, X. Yang, H. Liao, C. Rabino, P. Shah, K. Xega, Z. H. Yan, M. Sintchak, J. Bradley, H. Xu, M. Duffey, D. England, H. Mizutani, Z. Hu, J. Guo, R. Chau, L. R. Dick, J. E. Brownell, J. Newcomb, S. Langston, E. S. Lightcap, N. Bence and S. M. Pulukuri (2017). "Probing the roles of SUMOylation in cancer cell biology by using a selective SAE inhibitor." Nat Chem Biol **13**(11): 1164-1171.

Hideshima, T., P. Richardson, D. Chauhan, V. J. Palombella, P. J. Elliott, J. Adams and K. C. Anderson (2001). "The proteasome inhibitor PS-341 inhibits growth, induces apoptosis, and overcomes drug resistance in human multiple myeloma cells." Cancer Res **61**(7): 3071-3076.

Hipp, M. S., B. Kalveram, S. Raasi, M. Groettrup and G. Schmidtke (2005). "FAT10, a ubiquitin-independent signal for proteasomal degradation." Mol Cell Biol **25**(9): 3483-3491.

Hong, S. B., B. W. Kim, K. E. Lee, S. W. Kim, H. Jeon, J. Kim and H. K. Song (2011). "Insights into noncanonical E1 enzyme activation from the structure of autophagic E1 Atg7 with Atg8." Nat Struct Mol Biol **18**(12): 1323-1330.

Hori, T., F. Osaka, T. Chiba, C. Miyamoto, K. Okabayashi, N. Shimbara, S. Kato and K. Tanaka (1999). "Covalent modification of all members of human cullin family proteins by NEDD8." Oncogene **18**(48): 6829-6834.

Hsiang, T. Y., C. Zhao and R. M. Krug (2009). "Interferon-induced ISG15 conjugation inhibits influenza A virus gene expression and replication in human cells." J Virol **83**(12): 5971-5977.

Huang, S. C., S. Adhikari, J. E. Brownell, E. F. Calderwood, J. Chouitar, N. R. D'Amore, D. B. England, K. Foley, S. J. Harrison, P. J. LeRoy, D. Lok, A. Lublinsky, L. T. Ma, S. Menon, Y. Yang, J. Zhang and A. E. Gould (2020). "Discovery and optimization of pyrazolopyrimidine sulfamates as ATG7 inhibitors." Bioorg Med Chem **28**(19): 115681.

Huang, W. C., T. P. Ko, S. S. Li and A. H. Wang (2004). "Crystal structures of the human SUMO-2 protein at 1.6 A and 1.2 A resolution: implication on the functional differences of SUMO proteins." Eur J Biochem **271**(20): 4114-4122.

Huang, X. and V. M. Dixit (2016). "Drugging the undruggables: exploring the ubiquitin system for drug development." Cell Res **26**(4): 484-498.

Hyer, M. L., M. A. Milhollen, J. Ciavarri, P. Fleming, T. Traore, D. Sappal, J. Huck, J. Shi, J. Gavin, J. Brownell, Y. Yang, B. Stringer, R. Griffin, F. Bruzzese, T. Soucy, J. Duffy, C. Rabino, J. Riceberg, K. Hoar, A. Lublinsky, S. Menon, M. Sintchak, N. Bump, S. M. Pulukuri, S. Langston, S. Tirrell, M. Kuranda, P. Veiby, J. Newcomb, P. Li, J. T. Wu, J. Powe, L. R. Dick, P. Greenspan, K. Galvin, M. Manfredi, C. Claiborne, B. S. Amidon and N. F. Bence (2018). "A small-molecule inhibitor of the ubiquitin activating enzyme for cancer treatment." Nat Med **24**(2): 186-193.

Ichimura, Y., T. Kirisako, T. Takao, Y. Satomi, Y. Shimonishi, N. Ishihara, N. Mizushima, I. Tanida, E. Kominami, M. Ohsumi, T. Noda and Y. Ohsumi (2000). "A ubiquitin-like system mediates protein lipidation." Nature **408**(6811): 488-492.

Jin, J., X. Li, S. P. Gygi and J. W. Harper (2007). "Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging." Nature **447**(7148): 1135-1138.

Johnson, E. S. (2004). "Protein modification by SUMO." Annu Rev Biochem **73**: 355-382.

