TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie

Labor für Synthetische Biochemie

Development of a novel chemoenzymatic approach to ubiquitylate target proteins

Maximilian Fottner

Vollständiger Abdruck der, von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende/-r:	Prof. Dr. Matthias Feige
Prüfende/-r der Dissertation:	1. Prof. Dr. Kathrin Lang
	2. Prof. Dr. Cathleen Zeymer
	3. Prof. Dr. Henning Mootz

Die Dissertation wurde am 29.10.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 01.12.2020 angenommen.

The work presented within this thesis was carried out in the Lab for Synthetic Biochemistry, Technical University Munich, from October 2016 to August 2020, under the supervision of Prof. Dr. Kathrin Lang.

Danksagung

An dieser Stelle möchte ich all den Leuten danken, ohne die die letzten 4 Jahre ein unmögliches Unterfangen gewesen wären:

Liebe Kathrin, zu allererst möchte ich mich bei dir, für die Möglichkeit in deiner Gruppe promovieren zu dürfen, bedanken. Unser „Chefin - Angestellter“ Verhältnis hat in den letzten Jahren ja doch so einiges durchgemacht. Angefangen mit fast verpassten Retreats und viel „Quod licet Iovi, non licet bovi“. Ich bin dir unendlich dankbar, dass du auch in so manchen schwierigen Zeiten an mich geglaubt hast und mir den Rücken gestärkt hast. Mindestens genauso wichtig ist aber natürlich auch, dass du mich mit mahnenden Worten auf dem – hoffentlich – richtigen Weg gehalten hast. Ich werde die wissenschaftlichen Diskussionen mit dir mindestens genauso vermissen wie regelmäßigen gossipen über die ‚Academia‘ – aber noch ist es ja nicht vorbei. Danke für alles!

Und das bringt mich auch direkt zu den wunderbaren Kollegen/innen, die sich AK Lang nennen. Ich bin jetzt doch schon eine ganze Weile dabei und habe einige Menschen kommen und gehen sehen, aber so ne coole Gruppe wie unsere muss man erstmal finden. Anh, Kristina, Marie-Lena, Vera, Toni und natürlich auch die schon von uns geschiedenen: Marko, Susanne und Marie – Ihr seid die Besten! Ich glaube wir hatten nie sowas wie eine schlechte „Labatmosphäre“ und das ist schon etwas Außergewöhnliches – mindestens genauso „außergewöhnlich“ wie unsere abendlichen Zusammenkünfte über dem ein oder anderen Bier. Ich bin zwar noch nicht „Geschichte“ aber ich vermisse euch jetzt schon. Es war mir eine Ehre, dass ich mit euch die letzten Jahre verbringen durfte.

Des Weiteren möchte ich dem „Chemiegebäude“ danken, das in den letzten 9 Jahren meine zweite Heimat geworden ist; auch deinen brutalistischen Glanz werde ich vermissen, aber wahrscheinlich noch viel mehr die Leute, die das Gebäude zu dem gemacht haben, was es ist. Dazu gehören Kollegen und Freunde, die mich teils schon seit dem Studium begleiten wie „Die Buchners“: Chris (Prost!), Gordon, Pamina, Maxi, Moritz, Jannis, Bene, Tuan und Georg als auch Stephan und Masood vom AK Sattler. Natürlich möchte ich auch Costa und seiner Familie danken, die mich täglich genährt haben und mit denen ich den ein oder anderen schönen Abend verbringen durfte. Was wäre das Chemiegebäude aber schon wenn es kein Licht, funktionierende Aufzüge oder Rollläden geben würde. Danke Walter und Mario! Euch kenne ich jetzt ja wirklich schon ewig (seit Kesslerzeiten) und ich freue mich jedes Mal wieder mit euch quatschen zu dürfen.

Und nun zu den Leuten die wahrscheinlich den Großteil der Arbeit geleistet haben, die auf den nächsten knapp zweihundert Seiten zu sehen ist. Ich hatte das Glück von der Crème de la Crème der Masteranden/innen, Forschungspraktikanten/innen und Bacheloranden/innen unterstützt zu werden. Danke an alle meine Forschungspraktikanten: Konstantin, Dominik, Timon, Vera, Maxl, Flo, Carlos, Domi und Philipp (äh Thor), die es sich getraut haben mit mir ein paar Wochen (oder Monate) zu verbringen. Ein paar von euch haben ja nach einem Praktikum tatsächlich noch nicht genug von Ubiquitin und Sortasen gehabt und sind für eine Masterarbeit zurückgekommen. Liebe/r Vera, Maxl, Carlos und Flo, ich kann mir nur schwer vorstellen dass es bessere Masteranden/innen gegeben haben könnte als euch. Danke!

Nicht zu vergessen ist natürlich auch Maria, welche die erste Bachelorandin im AK Lang war und herausragende Arbeit geleistet hat und quasi im Alleingang die Grundlange für das dritte Kapitel dieser Doktorarbeit geliefert hat.

Größten Dank und Liebe will ich hier auch noch einmal Marko aussprechen. Nicht nur weil du ein unglaublicher Kollege warst, der mir den Großteil meiner organisch chemischen Kenntnisse beigebracht hat, sondern auch weil du als ständiger Proofreader das sprachliche Niveau dieser Doktorarbeit (und vieler anderer Arbeiten) extrem angehoben hast.

Die letzten Jahre fanden natürlich (und zum Glück) nicht ausschließlich im Labor statt und ohne die Unterstützung meiner Familie und Freunde wäre es mit Sicherheit nicht möglich gewesen diese Zeit, ohne einen Burnout, durchzustehen. Deshalb zuallererst: Danke Mama! Ohne deine Unterstützung in allen Lebenslagen wäre ich heute nicht da wo ich jetzt bin. Es ist gut zu wissen, dass es einen Menschen gibt auf den man sich immer verlassen kann.

Mindestens genauso viel Dank gilt natürlich Antonia, die wahrscheinlich am meisten unter den teils unorthodoxen Arbeitszeiten spät abends und an Wochenenden leiden musste. Danke dass du immer für mich da warst und es doch immer irgendwie mit mir aushältst (ich könnte es wahrscheinlich nicht). Es ist ein unglaublich gutes Gefühl dich an meiner Seite zu wissen.

Natürlich möchte ich mich hier auch noch bei Laura und Hansi bedanken. Wir sehen uns leider am häufigsten wenn es am meisten brennt, aber das schweißt ja schließlich zusammen und dafür ist Familie da. Mein Dank gilt ebenfalls Oma und Walter, ihr habt mich immer unterstützt auch wenn ihr nicht so genau wusstet was ich eigentlich den ganzen Tag so mache. Auch das ist viel wert.

Apropos Familie. Liebe Heidi, lieber Stefan (Anna, Luca und Vincent) ich hoffe dass euch diese Worte in irgendeiner Form erreichen werden. Ohne die stetige Unterstützung über die letzten Jahrzehnte wäre mit Sicherheit nichts von alledem möglich gewesen. Es ist schön euch Familie nennen zu dürfen und ich freue mich schon, wenn wir eine Gelegenheit finden anzustoßen.

Nicht zu vergessen sind natürlich auch Sybille, Seppi, Viktoria und Alex, ich bin froh dass ihr über Antonia in mein Leben gekommen seid und ich hoffe, dass wir noch viele schöne Abende miteinander verbringen dürfen.

Zu guter Letzt möchte ich noch meinen Freunden danken die mich jetzt schon über ein Jahrzehnt (oder länger) begleiten. Danke Felix, Tobi, Timo, Anja, Ayhan, Alex, Tim, Linse und Andi. Wir sehen uns alle leider viel zu selten, aber ich möchte niemanden von euch missen.

So und nachdem ich nun wahrscheinlich die Hälfte der Menschen vergessen habe, denen ich danken sollte oder wollte (was vielleicht meinem Übermüdungsgrad zu schulden ist), nochmals ein Dank an alle die sich angesprochen fühlen.

List of Publications

Parts of this thesis have been published as papers, reviews and patents or have been presented as posters at scientific conferences:

Journal Publication (peer reviewed)

Cigler, M.; Muller, T. G.; Horn-Ghetko, D.; von Wrisberg, M. K.; **Fottner, M.**; Goody, R. S.; Itzen, A.; Muller, M. P.; Lang, K. Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes In Vitro and In Vivo. *Angew Chem Int Ed Engl* **2017,** *56* (49), 15737.

Fottner, M.; Brunner, A. D.; Bittl, V.; Horn-Ghetko, D.; Jussupow, A.; Kaila, V. R. I.; Bremm, A.; Lang, K. Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat Chem Biol* **2019,** *15* (3), 276.

Mideksa, Y. G.; **Fottner, M.**; Braus, S.; Weiss, C. A. M.; Nguyen, T. A.; Meier, S.; Lang, K.; Feige, M. J. Site-Specific Protein Labeling with Fluorophores as a Tool To Monitor Protein Turnover. *Chembiochem* **2020,** *21* (13), 1861.

Fauser, J.[#]; Savitskiy, S.[#]; **Fottner, M.**; Trauschke, V.; Gulen, B. Sortase-Mediated Quantifiable Enzyme Immobilization on Magnetic Nanoparticles. *Bioconjug Chem* **2020,** *31* (8), 1883. [#]these author contributed equally

Fottner, M.; Lang, K. Decorating proteins with LACE. *Nat Chem* **2020***, 12* (11)*,* 980.

Manuscript in preparation

Fottner, M.; Cigler, M.; Lang, K. Genetic code expansion approaches to decipher the ubiquitin code.

Fottner, M.; Weyh, M. D.; Schwarz, D.; Lang, K. Generating complex ubiquitin architectures by using genetic code expansion and orthogonal sortases.

Poster presentations (peer reviewed)

Fottner, M.; Brunner, A. D; Bittl, V.; Horn-Ghetko, D.; Bremm, A.; Lang, K. Site-specific ubiquitylation and SUMOylation using genetic code expansion, poster presentation, EMBO Conference: Chemical Biology 2018. 29.08. – 01.09.2018, Heidelberg, Germany. Recognized with a poster award.

Fottner, M.; Brunner, A. D; Bittl, V.; Horn-Ghetko, D.; Bremm, A.; Lang, K. Sortylation – A novel approach to site-specific ubiquitylate and SUMOylate proteins in vitro and in living cells, poster presentation, Ubiquitin & Friends Symposium 2019. 20.05 – 21.05.2019, Vienna, Austria. Recognized with a poster award.

Fottner, M.; Brunner, A. D; Bittl, V.; Horn-Ghetko, D.; Bremm, A.; Lang, K. Sortylation – A novel approach to site-specific ubiquitylate and SUMOylate proteins in vitro and in living cells, poster presentation, LMB-GGNB Student Symposium 2019. 10.07 – 12.07.2019, Cambridge, UK.

Fottner, M.; Weyh, M.; Schwarz, D.; Lang, K. Building mixed and hybrid ubiquitin chains by using orthogonal sortases, poster presentation, EMBO Conference: The ubiquitin system: Biology, mechanisms and roles in disease. 13.09 – 17.09.2019, Cavtat, Croatia.

Patents

Lang, K; **Fottner, M.**; Brunner, A. D.; 2019 Means and methods for site-specific protein modification using transpeptidases. WO/2020/007899 A1

Table of Contents

Abstract

The covalent attachment of ubiquitin (Ub) to substrate proteins represents one of the most versatile and flexible post-translational modifications (PTMs) in eukaryotic cells and is central to the regulation of almost every cellular process. Conjugation of Ub to substrate proteins is mediated *via* a highly conserved enzymatic cascade involving E1/E2/E3 enzymes that attach Ub *via* its C-terminus to the ε-amino group of lysine residues within the target protein by forming an isopeptide bond (ubiquitylation). Once attached, Ub can serve as a platform for further ubiquitylation events *via* one of its seven lysine residues or *via* its N-terminus - a process called polyubiquitylation. Polymeric Ub architectures installed on target proteins can vary in length, linkages and even in branching points, thereby generating highly complex topologies that are translated into numerous functional outcomes by distinct classes of reader proteins bearing Ub-binding domains. Like many other PTMs, ubiquitylation is a reversible process which is tightly regulated by deubiquitylating enzymes (DUBs). The complex regulatory system comprising enzymes that install Ub, recognize Ub-patterns and finally cleave Ub-protein conjugates has been termed the Ub-code. This code is further expanded by a family of small proteins called ubiquitin-like proteins (Ubls). Ubls share common structural features with Ub, such as the beta-grasp fold and the flexible C-terminus, and they are conjugated to substrate proteins, recognized and cleaved in a similar manner as described for Ub.

Because of the pivotal role that ubiquitylation plays in nearly all aspects of eukaryotic biology, dysfunctions in ubiquitylation pathways can lead to severe pathologies such as cancer and neurodegenerative diseases. Our ability to decipher functional outcomes of specific ubiquitylation events is, however, limited by the difficulty to create defined Ub-substrate conjugates *in vitro* and by the impossibility to site-specifically ubiquitylate substrates within living cells. Over the last decades, the chemical biology community has developed a variety of methods to meet the challenge of generating defined Ub-substrate conjugates. Most of these methods rely, however, on harsh conditions and are therefore limited to *in vitro* applications and to simple refoldable proteins.

This thesis addresses two major challenges in the field of generating Ub-substrate conjugates: (i) Development of a novel and versatile method to site-specifically monoubiquitylate complex, non-refoldable target proteins – both *in vitro* and in living cells. (ii) Development of a versatile toolbox to build defined polymeric Ub architectures that can be charged site-specifically onto target proteins.

In the first part of this thesis, we describe the development of a novel method to site-specifically monoubiquitylate substrate proteins in an inducible fashion by combining genetic code expansion, sortase-mediated transpeptidation and Staudinger reduction. We exploit an optimized variant of the bacterial transpeptidase sortase A, which facilitates the transpeptidation between a five-amino acid sequence (sortase recognition motif), and a polyglycine acceptor nucleophile. In order to apply sortase-mediated transpeptidation to site-specific ubiquitylation, we install the sortase recognition motif at the C-terminus of Ub through the introduction of two point mutations (mutation from LRLRGG to LALTGG). In parallel, the corresponding glycine-based acceptor nucleophile is site-specifically installed at a user defined site within the target protein, by introducing the unnatural amino acid AzGGK *via* genetic code expansion. AzGGK displays an azide-bearing diglycine moiety attached to the ε-amino group

of lysine and is incorporated site-specifically into a protein of interest using an evolved variant of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase (PylRS)/$tRNA_{CUA}$ pair. The azide moiety of AzGGK serves as a masking group, which can be removed upon Staudinger reduction with water soluble phosphine reagents, resulting in GGK-modified proteins displaying the sortase-reactive diglycine moiety. The sortase-catalyzed transpeptidation between a GGK-modified substrate protein and Ub bearing the sortase motif at its C-terminus results in Ub-protein conjugates that are linked via a native isopeptide bond and that only differ from their native counterparts by two point mutations in the flexible C-terminal region of Ub (R72A, R74T). This approach, which we termed sortylation, allows to site-specifically ubiquitylate complex, non-refoldable, multimeric proteins under native conditions, as shown by site-specific mono-ubiquitylation of homotrimeric PCNA (proliferating cell nuclear antigen). By transferring the sortase motif to SUMO1 (small ubiquitin like modifier 1) we furthermore show that the scope of sortylation can be expanded towards Ubls. As sortylation is based on an enzymatic reaction that works under physiological conditions, and Staudinger reduction of AzGGK-bearing proteins to GGK-bearing proteins is compatible with living cells, sortylation can be used for site-specific ubiquitylation in living mammalian cells. Transient overexpression of all components needed for sortylation allows to site-specifically ubiquitylate and SUMOylate proteins in an inducible manner in living HEK293T cells.

In conclusion, sortylation allows the site-specific attachment of Ubls to nonrefoldable, multidomain proteins and enables for the first time the inducible and ubiquitin-ligase-independent ubiquitylation of proteins in mammalian cells, providing a powerful tool to dissect the biological functions of ubiquitylation with temporal control.

The second part of this thesis presents the expansion of sortylation towards complex polymeric Ub topologies. The transpeptidation reaction catalyzed by sortase is an inherently reversible reaction, as the sortase recognition motif is re-installed in the product after successful transpeptidation. This makes the sequential use of the same sortase for the creation of defined polymers impossible. To overcome this limitation, we have identified sortases that recognize different sortase motifs in an orthogonal manner (orthogonal sortases). Due to their orthogonal behavior, these sortases can be used one after another (iterative sortylation), thereby allowing the assembly of complex Ub architectures, including hybrid, mixed and branched Ub chains.

The recent discovery of Ub/SUMO hybrid chains (polymeric architectures in which Ub moieties are attached to a preexisting SUMO chain) and their involvement in DNA double strand break (DSB) repair through their binding to Rap80 raised several questions regarding the connectivity of these polymeric architectures. Rap80 contains two Ub-binding domains and one SUMO-binding domain, which facilitate the recruitment of the BRCA1 complex to DNA DSBs. While the SUMO-binding domain of Rap80 binds to SUMO2, the Ub-binding domains are well known to specifically recognize K63-linked diubiquitin (diUb), however, how the K63-diUb is connected to SUMO2 is still an enigma. Our pair of orthogonal sortases allow us to create in total seven differently linked Ub/SUMO hybrid chains by using sortylation in an iterative manner. We used these hybrid chains in pull-down assays with Rap80 to investigate how the connectivity affects their binding behavior, gaining novel insights into this intriguing regulatory process.

Additionally, we combine iterative sortylation with enzymatic Ub-chain assembly using E1/E2 enzymes to gain access to target proteins decorated with complex Ub topologies. The

attachment of a K48-linked tetraubiquitin (tetraUb) to proteins is well-known to facilitate the proteasomal degradation of proteins bearing these structures. However, structural information of proteasomal substrate engagement is limited, since access to target proteins site-specifically decorated with defined Ub chains has been restricted. By using E1/E2-mediated assembly of two K48-linked diUbs we are able to generate and charge a K48-linked tetraUb site-specifically onto a target protein by using iterative sortylation. Furthermore, we incorporate AzGGK at two different positions into one Ub-monomer, which allows us to build and charge K11/K48-linked branched Ub architectures onto target proteins, signals that have recently been described to enhance proteasomal degradation during mitosis.

To conclude, we successfully expand sortylation towards complex Ub/Ubl architectures by using a newly identified pair of orthogonal sortases. This enables us to create and study differently connected hybrid chain as well as to generate and charge complex Ub architectures onto target proteins. We envision that the developed strategies will be useful to shed light onto processes employing complex Ub topologies for example by using structural biology approaches such as cryo electron microscopy.

Zusammenfassung

Die kovalente Konjugation von Ubiquitin (Ub) an Substratproteine stellt eine der vielseitigsten und flexibelsten posttranslationalen Modifikationen (PTMs) in eukaryotischen Zellen dar und ist für die Regulation fast aller zellulären Prozesse von zentraler Bedeutung. Der Ubiquitylierungsprozess wird über eine hochkonservierte enzymatische Kaskade vermittelt, an der E1/E2/E3-Enzyme beteiligt sind, die Ub über seinen C-Terminus an die ε-Aminogruppe von Lysinen innerhalb des Zielproteins durch Ausbildung einer Isopeptidbindung knüpfen (Ubiquitylierung). Einmal gebunden, kann Ub über einen seiner sieben Lysinreste oder über seinen N-Terminus als Plattform für weitere Ubiquitylierungsereignisse dienen - ein Prozess, der als Polyubiquitylierung bezeichnet wird. Polymere Ub-Architekturen, die auf Zielproteinen installiert sind, können in Länge, Verknüpfungsart und in Verzweigungspunkten variieren und dadurch hochkomplexe Topologien erzeugen, die von verschiedenen Klassen von Leserproteinen, die Ub-Bindungsdomänen tragen, in zahlreiche funktionelle Ereignisse übersetzt werden. Wie viele andere PTMs ist die Ubiquitylierung ein reversibler Prozess, der durch deubiquitylierende Enzyme (DUBs) streng reguliert wird. Das komplexe regulatorische System aus Enzymen, die die Ubiquitylierung installieren, Ub-Muster erkennen und schließlich Ub-Proteinkonjugate spalten, wurde als Ub-Code bezeichnet. Der Ub-Code kann des Weiteren durch eine Familie kleiner Proteine, die Ubiquitin-ähnliche Proteine (Ubls) genannt werden, erweitert werden. Ubls und Ub haben gemeinsame strukturelle Merkmale, wie z.B. den sogenannten β-graspfold und den flexiblen C-Terminus, und werden in ähnlicher Weise installiert, erkannt und prozessiert wie bereits für Ub beschrieben.

Aufgrund der zentralen Rolle von Ubiquitylierungsprozessen in fast allen Aspekten der eukaryotischen Biologie können Fehlfunktionen in diesen Prozessen zu schweren Pathologien wie Krebs und neurodegenerativen Erkrankungen führen. Unsere Fähigkeit, die funktionellen Folgen spezifischer Ubiquitylierungsereignisse zu entschlüsseln, wird jedoch durch die Schwierigkeit, definierte Ub-Substrat-Konjugate *in vitro* herzustellen, und durch die Unmöglichkeit, Substrate innerhalb lebender Zellen ortsspezifisch zu ubiquitylieren, eingeschränkt. In den letzten Jahrzehnten haben chemische Biologen eine Vielzahl von Methoden entwickelt, um definierte Ub-Substrat-Konjugate zu erzeugen. Die meisten dieser Methoden beruhen jedoch auf chemisch-harschen Bedingungen und sind daher auf *in vitro*-Anwendungen und auf einfache, rückfaltbare Proteine beschränkt.

Diese Arbeit befasst sich mit zwei großen Herausforderungen auf dem Gebiet der Generierung von Ub-Substrat-Konjugaten: (i) Entwicklung einer neuartigen und vielseitigen Methode zur ortsspezifischen Monoubiquitylierung komplexer, nicht rückfaltbarer Zielproteine - sowohl *in vitro* als auch in lebenden Zellen. (ii) Entwicklung eines vielseitigen Werkzeugkastens zum Aufbau definierter polymerer Ub-Architekturen, die ortsspezifisch an Zielproteine konjugiert werden können.

Im ersten Abschnitt dieser Arbeit beschreiben wir die Entwicklung einer neuartigen Methode zur positionsspezifischen Monoubiquitylierung von Substratproteinen auf induzierbare Weise durch Kombination von genetischer Codeerweiterung, sortase-vermittelter Transpeptidierung und Staudinger-Reduktion. Wir nutzen eine optimierte Variante der bakteriellen Transpeptidase Sortase A, die die Transpeptidierung zwischen einer fünf Aminosäuren umfassenden Sequenz (Sortase-Motiv) und einem Polyglycin-Akzeptor-

Nukleophil ermöglicht. Um die Sortase-vermittelte Transpeptidierung auf die positionsspezifische Ubiquitylierung anzuwenden, installieren wir das Sortase-Motiv am C-Terminus von Ub durch die Einführung von zwei Punktmutationen (Mutation von LRLRGG zu LALTGG). Parallel dazu wird das entsprechende Glycin-basierte Akzeptor-Nukleophil an einer benutzerdefinierten Stelle innerhalb des Zielproteins durch Einbau der unnatürlichen Aminosäure AzGGK mittels genetischer Codeerweiterung installiert. AzGGK trägt eine Azid-maskierte Di-Glycin-Einheit, die über eine Isopeptidbindung an die ε-Aminogruppe von Lysin gebunden ist und kann mittels einer neu evolvierten Variante eines *Methanosarcina barkeri* Pyrrolysyl-tRNA-Synthetase (PylRS)/$tRNA_{CUA}$-Paares positionsspezifisch in Proteine eingeführt wird. Nach Einbau von AzGGK in ein Substratprotein kann die Azidgruppe durch Staudinger-Reduktion mit wasserlöslichen Phosphinreagenzien entfernt werden, so dass die entsprechenden GGK-modifizierten Proteine den Sortase-reaktiven Di-Glycinrest aufweisen. Die Sortase-katalysierte Transpeptidierung zwischen einem Substratprotein, das GGK trägt, und Ub, das das Sortase-Motiv an seinem C-Terminus trägt, führt zu Ub-Proteinkonjugaten, die eine native Isopeptidbindung tragen und sich von ihren nativen Gegenstücken nur durch zwei Punktmutationen in der flexiblen C-terminalen Region von Ub unterscheiden (R72A, R74T).
Diese Methode, die wir Sortylierung genannt haben, ermöglicht die ortsspezifische Ubiquitylierung komplexer, nicht faltbarer, multimerer Proteine unter nativen Bedingungen, wie zum Beispiel die ortsspezifische Mono-Ubiquitylierung des homotrimeren Proteins PCNA (proliferating cell nuclear antigen). Durch Übertragung des Sortase-Motivs auf SUMO1 (small ubiquitin like modifier 1) zeigen wir darüber hinaus, dass Sortylierung auch auf Ubls angewandt werden kann. Da Sortylierung auf einer enzymatischen Reaktion beruht, die unter physiologischen Bedingungen funktioniert, und die Staudinger-Reduktion von AzGGK-modifizierten Proteinen mit lebenden Zellen kompatibel ist, kann unsere Ubiquitylierungsmehode zur ortsspezifischen Ubiquitylierung in lebenden Säugetierzellen genutzt werden. Die transiente Überexpression aller für die Sortylierung erforderlichen Komponenten ermöglicht die ortsspezifische Ubiquitylierung und SUMOylierung von Proteinen auf induzierbare Weise in lebenden HEK293T-Zellen. Zusammenfassend lässt sich sagen, dass Sortylierung die ortsspezifische Modifikation komplexer, multimerer Proteine mit Ubls erlaubt und zum ersten Mal die induzierbare und Ubiquitin-Ligase-unabhängige Ubiquitylierung von Proteinen in Säugetierzellen ermöglicht.

Der zweite Abschnitt dieser Arbeit beschäftigt sich mit der Erweiterung des Sortylierungs-Ansatzes auf komplexe polymere Ub-Topologien. Da die durch die Sortase katalysierte Transpeptidierungsreaktion eine inhärent reversible Reaktion ist (das Sortase Erkennungsmotif findet sich im Endprodukt wieder), ist die sequentielle Verwendung derselben Sortase zur Herstellung definierter Polymere nicht möglich. Um diese Beschränkung zu überwinden, haben wir Sortase-Enzyme identifiziert, die verschiedene Sortase-Motive auf orthogonale Weise erkennen (orthogonale Sortasen). Aufgrund ihres orthogonalen Verhaltens können unsere Sortasen nacheinander verwendet werden (iterative Sortylierung), wodurch der Aufbau komplexerer Architekturen ermöglicht wird.

