Technische Universität München
TUM School of Natural Sciences

# DETERMINING THE UBIQUITYLATION MECHANISM OF UBR5

Laura Anna Hehl

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitz: Prof. Dr. Angela Casini

Prüfende der Dissertation:

1. Hon.-Prof. Brenda A. Schulman, Ph.D.
2. Prof. Dr. Danny Nedialkova
3. Priv.-Doz. Dr. Sonja Lorenz

Die Dissertation wurde am 06.02.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 23.04.2024 angenommen.

## ACKNOWLEDGEMENTS

First off, I want to thank Brenda for giving me the opportunity of doing my PhD in this awesome lab. I am extremely thankful for all of your guidance, your feedback, and your encouragement, but mostly for always having my back throughout this intense journey. You were a great mentor and you really changed the way I think about things. Thank you for making me a better scientist and shaping me as a person! I will always be grateful for that.

Next, I want to thank my great collaborator Daniel, who joined the project at a critical point to support me during publication. This study would not have been possible without all of our discussions, your help in the lab, your encouragement, and our chocolate-trade. Thank you so much!

Rajan, you greatly supported my study with your structure-building skills. Thank you so much for being the computational backbone of the lab and always supporting all of us.

Vielen Dank, für deine Unterstützung mit den ganzen Aufreinigungen, Ronnald! Ohne dich hätte ich den straffen Zeitplan bei der Publikation nicht geschafft. Vielen Dank auch, dass du dich so toll um die ÄKTAs kümmerst und uns allen dadurch so viel ermöglichst.

I want to thank my collaborators Monique, Gerbrand, David, and Tung for all the reagents they gave me and for the valuable feedback throughout drafting of our manuscript.

Petra Beli, Martin Eilers, and Gary Kleiger, thank you for being part of my thesis advisory committee, thanks for all the valuable feedback and the great discussions.

A big thank you to Asia and Ishi for being great friends and co-workers! I'm not sure how I would have managed my PhD without your scientific and especially personal support. Thank you so much for all the fun lunch breaks and the gaming nights! Also, I want to thank you, Asia for teaching me so much about activity-based probes and setting up the system for our di-ubiquitin probes.

I want to thank Josie and Linus for so many fun evenings and for very helpful feedback along the way. You guys were a great support and really contributed to my great memories!

Celine, thank you so much for joining the lab as my master student and for being so positive about everything. I appreciate the trust you put in me and I very much enjoyed mentoring you. I'm looking forward to what you'll accomplish in the future. You'll do great!

Thank you for getting me started in this lab, Christine, for being a fun bench-mate, and my mentor. Dawa, thank you for introducing me to Brenda, and thank you for being such a kind person, who is always happy to help! I am very thankful to all the other (former) members of the department for the nice discussions, the fun time, and the awesome memories, and especially to Kirby, Khee, Jakub, Arno, Jakob, and Leo, for the resources they provided, and for all the advice! I really appreciate it.

I am very grateful for all the committed technicians we have in this lab, without whom nothing would go as smoothly. Especially, I want to thank Susanne for handling the insect

cell facility in such a great way. Thank you for supplying me with so many P3s and also for being an amazing trash-partner! Also, I want to thank Maren and Judith for always keeping the lab stocked up, to enable all of our work.

A big shout-out goes to the facilities at the MPIB, and especially the people working there. Thank you, Daniel and Tillman for managing the cryo-EM facility in such a great way. Thank you for being so kind and very helpful even outside working hours. I also want to thank the members of the mass-spectrometry facility, especially Barbara for always being super responsive, happy to trouble shoot, and always very helpful during data analysis. I want to thank Stephan Uebel and Stefan Pettera from the biophysics core facility for synthesizing peptides for me and for helping me with some biophysical characterization.

Dear Sepp, thank you so much for being the good soul of this department, and being the number one person to go to with pretty much whatever issues we have!

Ich bin sehr dankbar für die viele Unterstützung und das Verständnis meiner Familie! Ich weiß, dass ich mich immer auf Euch verlassen kann und ihr immer für mich da seid. Vielen Dank dafür.

Lastly, I want to thank my partner Jonathan for always believing in me, supporting me, and being very understanding. I am extremely lucky to have you by my side and I'm looking forward to what the future will bring!

Thank you all for making my time in this lab an unforgettable experience!

# Table of Contents

# List of publications

Parts of this thesis were published as follows:

Research article (peer-reviewed)

Hehl, L. A. *et al.* Structural snapshots along K48-linked ubiquitin chain formation by the HECT E3 UBR5. *Nat Chem Biol*, doi:10.1038/s41589-023-01414-2 (2023).

Preview article (peer-reviewed)

Hehl, L. A. & Schulman, B. A. To be (in a transcriptional complex) or not to be (promoting UBR5 ubiquitylation): That is an answer to how degradation controls gene expression. *Mol Cell* **83**, 2616-2618, doi:10.1016/j.molcel.2023.07.010 (2023).

Additional publications (co-author):

Sherpa, D., et al., GID E3 ligase supramolecular chelate assembly configures multipronged ubiquitin targeting of an oligomeric metabolic enzyme, Molecular Cell, 2021, 81(11):2445-2459.e13

Perez-Berrocal, D.A., et al., A Pro-Fluorescent Ubiquitin-Based Probe to Monitor Cysteine-Based E3 Ligase Activity, Angewandte Chemie, 2023, 62(32):e202303319

Chrustowicz, J., et al., Multisite phosphorylation dictates selective E2-E3 pairing as revealed by Ubc8/UBE2H-GID/CTLH assemblies, Molecular Cell, 2023, doi:10.1016/j.molcel.2023.11.027

## Summary

The ubiquitin code intricately regulates nearly all eukaryotic cellular processes by modifying substrates with ubiquitin, eliciting varied molecular outcomes and cellular responses. Specific ubiquitin modifications dictate the fate of the substrates. The significance of this system resonates in its widespread deficiencies across diverse diseases and its potential in therapeutic research. Modification of a substrate with K48 ubiquitin chains to induce proteasomal degradation is the most vital type of ubiquitylation. Redirecting substrate ubiquitylation through small molecules can steer disease-associated targets toward proteasomal degradation, yet understanding the precise mechanism behind the formation of K48-linked ubiquitin chains is imperative for utilizing this system for therapeutic investigations. Central to substrate ubiquitylation are HECT E3 ligases, which not only determine substrate specificity but also chain specificity, defining the substrate and its subsequent fate. Despite substantial interest, the mechanism employed by HECT E3 ligases to create linkage-specific ubiquitin chains remains largely elusive due to the lack of structural investigations featuring full-length proteins and visualizing the different steps of catalysis.

UBR5 is a huge HECT E3 ligase of increasing public interest due to the multiplicity of substrates it regulates *in vivo*, which mostly have been identified in recent years. Additionally, it has garnered attention for its ability to modify existing ubiquitin chains with K48-linked ones, producing branched chains. Moreover, its dysregulation in various cancers adds to its clinical significance.

During this study, we reconstituted UBR5's proposed functions *in vitro*. We gained mechanistic insights into the distinct transition states of full-length UBR5 using several activity-based probes followed by single particle cryo-EM. We furthermore validated our proposed mechanism using various biochemical methods.

Using this pipeline, we were able to resolve a structure of full-length UBR5 in its resting state, which revealed the assistance of domains outside the catalytic HECT domain to support activity. We could derive models for the E2~$Ub^{D}$-bound transition state and the ubiquitin bound intermediate, revealing substantial conformational changes within the HECT domain and the supporting domains. A high-resolution map of polyubiquitylating UBR5 allowed us to understand the intricacies of UBR5-mediated polyubiquitylation and we were able to shed light on a long-lasting question, asking: how HECT E3 ligases can determine linkage-specificity.

This study allowed us to propose a comprehensive model elucidating catalysis mediated by HECT E3 ligases. Our structures and models support and expand prior knowledge of HECT E3 ligases, showing what conformational changes are required, how they can achieve linkage-specificity, and what contributes to the processivity. This knowledge could aid in future endeavors to use UBR5 as approach for targeted protein degradation as well as an approach to generally target HECT E3 ligases in drug discovery.

## Zusammenfassung

Der Ubiquitin-Code reguliert auf komplexe Weise fast alle zellulären Prozesse in Eukaryoten, indem Substrate mit Ubiquitin modifiziert und so unterschiedliche molekulare zelluläre Reaktionen ausgelöst werden. Spezifische Ubiquitin-Modifikationen entscheiden über das Schicksal der Substrate. Die Bedeutung dieses Systems spiegelt sich in seiner weit verbreiteten Fehlerhaftigkeit in verschiedenen Krankheiten und seinem Potenzial für die therapeutische Forschung wider. Die Modifizierung eines Substrats mit K48-Ubiquitinketten zur Herbeiführung des proteasomalen Abbaus ist die wichtigste Form der Ubiquitinylierung. Umlenken der Ubiquitinylierung von Substraten durch kleine Moleküle kann krankheitsassoziierte Zielproteine dem proteasomalen Abbau zuführen. Um dieses System für therapeutische Untersuchungen nutzen zu können, ist es jedoch unerlässlich, den zugrundeliegenden Mechanismus der Bildung von K48-gebundenen Ubiquitinketten genau zu verstehen. Von zentraler Bedeutung für die Ubiquitinylierung von Substraten sind HECT E3 Ligasen, die nicht nur die Substratspezifität, sondern auch die Kettenspezifität bestimmen und damit sowohl das Substrat als auch dessen weiteres Schicksal festlegen. Trotz des großen Interesses ist der Mechanismus, mit dem HECT E3 Ligasen Ubiquitinketten mit spezifischen Bindungen herstellen, nach wie vor unklar, da es an Strukturuntersuchungen mit Proteinen in voller Länge fehlt, die die verschiedenen Schritte der Katalyse sichtbar machen. UBR5 ist eine große HECT E3 Ligase, die aufgrund der Vielzahl der von ihr regulierten Substrate, die größtenteils erst in den letzten Jahren identifiziert wurden, zunehmend im Interesse der Öffentlichkeit steht. Ihre Fähigkeit, bestehende Ubiquitinketten mit K48-verknüpften Ketten zu modifizieren, wodurch verzweigte Ketten entstehen, hat weitere Aufmerksamkeit erregt. Darüber hinaus trägt die Dysregulation von UBR5 bei verschiedenen Krebsarten zur klinischen Bedeutung bei.

In dieser Studie haben wir die beschriebenen Funktionen von UBR5 *in vitro* rekonstituiert. Wir gewannen mechanistische Einblicke in die verschiedenen Übergangszustände von vollständigem UBR5, indem wir unterschiedliche aktivitätsbasierte Moleküle verwendeten, und anschließend Einzelpartikel-Kryo-EM machten. Darüber hinaus haben wir unseren vorgeschlagenen Mechanismus mit verschiedenen biochemischen Methoden validiert.

Mithilfe dieser Pipeline konnten wir eine Struktur von vollständigem UBR5 in seinem Ruhezustand auflösen, die eine Zusammenarbeit von Domänen außerhalb der katalytischen HECT-Domäne zur Unterstützung der Aktivität aufzeigte. Wir konnten Modelle für den E2~$Ub^{D}$-gebundenen Übergangszustand und das Ubiquitin-gebundene Zwischenprodukt ableiten, die erhebliche Konformationsänderungen innerhalb der HECT-Domäne und der unterstützenden Domänen aufzeigen. Eine hochauflösende Elektronendichtekarte von polyubiquitinylierendem UBR5 ermöglichte es uns, die Feinheiten der UBR5-herbeigeführten Polyubiquitinylierung zu verstehen, und wir waren in der Lage, Licht auf eine seit langem bestehende Frage zu werfen: wie HECT-E3-Ligasen Verknüpfungsspezifitäten bestimmen können.

Diese Studie ermöglichte es uns, ein umfassendes Modell vorzuschlagen, das die durch HECT E3 Ligasen vermittelte Katalyse aufklärt. Unsere Strukturen und Modelle unterstützen und erweitern das bisherige Wissen über HECT E3 Ligasen, indem sie zeigen, welche Konformationsänderungen erforderlich sind, wie sie Verknüpfungsspezifität erreichen und was zur Prozessivität beiträgt. Dieses Wissen könnte bei künftigen Bemühungen helfen, UBR5 als Ansatz für den gezielten Proteinabbau zu nutzen, sowie als Ansatz, um generell HECT E3 Ligasen bei der Entdeckung von Arzneimitteln ins Visier zu nehmen.

# Introduction

## Proteome regulation

Proteins are constantly synthesized and degraded, leading to a highly dynamic protein pool.[1] This allows the cell to quickly adapt to altered environmental conditions such as nutrient availability and stress conditions as well as to varying cellular states facilitating cell differentiation, growth and many other aspects.[2,3] However, tight regulation of the underlying processes is required to achieve a balance in the protein pool, termed protein homeostasis (**Figure 1a**). Disturbance of processes regulating homeostasis leads to the characteristics of the aging process and often contributes to a variety of diseases such as cancer and neurodegenerative diseases (**Figure 1b**).[3,4] Therefore, understanding the mechanisms controlling protein abundance has been of huge common interest for many decades. One key aspect of this regulation is protein degradation, which is most commonly achieved by target proteins being modified with the small protein ubiquitin. This modification then induces -amongst other cellular outcomes- proteasomal degradation of the target protein. Despite the huge and long-lasting common interest, many details of the ubiquitylation machinery and the various ubiquitin modifications termed ubiquitin code remain elusive.

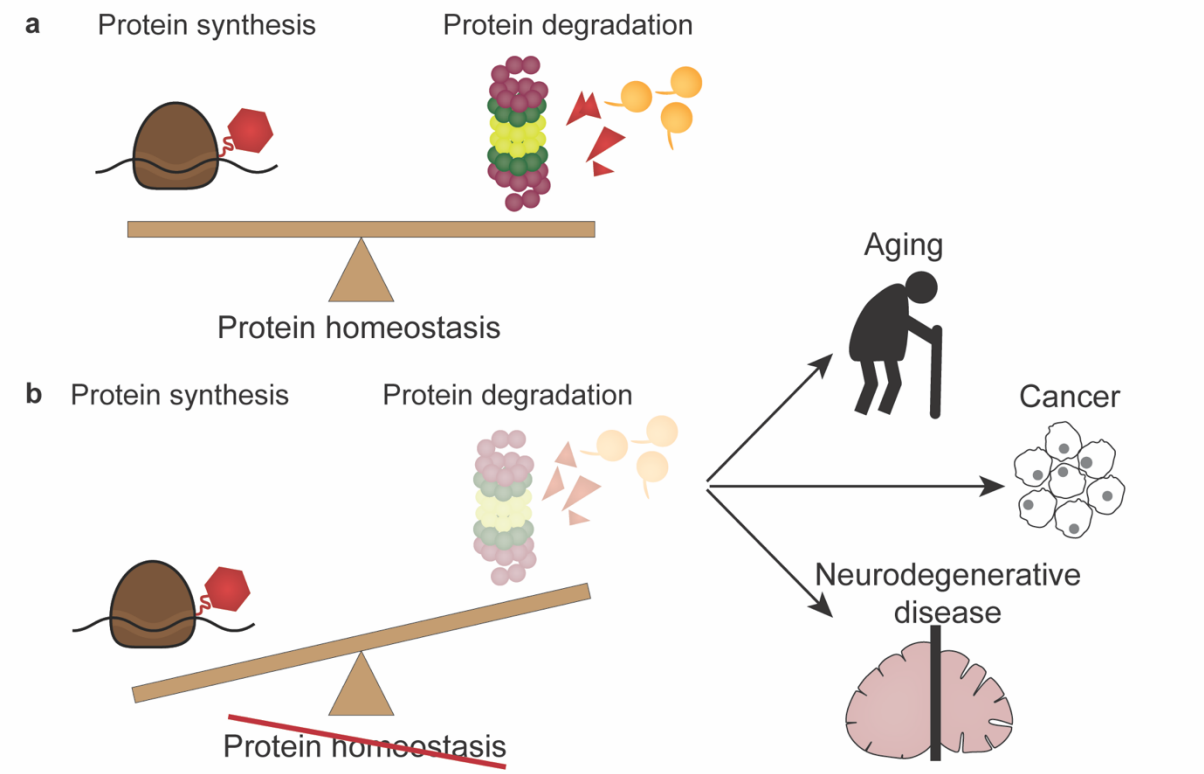


**Figure 1 Regulation of protein levels.**

**a:** Protein homeostasis achieved by balanced protein synthesis and protein degradation. **b:** Misregulated protein homeostasis and its effects.

## Ubiquitin-cascade

In eukaryotic cells, protein degradation is most often induced by modifying substrate proteins with the small protein ubiquitin in a process called ubiquitylation. Already two decades ago, a Nobel prize was awarded to Hershko, Rose, and Ciechanover for determination of the ubiquitin cascade comprising three classes of enzymes (**Figure 2**):[5] 1) Ubiquitin is linked with the carboxygroup of its C-terminal glycine (G76) to the catalytic cysteine of a ubiquitin-activating enzyme ("E1") in an ATP-dependent adenylation-reaction (**Figure 2 1**).[6-8] 2) From the catalytic cysteine of E1, ubiquitin is transferred to the catalytic cysteine of a ubiquitin-conjugating enzyme ("E2") in a transthiolation-reaction (**Figure 2 2**).[9] 3) Ubiquitin is transferred from the E2 to the substrate using one of the more than 600 ubiquitin "E3" ligases. This last step can happen in different ways:[10-12] The E3 ligase binds both the ubiquitin-loaded E2 and the substrate, and by orienting them specifically, ubiquitin is directly transferred onto the substrate. This mechanism applies to E3 ligases in the family of "**R**eally-**I**nteresting **N**ew **G**ene" (RING) (**Figure 2 3a**).[13] Another mechanism how substrates get modified, is applied by the family of "**H**omologous to **E**6AP **C**-**T**erminus" (HECT).[14,15] HECT E3 ligases contain a catalytic cysteine themselves, onto which ubiquitin is transferred from the E2 in another transthiolation reaction (**Figure 2 3b**). A substrate binds to a substrate-recognition motif and ubiquitin is then transferred from the E3 to the bound substrate in a transpeptidation reaction. The third main class of E3 ligases are the **R**ING-**B**etween-**R**ING ("RBR") E3s.[16,17] These combine properties of both other classes: They contain two RING-motifs that facilitate E2 binding and contain a catalytic cysteine, onto which ubiquitin is transferred in a transthiolation-reaction prior to passing it on to the substrate (**Figure 2 3c**).[18] Therefore, different classes of E3 ligases all work via different mechanisms to modify substrates with one or more ubiquitins (**Figure 2 4**). The different types of modification with ubiquitin can then induce various cellular responses (**Figure 2 5**).[19]

To add another layer of regulation, substrate-ubiquitylation is reversible: Ubiquitin attached to a substrate can also be removed again by another family of proteins, **De**-**UB**iquitylating enzymes "DUBs" (**Figure 2 6**). Approximately 100 different DUBs are encoded in the human genome and they fulfill various functions to terminate, modulate, or reverse the cellular response.[20] Different DUBs have different specificities regarding the ubiquitin chains they cleave, and DUBs themselves are also highly regulated on several levels, making polyubiquitylation even more complex.[21,22]

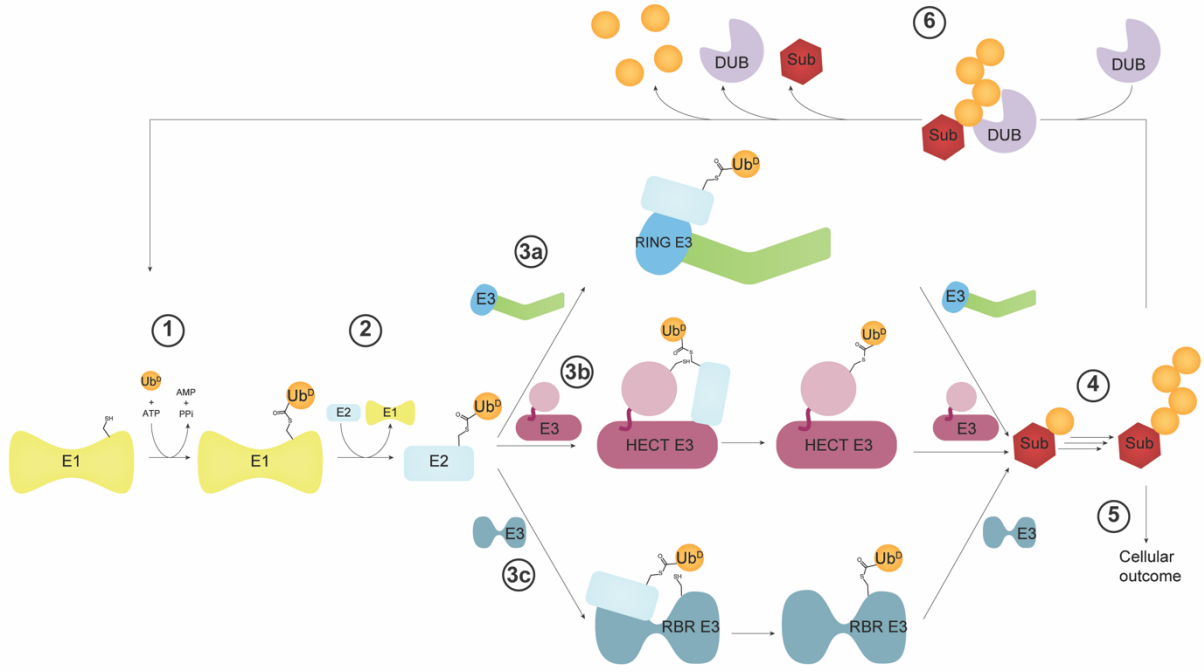


**Figure 2 Ubiquitylation cascade.**

The distinct steps of substrate-modification with ubiquitin are numbered.

## Substrate recognition

A crucial step within the cascade is the specific recognition of substrates. Distinct domains and regions of E3 ligases have evolved to recognize specific features on the substrate, so-called degrons (**Figure 3**).[23,24] Many different types of degrons exist, some of which are generally accessible, some are only conditionally accessible, and some are only formed upon distinct conditions.[25-28] If the degrons are generally accessible, this can appoint short half-lives to ensure rapid turnover.[29,30] This can be useful for substrates that have to be tightly regulated for example according to varying conditions. Such motifs can for example be found within the N-degron pathway, where specific N-terminal residues determine the half-life of proteins. Apart from recognition motifs that are always present and always accessible, there are also degrons, which are only exposed under certain conditions, or which are only generated upon specific circumstances.[25,26] Exposure of otherwise hidden surfaces can occur by misfolding of a protein.[31] A co-translational quality control causes stalling of translational machinery to facilitate correct folding, or it directly leads to degradation of misfolded proteins.[32] Proteins that underwent unfolding for example by heat-induced denaturation, also often expose specific hydrophobic residues or patches, that would normally be hidden.[33,34]

Another mechanism of substrate recognition relies on surfaces that would usually be hidden in the interface of a complex, which become available by complex-dissociation. Such a dissociation can suggest misregulation of different components of a big complex, or its malfunctioning. Such orphan quality control therefore ensures proper assembly of big molecular machines to facilitate proper function.[28]

In order to recognize all the different types of degrons, E3 ligases possess a multitude of different substrate recognition domains, which can specifically recognize single residues or longer peptides to facilitate modification of the correct substrate. In fact, many E3 ligases contain several different substrate recognition domains or interact with varying substrate recognition subunits and are therefore very versatile.[35]

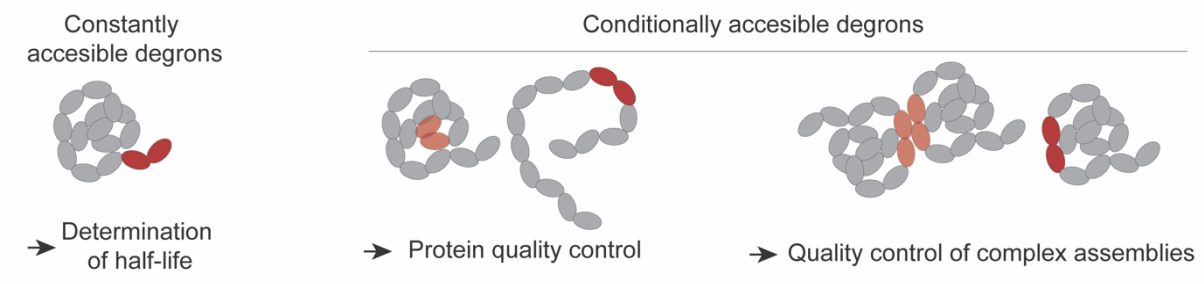


**Figure 3 Different types of substrate recognition motifs.**

Regions shaded dark red are accessible degrons, inaccessible degrons are colored light red.

## Ubiquitin modifications

### Homotypic chains

Once a specific substrate has been recognized by the E3 ligase, it gets ubiquitylated most commonly on a lysine residue.[36,37] This is achieved by passing on ubiquitin from the E3 in case of HECT or RBR E3 ligases, or from the E2 in case of RING E3 ligases.[38-42]

The most common type of modification is the sequential attachment of a ubiquitin chain to a substrate. In this case, several ubiquitins are linked to each other via isopeptide-bonds connecting the C-terminus of one ubiquitin to a primary amino group of another.[36,43] This amino group can be the side chain of the seven lysine residues present in ubiquitin: K6, K11, K27, K29, K33, K48, or K63, or the amino-terminus on methionine 1 (M1). The chains are then accordingly named by the connection of the ubiquitins with each other, such as K48 ubiquitin chains. Different chains cause different cellular responses, with most ubiquitin chains having the potential to induce various results themselves, often overlapping with other ubiquitin chains (**Figure 4**).[19,44]

M1-linked ubiquitin chains have been shown to be crucial for survival of Drosophila after bacterial infection by inducing expression of antimicrobial peptide genes, required for pathogen clearance.[45] Efficient mitochondrial clearance by mitophagy requires removal of K6-linked ubiquitin chains, making K6-linked chains mitophagy-regulators.[46] K11-linked ubiquitin chains are crucial regulators of the cell cycle. The **A**naphase **P**romoting **C**omplex/**C**yclosome “APC/C” targets different mitotic regulators for degradation by modifying them with K11-linked chains to drive cell cycle progression.[47] K27-linked ubiquitin chains have been shown to mediate antiviral immune response by modulating the activity of a kinase, which supports induction of selective autophagy.[48] Wnt-signaling is repressed by Smurf1, modulating the activity of axin by modifying it with K29-linked ubiquitin chains[49] and K33-linked chains play a vital role in the innate immune response

by modulating interferon-signaling. Exposure of cells with different viruses results in modulation of a kinase's activity by modification with K33-linked chains, inducing an antiviral response.[50] In return, interferon stimulation activates an E3 ligase, which modifies a signal with K33-linked chains to suppress ISG-expression.[51] K48-linked chains are the main cellular signal for degradation via the 26S proteasome, thereby being responsible for a majority of protein-turnover via 26S proteasomal degradation in the cell.[36,52] K63-linked ubiquitin chains play a crucial role in DNA damage response by acting as interaction platforms on the damaged DNA for a repair machinery to bind.[53,54]

A substrate can also be modified with only a single ubiquitin moiety or several single ubiquitins at different substrate lysines. Such mono-ubiquitylation can for example recruit an insulin receptor to the plasma membrane to enhance IGF signaling and mitogenic activity.[55]

## K48-linked ubiquitin chains

These exemplary functions of different types of ubiquitin chains concedes high importance to linkage-specificity during polyubiquitylation. Even though discovery of the first ubiquitin chain type -K48- dates back to three decades ago[56], how specific formation of these chains is achieved remains elusive. K48-linked chains have been shown to be the most abundant ubiquitin chains in humans and other eukaryotes[57,58], and lysine 48 is the sole essential residue in ubiquitin in yeast.[53] One of the first identified members of the ubiquitin system -Cdc34- is an E2 working with RING E3 ligases to rapidly form chains with high K48-linkage specificity[59,60], and the first HECT E3 ligase that was identified, also specifically forms K48-linked ubiquitin chains.[61] Furthermore, K48 chains were the first chain types, where the consequences of this modification were proposed[36,52] and investigated in detail over the last years.[62,63] Yet, also for these chains, it remains largely unknown how exactly they are formed specifically by different molecular machines.

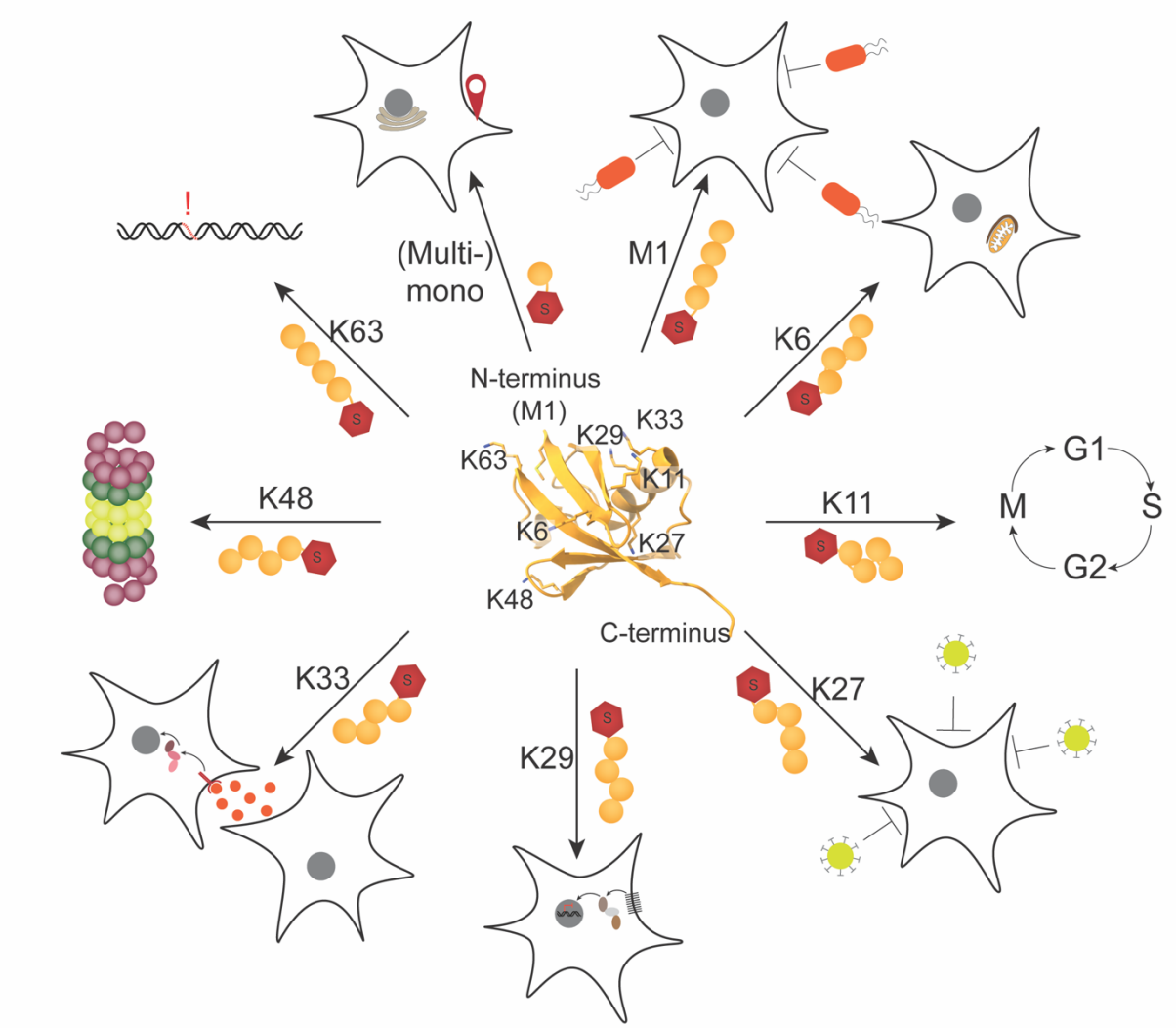


**Figure 4 Biological outcomes triggered by modification of substrate proteins with different ubiquitin chain-linkages.**

## Heterotypic chains

Additionally to ubiquitin chains that solely contain ubiquitins linked via one specific residue (homotypic chain), also heterotypic ubiquitin chains have been identified in recent years.[64] Ubiquitin moieties within these chains can either be linked via different residues, but each ubiquitin is only modified on one residue (mixed chain), or they can be linked via different residues and some ubiquitin moieties within this chain are modified on several sites by other ubiquitins (branched chains) (**Figure 5a**).[65,66] Methodological advances have paved the way to identify these heterotypic chains *in vivo*, however, the biological functions and purpose of these are still not completely understood and remain somewhat speculative.[67-69]

Generally, modifying a preassembled ubiquitin chain with another chain type could shift the cellular response to now induce another downstream-effect (**Figure 5b**).[70]

Combinations of different types of ubiquitin chains within a heterotypic chain could increase signal specificity by impairing interactions with downstream-effectors (**Figure**

**5c**).[65] Another purpose of branched chains could be the accumulation of ubiquitin that increases downstream signals by interacting with more potential effectors or by binding them more avidly (**Figure 5d**). This could be seen as priority signal for the cell under certain cellular stresses, in which the cell has to react rapidly to provide distinct modules or remove certain proteins in order to cope with a stress condition appropriately.[64,66,67,71,72]

Another potential outcome of branched chains is increased duration of the signal (**Figure 5e**). Heterotypic ubiquitin chains are more complex and might be less accessible substrates for deubiquitylating enzymes and therefore could be advantageous when a signal has to be maintained for a longer time.[73,74]

Another general advantage of heterotypic chain formation is that only few different enzymes have to be upregulated to modify a huge variety of different substrates and determine their fate.[75] The enzyme does not have to recognize the substrate specifically, but it could modify all proteins in a certain cellular compartment that are already modified with a distinct type of ubiquitin chain in order to target them all for degradation or induce another cellular outcome.

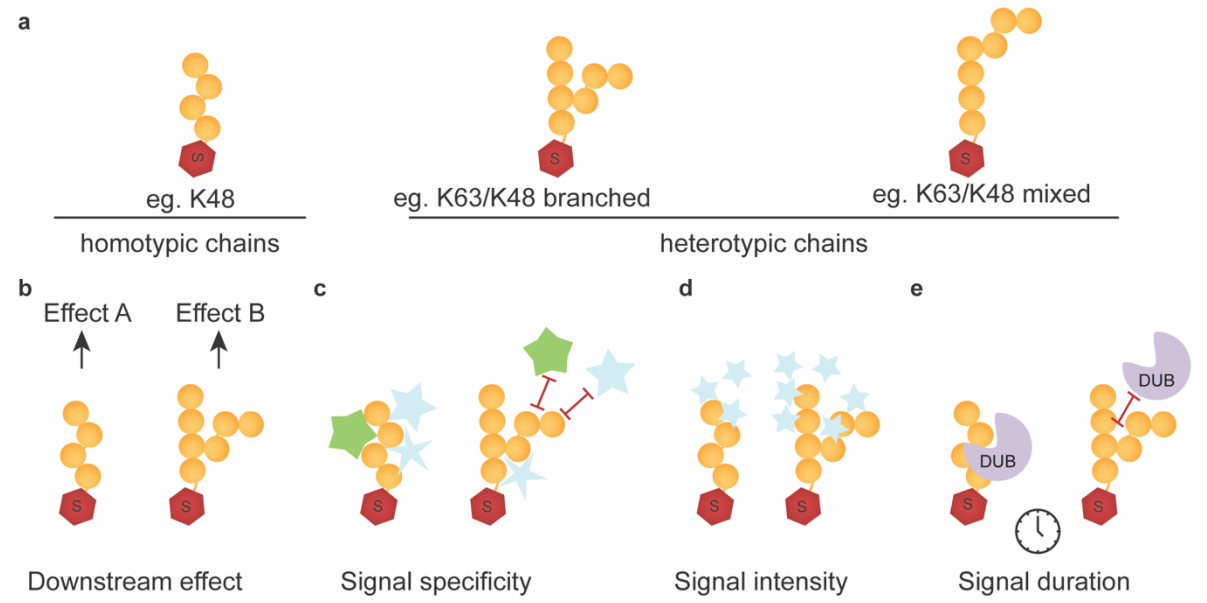


**Figure 5 Different types of ubiquitin modifications and their presumed mode-of-action.**

**a:** Different types of ubiquitin chains. Within homotypic chains, all ubiquitin-moieties are linked to one another via the same lysine residue. In branched ubiquitin chains some ubiquitin moieties are modified by other ubiquitins on two or more lysines, and in mixed ubiquitin chains each moiety is modified only on one lysine, however the moieties are linked via different lysines. **b**: Branching of the ubiquitin chain alters the fate of the substrate. **c**: The formation of a branched ubiquitin chain increases signal specificity. **d**: Branched chain formation increases signal intensity. **e**: The signal duration is increased through branched chains.

## HECT E3 ligases

Substrate-modification with ubiquitin is ultimately achieved by E3 ligases. The family of HECT E3 ligases was the first group of ligases ever discovered and members of this family share a relatively conserved catalytic domain: the HECT domain.[14,76,77] Almost 30 different

HECT E3 ligases have been found in humans, and they regulate vast biological processes.[78] Interestingly, HECT E3 ligases do not only play a huge role in eukaryotes, but other organisms also possess HECT-like proteins or use HECT E3 ligases. As such, invasive bacteria possess bacterial HECT-like proteins, which can hijack the human ubiquitin-proteasome system and redirect the hosts immune response.[79] Additionally, the first identified HECT E3 ligase E6AP can be hijacked by human papillomavirus to promote degradation of p53 by associating with the viral adapter-protein E6, thus inducing cervical carcinogenesis.[77,80,81]

## Comparison of HECT domains

HECT E3 ligases are defined by their catalytic HECT domain positioned in the far C-terminus. These consist of two subdomains: a larger, elongated N-terminal lobe ("N-lobe") and the smaller globular C-terminal lobe ("C-lobe"), which are flexibly tethered by a short linker.[82,83] Crystal structures have shown different isolated HECT domains to portray various conformations of the N- and C-lobe towards each other (**Figure 6**). The most extreme conformations are the "L-conformation" which describes the positioning of the C-lobe on the far edge of the N-lobe, and the "inverted T-conformation", in which the C-lobe sits central on the N-lobe. The first crystal structure of a HECT domain shows the isolated domain of E6AP in the L-conformation (**Figure 6a**).[82] A later crystal structure of WWP1's isolated HECT domain instead shows a conformation rather resembling the inverted T-conformation raising the question whether different HECT E3 ligases generally engage in different conformations, or whether the varying conformations facilitate distinct steps of the cascade (**Figure 6b**).[83]

Several studies aiming to determine the structural mechanisms required for activity, investigated distinct steps of HECT-catalysis. They found isolated HECT domains of different states positioned in miscellaneous conformations.[84-88] A crystal structure of NEDD4L's isolated HECT domain bound to E2~ubiquitin sheds light on the first step of HECT-activity:[84] In the inverted T-conformation, the E2 binds to a conserved E2-binding site on the N-lobe and ubiquitin is avidly bound to the E2 and the C-lobe (**Figure 6c)**.[82,84] In this structure, E2 and ubiquitin are linked to one another via an oxyester-linkage maintaining a native geometry. The structure revealed the catalytic cysteines of E2 and E3 pointing towards each other to facilitate Ub-transfer from E2 to E3. Interestingly, a crystal structure of E6AP's HECT domain solely bound to the E2, but not ubiquitin was found in the L-conformation, just like the E6AP HECT domain alone.[82] This indicated that the E2 enzyme is not sufficient to stabilize the inverted T-conformation, but both -binding of the E2 to the N-lobe, and binding of ubiquitin to the C-lobe- are required to keep the HECT domain in this conformation. This binding mode was described as conformational searching in which the E2 binds the N-lobe and conformational flexibility allows the HECT domain to then find the preferred binding position of ubiquitin.[84]

An intermediate state showing ubiquitin bound to the HECT domain has been visualized for NEDD4 as well as HUWE1. Interestingly, the structure of NEDD4's HECT domain

bound to ubiquitin depicts the HECT domain in the inverted T-conformation, similar to NEDD4L's E2~Ub-bound state (**Figure 6d**).[84,87] In contrast, HUWE1's HECT domain with ubiquitin bound is positioned in the L-conformation (**Figure 6e**).[88] These structures agree in binding of the donor ubiquitin via its I36 patch to a hydrophobic patch on the C-lobe. However, they do not agree in the overall conformation of the HECT domain, making it difficult to conclude how a HECT E3~Ub intermediate would be assembled.

Lastly, a crystal structure showing Rsp5's ubiquitin-bound HECT domain and a substrate peptide bound to a substrate-recognition domain provides a first glimpse into the transpeptidation-reactive state of a HECT E3 (**Figure 6f**).[86] In agreement with all prior crystal structures showing HECT domains bound to ubiquitin, this interaction takes place via ubiquitin's I36 patch and the C-lobe. The structure very much resembles the HUWE1$^{HECT}$~Ub intermediate suggesting that subsequently to Ub transfer from E2 to E3, the HECT domain swivels into the transpeptidation-reactive state to promote subsequent substrate-ubiquitylation.[88]

Many of these studies agree on certain HECT domain motifs, which are highly conserved, to play crucial roles during HECT-catalysis. As such, a phenylalanine in position -4 of the C-lobe is conserved amongst HECT domains. Despite apparent flexibility of the C-terminus of HECT E3 ligases, two crystal structures were able to attribute a function to this residue, which might be conserved amongst HECT E3 ligases: -4F anchors the C-terminal tail of the HECT domain to the N-lobe via hydrophobic contacts to orient the two lobes with respect to each other.[86,88]

These crystal structures provide important insights into the mechanism of HECT-mediated ubiquitylation (**Figure 6g**). Generally, HECT domains are flexible in their orientations. E2~Ub binding seems to occur in the inverted T-conformation. Subsequently to transthiolation and E2-dissociation, the C-lobe with ubiquitin bound to the catalytic cysteine can move to engage the L-conformation, in which ubiquitin can be transferred to a substrate.

Despite immense importance, none of the structures provide a rationale explaining how HECT E3s facilitate polyubiquitylation and achieve linkage-specificity. Additionally, all of these structures have been obtained in the context of only the HECT domain or at least significantly shortened HECT E3 constructs. This makes it very difficult to conclude a proper mechanism, as domains outside the HECT domain might also contribute to HECT-mediated ubiquitylation.

During our study, structures of full-length HECT E3 ligases have been published of human as well as Nematocida HUWE1, providing first insights into the overall architecture of a huge HECT E3 ligase.[89,90] HUWE1 contains an intricate backbone shaping an oval structure. It is decorated with substrate recognition domains as well as multiple ubiquitin binding domains that are invisible in the structures presumably due to high flexibility. Presence of an α-helical scaffold to orient the different domains in distinct conformations was also observed for another HECT E3 that was published very recently: human

HACE1.[91,92] HACE1 was found to dimerize and upon dimerization, to be organized within an oval as well, which results in its autoinhibition.

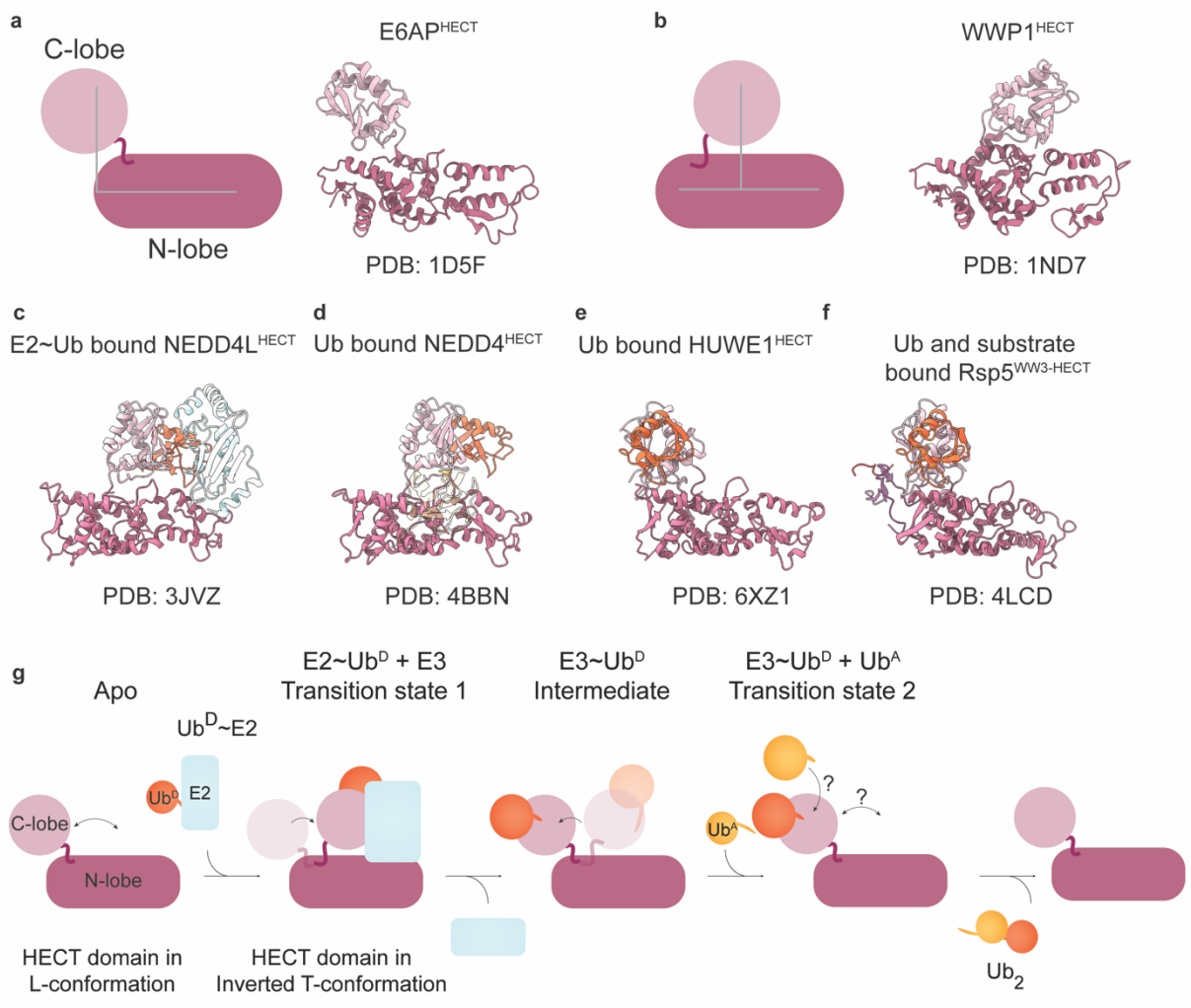


**Figure 6 Mechanistical insights based on previously published HECT domain structures of different HECT E3 ligases.**

**a**: Crystal structure of the E6AP-HECT domain in the L-conformation (PDB: 1D5F). **b**: Crystal structure of WWP1's HECT domain in the inverted T-conformation (PDB: 1ND7). **c**: Crystal structure showing NEDD4L's HECT domain bound to E2~Ub in inverted T-conformation (PDB: 3JVZ). **d**: $Ub^D$-bound state of NEDD4's HECT domain in inverted T-conformation (PDB: 4BBN). **e**: Crystal structure showing HUWE1's HECT domain bound to $Ub^D$ in L-conformation (PDB: 6XZ1). **f**: Sna3 substrate peptide bound to Rsp5's HECT domain and $Ub^D$ (crystal structure PDB: 4LCD). **g**: Proposed catalytic mechanism of HECT E3 ligases to yield polyubiquitylation.

## HECT families

HECT E3 ligases contain a multitude of different domains, besides the HECT domain accounting for different functions, such as substrate-recognition, autoinhibition, or aiding the processivity.[93] Different HECT E3 ligases are grouped into three subfamilies based on the presence of certain additional domains (**Figure 7**).[41,78,94]

The best studied family of HECT E3 ligases is the NEDD4-family consisting of nine members in humans.[95]

All members of this subfamily contain an N-terminal **C**a$^{2+}$-dependent binding motif "C2 domain" tethering the E3 to the membrane for substrate targeting.[96] Furthermore, this domain was shown to cause autoinhibition for some HECT E3s, such as SMURF2 by binding to the HECT domain, thereby impairing ubiquitin transfer from E2 to E3.[97,98] Additionally to the catalytic HECT domain and the C2 domain, NEDD4-family members also harbor varying numbers of WW domains that are responsible for substrate binding.[99-101] These WW domains are named based on two conserved tryptophanes, which bind to proline-rich regions, more specifically PPxY-motifs of the substrate. Apart from functioning as substrate-recognition domains, WW domains can also regulate the E3's catalytic activity by interacting with the N-lobe of the HECT domain or the flexible linker connecting N- and C-lobe to restrict flexibility of the two subdomains.[102] Autoinhibition mediated by WW domains can be lifted -amongst other mechanisms- by interaction of the WW domain with PY-motifs in another protein, the NEDD4-family interacting proteins.[103] C2 and WW domains already offer many regulation-options for NEDD4-family members, and yet, also unstructured linker regions connecting the WW domains seem to regulate the function of WWP2 by interacting with the HECT domain.[102] A non-covalent ubiquitin binding site within the HECT domain has furthermore been identified for several NEDD4-family members. This so-called exosite was shown to be crucial for processivity of the E3 ligase, presumably by stabilizing and orienting the distal Ub on the E3.[93]

Another subfamily of HECT E3 ligases is the HERC-family. Members of this group all contain -additionally to the C-terminal HECT domain- varying numbers of **R**CC1-**L**ike **D**omains "RLD".[104] These domains are 7-bladed propellers, with relatively low sequence-conservation.[105,106] The HERC-family can be divided into small and big HERC E3 ligases, which contain one RLD or more RLDs respectively.[107] Big HERC E3 ligases contain two or three RLDs and are amongst the biggest E3 ligases in humans. Recently, RLDs of HERC1 and HERC2 could be shown to recognize proteasomal orphan proteins or subcomplexes and lead to their degradation, thereby facilitating homeostasis of proteasome-subunits.[108,109]

The third subfamily of HECT E3 ligases summarizes all other HECT E3 ligases that could not be assigned to either of the two main subfamilies. Members of the group of "other HECT E3 ligases" do not share common domains and are very variable in size.[110] Due to the huge variance within this family, members of the "other" HECT E3 ligases are generally less well understood than members of the NEDD4- or the HERC-family. Interestingly, all HECT E3 ligases, which so far were shown to form branched ubiquitin chains -UBR5, HUWE1, TRIP12- are members of this group.[67,70,111] Domains that are portrayed in this subfamily are ubiquitin interacting domains, such as the ubiquitin-interaction motif "UIM" in HUWE1, or **Ub**iquitin-**A**ssociating domains "UBA" domains in HUWE1 or UBR5. They also harbor different substrate-recognition domains, such as the **U**biquitin **B**ox **R**ecognition domain "UBR" of UBR5 or the AZUL domain of E6AP.[112]

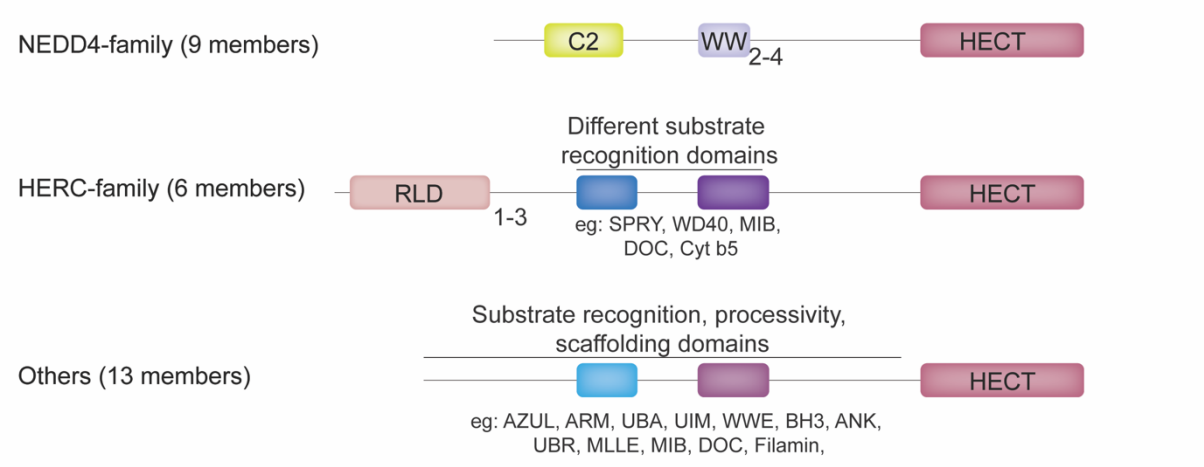


**Figure 7 Domain architecture of different families of HECT E3 ligases.**

## Linkage-specificity achieved by HECT E3s

Various different studies proposed several motifs to determine linkage-specificity within HECT E3 ligases. Testing chimeras of HECT E3s, which themselves have different linkage-specificities, showed that the specificity is encoded in the C-lobe of the respective HECT E3 (**Figure 8a**).[40,61,88]

Subsequent studies identified the far C-terminal tail of HECT E3 ligases to influence linkage-specificity.[87] All NEDD4-family members harbor an acidic residue as the ultimate residue (**Figure 8b**) and all of these E3s confer K63-specificity. Therefore, it was reasoned that the character of the ultimate residue might be sufficient to maintain this specificity.[111] This model could in parts be validated in mutational studies, where an altered C-terminal residue could partially shift linkage-specificity of HECT E3s.[61,111] However, the model also has limitations. As such, a C-terminal hydrophobic residue was proposed to result in K48-linkage formation. HUWE1, harboring a C-terminal alanine generates a mixture of K6-, K11- and K48-linked ubiquitin chains.[113,114] Controversially, HACE1 -also harboring a C-terminal alanine- was proposed to also form K27-linked ubiquitin chains.[115] Substituting the C-terminal tail of a HECT E3 with the tail from another HECT E3 conferring different linkages resulted in the formation of various chain-linkages, rather than completely switching the specificity.[87] How WWP1, a member of the NEDD4-family regulates linkage-specificity also disagrees with a model, where only the ultimate C-terminus determines linkage-specificity. WWP1 was shown to specifically form K63-linked ubiquitin chains at first, however, with increasing chain length, it also starts to assemble K11- and K48-linked chains thereby forming heterotypic ubiquitin chains.[116] This suggests a much more complex mechanism of linkage-determination, involving more aspects than a single residue but rather indicating also a role for the substrate within substrate-assisted catalysis.