Kannouche, P. L., J. Wing and A. R. Lehmann (2004). "Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage." Mol Cell **14**(4): 491-500.

Kerscher, O., R. Felberbaum and M. Hochstrasser (2006). "Modification of proteins by ubiquitin and ubiquitin-like proteins." Annu Rev Cell Dev Biol **22**: 159-180.

Keszei, A. F. and F. Sicheri (2017). "Mechanism of catalysis, E2 recognition, and autoinhibition for the IpaH family of bacterial E3 ubiquitin ligases." Proc Natl Acad Sci U S A **114**(6): 1311-1316.

Komander, D., M. J. Clague and S. Urbe (2009). "Breaking the chains: structure and function of the deubiquitinases." Nat Rev Mol Cell Biol **10**(8): 550-563.

Komander, D. and M. Rape (2012). "The ubiquitin code." Annu Rev Biochem **81**: 203-229.

Kuhn, D. J., Q. Chen, P. M. Voorhees, J. S. Strader, K. D. Shenk, C. M. Sun, S. D. Demo, M. K. Bennett, F. W. van Leeuwen, A. A. Chanan-Khan and R. Z. Orlowski (2007). "Potent activity of carfilzomib, a novel, irreversible inhibitor of the ubiquitin-proteasome pathway, against preclinical models of multiple myeloma." Blood **110**(9): 3281-3290.

Lake, M. W., M. M. Wuebbens, K. V. Rajagopalan and H. Schindelin (2001). "Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex." Nature **414**(6861): 325-329.

Lee, I. and H. Schindelin (2008). "Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes." Cell **134**(2): 268-278.

Lee, P. C., J. C. Dodart, L. Aron, L. W. Finley, R. T. Bronson, M. C. Haigis, B. A. Yankner and J. W. Harper (2013). "Altered social behavior and neuronal development in mice lacking the Uba6-Use1 ubiquitin transfer system." Mol Cell **50**(2): 172-184.

Leng, L., C. Xu, C. Wei, J. Zhang, B. Liu, J. Ma, N. Li, W. Qin, W. Zhang, C. Zhang, X. Xing, L. Zhai, F. Yang, M. Li, C. Jin, Y. Yuan, P. Xu, J. Qin, H. Xie, F. He and J. Wang (2014). "A proteomics strategy for the identification of FAT10-modified sites by mass spectrometry." J Proteome Res **13**(1): 268-276.

Li, P., W. Jiang, Q. Yu, W. Liu, P. Zhou, J. Li, J. Xu, B. Xu, F. Wang and F. Shao (2017). "Ubiquitination and degradation of GBPs by a Shigella effector to suppress host defence." Nature **551**(7680): 378-383.

Li, T., R. Santockyte, S. Yu, R. F. Shen, E. Tekle, C. G. Lee, D. C. Yang and P. B. Chock (2011). "FAT10 modifies p53 and upregulates its transcriptional activity." Arch Biochem Biophys **509**(2): 164-169.

Li, Y. Y., G. Y. Zhang, J. P. He, D. D. Zhang, X. X. Kong, H. M. Yuan and F. L. Chen (2017). "Ufm1 inhibits LPS-induced endothelial cell inflammatory responses through the NF-kappaB signaling pathway." Int J Mol Med **39**(5): 1119-1126.

Liang, J. R., E. Lingeman, T. Luong, S. Ahmed, M. Muhar, T. Nguyen, J. A. Olzmann and J. E. Corn (2020). "A Genome-wide ER-phagy Screen Highlights Key Roles of Mitochondrial Metabolism and ER-Resident UFMylation." Cell **180**(6): 1160-1177 e1120.

Liu, X., B. Zhao, L. Sun, K. Bhuripanyo, Y. Wang, Y. Bi, R. V. Davuluri, D. M. Duong, D. Nanavati, J. Yin and H. Kiyokawa (2017). "Orthogonal ubiquitin transfer identifies ubiquitination substrates under differential control by the two ubiquitin activating enzymes." Nat Commun **8**: 14286.

Liu, Y. C., J. Pan, C. Zhang, W. Fan, M. Collinge, J. R. Bender and S. M. Weissman (1999). "A MHC-encoded ubiquitin-like protein (FAT10) binds noncovalently to the spindle assembly checkpoint protein MAD2." Proc Natl Acad Sci U S A **96**(8): 4313-4318.