Die jüngste Entdeckung von Ub/SUMO-Hybridketten (polymere Architekturen, in denen Ub-Einheiten an eine bereits bestehende SUMO-Kette gebunden sind) und ihre Beteiligung an der DNA-Doppelstrangbruch-Reparatur durch ihre Bindung an Rap80 wirft mehrere Fragen bezüglich der Konnektivität dieser polymeren Architekturen auf. Rap80 enthält

zwei Ub-bindende Domänen und eine SUMO-bindende Domäne, die die Rekrutierung des BRCA1-Komplexes für DNA-Doppelstrangbruch-Reparatur ermöglichen. Während die SUMO-bindende Domäne von Rap80 an SUMO2 bindet, sind die Ub-bindenden Domänen dafür bekannt, dass sie spezifisch K63-verknüpftes Diubiquitin (diUb) erkennen. Wie das K63-diUb mit SUMO2 verbunden ist, ist nicht bekannt, da die entsprechenden Hybridketten nicht rekombinant hergestellt werden können. Mit unserem Paar orthogonaler Sortasen sind wir in der Lage, insgesamt sieben unterschiedlich verknüpfte Ub/SUMO-Hybridketten durch Sortylierung in iterativer Weise zu erzeugen. Diese Hybridketten werden in Pull-down-Assays mit Rap80 eingesetzt, um das Bindungsverhalten von unterschiedlich verknüpften Hybridketten zu untersuchen und so neue Einblicke in den regulatorischen Prozess der DNA-Reparatur zu erhalten.

Zusätzlich kombinieren wir die iterative Sortylierung mit enzymatischer Ub-Ketten-Assemblierung unter Verwendung von E1/E2-Enzymen, um Zugang zu Zielproteinen zu erhalten, die mit komplexen Ub-Topologien dekoriert sind. Die Anheftung eines K48-verknüpften Tetraubiquitins (tetraUb) an Proteine führt bekanntlich zum proteasomalen Abbau von Proteinen, welche diese Strukturen tragen. Allerdings ist die strukturelle Information über Substrat-Engagement im Proteasom begrenzt, da der Zugang zu Zielproteinen, die ortsspezifisch mit diesen Strukturen dekoriert sind, eingeschränkt ist. Durch die Verwendung einer E1/E2-vermittelten Assemblierung von zwei K48-verknüpften diUbs generieren wir mittels iterativer Sortylierung ein K48-verknüpftes tetraUb und installieren dieses darauffolgend auf einem Zielprotein. Darüber hinaus bauen wir AzGGK an zwei verschiedenen Positionen in ein und dasselbe Ub-Monomer ein, was es uns ermöglicht, K11/K48-verknüpfte verzweigte Ub-Architekturen auf Zielprotein zu installieren. Derartige Ub-Ketten sind optimale Signale für den proteasomalen Abbau während der Mitose, wie kürzlich gezeigt wurde.

Zusammenfassend zeigen wir, dass Sortylierung durch Verwendung eines neu identifizierten Paares orthogonaler Sortasen erfolgreich auf komplexe Ub/Ubl-Architekturen erweitert werden kann. Dies ermöglicht es uns, unterschiedlich verbundene Hybridketten zu erzeugen und deren Affinät für verschiedene Erkennungsproteine zu untersuchen. Des Weiteren erlaubt unser Ansatz die Modifikation von Zielproteinen mit komplexen Ub/Ubl-Architekturen, und bietet damit einen Grundstein für strukturbiologische Ansätze wie Kryo-Elektronenmikroskopie, um Licht auf biologische Prozesse zu werfen, die komplexe Ub-Topologien verwenden.

CHAPTER 1

Introduction

Parts of the following chapter are based on the manuscript:

Fottner, M.; Cigler, M.; Lang, K. Genetic code expansion approaches to decipher the ubiquitin code. *to be submitted*

1. Introduction

1.1 Posttranslational modifications – Diversification of the proteome

All living organisms respond to exogenous and endogenous signals on a cellular level by altering the abundance and status of the key players of life: proteins. Fundamental processes such as DNA replication, catalysis of metabolic reactions as well as the transport of molecules from one place to another would not be possible without the help of these extraordinary biomolecules, placing them right at the centre of cellular life as we know it. Proteins are macromolecules assembled from a distinct set of twenty building blocks, called amino acids, which are connected via peptide bonds. The different sidechains of the twenty proteinogenic amino acids vary in size, shape, charge and hydrophobicity, thereby conveying structural and chemical diversity to proteins[1]. As already stated in the central dogma of molecular biology in 1958[2], the molecular building plan of proteins is encoded in DNA regions called genes which are transcribed into RNA. A subtype of the latter, called mRNA, can then act as a template for protein synthesis facilitated by a molecular machine, the ribosome. Back in the early 2000's, when the world's largest collaborative biological project, the human genome project, ended and mankind had its blueprint in hand we were confronted with a rather surprising truth: Human complexity is based on as little as approx. 20,000 protein coding genes[3,4].

However, this is only half the truth. In contrary to the 1941 stated one-gene one-enzyme theory[5], nature has evolved an arsenal of mechanisms to diversify the composition of proteins beyond their genetically encoded blueprint. Proteomic diversification is achieved through complex processes both on transcriptional as well as on translational level resulting in the existence of individual molecular appearances of proteins, so-called "proteoforms"[6]. Beside intended mechanisms to create proteomic diversity, errors in DNA replication (10^{-8}), transcription of DNA into mRNA (10^{-5}) and translation of mRNA (10^{-4} – 10^{-3}) also have to be accounted for generating novel proteoforms (Figure 1.1)[7]. For example, with the given error rate for translation, 15 % of all average-length proteins will have at least one misincorporation[8].

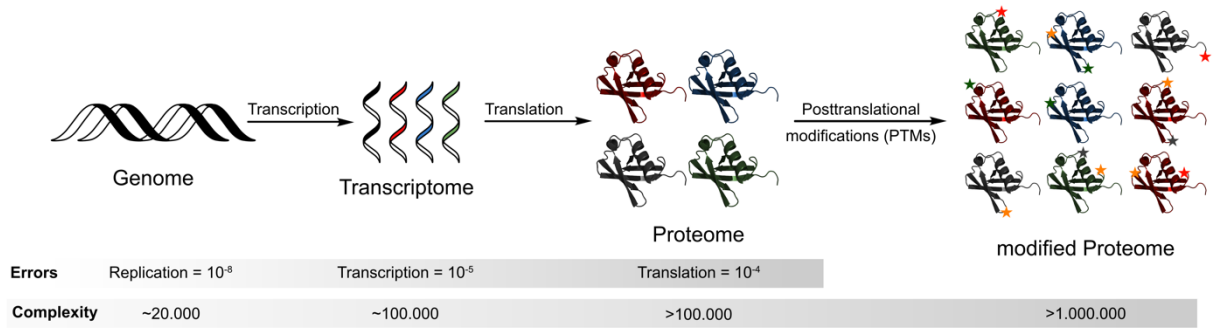


Figure 1.1 Overview of diversification mechanisms. Diversification of the approx. 20.000 human genes is mediated either by spontaneous errors in replication, transcription and translation or by intentional mechanisms on transcriptional and translational level. Figure is partially adapted from[9].

Once DNA is transcribed into pre-mRNA, a process called alternative splicing, is capable to create multiple proteins from one gene. This is facilitated by including and excluding certain exons of a pre-mRNA leading to the generation of different proteoforms. Bioinformatical studies revealed that up to 93 % of protein coding genes have at least one

alternatively spliced variant, increasing the number of human proteins to roughly 70,000[4,10]. Furthermore, RNA can be co- and post-transcriptionally modified for example by deamination -which is the most prevalent RNA modification - with about 4.5 million adenosine to inosine editing events occurring in RNA transcripts[11]. Importantly, only about 4,000 of these editing events are nonsynonymous and located in coding sequences, ultimately leading to mutations of the respective amino acids.

In addition to the above mentioned mechanisms acting on transcriptional level, nature's true escape from genetic imprisonment is represented by post-translational modifications (PTMs) of proteins[12]. Despite the sheer endless sequence space available from combining the 20 proteinogenic amino acids, PTMs are indispensable to adapt in a temporally flexible and economically reasonable manner to extra- and intracellular cues. PTMs convey a plethora of functions by acting as molecular switches, which fine-tune the interactome of certain modified proteins with DNA, metabolites or other proteins[13]. Nearly every eukaryotic protein is targeted by PTMs[4], and in many cases one protein is targeted by multiple PTMs at different residues, which can lead to the so called PTM crosstalk[13]. This is exemplified by histone H3 which can be modified by a vast array of PTMs such as phosphorylation, various types of acylations, mono-, di- and trimethylation, ADP-ribosylation and ubiquitylation giving rise to several hundred proteoforms for each H3 gene[4,14,15]

1.1.1 Backbone-targeting PTMs

PTMs can be classified into two categories: PTMs that either (1) target the peptide backbone or (2) target the amino acid side chain[16]. Backbone-targeting PTMs (bbPTMs) are introduced either via spontaneous maturation processes or via the action of specific enzymes[17].

The most prominent example of a post-translational maturation process might be the formation of the fluorophore of green fluorescent protein (GFP) through a sequence of successive cyclization, dehydration and oxidation reactions (Figure 1.2 a) [18]. Another common bbPTM involves asparagine and aspartate amino acid residues, in which the backbone amide nitrogen attacks the side-chain amide/carboxyl forming an unstable succinimide intermediate which can be hydrolysed to yield either aspartate or isoaspartate. Importantly, the succinimide intermediate is prone to epimerization, which can lead to the formation of D-aspartate and D-isoaspartate (Figure 1.2 b). Even though the formation of aspartate (from asparagine) and isoaspartate is generally thought to be a protein "aging" related process and has, for example, been described in the complementary-determining region (CDR) of antibodies[19], it has also been shown to exert specific functions in "young" proteins. For example, an isoaspartate residue was observed in MurA, an essential bacterial enzyme involved in cell-wall synthesis, and was proposed to trigger an unusual turn structure that stabilizes the protein (Figure 1.2 c)[20].

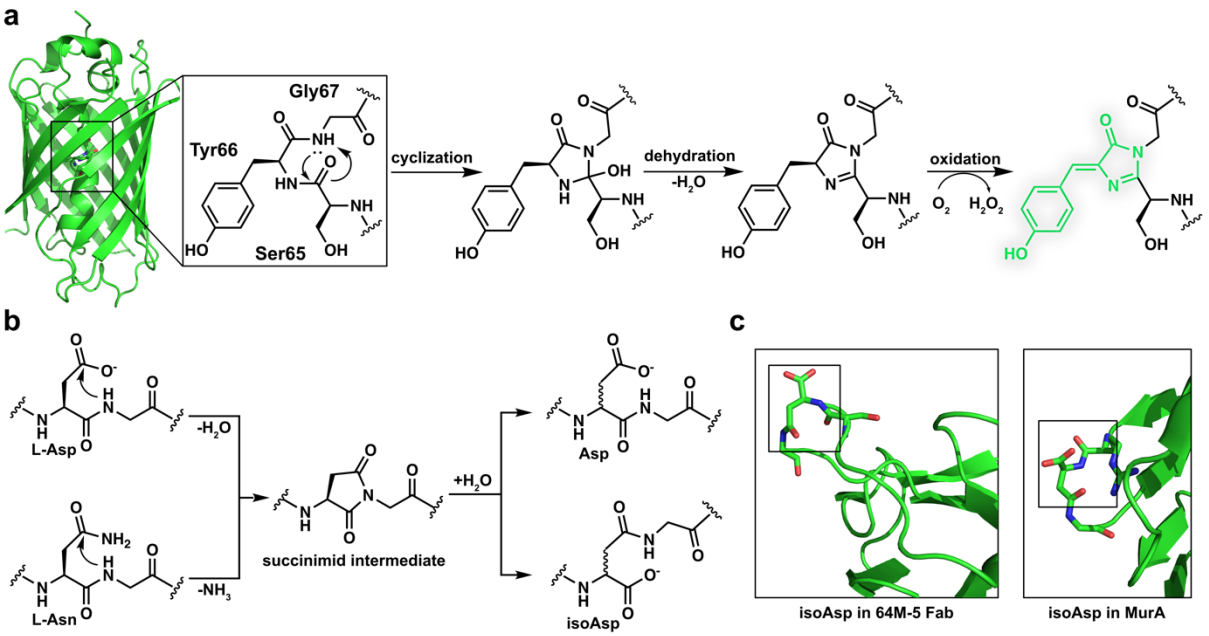


Figure 1.2 bbPTMs based on spontaneous maturation processes. a) The chromophore formation of green fluorescent protein (GFP, PDB 1GFL). **b)** Spontaneous bb rearrangement of L-aspartate and L-asparagine to D/L-isoaspartate and D/L-aspartate. **c)** Crystal structures of isoAsp residues in the Fab fragment 64M-5 (PDB 6KDH) and in MurA (PDB 1EJD).

Next to these spontaneous backbone modifying processes stands the modification of the peptide backbone by enzymes. A ubiquitous form of post-translational modification is proteolytic processing, which includes the removal of the N-terminal methionine, cleavage of signal peptides, processing of (pre)proproteins and cleavage of polyproteins as well as precursor proteins[1,21,22]. While the cleavage of the N-terminal methionine is facilitated co-translationally by the methionine aminopeptidase at the ribosome[23], many proteolytic processing events take place extracellularly, such as the activation of trypsinogen to trypsin, which is catalyzed by the enzyme enteropeptidase in the intestinal mucosa[22]. Most peculiar enzyme-mediated bbPTMs, such as didehydroaspartate[24], α–methylglutamine and thioglycine[25], have been observed close to the active site of an archaeal methyl-coenzyme M reductase (Figure 1.3 a). The exact enzymes involved in didehydroaspartate and α–methylglutamine formation have not been discovered yet, but recently the backbone kinase YcaO and an auxiliary protein TfuA have been proposed to catalyse the thioamide formation leading to the generation of thioglycine[26].

A special form of bbPTMs is the modification of the protein N- and C-termini. The primary amine of the protein N-terminus can be modified by a variety of acylations including the attachment of formyl, acetyl, propionyl, myristoyl and palmitoyl groups (Figure 1.3 b)[27]. It has been shown that N-terminal acetylation is involved in the regulation of protein half-life, protein folding as well as fine tuning PPIs[28], while myristylation and palmitoylation is mainly thought to serve as a membrane anchor[29]. Recent studies have shown that up to 90% of the protein N-termini in mammalian cells undergo co-translational acetylation[30,31]. Next to acylations, the modification of the N-terminus by alkylation, is represented by mono-, di- and trimethylation events (Figure 1.3 b).

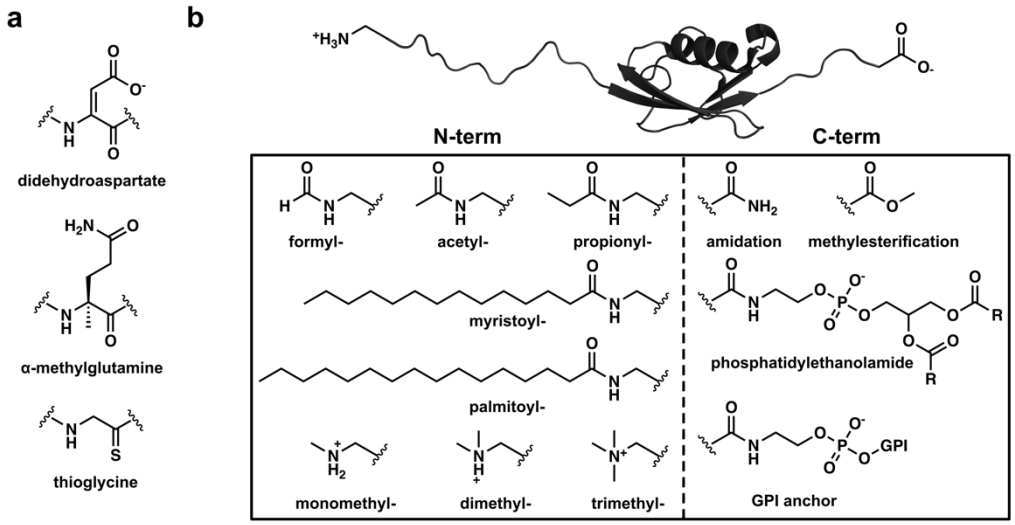


Figure 1.3 Enzymatic bb modification. a) Structures of rare bbPTMs within the peptide chain. **b)** Structures of N- and C-terminal bb PTMs (PDB 2N9E).

While the methylation of N-termini has been known for decades, the corresponding methyl transferases have only recently been identified and the modification is thought to be irreversible since N-terminal demethylases still remain elusive[27]. In contrast to the protein N-terminus, the C-terminus is often neglected regarding its PTMs, which is due to the inherently less reactive carboxyl group leading to a reduced extent of modification. Nevertheless, C-terminal PTMs can range from small modifications such as amidation via glycine oxidation and elimination of an N-glyoxylated peptide[32], and C-methyl esterification[33] to the attachment of large moieties like phosphatidylethanolamine[34] or glycosylphosphatidylinositol (GPI) anchors[35], which are usually involved in membrane targeting (Figure 1.3 b).

Another PTM which should be mentioned within the context of backbone modifications is protein splicing. Analogously to splicing events on RNA level, protein splicing involves the excision of an intervening polypeptide sequence (intein) from a protein, resulting in the fusion of the flanking polypeptides (exteins) via a peptide bond[36]. This exceptional process, which is independent of external factors, has found a myriad of applications in the field of protein chemistry, including the formation of N-terminally modified proteins, such as N-terminal thioesters which lay the foundation for native chemical ligation (NCL, see Chapter 1.4)[37,38].

In theory the peptide bond between two amino acids of a protein can adopt either a *cis* or a *trans* conformation, but for most amino acids the *trans* state is energetically extremely favoured compared to the *cis* state. Proline is unique in that regard. A study from the late 90s showed that in a set of 571 proteins only 0.03 % of X-nonP peptides, but 5.2 % of the X-P peptide bonds are in *cis*-conformation[39,40]. *Cis-trans* isomerization in X-P motifs results from the lower free energy difference between the two states and can occur spontaneously but is also catalyzed by a distinct class of enzymes, the so-called peptidyl proline *cis-trans* isomerases [41]. The structural difference between the *cis* and the *trans* state of a peptide bond within a protein acts as a molecular backbone switch that can fine-tune the interactome of a certain protein. Peptidyl proline *cis-trans* isomerases accelerate the *cis-trans*/*trans-cis* switches by several order of magnitudes and can be highly regulated. For example, the peptidyl proline *cis-trans* isomerases Pin1 exerts its action preferentially on peptide bonds where serine/threonine residues next to the proline residue are phosphorylated (pS/pT-P motifs, see chapter 1.1.2.1), which represents another excellent example of PTM crosstalk[42].

1.1.2 Side chain-targeting PTMs

The second broad category of PTMs are side chain targeting PTMs (scPTMs) with ~400 different modifications identified to date[4]. Nearly every side chain of the 20 natural amino acids can be targeted by a variety of modifications changing their chemical properties[16]. There are many ways to categorize scPTMs, either by the side chain targeted, their abundance, the complexity of the modification or by the functional outcome. The following paragraphs, however, will discuss the most abundant scPTMs subdivided in simple and complex modifications.

1.1.2.1 Simple side chain modifying PTMs

The most abundant scPTM, with approx. 38,000 annotated modification sites across around 20,000 human proteins is phosphorylation[4]. Protein phosphorylation is a reversible process, catalyzed by a class of phosphotransferases called kinases and reversed by the action of phosphatases[43]. The human proteome encodes for more than 500 kinases (also referred to the kinome), which facilitate the phosphorylation of mostly serine, threonine and tyrosine side chains in an ATP-dependent manner[44]. However, in recent years modern mass spectrometry approaches enabled the discovery of non-canonical phosphorylation sites on the side chains of histidine, arginine, lysine, asparagine, glutamate and cysteine, opening up new avenues in phosphorylation research (Figure 1.4)[45,46]. The addition of the negatively charged phosphate group to protein side chains acts as a molecular switch by fine tuning PPIs, for example serine and threonine phosphorylation is "read" by a specific class of phospho-binding domains, the so called 14-3-3 domains, which play a crucial role in eukaryotic signal transduction[47,48]. The fundamental biological importance of phosphorylation is exemplified by its involvement in numerous human diseases such as cancer, and kinase inhibitors represent well proven therapeutic agents for treating various diseases[49].

Next to phosphorylation, the second most abundant scPTM is acetylation (> 6,000 modification sites), which belongs to the group of acylations[50]. The most abundant type of acetylation is the attachment of an acetyl-group to the side chain of lysine (Figure 1.5 a).

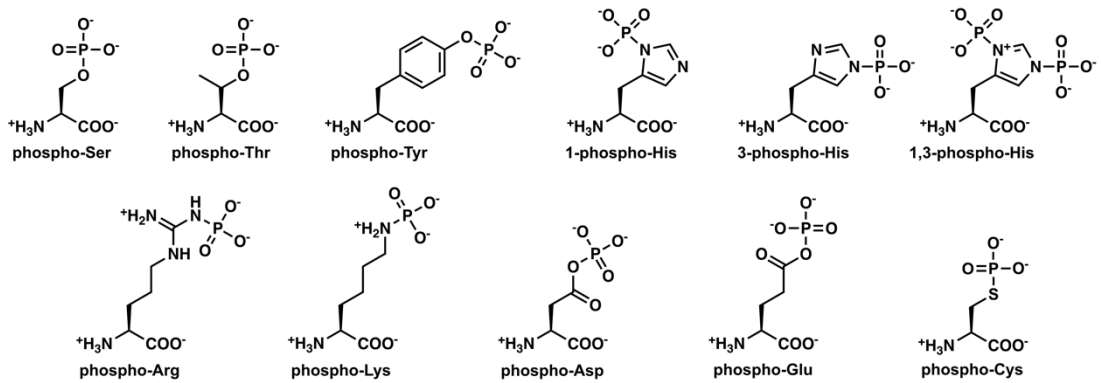


Figure 1.4 Overview of phospho-scPTMs.

Lysine acetylation is mediated by lysine acetyl transferases (KATs), which transfer an acetyl group from acetyl-coenzyme A (acetyl-CoA) to the Nε amine of lysine, a process which is reversed by the action of lysine deacetylases (KDACs). Originally identified in the context of epigenetic regulation of histones[51], acetylation is also involved in transcription, DNA damage repair and autophagy[52]. Downstream effects of protein acetylation are decoded by proteins containing so-called reader domains, such as bromodomains, YEATS and double/tandem PHD finger domains, which harbour a hydrophobic cavity that specifically recognizes the formerly positive charged, now acetylated lysine side chain[53].

As already indicated above, lysine acetylation belongs to the large group of lysine acylations. Analogous to backbone acylations, the side chain of lysine can also be modified by the addition of fatty acids such as myristoleic and palmitic acid, which play an essential role in the membrane localization of cytokines like interleukin 1α and TNFα[54-56]. Alongside the addition of fatty acids, lysine acylations comprise a variety of modifications including: formylation[57], propionylation[58], butyrylation[59], hydroxyisobutyrylation[60] and crotonylation[61] as well as the recently discovered malonylation[62], succinylation[62] and glutarylation[63], which are exceptionally fascinating since they reverse the net charge of the lysine side-chain from +1 to -1 (Figure 1.5 a). Besides the highly abundant acylations of lysine it is also important to mention acylations of hydroxyl/sulfhydryl groups yielding the respective oxo/thioester-modified amino acids. Acetylation of serine/threonine has only been identified ca.10 years ago, when it was discovered that a bacterial virulence factor of certain *Yersinia* species catalyzes the attachment of this PTM[64]. But soon it became clear that serine acetylation is also involved in histone regulation[65]. In contrast to serine/threonine acetylation, cysteine palmitoylation (Figure 1.5 b) is a well-studied scPTM that regulates the subcellular localization of proteins by increasing their membrane affinity[66].

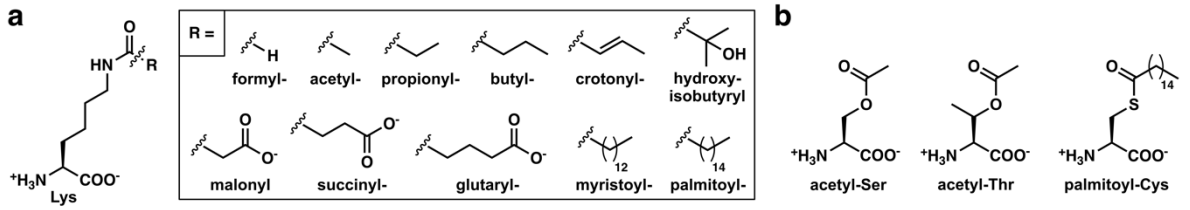


Figure 1.5 Overview of acylation based scPTMs. a) Acylation on lysine. **b)** Acylation of serine/threonine/cysteine.

A final class of simple scPTMs which should be mentioned within this chapter is protein alkylation. The transfer of alkyl groups to the side chains of amino acids can be grouped into the attachment of methyl groups and into the addition of the isoprenyl-based moieties such as the farnesyl and geranylgeranyl groups[16]. Protein methylation occurs primarily at the side chains of lysine and arginine and is installed by lysine methyltransferases (KMTs) and arginine methyltransferases (PRMTs), which use *S*-adenosylmethionine (SAM) as methyl donor (Figure 1.6)[67]. Both arginine and lysine methylation are recognized by specialized reader domains (e.g. chromodomains for methylated lysine[68] and tudor domains for methylated arginine[69]) and can be reversed by the action of distinct lysine/arginine demethylases[70].