A more recent study examined whether the HECT domain is sufficient to determine linkage-specificity by biochemically characterizing different truncations of the E3 ligase

E6AP.[117] This study claims that -at least in case of E6AP- additional N-terminal regions might be involved in conferring linkage-specificity.

Overall, many studies have attempted to identify the determinants of HECT E3s granting linkage-specificity. However, most studies examined HECT-mediated linkage-specificity only in the context of the HECT domain, rather than testing the entire protein. Even though several different determinants have been proposed in the many years, it is still not understood, which motifs are required for linkage-specificity due to the lack of structural studies showing the step of polyubiquitylation with some very recent exceptions. During the course of this study, a preprint showed how the yeast HECT E3 ligase Ufd4 attaches K29-linked ubiquitin chains to preassembled K48-linked chains providing a first glimpse to branched ubiquitin chain formation.[118]

Interestingly, pathogenic bacteria have evolved to harbor some HECT-like proteins, which can redirect a hosts ubiquitin system for their own needs. A recent study investigated ubiquitylation mediated by these HECT-like proteins and observed a state resembling polyubiquitylation.[119]

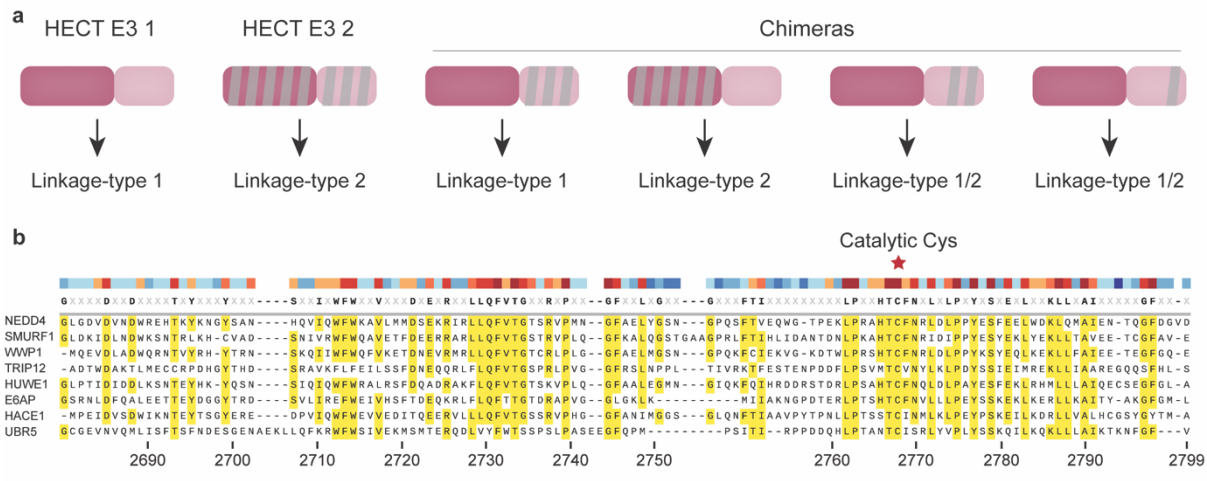


**Figure 8 Determinants of HECT-mediated linkage-specificity.**

**a:** Chimeras of different HECT E3 ligases were used to determine effects on ubiquitin linkages on the formed ubiquitin chains. **b:** Alignment of C-lobes of various HECT E3 ligases. Conserved residues are shaded yellow. The residue numbers of UBR5 are indicated and the position of the catalytic cysteine is indicated.

## Tools to understand and target HECT E3 ligases

Misregulation and malfunctioning of HECT E3s is involved in a variety of different diseases.[41,78] Due to the clinical significance, in-depth understanding of the catalytic mechanism of HECT E3 ligases would allow research in drug discovery, where malfunctioning HECT E3 ligases could be targeted or active HECT E3 ligases might be redirected to induce degradation of neo-substrates.[120-123] "**T**argeted **P**rotein **D**egradation" (TPD) is already used for disease-treatment, however, the lack of mechanistic understanding of HECT-catalysis significantly limits the research on HECT E3 ligases for TPD.[111]

Therefore, it has been a longstanding aim to find tools that could simplify research on HECT E3 ligases to provide further catalytic understanding of their mechanism.

One approach to study transthiolation-reactive members of the ubiquitylation-machinery, uses the properties of the present catalytic cysteine.[124] Detecting the fluorescent signal of a semi-chemically synthesized UbSRhodol, which changes upon reactivity to a cysteine allows to easily screen for molecules that affect activity of transthiolation-reactive enzymes.

Since members of the NEDD4-family of HECT E3 ligases are better understood than other HECT E3s, they were also particularly targeted in screens for activity-modulators. Several studies focused on targeting the ubiquitin-exosite present in NEDD4-family members. [125-128] This low-affinity ubiquitin binding region on the N-lobe enhances processivity of the respective E3 by fixing and orienting the distal ubiquitin of a chain on the HECT domain.[85,129-131] Targeting the exosite would be a tempting approach for drug design since the effect of the drug might increase with increasing chain length. One approach that was used to target HECT domains in general, but also specifically the exosites is a screen for ubiquitin variants ("UbVs").[127] HECT domains were speculated to have several low-affinity ubiquitin binding interfaces such as the ubiquitin exosite, the donor ubiquitin binding site, and potentially an acceptor ubiquitin binding site.[93,94,132] Therefore, one idea was to modulate HECT activity by finding ubiquitin mutations that would significantly increase the affinity of the UbV towards the HECT domain. This increased affinity could then cause a variety of different outcomes, depending on the binding site of the UbV on the HECT domain. A phage library screen was conducted to find UbVs that bind to isolated and immobilized HECT domains with high affinity and specificity. The identified UbVs inhibited the HECT activity by blocking the E2 binding site, activated or modulated the HECT function by binding the exosite, or were not further characterized. Initial *in vivo* experiments for a small subset of the identified UbVs targeting NEDD4L and SMURF could validate the obtained *in vitro* data and suggest the UbVs to be a useful tool. Design of UbVs to target specific surfaces could also be supported by novel deep learning tools.[133]

Overall, many approaches exist how HECT E3 ligases could be used for drug discovery, however, the main bottleneck so far is the lack of mechanistic insights into HECT-mediated substrate polyubiquitylation.

## UBR5

One HECT E3 of increasing interest is the huge E3 ligase UBR5, which was identified in Drosophila more than 45 years ago. It was shown to be involved in regulating cell proliferation in hyperplastic disc during development of Drosophila.[134] A human orthologue was then identified in human breast cancer tissue using differential display and termed EDD ("**E**3 identified by **D**ifferential **D**isplay") and was later renamed to UBR5 due to the high sequence similarity of a specific region to ubiquitin box recognition domains (**Figure 9d**).[135] UBR5 is commonly misregulated or mutated in various cancers suggesting that also human UBR5 is involved in regulating cell proliferation (**Figure 9a**).[136]

Several domains of the nuclear, 310 kDa big ligase could only be annotated very recently by us as well as several other groups in parallel, or their right boundaries were ambiguous previously (**Figure 9b**).[137-141] Overlapping findings that we published simultaneously with other groups will be discussed in the results. Domains that had been annotated previously are labelled, domains that only were verified recently are only indicated in the scheme but are not labelled.

## UBA domain

**U**biquitin **B**inding **D**omains ("UBD") are small domains that non-covalently bind to ubiquitin. UBDs are present in a huge variety of proteins ranging from proteins mediating polyubiquitylation such as E3 ligases to effectors of the ubiquitin-system like the proteasome up to sensors of the ubiquitin-state of a cell to perceive the cellular condition.[142,143]

The most abundant types of UBDs are UBA domains. These consist of three short α-helices that contact the hydrophobic I44-centered patch on ubiquitin. They are highly conserved on a structural level despite having low sequence similarity.[144] Tandem repeats of UBDs in other proteins were shown to grant linkage-specificity, chain length specificity or enable binding of complex ubiquitin chains.[145,146]

UBR5 contains one UBA domain close to the N-terminus. This UBA domain was shown to bind ubiquitin with low and different ubiquitin chains with slightly higher affinity and was proposed to be involved in polyubiquitylation.[147] A crystal structure of the UBA domain bound to ubiquitin showed canonical binding of ubiquitin via its I44 patch (**Figure 9c**).

## UBR domain

To facilitate precise discrimination between potential substrates, protein-features have evolved to be recognized by distinct E3 ligases and thereby to determine the half-life of the respective protein. As such, it was found that different N-terminal residues on the substrate confer different stabilities, termed N-degrons.[148] Several decades ago, the N-terminus of β-galactosidase in *S. cerevisiae* was found to affect the protein stability *in vivo*.[149] The UBR domain was identified to cause this effect. In the human genome, at least seven E3 ligases of different E3 families (UBR1-7) are encoded to contain a UBR domain (**Figure 9d**).[150] One of these E3 ligases is UBR5, which was shown to recognize N-terminal arginines of a substrate via this domain *in vitro*.[150,151] Despite many identified UBR5-substrates over the last years, no substrate could be shown yet to bind via the UBR domain. However, the high sequence conservancy amongst UBR domains of different E3 ligases allows to predict a mechanism for UBR5 to recognize N-terminal arginines:[152] acidic residues in one pocket of the UBR domain allow binding of the positively charged side chain of the arginine. A second pocket then encapsulates the bulky hydrophobic side chain of the second residue to facilitate stable binding.

## MLLE domain

Another potential substrate-recognition domain is the **M**ademoisel**le** (MLLE) domain, adjacent to the HECT domain. Strikingly, this domain had been identified as part of the Hyd gene of *Drosophila melanogaster* years before the HECT domain was annotated.[153] It was identified due to almost 60 % sequence identity with the **P**oly**a**denine **B**inding **C**-terminus (PABC).[154] It contains a conserved motif of the amino acids MLLE, responsible for the nomenclature.[155] This pattern binds to a distinct amino acid sequence, the **PA**BP-interacting **M**otif (PAM).[156] A model for MLLE-mediated substrate recognition of UBR5 has been proposed based on a crystal structure of the MLLE domain with a substrate peptide (**Figure 9e**):[157] A glycine residue on the MLLE domain creates a small cavity, which can be engaged by a conserved phenylalanine residue of the substrate peptide.[157,158] Hydrophobic pockets on the MLLE domain created by its four-helix bundle architecture, can be accommodated by further residues on the substrate. A PAM2-motif is found in the target protein PAIP2, and additionally, a PAM2-like motif in UBR5's N-lobe itself, suggests that the MLLE domain could have both, a role in substrate-recognition and catalytic regulation.[157]

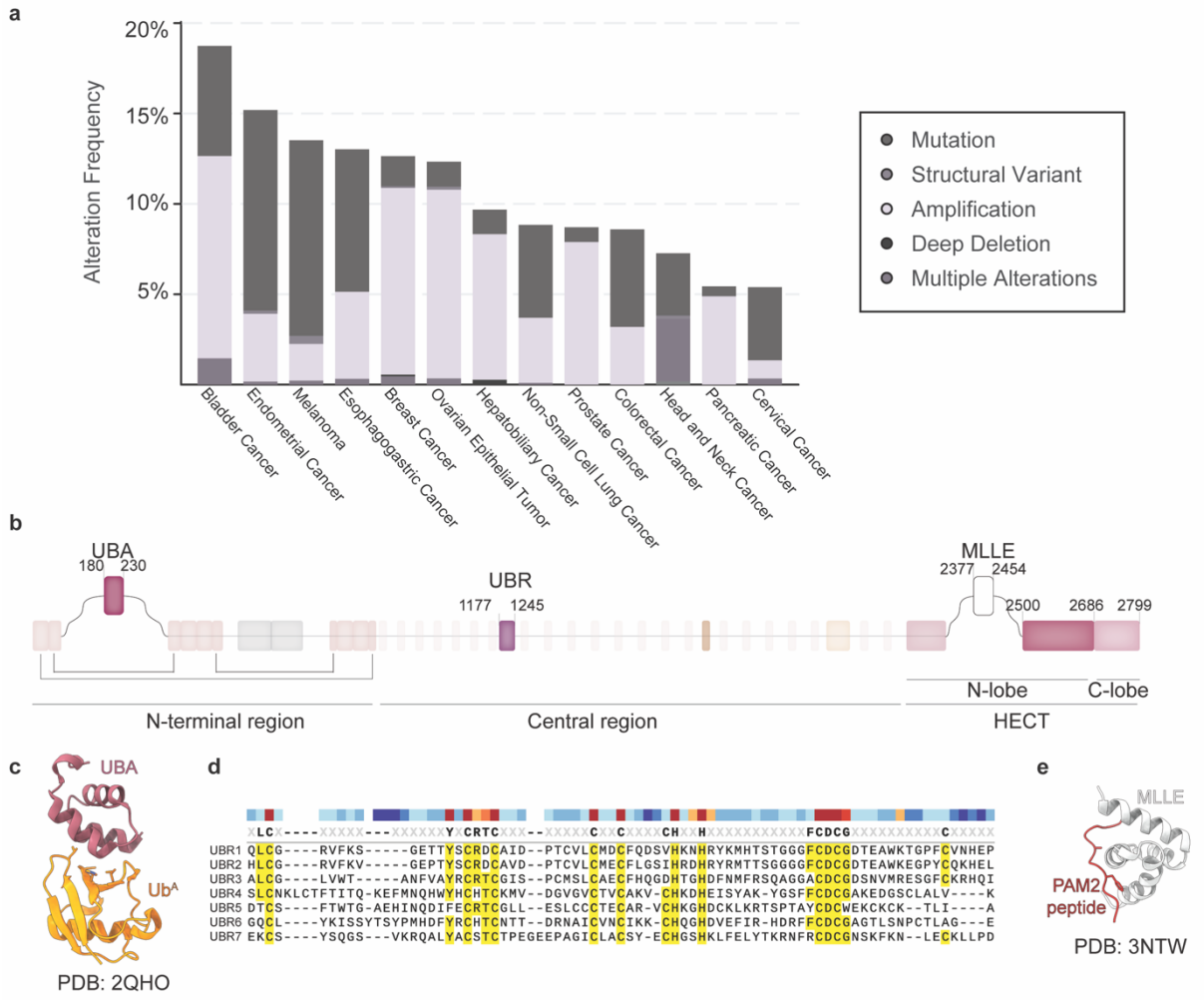


**Figure 9 Biological relevance of UBR5 and previously known insights into UBR5 domains.**

**a**: UBR5-alterations in different types of cancer based on cBioPortal. **b**: Initial domain annotation of UBR5. **c**: Crystal structure of UBR5's UBA domain bound to ubiquitin (PDB: 2QHO). **d**: Conservancy of UBR domains. **e**: Crystal structure of UBR5's MLLE domain bound to PAM2 substrate peptide (PDB: 3NTW).

## UBR5 substrates

In recent years, many proteins were found to be regulated by UBR5 and therefore be its substrates. As such, the E3 ligase RNF168, which gets recruited to DNA lesions to generate a platform for chromatin repair enzymes by K63-linked ubiquitin chain formation, was shown to be regulated by UBR5 and TRIP12 to prevent spreading of the chromatin repair machinery but rather localize it to the lesion-side for efficient repair.[159] UBR5 was shown to downregulate the level of the cell cycle inhibitor and tumor suppressor protein retinoblastoma in the early G1-phase to promote the G1/S-phase transition.[160] It was also proposed to play a role in regulating gluconeogenesis by ubiquitylating the rate-limiting gluconeogenic enzyme PEPCK1 upon glucose-induced PEPCK1-acetylation.[161] AKIRIN2, a mediator for proteasome-import into the nucleus, was identified to also be modified by UBR5 and targeted for degradation[138], just like the Drosophila homologue Akirin is a substrate to Hyd.[162] C-MYC was shown to often be co-amplified with UBR5 in various cancers and to be targeted for degradation by UBR5.[140,163,164] Lastly, a recent report identified chromatin-bound **N**uclear hormone **R**eceptors (NR) to be also regulated by UBR5, linking UBR5 malfunctioning to acute promyelocytic leukemia.[165]

Despite these and many more reports of potential UBR5-substrates,[166-172] only very recent studies were able to propose a mechanism how UBR5 recognizes two groups of substrates using distinct mechanisms.[173]

A study showed how different transcriptional regulators, including c-MYC are recognized by UBR5 (**Figure 10a**).[140] Active transcriptional regulators are usually incorporated in protein-complexes. However, in order to react to altered cellular conditions, some subunits within the complex get turned over rather rapidly in order to allow the cell to regulate the abundance of these complexes on an mRNA-level, and to also facilitate reorganization into different complexes. UBR5 can recognize and modify the resulting orphan substrates, which are lacking their binding partner via motifs that would be hidden in the interface of active transcription-regulating complexes (**Figure 3**, right). Therefore, UBR5 grants orphan protein quality control.

Another class of UBR5 substrates, where a recognition-mechanism was proposed, are NRs (**Figure 10b**).[139] These are transcription factors that react to hormonal signals by modulating protein expression in the cell. As such, the retinoic acid receptor α is crucial for white blood cell maturation, and controls cellular growth dependent on vitamin A availability. NR-misregulation is one of the main drivers of some cancer types: The estrogen receptor is often upregulated in breast cancer and is an important indicator to the healing-prognosis. NRs generally require a two-step activation: In an unbound-state, a flexible helix covers a rather conserved hydrophobic pocket of the NR. Binding of an agonist then leads to opening of this lid. In this state, the NR remains inactive and

transcription is not yet initiated, however, the lid opening allows for the second activation step: binding of a **N**uclear **Coa**ctivator (NCOA). Bound to the NCOA, the respective NR can now initiate transcription. Binding of UBR5 to the NR and its modification was also shown to be dependent on the lid-opening by agonist-binding. However, UBR5 competes with the NCOA by binding in the same hydrophobic pocket of the NR as the NCOA would bind.

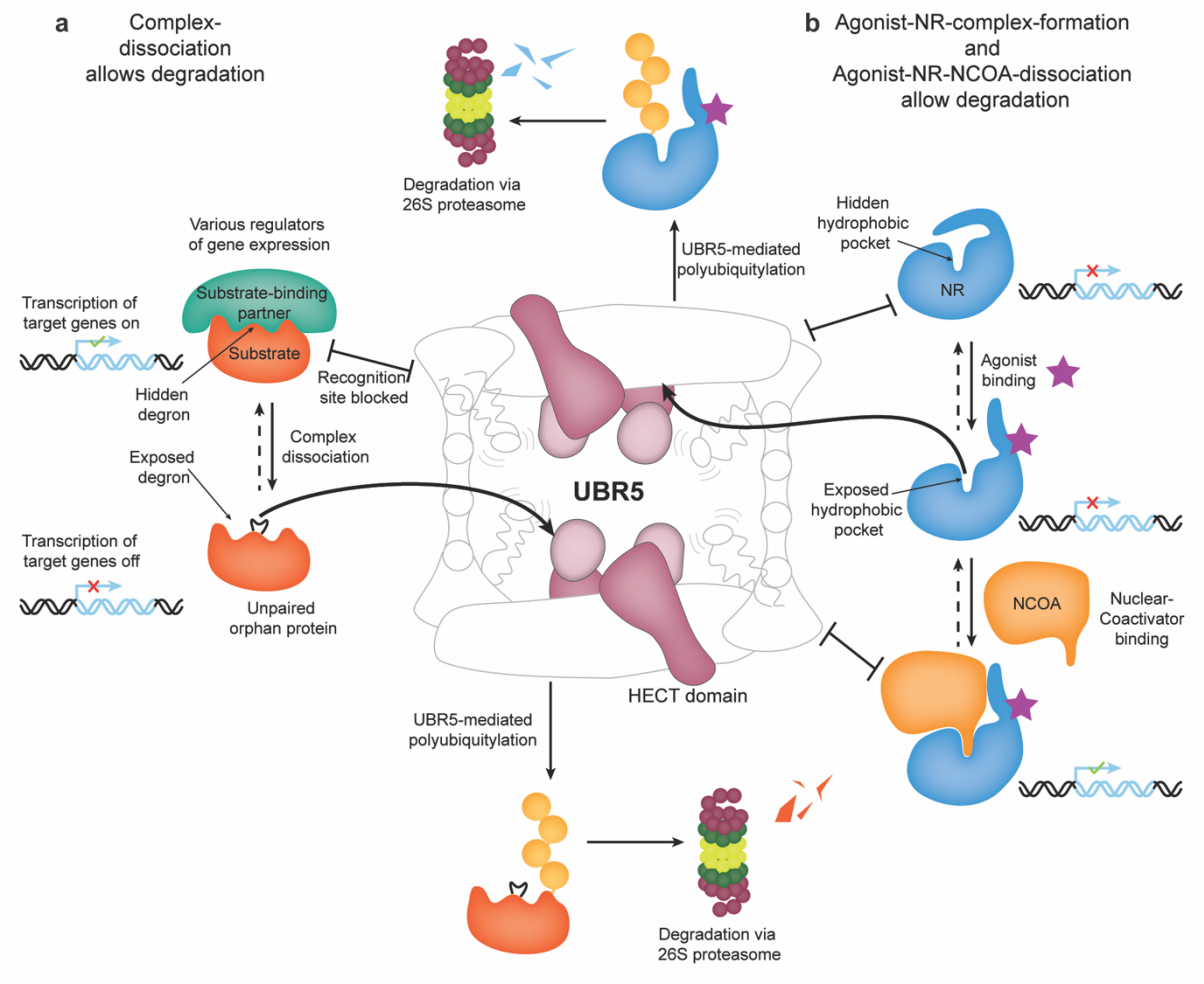


**Figure 10 Mechanism of UBR5 mediated substrate recognition.**

**a**: UBR5-mediated recognition of transcriptional regulators within orphan quality control. **b**: UBR5-mediated recognition of nuclear hormone receptors.

## Branched chain formation by UBR5

UBR5 was shown to specifically form K48-linked ubiquitin chains, but it could also forge such linkages on K11- or K63-linked chains preassembled by other E3 ligases.[67,70,174]

Modification of substrates decorated with K11-preassembled ubiquitin chains with K48-linked chains is massively upregulated under proteotoxic stress when the proteasome or heat-shock proteins are impaired.[67] Using K11/K48-bispecific antibodies and subsequent mass spectrometry analysis allowed identification of UBR5 to interact with such branched

ubiquitin chains. UBR5 could be identified to branch a K11-linked ubiquitin chain in order to generate a priority signal and push proteasomal degradation (**Figure 11a, Figure 5d**).

In contrast, branched chain formation by UBR5 adding K48-linked ubiquitin chains to preassembled K63-linked chains seems to be rather targeted and so far could only be shown for one specific substrate (**Figure 11b**).[70] The proapoptotic protein TXNIP gets ubiquitylated with K63-linked ubiquitin chains by the HECT E3 ligases WWP1 or ITCH. This modification does not induce proteolytic digest of TXNIP, however, it results in recruitment of UBR5 as well as another HECT E3 ligase -HUWE1- to the K63 ubiquitin chains and further modifies them with K48-linked chains. This novel modification can now switch the fate of the protein by inducing proteasomal degradation (**Figure 5b**).

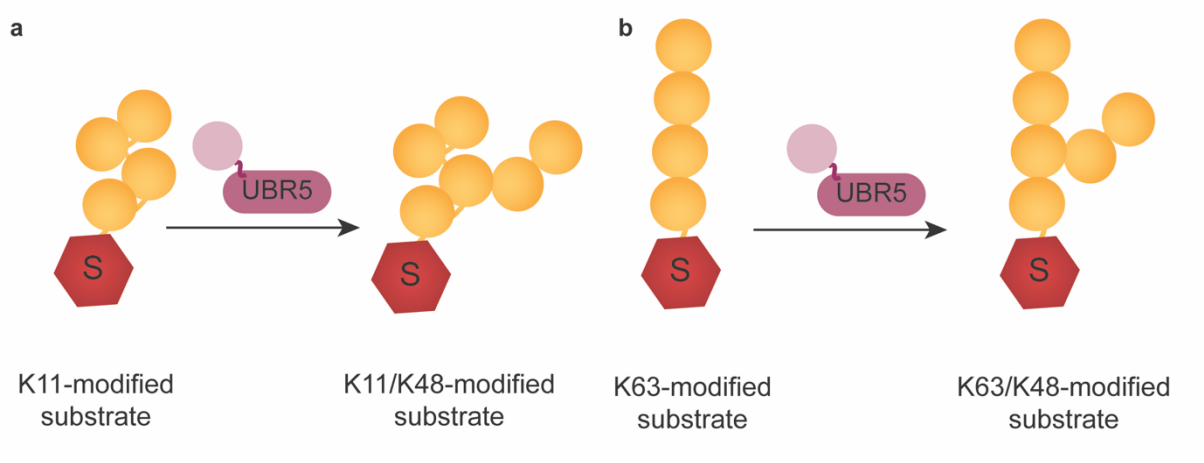


**Figure 11 UBR5-mediated branched chain formation.**

**a**: UBR5 branches K11-linked ubiquitin chains with K48-linked chains upon proteotoxic stress. **b**: UBR5 branches K63-linked ubiquitin chains on TXNIP with K48-linked chains to induce proteasomal degradation.

## Aim of the study

Since their discovery several decades ago, numerous studies have worked towards a mechanistical understanding of HECT E3 ligases.[40,61,82-88,97,98,175] Using different chemical tools, various studies have gained insight into distinct states of the catalysis of different HECT E3 ligases.[84,86-88]

These and other studies have revealed the HECT domains to be conformationally flexible due to the short flexible tether connecting the two HECT domain lobes, to bind ubiquitin via conserved mechanisms, and to be highly regulated through intra- and intermolecular interactions. However, a general mechanism explaining how HECT E3 ligases achieve their catalytic activity, what features contribute to this, and especially, how they facilitate linkage-specific ubiquitin chain formation, remains elusive. So far, several factors limit the mechanistical insight into HECT-mediated polyubiquitylation. 1) Prior studies investigating HECT-catalysis on a structural level used significantly truncated versions of HECT E3 ligases, most often even just the isolated HECT domain. This makes it very difficult to derive a mechanism, as regions that were omitted in these constructs might play crucial

roles during the native mechanism. 2) The different steps of the cascade were visualized for different HECT E3 ligases. This does not allow to unambiguously conclude what features might be specific for a distinct HECT E3 ligase, and what features are specific for a distinct step of the cascade. 3) No structural information is available showing a HECT E3 ligase during polyubiquitylation.

The aim of this study is to tackle these issues using a combination of biochemistry, chemistry, and structural biology. We reconstitute the HECT E3 ligase UBR5 *in vitro*, design and employ different chemically reactive probes to stabilize distinct transition states of UBR5, and we analyze the states using structural biology. Lastly, we validate the derived conclusions biochemically and we test them not only in the context of free ubiquitin chain formation, but we also try to apply the mechanism to substrate-ubiquitylation and to activity-modulation.

With this, we hope to gain comprehensive mechanistical insights into all the steps required for substrate-polyubiquitylation by the HECT E3 ligase UBR5 to be able to derive a common mechanism for HECT E3 ligases.

# Results

## 1. Reconstitution, biochemical, and structural characterization of UBR5

### UBR5 reconstitution

To recombinantly express and purify human UBR5, we employed the BacMam-system (**Figure 12a**).[176,177] In brief, a transfer plasmid carrying a cassette for mammalian expression, as well as the gene of interest fused to an affinity tag, is used to generate recombinant bacmids. These bacmids are utilized to infect Sf9-cells to yield recombinant baculovirus. Baculovirus can infect a variety of different mammalian cell lines. Due to the easy maintenance and high scalability, HEK293 suspension cells were employed for UBR5-expression. This was followed by lysis, streptavidin pull down performed in a gravity-flow set-up, and anion exchange chromatography. Since anion exchange chromatography did not significantly improve sample homogeneity as observed by SDS-PAGE analysis, but instead increased the level of aggregate-formation, this step was only performed during initial purifications and was omitted later on. The last step consisted of size exclusion chromatography and yielded relatively pure protein (**Figure 12b**).

Including a GFP-tag N-terminally of UBR5 made it easy to screen for constructs that resulted in high UBR5-expression and to also test different expression conditions. However, we aimed to test UBR5's activity with fluorescent ubiquitin, which is partially incompatible with UBR5's GFP-tag. For this reason, we tested different constructs for their ability to be cleaved by an HRV 3C protease (**Figure 12c**). The different constructs contained no linker between the 3C-cleavage site and UBR5, a linker of three amino acids, or a linker of seven amino acids that would remain on UBR5 after cleavage. We affinity-captured UBR5, took samples prior to protease treatment and then added protease to the resin. After incubation, we took samples of the resin and subsequently washed the resin to elute cleaved protein. After cleavage, UBR5 would not be fluorescent, or portray affinity to the resin anymore and therefore can be washed off. The large size of UBR5 makes it difficult to observe shifts in SDS-PAGE upon cleavage, but by visualizing the fluorescence and testing whether streptavidin-affinity remained, the cleavage-efficiency can be evaluated (**Figure 12d**). Construct #1 and #2 (no linker or 3 residue long linker) did not show significant cleavage as the fluorescent signal remains after protease treatment and only a small fraction of UBR5 was washed off the resin. However, for construct #3, the fluorescent signal on the height of UBR5 completely disappears after treatment and UBR5 could get washed off the resin, indicating that the proteolytic digest works very well. Therefore, all subsequent UBR5-constructs were designed to harbor a seven amino acid long linker between the 3C-cleavage site and UBR5.

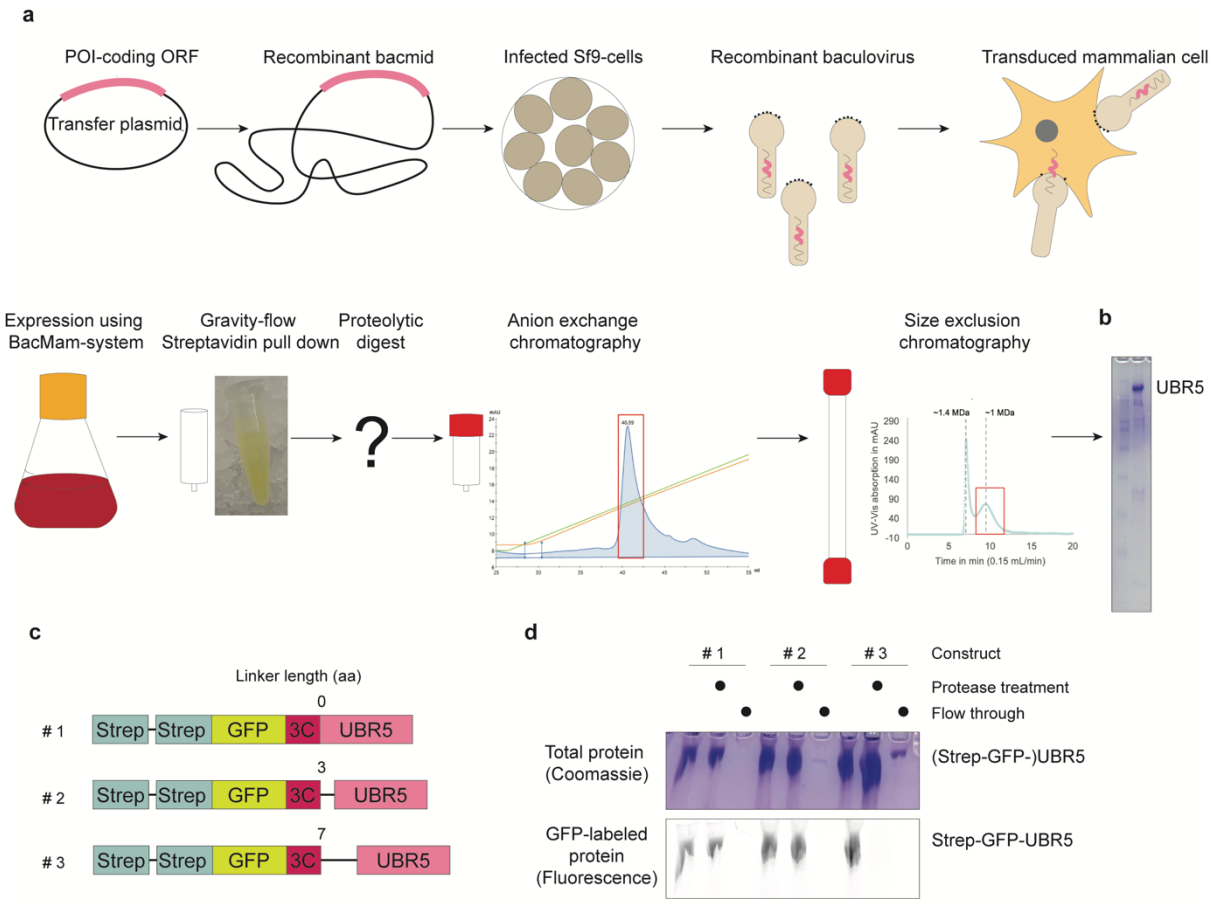


**Figure 12 Protocol-establishment for recombinant expression and purification of full-length UBR5.**

**a**: Expression and purification scheme established for UBR5. Using the BacMam-system (upper row), UBR5 was expressed in HEK293S-cells and was subsequently purified using affinity-purification and chromatography. **b**: SDS-PAGE showing resulting recombinantly expressed and purified full-length UBR5. **c**: Design of different expression-constructs that were tested. **d**: Proteolytic-digest of the different purification-constructs.

## Biochemical characterization of UBR5

Next, we sought to characterize UBR5 *in vitro*. Size exclusion chromatography that we had performed as a purification step, indicated that UBR5 runs at a much higher molecular weight as anticipated. The expected molecular weight of UBR5 would be 310 kDa, however, it eluted much earlier at a volume corresponding to >1 MDa, suggesting oligomerization of UBR5. To test whether this was reproducible and whether we could influence oligomerization with different conditions, we performed analytical HPLC runs with different buffer-compositions or different storage conditions. This would not only reveal features of the oligomerization, but also show how or whether we could store UBR5 for subsequent biochemical assays or cryo-EM.

First, we tested whether UBR5 oligomerizes in a salt-dependent manner (**Figure 13a**). To do so, we subjected purified UBR5 to the HPLC with different salt-concentrations. The overall pattern was similar regardless of the salt-concentration. A first very sharp peak

eluted at the void volume of the column, indicating aggregates, and a second peak eluted at a molecular weight corresponding to >1 MDa. However, higher salt seemed to increase the intensity of the aggregate-peak, which was the lowest in 100 mM salt.

We did not see a smaller peak appearing in high salt, which could have indicated disruption of an oligomerization-interface in a salt-dependent manner. However, disrupting charged interfaces using high salt concentrations could also instead increase aggregation if the intra- or intermolecular interface contributed to protein stability.

Next, we wanted to see whether the oligomerization depends on the concentration of reducing agent (**Figure 13b**). For this, we performed HPLC runs either in the presence of 1 mM DTT or 5 mM DTT. Strikingly, the aggregate-peak was significantly higher with high DTT compared to 1 mM DTT. This suggested that the protein might be stabilized by disulfide-bonds, but again we cannot judge whether this effect is caused by impaired oligomerization or affected tertiary structures.

Overall, we could conclude that targeting UBR5 oligomerization rather leads to increased protein aggregation than to the formation of stable smaller oligomers or monomers.

An ultimate set of HPLC runs was performed aiming to determine storage conditions for UBR5 (**Figure 13c**). We analyzed fresh UBR5 that was applied to the HPLC directly after purification, UBR5 that had been stored at 4°C, or UBR5 that had been frozen and thawed once. Again, the elution pattern resembled the previous runs, but it could be observed that the aggregate-peak is significantly smaller if fresh UBR5 was analyzed compared to UBR5 that had been stored previously.

As aggregation of UBR5 could pose difficulties for further analysis, we sought to gain more insight into UBR5's stability. For this reason, we performed **nano D**ifferential **S**canning **F**luorimetry (nanoDSF), which measures tryptophane-fluorescence at different conditions (**Figure 13d**).[178] Because of its hydrophobic nature, tryptophane is most often buried within the core of a protein and is not solvent-exposed. Denaturing the protein of interest leads to unfolding and solvent-exposure of the otherwise buried tryptophanes. This change of tryptophane-exposure can be measured in an altered fluorescence. If denaturing is achieved by applying heat to the protein, its stability can be measured by monitoring the tryptophane-fluorescence at increasing temperatures. Stable proteins will denature at higher temperatures whereas unstable proteins will reveal their tryptophanes already at lower temperatures. To test UBR5's stability, we performed nanoDSF at two different concentrations. Strikingly, the two different concentrations resulted in two different melting points that varied by ~5°C. The lower concentration of UBR5 already denatured around 54°C whereas the higher concentration only denatured around 59°C. Even though the curve for the low concentration is very flat, which might be caused by an overall low number of tryptophanes, or some of the tryptophanes could already be solvent-exposed at the start, the turning point is still quite distinct. A concentration-dependent stabilization of UBR5 could suggest the formation of different oligomers in a concentration-

dependent manner, which could have stabilizing effects as also discussed previously (**Figure 13a-c**).

This first analysis revealed that UBR5 seems to form a stable oligomer and that this oligomerization presumably contributes to UBR5's stability.

To examine catalysis mediated by UBR5, we first aimed to see which E2 UBR5 preferentially works with (**Figure 13e**). To do so, different E2 enzymes that are known to work with HECT E3 ligases, as well as UBE2S that does not collaborate with HECT E3 ligases are loaded with fluorescently labelled ubiquitin (*Ub) to yield the thioester-bound E2~*Ub in a pulse-reaction. This reaction is stopped by addition of EDTA. Next, E2~*Ub and UBR5 are mixed in the chase-reaction. Samples were retrieved at the indicated time points and analyzed using SDS-PAGE. Non-reducing samples show all intermediates of ubiquitin being covalently bound, whereas reducing samples do not show thioester-bound intermediates. Fluorescent imaging of the SDS-PAGE reveals whether an E2 can discharge its bound ubiquitin to UBR5 (only visible in non-reducing SDS-PAGE) and whether UBR5 can use this obtained ubiquitin to form free ubiquitin chains linked via isopeptide-bonds (visible in non-reducing and reducing SDS-PAGE). All UBE2D-members are able to discharge ubiquitin to UBR5 and UBR5 can then form ubiquitin chains. It is noteworthy that chain formation only initiates after addition of UBR5 and there are no ubiquitin chains present at the 0 sec time point. This shows that chain-formation is only achieved by the E3 and the chains are not formed by the E2 enzyme already. In case of UBE2L3, the chain formation is significantly weaker than for the UBE2D-family, which is in agreement with other publications.[138,179] In case of UBE2A, UBE2B, and UBE2S, no UBR5-mediated discharge can be observed and also no chain formation. However, these enzymes seem to autoubiquitylate to some extent, as it can be seen by the fluorescent signal on the height of E2~*Ub in the reducing SDS-PAGE. Since this is already present prior to UBR5-addition (time point 0 sec), this effect is not mediated by UBR5 but already occurs during the pulse-reaction. Because chain formation with UBE2D2 seemed to be the strongest, we continued using this E2 in subsequent assays.

Next, we aimed to determine whether the observed activity was dependent on UBR5's catalytic cysteine or whether UBR5 would rather serve as a scaffold to "activate" the respective E2 (**Figure 13f**). Again, a pulse-chase assay was employed. In this case, the fluorescently labelled ubiquitin had all its lysines mutated to arginine, and can't be used as acceptor, but only as **D**onor **Ub**iquitin ("*$Ub^{D}$"). This was done to focus on a single transpeptidation event rather than multiple as it would be the case with **W**ild**t**ype ("WT") ubiquitin. The pulse-reaction was performed as previously. For the chase-reaction, WT UBR5 or UBR5 with the catalytic cysteine C2768 mutated to alanine were supplemented with WT unlabeled **Ub**iquitin that could only serve as **A**cceptor ("$Ub^{A}$") due to the prior inactivation of the E1 loading reaction. Again, fluorescent scanning of the non-reducing SDS-PAGE reveal the passage of *$Ub^{D}$. If WT UBR5 was present, *$Ub^{D}$ could be readily transferred from the E2 to the E3 and subsequently to unlabeled $Ub^{A}$. In case of $UBR5^{C2768A}$, no discharge of E2~*$Ub^{D}$ could be observed similar to a chase-reaction

without any UBR5, showing that the ubiquitin chain formation seen in the presence of UBR5 indeed is mediated by UBR5's catalytic cysteine.

Previous publications showed UBR5 mediates K48 chain formation.[67] To test whether this feature is maintained in our *in vitro* reconstituted system, we used an unbiased and quantitative mass-spectrometry-based approach to identify what linkages were forged and to what extent (**Figure 13g**).[180] We first performed a polyubiquitylation assay with WT ubiquitin as well as WT UBR5. The obtained ubiquitin chains were then proteolytically digested to yield short peptides. This digest leaves a distinct remnant of two glycines (diGly) on the lysine that had been modified, and lysines that were not modified would not exhibit such a diGly-remnant. Using an antibody specifically binding to the diGly-motif, the peptides containing these modifications can be enriched. Specific amounts of heavy reference peptides for every potential ubiquitin-linkage are then added to the sample to allow **A**bsolute **Qua**ntification (AQUA) before subjecting it to **L**iquid **C**hromatography coupled to **M**ass **S**pectrometry (LC-MS).[180] The m/z ratio allows to unambiguously identify the linkage type a distinct peak represents and comparing the peak-intensity of the reference peptide to the sample peptide allows for quantification of the abundance. In agreement with *in vivo* data, UBR5 retains its K48-linkage-specificity also in our reconstituted system as it forms >> 99% K48-linked ubiquitin chains (**Figure 13h**).

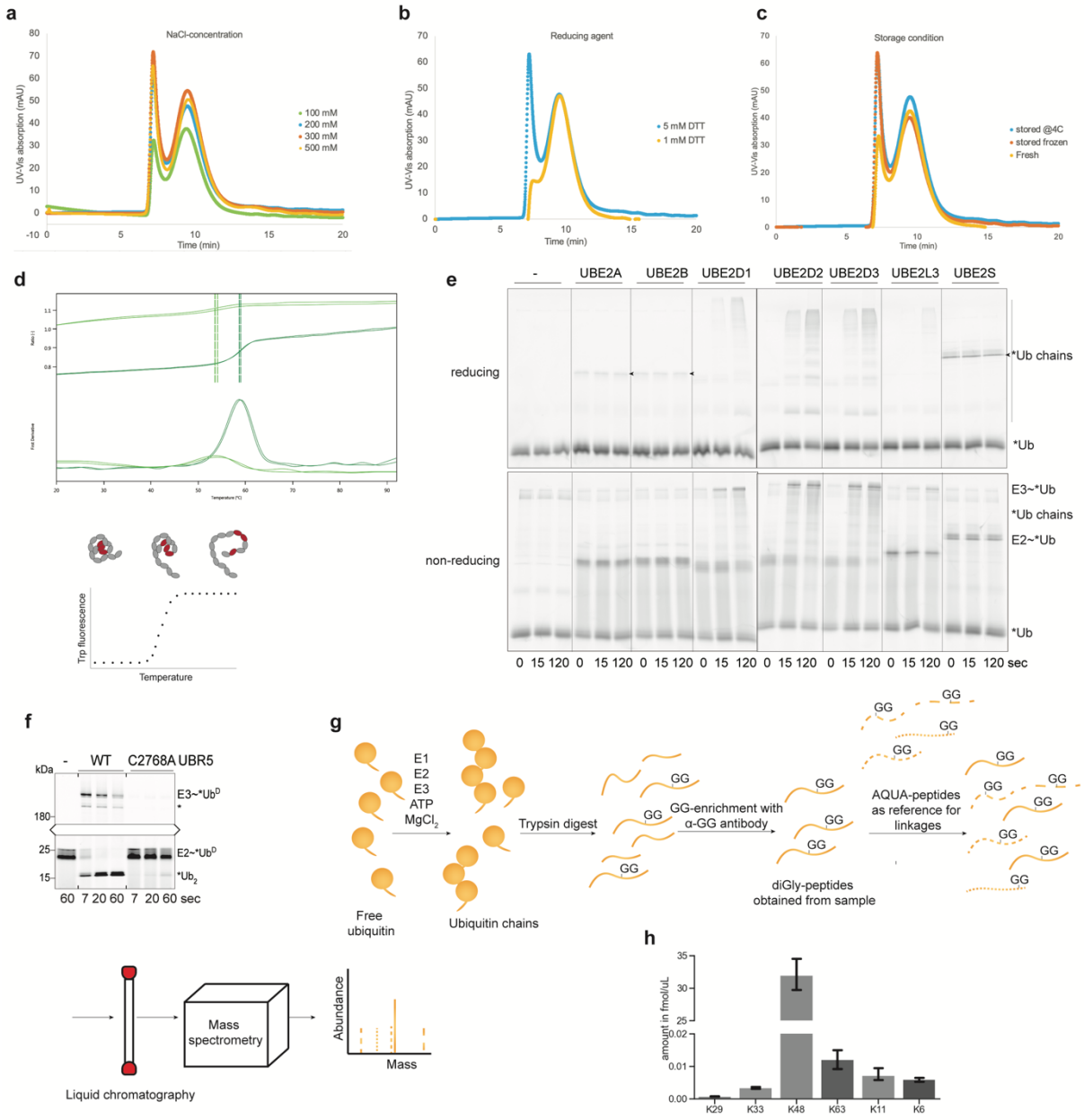


**Figure 13 Biophysical and biochemical characterization of UBR5**.

**a**: Analytical HPLC of UBR5 with various NaCl-concentrations to assess oligomerization. **b**: Analytical HPLC of UBR5 with various reducing reagent-concentrations to assess oligomerization. **c**: Analytical HPLC of UBR5 with various storage-conditions to assess protein stability. **d**: NanoDSF of UBR5 to assess protein stability. **e**: ubiquitylation assay in pulse-chase format to test different E2 enzymes with UBR5 and fluorescent ubiquitin. Reducing and non-reducing samples were taken at the indicated time points. Arrows mark UBR5-independent ubiquitin-bound species. **f**: Pulse-chase assay to test importance of UBR5's catalytic cysteine with fluorescently labeled K0 ubiquitin as donor and unlabeled WT acceptor ubiquitin. Non-reducing gel is shown, cropped for visualization. **g**: Principle of ubiquitin-linkage identification using AQUA peptides in a mass spectrometry-based approach. **h**: Identification of ubiquitin chain types formed by UBR5, measured with AQUA-MS.

## Structural investigation of UBR5

Since reconstituted UBR5 showed similar properties as proposed *in vivo*, we wanted to gain structural insights using cryo-EM.[67,179] Sufficient amounts were first obtained for UBR5$^{C2768A}$ and therefore, cryo-EM was initially performed with this mutant.

We obtained nicely distributed particles, that had relatively strong contrast and were easy to spot. However, severe preferred orientations and relatively strong sample heterogeneity, presumably caused by crashed UBR5, made it difficult to obtain a reliable initial model (**Figure 14a**). Addition of detergents improved the sample by allowing more conformations, and increasing the particle density (**Figure 14b-d**). Striking differences between the tested detergents were observed. Incubation with CHAPSO (**Figure 14d**) promised more distinct particle-orientations compared to β-OG (**Figure 14c**) as observed in 2D classes and the angular distribution of a 3D-reconstruction, while not inducing any detergent-based artifacts as FOM did (**Figure 14b**), and maintaining higher particle integrity compared to samples without detergents. Therefore, we chose to collect a bigger dataset of UBR5$^{C2768A}$ incubated with CHAPSO.