Lois, L. M. and C. D. Lima (2005). "Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1." EMBO J **24**(3): 439-451.

Lukasiak, S., C. Schiller, P. Oehlschlaeger, G. Schmidtke, P. Krause, D. F. Legler, F. Autschbach, P. Schirmacher, K. Breuhahn and M. Groettrup (2008). "Proinflammatory cytokines cause FAT10 upregulation in cancers of liver and colon." Oncogene **27**(46): 6068-6074.

Lv, Z., K. A. Rickman, L. Yuan, K. Williams, S. P. Selvam, A. N. Woosley, P. H. Howe, B. Ogretmen, A. Smogorzewska and S. K. Olsen (2017). "S. pombe Uba1-Ubc15 Structure Reveals a Novel Regulatory Mechanism of Ubiquitin E2 Activity." Mol Cell **65**(4): 699-714 e696.

Lv, Z., K. M. Williams, L. Yuan, J. H. Atkison and S. K. Olsen (2018). "Crystal structure of a human ubiquitin E1-ubiquitin complex reveals conserved functional elements essential for activity." J Biol Chem **293**(47): 18337-18352.

Lv, Z., L. Yuan, J. H. Atkison, K. M. Williams, R. Vega, E. H. Sessions, D. B. Divlianska, C. Davies, Y. Chen and S. K. Olsen (2018). "Molecular mechanism of a covalent allosteric inhibitor of SUMO E1 activating enzyme." Nat Commun **9**(1): 5145.

Matoso, E., J. B. Melo, S. I. Ferreira, A. Jardim, T. M. Castelo, A. Weise and I. M. Carreira (2013). "Insertional translocation leading to a 4q13 duplication including the EPHA5 gene in two siblings with attention-deficit hyperactivity disorder." Am J Med Genet A **161A**(8): 1923-1928.

Metlagel, Z., C. Otomo, G. Takaesu and T. Otomo (2013). "Structural basis of ATG3 recognition by the autophagic ubiquitin-like protein ATG12." Proc Natl Acad Sci U S A **110**(47): 18844-18849.

Metzger, M. B., J. N. Pruneda, R. E. Klevit and A. M. Weissman (2014). "RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination." Biochim Biophys Acta **1843**(1): 47-60.

Meulmeester, E. and F. Melchior (2008). "Cell biology: SUMO." Nature **452**(7188): 709-711.

Misra, M., M. Kuhn, M. Lobel, H. An, A. V. Statsyuk, C. Sotriffer and H. Schindelin (2017). "Dissecting the Specificity of Adenosyl Sulfamate Inhibitors Targeting the Ubiquitin-Activating Enzyme." Structure **25**(7): 1120-1129 e1123.

Mizushima, N., T. Noda, T. Yoshimori, Y. Tanaka, T. Ishii, M. D. George, D. J. Klionsky, M. Ohsumi and Y. Ohsumi (1998). "A protein conjugation system essential for autophagy." Nature **395**(6700): 395-398.

Mizushima, N., H. Sugita, T. Yoshimori and Y. Ohsumi (1998). "A new protein conjugation system in human. The counterpart of the yeast Apg12p conjugation system essential for autophagy." J Biol Chem **273**(51): 33889-33892.

Moscat, J., M. Karin and M. T. Diaz-Meco (2016). "p62 in Cancer: Signaling Adaptor Beyond Autophagy." Cell **167**(3): 606-609.

Muller, S., C. Hoege, G. Pyrowolakis and S. Jentsch (2001). "SUMO, ubiquitin's mysterious cousin." Nat Rev Mol Cell Biol **2**(3): 202-210.

Nahorski, M. S., S. Maddirevula, R. Ishimura, S. Alsahli, A. F. Brady, A. Begemann, T. Mizushima, F. J. Guzman-Vega, M. Obata, Y. Ichimura, H. S. Alsaif, S. Anazi, N. Ibrahim, F. Abdulwahab, M. Hashem, D. Monies, M. Abouelhoda, B. F. Meyer, M. Alfadhel, W. Eyaid, M. Zweier, K. Steindl, A. Rauch, S. T. Arold, C. G. Woods, M. Komatsu and F. S. Alkuraya (2018). "Biallelic UFM1 and UFC1 mutations expand the essential role of ufmylation in brain development." Brain **141**(7): 1934-1945.