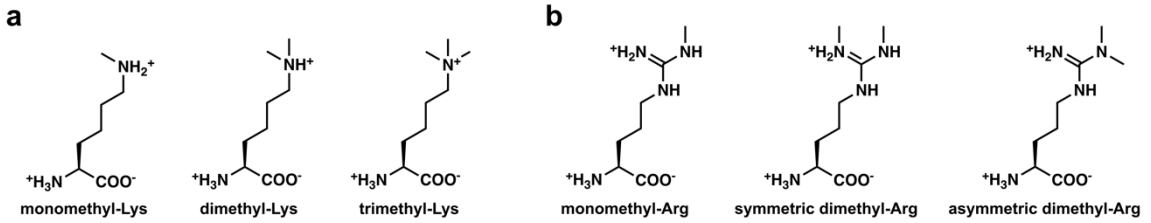


Figure 1.6 Overview of Lys and Arg methylation. a) Lys methylation. **b)** Arg methylation.

Together with acylation- and phosphorylation-based scPTMs, the methylation of arginine and lysine residues of histones sets the foundation for a complex regulatory network known as the histone code[71]. In contrast to the PTMs mentioned so far, lysine and arginine methylation have the ability to be installed in a progressive manner. The functionalization status of the Nε amine of lysine can range from mono- to di- and trimethylation[72] while the guanidinium group of arginine possesses the potential to be monomethylated as well as symmetrically and asymmetrically dimethylated (Figure 1.6b)[73]. These "methylation steps" add an additional layer of information and give rise to endless combinatorial possibilities for the PTM status of a target protein. While the addition of a methyl group is a rather small modification and is mainly thought to fine tune the interaction with the designated reader domains, the attachment of farnesyl and geranyl groups to the sulfhydryl side chain of cysteine is comparably large and can act as membrane anchor (Figure 1.7)[74]. Prenylations are mediated by the corresponding farnesyl/geranylgeranyl transferases under consumption of farnesyl/geranylgeranyl pyrophosphate. More than 100 prenylation sites have been experimentally proven to date[75]. Prenylation occurs typically at cysteine residues of a defined C-terminal CaaX motif (where *a* represents an aliphatic amino acid and X any amino acid), followed by the proteolytic cleavage of the aaX motif and methyl esterification of the C-terminus[76].

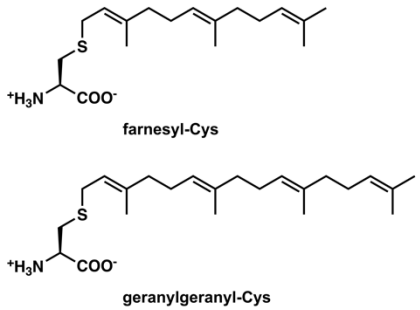


Figure 1.7 Cys farnesylation and geranylgeranylation.

Due to the large variety of exciting side chain PTMs it is unfortunately not possible to discuss all of them within this context. There are however some left which should be briefly highlighted. Collagen for example, which is the most abundant protein in mammals (approx. 30 % of whole-body protein content), owes its stability to several posttranslational hydroxylation events at specific amino acids. The most common is the hydroxylation of proline, which is mediated by a class of enzymes called prolyl hydroxylases, resulting in 3/4-hydroxyproline (Figure 1.8). Due to the high abundance of collagen, 4 -hydroxyproline comprises roughly 4 % of all amino acids found in animal tissue[77].

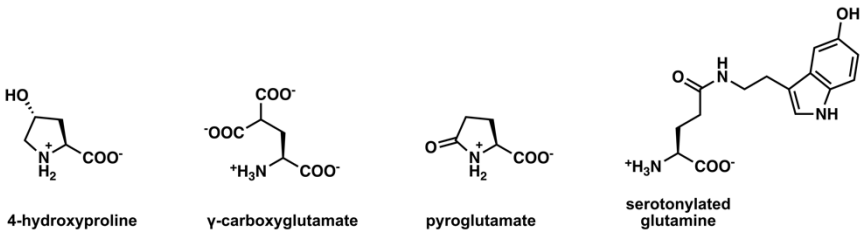


Figure 1.8 Structures of 4-hydroxyproline, γ-carboxyglutamate, pyroglutamate and serotonylated glutamine.

Two PTMs frequently found in extracellular proteins are 4-carboxyglutamate and pyrrolidone carboxylic acid (Figure 1.8). The attachment of the carboxy group to glutamate plays important roles in the coagulation cascade since the resulting 4-carboxyglutamate is able to bind Ca^{2+} with high affinity[78]. Pyrrolidone carboxylic acid on the other hand is formed by the cyclization of N-terminal glutamine residues in order to protect extracellular proteins from degradation by aminopeptidases (Figure 1.8)[79]. One last, just recently discovered PTM is the serotonylation of glutamine, which has been identified on a glutamine residue on Histone H3 and is thought to colocalize with a close by trimethyl lysine moiety[80]. Glutamine serotonylation is thought to influence transcription and thus further increases the complexity of the histone code.

1.1.2.2 Complex side chain modifying PTMs

In vast contrast to the above discussed rather simple scPTMs stand processes that install large, complex and sometimes polymeric structures to the side chains of amino acids. These modifications include the attachment of (1) carbohydrates (glycosylation), (2) linear or branched mono/polynucleotides (ADP-ribosylation) and (3) small proteins or complex architectures thereof (e.g. ubiquitylation).

Covalent glycosylation of amino acid side chains is a common process in eukaryotic cells and has many facets. The different types of glycosylation patterns can be classified by the nature of the linkage between the carbohydrate and the protein into N-glycosylation, O-glycosylation, C-glycosylation and S-glycosylation (Figure 1.9)[81].

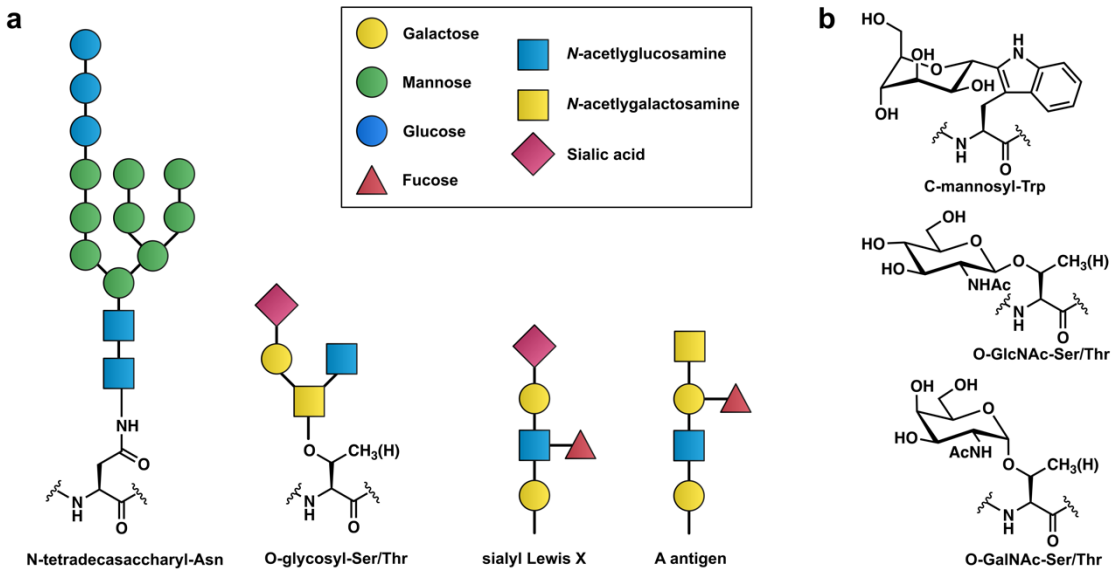


Figure 1.9 Overview of protein glycosylation. a) Examples of monosaccharide attachment. **b)** Examples of complex glyco-architectures. Partially adapted from [82].

The most prominent forms of glycosylation are N- and O-glycosylations, with several thousand predicted conjugation sites for N-glycosylation and more than 200 for O-glycosylation[83]. Protein N-glycosylation is a flexible process and usually starts with the *en bloc* attachment of a preassembled, branched tetradecasaccharyl fragment to the carboxamide nitrogen asparagine side chains in a N-X-S/T motif (Figure 1.9 a)[84]. This conjugation occurs co-translationally in the lumen of the endoplasmic reticulum (ER) and is catalyzed by a

membrane-bound multienzyme complex called oligosaccharyltransferase[85]. N-glycosylation plays an essential role in ER quality control where the nascent tetradecasaccharyl ($Glc_3Man_9GlcNAc_2$)-bearing protein undergoes glucose trimming to a dodecasaccharyl-bearing protein ($GlcMan_9GlcNAc_2$), initiating thereupon the folding process by the lectin chaperones calnexin and calreticulin. In case that protein folding fails, further glycan trimming leads to the recognition of the glycosylated protein by components of the ER-associated degradation machinery resulting in retro-translocation to the cytosol and ultimately to proteasomal degradation[86]. In contrast, successful folding promotes the transportation of the N-glycosylated protein to the Golgi apparatus. There, the glycosylation pattern is further modified in a process called "terminal glycosylation", followed by secretion of the protein [85].

O-glycosylation, on the other side, is generally less complex and typically occurs on the amino acid side chains of serine and threonine. The most common modifications are the attachment of the *N*-acetylglucosamine (GlcNAc) and *N*-acetylgalactosamin (GalNAc, Figure 1.9 a)[87]. GalNAc conjugation to proteins takes place in the Golgi apparatus after protein folding and prior to secretion, while GlcNAc is attached to proteins in the cytoplasm or in the nucleus[88]. In contrast to GlcNAc, which is a strictly monomeric scPTM, O-GalNAc modifications serve as a platform for the addition of further sugars forming complex mucine structures whose terminal sugar building blocks are essential for recognition by lectins[89]. Common structural elements formed by N- and O-glycans are antigens. Proteins bearing these glycan structures (e.g. A antigen/sialyl Lewis X, Figure 1.9 a) are presented on erythrocytes and thereby form the basis of our blood group system[82].

Another rather simple form of protein glycosylation represents C-mannosylation, where a monomeric α-mannopyranosyl is attached to the C2 carbon of the indole of tryptophane (Figure 1.9 b)[84]. C-mannosylation occurs at the first tryptophane in WxxW motifs and is quite rare with less than 100 reported conjugation sites, which are mainly found in secretory and membrane-bound proteins[90]. Even less abundant is S-glycosylation, where monomeric glucose is attached to the sulfhydryl side chain of cysteine, which has only been reported in glyocpeptides[91,92].

From a chemical perspective, a closely related complex scPTM is protein ADP-ribosylation, which plays important roles in DNA damage repair, metabolism and chromatin structure. Analogous to protein glycosylation, ADP-ribosylation results in a glycosidic linkage, which can either be N-, O- or S-glycosidic, whether the conjugation takes place to aspartate, glutamate, serine, tyrosine, arginine, cysteine, lysine, histidine or asparagine[93,94]. The attachment of an ADP-ribose (ADPr) group is carried out by a distinct class of ribosyl transferases. Mechanistically seen, ribosyl transferases cleave the nicotinamide moiety of the coenzyme NAD^+ resulting in a transition state of the ADPr group involving a ribaoxacarbenium ion which can be nucleophilically attacked by diverse amino acid side chains[16]. This process is reversed by ADP-ribose glycohydrolases. From a structural point of view, it is important to keep in mind that the addition of an ADPr group to a target protein can significantly influence its properties, since the ADPr group carries two negative charges, two ribose moieties and an adenine ring capable of hydrogen bonding and other interactions (Figure 1.10)[95].

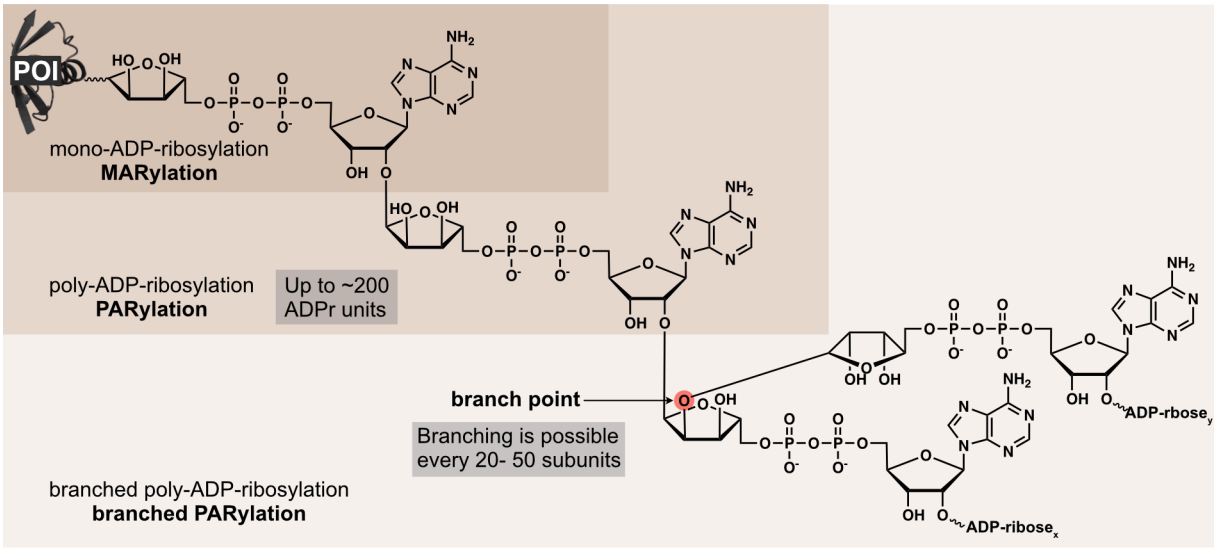


Figure 1.10 Overview of protein ADP-ribosylation. Partially adapted from [94].

Next to mono-ADP-ribosylation (MARylation) exists poly-ADP ribosylation (PARylation), that is facilitated by poly (ADP ribose) polymerases (PARPs), which are able to polymerize more than 200 ADPr groups (Figure 1.10)[96]. These polymeric structures gain additional complexity through branching events that confer distinct conformations[94]. The functional outcome of MARylation and PARylation is mediated by ADPr binding domains, such as macrodomains which specifically bind mono ADPr attached to target proteins or PBZ domains that bind to terminal poly ADPr structures by recognizing both, the terminal ADPr and the penultimate adenine residue[95].

While both glycosylation and ADP ribosylation are based on the attachment of a carbohydrate or a derivative via a glycosidic bond, the next complex scPTM to be discussed is based on the attachment of a whole protein to the side chain of amino acids of target proteins; a process called ubiquitylation.

1.2 The ubiquitin system

Ubiquitylation is characterized by the covalent attachment of the small protein ubiquitin (Ub) to target proteins and represents one of the most abundant and versatile PTMs. Ub is highly conserved throughout eukaryotic evolution and adopts a compact beta-grasp fold with a six amino acid long unstructured C-terminus. During ubiquitylation the C-terminal carboxylate of Ub is attached to the ε-amino group of a lysine residue within a substrate protein to form an isopeptide bond[97]. Several proteomic studies revealed that there are more than 10,000 ubiquitylation sites across ~5000 proteins[98,99], leading to the assumption that the majority of proteins experience ubiquitylation within their lifetime[100]. Famously studied in the context of proteasomal targeting[97], ubiquitylation controls numerous cellular processes including DNA-repair, endocytosis, autophagy, nuclear transport and chromosomal organisation[101]. Because of its pivotal role in nearly all aspects of eukaryotic biology, dysfunctions in ubiquitylation pathways can lead to severe pathologies such as cancer and neurodegenerative diseases[102]. Cells dedicate an orchestrated arsenal of enzymes to install (writer proteins), translate (reader proteins) and reverse ubiquitylation (eraser proteins) whose action will be discussed within the next subchapters.

1.2.1 The conjugation machinery

Site-specific ubiquitylation is accomplished through a multistep process employing a highly coordinated enzymatic cascade comprising E1, E2 and E3 enzymes that successively activate, conjugate and finally ligate Ub to target proteins. While the human genome encodes only two E1 enzymes, there are ~30 E2s and an estimated number > 600 E3 ligases that facilitate Ub transfer[103]. Ub is encoded in mammals by four genes and is synthesized as an inactive precursor. Therefore, processing by a class of proteases called deubiquitylating enzymes (DUBs, see chapter 1.2.3), is essential to reveal the characteristic C-terminal diglycine motif making it available for conjugation (Figure 1.11)[104]. Ubiquitylation is an ATP-dependent process, which is initiated by the generation of an activated Ub-AMP intermediate that is charged onto the active site cysteine of an E1 enzyme followed by transthioesterification onto the cysteine residue of an E2 enzyme. The final Ub-transfer to substrate proteins is facilitated by E3 ligases, which have historically been grouped into HECT and RING type E3 ligases. Their mode of action is inherently different: RING-type E3 ligases catalyze the direct transfer of Ub from the E2 enzyme to the substrate protein, whereas the mechanism of HECT-type E3 ligases involves formation of an obligate Ub-HECT thioester intermediate to transfer Ub to substrate proteins (Figure 1.11)[105]. Additionally and in contrast to the archetypal RING and HECT E3 ligases, the family of RBR (RING-BetweenRING-RING) E3s, exerts a hybrid mechanism in which the RING1 domain recognizes the E2-Ub intermediate and orchestrates the transfer of Ub onto the active site cysteine of the RING2 domain followed by transfer to the substrate protein[106]. Recently, a fourth family joined the group of E3-ligases, the RCR-type family, whose only member MYCBP2 contains two catalytic cysteines which relay Ub onto its substrate protein[107].

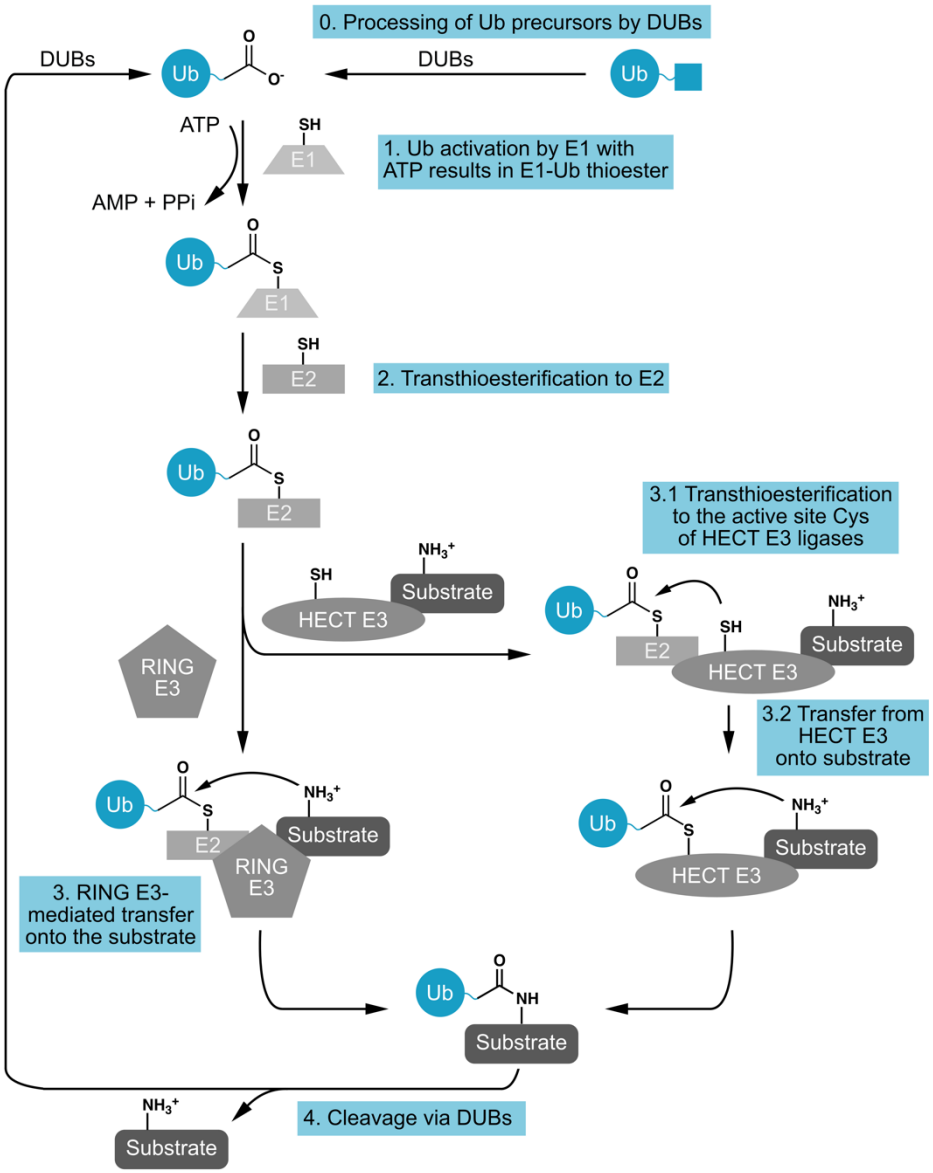


Figure 1.11 Overview of protein ubiquitylation and deubiquitylation. After proteolytic processing of one of the four Ub precursors by DUBs, Ub is charged in an ATP-dependent manner onto the active site of an E1 enzyme forming a thioester intermediate. After transthioesterification onto an E2 enzyme Ub is transferred to a lysine residue of a substrate protein by the action of an E3 ligase. In case of HECT E3 ligases Ub transfer to the substrate involves a Ub-E3 thioester intermediate. In contrast, RING E3 ligases merely mediate the transfer of Ub from the E2 enzyme to the substrate by forming a platform.

Most intriguingly, MYCBP2 is the first discovered human E3 ligase that preferentially mediates so-called non-canonical ubiquitylation. Until two decades ago, ubiquitylation was thought to happen exclusively at lysine residues but in recent years ubiquitylation at serine, threonine and cysteine residues has been identified, giving rise to novel cellular functions of Ub[108]. Alluringly, non-canonical ubiquitylation is often mediated by effector proteins of pathogens. This can be explained by the central role of the Ub system in many regulatory processes which forces bacteria and viruses either to bypass or exploit the Ub system to maximize their chance of infection[109,110].

Ubiquitylation events can be divided into monoubiquitylation and polyubiquitylation. In its simplest form, monoubiquitylation represents the attachment of one Ub to a single lysine residue within a substrate protein. This can be expanded by so-called multi-monoubiquitylation, where several Ub moieties are attached to different lysine residues within a target protein.

Attachment of Ub to the N-terminus or to one of the seven lysine residues (K6, K11, K27, K29, K33, K48 and K63) of another Ub forms the basis of polyubiquitylation, which can be divided into homotypic and heterotypic events. Ub chains linked via the same type of lysine residue are classified as homotypic chains, whereas heterotypic chains can be further subdivided into mixed and branched chains. Mixed chains are composed of Ub monomers linked via different types of lysine residues and thereby form the basis for branched chains, in which one Ub monomer acts as a platform for multiple Ub attachments leading to the formation of complex branched architectures (Figure 1.12)[100]. In a Ub chain the position of the donor Ub moiety, the one that is activated at its C-terminus, is defined as distal while the acceptor Ub moiety, the one that bears the target lysine residue, is called proximal.

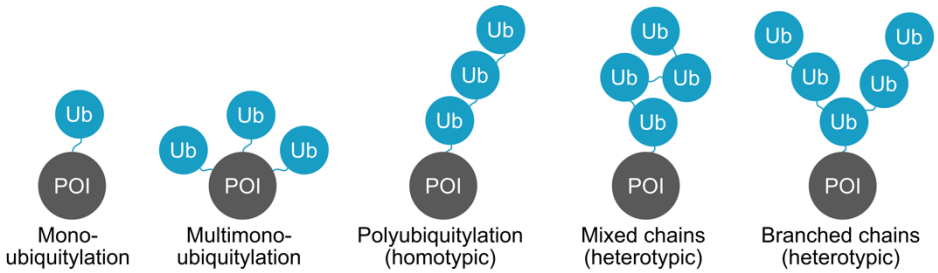


Figure 1.12 Overview of different Ub architectures.

1.2.2 Decoders of the code – Ubiquitin binding proteins

The various Ub architectures and topologies are translated into functional outcomes by reader proteins that interact with distinct patches on the surface of Ub through their Ub-binding domains (UBDs)[101]. Most of these interactions occur around an I44 centered patch but certain UBDs also recognize an acidic, hydrophilic patch around D58. In case of unconjugated Ub the free C-terminus can act as an additional recognition element (Figure 1.13 a)[111]. More than 20 different families of UBDs have been identified including UIMs (Ub-interacting motifs), UBAs (Ub-associated domains), and UBZs (Ub-binding zinc fingers) among many others[111]. UBDs interact non-covalently and usually with low affinity (10-500 µM) with Ub and can be found in many proteins. An exception is represented by the just recently identified UBD of the pathogen *Orientia tsutsugamushi* which shows an unprecedented dissociation constant of ~5 nM towards Ub[112]. The usually rather low affinity of a single UBD towards a single Ub is an important feature that allows avidity effects for proteins containing multiple UBDs and leads to fine-tuned selectivity towards distinct Ub topologies. NMR and X-ray studies showed that there is considerable structural variation depending on the linkage type of diubiquitins (diUbs), which represent the smallest type of Ub polymer. While K6-, K11- and K48-linked diUbs adopt a closed conformation, in which some interaction patches seem to be concealed, linear diUb as well as K63-linked diUbs show an extended structure[111]. These structural snapshots are, however, half-truths since polyubiquitin chains are inherently malleable, which allows UBDs to select specific conformations out of the available ensemble of polyUb chain conformations[113,114].