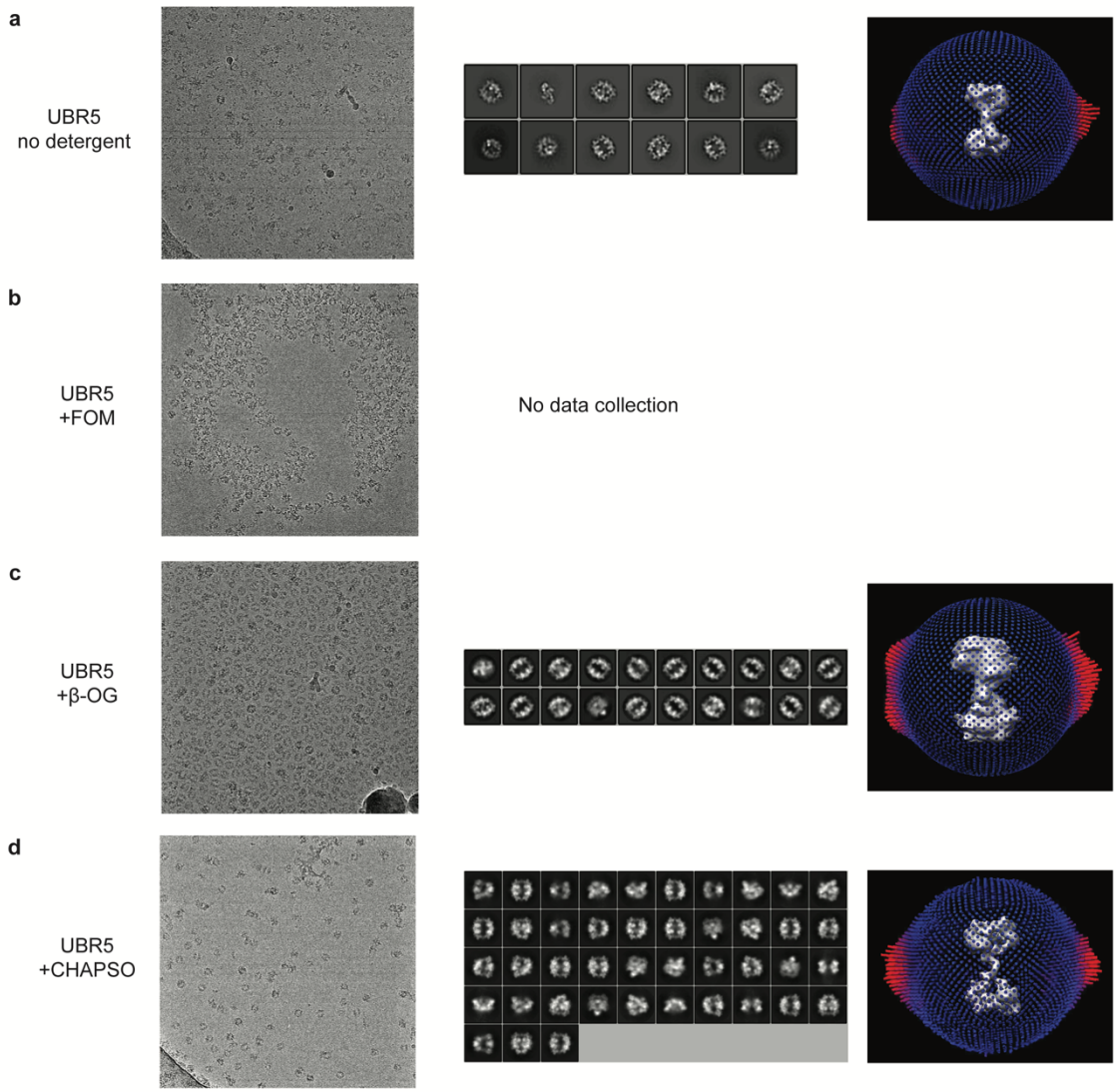

**Figure 14 Cryo-EM sample optimization of UBR5$^{C2768A}$ using detergent-screening.**

**a:** Representative micrograph of UBR5$^{C2768A}$ without detergent, 2D classification and angular distribution of a homogeneous refinement are depicted to visualize preferred orientations. **b:** Representative micrograph of UBR5$^{C2768A}$ plunged in the presence of FOM. **c:** Representative micrograph of UBR5$^{C2768A}$ with β-OG, 2D classification and angular distribution of a homogeneous refinement are shown. **d:** Representative micrograph of UBR5$^{C2768A}$ plunged with CHAPSO, 2D classification and angular distribution of a homogeneous refinement of a small dataset are depicted.

A larger dataset of UBR5$^{C2768A}$ derived in a reliable initial model and subsequently in a ~3.7 Å reconstruction (**Figure 15a**). Using AlphaFold2, we calculated a UBR5-model, which in part fit very nicely into the obtained density for UBR5$^{C2768A}$ (**Figure 15b**).[181] The AlphaFold2-model showed a propeller originating in the N-terminus and interrupted to yield two small β-barrels. The second main observed feature were α-helical repeats that were mostly coded for in the center of the UBR5-open reading frame. Lastly, the HECT domain is located at the far C-terminus. Splitting the model into these parts 1) propeller + small β-barrels, 2) α -helical scaffold, 3) HECT domain, allowed fitting them into the density very nicely. This provided several new insights: UBR5 forms a huge tetramer consisting of an upper and lower dimer, each stabilized by a major interface on the α-helical scaffold, and the two dimers interacting with each other via the two small β-barrels encoded in the N-terminal region (**Figure 15c**). During preparation of a manuscript[141], several other groups reported similar structures for UBR5, supporting that our obtained density map is not an artifact caused by the C2768A mutation (**Figure 15d**).[137-140]

In-depth analysis of obtained 3D-classes showed flexibility of the two dimers against one another as these revealed breathing motions of up to 20° with respect to each other, indicating that the dimer might be the minimal active unit for catalysis and explaining the comparably lower resolution of one half of the tetramer (**Figure 15e**). The huge extent of the interface between the two scaffolds indicates high stability and probably co-folding of this region, supporting our previous observation of an oligomerization-dependent protein-stabilization (**Figure 13d**). However, the interface between the two small β-barrels is a tempting target for mutagenesis. We designed a structure-based mutant in this area with the aim to disrupt the interface and generate stable dimeric UBR5, which might simplify further analysis (**Figure 15f**).

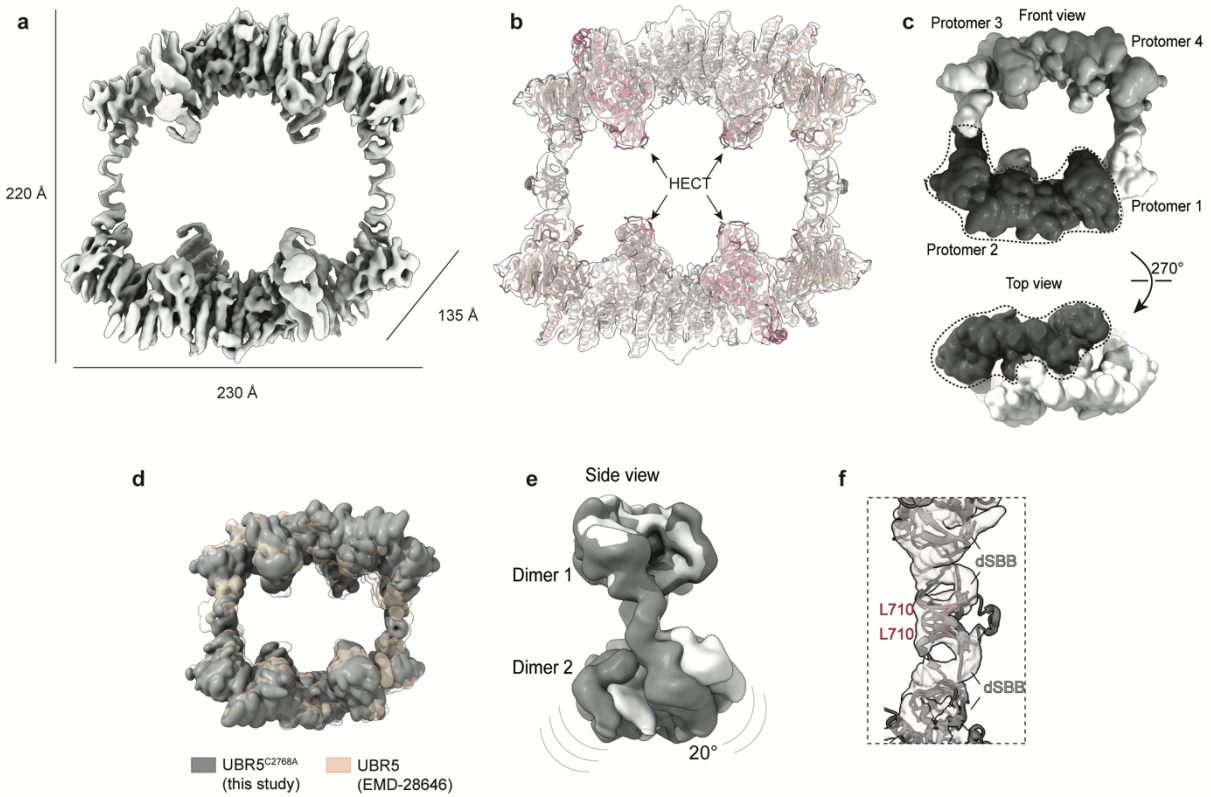


**Figure 15 Analysis of cryo-EM density map of UBR5$^{C2768A}$.**

**a**: Cryo-EM map of UBR5$^{C2768A}$ processed with C2-symmetry. Dimensions are depicted in Å. **b**: Fitting of parts of an AlphaFold2-model of UBR5 into the density map. Positions of the four HECT domains are indicated. **c**: Potential assembly of the four UBR5 monomers into a low-pass filtered tetrameric density map of UBR5$^{C2768A}$. The different monomers are shaded differently and one monomer is additionally highlighted with a dotted line. **d**: Overlay of our UBR5$^{C2768A}$-derived cryo-EM density map and a density map simultaneously published for UBR5. **e**: Movement of the two dimers within a tetramer against each other, observed during 3D-classification. **f:** Zoom into fitted barrels connecting upper and lower dimer. The residue selected for mutagenesis (L710) to disrupt the tetramer is annotated.

## Biochemical reconstitution of dimeric UBR5

Indeed, a L710D mutation, that would add repelling charges to this interface, succeeded in generating dimers of UBR5 rather than tetramers, as it could be observed using mass photometry (**Figure 16a**).[182] For this reason, UBR5 containing the L170D mutation will subsequently be referred to as UBR5$^{Dimer}$. To see whether the previously observed features of UBR5 are mediated by the tetrameric state, we again tested which ubiquitin linkages can be formed by the dimer (**Figure 16b**). A di-ubiquitin synthesis assay was performed with UBR5 or UBR5$^{Dimer}$ being mixed with different unlabeled ubiquitins, containing only the indicated lysine and all other lysines were mutated to arginine in the chase-reaction. Both, UBR5 as well as UBR5$^{Dimer}$ showed striking preference for K48, verifying that K48-specificity is an intrinsic UBR5-feature that is not mediated by the higher oligomerization-state. We furthermore wondered whether the tetrameric state is required to form longer ubiquitin chains (**Figure 16c**). For this reason, we performed a polyubiquitylation assay with labeled WT ubiquitin, that could be processed into ubiquitin chains by UBR5. Both, UBR5, and UBR5$^{Dimer}$ were able to generate long untethered

ubiquitin chains. In case of WT UBR5, increased signal at the top of the SDS-PAGE might indicate higher autoubiquitylation compared to $UBR5^{Dimer}$ or higher abundance of very long free ubiquitin chains. Yet, an increased signal at the top for UBR5 compared to $UBR5^{Dimer}$ can already be observed after 7 seconds, at which point there are barely other chains present yet. Therefore, it seems more likely that UBR5 undergoes increased autoubiquitylation. One possible explanation is that autoubiquitylation might happen in trans (**Figure 16d**). Since UBR5 has many unstructured regions, which might reach from one dimer to the other, these could preferentially get modified in the tetramer compared to the dimer. Another option is that the tetramer might be better in polyubiquitylating itself due to spatial accessibility, and therefore might not reach more or different sites, but it could rather attach more ubiquitins to distinct sites.

We furthermore tested whether both UBR5-versions can use differently long ubiquitin chains as acceptor (**Figure 16e**). We sought to understand whether the higher oligomerization-state of UBR5 is maybe required to stably bind longer ubiquitin chains to facilitate their modification. K63 chains up to penta-ubiquitin were used as acceptors in a pulse-chase assay and both UBR5-versions were able to use all differently long chains as acceptor with increasing efficiency. Since longer K63-linked chains provide more acceptor sites for UBR5, these chains can get modified multiple times on the distinct ubiquitin-moieties, thereby creating a ladder of ubiquitin-modifications. Interestingly, this effect seems to be slightly stronger for tetrameric UBR5 compared to dimeric UBR5. If K63-linked tri-ubiquitin is used as acceptor, the lane corresponding to fluorescent tetra-ubiquitin is stronger for dimeric UBR5 compared to tetrameric UBR5. However, in both samples, a similarly intense lane corresponding to fluorescent penta-ubiquitin can be observed. Also, only with tetrameric UBR5 and not with dimeric UBR5, a hexa-ubiquitin chain can be observed in this sample. We hypothesize that higher local concentration of UBR5 allows catching of a dissociated ubiquitin chain that had previously been modified and facilitate ubiquitylation of another ubiquitin-moiety within this chain. Another option is that tetrameric UBR5 can bind the chains tighter and thereby allow modification of these chains on several sites simultaneously.

Dimeric UBR5 did not show any deficiencies when modifying long chains with a single ubiquitin. This might indicate that the higher oligomeric state of UBR5 is not required to bind long chains to facilitate their modification, or that such an effect would only be visible with much longer chains.

With all different ubiquitin chains, dimeric UBR5 seemed to be a bit more active compared to tetrameric UBR5. This can most likely be explained by distribution-effects. For both UBR5-versions, the same number of UBR5 moieties was used rather than the same number of UBR5-molecules. $E2{\sim}Ub^{D}$ presumably would be distributed evenly in the solution. However, molecules of tetrameric UBR5 are distributed rather sparsely with higher local concentrations, dimeric UBR5 instead could be distributed also rather homogeneous in the solution with a lower local concentration. Overall, this would lead to more rapid encounter of dimeric UBR5 with $E2{\sim}Ub^{D}$ over tetrameric UBR5 with $E2{\sim}Ub^{D}$.

The respective acceptor ubiquitin is provided in excess and therefore the distribution-effect of UBR5 with respect to the acceptor can be neglected.

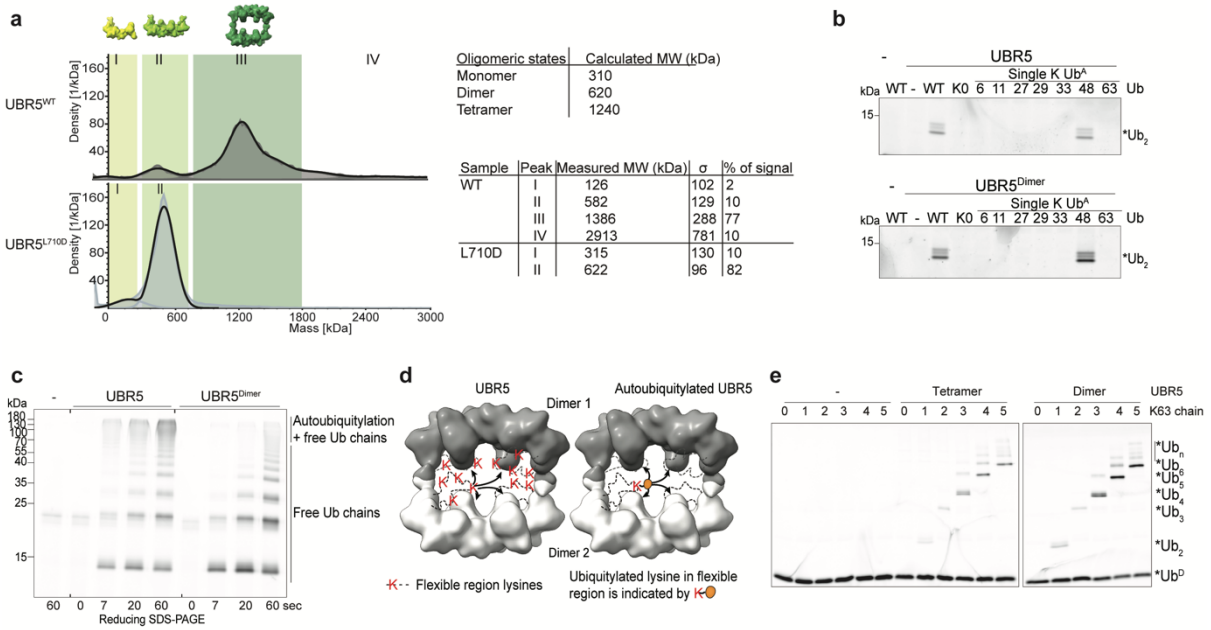


**Figure 16 Biophysical and biochemical characterization of UBR5 with a L710D point mutation.**

**a**: Mass photometry of UBR5$^{WT}$ and UBR5$^{L710D}$. **b**: Pulse-chase assay of UBR5$^{WT}$ and UBR5$^{Dimer}$ with single lysine-acceptor ubiquitin versions to assess the linkage-specificity of UBR5. **c**: Polyubiquitylation assay with UBR5 or UBR5$^{Dimer}$ and fluorescently labeled wildtype ubiquitin. **d**: Cartoon of potentially different autoubiquitylation-mechanisms of UBR5 and UBR5$^{Dimer}$. **e**: Pulse-chase assay UBR5 and UBR5$^{Dimer}$ with fluorescently labeled donor ubiquitin harboring a K48R mutation and various different lengths of K63-ubiquitin chains as acceptors.

## Structural investigation of UBR5$^{Dimer}$

Cryo-EM data obtained for UBR5$^{Dimer}$ yielded a 2-fold symmetric density map at 2.7 Å resolution. The high resolution allowed building a UBR5 structure on a side chain level in most places (**Figure 17a-b**). However, long disordered regions in UBR5 do not allow us to unambiguously conclude the dimerization on the huge interface (**Figure 17c**). The two monomers might either form an “S-shape” with the HECT domain of each monomer being distant from the propeller on the side, or the monomers might form a “C-shape” with the HECT domain directly neighboring the propeller of the respective monomer. Distances bridging the modelled residues in the scaffold, support the hypothesized S-shape and additionally, this orientation was proposed by other groups simultaneously working on the structure of UBR5 and will therefore be assumed throughout this study.[137-139]

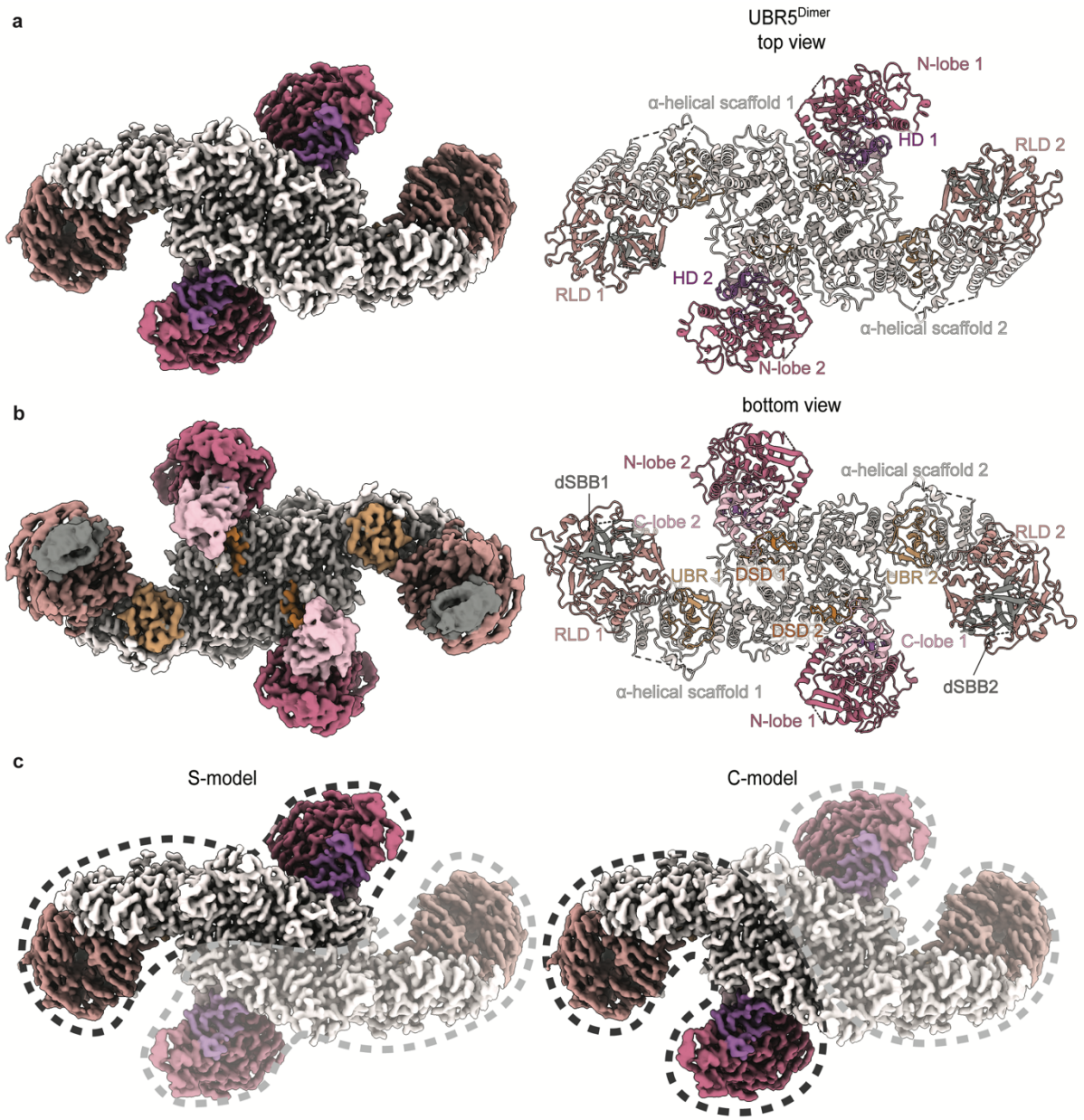


**Figure 17 Structural investigation of UBR5$^{Dimer}$ using cryo-EM.**

**a**: Top view of UBR5$^{Dimer}$. Left: density-map of UBR5$^{Dimer}$ with different domains colored. Right: Cryo-EM map derived structure of UBR5$^{Dimer}$. **b**: Bottom view of UBR5$^{Dimer}$. Left: density-map of UBR5$^{Dimer}$ with different domains colored. Right: Cryo-EM map derived structure of UBR5$^{Dimer}$. **c**: Potential assembly of the two monomers within dimeric UBR5.

The high-resolution structure allowed detailed analysis of distinct domains and new domain annotations in some cases (**Figure 18a**). Our findings on the UBR5 domains of apo UBR5 are in agreement with further reports describing the structure of UBR5 that were published alongside our study.[137-141]

The HECT domain N-lobe is interrupted by the MLLE domain that was proposed to have regulatory or substrate-recognition functions (**Figure 18b**).[157] The MLLE domain is tethered to the N-lobe by very long, presumably largely disordered stretches and is not visible in any of our or the other UBR5 cryo-EM maps.[137-141] Similar to most crystal

structures of HECT domains that are lacking the ultimate six residues of the C-lobe, also the C-terminus of UBR5 is not resolved in our structure. However, the C-lobe in our cryo-EM reconstruction closely resembles a previous crystal structure of UBR5's C-lobe.[183]

UBR5's C-terminal HECT domain is positioned in the L-conformation, as it had also been computed by AlphaFold2. This conformation is stabilized by multiple elements emerging from the scaffold (**Figure 18c**). The central region mediates dimerization and contains two meandering sequences that bind the HECT domain. We refer to these as **D**omain **S**wap **D**imerization (DSD) and **H**ECT **D**isplay (HD) domain. As part of the huge interface within the scaffold, one protomer's DSD domain is partly embedded in a groove of the other subunit. A peptide-like loop from this domain extends beyond the scaffold towards the C-lobe of one HECT domain, presumably to stabilize it in the L-conformation. The 6 kDa HD domain interacts with the scaffold on one side and the other side engages a concave surface on the N-lobe. Interestingly, a peptide-tiling screen yielded a hit located in the HD domain, termed "wedge" in the respective study.[139] This supports the importance of this domain in UBR5-mediated substrate-turnover.

Overall, the HECT domain arrangement is constituted by a seven-way interaction: 1) the HECT domain is encoded adjacent to the scaffold, and therefore the overall positioning of the HECT domain to UBR5 is defined, 2) the DSD domain orients the C-lobe, 3) the DSD domain is tethered to the scaffold, 4) the HD domain presents the N-lobe, 5) stabilization of the HD domain via interactions with the scaffold, 6) interaction of HD and DSD domains, presumably to harmonize presentation of N- and C-lobe, and 7) orientation of the N- and C-lobe to each other, connected by their flexible hinge.

N-terminal of the HECT domain, the immense α-helical scaffold is encoded (**Figure 18d**). It has several roles apart from its involvement in the HECT domain orientation: 1) It shapes the overall UBR5-architecture, by serving as scaffold and positioning all other domains with respect to each other, 2) it forms a huge dimerization interface, with one helix (1912-1929) being particularly inserted into the other scaffold, 3) it encompasses many differently long disordered insertions that might facilitate protein-protein interactions and serve as platforms.

The scaffold also incorporates UBR5's UBR domain that recognizes N-terminal arginines of substrates (**Figure 18e**).[150,151] The UBR domain adopts a triangular shape with three zinc ions chelated by C1179/C1208/C1211/C1232, C1211/C1215/C1234/C1240, and C1196/C1199/H1216/H1219. Our cryo-EM map revealed additional density in the substrate binding pocket. Overlaying the crystal structure of the conserved UBR domain of UBR2 bound to a substrate peptide[184], revealed that the additional density does not block the arginine-binding pocket, thereby potentially modulating substrate-specificity, but rather impairs peptide-binding overall. The low resolution of the additional density did not allow assignment of it. Nonetheless, several potential origins of the density are conceivable: it might be caused by co-purification of a binding partner or it might point towards a regulatory mechanism with the substrate-recognition site being blocked and only being made available under specific conditions. This could be achieved by

autoinhibition, peptide-binding, or small molecule binding. This hypothesis of a regulatory blocking of the UBR domain is supported by the fact that to date, no substrate that is recognized via an N-terminal arginine has been found for UBR5.

Ultimately, an RCC1-like domain is located at the N-terminus of UBR5 (**Figure 18f**). Notably, RLD domains are defining features of the HERC subfamily of HECT E3 ligases (**Figure 7**). Thus, the structure suggests that UBR5 could be classified as a distant member of the HERC class of E3 ligases. The 7-bladed propeller is severely interrupted at several positions: The first blade starts right at the N-terminus and is completed by two β-sheets from the C-terminal region of the propeller. The second blade harbors three antiparallel β-sheets followed by a ~270 amino acid long interruption that inserts the UBA domain connected by long linkers. High flexibility of the UBA domain is further demonstrated by the absence of cryo-EM density for this domain. The fourth sheet is added after the interrupting domain. The third to fifth blade are conventional blades that do not harbor any interruptions, however the loop connecting the fifth and the sixth blade again is interrupted by a ~240 amino acid long region that contains the barrels responsible for the small interaction.

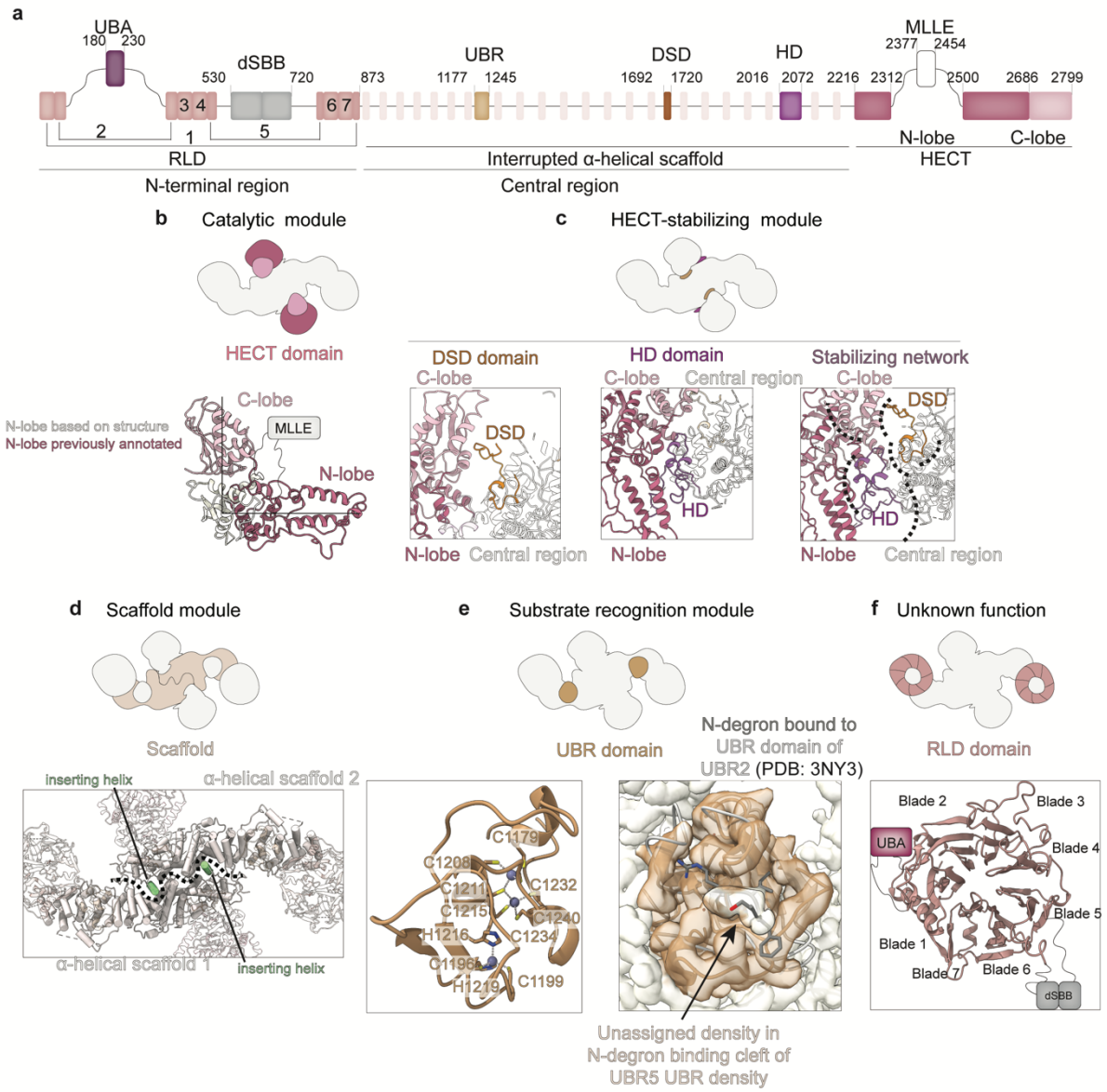


**Figure 18 Analysis of UBR5$^{Dimer}$ structure.**

**a**: Updated UBR5 domain-annotation with domain boundaries indicated by residue numbers. **b**: Zoom in on catalytic HECT domain. The previous HECT domain annotation and the updated structure-based annotation are indicated. **c**: Illustration of the different components of the HECT-stabilizing module. **d**: Zoom into scaffold module with particularly embedded α-helix highlighted. **e**: Analysis of substrate recognition module. Left: zinc-complexation by UBR domain elements. Right: Additional density found in cryo-EM derived map in substrate-binding pocket of UBR5$^{Dimer}$. **f**: Structure of UBR5's RLD with the seven blades and the interruptions harboring dSBB and UBA domains indicated.

## Role of UBR5 domains during polyubiquitylation

To understand the mechanism UBR5 employs for polyubiquitin chain formation, we took a structure-based approach to determine minimal UBR5-constructs that retained their polyubiquitylation-activity. Given that many HECT domains are sufficient to form ubiquitin chains -even though often less efficient than their full-length counterparts- we attempted to test different HECT-constructs side-by-side with full-length UBR5 (**Figure 19a**).[89,116] Ubiquitin-binding domains have been shown to be crucial for recruiting acceptor ubiquitins

to facilitate polyubiquitylation by other ubiquitin chain forming enzymes.[89,185-190] We thus considered that also UBR5's UBA domain could play a key role in ubiquitin chain formation. We generated a construct stretching over the entire HECT domain including the inserted MLLE domain, a construct occluding the MLLE domain to account for potential regulatory functions, a linear fusion of the UBA domain to the HECT domain, and a construct covering all of UBR5 as a tetramer (WT) or dimer (L710D mutant). Strikingly, di-ubiquitin formation could only be achieved by the full-length constructs ($UBR5^{Dimer}$ and UBR5) or by the UBA-HECT fusion, validating that the UBA domain indeed plays a crucial role in polyubiquitylation. The constructs covering the HECT domain or $HECT^{\Delta MLLE}$ remained catalytically active in that they can receive $*Ub^{D}$ from the E2 in the chase-reaction, however, they cannot pass it on to $Ub^{A}$ to form di-ubiquitins. We furthermore sought to understand whether the UBA-HECT fusion retained the linkage-specificity (**Figure 19b**). Therefore, we performed a similar assay than before, however, in this case $Ub^{A}$ contained a K48R mutation and could only serve as acceptor if $*Ub^{D}$ gets transpeptidated to another amino group than K48. Using this acceptor completely abolished di-ubiquitin-formation for all constructs, showing that also the minimal UBA-HECT-construct maintains the specificity for K48-linkages.

Even though omitting the MLLE domain in a construct covering only the HECT domain, did not affect polyubiquitylation at all, we still reasoned whether the MLLE domain would affect UBR5-mediated polyubiquitylation in an otherwise full-length construct. To test whether the MLLE domain might be modulating UBR5's catalytic activity somehow, we aimed to identify its reaction products in an unbiased way. Apart from a mass spectrometry-based approach, we had previously employed a pulse-chase assay with distinct lysines present in $Ub^{A}$ to identify the formed ubiquitin linkages. In this case only defects that already occur in the first isopeptide-formation can be observed. However, if specificity is only affected after several iterations, for example by allowing branched or mixed chain formation, you could not detect it using our pulse-chase-method.[116] **Ubi**quitin **C**hain **Res**triction assay (UbiCRest) is an approach to analyze what ubiquitin linkages a protein of interest generated, after allowing polyubiquitylation.[191] Similar to the mass-spectrometry approach, which is quite tedious and depends on the availability of the respective machines, UbiCRest analyses the linkages that have been formed rather than investigating the assembly directly. Adding different deubiquitylating enzymes that specifically cleave certain ubiquitin chain linkages then allows to recapitulate what type of linkages had been formed. We first established reaction conditions under which the respective DUBs could cleave the respective di-ubiquitin (**Figure 19c**).

We formed ubiquitin chains and subsequently analyzed them using this approach (**Figure 19d**). Both, UBR5 and $UBR5^{\Delta MLLE}$ were able to form comparably long ubiquitin chains, indicating that the MLLE domain is not crucial for formation of long ubiquitin chains. To identify the chain types, we then added the unspecific Usp2, K48-specific OTUB1, and K11-specific Cezanne. Since the K63-specific DUB AMSH appeared to be less active than the others, we omitted this sample. Both, chains formed by UBR5 and $UBR5^{\Delta MLLE}$ were completely digested by the unspecific Usp2 and also the K48-specific OTUB1, and could

not at all be processed by Cezanne. This demonstrates that deletion of the MLLE domain does not impair linkage-specificity even upon formation of long ubiquitin chains.

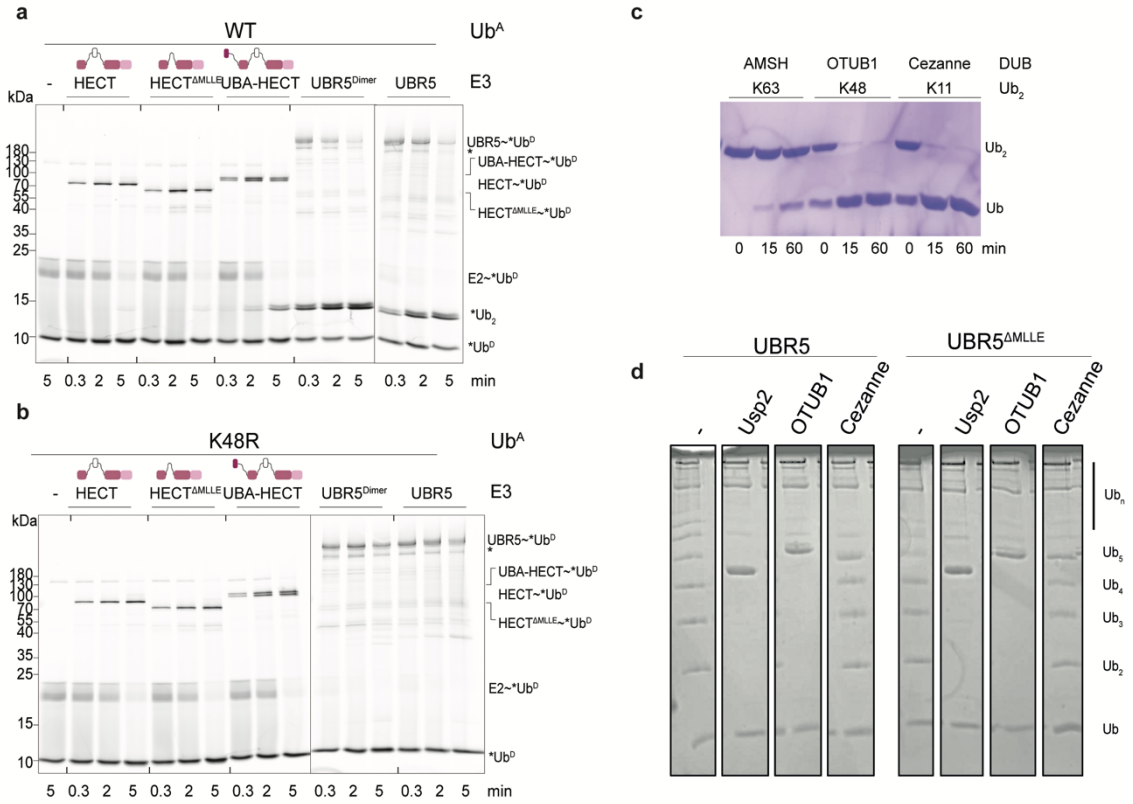


**Figure 19 Biochemically assessing roles of UBR5 domains during ubiquitin chain formation.**

**a**: Pulse-chase assay testing different versions of UBR5: the redefined HECT domain with (HECT), or without the interrupting MLLE domain (HECT$^{\Delta MLLE}$), a fusion of the UBA domain and the HECT domain (UBA-HECT), UBR5$^{Dimer}$, and UBR5. **b**: Pulse-chase assay to test linkage-specificity in different UBR5-versions. Unlabeled acceptor ubiquitin harboring a K48R mutation was used. **c**: Testing catalytic activity of different deubiquitylating enzymes. **d**: UbiCRest assay to test linkage-specificity of UBR5 or UBR5$^{\Delta MLLE}$ during polyubiquitylation. A multi-turnover assay was performed with the respective UBR5-version and the obtained ubiquitin chains were subsequently treated with an unspecific DUB (Usp2), a K48-specific DUB (OTUB1), or a K11-specific DUB (Cezanne).

## 2. Transition state 1: $Ub^D$ transfer from E2 to UBR5

### Visualization of transition state 1 with chemical probes and cryo-EM

Attempting to understand UBR5's extraordinary activity and K48-specificity, we tried to model UBR5's E2~$Ub^D$-bound state on our obtained apo structure (**Figure 20a**). As expected, based on prior publications showing HECT domains in the E2-bound state[84], apo UBR5 cannot accommodate E2~Ub and there are several clashes: if E2-binding occurs via the canonical E2-binding surface on HECT domains, the opposite side of the E2 would clash with the RLD in trans. Additionally, $Ub^D$ would clash with the α-helical scaffold in trans. To understand the conformational change UBR5 has to undergo in order to accommodate E2~Ub, we aimed to gain structural insight into this **T**ransition **S**tate **1** (TS1). Design of the employed activity-based probe, as well as cryo-EM for this complex was performed in close collaboration with Dr. D. Horn-Ghetko. A BmDPA-based probe that was adapted from one previously used to visualize the transition state of an RBR E3 ligase was used to visualize this state for UBR5.[192] The **A**ctivity-**B**ased **P**robe (ABP) mimics conjugation of $Ub^D$ to both, the catalytic cysteine of the E2 UBE2D2 as well as the catalytic cysteine of UBR5 at the same time, showing the transfer of ubiquitin from the E2 to the E3 (**Figure 20b**). Compared to the fleeting native state, the distance between the catalytic cysteines of E2 and E3 is increased by three atoms in this probe. The probe was obtained semi-synthetically in a two-step process (**Figure 20c**): first, a reactive ubiquitin-moiety was generated that could react with the catalytic cysteine of the E2. To provide specificity for this reaction, all other cysteines on the E2 were mutated.[193] Next, the obtained activated Michael-acceptor could react with a cysteine on UBR5. Due to the high number of cysteines in UBR5, it was not feasible to mutate all of them apart from the catalytic one. However, when reacting a fluorescent version of this probe with UBR5 or $UBR5^{C2768A}$, it became apparent that the probe only reacts significantly if UBR5's catalytic cysteine is present (**Figure 20d**).

For structural characterization of the transition state, we reacted the probe with $UBR5^{Dimer}$. The dimer was used for this, as it promised reduced preferred orientations, less particle heterogeneity, and it did not seem to have any defect in catalysis based on our previous characterization. Despite using $UBR5^{Dimer}$, the cryo-EM data did show considerable heterogeneity. Nonetheless, a map without symmetry could be obtained at 7.3 Å resolution (**Figure 20e**). Large fractions of the map strongly resembled $UBR5^{Dimer}$ in the apo-state. However, the density for one HECT domain was significantly changed and extra density next to it could be identified to harbor the E2 and ubiquitin. Density corresponding to the E2 sits on the N-lobe facing the RLD and the additional density corresponding to ubiquitin sits on top of the C-lobe and is corseted between E2, C-lobe, RLD, N-lobe and a loop in the scaffold.

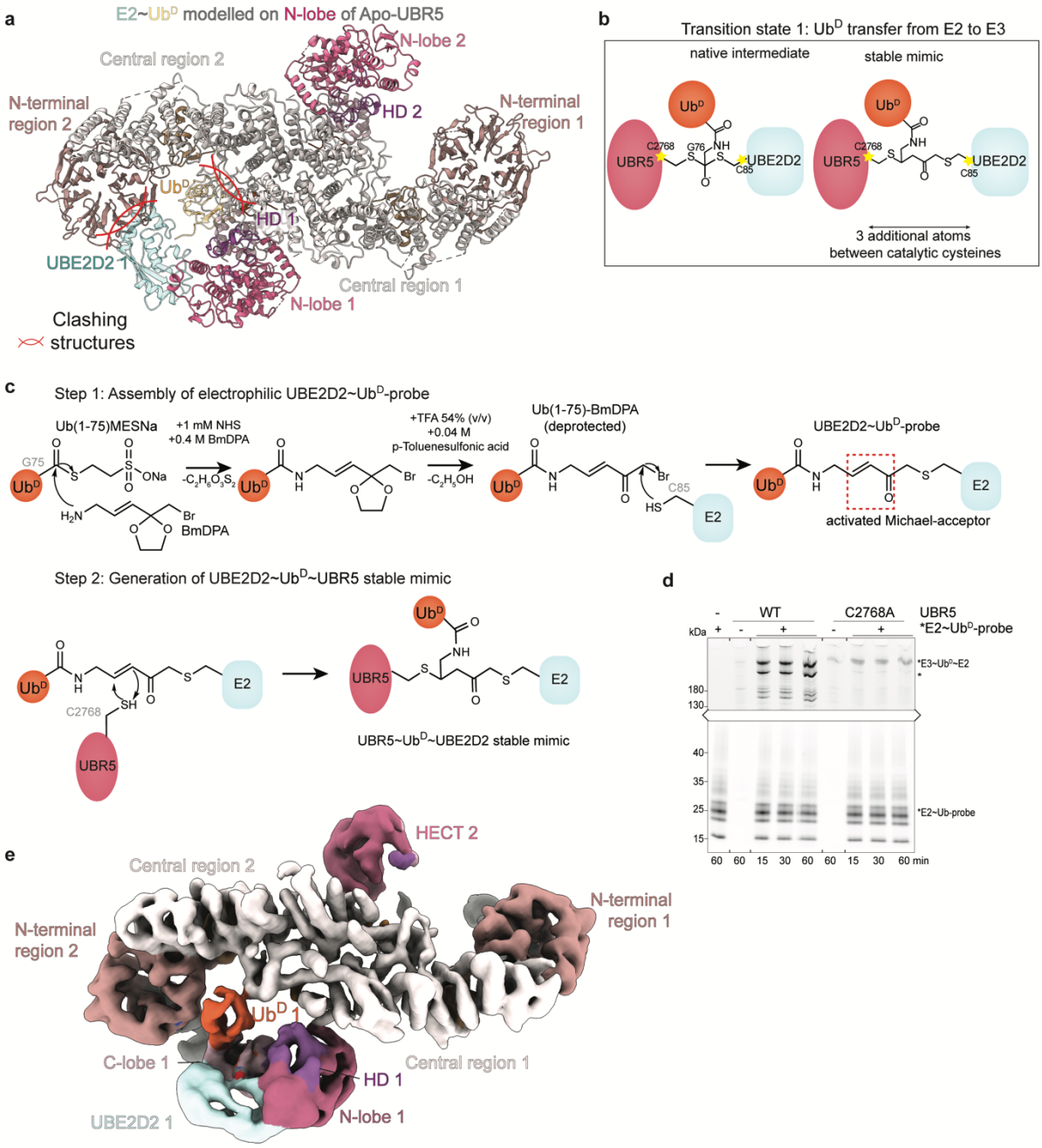


**Figure 20 Visualization of Ub$^D$ transfer from E2 to UBR5 using chemical probes and cryo-EM.**

**a**: Fitting E2~Ub$^D$ into the structure of apo UBR5$^{Dimer}$ shows incompatibility. Clashing regions are indicated with red arcs. **b**: Chemical structures of the native transition state and the chemical proxy employed to capture this transition state. Difference between the native geometry and the ABP is indicated. **c**: Reaction scheme to generate TS1-ABP. **d**: Probe-reactivity with UBR5 and UBR5$^{C2768A}$ was tested using a fluorescent version of the TS1-activity based probe. **e**: Density map obtained by cryo-EM of UBR5$^{Dimer}$ reacted with the TS1-probe.

## Structural analysis of transition state 1

The density of the other HECT domain is largely undefined and presumably portrays a mixture of conformations, which could not be abolished despite intense classification. Focusing on the HECT domain that showed clear density allowed us to generate a model of the E2 and ubiquitin bound to it by incorporating prior knowledge regarding

HECT~Ub$^{D}$~E2-complexes (**Figure 21a**).[84] In agreement with prior structures of isolated HECT domains in complex with E2~Ub$^{D}$, the HECT domain was positioned in the inverted T-conformation with the C-lobe sitting in the center of the N-lobe. The E2 binds via the canonical E2-binding site on the N-lobe. C-lobe and the E2 are oriented in a way that the catalytic cysteines are juxtaposed to facilitate transthiolation of ubiquitin's C-terminus from the E2 to the E3. Fitting the crystal structure of NEDD4L's HECT domain bound to UBE2D2~Ub$^{D}$ (PDB: 3JVZ) shows that the mechanism of transthiolation is conserved amongst HECT E3 ligases, as the structure fits very well into the UBR5-density (**Figure 21b**). In this crystal structure, E2 and ubiquitin are linked via an oxyester-bond that would maintain the native distance. The close resemblance of our density-derived model with this structure reassures us that the altered geometry of our employed probe did not significantly impact this transition state.

Analysis of the required movements UBR5's HECT domain has to undergo to bind the E2, revealed that the N-lobe has to rotate and tilt by 30° and 25 Å respectively relative to the scaffold, to ensure that the E2 would not clash with the RLD and ubiquitin would not clash with the scaffold (**Figure 21c**). Additionally, the C-lobe shifts and rotates significantly by >40 Å and 150° respectively around the interlobe tether to face the N-lobe-bound E2 (**Figure 21d**). These substantial conformational changes require disengagement or significant reorientation of the HECT-stabilizing module. Low resolution density of the HD domain suggests relocation of this domain (**Figure 21e**). This would be in agreement with the comparably little movement of the N-lobe, which the HD domain could potentially still cope with by minor movement. However, the low resolution of the map did not allow identification of density corresponding to the DSD domain. Yet, the substantial movement of the C-lobe, which the DSD domain interacts with, makes it unlikely that this domain remains bound. Overall, the required dissociation from DSD and potentially the HD domain destabilizes this conformation for UBR5 and one can imagine that UBR5 wants to revert into the L-conformation as soon as it would be spatially possible.

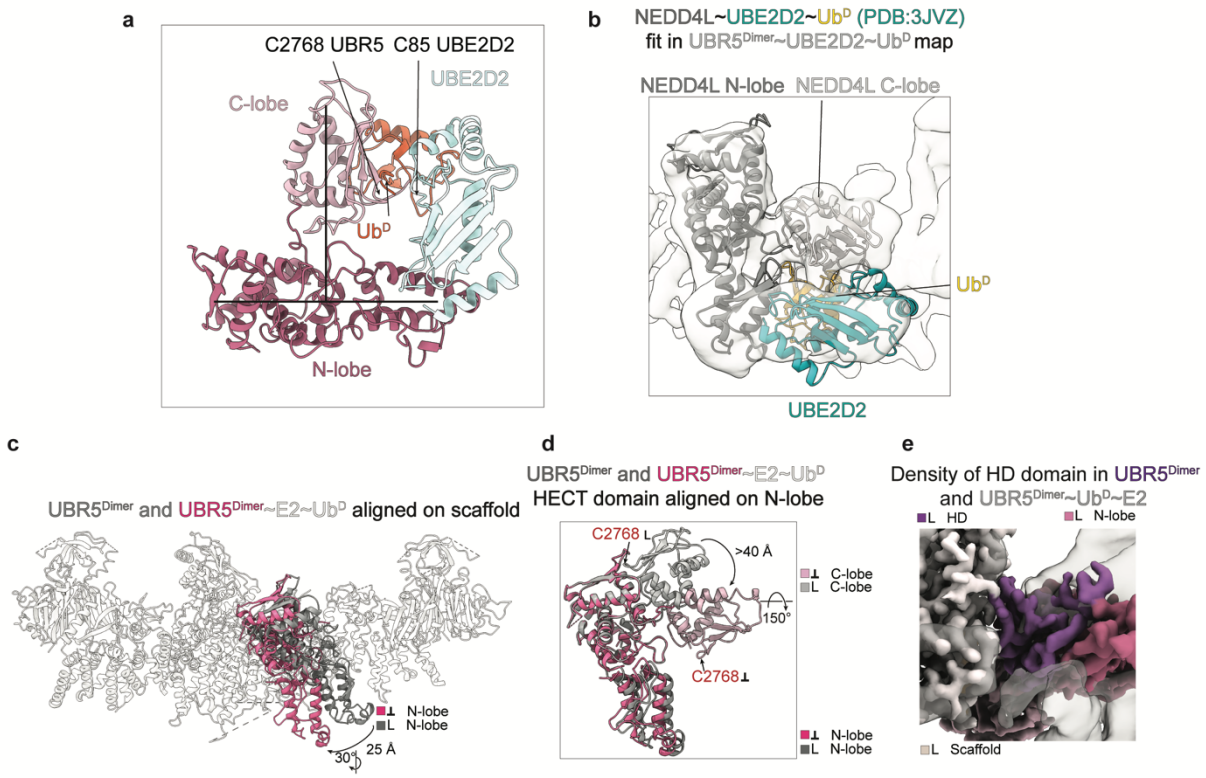


**Figure 21 Analysis of UBR5~Ub$^D$~E2 TS1 model.**

**a**: Model of E2~Ub$^D$ (PDB: 3JVZ) bound to HECT domain of UBR5 in the inverted T-conformation. Catalytic cysteines of UBR5 and UBE2D2 are indicated. **b**: Overlay of NEDD4L's HECT domain bound to UBE2D2~Ub$^D$ fit into density map of TS1 of UBR5. **c**: Scaffold-superposition of UBR5 and TS1 UBR5, depiction of UBR5 up to residue 2686. N-lobe movements between the different states are indicated. **d**: Superposition of apo UBR5 and TS1 UBR5 HECT domains, aligned on N-lobe. Rearrangement of the C-lobe is indicated. **e**: Density of UBR5$^{Dimer}$ HD domain (purple) overlayed with density of UBR5 in TS1 (transparent) showing necessary relocation of the HD domain.