Nakayama, K. I. and K. Nakayama (2006). "Ubiquitin ligases: cell-cycle control and cancer." Nat Rev Cancer **6**(5): 369-381.

Nawrocki, S. T., P. Griffin, K. R. Kelly and J. S. Carew (2012). "MLN4924: a novel first-in-class inhibitor of NEDD8-activating enzyme for cancer therapy." Expert Opin Investig Drugs **21**(10): 1563-1573.

Noda, N. N., Y. Fujioka, T. Hanada, Y. Ohsumi and F. Inagaki (2013). "Structure of the Atg12-Atg5 conjugate reveals a platform for stimulating Atg8-PE conjugation." EMBO Rep **14**(2): 206-211.

Olsen, S. K., A. D. Capili, X. Lu, D. S. Tan and C. D. Lima (2010). "Active site remodelling accompanies thioester bond formation in the SUMO E1." Nature **463**(7283): 906-912.

Olsen, S. K. and C. D. Lima (2013). "Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer." Mol Cell **49**(5): 884-896.

Oweis, W., P. Padala, F. Hassouna, E. Cohen-Kfir, D. R. Gibbs, E. A. Todd, C. E. Berndsen and R. Wiener (2016). "Trans-Binding Mechanism of Ubiquitin-like Protein Activation Revealed by a UBA5-UFM1 Complex." Cell Rep **16**(12): 3113-3120.

Pabis, M., M. Termathe, K. E. Ravichandran, S. D. Kienast, R. Krutyholowa, M. Sokolowski, U. Jankowska, P. Grudnik, S. A. Leidel and S. Glatt (2020). "Molecular basis for the bifunctional Uba4-Urm1 sulfur-relay system in tRNA thiolation and ubiquitin-like conjugation." EMBO J **39**(19): e105087.

Pelzer, C., I. Kassner, K. Matentzoglu, R. K. Singh, H. P. Wollscheid, M. Scheffner, G. Schmidtke and M. Groettrup (2007). "UBE1L2, a novel E1 enzyme specific for ubiquitin." J Biol Chem **282**(32): 23010-23014.

Peng, J., D. Schwartz, J. E. Elias, C. C. Thoreen, D. Cheng, G. Marsischky, J. Roelofs, D. Finley and S. P. Gygi (2003). "A proteomics approach to understanding protein ubiquitination." Nat Biotechnol **21**(8): 921-926.

Plechanovova, A., E. G. Jaffray, M. H. Tatham, J. H. Naismith and R. T. Hay (2012). "Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis." Nature **489**(7414): 115-120.

Popovic, D., D. Vucic and I. Dikic (2014). "Ubiquitination in disease pathogenesis and treatment." Nat Med **20**(11): 1242-1253.

Quintela, I., F. Barros, M. Fernandez-Prieto, R. Martinez-Regueiro, M. Castro-Gago, A. Carracedo, C. Gomez-Lado and J. Eiris (2015). "Interstitial microdeletions including the chromosome band 4q13.2 and the UBA6 gene as possible causes of intellectual disability and behavior disorder." Am J Med Genet A **167A**(12): 3113-3120.

Raasi, S., G. Schmidtke and M. Groettrup (2001). "The ubiquitin-like protein FAT10 forms covalent conjugates and induces apoptosis." J Biol Chem **276**(38): 35334-35343.

Rajagopalan, K. V. (1997). "Biosynthesis and processing of the molybdenum cofactors." Biochem Soc Trans **25**(3): 757-761.

Rani, N., A. Aichem, G. Schmidtke, S. G. Kreft and M. Groettrup (2012). "FAT10 and NUB1L bind to the VWA domain of Rpn10 and Rpn1 to enable proteasome-mediated proteolysis." Nat Commun **3**: 749.

Robert, X. and P. Gouet (2014). "Deciphering key features in protein structures with the new ENDscript server." Nucleic Acids Res **42**(Web Server issue): W320-324.

Ross, M. J., M. S. Wosnitzer, M. D. Ross, B. Granelli, G. L. Gusella, M. Husain, L. Kaufman, M. Vasievich, V. D. D'Agati, P. D. Wilson, M. E. Klotman and P. E. Klotman (2006). "Role of ubiquitin-like protein FAT10 in epithelial apoptosis in renal disease." J Am Soc Nephrol **17**(4): 996-1004.