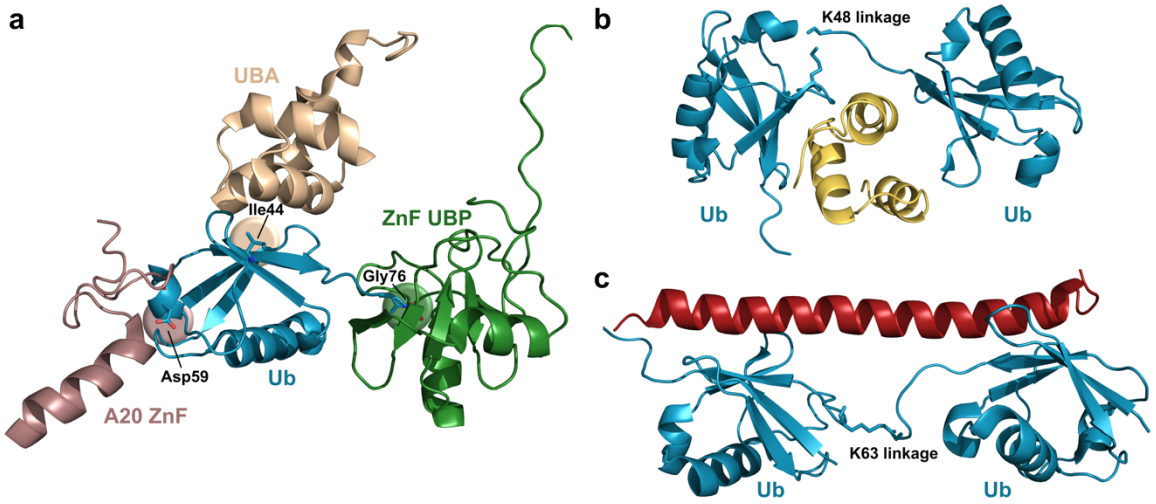


Figure 1.13 Overview of interactions between Ub and different UBDs. a) Alignment of different UBDs-Ub structures. UBDs recognize Ub via one of its different surface patches. The UBA domain of Dsk2p selectively interacts with the Ile44 patch (PDB 1WR1), while ZnF A20 domain recognizes the Asp59 patch (PDB 2L00). The C-terminus of unconjugated Ub can be recognized by the ZnF UBP domain (PDB 2G45). **b)** Structure of K48-linked diUb bound to the UBA domain of hHR23a (PDB 1ZO6). **c)** Structure of K63-linked diUb bound to the tUIM domain of Rap80 (PDB 2RR9). Partially adapted from[111].

The closed conformation of K48-linked diUb is for example recognized by the UBA2 domain of HR23a, which sandwiches in between the two Ub moieties (Figure 1.13 b)[115]. In contrast, the elongated conformation of K63-linked diUb is translated into functional outcomes by the interaction with the tandem UIMs (tUIMs) of Rap80, which adapts a helical conformation that precisely defines the distance and orientation of the Ub moieties (Figure 1.13c)[116].

These fine-tuned interactions between UBDs and Ub topologies serve as a platform for generating a plethora of different biological outcomes. Monoubiquitylation for example often regulates complex formation or allosteric regulation[117]. In the case of the homotrimeric DNA clamp PCNA (proliferating cell nuclear antigen), monoubiquitylation at K164 stimulates DNA translesion synthesis by recruiting a subset of DNA polymerases at DNA damage loci[118]. Alluringly, K63-linked polyubiquitination at the same lysine residue, facilitated by the E3-ligase Rad5, initiates a completely different mechanism of DNA repair, template switching. K63-linked chains, which represent the second most abundant type of Ub chain type in the cell, next to K48-linked chains[119], also play pivotal roles in processes such as NF-κB signaling[120] or DNA double strand break (DSB) repair mediated by Rap80 (see chapter 2 and 3)[121]. Next to the highly abundant K48 and K63 chains, less studied, atypical Ub chains including K6, K11, K27, K29, K33 and the M1-linked Ub chains, have in recent years been accounted for regulating processes such as angiogenesis, mitophagy, autophagy, epigenetics and many more[122]. For example, K27-linked polyubiquitylation of Y-box binding protein 1 (YB1), catalyzed by the HECT E3 ligase HACE1, has been shown to mediate the secretion of YB1[123]. Next to their fundamental role in proteasomal degradation[124], the well-studied K48-linked chains have recently gained increasing interest in the context of branched chains. The E3 ligase HUWE1 generates K48-linked branches on existing K63-chains in response to interleukin-1β. These branching events protect the K63 linkages from degradation by DUBs, while they still allow interaction with Tab2, and thereby NF-κB downstream signaling, which ultimately leads to an amplified downstream signal[125]. Besides their regulatory role in signaling events, branched chains also play an important role in host-pathogen interaction. This is exemplified by the Kaposi's sarcoma-associated herpesvirus, which encodes an E3 ligase (MIR2), that generates

K11/K63 branched chains on class I MHC receptors leading to internalization and thereby to immune evasion[126,127].

1.2.3 Deubiquitinases

Like most PTMs, ubiquitylation is a reversible process and tightly regulated by the action of DUBs, a class of proteases that hydrolyze the isopeptide bond between Ub and substrate proteins. Nearly a hundred DUB genes have hitherto been described in humans and are classified into seven structurally different superfamilies': six cysteine protease families (UCHs (ubiquitin C-terminal hydrolases); USPs (ubiquitin-specific proteases); OTUs (ovarian tumor proteases); Josephins, MINDYs (motif interacting with ubiquitin (MIU)-containing novel DUB family) and ZUP1 (Zinc Finger Containing Ubiquitin Peptidase 1) and a Zn-dependent metalloprotease family (JAMMs (JAB1/MPN/MOV34))[117].

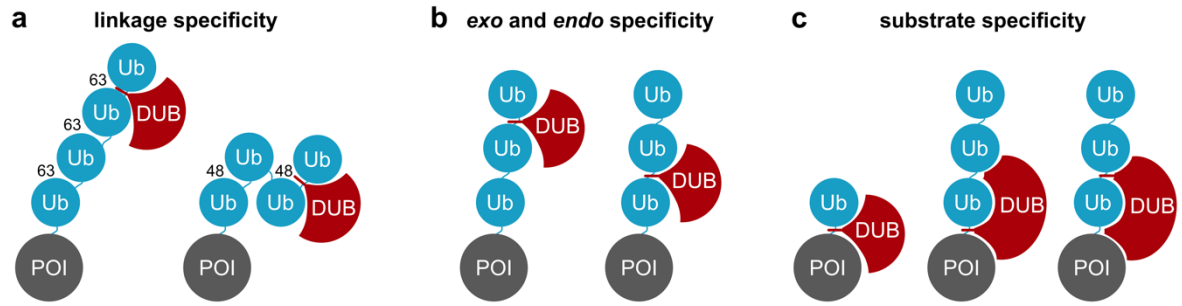


Figure 1.14 Overview of DUB specificities. a) Schematic representation of linkage specific DUBs. **b)** Schematic representation of DUBs with exo and endo specificity **c)** Schematic representation of substrate specific DUBs. Partially adapted from[128].

DUBs maintain the homeostasis of the ubiquitome by regulating the free ubiquitin pool and by fine tuning the ubiquitylation status of thousands of proteins. Due to the high complexity of ubiquitin chain architectures, DUBs display a wide variety of substrate specificities which allows them to distinguish between different Ub topologies[117,128]. A couple of DUBs show specificity towards the type of linkage between two Ubs (Figure 1.14 a), for example the DUB Cezanne which hydrolyzes preferentially K11-linkages[129]. Other DUBs, on the contrary, favor a certain chain length, for instance OTUD2 which prefers long Ub chains as substrate[130]. Besides preferences in linkage type and chain lengths, DUBs can be grouped into endo DUBs, which cleave within a Ub chain, and exo DUBs which preferentially cleave distal Ub subunits one after another (Figure 1.14 b)[131]. In addition to their specificities towards Ub architectures some DUBs also recognize the ubiquitylated substrate protein and remove either monoUb or whole polyUb topologies *en bloc* (Figure 1.14 c).

As already mentioned above, besides their antagonistic role in hydrolyzing Ub-substrate conjugates, DUBs are also responsible for processing Ub from one of its four expressed inactive precursors (Figure 1.11)[132]. Due to their central role in the Ub system, expression levels of DUBs, their specificity towards different Ub topologies and their distinct subcellular localization are key to cellular propagation, DNA-repair and innate immune signaling[133]. Therefore, alterations in DUB activity often leads to pathologies, which is the reason why DUBs are gaining increased attention in drug discovery[134].

1.2.4 Ubiquitin-like proteins

In analogy to the histone code, the entirety of the complex Ub regulatory system has been coined, the 'ubiquitin code'[135]. With E1/E2/E3 enzymes acting as writers, proteins bearing UBDs as readers and DUBs taking the role of erasers, Ub itself can be regarded as its ink. And as ink comes in different colors, so does Ub. The characteristic ß-grasp fold, also called Ub fold, and the flexible unstructured C-terminus displaying a terminal glycine residue are structural features that are not only found in Ub but also in so-called Ub-like proteins (Ubls, Figure 1.15 a). Most members of the Ubl family that comprises approximately 15 small proteins including SUMO (small ubiquitin-related modifier) and Nedd8 (neural-precursor-cell-expressed developmentally down-regulated 8)) are conjugated to substrate proteins, recognized and cleaved off in a similar manner as described for Ub[136].

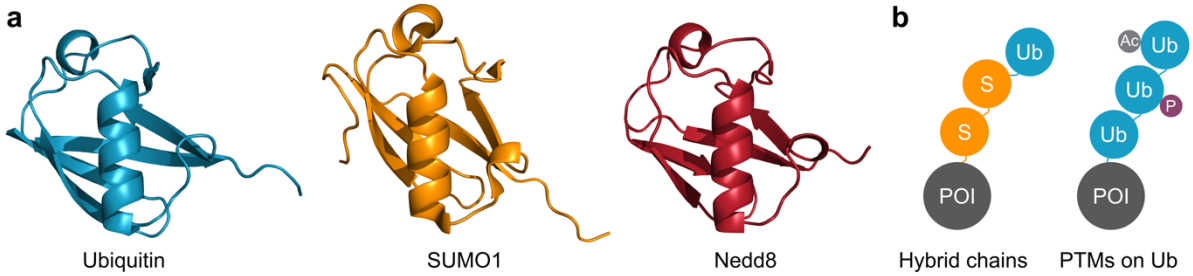


Figure 1.15 Structural representation of Ubls and complex Ub/Ubl topologies. a) Structures of Ubiquitin (PDB 1UBQ), SUMO1 (1Y8R) and Nedd8 (1NDD). **b)** Schematic representation of the next layer of complexity. Left: Ub/SUMO hybrid chain attached to a protein of interest (POI). Right: Modification of Ub with small PTMs such as acetylation and phosphorylation further increases the complexity of the Ub-code.

In contrast to Ub, SUMO has at least four paralogues (SUMO1-4) in humans[136] and represents after Ub, which has far more than 10,000 reported conjugations sites, the second most conjugated Ubl with ~300 sites for SUMO1 and ~150 sites for SUMO2[137]. Lysine residues that are typical SUMO conjugation sites are usually found within the distinct ψKXE (where ψ is a large hydrophobic residue and X is any residue) consensus motif[138]. Similarly to UBDs, so-called SUMO-interacting motif's (SIMs) interact non-covalently with SUMO to translate the conjugation into a functional outcome. SIM's are usually disordered, but upon binding in a distinct grove between the central α_1-helix and the β_2-strand of SUMO, the unstructured SIM adopts a beta strand fold that complements with the beta sheet of the β_2-strand[136]. While Ubls extend the concept of the ubiquitin code as they encode novel functional outcomes they can additionally lead to synergistic or antagonistic crosstalk with Ub. The interplay between different Ubls is intriguingly exemplified by the existence of so-called hybrid chains, where different Ubls are mixed in a polymeric Ubl chain. Alongside the recently reported Ub-Nedd8 hybrid chains, which are thought to play a role in the proteotoxic stress response[139], the discovery of Ub-SUMO hybrid chains and their role in DNA damage response is of great interest (Figure 1.15 b)[140].

1.2.5 The next layer of complexity

Due to its unique role as a protein-based PTM, Ub can itself be a target for other PTMs, creating a very complex new meta-level of modification patterns (Figure 1.15 b). As already discussed

above, all lysine residues of Ub can serve as a platform for further ubiquitylation events. In combination with the possibility of mixed and branched chains this gives rise to a plethora of different Ubl-topologies with uncountable different functional outcomes. But this still does not account for the full regulatory potential encoded in Ub. Beyond post-translational modification of Ub via covalent attachment of further Ubls, Ub is also target for PTMs, such as acetylation[141], phosphorylation[142], deamidation[143], ADP-ribosylation[144] and arginine phosphoribosylation[145], all of which have been found to play an important role in Ub signaling.
The most prominently studied example of a PTM on Ub might be the phosphorylation of serine 65 and its role in Parkinson's disease. Groundbreaking studies revealed that Ub is a substrate of the kinase PINK1[146], that phosphorylates serine 65 of Ub, which results in the activation of the E3-ligase parkin[147], leading to downstream effects that finally result in mitophagy[148]. Next to phosphorylation, all seven lysine residues of Ub have been shown to be targets for acetylation[149], which can for example lead to interference with Ub chain formation[112,] or may influence the interaction with E3-ligases[150]. Similarly, to Ub-acetylation, also SUMO-acetylation has been reported. Initial studies showed that acetylation in the interaction interface between SUMO and SIMs impairs binding of SUMO-binding proteins[151] and that SUMO acetylation can interfere with SUMO chain formation[152]. An intriguing example of a pathogen interfering with the Ub system is demonstrated by *Legionella pneumophila*, whose effector protein SdeA mediates NAD^+-dependent phosphoribosylation of arginine 42 of Ub[145]. Once phosphoribosylated at arginine 42, Ub-activation by E1/E2 enzymes is impaired, which depletes it from the cellular Ub pool. Furthermore, SdeA is able to catalyze non-canonical serine ubiquitylation by conjugating phosphoribosylated Ub to the hydroxy group of serine via a phosphodiester bond. Similarly, a recent report showed that the attachment of an ADPr group to threonine 66 of Ub, facilitated by the effector CteC from *Chromobacterium violaceum*, disrupts Ub signaling in the host[153]. Taken together, post translational modification of Ub and SUMO serves as an additional regulatory layer by influencing polymerization status as well as fine tuning downstream protein-Ubl interactions.

1.2.6 Deciphering the Ub code

Deciphering the Ub code in its entirety represents a major challenge and will most likely be a multidisciplinary adventure holding many unexpected surprises on its way as exemplified recently by the discovery of a human E3 ligase for non-canonical ubiquitylation at serine/threonine residues[107] and by the clever mechanism pathogens developed to hijack the host cell ubiquitylation system in order to impede host cells defense and to increase their virulence[12]. In order to gain further insight how Ub and Ubls orchestrate fundamental cellular processes and to understand how dysfunctions in ubiquitylation/deubiquitylation pathways are linked to pathologies, we need a molecular and mechanistic understanding how ubiquitylation affects target proteins and how different Ub/Ubl topologies are generated and decoded. As the specific combinations of enzymes that are used to ubiquitylate and deubiquitylate many proteins are however not known, understanding roles of ubiquitylation and deciphering specificities of writer, reader and eraser proteins represents a formidable challenge. The inherent complexity of the Ub code has attracted chemist and chemical biologist for almost two decades and has spurred creative solutions for the generation of defined Ub-protein of interest (POI) conjugates. Advances in chemical protein synthesis and protein semisynthesis paved the

way for studying structural and mechanistic features of Ub-POI conjugates, however, as these approaches rely in large parts on harsh chemical deprotection protocols they are mostly limited to *in vitro* ubiquitylation of simple, refoldable target proteins [154]. Moreover, in many cases, where we are challenged to create defined Ub-POI conjugates, the endogenous ubiquitylation machinery is either not fully understood/identified or inactive/unspecific under in *in vitro* conditions. Therefore, novel methodologies are needed, which are on the one hand capable of creating Ub-POI conjugates of increasing complexity including complex Ub topologies attached to non-refoldable multidomain substrate proteins and allow on the other hand the creation and the study of Ub-POI conjugates within their native environment; the inside of a eukaryotic cell. A powerful technology that may meet both of these challenges is genetic code expansion (GCE). GCE allows the installment of protein modifications by site-specifically incorporating artificial designer amino acids in living cells and thereby represents a promising tool to further dissect and study Ub signaling.

1.3 Genetic code expansion

Besides some rare exceptions such as selenocysteine[155] and pyrolysine[156], all proteins are built from the twenty natural amino acids. As extensively discussed above, further diversifications of the chemical space of the proteome can be achieved through a myriad of PTMs. However, to specifically manipulate biological processes, scientist have developed approaches to endow proteins with unique chemical functionalities by introducing unnatural amino acids (UAA) into proteins[157]. Early efforts to incorporate UAAs into proteins are solid-phase peptide synthesis (SPPS) [158], native chemical ligation (NCL) [159] and *in vitro* translation[160,161]. While SPPS allows the introduction of a large number of modifications it is limited to small, refoldable proteins[162]. In contrast, NCL-based approaches facilitate the modification of larger proteins, but modification sites must be located close to protein N- or C-termini (see chapter 1.4). Using nonsense suppressor tRNAs charged with an UAA, it is possible to modify proteins in a site-specific manner using *in vitro* translation systems. However, these approaches suffer from low yields and complex protocols regarding the generation of aminoacyl-tRNAs[163]. Due to these limitations, scientists developed approaches to harness the potential of the translational machinery of a host organism to introduce UAAs into proteins in living cells by expanding the genetic code. Such technologies include residue- and site-specific incorporation of UAAs during protein translation[164-168].

1.3.1 Residue-specific incorporation of UAAs

Already in the 1950's it has been demonstrated that, besides methionine, the bacterial methionine tRNA synthetase (MetRS) accepts the close structural analogue selenomethionine. By substituting methionine with selenomethionine in the growth media it was thereby possible to produce selenomethionine-bearing proteins, which later aided the structural elucidation using X-ray crystallography[169,170]. This concept was expanded over the years resulting in a widely used approach dubbed selective pressure incorporation, which allows global, proteome-wide decoration of proteins with specifically designed UAAs (Figure 1.16).

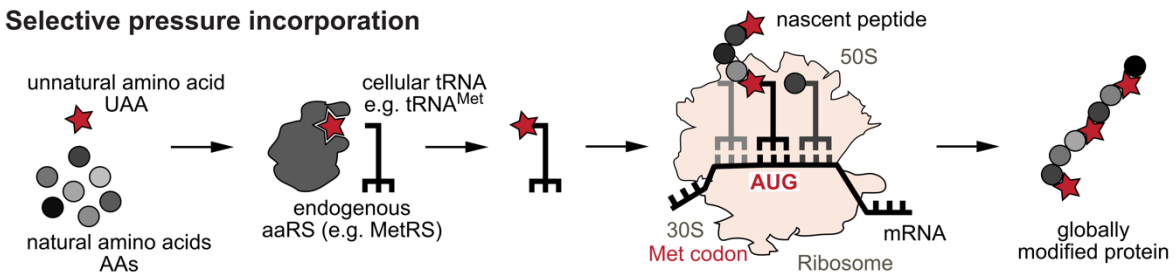


Figure 1.16 Selective pressure incorporation. Schematic representation of selective pressure incorporation. Under auxotrophic conditions an isostructural analogue of a natural amino acid (AA) is recognized by the endogenous aaRS and charged on its cognate canonical tRNA leading to residue specific incorporation by the ribosome resulting in a globally modified protein. UAA is represented by a red star, natural AA is represented by grey dots.

This method exploits the promiscuity of wild type (wt) aminoacyl-tRNA-synthetases (aaRS) to incorporate structural analogues of the endogenous substrates. In order to achieve quantitative replacement, auxotrophic strains, that lack the endogenous machinery for biosynthesis of the natural amino acid that is to be substituted, are used for protein production and the natural

amino acid is depleted from the growth media. SPI allows the residue specific installation of functional groups e.g. for selective protein labelling. The most widely used UAAs for this purpose are azidohomoalanine (Aha) and homopropargylglycine (Hpg), which can be applied for click-chemistry-based labelling approaches (see Chapter 1.4)[171].

Since the structural range of amino acid derivatives that are efficiently incorporated by wt aaRSs is limited, efforts to incorporate structurally more diverse analogues led to approaches altering the biosynthetic machinery. This was achieved by either mutating the editing domain of the aaRS[172] or by altering the amino acid binding pocket[173,174], which for example led to the incorporation of the lysine derivative azidonorleucine (Anl) by MetRS[175]. Even though selective pressure incorporation has proven its utility in applications ranging from material sciences[176] to whole proteome labelling approaches like bioorthogonal non-canonical amino acid tagging[177] and fluorescent non-canonical amino acid tagging [178], it suffers from structural limitations concerning the UAA and does not allow precise site-specific protein manipulation.

1.3.2 Site-specific incorporation of UAAs – Amber suppression

The second approach that is based on engineering the translation machinery is the site-specific incorporation of UAAs. This method relies on an orthogonal translation system, which consists of an orthogonal aaRS that specifically charges an UAA onto its cognate tRNA to ultimately promote the incorporation of the UAA into a protein in response to a blanc codon on the mRNA (Figure 1.17)[179]. Usually, the blank codon of choice is the amber stop codon, (thus the approach is often called amber suppression) as it is the least used in *E. coli* (9 %) and rarely terminates essential genes[180,181]. Orthogonality is central for an orthogonal translation system to function with high fidelity and it is of utmost importance that the aaRS exclusively charges the UAA onto the corresponding tRNA without interfering with the endogenous tRNAs, aaRSs and natural amino acids. In order to do so, most aaRS/tRNA pairs are derived from phylogenetically distant organisms, which helps to minimize the cross reactivity with the endogenous translational machinery system. Many of the orthogonal translation systems origin from archaea such as the *Methanocaldococcus jannaschi* tyrosyl-tRNA synthetase (*Mj*TyrRS)/tRNA$_{CUA}$ pair and the pyrrolysyl-tRNA synthetase (PylRS)/tRNA$_{CUA}$ pairs from *Methanosarcina mazei* (*Mm*) and *Methanosarcina barkeri* (*Mb*)[166,168,182]. While the *Mj*TyrR/tRNA pair is only orthogonal in eubacteria, the PylRS/tRNA pairs from *M. mazei* and *M. barkeri* are orthogonal in both prokaryotic and eukaryotic systems.

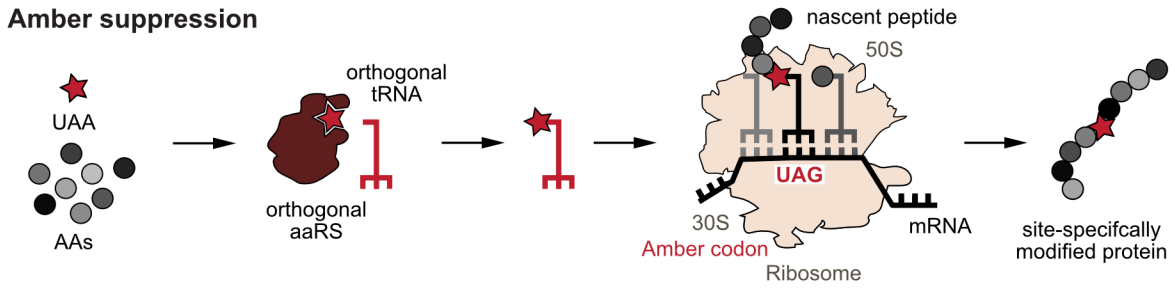


Figure 1.17 Schematic representation of amber suppression. Amber suppression makes use of an orthogonal aaRS that exclusively charges the UAA onto an orthogonal tRNA. In response to an amber stop codon introduced into the gene of interest the UAA is site-specifically incorporated by the ribosome. UAA is represented by a red star, natural AAs are represented by grey dots.

For the site-specific incorporation of designer UAAs, the active site of the designated aaRS needs to be engineered in order to selectively recognize the UAA. This is usually achieved via directed evolution - a two-step selection approach, which includes iterative rounds of positive and negative selections[166,167,183-185]. Therefore, a library with typically 10^7 - 10^8 members is created by randomizing active site residues using site saturation mutagenesis. In the first positive selection step the aaRS library is transformed in *E. coli* cells harboring a plasmid with a chloramphenicol acetyltransferase gene disrupted by an amber codon. The resulting transformants are then cultivated in the presence of the UAA and chloramphenicol in order to select for cells that bear an aaRS variant which incorporates either the UAA or and natural AA in response to the amber codon. To eliminate aaRS variants that accept a natural AA, the isolated aaRS variants from the positive selection are transformed into *E. coli* cells bearing a plasmid encoding the toxic ribonuclease barnase, which is disrupted by two amber codons. By omitting the UAA from the next selection step, aaRS that accept natural AA are removed from the aaRS pool. After iterative rounds of positive and negative selections, the remaining aaRS are isolated and transformed into cells bearing a GFP disrupted by an amber codon at position 150. Quantification of the amber suppression efficiency of the aaRS variants is performed by cultivating single clones in presence and absence of the UAA followed by fluorescence measurement (Figure 1.18).

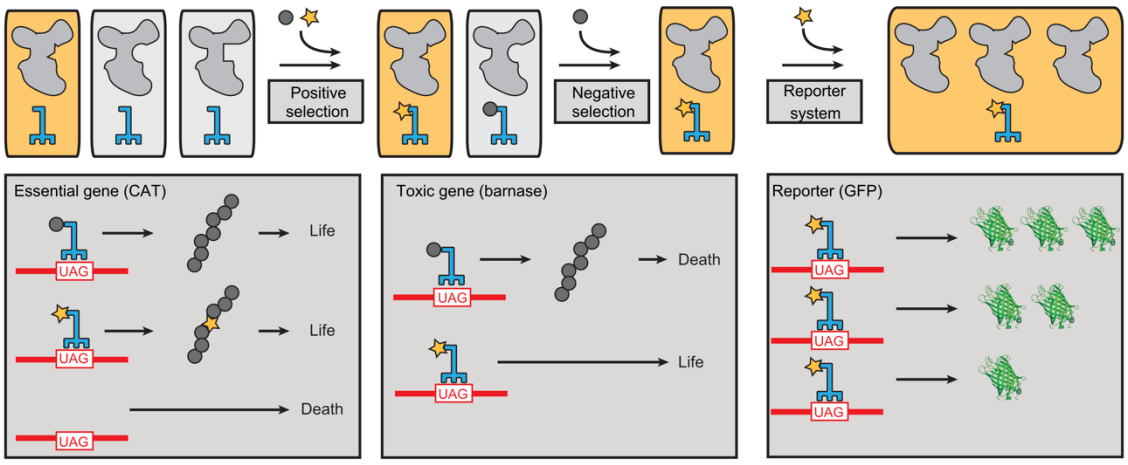


Figure 1.18 Schematic representation of directed evolution of aaRS. Positive selection with an essential gene disrupted by an amber codon in presence of the UAA results in aaRS variants that charge the UAA or a natural AA. Negative selection facilitated by the disruption of a toxic gene in the absence of the UAA leads to the removal of aaRS variants that accept a natural AA. Finally, a reporter system based on GFP fluorescence is used to quantify amber suppression efficiencies of derived aaRS variants.