## Biochemical characterization of transition state 1 model

To validate the generated model for TS1, we performed structure-based mutagenesis to test the effects *in vitro*. The first interface we analyzed was the one between the N-lobe and the E2 (**Figure 22a**). Our model as well as prior crystal structures suggested that the E2 binds mainly via its phenylalanine 62 to a hydrophobic and largely aromatic cleft on the N-lobe.[194] Reducing these E2-N-lobe contacts by introducing the F62A mutation into UBE2D3 impaired transthiolation and resulted in reduced di-ubiquitin formation. Next, we aimed to test the interaction between UBR5's C-lobe and Ub$^D$ (**Figure 22b**). Introducing a sterical clash between A2790 in the ultimate helix of UBR5's C-lobe and L71 and L73 in the C-terminus of Ub$^D$ by mutating A2790W, severely impaired discharge of E2~Ub$^D$ to the E3.

We also tested whether additional elements that have not been described for HECT E3s previously, might also contribute to this transition state. The map revealed close proximity of ubiquitin to one particular loop in the scaffold in trans. We termed this loop **S**caffold

**D**onor ubiquitin **A**pproaching (SDA, residues H1362-L1364) (**Figure 22c**). Our model suggests that the histidine of the SDA could form weak hydrogen bonds with Q2 of $Ub^D$ and a weak salt bridge with E64 of $Ub^D$. Furthermore, L1363 and L1364 of the SDA could approach $Ub^D$'s F4 to form hydrophobic contacts. We mutated all three residues of the SDA to aspartates and tested whether we could observe any deficiency in multi-turnover assays. Indeed, polyubiquitylation seems to be slightly impaired when comparing $SDA^{mut}$ to wildtype UBR5. However, when testing this interface in a pulse-chase format, the difference between the UBR5-versions was very subtle, suggesting that the SDA might slightly enhance processivity, but is not required for activity in general (**Figure 22d**).

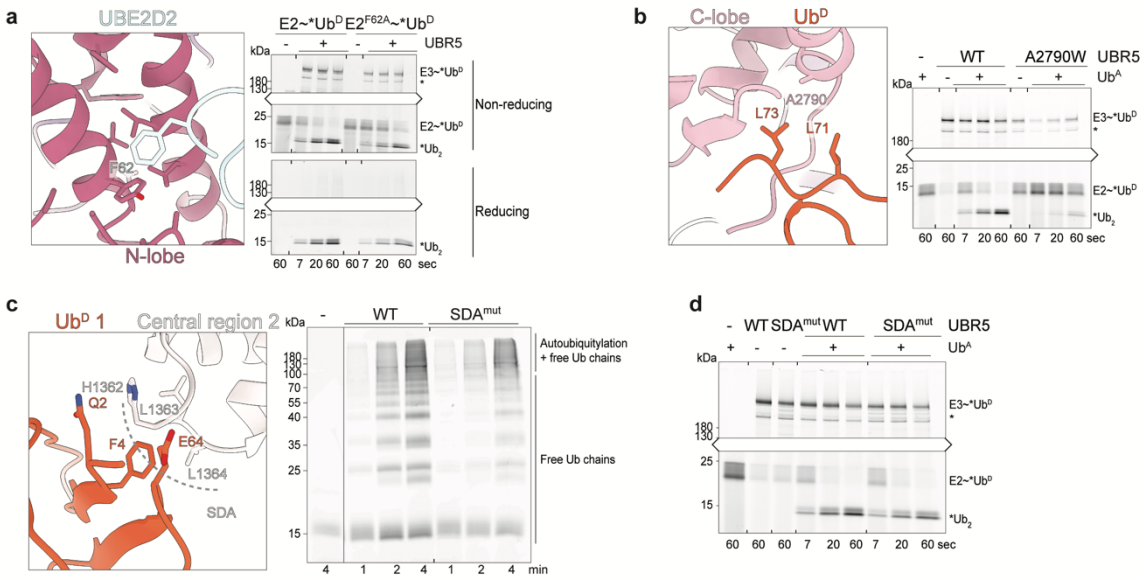


**Figure 22 Biochemical characterization of TS1.**

**a**: Di-ubiquitin synthesis assay in pulse-chase format to test the F62 mutation in the interface of the E2 UBE2D3 and the N-lobe. **b**: Di-ubiquitin synthesis assay to test the C-lobe-$Ub^D$-interaction. **c**: Multi-turnover assay to assess interaction between $Ub^D$ and the scaffold-donor ubiquitin approaching (SDA) loop reaching out of the central region in trans. **d**: Di-ubiquitin synthesis assay testing activity of the SDA mutant.

## Validation of transition state 1 model

We sought to validate our model with an unbiased approach as well. To do so, we employed cross-link mass-spectrometry (**Figure 23a**).[195] In brief, we cross-linked UBR5 with homobifunctional BS3. BS3 contains two amine-reactive groups that are connected with each other by a linker of 11.3 Å length. In the presence of proteins, the amine-reactive groups can react with the side chains of lysines. If two lysines are in close proximity (either intermolecular or intramolecular), both groups of a BS3-molecule can react with these and cross-link them. After a proteolytic digest, peptides that are cross-linked via their lysines, as well as non-cross-linked peptides are obtained. Measurement by LC-MS and subsequent analysis via *in silico* digest and database research then allows to identify what residues were linked with each other and therefore are somehow in close proximity to

each other. Mapping of these cross-links on a structural model can be used to verify such a model, detect new surfaces, identify flexible regions, or even conclude a mechanism.

We employed this analysis to gain insights into apo UBR5 in a resting condition as well as active UBR5 that is performing polyubiquitylation (**Figure 23b).**

Analysis of the obtained cross-links again demonstrates the difficulties of UBR5-structure investigation: Roughly one third of UBR5's amino acids could not be modeled due to highly flexible regions. Many of the identified cross-links are in such flexible regions, making it difficult to draw unambiguous conclusions (**Figure 23c-d**). However, the cross-links that map to regions, where a structure could be derived, are in agreement with our structural data. Generally, most of the identified links are found in or around RLD and dSBB domains. This region shapes the interaction between the two dimers. The multitude of cross-links that can be found there, pinpoints to three mechanisms: 1) it demonstrates significant movement in this area, again supporting that dimeric UBR5 is the main catalytic unit. 2) there are many long amino acid stretches interrupting the well-structured motifs in this region. These long flexible regions might support the interface between the two dimers, as they allow loose-contacts from one dimer to the other. 3) the flexibility of these unstructured regions might indicate a mechanism for substrate recognition and efficient modification, since unstructured regions in proteins often serve as interaction platforms. If the many unstructured regions in this area serve as substrate recognition motifs, the flexibility and localization of the recognition motifs would allow highly processive polyubiquitylation since catalytic domains in both dimers can be easily reached.

Comparison between the cross-links occurring only in the sample of active UBR5 (**Figure 23d**) compared to resting UBR5 (**Figure 23c**) yielded interesting observations: Despite the presence of all cascade-components in the active sample, only intramolecular cross-links within E2, ubiquitin and UBR5 were identified as well as intermolecular cross-links between ubiquitin and UBR5, as well as ubiquitin and the E2. No cross-links within or with the E1 enzyme UBA1 were identified, which might be caused by the high ubiquitin turnover by the E1 and the relatively low concentration.

Besides the many cross-links in the N-terminal region of UBR5, or in other highly disordered regions, only few additional cross-links could be identified. However, one very interesting cross-link was identified, which appeared only in the condition of active UBR5 and not in the sample of resting UBR5: a residue located in the RLD (residue 737) cross-linked to a residue located in UBR5's C-lobe (residue 2780) (**Figure 23d**). In apo UBR5, these two residues would be >30 Å apart (**Figure 23e**). In contrast, if UBR5 relocates into the inverted T-conformation to allow E2~Ub binding, these two residues come significantly closer to one another, and now have a distance of <20 Å (**Figure 23f**). It has to be noted that we did not obtain a high-resolution electron density map for TS1, and therefore are not able to evaluate whether the amino acids in this state engage as different rotamers or whether the organization of RLD or C-lobe maybe slightly differs compared to apo UBR5. The distances are therefore only approximations. Nonetheless, appearance of this cross-link upon addition of E2~Ub to UBR5 strongly supports our model of UBR5 repositioning

into the inverted T-conformation upon E2~Ub binding. Nevertheless, the measured distance in the inverted T-conformation would still be slightly too far for BS3 to reach both residues, indicating slight structural changes within the C-lobe or RLD, which the limited resolution of our map did not reveal.

Additionally, we observe many cross-links of ubiquitin to UBR5, in the active condition (**Figure 23d**). Due to the high processivity of UBR5, we have a mixture of differently long ubiquitin chains in this sample as well as unused ubiquitin (**Figure 23b**). For logistical reasons, we did not include a control of UBR5 with ubiquitin only, but without the other cascade-components. This makes it difficult to analyze the cross-links of UBR5 and ubiquitin. However, it is interesting that ubiquitin has several cross-links to the E2, and the E3, however no cross-links could be identified between the E2 and the E3. We previously established that UBR5-mediated polyubiquitylation requires E2-E3 interaction (**Figure 22a**). Yet with our cross-linking conditions, we were not able to capture this, suggesting that E2 binding to UBR5 is very transient and the E2 dissociates rapidly from UBR5 subsequent to transthiolation to allow reverting to the preferred conformation.

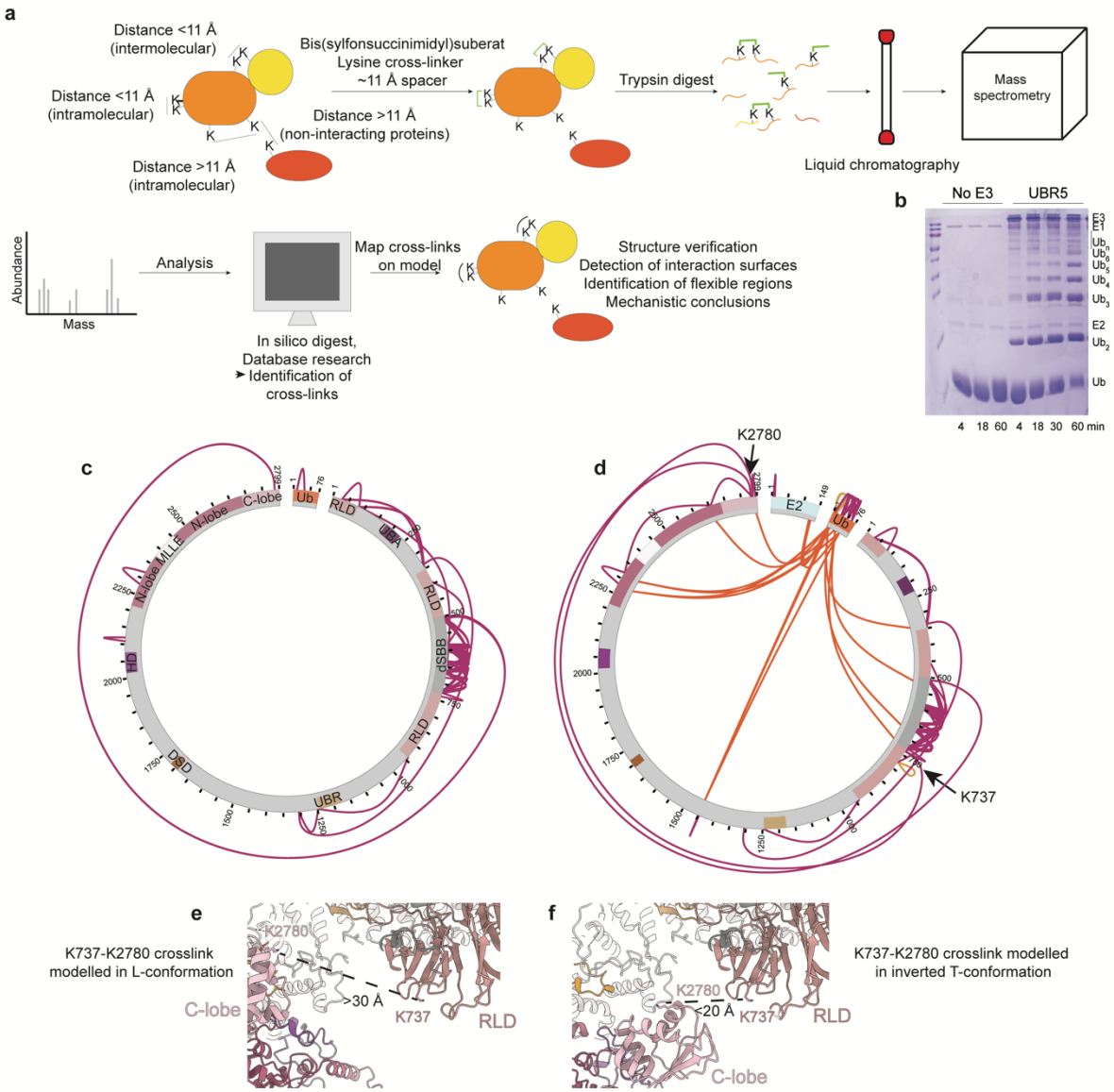


**Figure 23 Testing the TS1-model using cross-link mass spectrometry.**

**a**: Principle of cross-link mass spectrometry. **b**: Sample preparation of active UBR5 by performing a multi-turnover assay. **c**: Cross-link MS data obtained for recombinantly purified UBR5. Intramolecular cross-links are positioned outside of the circle. UBR5 domains are annotated. A low signal for co-purified ubiquitin was noted. **d**: Cross-link MS data obtained for active, recombinantly purified UBR5 after multi-turnover assay. Intramolecular cross-links are positioned outside of the circle, intermolecular cross-links are positioned inside the circle. E2, and ubiquitin (both added during the multi-turnover assay) are annotated. The intramolecular cross-link K737-K2780 is highlighted. **e**: Modeling of the K737-K2780 cross-link identified for active, polyubiquitylating UBR5 in the UBR5 apo-state (L-conformation). **f**: Modeling of the K737-K2780 cross-link identified for active, polyubiquitylating UBR5 in the UBR5 TS1 (inverted T-conformation).

## 3. Ubiquitin-bound intermediate

Subsequently to transferring $Ub^D$ from the E2 to the E3, $Ub^D$ is bound to UBR5's catalytic cysteine, awaiting its further transfer to a substrate or another ubiquitin. Prior crystal structures of other, mostly isolated HECT domains, have proposed different models regarding the overall conformation of this intermediate bridging transthiolation and transpeptidation.[87,88] Thus, we aimed to visualize the UBR5~$Ub^D$ intermediate in context of the full protein, in order to gain more insight into this state and the movements that would be required to facilitate it. We employed the stable proxy **Ub**iquitin-**V**inyl **M**ethyl **E**ster (Ub-VME) to react with UBR5, which would retain native distances between UBR5's catalytic cysteine and $Ub^D$'s C-terminus (**Figure 24a-b**).[196] The reactivity was dependent on UBR5's catalytic cysteine (**Figure 24c**). Subsequently, the obtained complex was subjected to cryo-EM to gain structural insights. Cryo-EM of this complex was done in close collaboration with Dr. D. Horn-Ghetko. A map at 5.3 Å could be derived that resembled apo $UBR5^{Dimer}$, with significant additional density neighboring the C-lobe (**Figure 24d**). We fitted the structure of $UBR5^{Dimer}$ into the density and saw very high consent of the apo-structure and the obtained map for the intermediate state. Additionally, we were able to confidently fit ubiquitin into the extra density (**Figure 24e**). The C-lobe-$Ub^D$ interface seems to be in agreement with the one observed previously for TS1. The C-lobe-$Ub^D$ binding also matches prior publications investigating HECT domains either bound to E2~$Ub^D$, or only bound to $Ub^D$ (**Figure 24f**). However, when paying respect to the entire HECT domain, rather than just the C-lobe~$Ub^D$, we can see that substantial rearrangement of the C-lobe~$Ub^D$-moiety is required compared to TS1 (UBR5~$Ub^D$~E2) (**Figure 24g**). The entire unit swivels around the interlobe-tether to revert into the L-conformation. During processing of the dataset, we barely saw particles that did not show additional density for $Ub^D$, however, the resolution of $Ub^D$ remained significantly lower than the resolution of UBR5 itself. This most likely can be explained by structural heterogeneity caused by flexibility of C-lobe-bound $Ub^D$. We addressed this issue using **3D**-**V**ariability **A**nalysis (3D-VA) implemented in CryoSparc.[197] This analysis revealed a spectrum of orientations of the C-lobe-$Ub^D$-unit ranging from an intermediate between the inverted T- and L-conformation (frame 1) and the other extreme being the L-conformation (frame 20) (**Figure 24h**).

Reduced interactions of E2 to $Ub^D$ subsequent to transthiolation, induce dissociation of the E2 from the complex. Without the bound E2, sterical clashes that would arise from the L-conformation are abolished and UBR5 with $Ub^D$ bound can revert back to the preferred orientation by rotating around the interlobe linker.

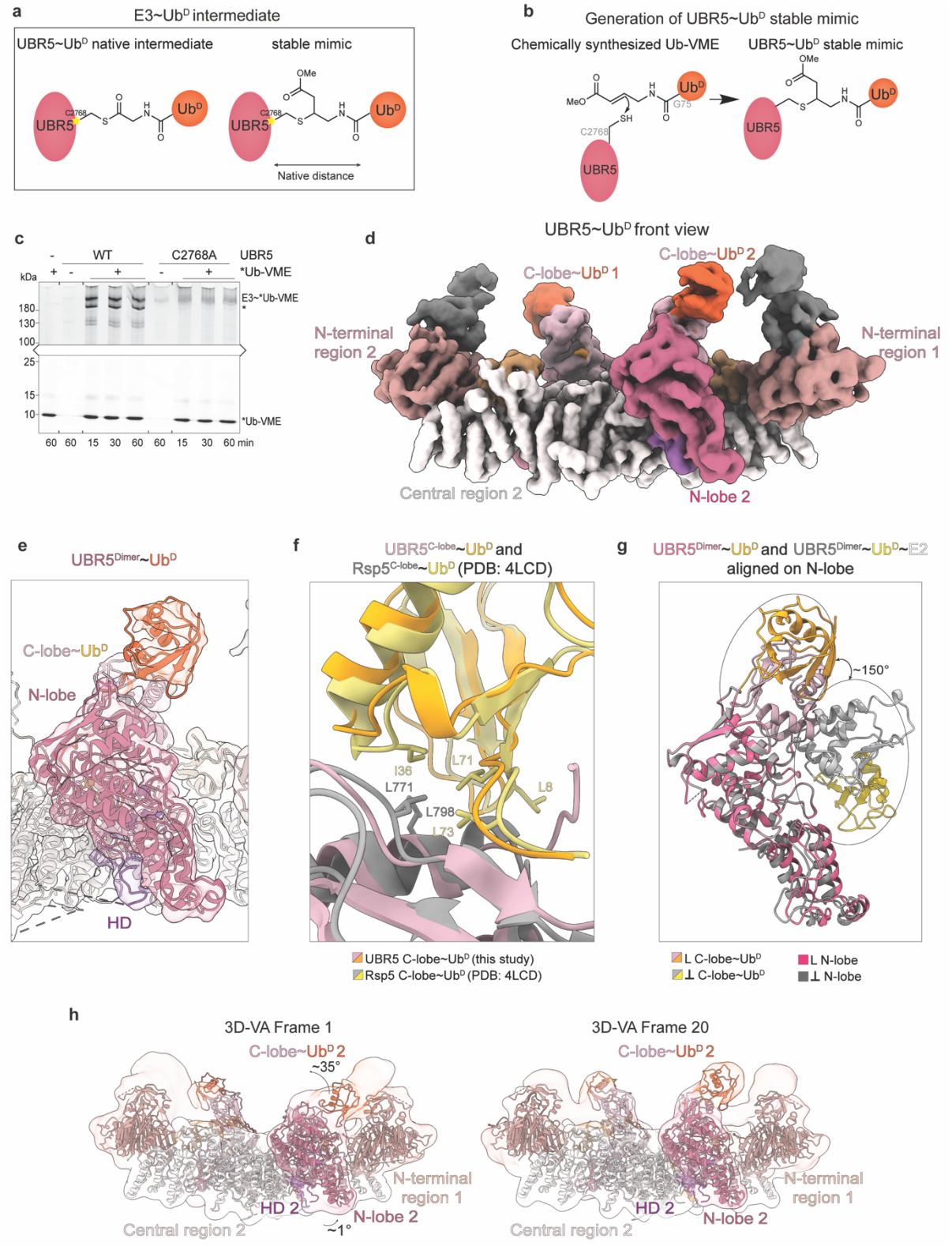


**Figure 24 Capturing UBR5~Ub$^D$ intermediate state.**

**a:** Comparison of the employed stable mimic (Ub-VME) compared to the native intermediate. **b**: Generation of the UBR5~Ub$^D$ intermediate. **c**: Probe-reactivity with UBR5 and UBR5$^{C2768A}$ was tested using a fluorescent version of Ub-VME. **d**: Density map obtained by cryo-EM of UBR5$^{Dimer}$ reacted with the activity-based probe. **e**: Zoom in to HECT~Ub$^D$ model. **f**: Comparison of modeled UBR5~Ub$^D$ interface and Rsp5~Ub$^D$ interface (PDB: 4LCD). **g**: overlay of HECT domain from TS1 (UBR5~Ub$^D$~E2) in inverted T-conformation and intermediate state (UBR5~Ub$^D$) in L-conformation. **h**: 3D-VA performed on dataset of UBR5~Ub$^D$. Left: frame 1, Right: frame 20.

## 4. UBR5-mediated polyubiquitylation

### Establishing geometric requirement for transition state 2

Despite gaining significant insights into the UBR5-mediated ubiquitylation mechanism already by studying TS1 and the ubiquitin-bound intermediate in context of full-length UBR5, we still did not yet understand how linkage-specificity is facilitated. To do so, we first asked what features the acceptor ubiquitin needs to have. We had already verified the requirement for K48 on the acceptor (**Figure 16b**), and now wondered whether the length of the K48-side chain is crucial. It had previously been reported that other systems require the precise length of the aliphatic side chain and therefore the native geometry for polyubiquitylation.[198] Using the same set of unnatural amino acids with variable side chain lengths on the acceptor's K48, we now tested whether this rule can be adapted to our system (**Figure 25a**). Indeed, UBR5 has a strong preference for native or close-to-native geometry of the acceptor lysine. If K48 of the acceptor was replaced by artificial amino acids harboring shorter aliphatic side chains with only one, two, or three methylene groups, UBR5 was not able to use this as acceptor to generate di-ubiquitin. However, if K48 on $Ub^A$ was replaced by a synthetically derived lysine with four (native length) or five methylene groups, di-ubiquitin could be formed. Knowing that UBR5 also requires native or at least close-to-native geometry of the acceptor side chain, we designed an ABP with these features (**Figure 25b**).[199]

In a two-step reaction, a BmDPA-based probe was generated (**Figure 25c**). First, $Ub^D$ with a C-terminal intein cleavage site was prepared. To facilitate native distances between the individual moieties in the final probe, $Ub^D$'s G76 was deleted. The obtained reactive group on the C-terminus was then modified with a molecule introducing two further reactive groups. The first reactive group, an α-bromo-ketone, reacted with $Ub^A$. The acceptor lysine of $Ub^A$ -K48- had been mutated to cysteine to facilitate this reaction. The absence of other cysteines in $Ub^A$ allowed to specifically form a di-ubiquitin probe with the desired linkage. Next, an activated Michael-acceptor could be used to react UBR5's catalytic cysteine with the probe while maintaining native distances between the moieties.

Originally, these ABPs were designed to investigate linkage-specificity of DUBs and to stabilize the transition state in which they cleave the ubiquitin chain.[199] We used this approach to test reactivity as well as geometry of our ABP: The ABP was incubated with a DUB that specifically binds and cleaves K48-linked ubiquitin chains (OTUB1) or a DUB that specifically binds and cleaves K63-linked ubiquitin chains (AMSH) (**Figure 25d**). Both DUBs were previously tested for their catalytic activity (**Figure 19c**). Both exhibited significant proteolytic activity, however, AMSH is considerably less active than OTUB1. Western blotting against the ABP and ABP-bound moieties then showed the reactivity of the trap: as intended, the ABP specifically reacted with OTUB1 and not at all with AMSH. This supports the geometry of our trap to closely resemble native K48-linked di-ubiquitin.

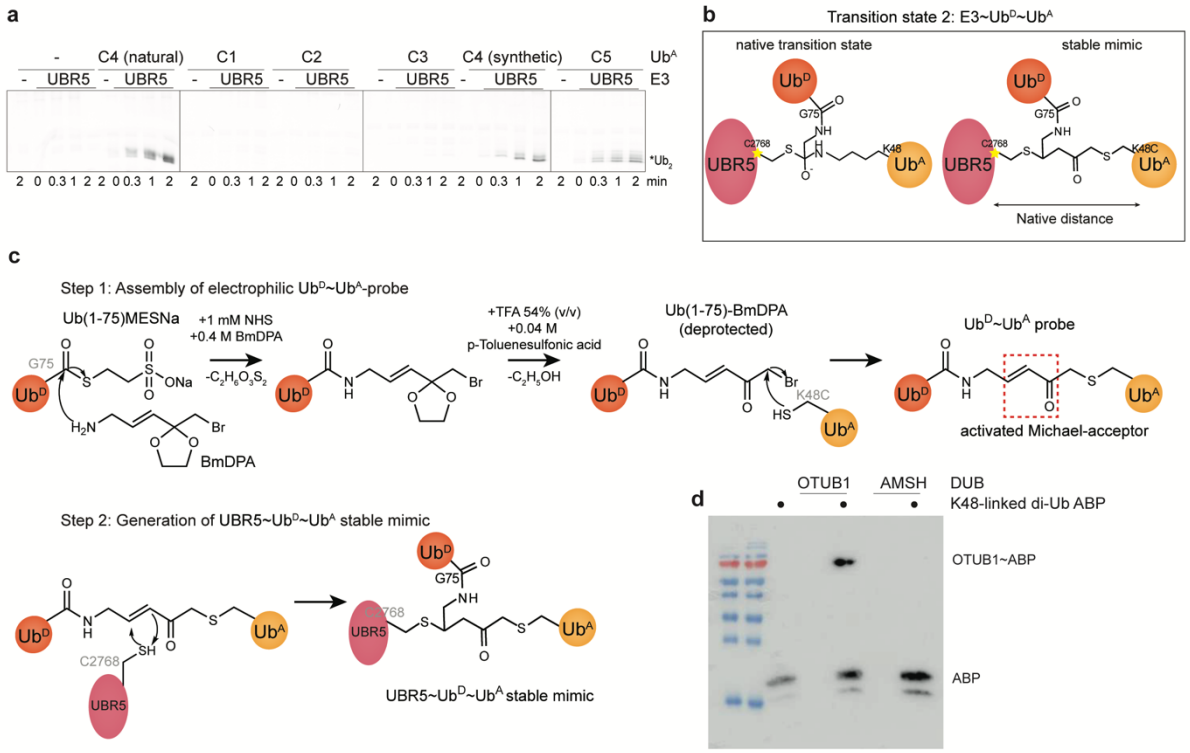


**Figure 25 Design of activity-based probe to capture transition state 2 of UBR5.**

**a**: Di-ubiquitin synthesis with synthetic acceptor ubiquitins. $Ub^{A'}$s side chain length of K48 was altered and the normal four methylene groups prior to the amino group were replaced by one to five methylene groups as indicated. **b**: Chemical structures of the transition state 2 activity-based probe and the native transition state. **c**: Two-step synthesis scheme of the transition state 2 activity-based probe. **d**: Reactivity test of the transition state 2-probe with DUB enzymes portraying different linkage-specificities. A western blot was performed and the His-tag remaining at the N-terminus of $Ub^D$ was used for visualization.

## Visualization of transition state 2 using chemical probes and cryo-EM

Next, the ABP was reacted with $UBR5^{Dimer}$ to form a stable mimic of UBR5's **T**ransition **S**tate **2** (TS2). This complex was subjected to single particle cryo-EM and initial processing on a global level already showed additional density next to the C-lobe, which could accommodate both, $Ub^D$, and $Ub^A$ (**Figure 26a**). Low resolution density next to $Ub^A$ suggested that it is furthermore bound by the UBA domain that had been invisible in all other structures of UBR5.[137-140] Focused classifications and refinements resulted in a high-resolution map of the HECT domain bound to both ubiquitins as well as the UBA domain (**Figure 26b**). This allowed building of the complex-structure or in case of the UBA domain, fitting and refining the published crystal structure with high confidence (**Figure 26c**).[147] The obtained structure revealed several interfaces contributing to catalysis (**Figure 26d**). The interactions stabilizing the HECT domain L-conformation through the HD and DSD domains seem to be consistent with those observed in the $UBR5^{Dimer}$ alone and in the UBR5~$Ub^D$ intermediate. As it had been suggested previously, the UBA domain binds $Ub^A$, thereby positioning its residue 48 at the active site. With the HECT domain in the L-

conformation, the catalytic cysteine linked to $Ub^D$ is not only adjacent to the acceptor but also situated at the junction with the N-lobe, poised to facilitate ubiquitin chain formation.

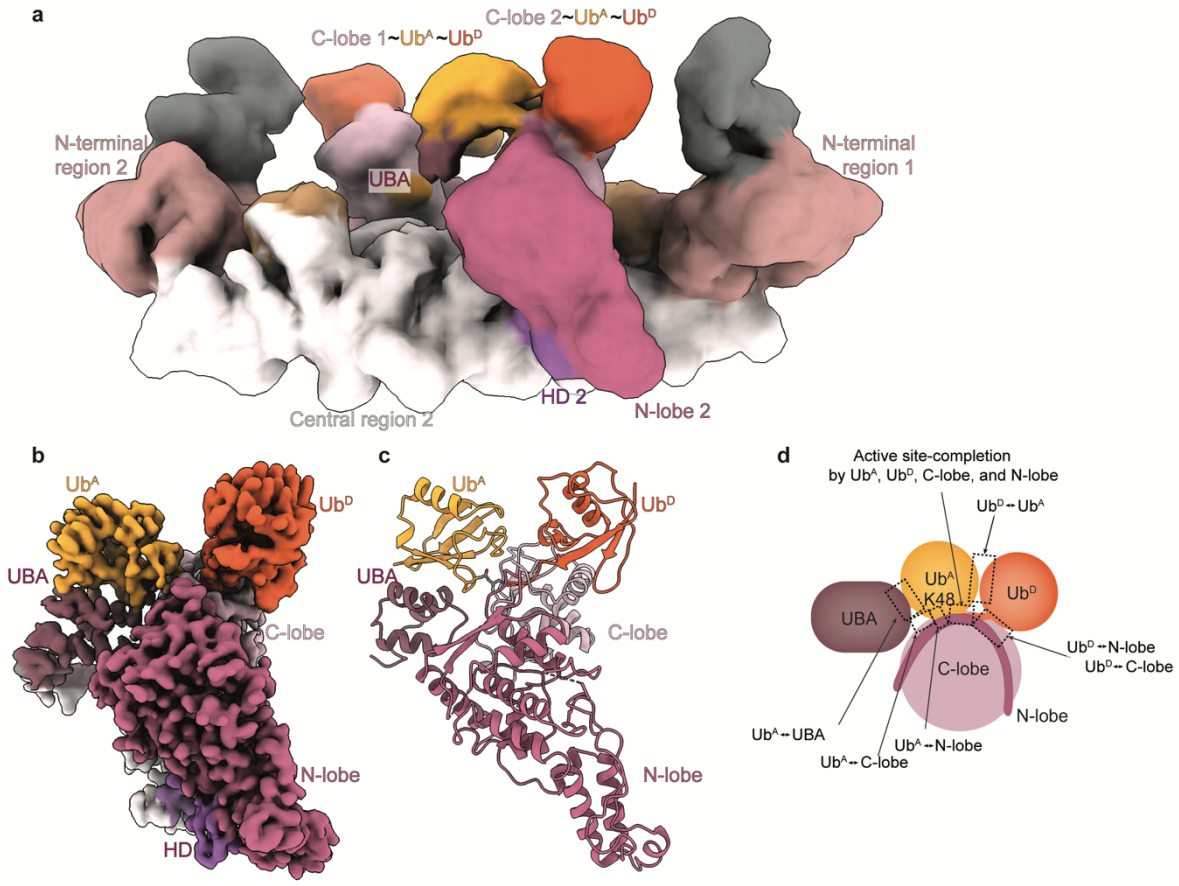


**Figure 26 Cryo-EM of UBR5~$Ub^D$~$Ub^A$.**

**a**: Initial low-resolution cryo-EM map of TS2 (UBR5~$Ub^D$~$Ub^A$). **b**: Focused-refined cryo-EM density map of TS2. **c**: Atomic model for HECT domain of UBR5 bound to $Ub^D$ and $Ub^A$, and recruitment of the UBA domain. **d**: Cartoon indicating the various observed and tested interfaces in TS2.

## $Ub^D$ positioning and active site configuration

First, in agreement with the UBR5~$Ub^D$ intermediate state, interactions between UBR5's C-lobe and $Ub^D$ position $Ub^D$ and sculpt the active site (**Figure 27**).

The map unveils the intricacies of the noncovalent interface between UBR5's C-lobe and $Ub^D$, observed across the cryo-EM maps for different intermediates along the cascade as well as prior publications of other HECT E3s (**Figure 27a**).[84,86-88] An intermolecular hydrophobic core forms between UBR5's F2732, L2762, and L2789, and A2790 and $Ub^D$'s I36, P37, L71, and L73. The hydrophobic interactions are reinforced through multiple polar interactions, involving T9, D39, Q40 of $Ub^D$, and UBR5's H2761, T2764, and K2792.

Second, the active site configuration is defined by extensive additional contacts between the HECT domain's C-terminus, $Ub^D$'s C-terminal tail, and $Ub^A$ (**Figure 27b**).

Both, $Ub^D$'s and UBR5's C-terminal tail, as well as the catalytic loop in the C-lobe wrap around each other in a 4-layered sandwich. C-terminal tails of the different HECT E3 ligases are very flexible and are most often not visible in structures.[82,84,86,87] This is reflected also in case of UBR5, where the C-terminus (2794-2799) was not visible in our apo-state but only in the polyubiquitylation-reactive state, suggesting that its flexibility is reduced upon binding of the ubiquitin-moieties. Interestingly, other recently published structures of apo-UBR5 have parts of the C-terminus built, even though the tail seemed to locate differently in one case.[138,139] This suggests that a subset of particles might have the C-terminus engaged in distinct orientations already in apo UBR5, and the transpeptidation-reactive orientation is then stabilized by binding of $Ub^D$ and $Ub^A$.

The structure showing TS2 shows parts of UBR5's C-terminus including the penultimate residue. The so-called -4 phenylalanine (F2796 for UBR5) is a conserved motif for HECT E3 ligases, which was shown to be crucial for catalytic activity of other HECT E3s.[86,88,200] In case of UBR5's TS2, the -4F nestles between the N-lobe and $Ub^D$'s C-terminus linked to the catalytic Cys. As it had been proposed before, our structure supports the -4F to assist in orienting the C-lobe with respect to the N-lobe.[86,88] Previous studies often had UBR5's C-lobe aligned to C-lobes of other HECT domains in different ways and per default, UBR5's -2F (F2798) would commonly be aligned to the -4F of other HECT domains (**Figure 8b**) due to other aligned residues. Our structure now clearly shows that UBR5's F2796 is the correct -4F, and F2798 would correspond to a -2F. The other end of UBR5's C-terminal tail -N2795- forms hydrogen bonds with an acidic loop found in the N-lobe. In contrast to N2795 and F2796 that interact with the N-lobe, the -2 phenylalanine (F2798) is intertwined between $Ub^D$ and $Ub^A$ and seems to function as a stabilizing scaffold between the two ubiquitin molecules to orient them in a certain position towards each other. Lastly, a salt bridge between $Ub^D$'s R72 and $Ub^A$'s D58 corset UBR5's penultimate F2798 thereby restricting flexibility of UBR5's C-terminus. Mutating different residues of UBR5's C-terminal tail show that indeed, the hydrophobic interactions mediated by F2796 and the size of F2798 are required for proper di-ubiquitin formation. Additionally, deletion of UBR5's ultimate residue V2799 also completely abolished di-ubiquitin formation without impairing transthiolation. Even though this residue was not resolved in our structure, we can extrapolate it's positioning based on a crystal structure that had the entire C-terminal tail of HUWE1 resolved (PDB: 6XZ1) (**Figure 27c**).[88] We hypothesize that impairment of transpeptidation upon deletion of UBR5's ultimate residue is caused by the introduction of a negative charge into the active site. Strikingly, all of these mutants did not have any defects in receiving $Ub^D$ from the E2, showing that UBR5's C-terminal tail is only required for transpeptidation, which is in agreement with prior literature.[86] We also tested the importance of the $Ub^D$-$Ub^A$ interaction (**Figure 27d**). Defects in E1-loading of ubiquitin containing R72 mutations did not allow to test this interface from the donor ubiquitin side, but we could examine it by mutating $Ub^A$.[201] We either introduced the rather conservative mutation D58A, or the repelling mutation D58R. Merely destroying the salt bridge resulted in very subtle defects in polyubiquitylation, but abolishing the close proximity of $Ub^D$ and $Ub^A$ using repelling charges, almost completely eliminates polyubiquitylation.

Overall, the organization of $Ub^D$'s C-terminus seems to be fundamental for efficient polyubiquitylation as it has major contributions in orienting UBR5's C-terminus. Restrained flexibility of $Ub^D$'s C-terminus during transpeptidation therefore is crucial to achieve transpeptidation. $Ub^D$'s C-terminus is partially fixated by noncovalent interactions with $Ub^A$, N-, and C-lobe, but it is mainly facilitated by covalent anchoring of $Ub^D$ to UBR5's active site. Subsequently to the transfer of $Ub^D$ from the C-lobe to $Ub^A$, this stabilization would be abolished and we hypothesize that this destabilizes the active site and leads to rapid release of the formed ubiquitin chain and return into the apo state.

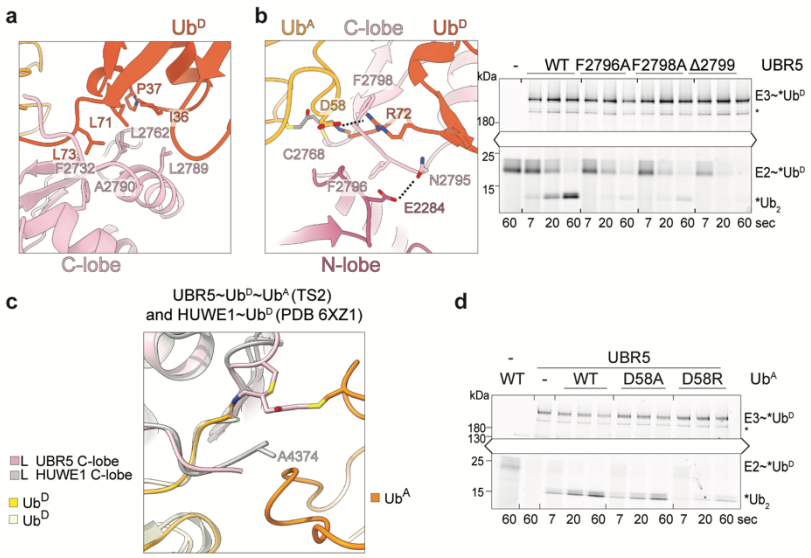


**Figure 27 Analysis of active site configuration.**

**a**: Interface between C-lobe and $Ub^D$. **b**: Intricate configuration of active site shaped by $Ub^A$, HECT domain N- and C-lobe, and $Ub^D$ (left) and di-ubiquitin synthesis assay testing the C-terminal tail of UBR5 (right). **c**: Overlay of UBR5's C-terminal tail with the published C-terminus of HUWE1 (PDB: 6XZ1). **d**: Di-ubiquitin synthesis assay testing interaction of $Ub^D$ and $Ub^A$ by mutagenesis of $Ub^A$.

## N-lobe -loop secures catalytic architecture and arranges $Ub^A$

Third, an acidic loop consisting of residues D2283-E2287 in UBR5's N-lobe, which we refer to as the **L**igation-**O**rganizing-**L**oop (LOL), secures the catalytic architecture by serving as a platform synergizing UBR5's catalytic cysteine linked to $Ub^D$'s C-terminus, as well as the $Ub^A$-conformation (**Figure 28a-b**).

This loop had been poorly resolved in our previous structures and also in some crystal structures of other HECT domains.[86,138,202,203] Upon completion of the active site by combining all required moieties, the loop is not only resolved better, suggesting reduced flexibility, but it also seems to be slightly pulled towards the active site compared to apo UBR5 (**Figure 28c**). Replacing the LOL sequence with alanines impairs polyubiquitylation without affecting transthiolation (**Figure 28a**). A similar effect had been observed previously for Rsp5. A crystal structure of Rsp5 with a donor ubiquitin as well as a substrate-peptide also failed to resolve this loop, however, alanine screening revealed that

mutating the acidic residues in this loop completely abolishes transpeptidation.[86] We tested whether the LOL is required for deprotonating the acceptor lysine to grant the nucleophilic attack of the primary amino group to $Ub^D$'s activated C-terminus (**Figure 28b**).[132,204] This was addressed by performing the chase-reaction under high pH-conditions, which would promote lysine-deprotonation. Regardless of the pH, only WT UBR5 but not the LOL mutant was able to form di-ubiquitin, demonstrating that the LOL is not directly involved in catalysis by lysine-preparation, but rather in structurally organizing the active site.

Fourth, $Ub^A$'s R54 projects toward E2287 in UBR5's LOL (**Figure 28d**).

Introducing a charge-repulsion by mutating $Ub^A$'s R54 to glutamate, severely impaired di-ubiquitin synthesis. If the charge-repulsion was lifted by mutating UBR5's E2287 to alanine or arginine, di-ubiquitin synthesis could be rescued. Interestingly, if the opposite charge-repulsion was introduced by mutating UBR5's E2287 to arginine and using WT $Ub^A$, di-ubiquitin synthesis could still take place. R54 is overall pointing towards UBR5's acidic LOL, whereas UBR5's E2287 only has R54 as opposing charge. Abrogation of di-ubiquitin-synthesis upon mutation of $Ub^A$'s R54 suggests that $Ub^A$ is held in place quite rigidly by the multitude of interactions, whereas the LOL is still somewhat flexible and can reduce the charge-repulsion if E2287 is mutated. Mutation of the entire LOL to alanine completely abrogated di-ubiquitin synthesis (**Figure 28a**), whereas mutating only E2287 to alanine did not have a significant effect (**Figure 28d, left**), but was able to rescue the charge-repulsion if combined with the $Ub^A$ R54E mutant. Overall, this shows that the LOL is required for chain formation. In return, E2287 itself is not required for chain formation but rather helps orienting $Ub^A$.

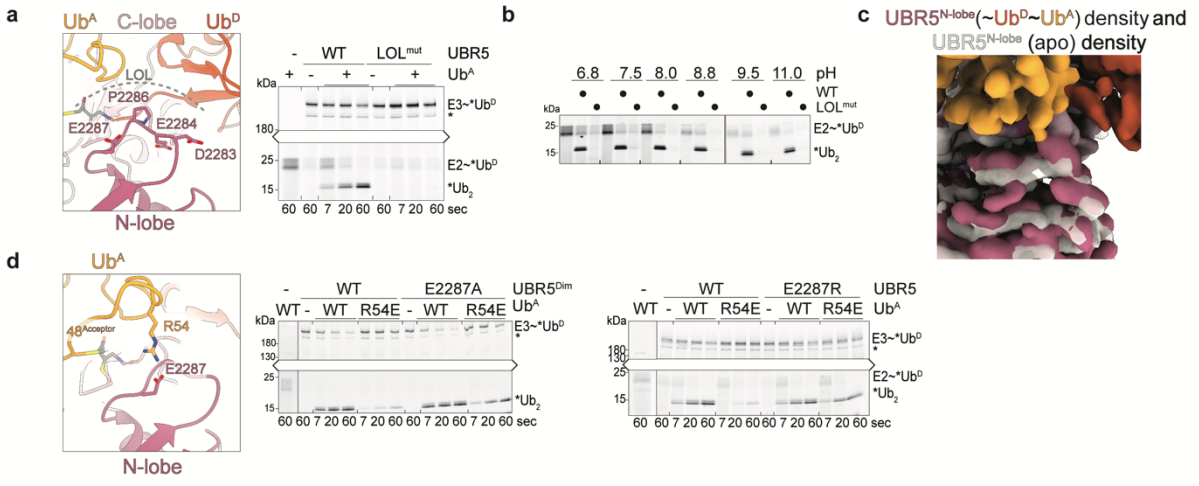


**Figure 28 Investigation of N-lobe - $Ub^A$ interface.**

**a**: N-lobes LOL reaches out to synergize $Ub^A$, $Ub^D$, and C-lobe (left). Right: di-ubiquitin synthesis assay testing the LOL mutant D2283-E2287 to alanine mutant for its ability to form di-ubiquitin. **b**: Di-ubiquitin synthesis assay with WT UBR5 or the LOL mutant with the chase-reaction performed at the indicated pH to test for lysine-deprotonation effects. **c**: Density of LOL in TS2 UBR5 (burgundy) and apo UBR5 (transparent) to visualize slight shift of LOL upon binding of the ubiquitins. **d**: Left: Zoom into part of the LOL interacting with $Ub^A$. Different combinations of E3 and $Ub^A$ mutants in di-ubiquitin synthesis-assay with $UBR5^{Dimer}$ (middle) and UBR5 (right).

### C-lobe – $Ub^A$ – interaction locks $Ub^A$ in transpeptidation-orientation

Fifth, $Ub^A$'s acceptor side chain is anchored in the active site through interactions between the C-lobe and residues adjacent to $Ub^A$'s K48 (**Figure 29**).

$Ub^A$'s A46 nests opposite Y2773 from UBR5's C-lobe (**Figure 29a**). The small size of A46 leaves a lot of space for UBR5's big Y2773. Mutating $Ub^A$'s A46 to aspartate or phenylalanine reduces the space from $Ub^A$'s side to different extents. The observed effects correlate with the introduced size-variation: The huge A46F mutation does not leave any space that could be accommodated by Y2773, whereas the slightly smaller A46D mutant still leaves a small cavity that Y2773 can fit into. Additionally, the A46D mutant offers a new interaction-possibility, which reduces the impairment caused by the size-clash: A46D and Y2773 can now form a hydrogen bond. Decreasing the size of Y2773 slightly by mutating it to phenylalanine, rescues di-ubiquitin formation in the presence of the A46F mutation on $Ub^A$ to some extent. Simply removing the hydroxyl group of Y2773 suffices to facilitate the required proximity of $Ub^A$ to the C-lobe. In return, the Y2773F mutant is not able to participate in a hydrogen bond with $Ub^A$'s A46D, which almost completely abrogates di-ubiquitin formation in this combination. We also tested effects of mutating UBR5's Y2773 to an aspartate. Combined with WT $Ub^A$, di-ubiquitin formation can be observed at a similar level than for WT UBR5. As expected, using the Y2773D mutant in combination with $Ub^A$ A46D, completely abrogates the di-ubiquitin formation because of the repelling charges. Interestingly, if the Y2773D mutant was combined with the A46F $Ub^A$ mutant, di-ubiquitin formation significantly exceeds the level compared to the combination of Y2773F UBR5 with A46D $Ub^A$ even though the same two residues are present in the interface, and are just swapped. This shows that more residues than these two are involved in the interface.

We aimed to investigate whether the observed effects of mutating $Ub^A$'s A46 residue would be conserved amongst HECT E3 ligases. To test this, we used the different $Ub^A$ mutants in a pulse-chase assay with either UBR5 or $NEDD4L^{\Delta C2}$ (**Figure 29b**). As expected, UBR5 is a lot more active and faster than NEDD4L in forming di-ubiquitin with WT $Ub^A$. Nevertheless, significant di-ubiquitin formation is also mediated by NEDD4L. If $Ub^A$ with the A46F mutation was used instead, UBR5-mediated di-ubiquitin formation was almost completely abolished while NEDD4L-mediated di-ubiquitin formation was barely affected. In contrast, UBR5 did not have issues using $Ub^A$ with the A46D mutation, while NEDD4L-mediated di-ubiquitin formation seemed to be affected more than previously. This shows that the observed effect is not a HECT-specific effect, but probably rather an effect of K48 chain formation and NEDD4L would interact with $Ub^A$ differently than UBR5.

Apart from A46, also F45 on $Ub^A$ is part of this $Ub^A$-C-lobe interface (**Figure 29a**). F45 is oriented in a parallel-displaced fashion towards Y2773, presumably undergoing weak staggered stacking. To test the importance of this interaction, we designed several mutants on $Ub^A$ and tested them in combination with the Y2773 mutations (**Figure 29c**). Generally, both, F45A and F45Y can be used as acceptor ubiquitin by UBR5 and there seem to barely be any differences if these mutations were combined with Y2773

mutations. However, we also tested an F45D mutation in $Ub^A$ and this mutation completely abrogated UBR5-mediated di-ubiquitin formation. We wondered whether this is due to a general defect of $Ub^A$ for example by impairing the overall fold. Therefore, we again tested whether this ubiquitin can be used by $NEDD4L^{\Delta C2}$ to form di-ubiquitin (**Figure 29d**). As before, the overall activity of $NEDD4L^{\Delta C2}$ is significantly worse than that of UBR5, but $NEDD4L^{\Delta C2}$-mediated di-ubiquitin formation is not at all impaired by the F45D mutation. This shows that the overall fold of ubiquitin is not affected. Very close proximity of this residue to the active site suggests that the observed effect might be caused by affecting the catalysis, rather than binding of $Ub^A$ to the C-lobe. NEDD4L is known to form K63-linked ubiquitin chains. Assuming that $Ub^A$-binding occurs to the same surface on the C-lobe, then $Ub^A$ would have to be significantly tilted to allow for the K63 side chain to approach the active site. In such an orientation, A46 as well as F45 would point away from the interface. Therefore, it seems reasonable that the effect observed for UBR5 can't be reproduced with $NEDD4L^{\Delta C2}$.