Rotin, D. and S. Kumar (2009). "Physiological functions of the HECT family of ubiquitin ligases." Nat Rev Mol Cell Biol **10**(6): 398-409.

Roverato, N. D., C. Sailer, N. Catone, A. Aichem, F. Stengel and M. Groettrup (2021). "Parkin is an E3 ligase for the ubiquitin-like modifier FAT10, which inhibits Parkin activation and mitophagy." Cell Rep **34**(11): 108857.

Rudolph, M. J., M. M. Wuebbens, K. V. Rajagopalan and H. Schindelin (2001). "Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation." Nat Struct Biol **8**(1): 42-46.

Ryu, K. S., Y. S. Choi, J. Ko, S. O. Kim, H. J. Kim, H. K. Cheong, Y. H. Jeon, B. S. Choi and C. Cheong (2008). "Direct characterization of E2-dependent target specificity and processivity using an artificial p27-linker-E2 ubiquitination system." BMB Rep **41**(12): 852-857.

Sarantopoulos, J., G. I. Shapiro, R. B. Cohen, J. W. Clark, J. S. Kauh, G. J. Weiss, J. M. Cleary, D. Mahalingam, M. D. Pickard, H. M. Faessel, A. J. Berger, K. Burke, G. Mulligan, B. J. Dezube and R. D. Harvey (2016). "Phase I Study of the Investigational NEDD8-Activating Enzyme Inhibitor Pevonedistat (TAK-924/MLN4924) in Patients with Advanced Solid Tumors." Clin Cancer Res **22**(4): 847-857.

Sarrio, D., S. M. Rodriguez-Pinilla, A. Dotor, F. Calero, D. Hardisson and J. Palacios (2006). "Abnormal ezrin localization is associated with clinicopathological features in invasive breast carcinomas." Breast Cancer Res Treat **98**(1): 71-79.

Scaglione, K. M., V. Basrur, N. S. Ashraf, J. R. Konen, K. S. Elenitoba-Johnson, S. V. Todi and H. L. Paulson (2013). "The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates." J Biol Chem **288**(26): 18784-18788.

Schafer, A., M. Kuhn and H. Schindelin (2014). "Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules." Acta Crystallogr D Biol Crystallogr **70**(Pt 5): 1311-1320.

Schelpe, J., D. Monte, F. Dewitte, T. K. Sixma and P. Rucktooa (2016). "Structure of UBE2Z Enzyme Provides Functional Insight into Specificity in the FAT10 Protein Conjugation Machinery." J Biol Chem **291**(2): 630-639.

Schlieker, C. D., A. G. Van der Veen, J. R. Damon, E. Spooner and H. L. Ploegh (2008). "A functional proteomics approach links the ubiquitin-related modifier Urm1 to a tRNA modification pathway." Proc Natl Acad Sci U S A **105**(47): 18255-18260.

Schulman, B. A. and J. W. Harper (2009). "Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways." Nat Rev Mol Cell Biol **10**(5): 319-331.

Schwarten, M., M. Stoldt, J. Mohrluder and D. Willbold (2010). "Solution structure of Atg8 reveals conformational polymorphism of the N-terminal domain." Biochem Biophys Res Commun **395**(3): 426-431.

Shimada, S., N. Okamoto, S. Nomura, M. Fukui, S. Shimakawa, N. Sangu, K. Shimojima, M. Osawa and T. Yamamoto (2013). "Microdeletions of 5.5 Mb (4q13.2-q13.3) and 4.1 Mb (7p15.3-p21.1) associated with a saethre-chotzen-like phenotype, severe intellectual disability, and autism." Am J Med Genet A **161A**(8): 2078-2083.

Soucy, T. A., L. R. Dick, P. G. Smith, M. A. Milhollen and J. E. Brownell (2010). "The NEDD8 Conjugation Pathway and Its Relevance in Cancer Biology and Therapy." Genes Cancer **1**(7): 708-716.

Soucy, T. A., P. G. Smith, M. A. Milhollen, A. J. Berger, J. M. Gavin, S. Adhikari, J. E. Brownell, K. E. Burke, D. P. Cardin, S. Critchley, C. A. Cullis, A. Doucette, J. J. Garnsey, J. L. Gaulin, R. E. Gershman, A. R. Lublinsky, A. McDonald, H. Mizutani, U. Narayanan, E. J. Olhava, S. Peluso, M. Rezaei, M. D. Sintchak, T. Talreja, M. P. Thomas, T. Traore, S. Vyskocil, G. S. Weatherhead, J. Yu, J. Zhang, L. R. Dick, C. F. Claiborne, M. Rolfe, J. B. Bolen and S. P. Langston (2009). "An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer." Nature **458**(7239): 732-736.