Engineering of orthogonal PylRS through directed evolution has hitherto facilitated the site-specific incorporation of more than 250 UAAs into proteins in prokaryotic[166,167] and/or eukaryotic cells[186-190], including multicellular organisms such as plants[191], *Caenorhabditis elegans*[192], *Drosophila melanogaster*[193] and embryonic mouse brain[194,195].

The ability to site-specifically introduce novel functionalities into POIs enabled chemical biologists to manipulate and investigate diverse biological processes in a hitherto unprecedented manner. Amber suppression has proven to be an excellent tool to site-specifically label proteins *in vitro* and *in vivo* with diverse biochemical and biophysical reporters by incorporating UAAs carrying bioorthogonal handles (e.g. azides, alkenes, alkynes and tetrazines amongst others) which undergo rapid and selective reactions with externally

added chemical probes (Figure 1.19 a)[168]. Furthermore, the incorporation of natural amino acids bearing photoremovable groups allowed the spatiotemporal control of protein localization and helped to dissect signaling networks (Figure 1.19 b)[189,196,197]. Another clever use of light as an external trigger is represented by photocrosslinking approaches to identify and map weak and transient protein-protein interactions (PPIs)[198,199]. The studies of PPIs have also been expanded towards UAAs bearing electrophilic crosslinking moieties which are inert under physiological conditions but whose reactivity can be triggered by proximal nucleophilic amino acid residues upon complex formation (Figure 1.19 c)[199,200]. These UAAs have been proven useful for stabilizing low-affinity protein complexes in their native environment[201,202].

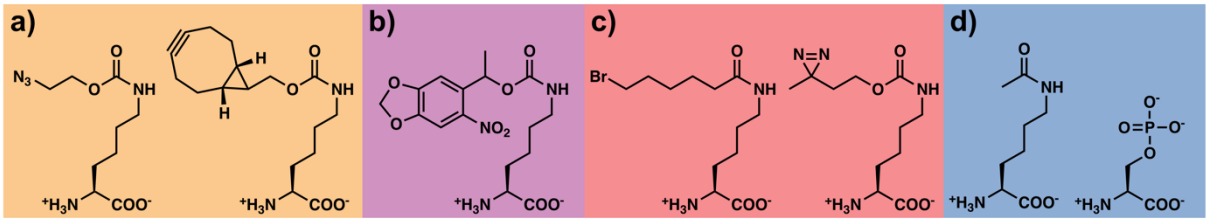


Figure 1.19 A selection of genetically encodable unnatural amino acids. a) UAAs carrying bioorthogonal handles such as azides and strained alkynes. **b)** Photocaged lysine. **c)** UAAs for chemical- and photo-crosslinking. **d)** UAAs bearing PTMs.

Regarding the study of PTMs, the striking advantage of amber suppression relies in its ability to install these modifications at user-defined residues without having to know the identity of the respective endogenous writer proteins, which are often elusive or inactive/unspecific *in vitro* (Figure 1.19 d). Over the last decade, numerous amber suppression-based approaches to install scPTMs or mimics have been developed, which now allow the generation of proteins modified with lysine acylations such as acetyl-[203], crotonyl-[204], propionyl-[205], butyryl-[205], hydroxyisobutyrl[206]-lysine and lysine-alkylation e.g. monomethyl-lysine[207-209]. Furthermore, amber suppression gives access to protein phosphorylation e.g. phospho-serine[210]/threonine[211]/tyrosine[212] and more exotic scPTMs such as nitration[213] and sulfation[214] of tyrosine. Amber suppression yields homogenously modified proteins and is in contrast to synthetic/semisynthetic approaches not limited to modification sites at N/C-termini of the substrate protein nor its size. Additionally, amber suppression is applicable to complex, non-refoldable and enables studies of the impact of PTMs in living cells, a feature that can hardly be addressed by other methods.

While small chemical scPTMs such as acetylation, methylation or phosphorylation can be directly encoded since UAAs resembling these modifications are suitable substrates for orthogonal aaRSs, scPTMs that include the attachment of a whole protein cannot be directly installed. In order to genetically encode protein-based scPTMs such as ubiquitylation, an UAA (most typically a lysine derivative) bearing a chemoselective reactive moiety for the conjugation of a modified Ub is incorporated. Post-translational site-specific conjugation of Ub is then mediated via chemical/chemoenzymatic procedures.

The following chapter will give a brief overview of the available genetic code expansion tools for generating Ub-POI conjugates in respect to their advantages and disadvantages. The methodologies will be categorized into: (1) methods that generate Ub-POI conjugates with artificial linkages and (2) methods resulting in native isopeptide-linked conjugates.

1.4 Genetic code expansion approaches to study the ubiquitin system

Over the past 15 years, chemical biologists have developed a toolbox comprised of methods to address the urgent need to site-specifically ubiquitylate proteins in order to study the Ub code. These methods can be divided into chemical approaches[215-217], including thiol-ene coupling[218], disulphide exchange[219], thioether ligation[220] and click chemistry[221], and approaches based on GCE, which can be categorized into methods resulting in conjugates linked via a native isopeptide bond and strategies that generate conjugates linked via artificial linkages. Approaches yielding natively linked Ub-POI conjugates are rather scarce and only encompass strategies based on native chemical ligation[222,223] as well as a method called "genetically encoded orthogonal protection and activated ligation" (GOPAL)[224]. On the other hand, there is a variety of approaches yielding Ub-POI conjugates with artificial linkages which are based on bioorthogonal reactions such as the thiazolidine ligation[225], oxime ligation[226] and click chemistry[227,228].

1.4.1 Approaches resulting in artificially linked Ub-POI conjugates

The following chapter will deal with the generation of Ub-POI conjugates displaying artificial linkages in place of the isopeptide bond which are produced by effectively employing bioorthogonal reactions. By definition, all reactions that are considered to be bioorthogonal must fulfill several essential requirements including high chemoselectivity between the reaction partners under physiological conditions (aqueous environment, ambient temperature and neutral pH) yielding close to quantitative yields with only inoffensive byproducts[229]. Importantly, also high robustness and fast reaction kinetics as well as metabolic, kinetic and thermodynamic stability of reacting functionalities are required for successful bioorthogonal reactions[230]. While many reactions have hitherto been (re)discovered and repurposed to fulfill aforementioned requirements, further considerations had to be made upon choosing a suitable bioorthogonal reaction for generating Ub-POI conjugates. The reaction of choice should ideally be fast, since target proteins do not withstand long incubation times at elevated temperatures and the resulting product should mimic the isopeptide bond in the best way possible. Alongside these criteria, researchers exploited the use of click reactions between azides and terminal alkynes[221,227], the oxime ligation between aldehydes and aminoxy moieties[226] as well as the thiazolidine ligation[225]. Another important challenge that researchers had to overcome included the installation of two reactive entities into Ub and the target protein which should be, in the best case, site-specific and applicable to all kind of proteins. The advantage of the resulting artificially linked Ub-POI conjugates lies within their resistance towards DUBs, opening up the opportunity to perform studies in complex environments such as cell lysates.

1.4.1.1 Click chemistry-mediated ubiquitylation

Over the last two decades the "click" reaction has proven to be one of the most powerful and versatile reactions in various chemical and biological settings[231]. The copper(I)-catalyzed 1,3-dipolar cycloaddition between azides and terminal alkynes (CuAAC) proceeds selectively and with fast kinetics in complex environments, affording selectively 1,4-disubstituted 1,2,3-triazoles as products. As mentioned above, genetic code expansion approaches have allowed

the site-specific installation of both azide and terminal alkyne moieties into recombinant proteins[232,233]. In order to apply click chemistry for the generation of defined Ub-POI-conjugates, two main strategies have evolved within the last 10 years (Figure 1.20). The first approach is based on a combination of selective pressure incorporation and amber suppression to introduce the azide and alkyne functionalities into Ub and POI, respectively[227]. Therefore, selective pressure incorporation employing the MetRS was used to substitute the terminal glycine of Ub with the azide bearing methionine analogue Aha[171], while the complementary terminal alkyne functionality was site-specifically incorporated at a chosen acceptor lysine site within the target protein in the form of Nε-(propargyloxycarbonyl)-L-lysine (PlK, Figure 1.20 a) using amber suppression and the wild type (wt) *Mb*PylRS/tRNA$_{CUA}$ pair. Subsequent click reaction allowed the production of all seven triazole-linked diUbs. Autoubiquitylation assays confirmed the functional integrity of the conjugates and incubation of triazole-linked diUbs in the presence of DUBs proved their resistance towards hydrolysis[234]. Apart from generating differently linked diUbs, the approach has been used for site-specifically ubiquitylating PCNA at lysine 164 (K164). PCNA is a ring-shaped homotrimeric protein that serves as a DNA clamp during replication and enhances the processivity of DNA polymerase delta (Pol δ)[235]. DNA lesions in the template strand lead to stalling of the replication fork, which triggers monoubiquitylation of PCNA-K164 via a specific E2/E3 combination. PCNA-Ub in turn leads to the recruitment of specialized translesion synthesis DNA polymerases (e.g. Pol η) to the DNA damage site in order to traverse the DNA damage. Triazole-linked PCNA-Ub showed stimulation of DNA synthesis by Pol δ and increased affinity towards Pol η, mimicking the behaviour of natively ubiquitylated PCNA and thereby confirming the suitability of triazole-linked POI-Ub conjugates for biochemical studies.[236]

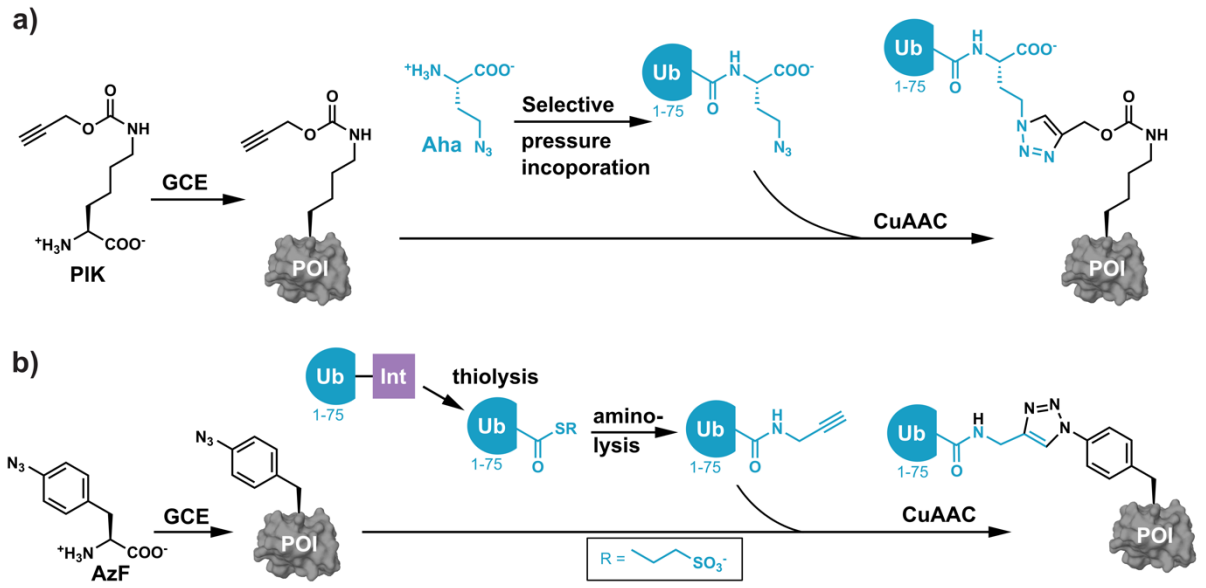


Figure 1.20 Click-chemistry based approaches for the generation of Ub-POI conjugates. a) Schematic representation of an approach combining GCE and SPI. **b)** Schematic representation of an approach combining GCE and intein (Int) technology.

By simultaneously incorporating Aha and PlK within the same Ub monomer, linkage-specific, triazole-linked Ub chains were accessible, as exemplified by formation of unanchored polymeric K11-, K27- and K29-linked Ub chains, as well as generation of poly-ubiquitylated pol β[234,237]. Furthermore, these DUB-resistant unanchored chains were used to study linkage-specific effects on meiotic progression in *Xenopus laevis* egg extract[234]. Investigation of the

Ca^{2+}-dependent cyclin B degradation showed that K11-linked chains inhibit degradation of cyclin B, most likely by proteasome binding. In contrast, K27- and K29-linked chains showed no influence on cyclin B levels.

The second click chemistry-based strategy to produce Ub- and Ubl-POI-conjugates combines intein technology and amber suppression. As already mentioned above, inteins are auto-processing protein domains that catalyze a chemical process known as protein splicing. During this process the intein excises itself from a larger polypeptide via a mechanism that involves a series of acyl transfer and transthioesterification reactions, leading to the ligation of the flanking extein peptides (N- and C-extein) via a new peptide bond[228]. Mutation of essential residues of the intein splicing cascade stalls the process after the initial thioester formation step between the C-terminus of the N-extein and the N-terminal cysteine of the intein. Transthioesterification using a water-soluble small molecule with a nucleophilic thiol moiety results in a C-terminal thioester of the N-extein[238]. As the process of protein splicing imposes no sequence restriction to the N- or C-extein, intein technology has been harnessed for generating C-terminal thioesters of virtually any protein by simply fusing the POI N-terminally to a mutated intein. The resulting C-terminal thioester can then undergo aminolysis using an appropriate small molecule with a nucleophilic amine moiety to install functional groups at the C-termini of proteins[37]. This was used to install an alkyne functionality at the C-terminus of a Ub-variant lacking its C-terminal G76 residue (UbΔG) by incubating an UbΔG-intein thioester with propargylamine. The complementary azide functionality was introduced into the target protein by amber suppression using an evolved *Mj*TyrRS/$tRNA_{CUA}$ pair that enables the site-specific incorporation of *p*-azido-L-phenylalanine (AzF, Figure 1.20 b)[233]. This approach was used for SUMOylation of the SUMO E2 conjugating enzyme Ubc9 (UBE2I) by incubating AzF-bearing Ubc9 with a C-terminally alkyne-modified SUMO2(ΔG) variant. Subsequent biochemical characterization in SUMOylation assays with Sp100 and RanGAP1 as substrates revealed an altered substrate preference of Ubc9-SUMO2 compared to unmodified Ubc9 in accordance with biochemical data on endogenously SUMOylated Ubc9. Furthermore, this approach facilitated the generation of all seven triazole-linked diUbs, which were shown to be refractory to DUB hydrolysis, but displayed unaltered binding properties towards the UBA domain of Mud1[239]. Structural and dynamic similarity of the triazole linkage within K48-linked diUb to its native isopeptide bond counterpart were additionally proven by theoretical calculations based on hybrid quantum mechanical and molecular mechanical (QM/MM) methods.[240] A recent follow-up study adopted this methodology for the generation and study of monoubiquitylated PEX5, which plays an essential role in peroxisomal import [241].

Click chemistry-based approaches strategies result in non-hydrolyzable conjugates and several triazole-linked diUbs were shown either experimentally or computationally to closely resemble their native isopeptide-linked counterparts, presenting therefore valuable tools to study Ub-POI conjugates in cell lysates. A drawback of the approach based on Aha incorporation at the C-terminus of Ub via selective pressure incorporation is however represented by its limitation to Ubls that do not contain methionine residues beside the one encoded by the start codon. Consequently, the approach is not applicable to Ubls such as SUMO1-4, Nedd8 and ISG15. Moreover, all of the described methods are dependent on cytotoxic Cu(I), excluding their use for *in vivo* settings, and *in vitro* applications on purified proteins may suffer from Cu(I)-induced protein precipitation and oxidation due to residual metal traces[242].

1.4.1.2 Ubiquitylation via oxime ligation

Carbonyl condensation reactions, such as the hydrazone, imine, and oxime bond-forming reactions are widely used for bioconjugation applications on purified proteins.[243]. One particular advantage of the oxime ligation between aldehyde and hydroxylamine functionalities consists in the higher stability of the resulting oxime-bond towards hydrolysis compared to hydrazones and imines[244]. In the context of studying Ubl-POI-conjugates, the oxime ligation offers additional benefits as the resulting linkage is isosteric to the native isopeptide bond but confers stability towards DUBs (Figure 1.21). In first attempts, a Ub variant, functionalized C-terminally with an aldehyde moiety was ligated to aminoxy-modified peptides synthesized by SPPS to get access to Ub-peptide-conjugates[245].

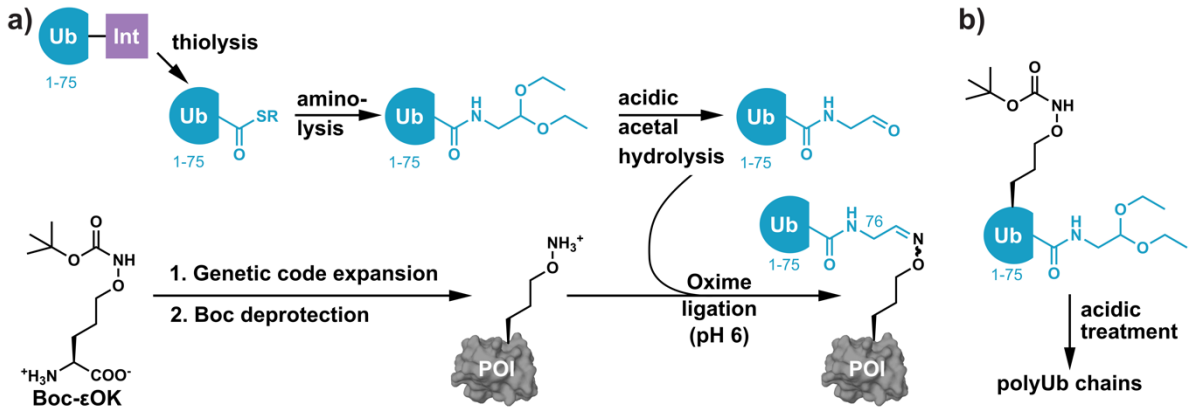


Figure 1.21 An oxime ligation-based approach for the generation of Ub-POI conjugates. a) Schematic overview of oxime ligation-based ubiquitylation of POIs. **b)** A bifunctional Ub building block displaying both, the Boc-protected oxime bearing UAA and the acetal group. Acidic treatment leads to unmasking of the oxime and the aldehyde moiety and therefor to the generation of polyUb chains.

In order to expand this strategy towards proteins, a UAA bearing an aminoxy moiety within the lysine side chain by substituting the ε-methylene group with an oxygen was designed and synthesized[226]. As specific recognition element for selective tRNA charging by PylRS, a *tert*-butoxycarbonyl (Boc) protection group was installed at the *N*ε amino group, creating Boc-εOK (Figure 1.21 a). Removal of the Boc protection group under acidic conditions after site-specific incorporation of Boc-εOK and protein purification revealed the reactive aminoxy moiety. The corresponding counterpart for oxime ligation, a Ubl containing a C-terminal aldehyde, was generated through intein chemistry. Aminolysis of the Ub-thioester with an acetal-bearing amine moiety, yielded an Ub-acetal, which was converted *in situ* to the corresponding aldehyde under acidic conditions[246]. Combining these technologies, it was possible to generate oxime-linked K6- and K48-diUbs as well as a SUMO(K11)-Ub conjugate bearing an oxime linkage. Structural elucidation of the oxime-linked K6-diUb revealed the trans regioisomer of the oxime-linkage as the prevalent species, mimicking accurately the native isopeptide counterpart. The isostery to the native isopeptide bond combined with the hydrolytic stability of the oxime linkage towards DUBs makes oxime-linked Ub-conjugates promising DUB inhibitors. IC_{50}-measurements revealed oxime-linked K6-diUb, SUMO(K11)-Ub as well as K48-diUb as potent inhibitors for the promiscuous DUB UCHL3. Additionally, oxime-linked K48-diUb showed nanomolar binding affinities towards USP2, in agreement with affinities determined for natively linked K48-diUb. Inspired by click-chemistry-based access

to triazole-linked poly-Ub chains[226], a bifunctional Ub bearing the Boc-protected aminoxy moiety at K6 and the acetal protected aldehyde moiety at the C-terminus was generated (Figure 1.21 b). Upon acidic deprotection with trifluoroacetic acid (TFA), oxime-linked K6 poly-Ub chains were generated that proved recalcitrant towards DUB hydrolysis as expected.

From a structural point of view, the oxime linkage may represent the "minimal perturbation motif" to confer DUB resistance to conjugates while at the same time showing high isostery to the native isopeptide linkage. Since the Ub-aldehyde can be generated via intein chemistry, it should easily be expandable towards other Ubls. Compared to other approaches (e.g. click chemistry), the oxime ligation may be more challenging as it requires complex multistep synthesis of the UAA and relies so far on acidic deprotection of incorporated Boc-εOK. This means that substrate proteins have to be refolded, limiting the approach to simple and easily refoldable proteins. Replacement of the Boc-protection group with photo-caging groups may prove advantageous and site-specific incorporation of a photo-protected aminoxy-bearing UAA has already been shown but has so far not been exploited for generating Ub-POI-conjugates.

1.4.2 Approaches resulting in natively linked Ub-POI conjugates

Next to the approaches yielding artificially linked Ub-POI conjugates stand approaches that allow the generation of native, isopeptide-linked conjugates. Since bioorthogonal reactions that result in native amide linkages such as the traceless Staudinger ligation[247] and the KAT ligation[248] have so far not been exploited for the generation of defined Ub-POI conjugates (e.g. due to the difficulty to install the reactive moieties site-specifically into target proteins), researches have, therefore, utilized nature's functional repertoire by adapting the mechanism of inteins to create new amide bonds. Additionally, development of complex protection/deprotection protocols gives access to proteins bearing single reactive lysine moieties that can be used for controlled chemical formation of isopeptide bonds. The resulting Ub-POI conjugates are cleavable by DUBs, due to the presence of the native isopeptide bond, which therefore allows to study the specificities of different DUBs towards distinct substrates.

1.4.2.1 Genetically encoded orthogonal protection and activated ligation (GOPAL)

To build POI-Ub conjugates with native isopeptide linkages, an elegant protection/deprotection strategy, dubbed GOPAL (genetically encoded orthogonal protection and activated ligation) was developed[224]. GOPAL uses a combination of genetic code expansion, intein chemistry and chemoselective ligation to build isopeptide-linked Ub-POI conjugates (Figure 1.22). What makes GOPAL strikingly different from approaches described so far is its independence on installing a chemical handle for biorthogonal chemistry (e.g. alkyne, aminoxy) into the protein. Instead, GOPAL uses the intrinsic reactivity of the *N*ε-amino group of lysine. To this end, an elegant orthogonal protection/deprotection strategy was developed based upon site-specific incorporation of *N*ε-(*tert*-butoxycarbonyl)-L-lysine (BocK) into target proteins using the wt *Mb*PylRS/$tRNA_{CUA}$ pair. The Boc group serves as a genetically encoded acid-labile protection group which is installed at the lysine residue of interest while the remaining amine moieties of the target protein (all lysines and the N-terminus) are orthogonally protected with benzyloxycarbonyl (Cbz) groups via treatment with *N*-(benzyloxycarbonyloxy)succinimide

(Cbz-OSu) (Figure 1.22). This orthogonal protection strategy allows the site-specific removal of the Boc protection group under acidic conditions resulting in a protein bearing a single reactive amino moiety. In parallel, intein-derived C-terminal Ub-thioester is treated with Cbz-OSu to globally protect all primary amino groups. The Ub-thioester is activated and converted *in situ* in the presence of Ag(I) to the corresponding *N*-hydroxysuccinimidyl ester to allow selective isopeptide formation with the target lysine in the substrate protein[249,250]. After successful ligation, the Cbz groups of the resulting Ub-POI conjugate are globally removed with an acidic cleavage cocktail[251], and the conjugate is subjected to refolding (Figure 1.22).

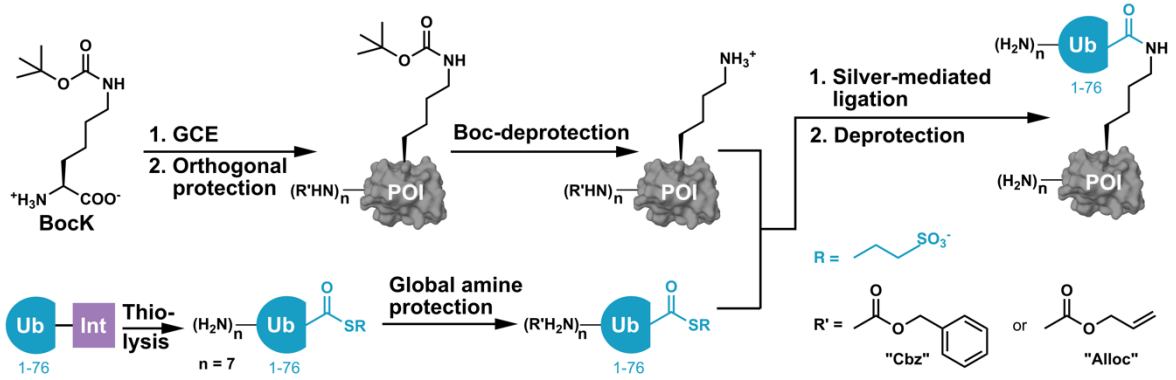


Figure 1.22 Schematic representation of GOPAL.