Amongst K63 chain forming HECT E3 ligases, the residue allocated at Y2773 is a highly conserved aspartate (**Figure 8**). We showed that the UBR5 Y2773D mutant generally is able to form di-ubiquitin and next we wondered whether mutating this single mutation would be sufficient to shift UBR5's linkage-specificity to form K63-linked ubiquitin chains (**Figure 29e**). As before, we employed a di-ubiquitin synthesis assay with single lysine ubiquitin mutants as $Ub^A$. As $UBR5^{WT}$, also $UBR5^{Y2773D}$ retains the K48-specificity and no di-ubiquitin is formed if only K63 is available, suggesting that all interfaces contributing to orient $Ub^A$ are required for linkage-specificity, rather than a single interface.

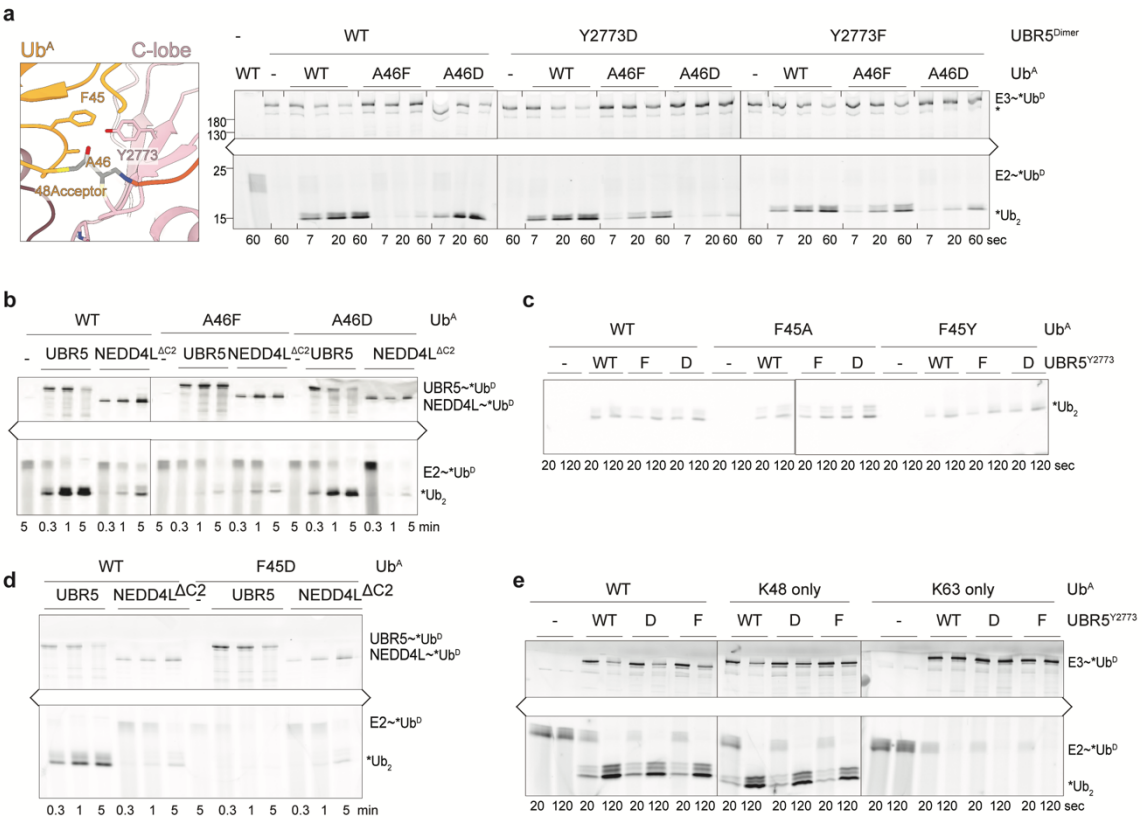

**Figure 29 Assessment of C-lobe - $Ub^A$ interface.**

**a**: Zoom into $Ub^A$ binding to C-lobe (left). Di-ubiquitin synthesis assay with different combinations of C-lobe mutants and $Ub^A$ mutants to test rescue of di-ubiquitin formation. **b**: Di-ubiquitin synthesis assay with UBR5 or NEDD4L$^{\Delta C2}$ and different mutants on the $Ub^A$ side. **c**: Di-ubiquitin synthesis assay with different combinations of C-lobe mutants and $Ub^A$ mutants to test rescue of the di-ubiquitin formation. **d**: Di-ubiquitin synthesis assay with UBR5 or NEDD4L$^{\Delta C2}$ and different $Ub^A$ mutants. **e:** Di-ubiquitin synthesis assay to assess linkage-specificity of different UBR5 C-lobe mutants in the interface with $Ub^A$.

## Recruitment of $Ub^A$ to the HECT domain by the UBA domain

Sixth, UBR5's UBA domain recruits the acceptor ubiquitin (**Figure 30**).

A prior crystal structure of UBR5's isolated UBA domain bound to ubiquitin could be fitted into the map with high confidence.[147] Despite relatively lower resolution in this area, the crystal structure allowed us to still investigate this interface in detail. The less defined density likely reflects conformational heterogeneity, presumably arising from the very long flexible tethers connecting the UBA domain to the scaffold. $Ub^A$'s I44 centered-patch nestles in between two of the three α-helices shaping the UBA domain. L8, I44, and V70 of $Ub^A$ create a hydrophobic surface, which V196, V216, L224 of the UBA domain weave into, creating an intricate hydrophobic network. This network is framed by weaker interactions of $Ub^A$'s H68 towards the backbone of helix 1 in the UBA domain.

Mutating UBR5's L224 to aspartate results in accumulation of the UBR5~$Ub^D$-intermediate and a severe defect in $Ub^D$ transfer to $Ub^A$ (**Figure 30a**). This suggests that the UBA domain indeed is not required for transthiolation and neither for stable binding of $Ub^D$ to the catalytic cysteine, but it is involved solely in transpeptidation. To exclude that the effect is caused by misfolding of UBR5, we inverted the assay and used $Ub^A$ mutations located in the UBA-$Ub^A$-interface, rather than mutating the UBA domain (**Figure 30b**). Impairing hydrophobic interactions by mutating L8/T9, I44, V70/L71 to alanine severely decreased di-ubiquitin formation, showing that not a single interaction between $Ub^A$ and the UBA domain, but the hydrophobic network is crucial for $Ub^A$ recruitment. Also interfering with the H68-mediated interaction affects di-ubiquitin formation, even though to a lesser extent than the hydrophobic interactions in the center.

Seventh, UBR5's UBA domain loosely approaches the C-lobe (**Figure 30c**).

A small loop connecting helix 1 and 2 of the UBA domain is positioned in close proximity to the HECT domain loop leading to the second β-sheet. The comparably low resolution of the UBA domain does not allow to build this loop de novo, but fitting the crystal structure suggests that G199 would be in closest proximity with the C-lobe residues P2748 and Q2747. As glycine would not participate in significant interactions, we reasoned that both, Q2747 and P2748 might make some contacts to the peptide-backbone of the UBA domain. Furthermore, we considered sterical effects arising from G199 in the UBA-loop to create a small binding pocket for Q2747. To test the different possible interactions, we introduced different mutations into UBR5$^{Dimer}$ and tested their ability for transpeptidation. Massively increasing the side chain of G199 by mutating it to tryptophane, did not impair di-ubiquitin formation at all, suggesting that binding of the UBA domain to the HECT

domain is not necessary or is not mediated by Q2747 binding to a G199-formed cavity. Mutation of P2748 to aspartate significantly reduced UBR5's ability for ligation, without impairing $Ub^D$-transfer from the E2 to the E3. Since E2 to E3 transthiolation was not affected by this mutation, it is unlikely that the mutation impaired folding of UBR5. However, the close proximity of the mutated residue to UBR5's catalytic cysteine could lead to variations in the microenvironment and might influence isopeptide-bond formation. Nonetheless, it is also possible that the effect arises from repelling forces of the negatively charged aspartate side chain with the partially negatively polarized carbonyl oxygen of the UBA backbone. We also tested whether removing the polar group of Q2747 by mutating it to leucine might affect UBR5's ligation-activity. Strikingly, this mutation almost completely abrogated di-ubiquitin formation in a pulse-chase assay. However, it has to be noted that also Ub-transfer from E2 to E3 seemed to be slower compared to WT UBR5. Again, the mutated residue is very close to the active site of UBR5, and might affect UBR5-catalysis rather than just UBA- and therefore $Ub^A$-recruitment to the HECT domain. The ambiguous results when disturbing the UBA-HECT interface that could all have several explanations, make it difficult to draw conclusions on whether the UBA domain has to bind the C-lobe or not. Since a G199W mutation did not interfere with di-ubiquitin formation, even though it should spatially impact this interface rather suggest that the UBA domain does not have to bind the C-lobe, but only catches acceptor ubiquitins to increase the local concentration. This might aid to establish the required contacts of N-lobe-$Ub^A$, C-lobe-$Ub^A$, and $Ub^D$-$Ub^A$. This could be tested upon titration of $Ub^A$. If the UBA domain is required to increase the local $Ub^A$-concentrations, high $Ub^A$-concentrations should make the UBA domain dispensable.

We furthermore aimed to understand whether the UBA domain gets recruited in cis or *trans*. For this, we analyzed cross-link MS data that was obtained as previously described (**Figure 23c, 31d**). As discussed before, most cross-links occurred within the RLD, the dSBB domains or in unstructured regions that would interrupt these. Apart from cross-links that occur because of flexibility of these unstructured regions, we identified one cross-link that could not simply be explained by small movements. This cross-link is between two unstructured loops, one of them in the RLD-dSBB-interruption (80-352) that incorporates the UBA domain and the other one in a loop extruding from the scaffold (1296-1314). The loop originating in the scaffold is a short, disordered region of <20 amino acids and therefore is restrained in its localization. However, the other cross-linked residue is positioned downstream of the UBA domain, leading back to the dSBB domain. It seems that the long stretches connecting the UBA domain with the dSBB domains make contacts with the scaffold when delivering the UBA domain towards the HECT domain (**Figure 30e**). Therefore, we suspect that the UBA domain approaches the HECT domain in cis, making some contacts with the scaffold on the way to restrain its flexibility and to facilitate guiding $Ub^A$ to the HECT domain.

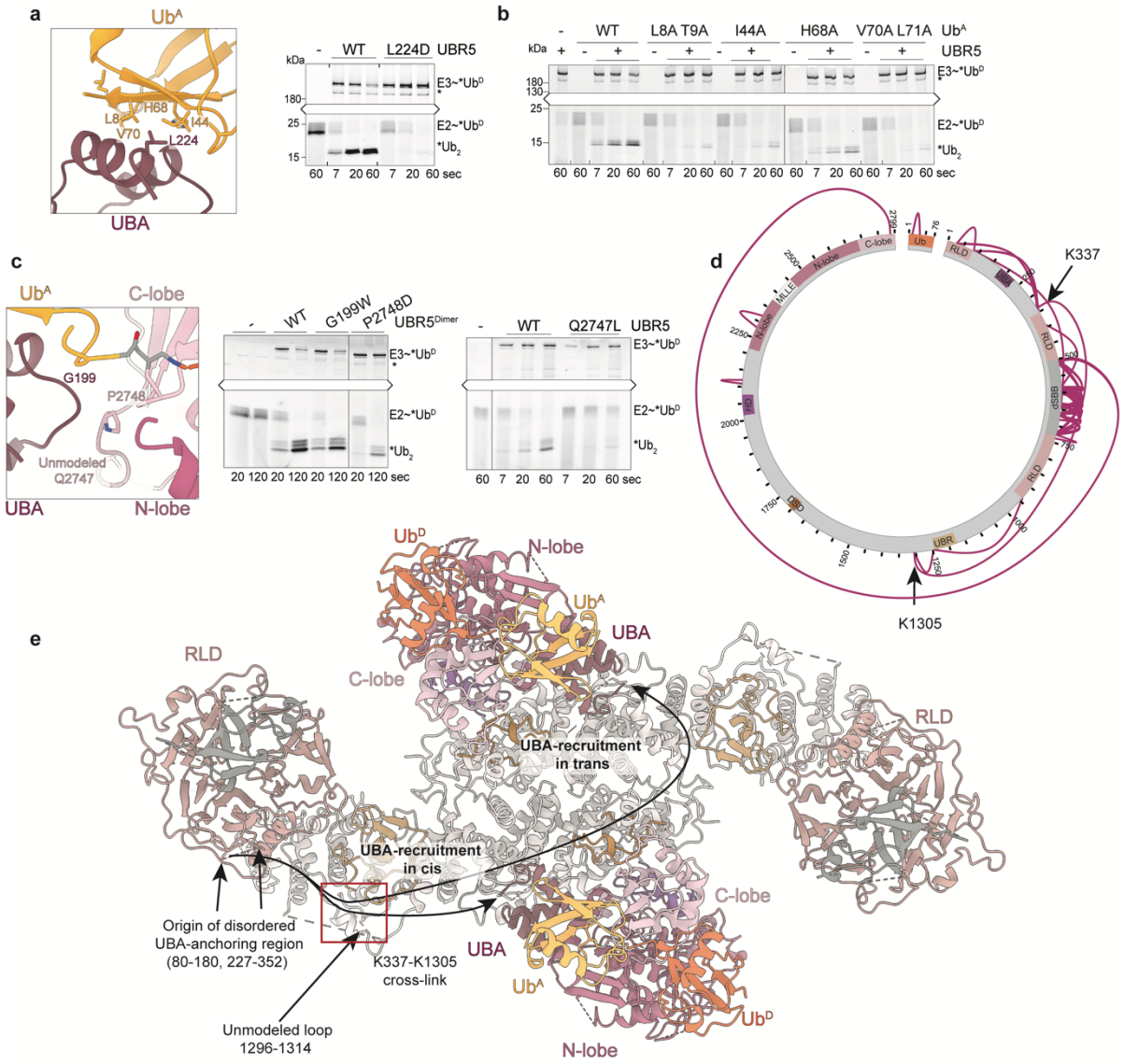


**Figure 30 Assessing the role of the UBA domain in UBR5-mediated polyubiquitylation.**

**a**: Zoom into $Ub^A$-UBA interface (left). Di-ubiquitin synthesis assay testing the UBA mutant L224D for its ability to form free di-ubiquitin (right). **b**: Di-ubiquitin synthesis assay testing various $Ub^A$ mutants located in the UBA-$Ub^A$ interface together with wildtype UBR5. **c**: Zoom into potential C-lobe – UBA interface (left). Di-ubiquitin synthesis assay testing the UBA mutation G199W and C-lobe mutations P2748D and Q2747L located at the supposed UBA-C-lobe interface with dimeric UBR5 (middle) or wildtype UBR5 (right). **d**: Cross-link mass-spectrometry results of UBR5 with one cross-link highlighted. **e**: Model of UBA recruitment to the HECT domain in cis or trans considering cross-link MS result, showing $UBR5^{Dimer}$ with $Ub^A$, $Ub^D$, and UBA domain in bottom view.

## Heterogeneous ubiquitin chain formation

Our structural studies show how the acceptor ubiquitin is avidly bound to UBR5. Despite this avid binding, most of $Ub^{A'}$s lysines are exposed, as well as the C- and N-terminus (**Figure 31a**). This would enable UBR5 to bind not only to free ubiquitin, but also to a pre-assembled ubiquitin chain, thereby facilitating extension or branching of such chains. To test whether UBR5 could modify more than just the previously reported K11- and K63-preassembled chains, we tested its ability to modify the differently linked di-ubiquitins.

Strikingly, both, UBR5 and dimeric UBR5 were able to modify either of the different di-ubiquitins substantially. Even though the different di-ubiquitins -except for the K48-linked di-ubiquitin- have two potential acceptor lysines (K48) available, we almost exclusively saw formation of tri-Ub, but not tetra-Ub. This suggested that UBR5 preferentially modifies one of the two acceptor ubiquitin moieties, thereby having a preference of forming either mixed (if the distal Ub moiety is preferred), or branched (if an internal of proximal Ub moiety is preferred) ubiquitin chains.

To test UBR5's preference for the acceptor ubiquitin position within a preassembled ubiquitin chain, we generated M1-linked tri-ubiquitin by linear fusion, harboring K48R mutations on distinct ubiquitin-moieties within the chain (**Figure 31b**). Ubiquitins that were colored green contain WT K48, red ubiquitins have the K48R mutation. Employing pulse-chase assays, we tested whether UBR5 has a preference for either position by determining whether mutating K48 on a specific ubiquitin impaired modification. Indeed, if only one ubiquitin within the tri-ubiquitin contained the K48R mutation, the biggest effect could be observed for the proximal ubiquitin. Vice versa, if two of the three ubiquitins contained this mutation, the labeling-efficiency was the highest if only the proximal ubiquitin exhibits K48. This shows that UBR5 much rather modifies the proximal ubiquitin within a preassembled chain and therefore rather forms branched chains than mixed chains.

With this in hand, we wondered whether UBR5 also has a preference regarding the ubiquitin-moiety the UBA domain binds. It has been established previously that UBA-$Ub^A$-binding occurs via $Ub^A$'s I44-centered patch.[147] Therefore, we used a similar assay as before (**Figure 31b**), but instead of mutating K48, we introduced an I44D mutation to the respective ubiquitin-moieties (**Figure 31c)**. As before, mutating only the proximal ubiquitin had the biggest effect on UBR5-mediated ubiquitylation of the chain. This shows that the UBA domain preferentially binds to the ubiquitin that directly gets targeted within a ubiquitin chain, rather than just bringing the ubiquitin chain into proximity of the active site. The UBA domain therefore does not seem to only increase the local $Ub^A$-concentration, but it also seems to orient or stabilize an acceptor ubiquitin of a specific chain position. We also tested the opposite effect: If two of the three ubiquitin-moieties have a mutated I44, can UBR5 still capture the ubiquitin chain efficiently and facilitate its modification? Strikingly, if only one "intact" UBA-binding site persisted in the ubiquitin chain, UBR5-mediated modification was almost completely abolished, regardless of which site was retained natively. Avid binding of UBA to $Ub^A$ might explain this effect. However, the UBA domain only has one canonical Ub-binding site and therefore presumably binds ubiquitin in a stoichiometric manner. The other option is that the UBA domain has two roles: It captures $Ub^A$ in solution and it directs it towards the HECT domain. This interaction can occur on either of the ubiquitin moieties and the presence of many intact UBA-binding sites on a Ub chain increases the chances of being captured by the UBA domain and brought to the HECT domain. The second function is that the UBA domain then precisely binds the ubiquitin moiety of interest to enable modification. Since the architecture of

different ubiquitin chains varies massively, this effect would have to be verified for other pre-assembled ubiquitin chains as well.

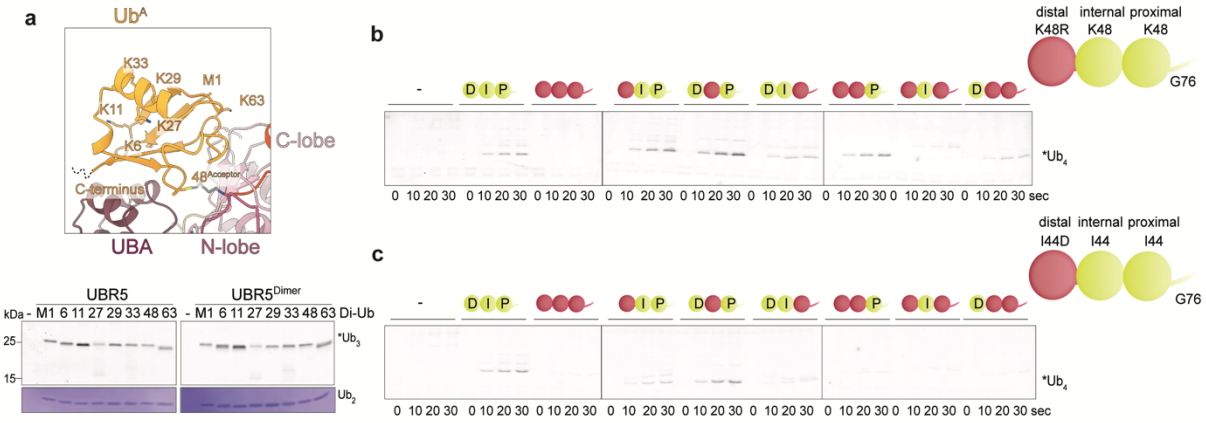


**Figure 31 Testing UBR5-mediated formation of heterotypic ubiquitin chains**.

**a**: Zoom into UBR5-bound $Ub^A$ with lysines highlighted (above). Tri-ubiquitin synthesis assay with all differently linked di-ubiquitins added as acceptor (below). **b**: Pulse-chase assay with M1-linked tri-ubiquitin used as acceptor to test preferred site of modification by UBR5. Green ubiquitins are wildtype ubiquitins, red ubiquitins harbor a K48R mutation and can't be modified by UBR5. **c**: Pulse-chase assay with M1-linked tri-ubiquitin used as acceptor to test preferred site of binding by UBR5's UBA domain. Green ubiquitins are wildtype ubiquitins and red ubiquitins harbor an I44D mutation, which can't be bound by UBR5's UBA domain.

## 5. UBR5-mediated substrate ubiquitylation

### Modification of MYC-peptide

Since we now understand the fundamentals of UBR5-mediated polyubiquitylation, we also sought to investigate this directly on a substrate rather than just free ubiquitin chains. Another study that aimed to ubiquitylate the UBR5-substrate AKIRIN2 *in vitro*, already pointed out that UBR5 much rather forms ubiquitin chains than it primes substrate.[138] We aimed to verify this with our reconstituted UBR5 and a substrate that has been identified several years ago: the transcription factor MYC.[163,164] One of the studies originally identifying it, also proposed a MYC-peptide with two lysines on the C-terminus that would get modified by UBR5 and a mutant peptide, that would not get modified by UBR5.[164] Controversially, the UBR5-version that had been employed in this study to test UBR5-mediated ubiquitylation consisted of only truncated HECT domain lacking crucial elements of the N-lobe structurally located close to the C-lobe, making it unlikely to be catalytically active. Nevertheless, we tested the fluorescently-labeled MYC-peptide as well as fluorescently-labeled peptide harboring the respective mutations in the presence of full-length UBR5. Additionally, we fused both of these peptides to the N-terminus of ubiquitin to mimic the substrate-peptides in their "primed" state (**Figure 32a**). In this case, UBR5 would only have to polyubiquitylate the substrates and not modify them directly on one of their lysines. We aimed to see whether fusion of the peptide to ubiquitin makes it a better substrate than just free ubiquitin, and whether UBR5 also manages to modify the peptide itself. To test this, we employed a pulse-chase assay to prevent use of the ubiquitin-moiety as donor (**Figure 32a**). Both peptide-ubiquitin fusions, as well as free ubiquitin serving as acceptor could get modified comparably well in our assay. If instead only the fluorescent peptides were used as acceptor, no modification could be observed suggesting that UBR5 indeed much rather ubiquitylates the fused ubiquitin than the peptide-moiety. Based on our assay it is not possible to judge whether fusing the peptide to ubiquitin enhances its modification, or whether there is no effect at all. Measuring the kinetics of the different substrates could provide more insight and allow proper evaluation. If instead of WT UBR5, the UBA mutant L224D containing UBR5 was used as E3, modification of free ubiquitin as well as the peptide-ubiquitin fusions was abolished. This demonstrates that the previously observed modification of the peptide-ubiquitin fusions is mediated by the UBA domain recruiting $Ub^A$.

Even though the mechanism of UBR5-mediated MYC-regulation has been further investigated,[140] it is not completely clear yet, where MYC would bind UBR5 and how this could facilitate substrate-ubiquitylation. It is noteworthy however, that substrate-peptides used in the recent study differ from the previously employed peptides, that we also used in this study. It will therefore have to be assessed again, whether use of the recently reported peptides would give different results.[140]

## Modification of N-degron substrates

We tried to assess substrate-ubiquitylation also in another way: We used an artificial substrate-peptide with a known binding site on UBR5, the UBR domain. However, to date no substrate has been found for UBR5 yet, that would get recognized via the N-degron pathway by an N-terminal arginine. To support UBR5 during the substrate-modification, we used primed substrate-peptides (**Figure 32b**). These consisted of an N-terminal arginine, then differently long extensions of the peptide used in the original study showing recognition of N-terminal arginines by UBR5[150], and for the long constructs additional glycine-serine linkers were included upstream of a ubiquitin-fusion. In a pulse-chase format, we evaluated whether these differently long peptides could be modified with labeled donor ubiquitin. Additionally, we aimed to understand whether this modification would be dependent on the UBA domain, as it is for free ubiquitin chain formation. Importantly, we do not know whether or to what extend the observed density in the UBR-binding cleft (**Figure 18e**) influences our attempt of modifying the peptide-substrates. The ubiquitin-fusion with the shortest N-terminal extension – just a single arginine- can be modified very well by UBR5 and seems to be modified better even than the following construct with an elongated extension of five additional amino acids (as can be observed in the short 7 second time point). This seems to be true also for the subsequent construct with 12 amino acids between the N-terminal arginine and the ubiquitin-fusion. This might be caused by the increased flexibility of $Ub^A$'s N-terminal extension causing instable and impaired $Ub^A$-binding to UBR5. In contrast, the next two constructs, especially the one with 35 amino acids between the N-terminal arginine and the fused ubiquitin, show a slightly higher level of modification again, especially visible in the first time point. If the previously observed effect was caused by destabilized binding due to the N-terminal extensions, the long extensions should impair modification even more. If we analyze our obtained structure of polyubiquitylating UBR5, we see that the distance between an arginine bound to the binding pocket (modeled by comparison with peptide-bound crystal structure of UBR2's UBR-box PDB: 3NY3)[184] (**Figure 18e**) to the N-terminus of $Ub^A$ bound in cis or trans is roughly 70 Å (**Figure 32c**). Therefore, a linker of minimum 24 amino acids between an arginine and a fused ubiquitin would be required so that both, arginine and ubiquitin can be bound simultaneously to the UBR and the HECT domain respectively, to increase affinity. This feature is provided partially in the construct with 25 amino acids in total connecting the arginine and the ubiquitin and is certainly provided in the longest construct. We performed the same assay with UBR5's UBA mutant L224D to test whether the UBA domain is absolutely crucial for modification of these constructs. As expected, the UBA mutant is severely impaired at modifying these substrate-peptide-fusions and we significantly had to increase the contrast to see any signal corresponding to formed product at all. Baseline amounts of the three short constructs could be modified within 60 seconds, but were not modified in the absence of UBR5. Substrate-modification was slightly increased for the UBA mutant if the two long constructs were used, which facilitate arginine-binding to the UBR domain as well as binding of $Ub^A$. We hypothesize that tethering the acceptor ubiquitin to UBR5 by fusing it to an adequate substrate or substrate-

peptide, decreases the requirement for the UBA domain to recruit acceptor ubiquitin to the HECT domain as $Ub^A$ already is stabilized in proximity to the HECT domain. This is in agreement with another study showing that substrates can be polyubiquitylated by UBR5 - even though to a lesser extent - without a functional UBA domain.[138]

Within our study we determined why UBR5 is so good at forming free polyubiquitin chains. However, we struggled to achieve robust ubiquitylation of a substrate. Considering our TS2-structure of UBR5 we can also reason why that is the case: UBR5 binds $Ub^A$ very avidly via multiple interfaces to facilitate efficient and precise polyubiquitylation. This shows that UBR5 is destined to bind an acceptor ubiquitin and form chains, and in order for a substrate to compete with free acceptor ubiquitin, it will need an intricate binding-mode. Such an intricate binding could be facilitated by UBR5's oligomerization or other factors that have yet to be determined.

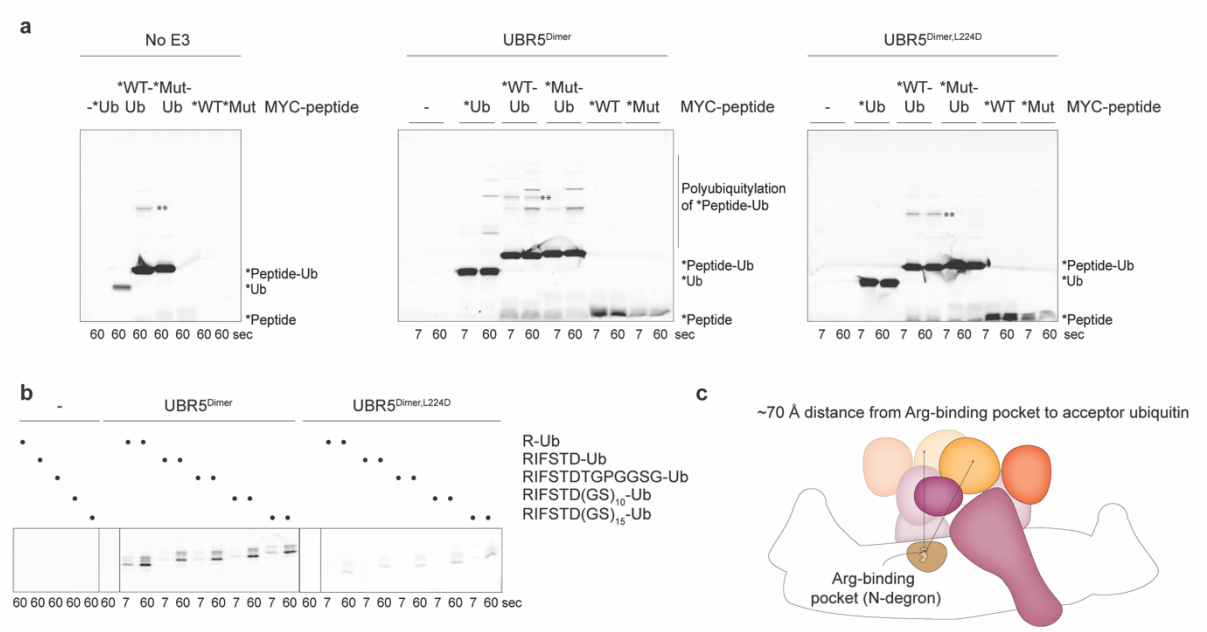


**Figure 32 Testing UBR5-mediated substrate ubiquitylation.**

**a**: Pulse-chase assay testing UBR5-mediated ubiquitylation of a MYC-peptide, a MYC mutant peptide, and ubiquitin fusions of both peptides. The substrates were fluorescently labeled and unlabeled wildtype ubiquitin was used as donor. Requirement of UBR5's UBA domain for this modification was tested (right). **b**: Substrate-ubiquitylation assay in pulse-chase format with dimeric UBR5 or dimeric UBR5 harboring a UBA domain mutation and differently long peptides with ubiquitin fusions targeting UBR5's UBR domain. **c**: Model illustrating the UBR domain as substrate recognition domain and the distance to the HECT domain for polyubiquitylation is indicated.

## 6. Targeting of UBR5 using UbVs

### Effect of UbVs on activity of truncated HECT domain of UBR5

The huge medical importance of UBR5 especially in various cancer types (**Figure 9a**), make it an important target for drug discovery as well as developing tools to further investigate it. One approach could be to use UBR5's polyubiquitylation activity to induce degradation of target proteins in cells. This would be beneficial considering the increased abundance of UBR5 in some diseases as it could translate into rapid degradation of target proteins as well as a potential tissue-specificity for disease tissue. However, if an increased UBR5 abundance directly or indirectly causes the disease, it might be more advantageous to impair UBR5's activity rather than redirect it. A tool to alter the activity of HECT E3 ligases that has been published several years ago are **Ub**iquitin **V**ariants (UbVs).[127] In a high-throughput screen, a huge variety of point mutations was introduced into ubiquitin, which was then selected for increased binding affinity to purified immobilized HECT domains.[127] In another selection round, the tight binders were tested for their specificity towards a specific HECT E3. For UBR5, two UbVs were identified that bind to UBR5 with high affinity. We aimed to test binding of these UbVs and assess their effect on UBR5's activity.

The two UbVs tested are termed UbV3 and UbV7 and have different interfaces mutated compared to WT ubiquitin (**Figure 33a**).

Identification of the two UbVs was performed using the last ~400 amino acids of UBR5 (**Figure 33b**). This area consists of the MLLE domain and parts of the HECT domain, however some N-terminal parts -α-helix 1-3 as well as β-sheet 1 & 2- are omitted in this construct (**Figure 33c**). Design of this construct was presumably inspired by prior misleading domain annotations of UBR5. In our characterization attempts, we therefore used the same construct (MLLE-HECT$^{\Delta N}$), a construct additionally omitting the MLLE domain (HECT$^{\Delta N}$), a construct only covering a part of the N-lobe$^{\Delta N}$, the entire C-lobe, and a construct covering the entirety of UBR5.

First, we aimed to verify binding of the UbVs in our system (**Figure 33d**). As expected, UbV3 can bind to the MLLE-HECT$^{\Delta N}$ construct. The binding site of the UbV could furthermore be narrowed by testing co-pull downs of UbV3 with various shorter constructs: Pulling on HECT$^{\Delta N}$ as well as the N-lobe$^{\Delta N}$ also led to enrichment of UbV3. Using only the isolated C-lobe in contrast, did not result in significant UbV3-binding, suggesting that the binding site is indeed found in the N-lobe. In return, pulling on UbV3 resulted in enrichment of MLLE-HECT$^{\Delta N}$, HECT$^{\Delta N}$, or N-lobe$^{\Delta N}$ but not in enrichment of the C-lobe. We tested the same for UbV7, but could not see significant enrichment for any of the constructs, suggesting weaker binding of this UbV to the HECT domain.

Since UbV3 seemed to be more promising, subsequent analyses focused on UbV3. Testing the effect of the UbV on UBR5 activity was performed initially with the HECT$^{\Delta N}$

construct since this seemed to be a minimal construct still showing binding and containing the E2 binding site as well as the C-lobe to allow transthiolation (**Figure 33e**). We tested whether $Ub^D$-transfer from the E2 to UBR5's catalytic cysteine is impaired in the presence of the UbVs. The truncated HECT domain alone can almost completely discharge E2~$Ub^D$ within our time course, however, the presence of UbV3 almost completely abolishes transthiolation. If instead UbV7 was added to the chase-reaction instead of UbV3, transthiolation was nearly unaffected. This could be expected in our system based on the supposedly weak binding of UbV7 to the $HECT^{\Delta N}$-construct (**Figure 33d**).

## Effect of UbVs on activity of full-length UBR5

If instead of the $HECT^{\Delta N}$-construct, full-length UBR5 was used in an assay, a similar effect could be observed: Without any UbV, or in presence of UbV7, E2~$Ub^D$ could be discharged and high molecular weight species appeared, indicating long ubiquitin chains (**Figure 33f**). However, if UbV3 was added to the chase-reaction, E2~$Ub^D$ discharge was severely impaired. One striking difference to the previous result was, however, the significant formation of a species that seems to correlate to a UbV3-*Ub "di-ubiquitin". This suggests UbV3 to serve as acceptor for UBR5, rather than just inhibiting UBR5's activity in general. We aimed to test whether we need an excess of UbV3 compared to the E3 construct. If the UbV would bind very tightly to UBR5, the effect it has on the activity should be maintained even in lower concentrations. Reduction of the UbV3 concentration but maintaining at least 5x molar excess of UbV3 over UBR5 resulted in partial recovery of the transthiolation reaction, suggesting that the UbV is inhibiting UBR5 partially also through crowding effects (**Figure 33g**).

We know that UBR5 requires both, its HECT domain as well as the UBA domain to form free ubiquitin chains (**Figure 30a-b**). Additionally, we know that the UBA domain interacts with an acceptor ubiquitin via ubiquitin's I44 patch. To test whether the observed UbV3-Ub formation mediated by UBR5 occurs canonically as would a normal di-ubiquitin formation, we disrupted potential interactions of the UbV and the UBA domain by introducing an I44D mutation into UbV3 (**Figure 33h**). This variation of UbV3 was then tested for its ability to impair UBR5-mediated E2~$Ub^D$ discharge. Interestingly, E2~$Ub^D$ discharge was less impaired compared to the original UbV3, which can be seen by the remaining E2~$Ub^D$ after two minutes. However, discharge without any UbVs present still exceeds. Even though E2-discharge is more efficient in the presence of $UbV3^{I44D}$ compared to UbV3, modification of the UbV is better with the original UbV compared to $UbV3^{I44D}$. This shows that $UbV3^{I44D}$ is a worse inhibitor or modulator than UbV3 itself. Noteworthy, UbVs that have been shown to bind to the exosite, present in the HECT domains of several members of the NEDD4-family, have been found to also bind there via their I44-patch.[127] No exosite has been identified for UBR5 so far, however, we can not exclude an exosite-like feature to also be present in UBR5's HECT domain.

Therefore, we wondered whether the altered effect of UbV3$^{I44D}$ compared to UbV3 really originated from impairment of its interaction with the UBA domain, or maybe from another site the UbV could bind to via its I44-patch. To test this, we used the original UbV3 but combined it with a UBA mutant of UBR5, which is deficient in free chain formation and was shown to be impaired in binding ubiquitin.[147] As seen before (**Figure 30a**), WT UBR5$^{Dimer}$ can form free ubiquitin chains, but the L224D mutant can't. Instead, the L224D mutant shows enrichment of UBR5~Ub$^{D}$ intermediates because transthiolation is not impacted but transpeptidation is. If we add UbV3 to both of these samples, E2~Ub$^{D}$ discharge is impaired even though not abolished (**Figure 33i**). In case of the L224D mutant, the E2-discharge seems to work a little bit better than for WT UBR5. Additionally, UbV3-Ub formation is significantly enhanced for WT UBR5 compared to the L224D mutant. Noteworthy, in absence of any UbV, the L224D mutant fails to form any free chains, however, in the presence of UbV3, significant amounts of UbV3-Ub can be formed.

Overall, our data suggests that UbV3, which had been identified as a UBR5 HECT domain binder might have a much more interesting mode of action. Based on our mutagenesis studies as well as binding experiments, it seems that the UbV binds in a way that is somewhat similar to a canonical acceptor ubiquitin, to the UBA domain and the HECT domain. The UbV seems to function as promising modulator especially if binding to the HECT domain and the UBA domain are facilitated, which becomes apparent in the mutagenesis assays interfering with its interaction to the UBA domain.

Based on this, we propose a model for UbV3 binding and its modulating effect on UBR5:

UbV3 binds to the UBA domain via its I44 patch, that slightly differs from the I44 patch of canonical ubiquitin. The interfaces of acceptor ubiquitin to N- or C-lobe that we identified previously (**Figure 26d**) are not altered significantly in UbV3 compared to canonical ubiquitin, insinuating a slightly different binding mode to the HECT domain. If UBR5 already was charged with Ub$^{D}$, this ubiquitin can then be transferred to the UbV. Unlike canonical acceptor ubiquitin, the UbV remains bound to UBR5 after being modified itself. By staying bound to UBR5 and somewhat complexing the HECT and the UBA domain, the UbV prevents the required movement of the HECT domain to undergo another transthiolation reaction. If the interactions of the UbV to the UBA domain are weakened, this complexation does not work and the HECT domain can undergo its movement for another round of transthiolation. In return, without the UbV-UBA interaction, the UbV can't be oriented as efficiently to facilitate its modification.

This model would explain the effects we saw in our assays, however, many aspects of this model remain unclear and will have to be evaluated in subsequent studies:

First, it will have to be tested whether the UbV indeed binds to the acceptor ubiquitin binding site, and if so, whether it does that with higher affinity than canonical ubiquitin. This can be achieved by performing competition assays to see whether the UbV can outcompete ubiquitin.

Second, it would be interesting to see whether the UbV preferentially binds to UBR5~Ub$^{D}$. This would be supported by our assays. We see that the UbV gets modified with donor

ubiquitin and that it significantly slows down transthiolation. If the affinity of the UbV to UBR5 is increased by $Ub^D$-binding, then UBR5 could undergo a first round of transthiolation and only afterwards the UbV would really start to impair the function. This could be addressed by analyzing the activity over a longer time course. If there is a strong decrease of E2~$Ub^D$ concentration in the beginning but later on, E2~$Ub^D$ is stabilized, this might suggest that the UbV-effect only comes to play with a UBR5~$Ub^D$ intermediate.

Third, it would be very interesting to see where the UbV gets modified. UBR5 is a highly specific K48 chain former and we showed that this specificity is achieved by a multitude of interfaces to avidly bind and orient $Ub^A$. In case of UbV3, most of these interfaces seem to remain untouched and therefore, this avid binding mode could persist and K48-specificity could be retained. However, if the UbV binds significantly tighter to UBR5 than canonical ubiquitin, this indicates involvement of additional interfaces that were not identified yet.

Additionally, in our assay we saw that the UBA mutant of UBR5 can modify the UbV with $Ub^D$. This suggests that the requirement of the UBA domain could be overcome with high affinities of an acceptor ubiquitin to UBR5 or with very high ubiquitin concentrations. It will be very informative to test whether transpeptidation with $UBR5^{L224D}$ could also occur if huge accesses of canonical ubiquitin were present instead of the UbV.

Overall, using UbV3 to modulate UBR5's activity might be an interesting approach for future studies. So far, it is only mediocre efficient as high concentrations of the UbV are required in order to significantly slow down E2~$Ub^D$ discharge. However, it might be a good foundation for further mutagenesis that might enhance the affinity even more and serve as a more potent modulator. If the UbV would be a preferred acceptor compared to ubiquitin, it could be used not necessarily to inhibit E2~$Ub^D$ discharge, but rather to redirect chain formation to UbV-Ub "chain" formation rather than substrate polyubiquitylation.

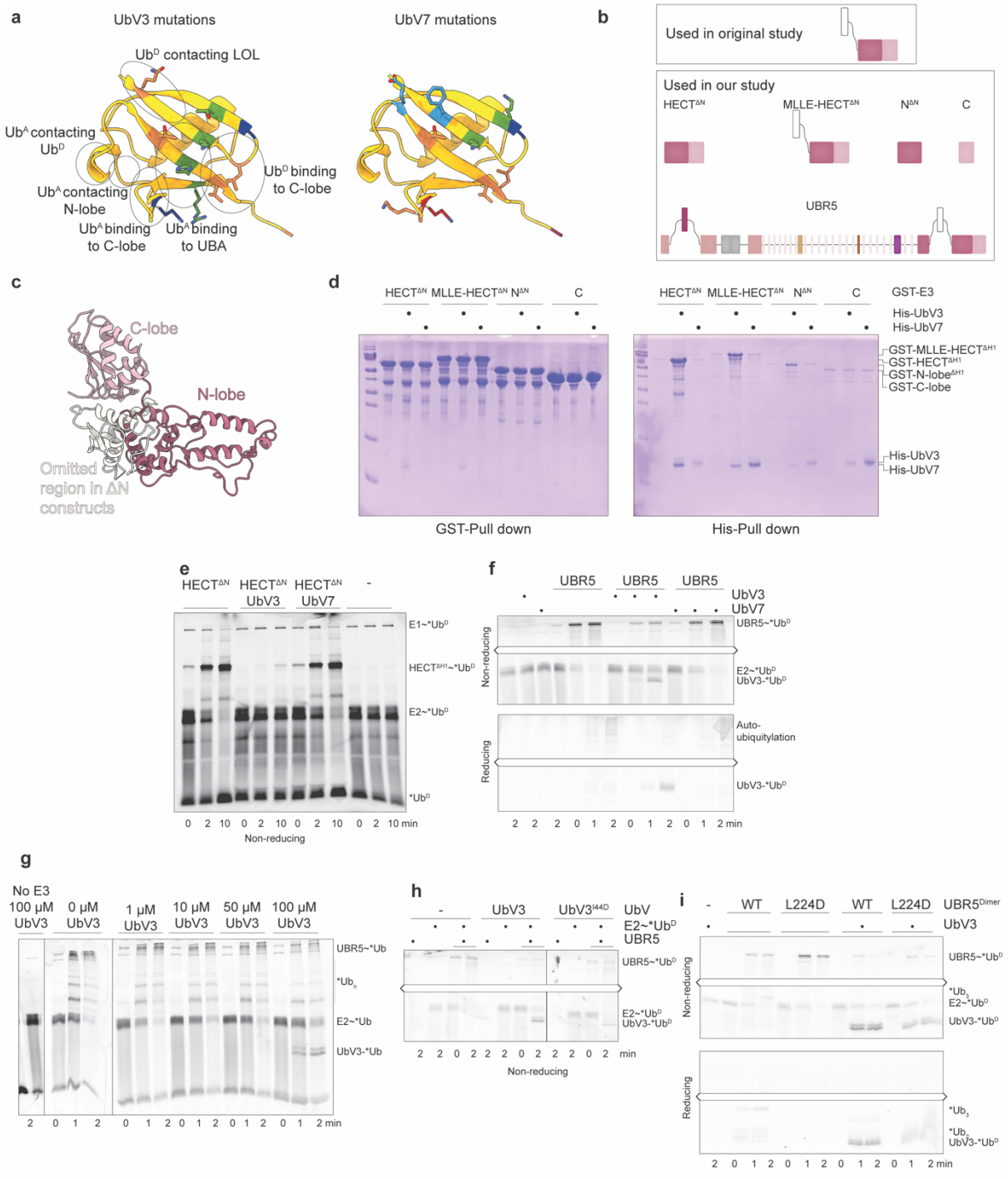


**Figure 33 Testing ubiquitin variants for UBR5.**

**a**: Structure of ubiquitin (PDB: 1UBQ) with previously established interaction sites during UBR5-mediated polyubiquitylation marked. Mutations of UbV3 (left) and UbV7 (right) compared to wildtype ubiquitin are shown. The side chains of wildtype ubiquitin are shown and alterations of the amino acid property are indicated by color coding: amino acids that are acidic in the UbV but not acidic in wildtype ubiquitin are shaded red, amino acids that are mutated to be basic in the UbVs are dark blue, hydrophilic amino acids are light blue, hydrophobic amino acids are shaded green, amino acids that are significantly larger in the UbVs are shaded orange. **b**: UBR5-construct used in the original study identifying the two UbVs (above) and constructs used in this study (below). **c**: Illustration of UBR5's HECT domain. The N-lobe region that is omitted in the original study and in this study in ΔN-constructs is shaded white. **d**: Co pull down experiments to identify interacting regions of UBR5 with UbV3 and UbV7. GST-tagged UBR5-truncations and His-tagged UbV-constructs were used. Left: Coomassie-stained SDS-PAGE of GST-pull down, right: Coomassie-stained SDS-PAGE of His-pull down. **e**: Pulse-chase assay with fluorescently labeled

wildtype ubiquitin with HECT$^{\Delta N}$ in the presence of UbV3 or UbV7 respectively. **f**: Pulse-chase assay with fluorescently labeled wildtype ubiquitin with UBR5 in the presence of UbV3 or UbV7 respectively. **g:** Pulse-chase assay titrating UbV3 to UBR5 in the chase-reaction. **h**: Pulse-chase assay with fluorescently labeled wildtype ubiquitin with UBR5 in the presence of UbV3 or UbV3$^{I44D}$ respectively. **i**: Pulse-chase assay with fluorescently labeled wildtype ubiquitin with UBR5 or UBR5$^{L224D}$ in the presence of UbV3 respectively.

# Discussion

## 1. Mechanism of UBR5-mediated polyubiquitylation

In the course of this study, we were able to visualize all different steps of HECT-mediated catalysis leading to polyubiquitylation. For this, we employed various activity-based probes to catch the huge HECT E3 ligase UBR5 in the act of undergoing a distinct transition or stabilizing an otherwise transient intermediate. With these complexes, we performed single-particle cryo-EM and verified the identified interfaces *in vitro*.

Modulation of the distinct steps allowed us to hypothesize a mechanism, explaining why UBR5 is so efficient at forming ubiquitin chains, while being extremely specific in the linkage-type it forms. The intricate activity can be itemized in three main parts:

### Conformational flexibility allows catalytic activity

Focusing on UBR5's HECT domain in our various structures and models, it becomes apparent that this has to undergo massive conformational changes in order to facilitate the distinct reaction (**Figure 34a**). Our structures can complement previous models in that we can not only see movements of the C-lobe with regards to the N-lobe, but we also visualize that the N-lobe has to relocate upon E2~Ub$^{D}$-binding to prevent clashing of the E2 with the RLD in trans and clashing of ubiquitin with the scaffold in trans. So far, this transition state had only been visualized for an isolated HECT domain, making it impossible to judge how the entire HECT domain has to reposition with respect to the remaining protein.[84] Dissociation of the E2 subsequently to transthiolation then allows the C-lobe~Ub$^{D}$ unit to swivel back into the preferred L-conformation, which is also the transpeptidation-reactive conformation.[86,88]

### A feed-forward mechanism grants UBR5 high processivity

Analysis of the various interactions within the different states suggests that different steps of UBR5-mediated catalysis are accelerated by promotion of the next intermediary state in a feed-forward mechanism (**Figure 34b**). The first process that is accelerated is dissociation of the E2 subsequently to transthiolation. Our structure of apo UBR5 clearly suggests that the preferred conformation of UBR5 is the L-conformation with the C-lobe sitting on the edge of the N-lobe. This conformation is stabilized by additional domains outside the HECT domain and is also the transpeptidation reactive state (**Figure 18b, 24e, 26c**). However, to accommodate E2~Ub$^{D}$, the HECT domain has to disengage from the DSD domain. The HD domain at least has to reposition, if not disengage as well. This disadvantageous state is stabilized by ubiquitin still being linked to the E2 enzyme, which holds the C-lobe in the inverted T-conformation. In this conformation, the HECT domain could be envisioned as a tensed spring and transfer of Ub$^{D}$'s C-terminus to UBR5's catalytic cysteine allows release of the tension: the E2 is not stably wedged in between the HECT domain and the RLD anymore and it can dissociate to allow HD and DSD

domains to renew their interactions with the HECT domain. Revisiting an early assay analyzing the activity of various UBR5-constructs shows that even though all required components for free ubiquitin chain formation are present in the UBA-HECT fusion, the processivity is significantly worse than for full-length UBR5 (**Figure 19a**). This effect might partially be caused by the lack of the HECT supporting domains HD and DSD that promote E2 dissociation after transthiolation and thereby are part of the feed-forward mechanism. The importance of stabilizing UBR5's HECT domain via additional motifs was also shown *in vivo*.[139] Using a CRISPR-tiling screen, it was identified that alterations in the HD domain result in stabilization of the UBR5 substrate RARA.

Transition back to the apo-state after transpeptidation is also promoted by negatively synergizing effects. Stabilization of the active site to achieve transpeptidation requires $Ub^D$'s C-terminus covalently tethered to UBR5's catalytic cysteine (**Figure 27b**). Upon transfer of $Ub^D$ to $Ub^A$ by transpeptidation, this stabilization of $Ub^D$'s C-terminus on UBR5 is lost and the active site becomes looser. Increased movements of the components contributing to the active site configuration then result in the collapse of the active site, the newly formed ubiquitin chain is released, UBR5's C-terminal tail is flexible again, and the UBA domain is set free. Thus, UBR5 is rapidly reset for another round of E2-binding, ubiquitin transthiolation, and ubiquitin transfer to a substrate or a growing ubiquitin chain. Determining the reaction kinetics of the distinct steps will be valuable tools to validate this mechanism.