Stewart, M. D., T. Ritterhoff, R. E. Klevit and P. S. Brzovic (2016). "E2 enzymes: more than just middle men." Cell Res **26**(4): 423-440.

Stratton, M. R. (1992). "The p53 gene in human cancer." Eur J Cancer **28**(1): 293-295.

Sun, X. X., K. B. Challagundla and M. S. Dai (2012). "Positive regulation of p53 stability and activity by the deubiquitinating enzyme Otubain 1." EMBO J **31**(3): 576-592.

Suzawa, M., D. A. Miranda, K. A. Ramos, K. K. Ang, E. J. Faivre, C. G. Wilson, L. Caboni, M. R. Arkin, Y. S. Kim, R. J. Fletterick, A. Diaz, J. S. Schneekloth and H. A. Ingraham (2015). "A gene-expression screen identifies a non-toxic sumoylation inhibitor that mimics SUMO-less human LRH-1 in liver." Elife **4**.

Swaim, C. D., A. F. Scott, L. A. Canadeo and J. M. Huibregtse (2017). "Extracellular ISG15 Signals Cytokine Secretion through the LFA-1 Integrin Receptor." Mol Cell **68**(3): 581-590 e585.

Swords, R. T., H. P. Erba, D. J. DeAngelo, D. L. Bixby, J. K. Altman, M. Maris, Z. Hua, S. J. Blakemore, H. Faessel, F. Sedarati, B. J. Dezube, F. J. Giles and B. C. Medeiros (2015). "Pevonedistat (MLN4924), a First-in-Class NEDD8-activating enzyme inhibitor, in patients with acute myeloid leukaemia and myelodysplastic syndromes: a phase 1 study." Br J Haematol **169**(4): 534-543.

Takemoto, M., Y. Kawamura, M. Hirohama, Y. Yamaguchi, H. Handa, H. Saitoh, Y. Nakao, M. Kawada, K. Khalid, H. Koshino, K. Kimura, A. Ito and M. Yoshida (2014). "Inhibition of protein SUMOylation by davidiin, an ellagitannin from Davidia involucrata." J Antibiot (Tokyo) **67**(4): 335-338.

Thrower, J. S., L. Hoffman, M. Rechsteiner and C. M. Pickart (2000). "Recognition of the polyubiquitin proteolytic signal." EMBO J **19**(1): 94-102.

Ungermannova, D., S. J. Parker, C. G. Nasveschuk, D. A. Chapnick, A. J. Phillips, R. D. Kuchta and X. Liu (2012). "Identification and mechanistic studies of a novel ubiquitin E1 inhibitor." J Biomol Screen **17**(4): 421-434.

Utine, G. E., G. Haliloglu, B. Volkan-Salanci, A. Cetinkaya, P. O. Kiper, Y. Alanay, D. Aktas, B. Anlar, M. Topcu, K. Boduroglu and M. Alikasifoglu (2014). "Etiological yield of SNP microarrays in idiopathic intellectual disability." Eur J Paediatr Neurol **18**(3): 327-337.

Walden, H., M. S. Podgorski, D. T. Huang, D. W. Miller, R. J. Howard, D. L. Minor, Jr., J. M. Holton and B. A. Schulman (2003). "The structure of the APPBP1-UBA3-NEDD8-ATP complex reveals the basis for selective ubiquitin-like protein activation by an E1." Mol Cell **12**(6): 1427-1437.

Wang, F. and B. Zhao (2019). "UBA6 and Its Bispecific Pathways for Ubiquitin and FAT10." Int J Mol Sci **20**(9).

Wang, X., R. A. Herr, M. Rabelink, R. C. Hoeben, E. J. Wiertz and T. H. Hansen (2009). "Ube2j2 ubiquitinates hydroxylated amino acids on ER-associated degradation substrates." J Cell Biol **187**(5): 655-668.

Wang, Y., D. Argiles-Castillo, E. I. Kane, A. Zhou and D. E. Spratt (2020). "HECT E3 ubiquitin ligases - emerging insights into their biological roles and disease relevance." J Cell Sci **133**(7).