GOPAL enabled for the first time the generation of atypically linked diUbs displaying a native isopeptide bond and allowed structural elucidation of K6-diUb. Access to all eight Ub-linkages facilitated extensive DUB profiling and unravelled the substrate specificity of the OTU family DUB TRABID towards K29- and K33-diUbs[252]. Similarly, hydrolysis preference of the viral vOTU of CCHFV towards K6-, K11-, K48- and K63-diUb was demonstrated, while K29- and linear diUb were found to be refractory towards cleavage by vOTU[253]. GOPAL-derived diUbs were also used for developing a MALDI-TOF DUB screening assay[254] and for investigating the substrate specificity of the UCH family of DUBs[255]. The orthogonal protection strategy of initial GOPAL was further expanded by blocking the primary amines within POI and Ub using an allyloxycarbonyl (Alloc) protection group instead of Cbz[256]. The advantage of Alloc over Cbz as a protection group for global amine protection lies in milder deprotection conditions, typically induced by catalytic amounts of Ru(II) complexes in the presence of an excess of thiophenol[256]. Furthermore, the Boc protection group is completely stable under Alloc deprotection conditions, allowing the iterative use of GOPAL to generate more complex Ub-chain topologies. Both Alloc- and Cbz-protection strategies were extensively used for creating differently linked diUbs[257], homotypic Ub-chains[256,258,259] as well as branched Ub-chains[260]. These conjugates paved the way for structural investigations by NMR spectroscopy and top-down mass spectrometry[261,262]. Since GOPAL involves multiple chemical protection and deprotection steps, it is limited to refoldable proteins and has only been used in two occasions on substrate proteins other than Ub. One follow-up study expanded GOPAL towards the Ubl Rub1, the yeast orthologue of human Nedd8[263], and enabled generation of Rub1-Ub conjugates. Biochemical analysis of Ub(K29)-Rub1 and Ub(K48)-Rub1 as well as Rub1(K48)-Ub showed that certain DUBs also possess derubylase activity[264]. Apart from work with Ubls, GOPAL enabled the ubiquitylation of the DIX domain of dishevelled 2 (Dvl2), a eukaryotic protein that plays a fundamental role in Wnt signaling[265]. Interestingly, Dvl2 site-specifically ubiquitylated

at K54 was no longer able to undergo head-to-tail polymerization, which is important for signalosome formation, while ubiquitylation at K58 remains polymerization competent. Additionally, DUB profiling revealed that DUBs display certain specificities towards K54-linked and K58-linked Dlv2-Ub conjugates indicating potential regulatory functions of DUBs in the context of Wnt signaling.

Taken together, GOPAL enabled first access to natively linked atypical Ub chains. These chains allowed structural and functional investigations of diUbs and numerous DUB profiling studies. Nevertheless, the rather complex multistep procedure which relies on denaturing conditions, limits GOPAL to simple refoldable proteins. Furthermore, the attachment of multiple hydrophobic Cbz protection groups may lead to problems with protein solubility and purification, limiting the number of possible target proteins even further.

1.4.2.2 Native chemical ligation using δ-thiolysine

One of the archetypical reactions for forming amide bonds under mild and aqueous conditions is represented by NCL, an efficient and robust method to ligate peptides and protein fragments via peptide linkages[266]. Over the past 20 years, NCL has enabled synthesis and semi-synthesis of numerous protein fragments and full length proteins site-specifically decorated with PTMs, D-amino acids and biophysical probes[267]. The archetypal NCL reaction links two unprotected polypeptide fragments, one containing a C-terminal thioester and the other an N-terminal cysteine residue. Mechanistically, NCL involves an initial transthioesterification, in which the nucleophilic thiol moiety of the N-terminal cysteine residue attacks the C-terminal thioester. The formed thioester intermediate then spontaneously undergoes an S-to-N-acyl shift resulting in a native peptide bond (Figure 1.23 a). Initially, NCL approaches were limited to the presence of a cysteine residue at the ligation site, narrowing the potential ligation junctions within a protein to sites N-terminal of cysteine. To overcome this limitation, desulfurization protocols were developed to convert cysteine to alanine residues which, due to high natural abundance of alanine, profoundly increased possible scar less ligation junctions in proteins[268,269]. This development formed the basis for various efforts to leverage the potential of NCL in order to generate isopeptide-linked POI-Ub-conjugates. Conceptually, all strategies rely on the installation of an 1,2-aminothiol moiety at the acceptor lysine site within the substrate protein, which then undergoes NCL with a C-terminal Ub thioester generated via intein chemistry. A complex multistep procedure developed by Muir and coworkers that enabled first access to site-specifically ubiquitylated histone H2B relies on a combination of SPPS, recombinant protein expression and two subsequent NCL reactions. For this, intein-derived C-terminal Ub-thioester was reacted with a SPPS-derived peptide corresponding to the N-terminus of H2B (117-125) (Figure 1.23b), carrying both a photocaged N-terminal cysteine and at the acceptor lysine position (K120), a lysine modified with a photolytically removable, thiol-bearing ligation auxiliary, functioning as 1,2-aminothiol mimic[270]. A first NCL reaction resulted in natively ubiquitylated H2B peptide that was, upon deprotection of the N-terminal cysteine, susceptible for a secondary NCL with the thioester-modified C-terminal part of H2B (1-116). Subsequent desulfurization to the native A117 and *in vitro* reconstitution of nucleosomes bearing H2B(K120)-Ub allowed biochemical analyses of chemically defined ubiquitylated nucleosomes[271].

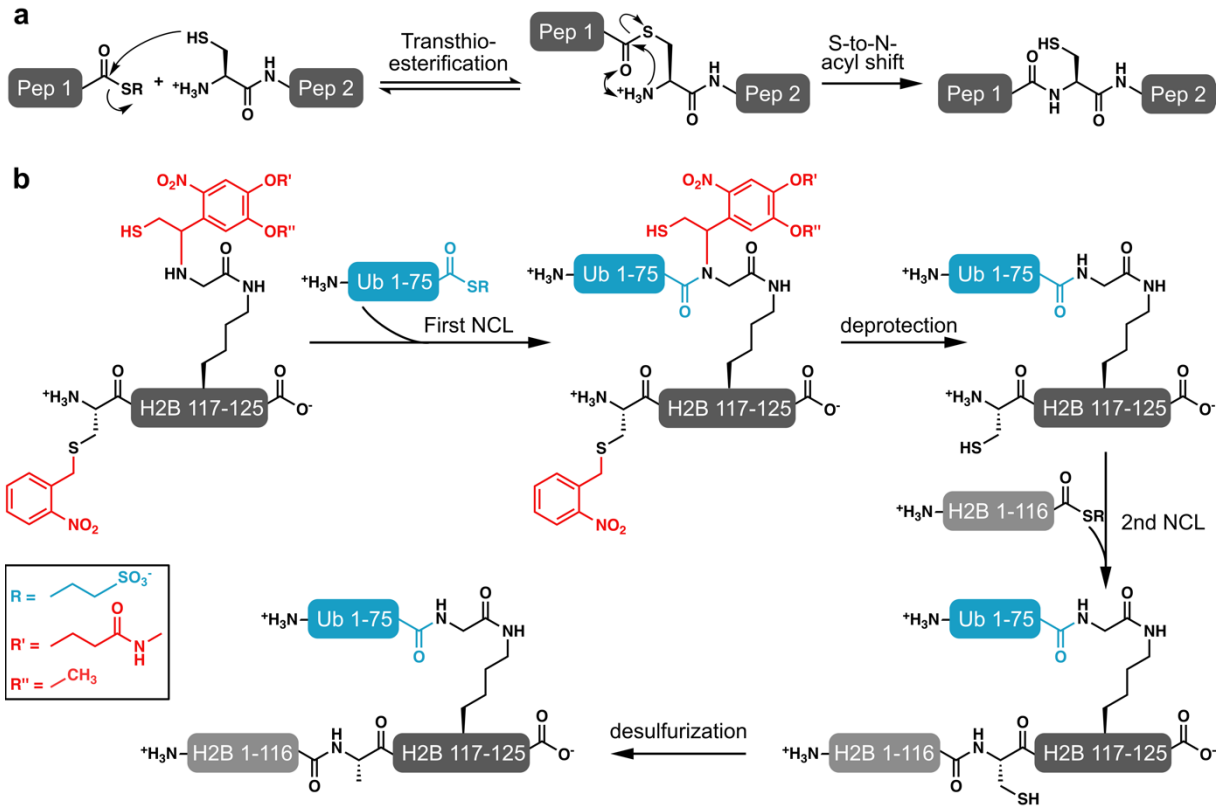


Figure 1.23 Schematic representation of NCL and the NCL-assisted semisynthesis of ubiquitylated H2B. a) Scheme showing an NCL between a peptide bearing a C-terminal thioester (Pep1) and a peptide displaying an N-terminal cysteine residue (Pep2)**. b)** Scheme displaying the semisynthesis of H2B ubiquitylated at K120 *via* two NCLs.

Further advancements of this approach placed the required 1,2-aminothiol or 1,3-aminothiol moiety within the side chain of lysine by attaching a sulfhydryl group to the γ- or δ-carbon of lysine (γ-ThioK and δ-ThioK) and facilitated thereby peptide monoubiquitylation[272,273], polyubiquitylation[274], the generation of diUbs[275,276], the assembly of a K48-linked tetraUb[277], as well as the synthesis of a tetraubiquitylated alpha synuclein[278] and ubiquitylated alpha globulin[279].

Strategies based on SPPS and SPPS in combination with NCL are limited to ubiquitylation of peptides and to ubiquitylation sites close to protein termini and thus require refolding of the Ub-POI conjugate. Consequently, attempts to site-specifically install 1,2-aminothiol moieties into proteins via genetic code expansion were envisioned to expand the number of potential target proteins. Therefore, a strategy combining genetic code expansion and NCL to generate POI-Ub-conjugates, which is based on the site-specific incorporation of the above described δ-thiolysine (δThioK)[223], that was first used by Brik and colleagues for ubiquitylating SPPS-derived peptides, was developed. To differentiate δThioK from lysine for selective recognition by a PylRS mutant, a UAA bearing a *p*-nitroCbz protection group at the Nε was designed and synthesized (δ-thiol-*N*ε-(*p*-nitrocarbobenzyloxy)-L-lysine (*p*-nitroCbz-δThioK))[223]. This protection group is removed post-translationally by reduction of the nitro group and concomitant elimination making additional deprotection steps obsolete. Interestingly, MS-analysis indicated that successful incorporation of *p*-nitroCbz-δThioK into a POI did not result in δThioK-modified proteins, but with a thiazolidine adduct stemming from condensation of the 1,2-aminothiol moiety with cellular pyruvate. The thiazolidine adduct could however be decaged by treating the protein with methoxyamine.

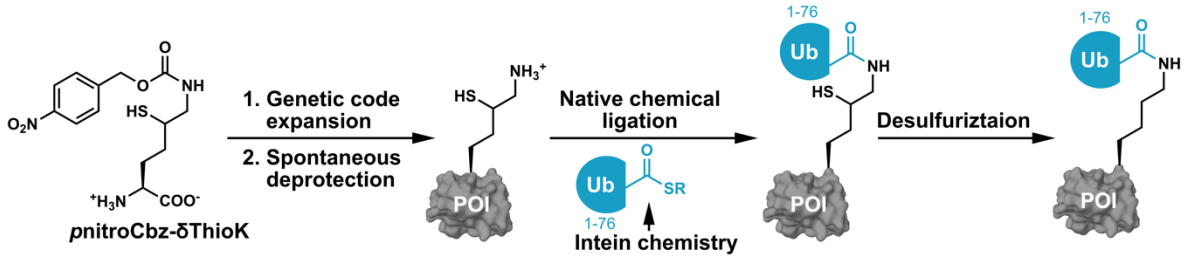


Figure 1.24 Schematic representation of an NCL-based approaches for the generation of Ub-POI conjugates.

To prove the functionality of resulting ubiquitylated proteins, δThioK was site-specifically incorporated into Ub at position K6 followed by thiazolidine cleavage and NCL using intein-derived Ub-thioester (Figure 1.24). K6-diUb bearing δThioK at the ligation site was then desulfurized (using a free radical method[268]) followed by refolding to give a completely native diUb. DUB screening assays confirmed NCL-derived K6-diUb as substrate for USP2 and USP5 and expansion of the approach towards ubiquitylation of SUMO at K11 demonstrated the applicability towards a non-Ub target protein.

In conclusion, NCL between δThioK-bearing proteins and a Ub-thioester is capable of producing completely native conjugates when a free radical desulfurization step is included after ubiquitylation, but the amino acid requires rather complex chemical synthesis. Additionally, the desulfurization requires denaturing conditions limiting the approach to refoldable proteins. Therefore, this approach has only been used for Ub and SUMO as target proteins. Since intein-mediated thiolysis, which is used to generate Ub thioester, can be easily transferred towards other Ubls, NCL-based approaches have the potential to be expanded to other Ubls.

1.5 Transpeptidases and their use in chemical biology

Protein semisynthesis is an emerging field in chemical biology and allows the precise construction of designer proteins. The methodological repertoire to selectively stitch peptide bonds between either unprotected synthetic peptides and expressed proteins or between two protein fragments is primarily based on adapting and evolving natural mechanisms[280]. This is exemplified on above discussed approaches like NCL and expressed protein ligation (EPL), which are inspired by naturally occurring protein splicing events[266,267]. Similarly, the discovery of protein trans splicing, a process where the intein is split in two separately translated peptides, which upon reassembly perform autocatalysis and thereby join the adjacent exteins via a peptide bond, forms the basis of a variety of approaches which now offer the opportunity to introduce synthetic fragments in a chemoenzymatic manner[281].

Another naturally occurring mechanism which includes the post-translational formation of a peptide bond is transpeptidation, which is defined as the process of transferring a peptide from one to another. Transpeptidation is catalyzed by a class of enzymes, known as transpeptidases, which recognize and cleave a certain amino acid motif to form a substrate-enzyme intermediate followed by transfer of the peptide onto a suitable amine acceptor. The most prominent example of this class of enzymes is sortase A (SrtA) from *Staphylococcus aureus*, which catalyses the transpeptidation between a five-amino acid recognition motif and a N-terminal polyglycine moiety. Endogenously, sortase A is responsible for the modification of the bacterial cell wall by anchoring proteins which display the LPXTG recognition motif to the peptidoglycan building block lipid II[282]. Since the discovery of SrtA, a variety of different sortases have been identified in organisms such as *Listeria monocytogenes*, *Streptococcus pyogenes*, *Streptococcus pneumoniae* and *Bacillus subtilis*[283]. Most of these bacteria encode, based on the primary sequence, several classes of sortases (A - F), which fulfil different function *in vivo*, including pilus polymerization[284]. However, the most frequently used sortase remains the Ca^{2+}-dependent *S. aureus* SrtA. From a mechanistic point of view, SrtA recognizes the LPXTG motif and cleaves the peptide bond between threonine and glycine by forming a thioester intermediate through its active site cysteine, which is dependent on base catalysis by a nearby histidine. This thioester intermediate is then nucleophilically attacked by the N-terminal amine moiety of a polyglycine substrate, yielding a new amide bond between threonine in the recognition motif and the N-terminal glycine of the polyglycine substrate (Figure 1.25 a). Importantly, it has been shown for *in vitro* reactions with the ~145 amino acid transpeptidase domain of SrtA, that the presence of only one glycine is sufficient for successful transpeptidation[285]. First applications of SrtA for protein semisynthesis included the decoration of the C-terminus of GFP with synthetic peptides[286]. Therefore, the sortase recognition motif was placed at the C-terminus of GFP and reacted with glycine-based peptides in presence of SrtA. This strategy, dubbed sortase-mediated protein ligation has found wide applications in protein semisynthesis as well as in protein engineering approaches to decorate proteins with a plethora of different moieties such as fluorophores, oligonucleotides, metal chelators, and click chemistry handles[287-289].

The sortase-catalyzed transpeptidation reaction is inherently a reversible reaction, as the recognition motif is reinstalled in the product, resulting in low conversion rates. In order to push the equilibrium to the product site, usually an excess of the acceptor glycyl-bearing reaction partner is used. Another possibility is the removal of the leaving group by either dialysis[290] or

by using centrifugal filters[291]. Next to these rather simple optimization approaches, several sophisticated alternative approaches for product maximization have been developed within recent years. For example, by installing iterative "W-T" motifs both at the donor and the acceptor site, the back reaction is sterically inhibited by the formation of a β-hairpin structure formed by W-T zippers at the ligation site (Figure 1.25 c)[292]. Besides masking the product through steric hindrance, another possibility is represented by rendering the leaving group unreactive. This is on the one hand possible by replacing the scissile amide bond of the recognition motif with a depsipeptide, which is however restricted to synthetic building blocks[293]. On the other hand, recent reports introduced a concept dubbed metal-assisted sortase-mediated ligation, which is based on transforming the leaving group into a Ni^{2+}-binding peptide. This is achieved by extending the LPXTG motif to LPXTGGH leading to GGH as leaving group, which then undergoes complex formation with Ni^{2+} and can thus be removed from the equilibrium (Figure 1.25 c)[294,295].

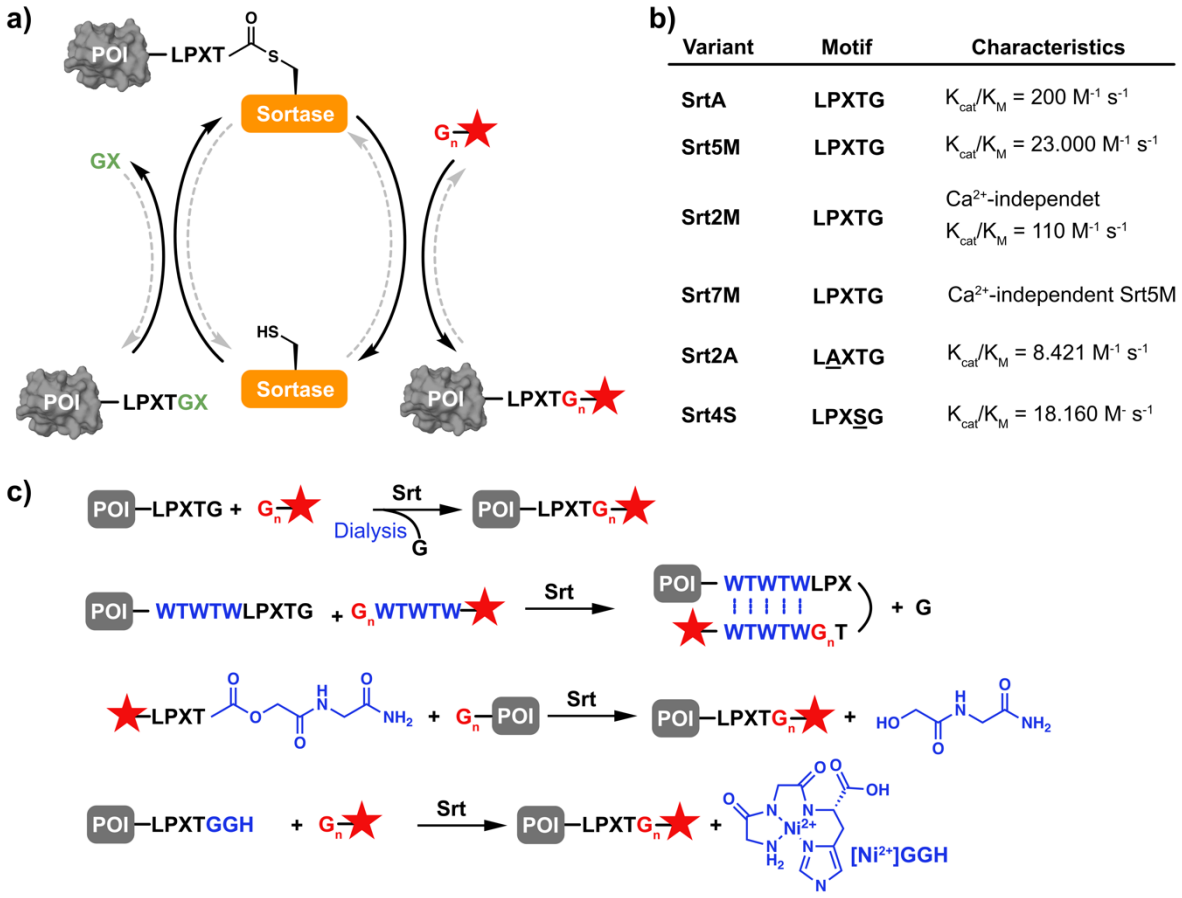


Figure 1.25 Sortase-mediated ligation. a) Schematic representation of sortase mediated ligation between a POI bearing the sortase recognition motif at its C-terminus and a polyglycine modified cargo (red star). **b)** Table showing different sortase variants and their characteristics.[296-300] **c)** Schematic representation of different approaches to shift the equilibrium towards the product site.

Due to several limitations of sortase-mediated ligation, such as the rather slow reaction kinetics of SrtA wt[297,298], the dependence on the LPXTG motif at the ligation site, as well as the calcium dependency of the enzyme, which prohibits its use *in vivo*[298,301], SrtA has been the target of numerous protein engineering and directed evolution approaches[297-300,302-307]. One of the milestones in directed evolution of sortases is represented by a yeast display approach

established by Liu and co-workers. By displaying a library of sortase mutants in proximity to a LPXTG bearing substrate on yeast cells, followed by an elegant cell sorting selection system with glycine-linked fluorophores, it was possible to identify a sortase mutant, dubbed sortase 5M (Srt5M), with five mutations that showed a 140-fold increased catalytic efficiency compared to the wt enzyme[297]. Using the same approach, it was also possible to reprogram the substrate specificity of Srt5M resulting in sortase variants that recognize the LPXSG (Srt4S) and the LAXTG (Srt2A) motif instead of the LPXTG (Srt5M) motif [300]. Furthermore, a rational engineering approach guided by the Ca^{2+}-independent SrtA enzyme from *S. pyogenes* resulted in a Ca^{2+}-independent *S. aureus* SrtA variant bearing two point mutations E105K and E108Q, dubbed Srt2M[299]. Importantly, it has been shown that the two point mutations conferring Ca^{2+}-independency are transferrable to Srt5M which led to the creation of Srt7M (Figure 1.25 b)[308]. Ca^{2+}-independent sortases expanded the scope of SML to living *E. coli*[298], and to *C. elegans*[309].

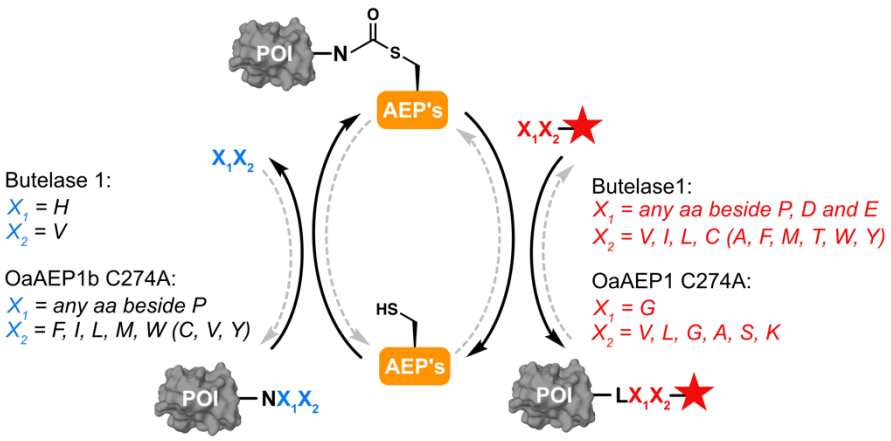


Figure 1.26 Transpeptidation reactions catalyzed by the AEP enzymes butelase 1 and OaAEP1.

Besides the well-established strategies utilizing the transpeptidase sortase, approaches employing peptide asparaginyl ligases (PAL), especially the subfamily of ligase-type asparaginyl endopeptidases (AEPs), have gained remarkable interest within recent years[280,310,311]. First reports showed that an enzyme derived from the butterfly pea *Clitoria ternatea* named butelase 1 is capable to catalyze the transpeptidation between an NHV motif and an acceptor dipeptide. Analogously to sortase, butelase 1 cleaves the peptide bond between N and H within the recognition motif, forming a thioester intermediate, followed by ligation with a variety of suitable acceptor dipeptides (Figure 1.26)[310]. Unfortunately, the recombinant expression of butelase1 remains difficult, thereby limiting its applicability. In contrast, the enzyme OaAEP1b can be recombinantly expressed in *E. coli* and was used for protein ligation and peptide cyclization. Alike butelase1, OaAEP1b recognizes an NXX motif (where XX is usually GL but can also be AL and CL) and cleaves the peptide bond adjacent to N followed by ligation with GL, AL or CL acceptor peptides[312]. While the wt OaAEP1b shows rather low catalytic efficiency (~200 M^{-1} s^{-1}) point mutation of the gatekeeper cysteine residue 274 to alanine increased the catalytic efficiency (~35000 M^{-1} s^{-1}) and substrate specificity paving the way for efficient protein modifications through transpeptidation including the modification of the Ub C-terminus with synthetic peptides (Figure 1.26)[313].

One last class of enzymes, which should be mentioned within the context of transpeptidases, are peptide ligases. Peptide ligases are often engineered variants of proteases, which favour the catalysis of an amide bond over hydrolysis. This was first exemplified on the serine protease subtilisin.

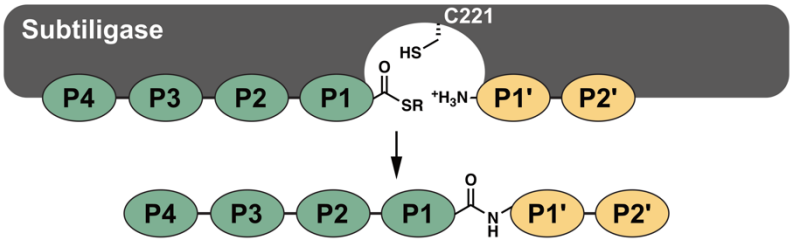


Figure 1.27 Scheme displaying subtiligase-mediated transpeptidation

Mutation of the active site serine to cysteine as well as the P225A mutation resulted in the enzyme subtiligase that catalyzes peptide bond formation over peptide hydrolysis[314]. Further optimization yielded stabiligase which showed increased stability in presence of guanidine hydrochloride[315]. Initially used with peptide oxoesters as substrates, recent studies have shown that subligase prefers peptide thioesters over oxoesters, improving the aminolysis reaction[316]. This preference further increases the applicability of subiligase since protein thioesters are easily accesible via intein technology. In contrast to transpeptidases like sortases, peptide ligases derived from proteases possess well-defined P and P' sites. For example subtiligase prefers large hydrophobic residue at the N-terminal P1 – P4 sites while the specificity at the P1' and P2' sites is rather flexible (Figure 1.27)[314]. Further engineering of subtiligase and variants thereof led to a broadened substrate scope[317,318], which allowed the semisynthesis of phosphorylated PTEN[319]. Next to subtiligase, subtilisin engineering also resulted in an enzyme called peptiligase[320] and a variant dubbed omniligase[321], which has an even broader tolerance for both the P1' and the P2' positions.