### Avid binding of the acceptor ubiquitin achieves high specificity

To achieve high linkage-specificity, the acceptor ubiquitin is bound by multiple different elements on several sides (**Figure 34c**). Disrupting either of them is not sufficient to alter linkage-specificity, which would have become apparent in our assays due to the use of donor ubiquitin with a K48R mutation, but it can be compensated by the various other interactions, if polyubiquitylation is not completely abrogated (**Figure 27-30**).

First, $Ub^A$ gets recruited by the flexibly tethered UBA domain to bring it into closer proximity to the HECT domain. Loose contacts between $Ub^A$ and the N-lobe, more specifically parts of the LOL, then roughly orient $Ub^A$. The acceptor ubiquitin then gets latched in a position allowing for polyubiquitylation by the large hydrophobic C-lobe residue Y2773 nestling into a hydrophobic pocket on $Ub^A$ created by A46 and the adjacent F45. These interactions facilitate precise binding of the acceptor ubiquitin and allow luring K48 to the active site. An additional salt bridge between acceptor and donor ubiquitin aids in stabilizing the active site configuration to support transpeptidation.

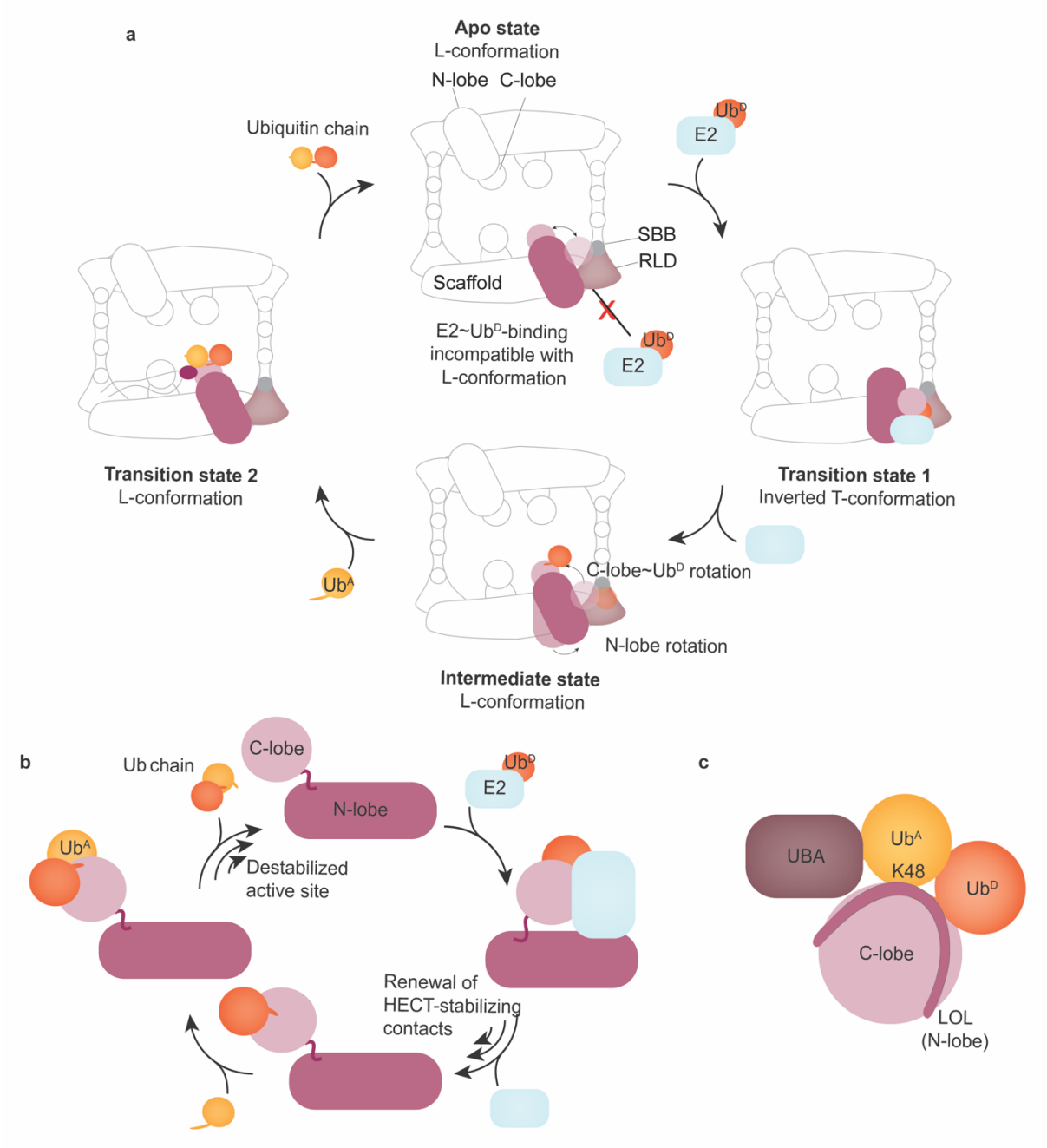


**Figure 34 Mechanism of UBR5-mediated polyubiquitylation.**

**a:** Model of tetrameric UBR5 in the distinct states of catalysis. Flexibility of UBR5's HECT domain allows for catalytic activity. **b:** HECT domain in different catalytic states showing a feed-forward mechanism that allows high processivity in UBR5-mediated polyubiquitylation. **c:** Cartoon showing avid binding of $Ub^A$ by several components, which results in high linkage-specificity.

## 2. Generality of the mechanism

HECT domains are highly conserved catalytic regions and many publications have proposed a similar mode-of-action for the different HECT E3 ligases. Considering the entirety of UBR5 during the respective conformational changes allowed us to expand the common knowledge and by visualizing all steps for the same HECT E3 ligase, we can

now distinguish what effects are specific to a certain step and which features might be specific for a distinct HECT E3.

### Conservancy of HECT domain movement along trajectory

Flexibility of the two subdomains of UBR5's HECT domain around the flexible interlobe tether is required for catalytic activity. This is in great agreement with the literature, showing various conformations for different HECT domains during distinct steps.[82,83,205] Knowing the required movements of UBR5 to facilitate distinct steps allows us to compare prior crystal structures to be able to conceptualize them even better.

Different HECT E3 ligases engage in different preferred orientations in the apo state.[82,83,206,207] This is especially clear when considering full-length HECT E3 ligases. In case of UBR5, the apo HECT domain preferentially engages in the L-conformation, stabilized by additional domains. Instead, the HECT domains within full-length HUWE1 or full-length HACE1 are preferentially oriented in the inverted T-conformation in an apo state.[89-92]. Why HECT domains of different E3 ligases accommodate different preferred orientations remains elusive. We hypothesize that in case of UBR5, the preferred L-conformation contributes to the high processivity by promoting E2-dissociation after the transthiolation and by resetting the system after transpeptidation. However, the preferred orientations might also be a molecular response to limited protein pools, for example of E2~Ub, to bind this moiety more efficiently, or it might be a tool regulating the rate-limiting step.

The prior structure of a HECT domain in TS1 agrees in the binding of E2~$Ub^D$ to be facilitated in the inverted T-conformation.[84] Strikingly, structural alignment of NEDD4L's HECT domain in complex with E2~$Ub^D$ and UBR5's HECT domain bound to E2~$Ub^D$ shows great consensus (**Figure 35a**). E2 binding on the N-lobe as well as $Ub^D$ binding on the C-lobe seem to occur in a similar fashion.

Subsequently to transthiolation of ubiquitin to the E3's catalytic cysteine, the E2 can dissociate and ubiquitin remains bound to the C-lobe. A crystal structure of NEDD4's HECT domain bound to ubiquitin but not the E2 shows the HECT domain arranged in the inverted T-conformation. This structure of post transition state 1 aligns very well with UBR5's HECT domain in TS1 (**Figure 35b**). We hypothesize that in case of NEDD4, relocation of the C-lobe~$Ub^D$ into the transpeptidation-reactive L-conformation after transthiolation is not promoted by a feed-forward mechanism, causing it to remain in the inverted T-conformation for a longer time. It would be interesting to determine whether this potentially delayed relocation into the L-conformation has any regulatory effects on NEDD4 or whether it might be slowed down by the additional ubiquitin bound to NEDD4's exosite in this structure.

In anticipation of the transpeptidation-reaction, the C-lobe~$Ub^D$ unit swivels around the N-lobe subsequent to the transthiolation and E2-dissocation. The HECT domain with bound ubiquitin engages in the L-conformation, as it was previously shown in a crystal structure

of HUWE1's HECT domain bound to ubiquitin, and is in agreement with our UBR5~Ub$^{D}$ model (**Figure 35c**).[88]

Lastly, one crystal structure showed a HECT E3 transferring bound Ub$^{D}$ to a substrate peptide.[86] This conformation of Rsp5 (cross-linked to Ub$^{D}$ and a Sna3 substrate peptide) agrees with UBR5's HECT domain conformation in the Ub$^{D}$-bound intermediate state and with TS2 (**Figure 35d**).

Very close resemblance of the different UBR5 conformations with various other HECT E3 ligases during different steps verifies that different conformations facilitate the different steps of HECT-mediated catalysis not only for specific HECT E3 ligases but generally.

Additionally, mutations in the short flexible tether connecting N- and C-lobe were shown to massively impair activity of WWP1 supporting the notion that the relay-mechanism we observed in our structures is highly conserved amongst HECT E3 ligases.[83]

An exception to this mechanism might be HECT E3 ligases that perform en bloc transfer rather than sequential ubiquitylation.[61] This had been shown for E6AP, which preassembles ubiquitin chains linked to its catalytic cysteine prior to transferring the entire chain to a substrate. A recent publication showed that this feature was abolished upon mutation of the flexible tether and E6AP could now form ubiquitin chains in a sequential mode.[205]

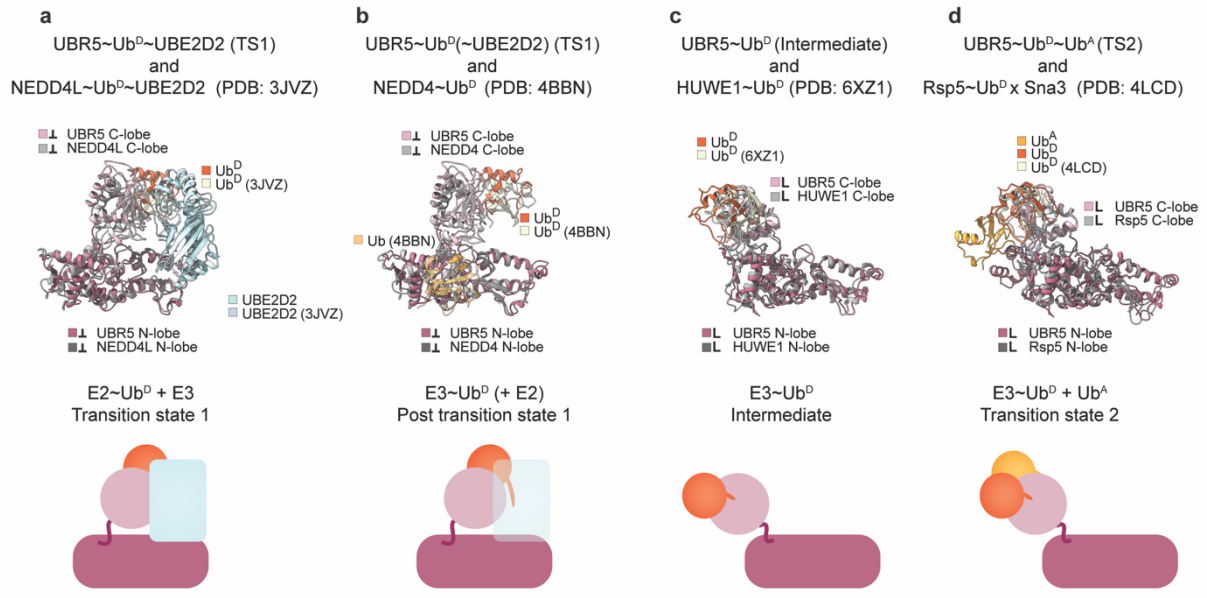


**Figure 35 Comparison of HECT domain structures of different steps with UBR5.**

**a:** Comparison of NEDD4L$^{HECT}$~Ub$^{D}$~UBE2D2 crystal structure (PDB: 3JVZ) with model of TS1 of UBR5. **b:** Comparison of NEDD4$^{HECT}$~Ub$^{D}$ crystal structure (PDB: 4BBN) with model of TS1 of UBR5. **c:** Comparison of HUWE1$^{HECT}$~Ub$^{D}$ crystal structure (PDB: 6XZ1) with model of Ubiquitin bound intermediate of UBR5. **d:** Comparison of Rsp5$^{HECT}$~Ub$^{D}$ x Sna3 crystal structure (PDB: 4LCD) with structure of TS2 of UBR5.

## Determinants of linkage-specificity

It had been proposed for a long time already that the acceptor ubiquitin binds to the HECT domain's C-lobe and thereby determines linkage-specificity of HECT E3 ligases.[40,61,88] This hypothesis was further elaborated and it was suggested that the ultimate C-terminal residue determines linkage-specificity.[87,111] However, this model can not fully explain HECT-mediated linkage-specificity as discussed earlier. We hypothesize that the ultimate C-terminus contributes to linkage-specificity rather than determining it.

Even though, none of our maps shows density for UBR5's ultimate residue, and we did not mutate the ultimate residue to determine a potentially altered linkage-specificity, we speculate that more factors than just the ultimate residue contribute to the linkage-specificity in case of UBR5. Our structure of polyubiquitylating UBR5 reveals several interfaces of the acceptor ubiquitin with UBR5 and also with $Ub^D$, by which $Ub^A$ is oriented. If an altered C-terminal residue would result in a shifted linkage-specificity, the other interactions could not be engaged properly anymore. However, our observation that mutating these interfaces can abrogate polyubiquitylation (**Figure 28a, 29a, 29d**), shows that they are required for efficient polyubiquitylation and have to be engaged.

Therefore, we suppose that while for some HECT E3s, the ultimate residue might be sufficient for linkage-specificity, the high processivity of some HECT E3 ligases, and the concomitant short-lived and rather weak interactions with $Ub^A$ require multiple factors ensuring specificity during polyubiquitylation.

This hypothesis will have to be tested in subsequent studies by mutating UBR5's ultimate residue and -assuming that UBR5 still performs polyubiquitylation- analyze whether the linkage-specificity is shifted, abrogated, or maintained.

## Ligation-organizing-loop

Resolving the structure of UBR5 undergoing polyubiquitylation, allowed us to identify a crucial element in HECT-mediated ubiquitin chain formation. An acidic loop located in the N-lobe, which is relatively conserved amongst HECT E3 ligases, reaches out from the N-lobe to synergize $Ub^D$, $Ub^A$, and the C-lobe (**Figure 28a**). Previous studies on other HECT E3 ligases ascribe this loop an important role during transpeptidation[86,88], however, the lack of structural data visualizing this step, did not allow to assign a proper function to it yet.

We propose that the LOL assists in recruiting $Ub^A$ to the C-lobe via a salt bridge and then helps orienting it with respect to the C-lobe and $Ub^D$. This easy-to-establish bond can aid in approximating the acceptor ubiquitin to the C-lobe to facilitate the $Ub^A$-C-lobe interaction, which locks $Ub^A$ in a position allowing transpeptidation.

The presence of such a loop in other HECT E3 ligases further reinforces its importance not only during polyubiquitylation, but generally during isopeptide-bond formation. Mutation of this loop has led to impaired substrate modification as well as impaired K63-linked di-ubiquitin formation for the yeast HECT E3 ligase Rsp5.[86] This broad involvement of the LOL in substrate ubiquitylation as well as in polyubiquitylation with various chain

types suggests a role in preparing an acceptor lysine to allow its nucleophilic attack on the donor ubiquitins C-terminus. However, assessing its role at varying pH shows that the LOL is not involved in lysine deprotonation but must have another role. Interestingly, a similar mechanism had been proposed for bacterial HECT-like enzymes, which might recruit or orient acceptor ubiquitins using similar mechanisms.[208]

### Flexibility of HECT domain's C-terminus

Previous studies on other HECT E3 ligases suggested that the ultimate C-terminus determines or at least supports linkage-specificity. Additionally, a HECT domain's C-lobe contains the highly conserved -4 Phe, which was proposed to make contacts with the N-lobe to contribute to the overall HECT conformation.[86,88] However, the limited availability of structural data visualizing the C-terminal tail of HECT E3 ligases posed some difficulties to derive a mechanism for these residues.

Analysis of our cryo-EM reconstructions of apo $UBR5^{Dimer}$ as well as $UBR5^{Dimer}$ in TS2, reveal why the C-terminal tail is most often not visible in structures of HECT E3 ligases: The C-terminus of the HECT domain is highly flexible and an intricate web of interactions between the different binding partners is required to stabilize it in the transpeptidation-reactive orientation (**Figure 27b**). The C-terminus in return organizes the active site, facilitating transpeptidation by ensuring the proper arrangement of $Ub^{D}$ and $Ub^{A}$ with respect to each other, and by stabilizing the C-lobe in the transpeptidation-reactive L-conformation by contacting the N-lobe.

To date, only one structure of a HECT domain was published with the ultimate residue built.[88] However, the lack of an acceptor in this structure, does not allow to conclude how the ultimate residue in a HECT domain might influence linkage-specificity and further studies will be required to do so.

### Flexible ubiquitin binding domains during branched chain formation

Our structural investigation of UBR5's TS2, as well as the follow-up biochemical characterization demonstrates the importance of the flexibly-tethered UBA domain to bind $Ub^{A}$ (**Figure 30a**). If the interaction with the UBA domain was impaired from either side, formation of ubiquitin chains was significantly impaired. In return, a minimal construct with the UBA domain fused to the HECT domain was sufficient to form free ubiquitin chains. Involvement of the UBA domain in UBR5-mediated polyubiquitylation of canonical substrates *in vitro* and *in vivo* was also observed in other studies.[139,140] Noteworthy, one substrate could be modified by UBR5 to some extent also in the absence of an intact UBA domain.[138] We think that substrate-modification with homotypic ubiquitin chains by UBR5 relies on the UBA domain to some extent. However, if a substrate binds in a way that a growing ubiquitin chain would already be in close proximity to the HECT domain, the UBA domain might not be critical anymore, as the acceptor ubiquitin is close enough at the HECT domain to form the other required interactions.

Another HECT E3 ligase, which was proposed to efficiently form branched ubiquitin chains is HUWE1, which is also grouped in the family of "other" HECT E3 ligases.[67,70] Full-length structures of HUWE1 showed that it forms an oval with the HECT domain being exposed towards the outside.[89,90] Strikingly, the HUWE1 structures were also missing distinct features: the flexibly-tethered ubiquitin-interacting UBA, UIM, and UBM domains, which were also be shown to be crucial for HUWE1-mediated polyubiquitylation *in vitro*.[89]

The significance of flexibly-tethered ubiquitin-binding domains in both major branched chain-forming HECT E3 ligases led us to believe that these domains might contribute to heterotypic ubiquitin chain formation. Such heterotypic chain formation can be induced upon certain cellular stress conditions, when the cell quickly has to turn over large amounts of proteins.[67] In such a case, the ubiquitin-moiety, rather than the substrate-portion itself gets recognized by the branching E3 enzyme. Ubiquitin binding domains could then serve as "recognition" domains to generally recruit substrates that are already modified with distinct chains to allow rapid degradation.

### Overall HECT E3 ligase structures

In the course of our study, full-length structures of other HECT E3 ligases were released. Structures of human[89], and Nematocida HUWE1[90] as well as human HACE1[91,92] were published apart from several structures portraying UBR5[137-141]. Despite all of these HECT E3 ligases being present in different oligomeric states – HUWE1 being a monomer, HACE1 being found as a head-to-tail assembled dimer, and UBR5 being a dimer of dimers – all of the structures share striking similarities: In each case, an α-helix-rich scaffold dictates the overall shape of the protein and orients the other domains with respect to each other. Additionally, as observed for many other E3 ligases, all of these HECT E3 ligases are found in an oval or round shape. In case of HUWE1, the catalytic domain is displayed towards the outside of the circle[89,90], whereas in case of UBR5 all the catalytic domains are positioned inwards[137-141], suggesting collaboration of these. Furthermore, formation of the oval by dimerization results in autoinhibition of HACE1[91,92]. Therefore, it seems unlikely that the oval shape of these three ligases confers similar mechanistic intricacies.

## 3. Future directions

Analyzing the distinct steps of UBR5-mediated ubiquitylation allowed us to conclude a mechanism how UBR5 achieves its striking activity. However, many features of UBR5 await further investigation.

### Role of oligomerization

During purifications, we noted early on that UBR5 is a tetramer and that the main unit is a dimer. This is in agreement with simultaneous findings of other groups, who mostly described a tetrameric oligomeric state for UBR5 as well.[137-140] Additionally, one group

proposed a subset of UBR5 to also be present in a dimeric or an octameric state apart from the predominant tetramer.[138] When disrupting the tetramer to obtain stable dimers, polyubiquitylation activity did not seem to be significantly affected *in vitro*.

This led us to question why an anyway big protein would go through the hassle of forming a major oligomer, which is almost as big as prokaryotic ribosomes. Or more precisely: what does UBR5 need to be a tetramer for? Our biochemical and structural analysis can easily explain why UBR5 needs to be a dimer: many aspects of UBR5-mediated ubiquitylation require the two monomers within a dimer to collaborate. As such, the DSD domain supports UBR5 in the L-conformation in trans (**Figure 18c**), the RLD and dSBB domains in trans render it indispensable for the HECT domain to actively relocate to allow E2~Ub$^{D}$ binding (**Figure 20a**) and doing so, the feed-forward mechanism is facilitated by UBR5 features in trans. Additionally, the huge interface between the two monomers within a dimer imparts stability (**Figure 13d**).

Tetrameric UBR5, instead might be beneficial to avidly bind oligomeric substrates (**Figure 36a**). A similar mechanism has been shown for the multisubunit RING E3 ligase glucose-induced degradation deficiency ("GID"), which can engage in a monomeric complex or in a dimeric complex. Depending on which oligomeric state the GID complex is in, it can facilitate ubiquitylation of different substrates, which are bound avidly and modified from several sites simultaneously.[209] A similar mechanism could be considered for UBR5.

Another potential advantage of tetrameric UBR5 over dimeric UBR5 could be the presence of multiple flexible UBA domains to assist in the assembly of long ubiquitin chains (**Figure 36b**). We could rule out that the higher number of UBA domains confers a specificity of an acceptor ubiquitin chain (**Figure 31a**). With increasing length of a formed ubiquitin chain, the flexibility of this chain increases. This makes it more difficult for the E3 to stably bind and orient the chain in a way that facilitates further modification. The flexibly tethered UBA domains could then collaborate to stabilize the growing chain. We could show that both, tetrameric and dimeric UBR5 can modify ubiquitin chains of up to five ubiquitins comparably well (**Figure 16e**), but it is possible that such an effect would only be visible with significantly longer chains.

A third potential mechanism how tetrameric UBR5 might be beneficial over dimeric UBR5 is when previously assembled ubiquitin chains are branched (**Figure 36c**). Again, the higher number of UBA domains might be advantageous here, and a preexisting ubiquitin chain could potentially be bound and modified on several moieties simultaneously, thereby achieving more branching events, which in return could accelerate the cell's response to this modification. Such a mechanism could explain why tetrameric UBR5 modified an acceptor chain on several sites rather than dimeric UBR5 (**Figure 16e**).

Lastly, one potential advantage of tetrameric UBR5 is inspired by the mechanism of fatty acid synthase: linked to an acyl-carrier protein, the growing fatty acid substrate is shuttled from one active site to the next to allow rapid and efficient synthesis.[210] Something comparable could also be imagined for UBR5: a substrate bound to the flexible MLLE

domain, or an already modified substrate bound to the UBA domain could be shuttled from one HECT domain to the next to promote its polyubiquitylation (**Figure 36d**).

Whether one of these hypotheses is correct, will have to be addressed in future studies to understand UBR5-mediated substrate ubiquitylation *in vivo* in more depth. Determining the kinetics of tetrameric and dimeric UBR5 at different reactions, such as substrate-priming, substrate polyubiquitylation, and branched chain formation, can help to assess differences between the oligomeric states and might aid in assigning a specific function to this state.

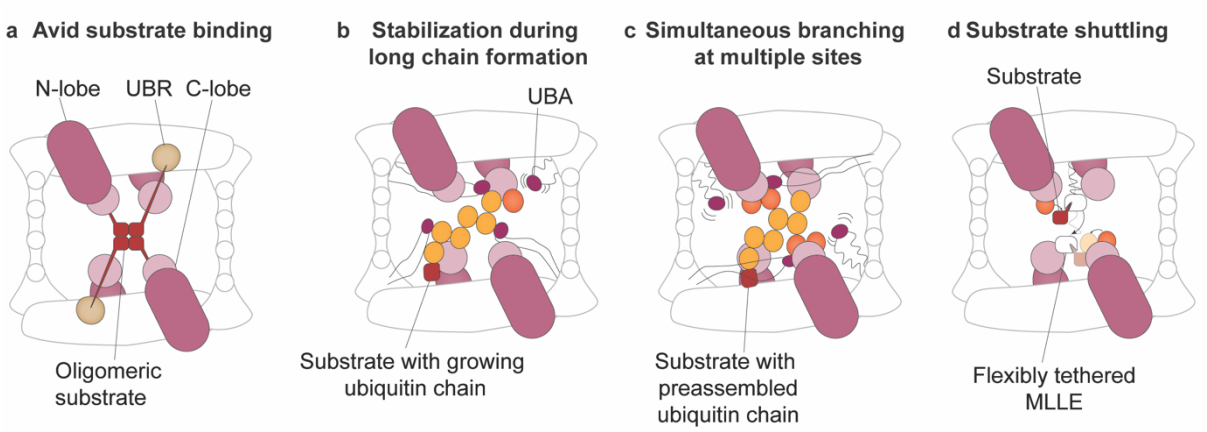


**Figure 36 Models for role of UBR5's oligomerization state in substrate ubiquitylation.**

**a:** Model of tetrameric UBR5 facilitating avid substrate binding. The catalytic HECT domains are indicated. **b:** Model of tetrameric UBR5 allowing for stable binding of growing ubiquitin chain to facilitate formation of long chains. **c:** Model of tetrameric UBR5 simultaneously branching a preassembled ubiquitin chains on several sites. **d:** Model of tetrameric UBR5 with substrate shuttling between different catalytic domains indicated.

## UBR5-mediated substrate-modification

Our data (**Figure 32**) and another study investigating UBR5-mediated substrate ubiquitylation agree that UBR5 is much better at modifying a substrate that is already primed with ubiquitin, rather than priming it itself. Since UBR5 interacts with an acceptor ubiquitin on many levels (**Figure 26d**), it is obvious that ubiquitin itself is a great substrate and it explains why UBR5 can readily modify preassembled ubiquitin chains *in vivo*.[67,70] However, future studies will have to determine whether additional factors such as cellular localization, other proteins to support substrate-binding, or activity-modulating proteins assist in UBR5-mediated substrate-priming *in vivo*.

## Regulation of UBR5 activity *in vivo*

Our *in vitro* reconstitution of UBR5-mediated polyubiquitylation revealed extraordinary chain-formation abilities. While being highly processive due to the feed-forward mechanism, UBR5 is very specific in which linkage-type it forms (**Figure 13h**). At the same time, UBR5 is very versatile in which preassembled ubiquitin chains it further modifies

(**Figure 31a**). We suspect that this extraordinary activity is somehow regulated *in vivo*. The huge size of UBR5 makes it unlikely that UBR5-regulation solely depends on altering its stability and expression, but we suspect that UBR5 might be bound to additional proteins *in vivo* that impose a certain specificity on which chains are modified, or a binding partner that could keep UBR5 in an inhibited state. Previous studies suggested that UBR5's MLLE domain might have a regulatory function itself by binding to a PAM2-like motif on UBR5's N-lobe.[157] However, neither of the structural analysis published alongside our study, nor any of the distinct UBR5-steps within our study revealed density that could be identified as MLLE domain.[137-141]

Future studies, especially with a focus on *in vivo* characterization of UBR5-activity will be required to determine whether or how UBR5 is regulated in a cell.

## Tools for UBR5 investigation and targeting

The common misregulation of UBR5 in different types of cancer (**Figure 9a**) as well as its striking ability to form ubiquitin chains in a highly processive and specific manner make it a very promising target for drug discovery.

One approach in clinical research would be to impair UBR5's catalytic activity *in vivo* in order to compensate for upregulated UBR5. This could be achieved by different types of small molecules, or by UbVs. Discovery of such molecules could be aided by UbSRhodoI, which facilitates straight-forward measurement of the catalytic activity of HECT E3 ligases.[124]

Since the UBR5-abundance is often massively increased in cancer, using UBR5 to induce targeted protein degradation could be a very efficient way to remove access disease-driving proteins. Understanding UBR5's polyubiquitylation-mechanism[141] and the mechanism of substrate-recognition[139,140] pave the way to use UBR5 for future studies attempting targeted protein degradation.[211] Another HECT E3 ligase -TRIP12- was already part of a study investigating ubiquitylation of neo-substrates upon the treatment of small molecules to induce targeted protein degradation.[111] We think that the more profound understanding of HECT-mediated polyubiquitylation will allow for more HECT E3 ligases to be used as tools for TPD more commonly.

# Methods

## Construct design and cloning

Constructs used in this study were generated by the use of standard biology techniques and verified with Sanger sequencing. The constructs that were used for recombinant protein expression and protein purification are listed in Table 1.

**Table 1 Constructs used for recombinant protein expression**

| Protein | Construct | Vector | Expression system |
|---|---|---|---|
| E1 | GST-TEV-UBA1 | pLIB | High-Five insect cells |
| E2 | GST-TEV-UBE2D2 | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-UBE2D3 | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-UBE2D3$^{F62A}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-UBE2D2$^{C21A,C107A,C111S}$ (for ABP) | pGEX | BL21(DE3) RIL cells |
| Donor ubiquitin | GST-3C-Cys-Ub | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Cys-Ub$^{K0}$ | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Cys-Ub$^{K48R}$ | pGEX | BL21(DE3) RIL cells |
| Single K ubiquitin | GST-3C-Ub | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Ub$^{K0}$ | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Ub$^{K6only}$ | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Ub$^{K11only}$ | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Ub$^{K27only}$ | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Ub$^{K29only}$ | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Ub$^{K33only}$ | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Ub$^{K48only}$ | pGEX | BL21(DE3) RIL cells |
| | GST-3C-Ub$^{K63only}$ | pGEX | BL21(DE3) RIL cells |
| M1-linked chains | GST-TEV-$Ub_2$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{WT,WT,WT}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{K48R,K48R,K48R}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{K48R,WT,WT}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{WT,K48R,WT}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{WT,WT,K48R}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{K48R,K48R,WT}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{K48R,WT,K48R}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{WT,K48R,K48R}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{I44D,I44D,I44D}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{I44D,WT,WT}$ | pGEX | BL21(DE3) RIL cells |

| | | | |
|---|---|---|---|
| | GST-TEV-$Ub_3^{WT,I44D,WT}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{WT,WT,I44D}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{I44D,I44D,WT}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{I44D,WT,I44D}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-$Ub_3^{WT,I44D,I44D}$ | pGEX | BL21(DE3) RIL cells |
| Acceptor ubiquitin | Ub(1-76)-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{L8A,T9A}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{I44A}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{V70A\ L71A}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{H68A}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{R54E}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{A46F}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{A46D}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{F45A}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{F45Y}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{F45D}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{D58A}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | $Ub(1-76)^{D58R}$-6xHis | pRSF | BL21(DE3) RIL cells |
| | Ub | pET22b | BL21(DE3) RIL cells |
| HECT domain constructs | GST-TEV-HECT (2216-2799) | pFLN | High-Five insect cells |
| | GST-TEV-$HECT^{\Delta MLLE}$ (2216-2376 –6 aa linker– 2455-2799) | pFLN | High-Five insect cells |
| | GST-TEV-UBA-$(GSG)_5$-HECT (179-230 - 15 aa linker- 2216-2799) | pFLN | High-Five insect cells |
| | GST-SUMO-MLLE-$HECT^{\Delta N}$ (2377-2799) | pGEX | BL21(DE3) RIL cells |
| | GST-SUMO-$HECT^{\Delta N}$ (2499-2799) | pGEX | BL21(DE3) RIL cells |
| | GST-SUMO-$N^{\Delta N}$ (2499-2686) | pGEX | BL21(DE3) RIL cells |
| | GST-SUMO-C (2687-2799) | pGEX | BL21(DE3) RIL cells |
| UBR5 mutants | TwinStrep-GFP-3C-UBR5 | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-$UBR5^{C2768A}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-$UBR5^{L710D}$ ($UBR5^{Dimer}$) | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-$UBR5^{HLL1362-1364DDD}$ ($SDA^{mut}$) | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-$UBR5^{A2790W}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-$UBR5^{L224D}$ | pEG | HEK293S, BacMam-system |

| | | | |
|---|---|---|---|
| | TwinStrep-GFP-3C-UBR5$^{D2283\text{-}2287A}$ (LOL$^{mut}$) | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{E2287R}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{Dimer,E2287A}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{F2796A}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{F2798A}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{\Delta 2799}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{Dimer,Y2773D}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{Dimer,Y2773F}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{Y2773D}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{Y2773F}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{Dimer,G199W}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{Dimer,P2748D}$ | pEG | HEK293S, BacMam-system |
| | TwinStrep-GFP-3C-UBR5$^{Q2747L}$ | pEG | HEK293S, BacMam-system |
| ABP-constructs | His-Ub(1-75)-Intein | pET22b | BL21(DE3) RIL cells |
| | Ub$^{K48C}$ | pET22b | BL21(DE3) RIL cells |
| Substrates | GST-TEV-MYC_peptide(174-196)-KK-Ub$^{GGR}$ | pGEX | BL21(DE3) RIL cells |
| | GST-TEV-MYC_mutant_peptide-KK-Ub$^{GGR}$ | pGEX | BL21(DE3) RIL cells |
| | GST-SUMO-R-Ub$^{\Delta GG}$ | pGEX | BL21(DE3) RIL cells |
| | GST-SUMO-RIFSTD-Ub$^{\Delta GG}$ | pGEX | BL21(DE3) RIL cells |
| | GST-SUMO-RIFSTD(GS)$_{10}$-Ub$^{\Delta GG}$ | pGEX | BL21(DE3) RIL cells |
| | GST-SUMO-RIFSTD(GS)$_{15}$-Ub$^{\Delta GG}$ | pGEX | BL21(DE3) RIL cells |
| UbVs | 6xHis-FLAG-UbV3 | pRSF | BL21(DE3) RIL cells |
| | 6xHis-FLAG-UbV7 | pRSF | BL21(DE3) RIL cells |
| | 6xHis-FLAG-UbV3$^{I44D}$ | pRSF | BL21(DE3) RIL cells |

## Protein expression

### Protein expression in *E. coli*

Expression of ubiquitin (mutants), E2 constructs, substrates, UbVs, and truncations of UBR5's HECT domain was conducted in *E. coli* BL21(DE3) RIL cells harboring the respective expression plasmids. Cultivation took place at 37°C in TB medium supplemented with suitable antibiotics until the culture reached an optical density of 0.8. Subsequently, the temperature was reduced to 18°C, and IPTG was added to a final concentration of 0.6 mM to induce expression, which continued for 18 hours. Cell harvesting was done by centrifugation at 4°C for 15 minutes at 4,500 x g.

### Protein expression in insect cells

Expression of different versions of UBR5's HECT domain with N-terminal GST-tags, or expression of the human E1 enzyme UBA1 also containing an N-terminal GST-tag and a subsequent TEV cleavage site, was performed using insect cells.

First, Sf9-cells were infected with Bacmids that were prepared to carry the respective gene expression cassette. Next, the obtained Baculovirus was used to infect Hi5-cells to recombinantly express the protein of choice.

After 3 days of expression at 27°C, cells were harvested by centrifugation at 4°C for 15 minutes at 1,000 x g.

### Protein expression in mammalian cells (BacMam-system)

Full-length UBR5 as well all UBR5 point mutations were expressed using the BacMam-system.[212] All UBR5-constructs derived from a GFP-UBR5 plasmid kindly gifted by Darren Saunders (Addgene plasmid #52050; http://n2t.net/addgene:52050; RRID: Addgene_52050).[159] The UBR5 open-reading frame was recloned into a pEG-vector that can be used for recombinant expression via the BacMam-system in HEK293S-cells. Noteworthy, the original plasmid contained the UBR5 point mutation K503R and all UBR5-constructs used in this study therefore also carry this mutation even though we refer to them as wildtype and do not mention the mutation throughout the study for clarity reasons. Baculovirus of the respective construct was prepared as described earlier and used to infect HEK293S cells at 10 % final ratio. HEK293S cells were cultured to a cellular density of ~3 Mio cells/mL in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10 % fetal calf serum and a mix of antibiotics. To improve UBR5 expression, a final of 100 µM $ZnSO_4$ was added alongside the Baculovirus. 8 hours post-infection, 10 mM Na-Butyrate was added and the cells were then cultured for 60 hours at 30°C.

Cell harvesting was performed by centrifugation at 4°C for 15 minutes at 450 x g.

## Protein purification

The resulting pellet was resuspended in ice-cold lysis buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 5 mM DTT or instead 5 mM β-mercaptoethanol for His-tagged constructs, and 2.5 mM PMSF). If protein expression was conducted in insect cells or mammalian cells, 10 µg/mL leupeptin, 20 µg/mL aprotinin, and 10 µg/mL DNase were added as well. Only the lysis buffer for His-tagged Ub-Intein used to generate Ub-BmDPA differed significantly as described below.

Cells were lysed through sonication on ice, followed by pre-clearing the lysate via centrifugation at 4°C for 40 minutes at 20,000 x g.

### Acceptor ubiquitin mutants and UbVs

Acceptor ubiquitin containing a C-terminal His-tag ($Ub^{WT}$, $Ub^{D58A}$, $Ub^{D58R}$, $Ub^{R54E}$, $Ub^{A46F}$, $Ub^{A46D}$, $Ub^{F45A}$, $Ub^{F45Y}$, $Ub^{F45D}$, $Ub^{L8A,T9A}$, $Ub^{I44A}$, $Ub^{H68A}$, $Ub^{V70A,L71A}$), and the UbVs underwent purification through Ni-NTA-affinity chromatography using a gravity flow column setup. Washing the resin with buffer A (50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 5 mM β-mercaptoethanol) containing 20 mM imidazole facilitated the removal of nonspecifically bound proteins, and elution of specifically bound proteins was achieved using an elution buffer containing a high imidazole-concentration (buffer A + 300 mM imidazole). Size exclusion chromatography at 4°C in a suitable buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl yielded pure His-tagged proteins.

### E2 enzymes

GST-tagged proteins, such as E2 enzymes (UBE2D2, UBE2D3, $UBE2D3^{F62A}$, $UBE2D2^{AAS}$ for activity-based probe), single lysine acceptor ubiquitins, and M1-linked ubiquitin chains were purified using GST-sepharose resin. After incubating the pre-cleared lysate with the resin, extensive washing with buffer B (50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 5 mM DTT) followed by elution using 10 mM reduced glutathione in buffer B helped obtain the GST-fusion proteins. Subsequent cleavage of the GST-tag employed specific proteases overnight at 4°C, namely His-tagged HRV 3C protease for single lysine ubiquitin-constructs or His-tagged TEV protease for E2 enzymes and M1-linked ubiquitin chains. Further purification of E2s and M1-linked chains involved ion-exchange chromatography and size exclusion chromatography. Single lysine ubiquitins were additionally purified using size exclusion chromatography (buffer C: 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT) subsequent to affinity purification.

**Fluorescently-labeled ubiquitin**

Fluorescently-labeled ubiquitins ($Ub^{WT}$, $Ub^{K0}$ with all lysines mutated to arginine, and $Ub^{K48R}$), intended for fluorescence experiments, were generated by expressing them as GST-3C-fusions in a pGEX-vector with an N-terminal cysteine.[213] This process involved GST-affinity chromatography, HRV 3C protease cleavage, and subsequent size exclusion chromatography into 50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM DTT to ensure complete reduction of the N-terminal cysteine before labeling. Removal of the reducing agent right before the labeling reaction was achieved by two desalting steps using ZebaTM Spin desalting columns in 50 mM HEPES pH 7.5, 150 mM NaCl. The labeling process involved the addition of fluorescein-5-maleimide or tetramethylrhodamine-5-maleimide (TAMRA) resuspended in anhydrous DMSO to reduced and desalted ubiquitin, followed by 2 h incubation and subsequent quenching with 10 mM DTT to inactivate unused maleimide. Purification of the labeled ubiquitin conjugates included desalting and two runs of size exclusion chromatography to ensure high purity.

**Untagged ubiquitin**

Untagged ubiquitin was the basic module for many different ubiquitin chains employed in this study and the $Ub^{K48C}$ variant was furthermore required to form the activity-based probe of transition state 2. To isolate ubiquitin from other proteins, acetic acid (glacial) was gradually added to the lysate until reaching a pH of approximately 4.5, selectively precipitating proteins other than the stable ubiquitin. Subsequent steps involved ion exchange chromatography in a gravity-flow set-up and size exclusion chromatography at 4°C into 25 mM HEPES pH 7.5, 150 mM NaCl (additionally 1 mM DTT was used in case of $Ub^{K48C}$), resulting in the isolation of untagged ubiquitin and its mutants.

**Ubiquitin chains**

The di-Ub chains and higher-order Ub chains were generated using various methods.

K27, K29, and K33-linked di-Ubs were chemically synthesized by Dr. G. van der Heden van Noort using solid phase peptide synthesis.[214]

M1-linked di- and Tri-Ub was expressed as a linear fusion in *E. coli* and expressed and purified as described above.

K6, K11, K48, and K63-linked di-ubiquitins and longer chains were enzymatically assembled using tagless Ub.

K6-linked di-Ub was generated by incubating 2.5 mM Ub 0.1 µM E1, 0.6 µM UBE2L3, 10 µM NleL in 40 mM Tris-HCl pH 8.8, 10 mM $MgCl_2$, 1 mM DTT in the presence of 10 mM

ATP at 37°C for 3 hours. K48-linked ubiquitin chains that were assembled as side-product were eliminated by subsequent incubation with OTUB1.

K11-linked Ub chains were obtained using the E2-fusion Ube2S-UBA-IsoT.[185] 0.5 mM Ub were incubated with 0.25 µM E1, and 5 µM Ube2S-UBA-IsoT and ATP for 2 hours at 37°C.

K48-linked di-Ub was obtained by incubating 2.5 mM ubiquitin with E1 (1 µM), and UBE2R1 (25 µM) in the presence of 10 mM ATP for 3 h at 37°C. The reaction was quenched by adding 10 mM DTT. Subtle amounts of contaminating ubiquitin chains were removed by incubation with 1 µM AMSH.

To generate K63-linked di-Ub as well as longer K63 chains ($Ub_2$, $Ub_3$, $Ub_4$, $Ub_5$), 1 mM ubiquitin with 0.5 µM E1, 8 µM Ube2N and 8 µM Ube2V1 were incubated in 40 mM Tris-HCl pH 8.5, 10 mM $MgCl_2$, 0.5 mM DTT, and 10 mM ATP for 30 min at 37°C. The reaction was then stopped by addition of DTT to a final concentration of 10 mM.

Subsequently, different chain lengths for the various types of chains were separated through successive rounds of cation exchange chromatography, followed by size exclusion chromatography in a final buffer of 25 mM HEPES pH 7.5.

**UBR5 HECT constructs and UBA1**

Proteins derived from insect cells all were cloned to contain an N-terminal GST-tag and they were purified by gravity flow affinity purification using GST-resin (buffer B for wash, buffer B + 10 mM reduced glutathione for elution), followed by proteolytic cleavage of GST-fusion proteins employing His-tagged TEV-protease. Finally, ion exchange chromatography and size exclusion chromatography were conducted in buffer C.

**UBR5 and UBR5 mutants**

UBR5 variants expressed in mammalian cells all contained an N-terminal TwinStrep-tag followed by a GFP-tag and an HRV 3C cleavage site. Cell lysis and pre-clearance was performed as described earlier. The protein was isolated using Strep-affinity chromatography, followed by HRV 3C protease-mediated cleavage overnight and size exclusion chromatography. Varied UBR5 constructs were purified in distinct buffers tailored for specific experiments, including buffers containing either 1 mM DTT ($UBR5^{C2768A}$ and $UBR5^{Dimer}$ for which cryo-EM datasets of the apo-state were collected) or 1 mM TCEP for all other constructs as well as later $UBR5^{Dimer}$ purifications to allow reactivity with the activity-based probes. Despite rigorous efforts to prevent it, SDS-PAGE gel analysis revealed UBR5 degradation products after size exclusion chromatography, which retained some catalytic activity. Mass spectrometry confirmed these truncated species as UBR5. Figures denote the SDS-PAGE band(s) corresponding to these modified truncations with an asterisk (*).

**Deubiquitylating enzymes**

Various DUBs were used within this study to analyze ubiquitin chain properties. OTUB1, AMSH, and Cezanne were kindly provided by Dr. K. Swatek. The catalytic core of USP2 was expressed in *E. coli* with an N-terminal 6xHis-tag and a subsequent thrombin cleavage site. Protein expression, lysis, and Ni-affinity purification was performed as previously described. After Ni-affinity purification, ion exchange chromatography was performed and lastly, size exclusion chromatography yielded pure and active protein. The protein was not proteolytically digested.

**MYC-peptide substrates**

N-terminally rhodamine-labeled peptide substrates for MYC as well as the mutated MYC-peptide, both carrying two lysines at the C-terminus were synthesized in the core facility at MPIB with solid phase peptide synthesis using Fmoc chemistry (WT: Rho-GGGSSLYLQDLSAAASECIDPSVVFPKK-KK, Mut: Rho-GGGSSLYLQDVTAAATDSLDPSVVFPKK-KK). Peptide-ubiquitin fusions were expressed, purified, and fluorescently labeled by R. Vollrath and derived from GST-TEV-(MYC-peptide/mutant peptide)-KK-$Ub^{GGR}$-constructs. *E. coli* expression, cell lysis, purification involving GST-affinity chromatography, proteolytic digest, ion exchange chromatography, and size exclusion chromatography were performed as previously described. Fluorescent labeling of the obtained protein was achieved by utilizing the N-terminal Gly-remnant, left after proteolytic cleavage with the TEV-protease. Using a synthesized fluorescent peptide of TAMRA-$(PEG)_5$-LPETGG and mixing it in a 5 x excess with the purified protein (50 µM) as well as Sortase A (2.5 µM) in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM $CaCl_2$.

**N-degron substrates and short HECT constructs**

Various lengths of N-degron substrate peptides fused to ubiquitin were expressed as fusions of GST-SUMO-R(xxx)-$Ub^{\Delta GG}$, with the “xxx” representing the different additional residues and the fused ubiquitin lacking the ultimate two glycine-residues. The truncated HECT domain ($HECT^{\Delta N}$) was also expressed as GST-SUMO fusion. The constructs were expressed in *E. coli*, and initial purification steps including a GST-affinity capturing were performed accordingly. Proteolytic cleavage using the Senp2-protease was performed overnight on the resin. Cleaved protein could be washed off the resin and size exclusion chromatography resulted in pure protein.

**Ub-BmDPA**

The synthesis of Ub-BmDPA, a crucial element for our activity-based probes, facilitated the acquisition of structural representations showcasing the assemblies involved in transferring $Ub^D$ from E2 to E3, and transferring $Ub^D$ onto an acceptor ubiquitin. First, His-

Ub(1-75)-intein-chitin-binding domain (CBD) was expressed in *E. coli*, and cell lysis occurred in a buffer comprising 20 mM Tris-HCl pH 6.8, 50 mM NaOAc, 100 mM NaCl, and 2.5 mM PMSF. Purification of His-Ub(1-75)-intein-CBD involved Ni-NTA affinity chromatography. Importantly, the pH of the wash buffer as well as the elution buffer (20 mM Tris-HCl pH 6.8, 50 mM NaOAc, 100 mM NaCl, 300 mM imidazole) had to be adjusted to maintain a low pH of 6.8. Intein-based cleavage using 100 mM MESNa was induced subsequently. The resulting His-Ub(1-75)-MESNa underwent further purification via size exclusion chromatography in of 25 mM HEPES pH 6.8, 100 mM NaCl, 20 mM NaOAc. The ratio of MESNa-hydrolysis was evaluated using liquid chromatography coupled to mass spectrometry (LC-MS) with intact proteins, and was taken into account during subsequent procedures. Converting His-Ub(1-75)-MESNa into chemically reactive proxies involved modifying the thioester group. His-Ub(1-75)-MESNa (10 mg/mL) was coupled with 0.4 M (E)-3-[2-(bromomethyl)-1,3-dioxolan-2-yl]prop-2-en-1-amine (BmDPA) (ChiroBlock) in 1 mM N-hydroxysuccinimide, 10 % (v/v) DMSO and 50 mM HEPES pH 6.8. Following overnight incubation at 30°C, 300 rpm, desalting into 25 mM HEPES pH 6.8, 100 mM NaCl confirmed completion of the reaction through LC-MS. Subsequently, deprotection of the product involved incubation at a concentration of ~1 mg/mL in 40 mM p-Toluenesulfonic acid and 54 % (v/v) TFA for 1 hour at room temperature. TFA removal included multiple washes with ice-cold diethyl ether. The resulting Ub flakes were air-dried and resuspended in 100 mM $Na_2HPO_4$ pH 6.0, 500 mM NaCl, 8 M urea, followed by protein refolding via overnight dialysis in 20 mM $Na_2HPO_4$ pH 6.0 and 100 mM NaCl at 4°C.

**Fluorescently-labeled UBE2D2$^{AAS}$**

Since a mass shift on an SDS-PAGE is not well suited to judge reactivity of an ABP with UBR5 due to UBR5's large size, we aimed to visualize this reactivity otherwise. Our goal was to incorporate fluorescence for visualization, specifically enabling the detection of probe conjugates using a Typhoon Scanner. To achieve this, our strategy involved designing a fluorescein-labeled LPETGG peptide, incorporating a PEG5 linker between the fluorophore and the sortase recognition sequence (CF-PEG5-LPETGG, synthesized by the MPIB core facility). Utilizing the Gly-Ser remnant present at the N-terminus of UBE2D2$^{C21A,C107A,C111S}$ that remains after the proteolytic TEV-cleavage, we exploited this sequence recognized by sortase A as a substrate for the transpeptidation reaction with the labeled peptide. Consequently, an incubation process involved a 6x fold molar excess of labeled-peptide (300 µM) with 50 µM UBE2D2 and 5 µM His-tagged sortase A for 1 hour at room temperature in a solution comprising 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 10 mM $CaCl_2$. The subsequent purification steps entailed the removal of Sortase A using its affinity to Ni-NTA, while collecting the flow-through containing the E2 enzyme. Further purification of the labeled E2 enzyme was accomplished through size exclusion chromatography in 50 mM HEPES pH 7.5, 150 mM NaCl.