Wenzel, D. M., A. Lissounov, P. S. Brzovic and R. E. Klevit (2011). "UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids." Nature **474**(7349): 105-108.

Wilkinson, K. D. (2005). "The discovery of ubiquitin-dependent proteolysis." Proc Natl Acad Sci U S A **102**(43): 15280-15282.

Williams, K. M., S. Qie, J. H. Atkison, S. Salazar-Arango, J. Alan Diehl and S. K. Olsen (2019). "Structural insights into E1 recognition and the ubiquitin-conjugating activity of the E2 enzyme Cdc34." Nat Commun **10**(1): 3296.

Xirodimas, D. P., M. K. Saville, J. C. Bourdon, R. T. Hay and D. P. Lane (2004). "Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity." Cell **118**(1): 83-97.

Xu, G. W., M. Ali, T. E. Wood, D. Wong, N. Maclean, X. Wang, M. Gronda, M. Skrtic, X. Li, R. Hurren, X. Mao, M. Venkatesan, R. Beheshti Zavareh, T. Ketela, J. C. Reed, D. Rose, J. Moffat, R. A. Batey, S. Dhe-Paganon and A. D. Schimmer (2010). "The ubiquitin-activating enzyme E1 as a therapeutic target for the treatment of leukemia and multiple myeloma." Blood **115**(11): 2251-2259.

Xu, P., D. M. Duong, N. T. Seyfried, D. Cheng, Y. Xie, J. Robert, J. Rush, M. Hochstrasser, D. Finley and J. Peng (2009). "Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation." Cell **137**(1): 133-145.

Yamamoto, N., T. Shimizu, K. Yonemori, S. Kitano, S. Kondo, S. Iwasa, T. Koyama, K. Sudo, J. Sato, K. Tamura, J. Tomomatsu, M. Ono, N. Fukuda and S. Takahashi (2021). "A first-in-human, phase 1 study of the NEDD8 activating enzyme E1 inhibitor TAS4464 in patients with advanced solid tumors." Invest New Drugs.

Yan, J., J. Lei, L. Chen, H. Deng, D. Dong, T. Jin, X. Liu, R. Yuan, Y. Qiu, J. Ge, X. Peng and J. Shao (2018). "Human Leukocyte Antigen F Locus Adjacent Transcript 10 Overexpression Disturbs WISP1 Protein and mRNA Expression to Promote Hepatocellular Carcinoma Progression." Hepatology **68**(6): 2268-2284.

Yang, S. W., X. Huang, W. Lin, J. Min, D. J. Miller, A. Mayasundari, P. Rodrigues, E. C. Griffith, C. T. Gee, L. Li, W. Li, R. E. Lee, Z. Rankovic, T. Chen and P. R. Potts (2020). "Structural basis for substrate recognition and chemical inhibition of oncogenic MAGE ubiquitin ligases." Nat Commun **11**(1): 4931.

Yang, Y., J. Kitagaki, R. M. Dai, Y. C. Tsai, K. L. Lorick, R. L. Ludwig, S. A. Pierre, J. P. Jensen, I. V. Davydov, P. Oberoi, C. C. Li, J. H. Kenten, J. A. Beutler, K. H. Vousden and A. M. Weissman (2007). "Inhibitors of ubiquitin-activating enzyme (E1), a new class of potential cancer therapeutics." Cancer Res **67**(19): 9472-9481.

Yoshimura, C., H. Muraoka, H. Ochiiwa, S. Tsuji, A. Hashimoto, H. Kazuno, F. Nakagawa, Y. Komiya, S. Suzuki, T. Takenaka, M. Kumazaki, N. Fujita, T. Mizutani and S. Ohkubo (2019). "TAS4464, A Highly Potent and Selective Inhibitor of NEDD8-Activating Enzyme, Suppresses Neddylation and Shows Antitumor Activity in Diverse Cancer Models." Mol Cancer Ther **18**(7): 1205-1216.

Yu, Q., Y. Jiang and Y. Sun (2020). "Anticancer drug discovery by targeting cullin neddylation." Acta Pharm Sin B **10**(5): 746-765.

Zhang, D. W., K. T. Jeang and C. G. Lee (2006). "p53 negatively regulates the expression of FAT10, a gene upregulated in various cancers." Oncogene **25**(16): 2318-2327.