CHAPTER 2

Development of a novel approach for site-specific ubiquitylation (Sortylation)

The work in the following chapter is published as:

Fottner, M.; Brunner, A. D.; Bittl, V.; Horn-Ghetko, D.; Jussupow, A.; Kaila, V. R. I.; Bremm, A.; Lang, K. Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat Chem Biol* **2019,** *15* (3), 276.

The chapter is strongly based on the publication above. Display items and text were reprinted with permission from Springer Nature.

2. Development of a novel approach for site-specific ubiquitylation (Sortylation)

2.1 Aim & Background

As already discussed in Chapter 1, various methods have been developed over the last decade to address the challenge of creating Ub-POI conjugates to study this intriguing PTM[226]. However, targeted site-specific ubiquitylation of target proteins remains challenging *in vitro* and impossible in living cells since most of the available approaches rely on harsh conditions and are therefore limited to simple, refoldable proteins.

Here we therefore aimed to develop a novel enzyme-mediated approach to overcome these limitations. Among the different amide bond-forming enzymes (see Chapter 1.5) we decided to harness the potential of the transpeptidase SrtA. In contrast to other transpeptidases or engineered proteases, such as OaAEP1b or subtiligase, SrtA has already been used successfully in living *E. coli*[298] and *C. elegans*[309] and is therefore ideally suited for the development of a ubiquitylation approach that is supposed to function in living cells. SrtA catalyzes the transpeptidation between a five-amino acid motif (LPLTG), dubbed sortase motif, by cleaving the peptide bond between threonine and glycine resulting in a thioester intermediate that is subsequently nucleophilically attacked by an N-terminal polyglycine moiety generating a new peptide bond. During our analysis of the Ub C-terminus (LRLRGG) we noticed that only two point mutations are needed to introduce the archetypal SrtA motif (LPLTGG). We therefore reasoned that incubation of Ub bearing the sortase motif at its C-terminus with SrtA will lead to formation of the thioester intermediate between the active site cysteine in sortase and the carboxy group of threonine within the mutated Ub C-terminus (Figure 2.1). In order to attach the Ub variant site-specifically to a substrate protein, we designed a novel UAA called GGK, in which a glycylglycine moiety is coupled to the ε-amino group of lysine *via* an isopeptide-bond. We envisioned to site-specifically incorporate GGK into a POI via genetic code expansion. A GGK-bearing POI should then be capable to nucleophilically attack the thioester intermediate between Ub and sortase, resulting in a Ub-POI conjugate that only differs from its native counterpart by two point mutations (R72P and R74P, Figure 2.1).

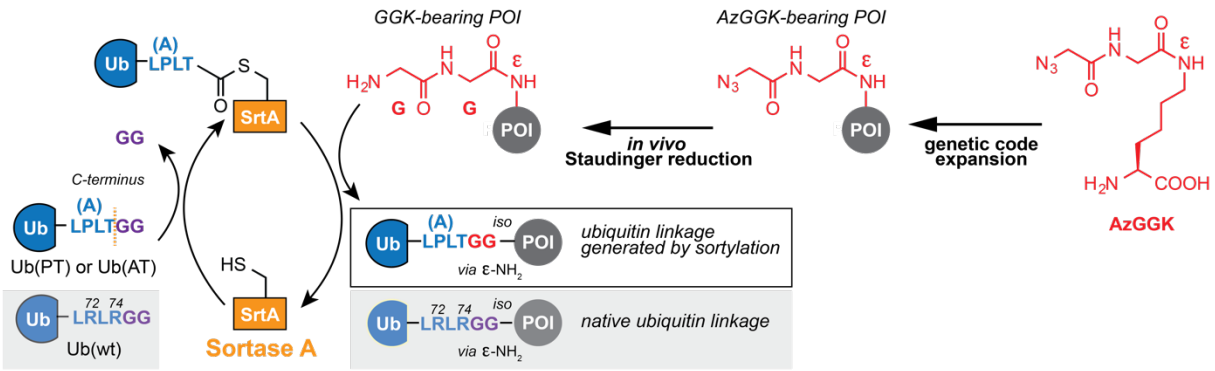


Figure 2.1 General scheme showing sortase-mediated ubiquitylation (sortylation). The unnatural amino acid AzGGK is site-specifically incorporated into proteins *via* genetic-code expansion. *In vivo* Staudinger reduction converts AzGGK-bearing proteins to GGK-bearing proteins, which in turn undergo transpeptidation with a modified Ub bearing a sortase recognition motif (LPLTG or LALTG) *via* Srt5M/SrtA. Generated ubiquitylated proteins display a native isopeptide bond and two point mutations (R72P or R72A and R74T) in the linker region.

To expand our approach towards living cells, we decided to mask the primary amino group of glycine within GGK in order to confer inducibility over the transpeptidation reaction. Azides can be quantitatively reduced to the corresponding amines *via* Staudinger reduction with water-soluble and cell-permeable phosphine reagents such as 2DPBA (2-(diphenylphosphino)benzoic acid)[322]. We therefore replaced the primary amine of GGK with an azide moiety, creating the UAA AzGGK (Figure 2.1). We hypothesized that AzGGK may not take part in a sortase-catalyzed transpeptidation with modified Ub until the azide is reduced by Staudinger reduction, which will reveal the amine nucleophile. Thereby we would be able to selectively trigger ubiquitylation events within living cells. We called our approach sortylation.

2.2 A strategy for sortase-mediated peptide ligation employing a novel UAA

In order to investigate if the UAA GGK may act as an acceptor in a sortase-mediated transpeptidation reaction, we decided to test our hypothesis in a peptide assay. We used standard Fmoc-SPPS to synthesize a seven amino acid long peptide resembling the sortase-compatible C-terminus of Ub (Fmoc-VLPLTGG, Figure 2.2 b). The N-terminal Fmoc-protection group was included in the peptide to facilitate monitoring of the reaction via LC-MS at 280 nm. In a similar manner, we synthesized GGK by coupling *tert*-butyloxycarbonyl protected diglycine to the ε-amino group of resin bound Nα-protected lysine (see chapter 4.1). Since wt SrtA is a rather slow enzyme, we decided to use an evolved variant, called Srt5M, which shows enhanced kinetics (Chapter 1.5)[297]. After expression and purification of Srt5M *via* affinity chromatography using Ni-NTA, we obtained up to 20 milligrams Srt5M per litre of pure protein. Finally, we incubated the Fmoc-VLPLTGG peptide with a 10-fold excess of GGK in the presence of Srt5M. To our delight, LC-MS analysis showed nearly quantitative formation of the transpeptidation product (Fmoc-VLPLTGGK) within 30 minutes (Figure 2. 2 a, b and c). We subsequently wanted to investigate if we can substitute Fmoc-VLPLTGG with a Ub variant bearing two point mutations in its C terminus and thereby displaying the sortase motif (Ub-LPLTGG (Ub(PT)). Analogously to the reaction setup described above, we incubated Ub(PT) with 100-fold excess of GGK in the presence of Srt5M. Excitingly, LC-MS analysis after overnight incubation showed complete conversion of Ub-LPLTGG to Ub-LPLTGGK (Figure 2.2 d). To evaluate our hypothesis that the azide group of AzGGK blocks the transpeptidation and that reduction using 2DPBA restores the reactivity, we synthesized AzGGK using SPPS by coupling glycine and azidoacetic acid subsequently onto resin-bound Nα-protected lysine (see chapter 4.1). We incubated Fmoc-VLPLTGG with 10-fold excess of AzGGK in the presence of Srt5M. As expected, we could not observe any product formation. However, after addition of 2DPBA to the reaction mixture, sortase-mediated transpeptidation proceeded smoothly, proving that the azide group in AzGGK effectively blocks the UAA to act as an acceptor nucleophile in sortase-mediated transpeptidation (Figure 2.2 c).

On the basis of our peptide assays, we envisioned that the site-specific incorporation of AzGGK into a POI and its subsequent conversion to GGK *via* Staudinger reduction will be well suited to lay the foundation for a chemoenzymatic approach for site-specific ubiquitylation of proteins with temporal control and the next challenge therefore consisted in site-specifically incorporating AzGGK into proteins via genetic code expansion.

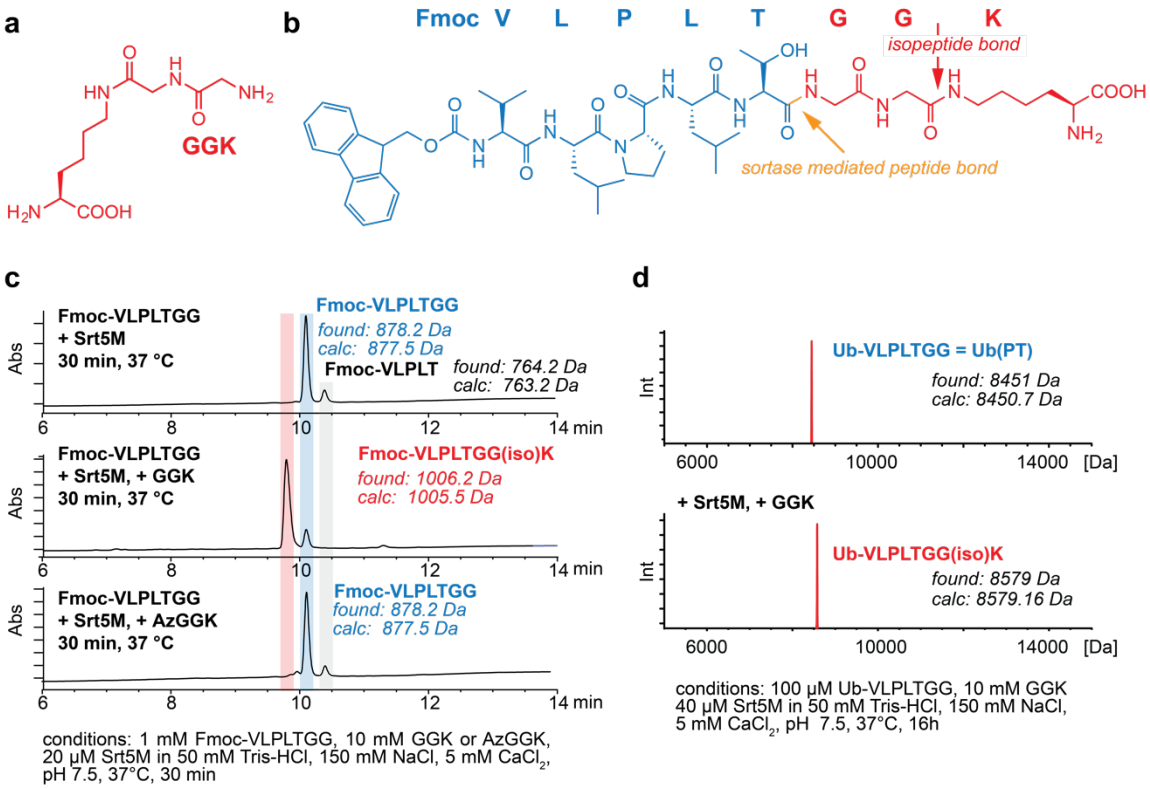


Figure 2.2 Sortase-mediated transpeptidation between GGK and Fmoc-VLPLTGG or Ub-VLPLTGG. a) Structural formula of the unnatural amino acid GGK. **b)** Structural formula of the sortase-mediated transpeptidation product that is formed between peptide Fmoc-VLPLTGG and GGK. **c)** LC-MS analysis of Srt5M-mediated transpeptidation between Fmoc-VLPLTGG and GGK shows close to quantitative formation of product within 30 minutes (red box), while incubation of Fmoc-VLPLTGG with AzGGK in the presence of Srt5M did not yield the transpeptidation product, as expected (blue box). In the absence of GGK, hydrolysis of the thioester, formed between Srt5M and Fmoc-VLPLTGG, did lead to small amounts of hydrolyzed peptide Fmoc-VLPLT (grey box). Chromatogram shows the absorption at 280 nm. **d)** Overnight-incubation of a Ub mutant displaying two mutations in its C-terminus (=Ub(PT)) with GGK and Srt5M leads to formation of the expected product.

2.3 Site-specific incorporation of AzGGK

Encouraged by our sortase-mediated peptide ligation assays, we next set out to site-specifically incorporate AzGGK into proteins. As described in Chapter 1.3, a large variety of structurally diverse UAAs have been genetically encoded in response to an amber codon introduced into a gene of interest by employing engineered pyrrolysyl tRNA synthetase (PylRS)/$tRNA_{CUA}$ pairs from *Methanosarcina* species, including *M. barkeri* (*Mb*) and *M. mazei* (*Mm*)[323,324].

However, many of the lysine-based UAAs which can be incorporated *via* PylRS variants display an alkyl side chain with a bulky and often hydrophobic moiety that is attached to lysine *via* a carbamate linkage[325,326]. Only in a few lysine-derivatives, for which PylRS mutants have been engineered, the side chain is attached *via* an isopeptide-bond to the ε-amino group of lysine, similar as in the wt substrate pyrrolysine. In contrast to AzGGK, none of the so far incorporated lysine-derivatives carries a planar and polar peptide-bond in its side chain. Guided by structural analyses of the C-terminal catalytic centre of wt PylRS and its mutants[327], we screened a panel of > 25 different *Mb*PylRS mutants that accepted lysine-derivatives with long and bulky side chains for their ability to direct the selective and site-specific incorporation of AzGGK.

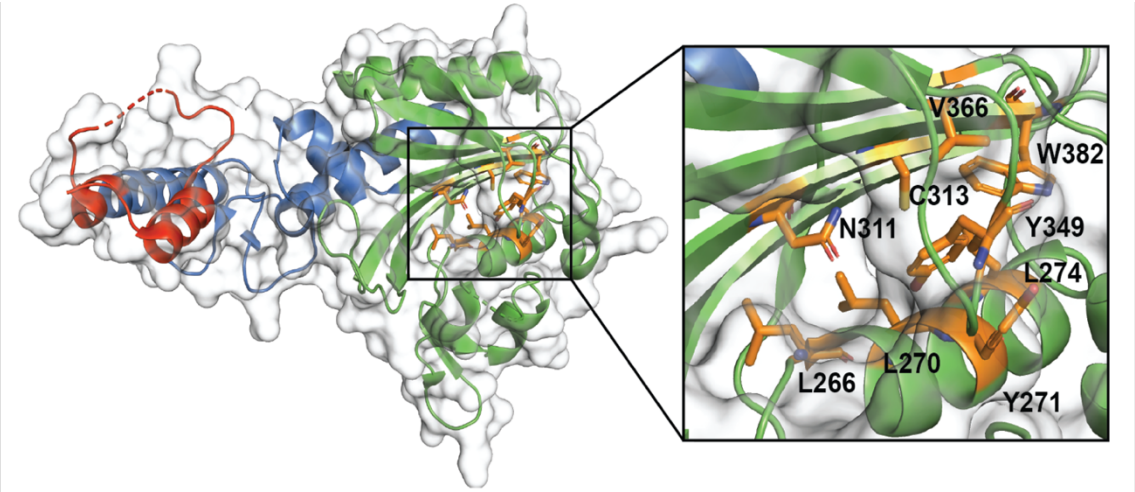


Figure 2.3 Structural analysis of wild type PylRS. Crystal structure of *Mm*PylRS (PDB: 2Q7H)[327]. The PylRS C-lobe is shown in cartoon model in red, blue and green. Key amino acid positions in the active site are displayed in stick model in orange. The blue and green part (723 bp) represent the part subjected to DNA-shuffling, the green part (495 bp) was subjected to error-prone PCR. The PylRS variants used for the shuffling approach contained amino acid mutations at positions highlighted in orange.

Unfortunately, none of the tested PylRS variants charged AzGGK onto its respective tRNA. Therefore, we decided to create a new PylRS library by combining DNA-shuffling of the PylRS C-terminal domain of 17 known synthetases with an unbiased error-prone PCR product of the catalytic wt-PylRS C-lobe (Figure 2.3)[328]. With the tailor-made library in hand we carried out directed evolution towards the incorporation of AzGGK by performing several rounds of positive and negative selection in *E. coli* (as described in chapter 1.3)[329]. In the last round we combined the positive selection step with a fluorescence readout by co-transforming the library DNA isolated from the surviving clones from the negative selection with a reporter plasmid bearing both a chloramphenicol acetyltransferase gene interrupted by an amber codon, as well as an sfGFP gene interrupted by an amber codon. Gratifyingly, we observed sfGFP-expressing colonies that grew in the presence of chloramphenicol. We picked these clones and evaluated the efficiency and selectivity for the incorporation of AzGGK, by measuring GFP fluorescence intensity in the presence and absence of AzGGK. Mutants that showed at least 5-fold fluorescence intensity increase in the presence of AzGGK were used for preparative scale expression of C-terminal His-tagged sfGFP bearing an amber codon at position 150 (sfGFP-N150TAG-His). To our delight, *E. coli* containing a PylRS mutant with the three mutations L274A, N311Q and C313S in the enzyme active site (dubbed AzGGKRS) and a plasmid that encodes *Mb*tRNA$_{CUA}$ and sfGFP-N150TAG-His6 led to the amino acid-dependent synthesis of full-length sfGFP (Figure 2.4 a).

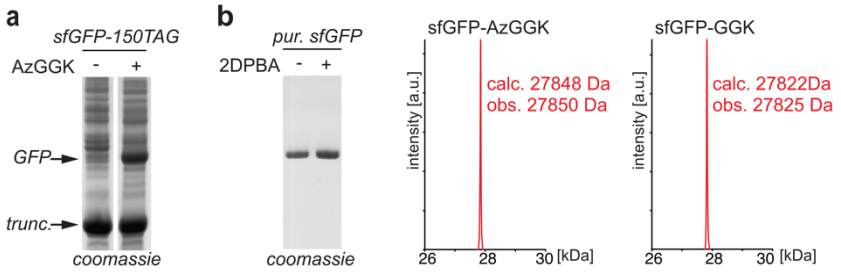


Figure 2.4 Site-specific incorporation of AzGGK into proteins in *E. coli*. a) SDS-PAGE analysis of AzGGK-dependent expression of full-length sfGFP bearing an amber codon at position 150. **b)** Purified sfGFP-N150AzGGK can be reduced to sfGFP-N150GGK on protein by treatment with 2DPBA. LC-MS-analysis confirms the integrity of sfGFP-N150AzGGK and sfGFP-N150GGK.

LC-MS analysis confirmed incorporation of AzGGK. Apart from the mass peak corresponding to sfGFP-AzGGK, we observed a further low-intensity peak (approximately 10-20%) in the LC-MS analysis, which we could assign to misincorporation of phenylalanine (Figure S2.1 a and b). In order to abolish the misincorporation, we performed sfGFP expression under auto-induction conditions[330] without phenylalanine. Purification of sfGFP *via* affinity chromatography using Ni-NTA followed by LC-MS analysis led to a clean mass without misincorporation (Figure S2.1 c). Next we investigated if the azide moiety of AzGGK can be reduced to the corresponding amine moiety by Staudinger reduction. Gratifyingly, incubation of sfGFP-AzGGK with 2 equivalents of 2DPBA followed by LC-MS analysis revealed quantitative reduction to sfGFP-GGK within 10 minutes at room temperature (Figure 2.4 b). Directed evolution and library generation was performed by the former master student Andreas-David Brunner.

2.4 Ubiquitylation of GGK-bearing proteins *via* sortase-mediated transpeptidation

With GGK-containing proteins in hand, we next investigated if it was possible to site-specifically ubiquitylate these modified proteins using Srt5M. As a first experiment, we incubated sfGFPN150GGK with an excess of Ub(PT) in the presence of Srt5M at 37 °C. Gratifyingly, we observed formation of the expected sfGFPN150-Ub(PT) conjugate, which was also confirmed by LC-MS (Figure 2.5 a and b). Importantly, when we used sfGFP bearing the unreactive UAA BocK instead of GGK, we could not observe any unspecific ubiquitylation. Similarly, incubation of sfGFP-GGK and Ub(PT) in the absence of Srt5M did not result in sfGFP-Ub conjugate formation (Figure 2.5 a and Figure S2.2 for full gels). Even though ubiquitylation was selective for GGK-containing proteins and LC-MS confirmed the correct conjugate, only 10-20% of sfGFP-GGK were successfully ubiquitylated after three hours of incubation (Figure 2.5 a and Figure S2.2).

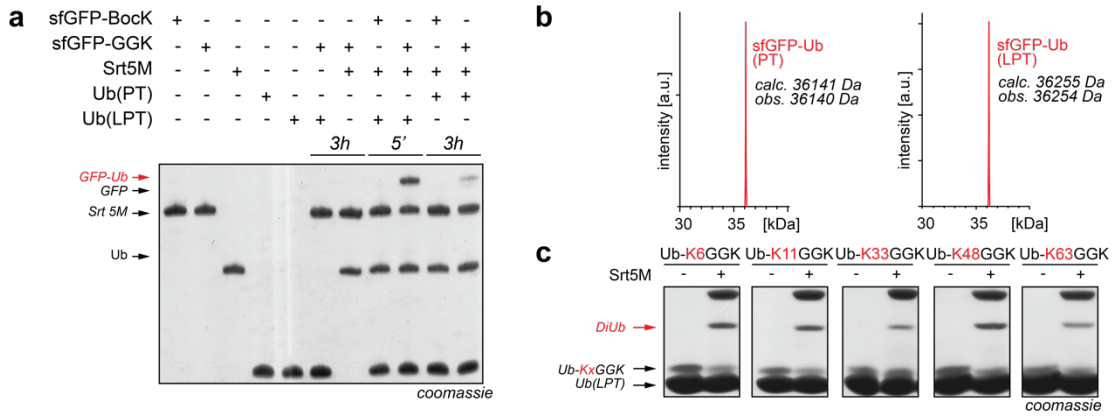


Figure 2.5 Srt5M-mediated ubiquitylation of GGK-bearing proteins. a) SDS-PAGE analysis of Srt5M-mediated formation of sfGFP-Ub conjugates confirms the specificity of the approach for GGK-bearing sfGFP. **b)** MS-analysis of sfGFP-Ub(PT) and sfGFP-Ub(LPT) conjugates. **c)** SDS-PAGE analysis of Srt5M-mediated diUb formation between Ub-KxGGK and Ub(LPT) at 37 °C, 1 hour (Ubiquitylation assays were carried out as described in Chapter 4.3; Sortylation yields are given in Figure S2.3).

We reasoned that the low conversion may stem from restricted access of Srt5M to the unstructured C-terminus of Ub(PT), as the first leucine of the sortase-motif LPLTGG (L71 in Ub) is engaged in the C-terminal ß-sheet of Ub. In order to confer better accessibility of Srt5M to its recognition motif, we introduced a leucine spacer within the LPLTGG sequence between L71 and P72, generating Ub(LPT) with an LLPLTGG C-terminal sequence to shift the sortagging-motif one amino acid away from the compact Ub ß-grasp fold, making it more accessible (Figure 2.6). To our delight, incubation of Ub(LPT) with sfGFP-GGK in the presence of Srt5M afforded ca. 50% of ubiquitylated sfGFP within five minutes (Figure 2.5 a and Figure S2.2) and LC-MS analysis confirmed the identity of the sfGFP-Ub(LPT) conjugate (Figure 2.5 b). Importantly, control experiments with sfGFP-BocK in the presence of Ub(LPT) and Srt5M did not show any formation of ubiquitylated sfGFP, confirming the selectivity of the transpeptidation reaction (Figure 2.5 a)

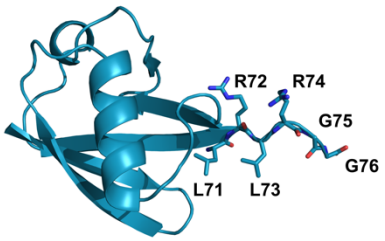


Figure 2.6 Structural representation of Ub. The six C-terminal amino acids are displayed as sticks.

Encouraged by these results we next set out to generate diUbs with different linkages using sortylation. For this, we individually expressed C-terminal His6-tagged ubiquitins with AzGGK at position K6, K11, K27, K29, K33, K48 and K63, respectively. The Ub-AzGGK variants were reduced to their Ub-GGK counterparts using 2DPBA and were characterized by SDS-PAGE and LC-MS (Figure S2.3 a and b). Incubation of five of the Ub-GGK variants (Ub-K6GGK, Ub-K11GGK, Ub-K33GGK, Ub-K48GGK and Ub-K63GGK) with a five-fold excess of Ub(LPT) in the presence of Srt5M resulted in efficient (up to 62 % conversion rate) and selective formation of site-specifically isopeptide-linked diUbs within an hour (Figure 2.5 c) Unfortunately, K27-diUb and K29-diUb could not be generated using sortylation (Figure S2.3 c). Since K27 and K29 are less exposed than the other lysine residues in the Ub fold we speculated that Ub-K27GGK and Ub-K29GGK may not be able to nucleophilically attack the thioester intermediate formed between Ub(LPT) and Srt5M due to steric clashes.