## Biophysical characterization

### HPLC

A standard curve was generated using a protein mixture of Thyroglobulin (670 kDa), γ-globulin (158 kDa), Ovalbumin (44 kDa), Myoglobin (17 kDa), and Vitamin B12 (1.35 kDa) and subjecting it to HPLC at ~1.5 mg/mL total protein concentration onto a Superose 6 column at 0.15 mL/min.

UBR5 was diluted to a final concentration of 1 mg/mL. The buffer UBR5 was diluted into was also used as running buffer and contained different concentrations of salt (25 mM HEPES pH 7.5, 5 mM DTT, 100 / 200 / 300 / 500 mM NaCl) (**Figure 13a**), different concentrations of reducing agent (25 mM HEPES pH 7.5, 200 mM NaCl, 1 / 5 mM DTT) (**Figure 13b**), or UBR5 was stored in different conditions (25 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT) (**Figure 13c**).

### NanoDSF

NanoDSF was performed to evaluate the stability of recombinantly expressed and purified $UBR5^{WT}$ (**Figure 13d**).[178] Two samples of pure protein diluted to 25 µg/mL and 120 µg/mL in a buffer of 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM DTT were subjected to duplicate measurements using a NanoTemper Prometheus. Due to the subtle signal of the lower concentrated sample, laser intensity had to be increased by 25 % compared to the other sample to receive a reliable signal. Thermal unfolding was induced by gradually increasing the sample temperature from 20°C to 90°C and the obtained red shift of tryptophane-fluorescence was detected by fluorescence measurements at 350 and 330 nm.

### Mass photometry

The determination of UBR5's oligomeric state involved conducting mass photometry measurements using the Refeyn TwoMP and the Refeyn AcquireMP 2.3.0 software (**Figure 16a**).[182] A mass calibration curve was generated using a protein mixture encompassing a molecular mass spectrum—Conalbumin (75 kDa), Aldolase (158 kDa), Ferritin (440 kDa), and Thyroglobulin (669 kDa). The mix was measured with each component at a final concentration of approximately 50 nM. Subsequent measurements were carried out on either UBR5 or $UBR5^{Dimer}$, with UBR5 diluted to a final concentration of approximately 140 nM in the buffer identical to that used for focus-finding (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT). Movies were captured for a duration of one minute. Data analysis was conducted using the DiscoverMP 2.3.0 (Refeyn) software, referencing the collected mass calibration for accurate assessment.

## General biochemical characterization

### Polyubiquitylation for UbiCRest

To analyze what ubiquitin chains a UBR5 formed, chain formation was followed by treatment with various linkage-specific DUBs that would break down chains of distinct linkages (UbiCRest).[191] First, the reaction conditions for these DUBs were tested (**Figure 19c**). To ensure reduction and therefore activity of the DUBs, they were first diluted into a buffer containing 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM DTT. After incubation at 37°C for 10 min, the DUBs were mixed at a final concentration of 1 µM with 20 µM of the respective di-ubiquitin (K48 for OTUB1, K11 for Cezanne, K63 for AMSH) in 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM DTT. The samples were incubated at 37°C for the indicated duration and progression of the cleavage was analyzed using SDS-PAGE and subsequent Coomassie Brilliant Blue staining. Due to the comparably low reactivity of AMSH, this sample was omitted in the actual assay.

For the UbiCRest-assay, UBR5-mediated polyubiquitylation was first required. A multi-turnover assay with 0.5 µM E1, 5 µM E2, 30 µM wildtype unlabeled ubiquitin, and 1 µM UBR5 or UBR5$^{\Delta MLLE}$ with 5 mM ATP and 10 mM $MgCl_2$ was prepared and incubated for 30 min at room temperature. The reaction was quenched with EDTA and subsequently, 2 µM Usp2, OTUB1, or Cezanne were added before incubating it at 37°C for 1 h. Again, progression of the reaction was visualized using SDS-PAGE and Coomassie Brilliant Blue staining (**Figure 19d**).

### Polyubiquitylation for AQUA-MS

A multi-turnover assay was performed to generate ubiquitin chains, which could then be analyzed for their linkages using AQUA-MS (**Figure 13g,h**).[180] 0.5 µM E1, 2 µM E2, and 0.5 µM UBR5 were incubated with 50 µM wildtype ubiquitin in a buffer containing 3 mM ATP, and 10 mM $MgCl_2$ for 60 min at room temperature. Reactivity of the E1 was subsequently quenched by adding EDTA. Subsequent sample digest and MS-measurement was performed by Dr. F. Hansen and D. Tung Vu.

### Polyubiquitylation for cross-link MS

To prepare a sample with UBR5 “in-action” for subsequent cross-link MS analysis, 0.5 µM E1, 2 µM E2, and 2 µM UBR5 were incubated with 20 µM unlabeled wildtype ubiquitin in a buffer containing 2 mM ATP, and 10 mM $MgCl_2$. After incubating the sample for 30 min at room temperature, cross-linking was induced by adding 0.1 mM BS3. Cross-linking was performed for 30 min at room temperature and then quenched by addition of 50 mM Tris-HCl pH 7.5 Subsequent sample digest, measurement, and initial data analysis was performed by Dr. B. Steigenberger in the MPIB mass spectrometry core facility (**Figure 23**).[195]

In brief, the samples were denatured, reduced, and alkylated and subsequently proteolytically digested. Peptides were applied to liquid chromatography, they were fractionated and then measured using an Exploris 480 or a QExactive HF mass

spectrometer (Thermo Fisher Scientific) with distinct settings for data acquisition of crosslinked peptides. The raw data were initially analyzed using Proteome Discoverer. In-depth data analysis and visualization was performed using xiView.[215]

## Cryo-EM: Sample preparation, data collection and processing

Cryo-EM grids were prepared by applying the sample to freshly glow-discharged R1.2/1.3, Cu 200 mesh holey carbon grids (Quantifoil). Plunge-freezing of the grids into liquid ethane was conducted using a Vitrobot Mark IV. Cryo-EM of all described samples was performed on a Glacios, Arctica, or Titan Krios transmission electron microscope (TEM). Data acquisition was set up using S erialEM v.3.8.0-b5 and FEI EPU v2.7.0.

### Tetrameric UBR5

In the process of this study, obtaining purified catalytically-inactive UBR5$^{C2768A}$ protein preceded a large-scale purification of WT UBR5. Consequently, we opted to proceed with cryo-electron microscopy using UBR5$^{C2768A}$ to provide initial guidance for subsequent investigations. First attempts to gain structural insight into UBR5 using cryo-EM revealed several difficulties: Only low concentrations of UBR5 could be used for cryo-EM, as UBR5 would otherwise aggregate. Additionally, oligomeric UBR5 seemed to fall apart in the air-ice interface, making it difficult to obtain a homogeneous dataset. Furthermore, initial reconstructions revealed strong effects caused by preferred orientations. To overcome these issues, we tested different detergents to help acquire a homogeneous dataset of UBR5$^{C2768A}$ portraying intact UBR5 and revealing different orientations of UBR5.

### Detergent screen

Different types of detergents were used to determine which one yields the most informative dataset (**Figure 14**).

Fluorinated Octyl Maltoside (FOM) was used at its critical micelle concentration of 0.7 mM. UBR5 at a final concentration of 0.6 mg/mL was supplemented with FOM several minutes before applying the sample to the grids and plunge-freezing them. Screening of the grids was performed using a Glacios electron microscope. Significant FOM-induced particle distribution effects were clearly visible and therefore no dataset was collected.

n-Octyl-β-D-glucopyranoside (β-OG) was also tested for positive effects on UBR5 particle distribution, stability, and orientations. A final concentration of 0.1 % (w/v) of β-OG was added to UBR5 at 3.5 mg/mL before plunging. In the presence of β-OG, UBR5 was well distributed and a majority of particles seemed to be intact. Analysis of the angular distribution of an initial 3D-reconstruction revealed slightly more orientations of UBR5, however UBR5 would still mostly lay flat in the ice and just be slightly tilted, hampering a high-resolution reconstruction.

Lastly, we tested effects of 3-[(3-Cholamidopropyl)-dimethylammonio]-2-hydroxy-1-propansulfonat (CHAPSO) on UBR5. Again, CHAPSO at a final concentration of 8 mM was added to UBR5 shortly before plunging (2.5 mg/mL). The particles seemed to be sparsely distributed and mostly were pushed onto the carbon backbone. Nevertheless, the particles seemed to engage more distinct conformations. A small dataset revealed many novel orientations already in initial 2D-classifications, which was not achieved with β-OG. Despite the lower particle number, the angular distribution revealed slightly more views compared to the prior sample. Therefore, a larger dataset of $UBR5^{C2768A}$ with CHAPSO was collected.

**$UBR5^{C2768A}$ data collection**

A larger dataset was collected using a Titan Krios microscope operating at 300 kV with a Gatan K3 Summit direct electron detector in counting mode. 10,091 micrographs were collected at a nominal magnification of 64,000 x, a pixel size of 1.384 Å/pixel, a target defocus range of -2.4 and -0.8 µm and total exposure of 56.28 $e^-/Å^2$. Pre-processing consisted of MotionCorr2 for alignment and dose-weighing, GCTF was used for Contrast Transfer Function (CTF) estimation. [216,217] Particle picking was performed using template-based picking with Gautomatch (K. Zhang, MRC Laboratory of Molecular Biology, Cambridge). Subsequent processing was performed using RELION 3.1.1 for particle selection, extensive 2D- and 3D-classification, refinement, CTF refinement, particle polishing, and post-processing. An initial reconstruction that derived from the small screening dataset of $UBR5^{C2768A}$ was used to generate templates for picking and was used as initial model. Binning of particles was progressively lifted during 3D classifications. During the final refinement, C2 symmetry was applied, and the resulting map underwent sharpening using DeepEMhancer or post-processing, yielding a map resolution of 3.7 Å.[218] Notably, the intrinsic flexibility of the dimeric units affected the quality of the lower portion of the map. Comparing different 3D classes revealed breathing motions between the upper and lower dimeric units.

**$UBR5^{Dimer}$**

The L710D mutant of UBR5, termed $UBR5^{Dimer}$, was treated with n-Octyl-β-D-Glucopyranoside (β-OG) at a concentration of 0.1 % (w/v) before vitrification at 1.3 mg/mL. Data was collected at 105,000 x magnification, with a pixel size of 0.8512 Å/pixel, using a target defocus range of -3.0 to -0.5 µm and a total exposure of 67.8 $e^-/Å^2$. 21,270 micrographs were collected and subjected to alignment and dose-weighing as described earlier. An initial model derived from a screening dataset obtained from the Glacios microscope for $UBR5^{Dimer}$ was used to generate templates for particle picking by Gautomatch. 762,722 particles were picked and further processed using RELION 4.0.[219] 2D classifications were performed as first step to clean up the dataset. The Glacios-derived reconstruction also served as initial 3D reference for a first 3D refinement. 3D classification, two iterative rounds of CTF refinement, and particle polishing yielded a 2.7

Å resolution map. This was obtained by applying C2 symmetry during 3D refinement and map-sharpening using post-processing or DeepEMhancer.

Converting UBR5 to a dimeric state significantly alleviated preferred orientations. The final refined particles from RELION were transferred to CryoSparc4.2.0 to enhance density around specific regions like the RLD and DSD domains. Non-uniform and local refinements were performed, focusing on the RLD, a section of the scaffold, and the proximal HECT domain.[220,221] Utilizing the local refinement map in CryoSparc, DeepEMhancer was employed to generate a map with a resolution of 2.98 Å. While the overall resolution is slightly lower, the local resolutions within the RLD-region exhibit significant improvement compared to the entire map.

**Transition state 1: UBR5$^{Dimer}$~Ub$^{D}$~UBE2D2**

Since structural investigation of apo UBR5$^{Dimer}$ did not explain how exactly an E2~Ub moiety could bind to UBR5, we sought to mimic this state and more precisely, the transfer of Ub$^{D}$ from the E2's catalytic cysteine to UBR5's catalytic cysteine using a stable mimic.[192] Generation of this activity-based probe as well as subsequent sample preparation and cryo-EM was performed in close collaboration with Dr. D. Horn-Ghetko.

**Generating the TS1-ABP**

To generate the activity-based probe, a UBE2D2-variant with all cysteines apart from the catalytic cysteine being mutated (C21A, C107A, C111S) was generated.[193] The E2 was incubated with 1 mM TCEP for 30 minutes at room temperature to ensure reduction of the catalytic cysteine. Directly prior to reacting the E2 with Ub-BmDPA, the E2 was desalted into buffer without any reducing agent. Approximately 100 µM of Ub-BmDPA were reacted with a 5 x molar excess of the reduced and desalted E2 for 2 h at 30˚C. Nickel affinity of Ub-BmDPA and the reacted E2~Ub moieties was used to remove excess unreacted E2. Subsequent size exclusion chromatography into 25 mM HEPES pH 7.5, 150 mM NaCl yielded relatively pure E2~Ub$^{D}$-ABP.

To obtain a fluorescent version of this ABP, which could be used to test reactivity as well as specificity, fluorescein-labeled UBE2D2 was used instead of unlabeled UBE2D2. To generate an ABP with this E2, the molar excess of E2 and Ub-BmDPA was reversed due to the limited yield of fluorescent E2. 50 µM of labeled UBE2D2 were reduced and desalted as described above before being incubated with a 5 x molar excess of Ub-BmDPA for 2 h at 30˚C. The obtained reaction product was purified using size exclusion chromatography in 50 mM HEPES pH 7.5, 150 mM NaCl.

### Generating UBR5$^{Dimer}$~Ub$^{D}$~UBE2D2

To evaluate whether the TS1-ABP would react dependent on UBR5's catalytic cysteine, we tested whether the fluorescently labeled TS1-ABP prefers with UBR5 and with UBR5$^{C2768A}$ (**Figure 20d**). Both UBR5 versions were incubated with a 5 x molar excess of the fluorescent TS1-ABP and reacted at room temperature for the indicated time points. SDS-PAGE analysis and scanning of the in-gel fluorescence showed the progression of the probe reaction. Base-level fluorescence of UBR5$^{C2768A}$ also in the absence of the ABP suggests remnants of uncleaved GFP-tagged UBR5.

To generate the TS1-complex for cryo-EM, UBR5$^{Dimer}$ was incubated with equimolar amounts of K63-linked tetra-ubiquitin as well as 2 x molar excess of the UBE2D2~Ub$^{D}$ probe for 2 h on ice. Without any further purification, the sample was supplemented with β-OG and plunged at 2 mg/mL as described previously.

### Data collection and processing of TS1

Data collection of the TS1-sample was performed on an Arctica electron microscope, which is equipped with a Falcon III electron detector in linear mode. 40 frames at a nominal magnification of 73,000 x and a pixel size of 1.997 Å/pixel were collected for each micrograph with a total exposure of ~70 $e^{-}$/Å$^{2}$. Data collection was performed with a target defocus of -3.5 and -1.0 μm and 1,740 micrographs were collected in total.

Processing of this dataset was performed using RELION 4.0. Motion correction and estimation of the contrast transfer function using CTFFIND-4.1 were performed for pre-processing.[222] A low-pass filtered model of UBR5$^{Dimer}$ was used for template picking with Gautomatch and also as initial model. 1.4 million particles were subjected to extensive 3D classification. Several iterations of classifications revealed one HECT domain in an inverted T-conformation with extra density, presumably incorporating ubiquitin, next to the C-lobe and additional density next to the N-lobe presumably corresponding to the E2. Significant sample heterogeneity, do not allow to properly resolve both HECT domains, but only one can be resolved to an extend that allows fitting the HECT domain, the E2, and ubiquitin.

A final reconstruction at 7.3 Å was obtained and sharpened using DeepEMhancer.

### Intermediate state: UBR5$^{Dimer}$~Ub$^{D}$

Prior published structures aiming to describe the intermediate state of a ubiquitin-bound HECT E3 ligase, are controversial and propose somewhat different mechanisms. Therefore, we aimed to investigate this state in the context of a full-length HECT E3 ligase. To do so, Ub-VME that was synthesized by Dr. M. Mulder was reacted with UBR5$^{Dimer}$

since this would retain the native distance. [196,214,223] Sample preparation as well as cryo-EM of this state again was achieved collaboratively with Dr. D. Horn-Ghetko.

### Generating $UBR5^{Dimer}$~$Ub^{D}$

As with the TS1-ABP, we sought to test whether Ub-VME would react with UBR5 dependent on UBR5's catalytic cysteine (**Figure 24c**). A fluorescently-labeled rhodamine-Ub-VME was incubated at 5 x molar excess with UBR5 or $UBR5^{C2768A}$ at room temperature for the respective time. SDS-PAGE with subsequent in-gel fluorescence then revealed the reacted species. Again, background signal of UBR5 and $UBR5^{C2768A}$ even in the absence of the probe indicate incompletely cleaved GFP-tagged UBR5.

Sample preparation for cryo-EM was performed by incubating $UBR5^{Dimer}$ with equimolar K63-linked tetra-Ub and a 10 x molar excess of Ub-VME for 2 h at room temperature. Subsequent size exclusion chromatography into 25 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP was performed to remove excess probe and ubiquitin chains. A sample concentrated to 1.5 mg/mL was supplemented with CHAPSO directly before plunge-freezing as described earlier.

### Data collection and processing of E3~$Ub^{D}$ intermediate state

Using a Glacios electron microscope with a K2 Summit direct electron detector in counting mode, a dataset was collected. Data collection was performed at a nominal magnification of 22,000 x with a pixel size of 1.885 Å/pixel and a total exposure of ~60 $e^-/Å^2$ distributed over 40 frames. A target defocus of -2.6 and -0.8 μm was used to collect 1,808 micrographs. RELION 4.0 was used for processing of the obtained data. Template-free particle picking with Gautomatch was performed after motion correction, dose weighing and estimation of the contrast transfer function using CTFFIND-4.1. ~835,000 particles were picked and 3D classification was performed with a low-pass filtered model of $UBR5^{Dimer}$ serving as initial reference. Significant density next to the C-lobes could be identified early on, however, the resolution of this density was considerably lower compared to the rest of the map, indicating flexibility of this region. 3D variability analysis implemented in CryoSparc was used to visualize this flexibility (**Figure 24h**).[197] Default parameters were used with the maps being low-pass filtered to 9 Å.

A final 3D reconstruction at 5.3 Å could be obtained, allowing to confidently fit UBR5 as well as a C-lobe-bound ubiquitin.

### Transition state 2: $UBR5^{Dimer}$~$Ub^{D}$~$Ub^{A}$

In order to understand HECT-mediated polyubiquitylation and to gain insight into the linkage-specificity of HECT E3 ligases, we attempted to visualize the step of a di-ubiquitin formation mediated by UBR5. For this, we designed a di-ubiquitin probe that mimics a

K48-linked chain. The employed probe retains native geometry between UBR5's catalytic cysteine, the α-carbon of the acceptor ubiquitin and the C-terminus of a donor ubiquitin. This stable mimic represents the fleeting transition state 2: the isopeptide bond formation between $Ub^D$ and $Ub^A$, mediated by UBR5.

**Generating the TS2-ABP**

A stable proxy of TS2 was generated with $Ub^D$'s C-terminus, $Ub^A$'s target side chain (natively K48, mutated to cysteine), and UBR5's catalytic cysteine all connected at one atom.[199] To obtain this probe, Ub-BmDPA was used to mimic $Ub^D$, and in a first step, this was reacted with an acceptor ubiquitin. $Ub^A$ was incubated with 1 mM TCEP directly prior to the trap formation to ensure complete reduction of the K48C mutation. $Ub^A$ was desalted into 25 mM HEPES pH 7.5 and 150 mM NaCl and was incubated with a 5 x molar excess of Ub-BmDPA for 1 h at 30°C. Excess unreacted protein was removed by size exclusion chromatography with the final buffer being 25 mM HEPES pH 7.5, 150 mM NaCl.

**Probe reactivity and geometry**

First, we wanted to assess whether the obtained TS2-probe is reactive and whether it maintains a native geometry. Since these probes were originally developed to investigate linkage-specificity of different DUBs, we reacted our TS2-ABP with two DUBs containing different linkage-specificities (**Figure 25d**). For this, we incubated the probe in a 2 x molar excess with either the K48-specific DUB OTUB1, or the K63-specific AMSH. Samples were taken after 1 h incubation at 30°C and applied to an SDS-PAGE for subsequent Western blotting. The $Ub^D$-moiety of the ABP contains an N-terminal His-tag, and therefore blotting against this tag can be used to visualize any species that reacted with this probe. As anticipated based on the probe-design, the ABP specifically reacts with OTUB1, but not with AMSH.

To generate the $UBR5^{Dimer}$~$Ub^D$~$Ub^A$ complex for cryo-EM, we incubated $UBR5^{Dimer}$ with a ~50 x molar excess of the TS2-ABP in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP for 2 h at room temperature. Using size exclusion chromatography with the same buffer, we could remove excess ABP. The peak fractions were concentrated to 0.6 mg/mL and supplemented with CHAPSO before being plunge-frozen.

**Data collection and processing of TS1**

An initial small dataset of 705 micrographs was collected on a Glacios screening microscope with a target defocus of -3.0 and -0.3 μm, a pixel size of 1.885 with a nominal magnification of 22,000 x, and a total exposure of ~60 $e^-/Å^2$ distributed on 40 frames. RELION 3.1.1 was used for pre-processing using MotionCorr2 and dose weighing as well as estimation of the contrast transfer function with GCTF. Template-based particle picking with Gautomatch and subsequent 2D and 3D classifications yielded a 3D reconstruction

at 8.3 Å, showing significant density next to the C-lobe presumably corresponding to $Ub^D$, $Ub^A$, and the UBA domain. To improve map quality, a larger dataset was collected using a Titan Krios electron microscope. A nominal magnification of 105,000 x with a pixel size of 0.8512 Å/pixel was used to collect 17,689 micrographs with a defocus range of -2.2 to -0.6 µM. Again, template-based picking with Gautomatch was performed and 1.7 million particles could be extracted and subjected to 2D and 3D classifications using RELION 3.1.1. Focused classifications with a mask covering the HECT domain, $Ub^A$, $Ub^D$, and the neighboring UBA domain. The obtained classified particles were then imported into CryoSparc for non-uniform and local refinement. A final reconstruction at 3.3 Å resolution was obtained and sharpened using DeepEMhancer.

**Model building**

Structure building of $UBR5^{Dimer}$ and UBR5~$Ub^D$~$Ub^A$ (TS2) was performed together with Dr. R. Prabu.

**Model for tetrameric UBR5**

Using AlphaFold2, an initial model of UBR5 was generated.[181] Splitting the model into smaller parts of 1) the N-terminal region containing RLD and dSBB domains, 2) the α-helical scaffold, and 3) the HECT domain, allowed to dock large parts of the folded UBR5 regions into the density with UCSF Chimera.[224] Analysis of the docked parts in the density map allowed identification of the residue L710 to be suitable for mutagenesis.

**Structure for dimeric UBR5**

The high-resolution density of $UBR5^{Dimer}$ enabled detailed construction of the protein backbone along with most side chains using COOT.[225] However, due to the lower resolution in the dSBB domains, only one barrel derived from the AlphaFold2-model could be docked with confidence and barely any density was resolved for the second barrel.

How the residues 1523-1773 are connected at the heterodimerization interface, could not be unambiguously determined. We therefore split the UBR5-structure into separate chains to avoid false annotation of distinct regions. Nevertheless, in agreement with simultaneously published UBR5-structures from other groups and with AlphaFold2 predictions, we anticipate the protomers to engage in an “S-model” and illustrate it like this throughout the study (**Figure 17c**).

Better map quality in the focused $UBR5^{Dimer}$ density around the RLD and DSD domains facilitated building of these specific regions. The missing density for the connection of the DSD domain to the scaffold allowed for it to potentially originate from either monomer, hence it was kept as a separate chain. However, closer spatial proximity and the published

structures from other groups suggest that the DSD domain likely originates from one monomer and integrates into the other to reinforce the dimerization-interface.

Initially, a two-fold symmetry was applied to the structure of a monomer to derive the dimeric structure during early refinement cycles. Finally, multiple rounds of real-space refinement were conducted using PHENIX v1.19.2.[226] Using Molprobity v.4.2, we validated the resultant atomic model of $UBR5^{Dimer}$.[227]

**Model for TS1**

A model for UBR5~$Ub^{D}$~UBE2D2 was generated by splitting the structure of $UBR5^{Dimer}$ into several parts: The N-terminal region as well as the scaffold could be easily docked into the density. The HD domain and the HECT domain N-lobe had to be tilted compared to the apo-structure. It has to be noted that the rather small HD domain can not be fit into the low-resolution density unambiguously and we can't exclude that the HD domain engages slightly different. The HECT domain's C-lobe has to be massively rearranged to be docked into the respective density. The E2~$Ub^{D}$ moiety was extracted from a prior crystal structure (PDB: 3JVZ) of NEDD4L's HECT domain bound to UBE2D2~Ub. Docking of the respective parts into the density was performed using UCSF Chimera. It has to be noted that UBE2D2 from the published crystal structure has the catalytic cysteine mutated to serine for oxyester bond formation. UBE2D2 used in our study however, has the catalytic cysteine in its original form, however, the three other cysteines are mutated (C21A/C107A/C111S).

As mentioned above, only one of the two HECT domains in our 3D reconstruction shows clear density and the other HECT domain presumably is a mix of L- and inverted T-conformations. Therefore, the described model only applies to HECT domain 1 in dimeric UBR5 and we can't make any statements for the respective other HECT domain.

**Model for UBR5~$Ub^{D}$ intermediate state**

A model portraying $UBR5^{Dimer}$~$Ub^{D}$ was generated using our structure of TS2. The HECT~$Ub^{D}$ moiety of TS2 fit very nicely into the density of the UBR5~$Ub^{D}$ intermediate state. These parts were docked into the density of the intermediate state alongside the remaining parts of UBR5, which were used from the apo $UBR5^{Dimer}$ structure. Due to the lower resolution of the intermediate state map, we can't assign density to the ultimate C-terminus of UBR5 and also not to the C-terminus of $Ub^{D}$. For this reason, these parts were truncated for the final model of $UBR5^{Dimer}$~$Ub^{D}$.

**Structure for TS2**

Structure building of UBR5's TS2 was based on docking a preliminary model of a donor ubiquitin on UBR5's C-lobe (considering the published crystal structure of Ub-bound Rsp5

HECT domain PDB: 4LCD) and docking a crystal structure of UBR5's UBA domain bound to ubiquitin (PDB: 2QHO) into the focus refined density. Using COOT, the HECT domain could then be built mostly on a side chain level. The structure of $Ub^A$ was edited to contain the point mutant K48C, which was used to generate the ABP. Clear density for UBR5's residues 2796-2798 (until the penultimate residue) guides the backbone of the C-terminal tail, however, the side chain positions are somewhat ambiguous due to smeared density (especially for F2796). This might indicate heterogeneity of the side chain conformation. Since no density could be identified for the ultimate residue V2799, also the side chain conformation for F2798 can't be determined unambiguously as the density could also harbor the ultimate residue rather than the side chain of F2798. Similarity with a published structure of HUWE1's HECT domain bound to a donor ubiquitin (PDB: 6XZ1), supported our assignment of the density as side chain density. Additionally, the observed biochemical effects also align with this configuration of UBR5's C-terminus. Side chains of the donor ubiquitins R72 and R74 were slightly displaced to accommodate UBR5's C-terminus. A $Ub^A$ R54 rotamer slightly differing from the cryo-EM density was built in consideration of observed biochemical effects. However, future studies will have to determine the precise location of this side chain. Knowing the physical nature and geometry of the employed ABP allowed building the probe without seeing the high-resolution density for it. A three-way cross link connecting UBR5's C2768, $Ub^D$'s G75, and $Ub^A$'s C48 was built.

Several iterations of real-space refinement in PHENIX were performed and the atomic model was validated using Molprobity.

### (Structure-based) biochemical assays

To evaluate activity of various UBR5 constructs, investigate how well a certain ubiquitin could accept another ubiquitin to form a di-ubiquitin or a longer chain, or whether UBR5 could use a distinct substrate and modify it with ubiquitin, was tested in these biochemical assays. Generally, within these assays a donor ubiquitin ($Ub^D$) is loaded onto an E1 enzyme in the presence of ATP. From there it gets passed on to the E2 enzyme by transthiolation. Next, in another transthiolation reaction, $Ub^D$ gets transferred to UBR5's catalytic cysteine before being finally passed on to an acceptor ubiquitin ($Ub^A$) (chain formation), a substrate (substrate ubiquitylation), or a lysine on UBR5 itself (autoubiquitylation).
In most cases, fluorescently labeled $Ub^D$ was used, whose progression through the cascade could be tracked by different migration properties of the respective intermediates on an SDS-PAGE. Two different sets of SDS-PAGE were performed: Non-reducing SDS-PAGE, where the sample was mixed at the respective time point with a denaturing Laemmli buffer that is not reducing and reducing SDS-PAGE in which the sample had been mixed with a denaturing and reducing Laemmli buffer (100 mM DTT as final concentration). Non-reducing SDS-PAGE would reveal all $Ub^D$ reaction products that are

connected via isopeptide bonds (ubiquitylated substrate, formed ubiquitin chains, autoubiquitylated UBR5) as well as all intermediate products linked via a thioester bond (E1~$Ub^D$, E2~$Ub^D$, UBR5~$Ub^D$). Isopeptide bonds will be indicated with dashes whereas thioester bonds will be denoted with a tilde.
In reducing SDS-PAGE however, no thioester bound intermediates will be visible.
For most assays, the fluorescent scans of the SDS-PAGE are cropped into two parts for illustrative reasons. These parts of the same scan are connected on the side for clarity.

## Polyubiquitylation in pulse-chase format

### E2 preference

As a first characterization of the catalytic activity of UBR5, we aimed to identify the most compatible E2 enzyme to work with UBR5. A set of E2 enzymes was tested by monitoring the transthiolation activity of $Ub^D$ from E2 to E3 and analyzing the amount of ubiquitin chains formed by UBR5 subsequently (**Figure 13e**). A pulse-chase assay was employed to avoid effects of different kinetics in E2~$Ub^D$-loading. Fluorescently-labeled wildtype ubiquitin was used for this assay. In a pulse-reaction, 15 µM fluorescent ubiquitin were incubated with 15 µM of the respective E2 (UBE2A, UBE2Bm UBE2D1, UBE2D2, UBE2D3, UBE2L3, UBE2S), and 0.3 µM UBA1 in a buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl, 2.5 mM $MgCl_2$, and 1 mM ATP for 30 min at room temperature. The reaction was subsequently quenched by addition of 50 mM EDTA final concentration, and the pulse-mix was diluted to a final concentration of 2 µM into the chase-mix with final concentrations of 0.2 µM UBR5. Samples were taken at the indicated time points and mixed with reducing or non-reducing Laemmli buffer.

### UBR5 vs $UBR5^{Dimer}$

To evaluate whether $UBR5^{Dimer}$ might have any defects during polyubiquitylation compared to wildtype UBR5, a polyubiquitylation assay in a pulse-chase format was employed (**Figure 16c**). The pulse-reaction was performed by incubating 30 µM fluoresceine-labeled wildtype ubiquitin with 20 µM UBE2D2, and 0.5 µM UBA1 in 25 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 2 mM ATP, and 0.04 mg/mL BSA ("pulse-buffer") for 30 min at room temperature.[179] 50 mM EDTA were added to stop the reaction and the mix was then diluted to a final concentration of 1 µM into the chase-reaction with either UBR5 or $UBR5^{Dimer}$ at a final concentration of 0.2 µM. Samples were taken as indicated and mixed with reducing Laemmli buffer.

## Di-ubiquitin synthesis assays (pulse-chase format)

Pulse-chase assays were performed to examine effects of different UBR5 and $Ub^A$ mutations. For the di-ubiquitin synthesis assays, fluorescently labeled $Ub^D$ that carried a K48R mutation to prevent its use as acceptor and unlabeled $Ub^A$ were used if not stated otherwise.
During the pulse-reaction, 30 µM fluorescent donor $Ub^{K48R}$ were incubated with 20 µM of UBE2D2 (unless stated otherwise), and 0.5 µM UBA1 in pulse-buffer for 30 min at room

temperature. Subsequently, the pulse-reaction was quenched by addition of 50 mM EDTA to complex the $Mg^{2+}$ ions and stop the E1-loading reaction (pulse-mix). This was followed by a chase-reaction. In the chase-reaction, the respective UBR5 construct premixed with the indicated $Ub^A$ (wildtype if not indicated otherwise) were combined (chase-mix) and the pulse-mix was added (with E2~$Ub^D$). Unless stated otherwise, this was done with a final concentration of 0.2 µM E3, 2 µM $Ub^A$, and 0.2 µM E2~$Ub^D$ (pulse-mix was diluted 100 x into chase-mix) in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT. Samples were taken at the indicated time and supplemented with non-reducing or reducing Laemmli buffer respectively.

**UBR5 and $Ub^A$ variants**

Various UBR5 point mutations were tested either for tetrameric UBR5 (wildtype background), or for dimeric UBR5 (L710D mutant background). Additionally, different acceptor ubiquitin variations were tested. In some cases, UBR5 variations were combined with $Ub^A$ mutations to test for additive or compensatory effects. Non-reducing SDS-PAGE are shown for all of the following assays.

Whether the observed ubiquitin chain forming activity is dependent on UBR5's catalytic cysteine was tested using tetrameric UBR5 or UBR5$^{C2768A}$ in the presence of wildtype $Ub^A$-6xHis (**Figure 13f**), however the donor ubiquitin used for this particular assay was fluorescently labeled K0 ubiquitin (all lysines mutated to arginine) rather than K48R ubiquitin. Effects of the C-lobe-$Ub^D$ interaction by mutating A2790 to Trp (**Figure 22b**), as well as potential effects of an additional $Ub^D$ binding site during the E2-E3 transthiolation tested by mutating the SDA (H1362A / L1363A / L1364A) (**Figure 22d**) were examined with wildtype $Ub^A$-6xHis present in the chase-mix. Effects of mutating UBR5's C-terminus (F2796A, F2798A, Δ2799) were tested with tetrameric UBR5 and wildtype $Ub^A$-6xHis present in the chase-mix (**Figure 27b**). Whether catalysis would be impaired by disturbing the $Ub^D$-$Ub^A$-interface was investigated using different versions of $Ub^A$-6xHis in the chase-reaction alongside wildtype tetrameric UBR5: $Ub^A$-versions carrying D58A or D58R mutations were tested (**Figure 27d**). Effects of mutating the LOL (D2283A/E2284A/G2285A/P2286A/E2287A) were investigated with tetrameric UBR5 or UBR5 with the respective mutations as well as wildtype $Ub^A$-6xHis (**Figure 28a**). One residue of the LOL was tested more particular for specific interactions with $Ub^A$: in a rescue experiment, dimeric UBR5 or dimeric UBR5 carrying the E2287A mutation were combined with wildtype $Ub^A$-6xHis or $Ub^A$-6xHis carrying an R54E mutation (**Figure 28d**, left). Additionally, an E2287R mutation of UBR5 was tested in the tetrameric background: wildtype tetrameric UBR5 or tetrameric UBR5 with the E2287R mutation were mixed with wildtype $Ub^A$-6xHis or $Ub^A$-6xHis with the R54E mutation (**Figure 28d**, right). Various mutations on both, the UBR5 C-lobe (Y2773D and Y2773F in dimeric UBR5) as well as the acceptor ubiquitin in $Ub^A$-6xHis (wildtype, A46F, or A46D) were tested for compensatory rescuing effects (**Figure 29a**). A reducing SDS-PAGE was prepared when testing compensatory effects of different UBR5 mutations in tetrameric UBR5 (Y2773F, Y2773D) with the $Ub^A$-6xHis mutants F45A and F45D (**Figure 29c**). Interrupting the UBA-

$Ub^A$ interface was also addressed from both sides: in a tetrameric UBR5 background, the UBA mutant L224D was tested together with wildtype $Ub^A$-6xHis as acceptor (**Figure 30a**) and different $Ub^A$-6xHis mutations were also tested (L8A / T9A, I44A, H68A, V70A / L71A) (**Figure 30b**). The potential UBA-C-lobe interface of UBR5 was tested by mutating the UBA residue G199W or the C-lobe residue P2748D in dimeric UBR5. The C-lobe residue Q2747 was mutated to leucine in tetrameric UBR5. All of these mutations were tested for potential effects on receiving and passing on $Ub^D$ to $Ub^A$ with wildtype $Ub^A$-6xHis present in the chase-mix.

**Linkage-specificity dependency on oligomeric state**

To assess whether both, UBR5 and $UBR5^{Dimer}$ maintain the same linkage-specificity, a di-ubiquitin synthesis assay was performed with the respective version of UBR5 in the chase-mix as well as untagged single lysine acceptor ubiquitin mutants. In these $Ub^A$-variants, all lysines are intact (WT), all lysines are mutated to arginine (K0), or all lysines apart from the depicted lysine are mutated to arginine and only one lysine remains intact ("6", "11", "27", "29", "33", "48", "63"). Samples were taken after 1 min and mixed with reducing Laemmli buffer (**Figure 16b**).

**Linkage-specificity dependency on $Ub^A$-C-lobe interaction**

Whether a slightly altered C-lobe-$Ub^A$-interface affects the linkage-specificity of UBR5, we performed di-ubiquitin synthesis assays with the UBR5 mutants Y2773D and Y2773F. During the chase-reaction, we incubated the respective E3 version (0.2 µM final concentration) with either wildtype untagged $Ub^A$, or $Ub^A$ mutants (2 µM final concentration) with all lysines mutated to arginine apart from K48 or K63 respectively. E2~$Ub^D$ was added (0.2 µM final concentration) and samples were taken at the indicated time points and quenched with non-reducing Laemmli buffer (**Figure 29e**).

**UBR5 truncations**

Various UBR5 truncations were also tested for their ability to form di-ubiquitin (**Figure 19a**). During the chase-reaction tetrameric (WT UBR5) or dimeric ($UBR5^{Dimer}$) full-length UBR5 were used, or several truncations: The C-terminal region of UBR5 containing the entire HECT domain as well as the interrupting MLLE domain ("HECT") (residues 2216-2799), the C-terminal HECT domain of UBR5 (2216-2799) occluding the MLLE domain (residues 2377-2454) and replacing it with a 6 amino acid long linker ("$HECT^{\Delta MLLE}$"), or a fusion of the UBA domain (residues 179-230) connected to the N-terminus of the HECT domain (including the MLLE domain) with a 15 amino acid long linker ("UBA-HECT"). In the chase-reaction, the respective E3 variant was mixed at a final concentration of 0.5 µM with $Ub^A$ (**Figure 19a**), or $Ub^A$ carrying a K48R mutation (**Figure 19b**) at 5 µM final concentration and supplemented with the pulse-mix at 1 µM (20 x ). Samples were taken at the indicated time points and mixed with non-reducing Laemmli buffer.

**Dependency on E2-N-lobe interaction**

To test the importance of the interaction between the E2 and UBR5's N-lobe, we tested whether an E2 carrying a F62A mutation could still transfer its ubiquitin on to UBR5 (**Figure 22a**). For this, we used the E2 enzyme UBE2D3 and UBE2D3$^{F62A}$ in the pulse-reaction with the other conditions maintained as previously described. Again, the pulse-reaction was quenched by addition of 50 mM EDTA and the pulse-mix was then added to the chase-reaction containing 0.2 µM UBR5 and wildtype 2 µM Ub$^{A}$-6xHis to a final concentration of 0.2 µM. Samples were taken as indicated and non-reducing as well as reducing SDS-samples were prepared.

**Lysine ruler**

We aimed to assess what properties the target side chain of the acceptor ubiquitin needs to have. After establishing that a lysine at position 48 of the acceptor ubiquitin is absolutely required, we next asked whether the precise geometry of the side chain is required or whether differently long side chains connecting to the amino group are compatible with UBR5 (**Figure 25a**).

The pulse-reaction was performed as described previously, with the exception that K0 ubiquitin was used as donor instead of K48R ubiquitin. Synthetically derived ubiquitins harboring differently long side chains to connect to the amino group (C1-C5) or a natural ubiquitin, recombinantly expressed and purified (C4 natural) as control were used as acceptor ubiquitin. The number of methylene groups between the α-carbon of residue 48 of the acceptor ubiquitin and the amino group are depicted for example as C1 containing one methylene group. These ubiquitin versions were kindly provided by Dr. J. Liwocha.

Tetrameric UBR5 (0.2 µM final concentration) was mixed with the respective acceptor ubiquitin version (10 µM final concentration) before adding the pulse-mix at a final concentration of 1 µM. Samples were taken at the indicated time points and mixed with reducing Laemmli buffer.

**pH-dependency of LOL**

We sought to understand whether the observed effect of mutating the LOL was caused by impaired deprotonation of the acceptor lysine.[204] For this reason, we performed a di-ubiquitin synthesis assay with wildtype UBR5 or the LOL mutant (0.2 µM final concentration), both in combination with wildtype Ub$^{A}$-6xHis (2 µM final concentration). Differently than before, the chase-reaction was not performed in the previously described buffer, but the pH of the mix was altered for the respective samples. All sample buffers contained 150 mM NaCl, and 1 mM DTT. For the reactions at pH 6.8, 7.5, and 8.8, 25 mM Tris-HCl titrated to the respective pH were used. To obtain higher pH, the buffer component had to be changed. To realize a pH of 9.5, 25 mM CAPSO were used and pH 11 was obtained using 25 mM CAPS. The pulse-reaction was performed as described before and subsequently to quenching the reaction, E2~Ub$^{D}$ was diluted into the chase-reaction 20 x (final concentration 1 µM). Samples were taken after 1 min and mixed with non-reducing Laemmli buffer (**Figure 28b**).

**UBR5 vs. NEDD4L**

We sought to understand whether the effects we observed by mutating residues of $Ub^A$, located in the interface of UBR5's C-lobe to $Ub^A$ are conserved amongst HECT E3 ligases. For this reason, we employed a pulse-chase assay with the E3 ligase added in the chase-reaction being either UBR5, or NEDD4L$^{\Delta C2}$, which was kindly provided by J. Botsch. NEDD4L is known to form K63 chains rather than K48-linked ubiquitin chains. The chase-reaction was performed with a final concentration of 0.5 µM for the respective E3 ligase, 5 µM for the respective $Ub^A$-6xHis version (wildtype, A46F, A46D in **Figure 29b**, or wildtype, and F45D for **Figure 29d**), and 1 µM E2~$Ub^D$. Samples were collected at the indicated time points and mixed with non-reducing Laemmli buffer.

## Heterogeneous chain formation (pulse-chase format)

### Modification of different di-ubiquitins

We assessed how well UBR5 can modify di-ubiquitins linked via their different lysines or the N-terminal amino group (**Figure 31a**). We aimed to analyze whether the compatibility of UBR5 with the respective di-ubiquitin correlates with how well the respective lysine would be accessible in our TS2 structure. Again, a pulse-reaction was performed as described previously. The chase-mix was prepared by mixing UBR5 or UBR5$^{Dimer}$ (0. 2 µM final concentration) with the various di-ubiquitins (2 µM final concentration) linked via the indicated residue. Samples were taken after 20 sec after adding E2~$Ub^D$ (0. 2 µM final concentration) and quenched with reducing Laemmli buffer.

### Position-dependent modification of tri-ubiquitin

Whether UBR5 preferentially modifies a specific position within a pre-assembled ubiquitin chain and thereby would form rather mixed or branched ubiquitin chains, was tested using tri-ubiquitin as acceptor ubiquitin chain with the acceptor lysine only present in distinct ubiquitin moieties within the chain (**Figure 31b**). Different M1-linked tri-ubiquitins were generated with K48 in the different ubiquitin moieties (proximal, internal, distal) mutated to arginine (K48R mutated ubiquitin in the chain is illustrated red, wildtype K48-carrying ubiquitin is illustrated as green moiety). After the pulse-reaction, E2~$Ub^D$ was added to a mix of UBR5 (0.2 µM) and the respective tri-ubiquitin (5 µM) at a final concentration of 1 µM, samples were taken at the indicated time points, and quenched with reducing Laemmli buffer.

### Position-dependent binding and modification of tri-ubiquitin

We did not only want to know which ubiquitin within a preassembled chain UBR5 preferentially modifies, but also which ubiquitin moiety would be bound by the UBA to subsequently be modified by UBR5's HECT domain (**Figure 31c**). For this, we again used different versions of M1-linked tri-ubiquitin. This time, all ubiquitin moieties within the acceptor chain had intact acceptor lysines (K48), however, the UBA binding site was mutated on different moieties within the chain (I44D). Red ubiquitin moieties show the position with an I44D mutation, green ubiquitins have an intact UBA binding site. UBR5

(0.2 µM) was mixed with the respective acceptor chain (5 µM) in the chase-mix, and E2~Ub$^{D}$ (1 µM) was added. Samples were taken as indicated and quenched with reducing Laemmli buffer.

**Modification of differently long acceptor ubiquitin chains**

Using a pulse-chase assay with differently long K63-linked ubiquitin chains serving as acceptor, we tested whether UBR5$^{Dimer}$ would be deficient compared to tetrameric UBR5 when modifying various lengths of ubiquitin chains (**Figure 16e**). The pulse-reaction was performed as previously described. For the chase-reaction, UBR5 or UBR5$^{Dimer}$ were mixed at a final concentration of 0.2 µM with the respective K63-ubiquitin chain (5 µM) and the pulse-mix (1 µM). Samples were taken after 20 sec and quenched with reducing Laemmli buffer.

**Multi-turnover assays**

**Polyubiquitylation mediated by SDA mutant**

Whether the SDA mutant would be deficient during several rounds of catalysis, was tested in a multi-turnover assay (**Figure 22c**). To do so, 20 µM fluoresceine-labeled wildtype ubiquitin was mixed with 5 µM UBE2D2, and 0.5 µM WT UBR5 or the SDA mutant respectively. Adding 0.5 µM E1 enzyme started the cascade and samples were taken at the indicated time points and quenched with reducing Laemmli buffer.

## Substrate ubiquitylation

**Ubiquitylation of MYC-substrate in pulse-chase format**

Whether UBR5 can modify a MYC-peptide, that had previously been published to be recognized by UBR5$^{164}$, or this MYC-peptide fused to one ubiquitin to mimic a primed state, was tested in a pulse-chase format (**Figure 32a**). The peptide as well as the peptide fused to ubiquitin were labeled with TAMRA. For the pulse-reaction, unlabeled K0 ubiquitin was used while the other pulse-conditions were maintained as previously described. In the chase-reaction, 10 µM of the labeled substrate (WT-peptide fused to ubiquitin, a mutant peptide fused to ubiquitin, or the wildtype or mutant peptide alone) as well as free labeled ubiquitin as a control, were pre-mixed with a final concentration of 0.2 µM dimeric UBR5 or UBR5$^{L224D}$. The UBA mutant was included to test whether any observed ubiquitylation is mediated by the UBA domain. 1 µM of the pulse-mix (E2~Ub$^{D}$) was added, samples were taken at the indicated time points, and quenched with reducing Laemmli buffer.

**Ubiquitylation of N-end degron substrate in pulse-chase format**

We aimed to determine whether dimeric UBR5 could modify short substrate-peptides that would presumably be recognized via the N-end rule pathway (**Figure 32b**). UBR5 was shown to recognize N-terminal arginines using its UBR domain.[150] Therefore, differently long peptides were designed, all of them fused to the N-terminus of ubiquitin. The peptides ranged from adding only one extra residue (N-terminal arginine) to Ub, to several residues

(RIFSTD, or RIFSTDTGPGGSG, RIFSTD$(GS)_{10}$, RIFSTD$(GS)_{15}$) with the longest peptide containing 35 residues between the N-terminal arginine and the fused N-terminus of ubiquitin. A pulse-reaction was prepared as discussed above with fluorescently labeled $Ub^D$ carrying a K48R mutation. The chase-mix was prepared by mixing the respective substrate at a final concentration of 10 µM and dimeric UBR5 or dimeric UBR5 with the L224D UBA mutation at 0.2 µM. E2~$Ub^D$ was added at a final concentration of 1 µM. Samples were collected at the indicated time points and quenched with reducing Laemmli buffer.

## Characterization of UbV-effects on UBR5's activity

**Effects of UbVs on catalytic activity of truncated HECT domain (pulse-chase format)**
We assessed whether UbV3 and UbV7 have an effect on the catalytic activity of UBR5's truncated HECT domain by employing a pulse-chase assay (**Figure 33e**). The pulse-reaction was performed as previously described with fluorescently-labeled wildtype ubiquitin loaded onto the E2. The chase-reaction was prepared by mixing a final of 2.5 µM $HECT^{\Delta N}$ with 50 µM of the respective UbV. After adding E2~$Ub^D$ at a final concentration of 1 µM, samples were taken and non-reducing Laemmli buffer was used to quench the reaction.

**Effects of UbVs on catalytic activity of UBR5 (pulse-chase format)**
Whether the UbVs would have an effect on the activity of full-length UBR5 was investigated by employing a pulse-chase assay (**Figure 33f)**. Again, fluorescently labeled wildtype ubiquitin was loaded onto the E2 in a pulse-reaction. The chase-mix was prepared by combining 1 µM UBR5 with 100 µM of the respective UbV. 1 µM E2~$Ub^D$ was added and non-reducing samples were collected at the indicated time points.

**Titration of UbV3 to UBR5 (pulse-chase format)**
To determine the minimum concentration UbV3 has to be used at in order to have an inhibitory effect on UBR5's activity, we performed a titration assay in a pulse-chase format (**Figure 33g**). Fluorescently-labeled wildtype ubiquitin was loaded onto the E2 in a pulse-reaction as described earlier. For the chase-reaction, a final of 0.2 µM of UBR5 was mixed with no UbV3, 1 µM, 10 µM, 50 µM, or 100 µM UbV3. E2~$Ub^D$ was added to a final concentration of 0.3 µM and samples were collected at the respective time points and quenched with non-reducing Laemmli buffer.

**Role of UbV3's I44 on UBR5's activity (pulse-chase format)**
We know that acceptor ubiquitins are recruited to UBR5 by its UBA domain binding the I44 patch of $Ub^A$.[147] To test whether this interaction is also important for the UbV3-UBR5 interaction, we performed a pulse-chase assay with fluorescently-labeled wildtype ubiquitin in the pulse-reaction and UBR5 (0.5 µM final concentration) along with UbV3 or

a UbV3$^{I44D}$ mutation (100 µM final) in the chase-reaction (**Figure 33h**). E2~Ub$^{D}$ was added to a final concentration of 1 µM. Non-reducing samples were prepared at the indicated time points.

**Influence of UBR5's UBA domain on UbV3's effect (pulse-chase format)**

We attempted to verify the effect observed by mutating UbV3's I44 to aspartate (**Figure 33h**), by disrupting the presumed interface from the other side: In a pulse-chase assay, we sought to identify what effect an L224D mutation in UBR5's UBA domain would have on the formation of the UbV3-Ub products (**Figure 33i**). As before, a pulse-mix was prepared with fluorescently-labeled wildtype ubiquitin. The chase-mix consisted of dimeric UBR5 or dimeric UBR5$^{L224D}$ at a final concentration of 0.2 µM, as well as UbV3 at a final concentration of 100 µM. Adding E2~Ub$^{D}$ at 1 µM to the mix started the reaction and samples were collected and quenched with non-reducing Laemmli buffer at the indicated time points.

## Interaction-studies

**Co-pull down of truncated HECT constructs and UbVs**

To test whether the different UbVs would interact with UBR5's truncated HECT domain, and more specifically, with which part of the HECT domain, we performed co-pull down experiments (**Figure 33d**). For this, the His-tagged UbVs were co-expressed with the GST-tagged HECT construct in *E. coli* in the respective combinations. The pellets were resuspended in lysis buffer and sonication was performed to disrupt the cells. The supernatant was pre-cleared and normalized using Bradford-measurements. Subsequently, each supernatant was split in two and either incubated with GST-resin or Ni-NTA resin to pull on the HECT-moiety or the UbV-moiety respectively. SDS-samples were prepared of the resin subsequently to gravity-flow pull down and extensive washing (25 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 mM β-mercaptoethanol, 20 mM imidazole). Coomassie-staining of the SDS-PAGE was then used to visualize the different protein moieties.

# List of Abbreviations

ABP - Activity-based Probe
APC/C - Anaphase-Promoting Complex/Cyclosome
AQUA-MS – Absolute quantification mass spectrometry
C2 domain – $Ca^{2+}$-dependent binding motif
C-lobe – C-terminal lobe
Cross-link MS – Cross-link mass spectrometry
cryo-EM – cryo electron microscopy
DMEM – Dulbecco's Modified Eagle Medium
dSBB – Double small β-barrels
DSD domain – Domain swap dimerization domain
DUB - Deubiquitylating enzyme
E1 - Ubiquitin-activating enzyme
E2 - Ubiquitin-conjugating enzyme
E3 - Ubiquitin ligase
E6AP - E6-Associated Protein
EDD – E3 identified by differential display
FL – Full-length
FT - Flow-Through
GFP – Green fluorescent protein
GST - Glutathione S-Transferase
HD domain – HECT display domain
HECT - Homologous to E6AP C terminus
HPLC – High pressure liquid chromatography
HRV – Human rhinovirus
IEX - Ion-Exchange
kDa - Kilodalton
LC-MS - Liquid Chromatography- Mass Spectrometry
LOL _ Ligation-organizing-loop
MLLE domain – Mademoiselle domain
MS - Mass Spectrometry
nanoDSF – Nano differential scanning fluorimetry
NCOA – Nuclear coactivator
N-lobe – N-terminal lobe
NR – Nuclear hormone receptor
PABC – Polyadenine binding C-terminus
PABP – Polyadenine binding protein
PAM – PABP interacting motif
PD – Pull down
POI – Protein of interest
RBR - RING-between-RING

RING - Really Interesting New Gene
RLD – RCC1-like domain
SDA – Scaffold donor ubiquitin approaching
SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC - Size-Exclusion Chromatography
Smurf1 – Smad ubiquitylation regulatory factor 1
TEV - Tobacco Etch Virus Protease
TPD – Targeted protein degradation
TS1 – Transition state 1
TS2 – Transition state 2
Ub - Ubiquitin
$Ub^{A}$ - Acceptor Ubiquitin
UBA domain – Ubiquitin associated domain
$Ub^{D}$ - Donor Ubiquitin
UBD – Ubiquitin binding domain
UBM – Ubiquitin binding motif
UbiCRest – Ubiquitin chain restriction
UBR domain – Ubiquitin box recognition domain
UBR5 – Ubiquitin box recognition 5
UbV – Ubiquitin variant
Ub-VME – Ubiquitin vinyl methyl ester
UIM – Ubiquitin interacting motif
UPS - Ubiquitin Proteasome System
USP - Ubiquitin-Specific Protease
WT - Wildtype
ZF - Zinc-Finger

# References


1 Schoenheimer, R. The dynamic state of body constituents. *Harvard University Press* (1942).

2 Alber, A. B. & Suter, D. M. Dynamics of protein synthesis and degradation through the cell cycle. *Cell Cycle* **18**, 784-794, doi:10.1080/15384101.2019.1598725 (2019).

3 Chen, X. Q. *et al.* Protein homeostasis in aging and cancer. *Front Cell Dev Biol* **11**, 1143532, doi:10.3389/fcell.2023.1143532 (2023).

4 Morimoto, R. I. & Cuervo, A. M. Protein homeostasis and aging: taking care of proteins from the cradle to the grave. *J Gerontol A Biol Sci Med Sci* **64**, 167-170, doi:10.1093/gerona/gln071 (2009).

5 Callis, J. The ubiquitination machinery of the ubiquitin system. *Arabidopsis Book* **12**, e0174, doi:10.1199/tab.0174 (2014).

6 Hershko, A., Ciechanover, A. & Rose, I. A. Identification of the active amino acid residue of the polypeptide of ATP-dependent protein breakdown. *J Biol Chem* **256**, 1525-1528 (1981).

7 Haas, A. L. & Rose, I. A. The mechanism of ubiquitin activating enzyme. A kinetic and equilibrium analysis. *J Biol Chem* **257**, 10329-10337 (1982).

8 Schulman, B. A. & Harper, J. W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* **10**, 319-331, doi:10.1038/nrm2673 (2009).

9 Stewart, M. D., Ritterhoff, T., Klevit, R. E. & Brzovic, P. S. E2 enzymes: more than just middle men. *Cell Res* **26**, 423-440, doi:10.1038/cr.2016.35 (2016).

10 Berndsen, C. E. & Wolberger, C. New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol* **21**, 301-307, doi:10.1038/nsmb.2780 (2014).

11 Buetow, L. & Huang, D. T. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nat Rev Mol Cell Biol* **17**, 626-642, doi:10.1038/nrm.2016.91 (2016).

12 Zheng, N. & Shabek, N. Ubiquitin Ligases: Structure, Function, and Regulation. *Annu Rev Biochem* **86**, 129-157, doi:10.1146/annurev-biochem-060815-014922 (2017).

13 Metzger, M. B., Pruneda, J. N., Klevit, R. E. & Weissman, A. M. RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim Biophys Acta* **1843**, 47-60, doi:10.1016/j.bbamcr.2013.05.026 (2014).

14 Huibregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. M. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc Natl Acad Sci U S A* **92**, 2563-2567, doi:10.1073/pnas.92.7.2563 (1995).

15 Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81-83, doi:10.1038/373081a0 (1995).

16 Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* **474**, 105-108, doi:10.1038/nature09966 (2011).

17 Riley, B. E. *et al.* Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. *Nat Commun* **4**, 1982, doi:10.1038/ncomms2982 (2013).

18 Wang, X. S. *et al.* The unifying catalytic mechanism of the RING-between-RING E3 ubiquitin ligase family. *Nat Commun* **14**, 168, doi:10.1038/s41467-023-35871-z (2023).

19 Komander, D. & Rape, M. The ubiquitin code. *Annu Rev Biochem* **81**, 203-229, doi:10.1146/annurev-biochem-060310-170328 (2012).

20 Komander, D., Clague, M. J. & Urbe, S. Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol* **10**, 550-563, doi:10.1038/nrm2731 (2009).

21 Mevissen, T. E. T. & Komander, D. Mechanisms of Deubiquitinase Specificity and Regulation. *Annu Rev Biochem* **86**, 159-192, doi:10.1146/annurev-biochem-061516-044916 (2017).

22 Snyder, N. A. & Silva, G. M. Deubiquitinating enzymes (DUBs): Regulation, homeostasis, and oxidative stress response. *J Biol Chem* **297**, 101077, doi:10.1016/j.jbc.2021.101077 (2021).
23 Varshavsky, A. Naming a targeting signal. *Cell* **64**, 13-15, doi:10.1016/0092-8674(91)90202-a (1991).
24 Kraft, C., Peter, M. & Hofmann, K. Selective autophagy: ubiquitin-mediated recognition and beyond. *Nat Cell Biol* **12**, 836-841, doi:10.1038/ncb0910-836 (2010).
25 Won, K. A. & Reed, S. I. Activation of cyclin E/CDK2 is coupled to site-specific autophosphorylation and ubiquitin-dependent degradation of cyclin E. *EMBO J* **15**, 4182-4193 (1996).
26 Diehl, J. A., Zindy, F. & Sherr, C. J. Inhibition of cyclin D1 phosphorylation on threonine-286 prevents its rapid degradation via the ubiquitin-proteasome pathway. *Genes Dev* **11**, 957-972, doi:10.1101/gad.11.8.957 (1997).
27 Farrell, A. S. & Sears, R. C. MYC degradation. *Cold Spring Harb Perspect Med* **4**, doi:10.1101/cshperspect.a014365 (2014).
28 Juszkiewicz, S. & Hegde, R. S. Quality Control of Orphaned Proteins. *Mol Cell* **71**, 443-457, doi:10.1016/j.molcel.2018.07.001 (2018).
29 Timms, R. T. & Koren, I. Tying up loose ends: the N-degron and C-degron pathways of protein degradation. *Biochem Soc Trans* **48**, 1557-1567, doi:10.1042/BST20191094 (2020).
30 Sherpa, D., Chrustowicz, J. & Schulman, B. A. How the ends signal the end: Regulation by E3 ubiquitin ligases recognizing protein termini. *Mol Cell* **82**, 1424-1438, doi:10.1016/j.molcel.2022.02.004 (2022).
31 Nakatsukasa, K. & Brodsky, J. L. The recognition and retrotranslocation of misfolded proteins from the endoplasmic reticulum. *Traffic* **9**, 861-870, doi:10.1111/j.1600-0854.2008.00729.x (2008).
32 Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354, doi:10.1126/science.aac4354 (2016).
33 Ciechanover, A. & Schwartz, A. L. The ubiquitin-dependent proteolytic pathway: specificity of recognition of the proteolytic substrates. *Revis Biol Celular* **20**, 217-234 (1989).
34 Ciechanover, A. & Schwartz, A. L. How are substrates recognized by the ubiquitin-mediated proteolytic system? *Trends Biochem Sci* **14**, 483-488, doi:10.1016/0968-0004(89)90180-1 (1989).
35 Harper, J. W. & Schulman, B. A. Cullin-RING Ubiquitin Ligase Regulatory Circuits: A Quarter Century Beyond the F-Box Hypothesis. *Annu Rev Biochem* **90**, 403-429, doi:10.1146/annurev-biochem-090120-013613 (2021).
36 Chau, V. *et al.* A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**, 1576-1583, doi:10.1126/science.2538923 (1989).
37 Laney, J. D. & Hochstrasser, M. Substrate targeting in the ubiquitin system. *Cell* **97**, 427-430, doi:10.1016/s0092-8674(00)80752-7 (1999).
38 Kirisako, T. *et al.* A ubiquitin ligase complex assembles linear polyubiquitin chains. *EMBO J* **25**, 4877-4887, doi:10.1038/sj.emboj.7601360 (2006).
39 Christensen, D. E., Brzovic, P. S. & Klevit, R. E. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat Struct Mol Biol* **14**, 941-948, doi:10.1038/nsmb1295 (2007).
40 Kim, H. C. & Huibregtse, J. M. Polyubiquitination by HECT E3s and the determinants of chain type specificity. *Mol Cell Biol* **29**, 3307-3318, doi:10.1128/MCB.00240-09 (2009).
41 Rotin, D. & Kumar, S. Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* **10**, 398-409, doi:10.1038/nrm2690 (2009).
42 Mattiroli, F. & Sixma, T. K. Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nat Struct Mol Biol* **21**, 308-316, doi:10.1038/nsmb.2792 (2014).

43 van Nocker, S. & Vierstra, R. D. Multiubiquitin chains linked through lysine 48 are abundant in vivo and are competent intermediates in the ubiquitin proteolytic pathway. *J Biol Chem* **268**, 24766-24773 (1993).
44 Husnjak, K. & Dikic, I. Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu Rev Biochem* **81**, 291-322, doi:10.1146/annurev-biochem-051810-094654 (2012).
45 Aalto, A. L. *et al.* M1-linked ubiquitination by LUBEL is required for inflammatory responses to oral infection in Drosophila. *Cell Death Differ* **26**, 860-876, doi:10.1038/s41418-018-0164-x (2019).
46 Durcan, T. M. *et al.* USP8 regulates mitophagy by removing K6-linked ubiquitin conjugates from parkin. *EMBO J* **33**, 2473-2491, doi:10.15252/embj.201489729 (2014).
47 Wickliffe, K. E., Williamson, A., Meyer, H. J., Kelly, A. & Rape, M. K11-linked ubiquitin chains as novel regulators of cell division. *Trends Cell Biol* **21**, 656-663, doi:10.1016/j.tcb.2011.08.008 (2011).
48 Sparrer, K. M. J. *et al.* TRIM23 mediates virus-induced autophagy via activation of TBK1. *Nat Microbiol* **2**, 1543-1557, doi:10.1038/s41564-017-0017-2 (2017).
49 Fei, C. *et al.* Smurf1-mediated Lys29-linked nonproteolytic polyubiquitination of axin negatively regulates Wnt/beta-catenin signaling. *Mol Cell Biol* **33**, 4095-4105, doi:10.1128/MCB.00418-13 (2013).
50 Lin, M. *et al.* USP38 Inhibits Type I Interferon Signaling by Editing TBK1 Ubiquitination through NLRP4 Signalosome. *Mol Cell* **64**, 267-281, doi:10.1016/j.molcel.2016.08.029 (2016).
51 Liu, S. *et al.* Nuclear RNF2 inhibits interferon function by promoting K33-linked STAT1 disassociation from DNA. *Nat Immunol* **19**, 41-52, doi:10.1038/s41590-017-0003-0 (2018).
52 Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. Recognition of the polyubiquitin proteolytic signal. *EMBO J* **19**, 94-102, doi:10.1093/emboj/19.1.94 (2000).
53 Spence, J., Sadis, S., Haas, A. L. & Finley, D. A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol Cell Biol* **15**, 1265-1273, doi:10.1128/MCB.15.3.1265 (1995).
54 Liu, P. *et al.* K63-linked polyubiquitin chains bind to DNA to facilitate DNA damage repair. *Sci Signal* **11**, doi:10.1126/scisignal.aar8133 (2018).
55 Fukushima, T. *et al.* Nedd4-induced monoubiquitination of IRS-2 enhances IGF signalling and mitogenic activity. *Nat Commun* **6**, 6780, doi:10.1038/ncomms7780 (2015).
56 Finley, D. *et al.* Inhibition of proteolysis and cell cycle progression in a multiubiquitination-deficient yeast mutant. *Mol Cell Biol* **14**, 5501-5509, doi:10.1128/mcb.14.8.5501-5509.1994 (1994).
57 Xu, P. *et al.* Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137**, 133-145, doi:10.1016/j.cell.2009.01.041 (2009).
58 Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* **44**, 325-340, doi:10.1016/j.molcel.2011.08.025 (2011).
59 Goebl, M. G. *et al.* The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme. *Science* **241**, 1331-1335, doi:10.1126/science.2842867 (1988).
60 Petroski, M. D. & Deshaies, R. J. Mechanism of lysine 48-linked ubiquitin-chain synthesis by the cullin-RING ubiquitin-ligase complex SCF-Cdc34. *Cell* **123**, 1107-1120, doi:10.1016/j.cell.2005.09.033 (2005).
61 Wang, M. & Pickart, C. M. Different HECT domain ubiquitin ligases employ distinct mechanisms of polyubiquitin chain synthesis. *EMBO J* **24**, 4324-4333, doi:10.1038/sj.emboj.7600895 (2005).
62 Grice, G. L. & Nathan, J. A. The recognition of ubiquitinated proteins by the proteasome. *Cell Mol Life Sci* **73**, 3497-3506, doi:10.1007/s00018-016-2255-5 (2016).

63 Du, J. *et al.* A cryptic K48 ubiquitin chain binding site on UCH37 is required for its role in proteasomal degradation. *Elife* **11**, doi:10.7554/eLife.76100 (2022).
64 Meyer, H. J. & Rape, M. Enhanced protein degradation by branched ubiquitin chains. *Cell* **157**, 910-921, doi:10.1016/j.cell.2014.03.037 (2014).
65 Wang, Y. S., Wu, K. P., Jiang, H. K., Kurkute, P. & Chen, R. H. Branched Ubiquitination: Detection Methods, Biological Functions and Chemical Synthesis. *Molecules* **25**, doi:10.3390/molecules25215200 (2020).
66 Kolla, S., Ye, M., Mark, K. G. & Rape, M. Assembly and function of branched ubiquitin chains. *Trends Biochem Sci* **47**, 759-771, doi:10.1016/j.tibs.2022.04.003 (2022).
67 Yau, R. G. *et al.* Assembly and Function of Heterotypic Ubiquitin Chains in Cell-Cycle and Protein Quality Control. *Cell* **171**, 918-933 e920, doi:10.1016/j.cell.2017.09.040 (2017).
68 Rana, A., Ge, Y. & Strieter, E. R. Ubiquitin Chain Enrichment Middle-Down Mass Spectrometry (UbiChEM-MS) Reveals Cell-Cycle Dependent Formation of Lys11/Lys48 Branched Ubiquitin Chains. *J Proteome Res* **16**, 3363-3369, doi:10.1021/acs.jproteome.7b00381 (2017).
69 Swatek, K. N. *et al.* Insights into ubiquitin chain architecture using Ub-clipping. *Nature* **572**, 533-537, doi:10.1038/s41586-019-1482-y (2019).
70 Ohtake, F., Tsuchiya, H., Saeki, Y. & Tanaka, K. K63 ubiquitylation triggers proteasomal degradation by seeding branched ubiquitin chains. *Proc Natl Acad Sci U S A* **115**, E1401-E1408, doi:10.1073/pnas.1716673115 (2018).
71 Liu, C., Liu, W., Ye, Y. & Li, W. Ufd2p synthesizes branched ubiquitin chains to promote the degradation of substrates modified with atypical chains. *Nat Commun* **8**, 14274, doi:10.1038/ncomms14274 (2017).
72 French, M. E., Koehler, C. F. & Hunter, T. Emerging functions of branched ubiquitin chains. *Cell Discov* **7**, 6, doi:10.1038/s41421-020-00237-y (2021).
73 Wertz, I. E. *et al.* Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation. *Nature* **528**, 370-375, doi:10.1038/nature16165 (2015).
74 Ohtake, F., Saeki, Y., Ishido, S., Kanno, J. & Tanaka, K. The K48-K63 Branched Ubiquitin Chain Regulates NF-kappaB Signaling. *Mol Cell* **64**, 251-266, doi:10.1016/j.molcel.2016.09.014 (2016).
75 Hoppe, T. Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all. *Trends Biochem Sci* **30**, 183-187, doi:10.1016/j.tibs.2005.02.004 (2005).
76 Huibregtse, J. M., Scheffner, M. & Howley, P. M. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. *EMBO J* **10**, 4129-4135, doi:10.1002/j.1460-2075.1991.tb04990.x (1991).
77 Scheffner, M., Huibregtse, J. M., Vierstra, R. D. & Howley, P. M. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75**, 495-505, doi:10.1016/0092-8674(93)90384-3 (1993).
78 Scheffner, M. & Kumar, S. Mammalian HECT ubiquitin-protein ligases: biological and pathophysiological aspects. *Biochim Biophys Acta* **1843**, 61-74, doi:10.1016/j.bbamcr.2013.03.024 (2014).
79 Lin, D. Y., Diao, J. & Chen, J. Crystal structures of two bacterial HECT-like E3 ligases in complex with a human E2 reveal atomic details of pathogen-host interactions. *Proc Natl Acad Sci U S A* **109**, 1925-1930, doi:10.1073/pnas.1115025109 (2012).
80 Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. & Howley, P. M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**, 1129-1136, doi:10.1016/0092-8674(90)90409-8 (1990).
81 Simonson, S. J., Difilippantonio, M. J. & Lambert, P. F. Two distinct activities contribute to human papillomavirus 16 E6's oncogenic potential. *Cancer Res* **65**, 8266-8273, doi:10.1158/0008-5472.CAN-05-1651 (2005).
82 Huang, L. *et al.* Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **286**, 1321-1326, doi:10.1126/science.286.5443.1321 (1999).

83 Verdecia, M. A. *et al.* Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. *Mol Cell* **11**, 249-259, doi:10.1016/s1097-2765(02)00774-8 (2003).
84 Kamadurai, H. B. *et al.* Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. *Mol Cell* **36**, 1095-1102, doi:10.1016/j.molcel.2009.11.010 (2009).
85 Maspero, E. *et al.* Structure of the HECT:ubiquitin complex and its role in ubiquitin chain elongation. *EMBO Rep* **12**, 342-349, doi:10.1038/embor.2011.21 (2011).
86 Kamadurai, H. B. *et al.* Mechanism of ubiquitin ligation and lysine prioritization by a HECT E3. *Elife* **2**, e00828, doi:10.7554/eLife.00828 (2013).
87 Maspero, E. *et al.* Structure of a ubiquitin-loaded HECT ligase reveals the molecular basis for catalytic priming. *Nat Struct Mol Biol* **20**, 696-701, doi:10.1038/nsmb.2566 (2013).
88 Nair, R. M. *et al.* Reconstitution and Structural Analysis of a HECT Ligase-Ubiquitin Complex via an Activity-Based Probe. *ACS Chem. Biol.* **16**, 1615-1621, doi:10.1021/acschembio.1c00433 (2021).
89 Hunkeler, M. *et al.* Solenoid architecture of HUWE1 contributes to ligase activity and substrate recognition. *Mol Cell* **81**, 3468-3480 e3467, doi:10.1016/j.molcel.2021.06.032 (2021).
90 Grabarczyk, D. B. *et al.* HUWE1 employs a giant substrate-binding ring to feed and regulate its HECT E3 domain. *Nat Chem Biol* **17**, 1084-1092, doi:10.1038/s41589-021-00831-5 (2021).
91 Singh, S. *et al.* Structural Basis for the Enzymatic Activity of the HACE1 HECT-Type E3 Ligase Through N-Terminal Helix Dimerization. *Adv Sci (Weinh)* **10**, e2207672, doi:10.1002/advs.202207672 (2023).
92 Duering, J. *et al.* Structural mechanisms of autoinhibition and substrate recognition by the ubiquitin ligase HACE1. doi:https://doi.org/10.21203/rs.3.rs-3220888/v1 (2023).
93 Weber, J., Polo, S. & Maspero, E. HECT E3 Ligases: A Tale With Multiple Facets. *Front Physiol* **10**, 370, doi:10.3389/fphys.2019.00370 (2019).
94 Sluimer, J. & Distel, B. Regulating the human HECT E3 ligases. *Cell Mol Life Sci* **75**, 3121-3141, doi:10.1007/s00018-018-2848-2 (2018).
95 Harvey, K. F. & Kumar, S. Nedd4-like proteins: an emerging family of ubiquitin-protein ligases implicated in diverse cellular functions. *Trends Cell Biol* **9**, 166-169, doi:10.1016/s0962-8924(99)01541-x (1999).
96 Knopf, J. L. *et al.* Cloning and expression of multiple protein kinase C cDNAs. *Cell* **46**, 491-502, doi:10.1016/0092-8674(86)90874-3 (1986).
97 Wiesner, S. *et al.* Autoinhibition of the HECT-type ubiquitin ligase Smurf2 through its C2 domain. *Cell* **130**, 651-662, doi:10.1016/j.cell.2007.06.050 (2007).
98 Mari, S. *et al.* Structural and functional framework for the autoinhibition of Nedd4-family ubiquitin ligases. *Structure* **22**, 1639-1649, doi:10.1016/j.str.2014.09.006 (2014).
99 Staub, O. *et al.* WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na+ channel deleted in Liddle's syndrome. *EMBO J* **15**, 2371-2380 (1996).
100 Kasanov, J., Pirozzi, G., Uveges, A. J. & Kay, B. K. Characterizing Class I WW domains defines key specificity determinants and generates mutant domains with novel specificities. *Chem Biol* **8**, 231-241, doi:10.1016/s1074-5521(01)00005-9 (2001).
101 Kanelis, V., Rotin, D. & Forman-Kay, J. D. Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nat Struct Biol* **8**, 407-412, doi:10.1038/87562 (2001).
102 Chen, Z. *et al.* A Tunable Brake for HECT Ubiquitin Ligases. *Mol Cell* **66**, 345-357 e346, doi:10.1016/j.molcel.2017.03.020 (2017).
103 Mund, T. & Pelham, H. R. Control of the activity of WW-HECT domain E3 ubiquitin ligases by NDFIP proteins. *EMBO Rep* **10**, 501-507, doi:10.1038/embor.2009.30 (2009).
104 Rosa, J. L., Casaroli-Marano, R. P., Buckler, A. J., Vilaro, S. & Barbacid, M. p619, a giant protein related to the chromosome condensation regulator RCC1, stimulates

guanine nucleotide exchange on ARF1 and Rab proteins. *EMBO J* **15**, 4262-4273 (1996).
105 Ohtsubo, M. *et al.* Isolation and characterization of the active cDNA of the human cell cycle gene (RCC1) involved in the regulation of onset of chromosome condensation. *Genes Dev* **1**, 585-593, doi:10.1101/gad.1.6.585 (1987).
106 Renault, L. *et al.* The 1.7 A crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller. *Nature* **392**, 97-101, doi:10.1038/32204 (1998).
107 Hochrainer, K. *et al.* The human HERC family of ubiquitin ligases: novel members, genomic organization, expression profiling, and evolutionary aspects. *Genomics* **85**, 153-164, doi:10.1016/j.ygeno.2004.10.006 (2005).
108 Zavodszky, E., Peak-Chew, S. Y., Juszkiewicz, S., Narvaez, A. J. & Hegde, R. S. Identification of a quality-control factor that monitors failures during proteasome assembly. *Science* **373**, 998-1004, doi:10.1126/science.abc6500 (2021).
109 Yagita, Y., Zavodszky, E., Peak-Chew, S. Y. & Hegde, R. S. Mechanism of orphan subunit recognition during assembly quality control. *Cell* **186**, 3443-3459 e3424, doi:10.1016/j.cell.2023.06.016 (2023).
110 Singh, S., Ng, J. & Sivaraman, J. Exploring the "Other" subfamily of HECT E3-ligases for therapeutic intervention. *Pharmacol Ther* **224**, 107809, doi:10.1016/j.pharmthera.2021.107809 (2021).
111 Kaiho-Soma, A. *et al.* TRIP12 promotes small-molecule-induced degradation through K29/K48-branched ubiquitin chains. *Mol Cell* **81**, 1411-1424 e1417, doi:10.1016/j.molcel.2021.01.023 (2021).
112 Lemak, A., Yee, A., Bezsonova, I., Dhe-Paganon, S. & Arrowsmith, C. H. Zn-binding AZUL domain of human ubiquitin protein ligase Ube3A. *J Biomol NMR* **51**, 185-190, doi:10.1007/s10858-011-9552-y (2011).
113 Michel, M. A., Swatek, K. N., Hospenthal, M. K. & Komander, D. Ubiquitin Linkage-Specific Affimers Reveal Insights into K6-Linked Ubiquitin Signaling. *Mol Cell* **68**, 233-246 e235, doi:10.1016/j.molcel.2017.08.020 (2017).
114 Zhou, M. *et al.* HUWE1 Amplifies Ubiquitin Modifications to Broadly Stimulate Clearance of Proteins and Aggregates. *bioRxiv*, doi:10.1101/2023.05.30.542866 (2023).
115 Palicharla, V. R. & Maddika, S. HACE1 mediated K27 ubiquitin linkage leads to YB-1 protein secretion. *Cell Signal* **27**, 2355-2362, doi:10.1016/j.cellsig.2015.09.001 (2015).
116 French, M. E. *et al.* Mechanism of ubiquitin chain synthesis employed by a HECT domain ubiquitin ligase. *J Biol Chem* **292**, 10398-10413, doi:10.1074/jbc.M117.789479 (2017).
117 Kobayashi, F., Nishiuchi, T., Takaki, K. & Konno, H. Ubiquitin chain specificities of E6AP E3 ligase and its HECT domain. *Biochem Biophys Res Commun* **496**, 686-692, doi:10.1016/j.bbrc.2017.12.076 (2018).
118 Mao, J. *et al.* Structural Visualization of HECT-E3 Ufd4 accepting and transferring Ubiquitin to Form K29/K48-branched Polyubiquitination on N-degron. *bioRxiv*, 2023.2005.2023.542033, doi:10.1101/2023.05.23.542033 (2023).
119 Franklin, T. G., Brzovic, P. S. & Pruneda, J. N. Bacterial mimicry of eukaryotic HECT ubiquitin ligation. *bioRxiv*, 2023.2006.2005.543783, doi:10.1101/2023.06.05.543783 (2023).
120 Luh, L. M. *et al.* Prey for the Proteasome: Targeted Protein Degradation-A Medicinal Chemist's Perspective. *Angew Chem Int Ed Engl* **59**, 15448-15466, doi:10.1002/anie.202004310 (2020).
121 Burslem, G. M. & Crews, C. M. Proteolysis-Targeting Chimeras as Therapeutics and Tools for Biological Discovery. *Cell* **181**, 102-114, doi:10.1016/j.cell.2019.11.031 (2020).
122 Hanzl, A. & Winter, G. E. Targeted protein degradation: current and future challenges. *Curr Opin Chem Biol* **56**, 35-41, doi:10.1016/j.cbpa.2019.11.012 (2020).

123 Kozicka, Z. & Thomä, N. H. Haven't got a glue: Protein surface variation for the design of molecular glue degraders. *Cell Chemical Biology* **28**, 1032-1047, doi:10.1016/j.chembiol.2021.04.009 (2021).
124 Perez Berrocal, D. A. *et al.* A Pro-Fluorescent Ubiquitin-Based Probe to Monitor Cysteine-Based E3 Ligase Activity. *Angew Chem Int Ed Engl* **62**, e202303319, doi:10.1002/anie.202303319 (2023).
125 Aronchik, I., Kundu, A., Quirit, J. G. & Firestone, G. L. The antiproliferative response of indole-3-carbinol in human melanoma cells is triggered by an interaction with NEDD4-1 and disruption of wild-type PTEN degradation. *Mol Cancer Res* **12**, 1621-1634, doi:10.1158/1541-7786.MCR-14-0018 (2014).
126 Kathman, S. G. *et al.* A Small Molecule That Switches a Ubiquitin Ligase From a Processive to a Distributive Enzymatic Mechanism. *J Am Chem Soc* **137**, 12442-12445, doi:10.1021/jacs.5b06839 (2015).
127 Zhang, W. *et al.* System-Wide Modulation of HECT E3 Ligases with Selective Ubiquitin Variant Probes. *Mol Cell* **62**, 121-136, doi:10.1016/j.molcel.2016.02.005 (2016).
128 Quirit, J. G. *et al.* Indole-3-carbinol (I3C) analogues are potent small molecule inhibitors of NEDD4-1 ubiquitin ligase activity that disrupt proliferation of human melanoma cells. *Biochem Pharmacol* **127**, 13-27, doi:10.1016/j.bcp.2016.12.007 (2017).
129 French, M. E., Kretzmann, B. R. & Hicke, L. Regulation of the RSP5 ubiquitin ligase by an intrinsic ubiquitin-binding site. *J Biol Chem* **284**, 12071-12079, doi:10.1074/jbc.M901106200 (2009).
130 Ogunjimi, A. A. *et al.* The ubiquitin binding region of the Smurf HECT domain facilitates polyubiquitylation and binding of ubiquitylated substrates. *J Biol Chem* **285**, 6308-6315, doi:10.1074/jbc.M109.044537 (2010).
131 Kim, H. C., Steffen, A. M., Oldham, M. L., Chen, J. & Huibregtse, J. M. Structure and function of a HECT domain ubiquitin-binding site. *EMBO Rep* **12**, 334-341, doi:10.1038/embor.2011.23 (2011).
132 Lorenz, S. Structural mechanisms of HECT-type ubiquitin ligases. *Biol Chem* **399**, 127-145, doi:10.1515/hsz-2017-0184 (2018).
133 Kao, H. W. *et al.* Robust Design of Effective Allosteric Activators for Rsp5 E3 Ligase Using the Machine Learning Tool ProteinMPNN. *ACS Synth Biol* **12**, 2310-2319, doi:10.1021/acssynbio.3c00042 (2023).
134 Martin, P., Martin, A. & Shearn, A. Studies of l(3)c43hs1 a polyphasic, temperature-sensitive mutant of Drosophila melanogaster with a variety of imaginal disc defects. *Dev Biol* **55**, 213-232, doi:10.1016/0012-1606(77)90168-3 (1977).
135 Callaghan, M. J. *et al.* Identification of a human HECT family protein with homology to the Drosophila tumor suppressor gene hyperplastic discs. *Oncogene* **17**, 3479-3491, doi:10.1038/sj.onc.1202249 (1998).
136 Shearer, R. F., Iconomou, M., Watts, C. K. & Saunders, D. N. Functional Roles of the E3 Ubiquitin Ligase UBR5 in Cancer. *Mol Cancer Res* **13**, 1523-1532, doi:10.1158/1541-7786.MCR-15-0383 (2015).
137 Wang, F. *et al.* Structure of the human UBR5 E3 ubiquitin ligase. *Structure* **31**, 541-552 e544, doi:10.1016/j.str.2023.03.010 (2023).
138 Hodakova, Z. *et al.* Cryo-EM structure of the chain-elongating E3 ubiquitin ligase UBR5. *EMBO J*, e113348, doi:10.15252/embj.2022113348 (2023).
139 Tsai, J. M. *et al.* UBR5 forms ligand-dependent complexes on chromatin to regulate nuclear hormone receptor stability. *Molecular Cell* (2023).
140 Mark, K. G. *et al.* Orphan quality control shapes network dynamics and gene expression. *Cell* **186**, 3460-3475 e3423, doi:10.1016/j.cell.2023.06.015 (2023).
141 Hehl, L. A. *et al.* Structural snapshots along K48-linked ubiquitin chain formation by the HECT E3 UBR5. *Nat Chem Biol*, doi:10.1038/s41589-023-01414-2 (2023).
142 Dikic, I., Wakatsuki, S. & Walters, K. J. Ubiquitin-binding domains - from structures to functions. *Nat Rev Mol Cell Biol* **10**, 659-671, doi:10.1038/nrm2767 (2009).
143 Trempe, J. F. Reading the ubiquitin postal code. *Curr Opin Struct Biol* **21**, 792-801, doi:10.1016/j.sbi.2011.09.009 (2011).

144 Mueller, T. D. & Feigon, J. Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. *J Mol Biol* **319**, 1243-1255, doi:10.1016/S0022-2836(02)00302-9 (2002).
145 Wang, Q., Young, P. & Walters, K. J. Structure of S5a bound to monoubiquitin provides a model for polyubiquitin recognition. *J Mol Biol* **348**, 727-739, doi:10.1016/j.jmb.2005.03.007 (2005).
146 Zhang, N. *et al.* Structure of the s5a:k48-linked diubiquitin complex and its interactions with rpn13. *Mol Cell* **35**, 280-290, doi:10.1016/j.molcel.2009.06.010 (2009).
147 Kozlov, G. *et al.* Structural basis of ubiquitin recognition by the ubiquitin-associated (UBA) domain of the ubiquitin ligase EDD. *J Biol Chem* **282**, 35787-35795, doi:10.1074/jbc.M705655200 (2007).
148 Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci* **20**, 1298-1345, doi:10.1002/pro.666 (2011).
149 Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179-186, doi:10.1126/science.3018930 (1986).
150 Tasaki, T. *et al.* A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol* **25**, 7120-7136, doi:10.1128/MCB.25.16.7120-7136.2005 (2005).
151 Tasaki, T. *et al.* The substrate recognition domains of the N-end rule pathway. *J Biol Chem* **284**, 1884-1895, doi:10.1074/jbc.M803641200 (2009).
152 Sherpa, D. *et al.* Modular UBE2H-CTLH E2-E3 complexes regulate erythroid maturation. *Elife* **11**, doi:10.7554/eLife.77937 (2022).
153 Mansfield, E., Hersperger, E., Biggs, J. & Shearn, A. Genetic and molecular analysis of hyperplastic discs, a gene whose product is required for regulation of cell proliferation in Drosophila melanogaster imaginal discs and germ cells. *Dev Biol* **165**, 507-526, doi:10.1006/dbio.1994.1271 (1994).
154 Lim, N. S. *et al.* Comparative peptide binding studies of the PABC domains from the ubiquitin-protein isopeptide ligase HYD and poly(A)-binding protein. Implications for HYD function. *J Biol Chem* **281**, 14376-14382, doi:10.1074/jbc.M600307200 (2006).
155 Xie, J., Kozlov, G. & Gehring, K. The "tale" of poly(A) binding protein: the MLLE domain and PAM2-containing proteins. *Biochim Biophys Acta* **1839**, 1062-1068, doi:10.1016/j.bbagrm.2014.08.001 (2014).
156 Kozlov, G., Menade, M., Rosenauer, A., Nguyen, L. & Gehring, K. Molecular determinants of PAM2 recognition by the MLLE domain of poly(A)-binding protein. *J Mol Biol* **397**, 397-407, doi:10.1016/j.jmb.2010.01.032 (2010).
157 Munoz-Escobar, J., Matta-Camacho, E., Kozlov, G. & Gehring, K. The MLLE domain of the ubiquitin ligase UBR5 binds to its catalytic domain to regulate substrate binding. *J Biol Chem* **290**, 22841-22850, doi:10.1074/jbc.M115.672246 (2015).
158 Devan, S. K. *et al.* A MademoiseLLE domain binding platform links the key RNA transporter to endosomes. *PLoS Genet* **18**, e1010269, doi:10.1371/journal.pgen.1010269 (2022).
159 Gudjonsson, T. *et al.* TRIP12 and UBR5 suppress spreading of chromatin ubiquitylation at damaged chromosomes. *Cell* **150**, 697-709, doi:10.1016/j.cell.2012.06.039 (2012).
160 Zhang, S., Valenzuela, L. F., Zatulovskiy, E. & Skotheim, J. M. The G1/S transition is promoted by Rb degradation via the E3 ligase UBR5. *bioRxiv*, doi:10.1101/2023.10.03.560768 (2023).
161 Jiang, W. *et al.* Acetylation regulates gluconeogenesis by promoting PEPCK1 degradation via recruiting the UBR5 ubiquitin ligase. *Mol Cell* **43**, 33-44, doi:10.1016/j.molcel.2011.04.028 (2011).
162 Cammarata-Mouchtouris, A. *et al.* Hyd ubiquitinates the NF-kappaB co-factor Akirin to operate an effective immune response in Drosophila. *PLoS Pathog* **16**, e1008458, doi:10.1371/journal.ppat.1008458 (2020).

163 Qiao, X. *et al.* UBR5 Is Coamplified with MYC in Breast Tumors and Encodes an Ubiquitin Ligase That Limits MYC-Dependent Apoptosis. *Cancer Res* **80**, 1414-1427, doi:10.1158/0008-5472.CAN-19-1647 (2020).
164 Schukur, L. *et al.* Identification of the HECT E3 ligase UBR5 as a regulator of MYC degradation using a CRISPR/Cas9 screen. *Sci Rep* **10**, 20044, doi:10.1038/s41598-020-76960-z (2020).
165 Tsai, J. M. *et al.* UBR5 Is a Hect E3 Ubiquitin Ligase That Regulates Chromatin Bound Nuclear Hormone Receptor Stability. *Blood* **140**, 2968-2969, doi:10.1182/blood-2022-167615 (2022).
166 Cojocaru, M. *et al.* Transcription factor IIS cooperates with the E3 ligase UBR5 to ubiquitinate the CDK9 subunit of the positive transcription elongation factor B. *J Biol Chem* **286**, 5012-5022, doi:10.1074/jbc.M110.176628 (2011).
167 Hay-Koren, A., Caspi, M., Zilberberg, A. & Rosin-Arbesfeld, R. The EDD E3 ubiquitin ligase ubiquitinates and up-regulates beta-catenin. *Mol Biol Cell* **22**, 399-411, doi:10.1091/mbc.E10-05-0440 (2011).
168 Wang, X. *et al.* HIV-1 Vpr protein inhibits telomerase activity via the EDD-DDB1-VPRBP E3 ligase complex. *J Biol Chem* **288**, 15474-15480, doi:10.1074/jbc.M112.416735 (2013).
169 Zhang, T., Cronshaw, J., Kanu, N., Snijders, A. P. & Behrens, A. UBR5-mediated ubiquitination of ATMIN is required for ionizing radiation-induced ATM signaling and function. *Proc Natl Acad Sci U S A* **111**, 12091-12096, doi:10.1073/pnas.1400230111 (2014).
170 Koyuncu, S. *et al.* The ubiquitin ligase UBR5 suppresses proteostasis collapse in pluripotent stem cells from Huntington's disease patients. *Nat Commun* **9**, 2886, doi:10.1038/s41467-018-05320-3 (2018).
171 Chen, L. *et al.* E3 ubiquitin ligase UBR5 promotes pancreatic cancer growth and aerobic glycolysis by downregulating FBP1 via destabilization of C/EBPalpha. *Oncogene* **40**, 262-276, doi:10.1038/s41388-020-01527-1 (2021).
172 Dhingani, N. *et al.* The E3 ubiquitin ligase UBR5 interacts with TTC7A and may be associated with very early onset inflammatory bowel disease. *Sci Rep* **10**, 18648, doi:10.1038/s41598-020-73482-6 (2020).
173 Hehl, L. A. & Schulman, B. A. To be (in a transcriptional complex) or not to be (promoting UBR5 ubiquitylation): That is an answer to how degradation controls gene expression. *Mol Cell* **83**, 2616-2618, doi:10.1016/j.molcel.2023.07.010 (2023).
174 Julg, J., Edbauer, D. & Behrends, C. C9orf72 protein quality control by UBR5-mediated heterotypic ubiquitin chains. *EMBO Rep* **24**, e55895, doi:10.15252/embr.202255895 (2023).
175 Sander, B., Xu, W., Eilers, M., Popov, N. & Lorenz, S. A conformational switch regulates the ubiquitin ligase HUWE1. *Elife* **6**, doi:10.7554/eLife.21036 (2017).
176 Boyce, F. M. & Bucher, N. L. Baculovirus-mediated gene transfer into mammalian cells. *Proc Natl Acad Sci U S A* **93**, 2348-2352, doi:10.1073/pnas.93.6.2348 (1996).
177 Condreay, J. P., Witherspoon, S. M., Clay, W. C. & Kost, T. A. Transient and stable gene expression in mammalian cells transduced with a recombinant baculovirus vector. *Proc Natl Acad Sci U S A* **96**, 127-132, doi:10.1073/pnas.96.1.127 (1999).
178 Chattopadhyay, G. & Varadarajan, R. Facile measurement of protein stability and folding kinetics using a nano differential scanning fluorimeter. *Protein Sci* **28**, 1127-1134, doi:10.1002/pro.3622 (2019).
179 Pao, K. C. *et al.* Activity-based E3 ligase profiling uncovers an E3 ligase with esterification activity. *Nature* **556**, 381-385, doi:10.1038/s41586-018-0026-1 (2018).
180 Kirkpatrick, D. S., Gerber, S. A. & Gygi, S. P. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **35**, 265-273, doi:10.1016/j.ymeth.2004.08.018 (2005).
181 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

182 Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423-427, doi:10.1126/science.aar5839 (2018).
183 Matta-Camacho, E., Kozlov, G., Menade, M. & Gehring, K. Structure of the HECT C-lobe of the UBR5 E3 ubiquitin ligase. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 1158-1163, doi:10.1107/S1744309112036937 (2012).
184 Matta-Camacho, E., Kozlov, G., Li, F. F. & Gehring, K. Structural basis of substrate recognition and specificity in the N-end rule pathway. *Nat Struct Mol Biol* **17**, 1182-1187, doi:10.1038/nsmb.1894 (2010).
185 Bremm, A., Freund, S. M. & Komander, D. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nat Struct Mol Biol* **17**, 939-947, doi:10.1038/nsmb.1873 (2010).
186 Stieglitz, B. *et al.* Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. *Nature* **503**, 422-426, doi:10.1038/nature12638 (2013).
187 Branigan, E., Plechanovova, A., Jaffray, E. G., Naismith, J. H. & Hay, R. T. Structural basis for the RING-catalyzed synthesis of K63-linked ubiquitin chains. *Nat Struct Mol Biol* **22**, 597-602, doi:10.1038/nsmb.3052 (2015).
188 Pan, M. *et al.* Structural insights into Ubr1-mediated N-degron polyubiquitination. *Nature* **600**, 334-338, doi:10.1038/s41586-021-04097-8 (2021).
189 Cotton, T. R. *et al.* Structural basis of K63-ubiquitin chain formation by the Gordon-Holmes syndrome RBR E3 ubiquitin ligase RNF216. *Mol Cell* **82**, 598-615 e598, doi:10.1016/j.molcel.2021.12.005 (2022).
190 Nakasone, M. A. *et al.* Structure of UBE2K-Ub/E3/polyUb reveals mechanisms of K48-linked Ub chain extension. *Nat Chem Biol* **18**, 422-431, doi:10.1038/s41589-021-00952-x (2022).
191 Hospenthal, M. K., Mevissen, T. E. T. & Komander, D. Deubiquitinase-based analysis of ubiquitin chain architecture using Ubiquitin Chain Restriction (UbiCRest). *Nat Protoc* **10**, 349-361, doi:10.1038/nprot.2015.018 (2015).
192 Horn-Ghetko, D. *et al.* Ubiquitin ligation to F-box protein targets by SCF-RBR E3-E3 super-assembly. *Nature* **590**, 671-676, doi:10.1038/s41586-021-03197-9 (2021).
193 Baek, K. *et al.* NEDD8 nucleates a multivalent cullin-RING-UBE2D ubiquitin ligation assembly. *Nature* **578**, 461-466, doi:10.1038/s41586-020-2000-y (2020).
194 Nuber, U. & Scheffner, M. Identification of determinants in E2 ubiquitin-conjugating enzymes required for hect E3 ubiquitin-protein ligase interaction. *J Biol Chem* **274**, 7576-7582, doi:10.1074/jbc.274.11.7576 (1999).
195 Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* **173**, 530-540, doi:10.1016/j.jsb.2010.10.014 (2011).
196 Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem Biol* **9**, 1149-1159, doi:10.1016/s1074-5521(02)00248-x (2002).
197 Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol* **213**, 107702, doi:10.1016/j.jsb.2021.107702 (2021).
198 Liwocha, J. *et al.* Linkage-specific ubiquitin chain formation depends on a lysine hydrocarbon ruler. *Nat Chem Biol* **17**, 272-279, doi:10.1038/s41589-020-00696-0 (2021).
199 Li, G., Liang, Q., Gong, P., Tencer, A. H. & Zhuang, Z. Activity-based diubiquitin probes for elucidating the linkage specificity of deubiquitinating enzymes. *Chem Commun (Camb)* **50**, 216-218, doi:10.1039/c3cc47382a (2014).
200 Salvat, C., Wang, G., Dastur, A., Lyon, N. & Huibregtse, J. M. The -4 phenylalanine is required for substrate ubiquitination catalyzed by HECT ubiquitin ligases. *J Biol Chem* **279**, 18935-18943, doi:10.1074/jbc.M312201200 (2004).
201 Souphron, J. *et al.* Structural dissection of a gating mechanism preventing misactivation of ubiquitin by NEDD8's E1. *Biochemistry* **47**, 8961-8969, doi:10.1021/bi800604c (2008).

202 Zhu, K. *et al.* Allosteric auto-inhibition and activation of the Nedd4 family E3 ligase Itch. *EMBO Rep* **18**, 1618-1630, doi:10.15252/embr.201744454 (2017).
203 Singh, S. & Sivaraman, J. Crystal structure of HECT domain of UBE3C E3 ligase and its ubiquitination activity. *Biochem J* **477**, 905-923, doi:10.1042/BCJ20200027 (2020).
204 Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nat Struct Mol Biol* **13**, 491-499, doi:10.1038/nsmb1104 (2006).
205 Takeda, K. *et al.* Structural dynamics of E6AP E3 ligase HECT domain and involvement of flexible hinge loop in ubiquitin chain synthesis mechanism. *bioRxiv*, 2022.2011.2018.516873, doi:10.1101/2022.11.18.516873 (2022).
206 Pandya, R. K., Partridge, J. R., Love, K. R., Schwartz, T. U. & Ploegh, H. L. A structural element within the HUWE1 HECT domain modulates self-ubiquitination and substrate ubiquitination activities. *J Biol Chem* **285**, 5664-5673, doi:10.1074/jbc.M109.051805 (2010).
207 Singh, S., Ng, J., Nayak, D. & Sivaraman, J. Structural insights into a HECT-type E3 ligase AREL1 and its ubiquitination activities in vitro. *J Biol Chem* **294**, 19934-19949, doi:10.1074/jbc.RA119.010327 (2019).
208 Franklin, T. G., Brzovic, P. S. & Pruneda, J. N. Bacterial ligases reveal fundamental principles of polyubiquitin specificity. *Mol Cell*, doi:10.1016/j.molcel.2023.11.017 (2023).
209 Sherpa, D. *et al.* GID E3 ligase supramolecular chelate assembly configures multipronged ubiquitin targeting of an oligomeric metabolic enzyme. *Mol Cell* **81**, 2445-2459 e2413, doi:10.1016/j.molcel.2021.03.025 (2021).
210 Jenni, S. *et al.* Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254-261, doi:10.1126/science.1138248 (2007).
211 Taherbhoy, A. M. & Daniels, D. L. Harnessing UBR5 for targeted protein degradation of key transcriptional regulators. *Trends Pharmacol Sci* **44**, 758-761, doi:10.1016/j.tips.2023.09.001 (2023).
212 Dukkipati, A., Park, H. H., Waghray, D., Fischer, S. & Garcia, K. C. BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Expr Purif* **62**, 160-170, doi:10.1016/j.pep.2008.08.004 (2008).
213 Scott, D. C. & Schulman, B. A. Dual-color pulse-chase ubiquitination assays to simultaneously monitor substrate priming and extension. *Methods Enzymol* **618**, 29-48, doi:10.1016/bs.mie.2019.01.004 (2019).
214 El Oualid, F. *et al.* Chemical synthesis of ubiquitin, ubiquitin-based probes, and diubiquitin. *Angew Chem Int Ed Engl* **49**, 10149-10153, doi:10.1002/anie.201005995 (2010).
215 Graham, M., Combe, C., Kolbowski, L. & Rappsilber, J. xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data. *bioRxiv*, 561829, doi:10.1101/561829 (2019).
216 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).
217 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).
218 Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol* **4**, 874, doi:10.1038/s42003-021-02399-1 (2021).
219 Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J* **478**, 4169-4185, doi:10.1042/BCJ20210708 (2021).
220 Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214-1221, doi:10.1038/s41592-020-00990-8 (2020).

221 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).
222 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).
223 de Jong, A. *et al.* Ubiquitin-based probes prepared by total synthesis to profile the activity of deubiquitinating enzymes. *Chembiochem* **13**, 2251-2258, doi:10.1002/cbic.201200497 (2012).
224 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
225 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
226 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
227 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315, doi:10.1002/pro.3330 (2018).

## Website

cbioportal, last accessed 19.01.2024:

https://www.cbioportal.org/results/cancerTypesSummary?case_set_id=all&gene_list=UBR5&cancer_study_list=5c8a7d55e4b046111fee2296