In order to introduce the Srt5M motif, we modified the C-terminus of Ub to LPLTGG (or LLPLTGG). Although we were able to produce monoubiquitylated sfGFP and diUbs, we worried that especially the proline residue introduced at position 72 of Ub might be a poor mimic for the native C-terminus, as it could result in an unusual conformational rigidity and might increase the likelihood of the cis-isomer of the L-P peptide-bond as opposed to the native L-R peptide-bond. We therefore turned our attention to a recently evolved sortase, Srt2A which shows a reprogrammed substrate specificity (LAXTG recognition motif instead of LPXTG, see chapter 1.5)[300]. In order to implement Srt2A into our approach, we expressed and purified Srt2A and Ub bearing an LALTGG (dubbed Ub(AT)) or an LLALTGG (dubbed Ub(LAT)) C-terminus. Incubation of both Ub(AT) and Ub(LAT) with sfGFP-GGK in the presence of Srt2A yielded the corresponding sfGFP-Ub conjugates whose identity was confirmed by LC-MS (Figure S2.4 a and b). The sfGFP-Ub(LAT) conjugate formed with conversion rates of ca. 50% within five-minute incubation at 37 °C. Analogously to experiments with Srt5M, control experiments with sfGFP-BocK in the presence of Srt2A and Ub(AT) or Ub(LAT) showed no formation of sfGFP-Ub conjugate, demonstrating the specificity and selectivity of Srt2A-mediated ubiquitylation.

2.5 Srt2A-mediated generation of non-hydrolyzable diubiquitins

After establishing the site-specific ubiquitylation of sfGFP using Srt2A, we decided to turn our attention on creating differently linked diUbs. Therefore we incubated all seven Ub-GGK-His6 acceptor ubiquitins with donor ubiquitins Ub(AT) and Ub(LAT) in the presence of Srt2A. As already observed for Srt5M, K6-, K11-, K33-, K48- and K63-linked diUbs were formed in very good conversion yields (up to 75 %, Figure S2.5 and S2.6) employing either Ub(AT) or Ub(LAT) (Figure 2.7). Importantly, incubation of Ub-K6GGK with Ub(wt), displaying the native C-terminal sequence LRLRGG did not lead to diUb formation (Figure S2.5 a). As already described in Chapter 1.5, sortase-mediated transpeptidation is an inherently reversible process, since the sortase recognition motif is re-installed in the product. Thus, sortase-generated diUbs can in principle be hydrolysed back to the corresponding mono-Ubs. In order to identify optimal reaction conditions for preparative purposes, we set up a panel of K6-diUb assays with varying concentrations of Srt2A. As expected, highly increased Srt2A concentrations accelarated product formation but also the hydrolyzation rate (Figure S2.7).

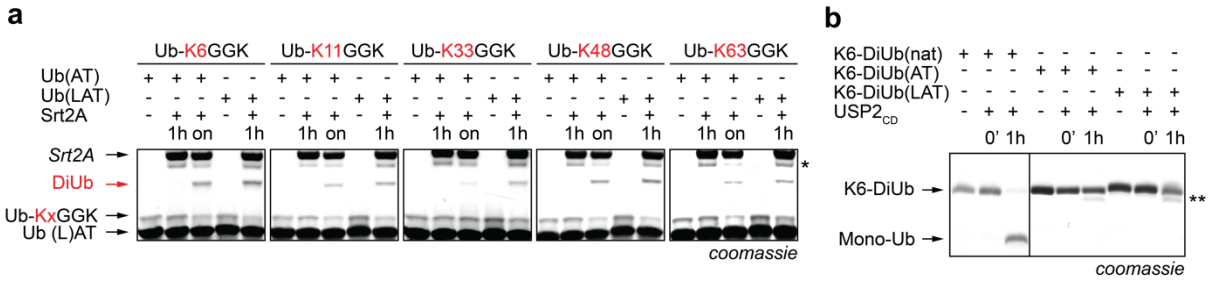


Figure 2.7. Srt2A-mediated generation of diUbs and deubiquitylation assays. **a)** Srt2A-mediated formation of diUbs analyzed by SDS-PAGE. *denotes an impurity from Srt2A (Ubiquitylation assays were carried out as described in Chapter 4.1). **b)** Incubation of diUbs with $USPD_{CD}$ shows that Srt2A-generated K6-diUb(AT) and K6-diUb(LAT) are stable towards DUB cleavage, while natively linked K6-diUb is quantitatively cleaved within an hour at 37 °C. **denotes bands that correspond to diUbs where the C-terminal His6-tag of the acceptor Ub has been cleaved off.

Accordingly, we upscaled diUb formation reactions, which were stopped at maximum product formation by adding the cysteine protease inhibitor phenylvinylsulfone (PVS). The resulting Srt2A-generated diUbs were purified *via* affinity chromatography using Ni-NTA, followed by size-exclusion chromatography (SEC). To our delight, we were able to obtain all diUbs in milligram scale and confirm their identity by LC-MS (Figure S2.7). As sortylation works under native conditions in aqueous buffers, refolding of the generated diUbs is not necessary, which is an advantage over described chemical methods that often rely on acidic deprotection conditions and desulfurization protocols.[223,224,226,331]

With differently linked diUbs in hand, we set out to investigate if our isopeptide-linked diUbs are resistant to DUBs. It has been shown previously that the mutation L73P in the Ub C-terminus confers resistance to various DUB families.[332] Since we introduced two mutations in the flexible C-terminus of Ub R72A and R74T we were curious if our diUbs would also be refractory to cleavage by DUBs. Therefore, we incubated the purified sortase-derived diUbs and commercially available natively linked diUbs with the catalytic domain of Ub carboxyl terminal hydrolase 2 ($USP2_{CD}$). SDS-PAGE analysis revealed, that all native diUbs are efficiently cleaved to the corresponding monoUbs within one hour, as expected, while all of the sortase-generated diUbs are resistant to $USP2_{CD}$ cleavage (Figure 2.7 b and Figure S2.8). Importantly, we could observe a faint band that migrated slightly further than our diUbs in SDS-

PAGE gels. Using LC-MS we could prove that this band corresponds to cleavage of the C-terminal His-tag of the acceptor Ub. Overnight incubation of sortase-generated diUbs with $USP2_{CD}$ led to quantitative cleavage of the C-terminal His6-tag, while the isopeptidic linkage was resistant to cleavage (Figure S2.8).

2.6 Structure and biological function of sortase-generated diUbs

After proving that diUbs produced *via* sortylation are resistant towards DUBs we were interested to further investigate their structural properties. As already introduced in Chapter 1.2, Ub chain topology is a crucial determinant of Ub-mediated cellular signaling. To translate the Ub code, effector proteins bearing single or multiple UBDs convert the different types of Ub linkages into functional outcomes[101,135,333]. In order to evaluate structural and dynamic properties of monoUb variants containing the sortase recognition motif, as well as sortase-generated diUbs, we performed atomistic molecular dynamic (MD) simulations initiated from the experimental structures of monoubiquitin (PDB: 1UBQ)[334], K48- (PDB: 1AAR)[335] and K63-linked diUb (PDB: 2JF5)[336] with *in silico* modelled AT, and LAT linkers.

MD simulations were performed by our cooperation partner Alex Jussupow from the group of Ville Kaila and showed that internal dynamics of the individual Ub monomers do not alter upon AT and LAT substitutions in the C-terminus (Supporting Figure S2.9). Furthermore, we could not observe dynamical differences between the flexible open form of K63-linked diUb (K63-diUb(wt)) and the sortase-generated K63-linked diUbs (K63-diUb(AT) and K63-diUb(LAT)). For the more compact K48-diUb we observed minor structural rearrangements with the sortase-generated diUb containing the spacer leucine amino acid (LAT), but not for the variant bearing the AT substitutions in the linker. These structural arrangements do, however, not affect the relative orientation of the hydrophobic patches at the diUb surface that are essential for recognitions by UBDs. In conclusion, our simulations suggest that the AT and LAT substitutions in sortase-generated diUbs result in unaltered diUb dynamics, closely mimicking the behavior of the native linker.

Encouraged by the results of this computational study, we set out to validate experimentally if sortase-generated diUbs bearing AT or LAT substitutions in the linker region are selectively recognized by specific Ub binding proteins. Therefore, we first probed our sortase-generated K63-diUb with a K63-linkage specific antibody. To our delight, both K63-diUb(AT) and K63-diUb(LAT) were recognized to a similar extent as native K63-diUb in anti-K63 western blots, indicating that AT or LAT substitutions in the linker region do not compromise binding to the linkage specific antibody (Figure 2.8 a). UBDs recognizing K63-linked Ub topologies play important roles in the regulation of several cellular functions, including activation of NF-κB signaling and DNA damage repair[101]. An important upstream event in NF-κB signaling requires activation of the protein kinase TAK1 through binding to K63-linked Ub *via* its adaptor subunit TAB2 that contains an NZF UBD[337]. As seen in crystal structures of TAB2-NZF bound to K63-diUb, the I44 patches of each Ub interact with two distinct binding sites on TAB2-NZF[338].

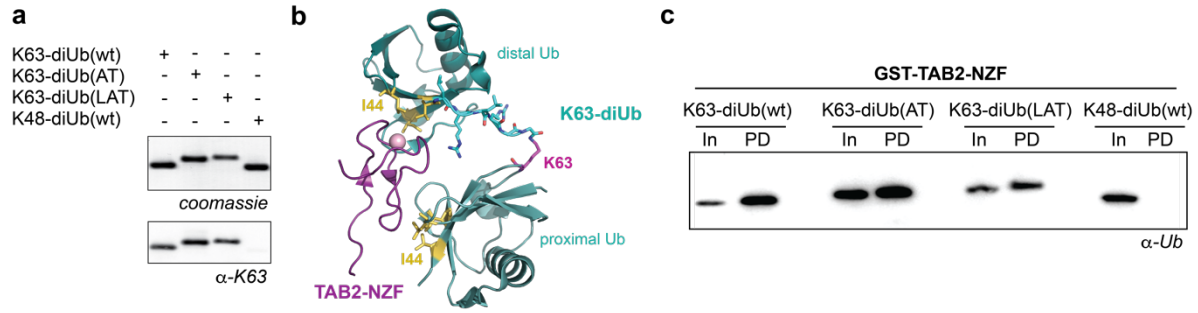


Figure 2.8 Structural and physiological integrity of sortase-generated K63-diUbs. a) Sortase-generated K63-diUbs are recognized by linkage-specific anti-K63 antibody. **b)** Structure of K63-diUb bound to TAB2-NZF shows interactions of the I44 patches of proximal and distal diubiquitins with TAB2-NZF (PDB: 2wwz).[338] **c)** GST-fused TAB2-NZF was used for pull-down assay with K63-diUb(wt), K63-diUb(AT), K63-diUb(LAT) and K48-diUb(wt). We could not observe interactions of K48-linked diubiquitin and TAB2-NZF in pull-down analyses, while K63-linked sortase-generated diUbs could be pulled-down with similar efficiency as K63-diUb(wt). 1% of protein was loaded as input for pull-down experiments.

The TAB2-NZF domain does not contact the isopeptide linker but senses the conformational features of K63 linkages for specific recognition (Figure 2.8 b). To analyze the binding capability of sortase-generated K63-diUb(AT) and K63-diUb(LAT) we performed *in vitro* pull down assays using glutathione-S-transferase (GST)-fused TAB2-NZF[339]. Therefore we incubated purified K63-diUb(AT), K63diUb(LAT) and natively linked K63-diUb together with GST-tagged TAB2-NZF, followed by washing and western blot analysis. Both K63-diUb(AT) and K63-diUb(LAT) retained the ability to bind TAB2-NZF, indicating that sortase-generated K63-linked diUbs are able to adopt conformational features similar to wt K63-linked diUbs upon specific recognition by TAB2 (Figure 2.8 c). Under the same experimental setup, natively K48-linked diUb could not be pulled down, in accordance to what has been shown before[338,339].

Next we tested if also the more compact sortase-generated K48-linked diUbs retain their ability to bind their designated UBDs. The UBA domain of proteasomal shuttling factor hHR23A-UBA2, sandwiches between the Ub moieties of K48-linked diUbs, making close contact with the I44 patches and the linker region that connects the two moieties, an event that requires dynamic opening of compact K48-linked chains[115]. Similarly to the Tab2-NZF pull-downs described above, we incubated K48-diUb(AT), K48-diUb(LAT) and natively linked K48-diUb together with GST-fused hHR23A-UBA2. Gratifyingly, western blot analysis showed that K48-diUb(AT) and K48-diUB(LAT) have pull-down efficiencies comparable to those of natively linked K48-linked diUb (Figure 2.9).

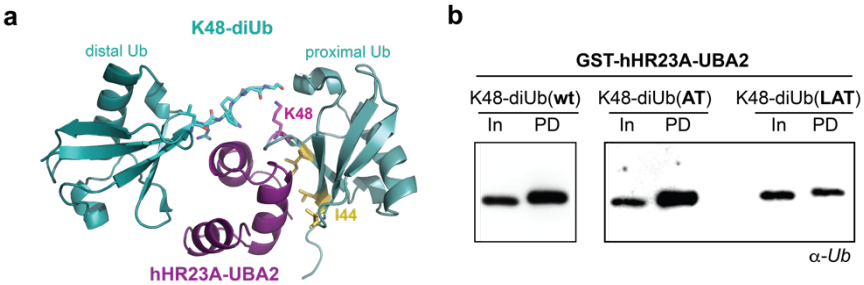


Figure 2.9 Structural and physiological integrity of sortase-generated K48-diUbs. a) NMR structure of the hHR23A UBA2 domain bound to K48-linked diUb (PDB: 1ZO6) shows multiple distinct interactions between proximal and distal Ub and hHR23A-UBA2[115]. **b)** Pull down assays with GST-fused hHR23A-UBA2 shows that sortase-generated K48-diUbs are still able to interact with hHR23A-UBA2. 1% of protein was loaded as input for pull-down experiments.

Many Ub binding proteins contain multiple UBDs connected by a defined spacer to confer binding selectivity for specific Ub chain types (see Chapter 1.2)[101,135]. For example, Rap80 targets BRCA1 to DNA damage foci through its tUIMs that are separated by a seven-residue helix to specifically recognize K63-linked chains.[340,341] As Rap80 tUIMs sense the linker length within K63-diUb and thereby the distance between the two connected Ubs, we reasoned that binding to Rap80 tUIMs may represent a thorough and valuable proof of principle target for binding our sortase-generated conjugates (Figure 2.10 a).

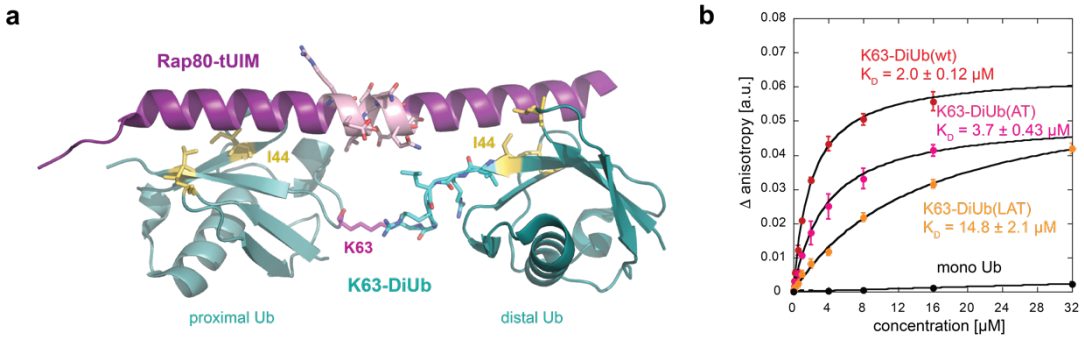


Figure 2.10 Interaction of sortase-generated K63-diUbs with Rap80-tUIM. a) Crystal structure of the Rap80 tandem Ub-interactin motif (tUIM) bound to K63-linked diUb (PDB: 3a1q).[341] b) To validate binding affinity of differently linked K63-diubiquitins towards Rap80-7A-tUIM we performed fluorescence anisotropy measurements with fluorescently labelled Rap80-7A-tUIM to determine the K_D of the interaction.[340] Experimental details can be found in Chapter 4.3.

In order to examine binding affinity in a more quantitative way, we expressed and purified a Rap80 construct bearing the tUIMs including a single reactive cysteine residue, which we used for fluorescence labeling using Atto488 maleimide. LC-MS confirmed quantitative labeling and excess of fluorophor was removed using SEC. Using fluorescence anisotropy, we were able to determine binding constants for K63-diUb(AT), K63-diUb(LAT), natively linked K63-diUb and Ub(wt). In accordance with literature values, we measured a K_D of 2.0 µM for K63-diUb(wt). Binding affinity of K63-diUb(AT) was slightly lower (K_D = 3.7 µM), and the sortase-generated K63-diUb with the LAT linker bound Rap80 with four-fold lower affinity (K_D = 14.8 µM), but still 60 times tighter than monoUb (Figure 2.10 b and Supporting Figure S2.10).

In conclusion, we could successfully show that sortase-generated diUbs, despite being resistant towards DUB cleavage, retain their binding affinity towards linkage-specific UBDs, a requirement for triggering various biological signaling pathways, including protein degradation, DNA damage repair and protein kinase activation.

2.7 Generating site-specific ubiquitylated PCNA *via* sortylation

Since sortylation is based on an enzymatic reaction which is performed using mild conditions, it is applicable to complex and non-refoldable proteins, including multi-domain proteins. For this reason, we decided to generate site-specifically ubiquitylated PCNA *via* sortylation. As already mentioned in Chapter 1, PCNA is a doughnut-shaped, homotrimeric protein that functions as a sliding clamp during DNA replication and enhances the processivity of DNA pol δ[342]. Studies have shown, that DNA lesions in the template strand lead to stalling of the replication fork, which triggers mono-ubiquitylation of PCNA at K164 (PCNA-Ub) *via* specific E2/E3 enzymes (Figure 2.11 a)[343]. PCNA-Ub in turn facilitates to recruitment of specialized

translesion synthesis (TLS) polymerases to the DNA damage site in order to traverse the damage[344]. Most TLS polymerases contain conserved UBDs for recognition of PCNA-Ub. Ubiquitylated PCNA with non-native disulphide or triazole linkages has been obtained *via* different chemical approaches[219,236]. To generate PCNA-Ub conjugates *via* sortylation, we introduced an amber codon at position 164 into the gene coding for PCNA and expressed it in the presence of AzGGK and its specific tRNA/synthetase pair. After affinity chromatography using Ni-NTA and cleavage of the CPD tag, we isolated several milligrams of PCNA-K164AzGGK per litre of culture, which was subsequently reduced to PCNA-K164GGK using 2DPBA. For site-specific ubiquitylation, we incubated PCNA-K164GGK with Ub(LAT) and Ub(AT) in the presence of Srt2A. To our delight we could observe ubiquitylation yields > 50% for GGK bearing PCNA, while we could not detect any unspecific ubiquitylation for PCNA displaying BocK (Figure 2.11 and Figure S2.11). Ubiquitylation of PCNA is endogenously reversed by the action of a DUB called Usp1 in combination with UAF1, which serves as an activator[345]. Analogously to our sortase-generated diUbs we wanted to investigate if PCNA ubiquitylated *via* sortylation is processed by the DUB complex Usp1/UAF1. To this end, we purified homogenously ubiquitylated PCNA *via* affinity chromatography using Ni-NTA followed by SEC. While sortase-generated PCNAK164-Ub(AT) was recalcitrant towards cleavage by Usp1/UAF1, natively ubiquitylated PCNA was completely cleaved within one hour under the same conditions (Figure 2.10 c). Natively ubiquitylated PCNA was provided by our collaboration partner Christian Biertümpfel.

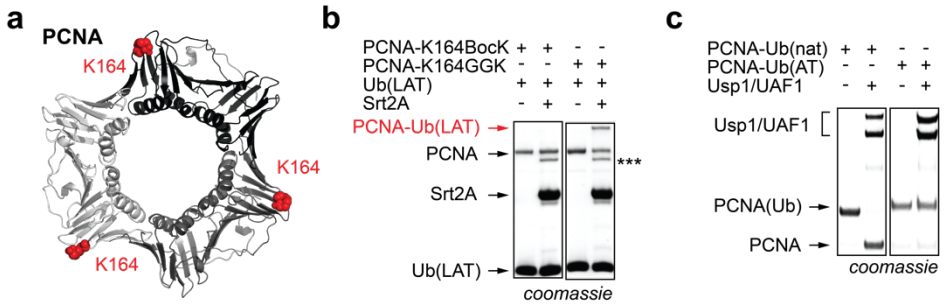


Figure 2.11 Ubiquitylation of PCNA. a) PCNA (PDB: 1axc) is a homotrimeric DNA-repair protein that is ubiquitylated at position K164. **b)** SDS-PAGE analysis of Srt2A-mediated formation of PCNA-Ub(LAT) conjugate. *denotes the generated thioester intermediate between Srt2A and Ub(AT). **c)** Incubation of natively ubiquitylated PCNA-Ub(nat) and Srt2A-generated PCNA-Ub(AT) with Usp1/UAF1 shows that Srt2A-generated mono-ubiquitylated PCNA is resistant to DUB-cleavage. Ubiquitylation reactions and deubiquitylation assays were carried out as described in Chapter 4.3 and sortylation yields are annotated in Supporting Figure S2.11).

2.8 Sortase-mediated SUMOylation

As described in Chapter 1, SUMO proteins belong to the family of Ubls and display the common ß-grasp fold with a flexible six-residue C-terminal tail and the characteristic GG motif that is exposed after proteolytic maturation[346]. We therefore envisioned that our sortylation approach might be expandable towards the site-specific formation of SUMO conjugates.

We cloned, expressed and purified SUMO1 bearing the Srt2A-compatible C-terminus LAQTGG (dubbed SUMO(AT)) and an N-terminal His6-TEV site in order to facilitate purification. The native C-terminal sequence of mature, proteolytically processed SUMO1 is QEQTGG. Introduction of the sortagging-motif into the SUMO1-C-terminus results in two

point mutations: Q92L and E93A (Figure 2.12 a). When we incubated sfGFPN150GGK with SUMO(AT) in the presence of Srt2A, SDS-PAGE analysis confirmed formation of the sfGFPN150SUMO(AT) conjugate. Analogously to our observations for site-specific ubiquitylation, introduction of an extra leucine spacer in the C-terminus of SUMO(AT) resulting in SUMO(LAT), led to more rapid SUMOylation with ca. 40% product formation within 15 minutes. Importantly, control experiments with sfGFP-BocK confirmed specificity of SUMOylation (Figure 2.12 b and Figure S2.12 a).

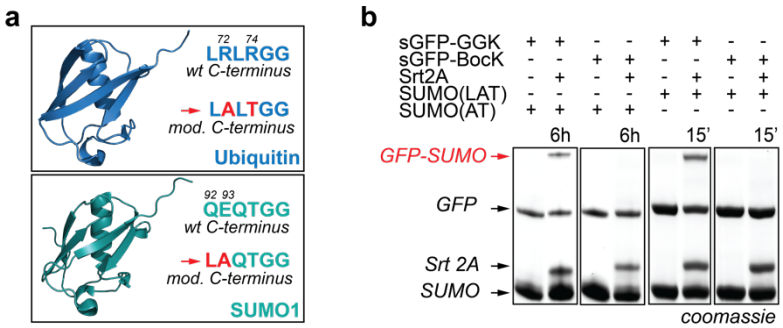


Figure 2.12 Site-specific SUMOylation of GGK-bearing proteins. a) Ub (PDB: 1ubq) and SUMO1 (PDB: 1y8r) show a similar globular ß-grasp fold with an unstructured C-terminus. The wild type C-terminal sequence of SUMO1 is QEQTGG. For recognition by Srt2A two point mutations are introduced: Q92L and E93A. **b)** SDS-PAGE analysis of Srt2A-catalyzed SUMOylation shows specificity for sfGFP-GGK.

2.9 Rational design of a Ca^{2+}-independent Srt2A mutant for sortylation in living *E. coli*

As already mentioned in Chapter 1, the activity of *S. aureus* SrtA-derived sortase mutants, including Srt5M and Srt2A, is strongly dependent on Ca^{2+}. The basis of this dependency is formed by the binding of Ca^{2+} to glutamate residues in the ß3/ß4 loop, which enhances substrate binding by stabilizing a closed conformation of the active site ß6/ß7 loop in *S. aureus* SrtA (Figure 2.13)[301]. We propose, that the strong Ca^{2+}-dependency may make it difficult to use Srt2A in conditions with low Ca^{2+} concentrations, including in the cytosol of living cells. While the SrtA superfamily shows a conserved active site among different Gram-positive bacteria, the amino acids residues in the ß3/ß4 loop that bind Ca^{2+} are however not conserved. In fact, *S. pyogenes* SrtA and *B. anthracis* SrtA are active in environments lacking Ca^{2+} [347,348].

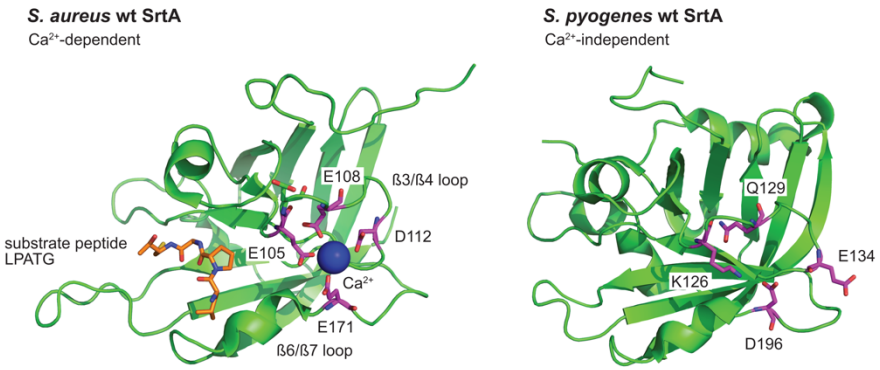


Figure 2.13 Comparison of different sortase A enzymes. Comparison of *S. aureus* wt SrtA (PDB: 2kid)[5] and S. *pyogenes* wt SrtA (PDB: 3fn5)[6]. *S. aureus* SrtA is strongly Ca^{2+}-dependent. Binding of Ca^{2+} to glutamate residues in the ß3/ß4 loop, distal to the active site enhances substrate binding by stabilizing a closed conformation of the active site ß6/ß7 loop. *S. pyogenes* SrtA is Ca^{2+}-independent and the ß3/ß4-loop and ß6/ß7-loop are kept in a closed conformation through hydrogen-bonding between K126 and D196.

Recently, a Ca^{2+}-independent Srt5M mutant (Srt7M) was published which was derived by substituting two glutamate residues within the ß3/ß4 loop with neutral or positively charged amino acids[299]. In order to render our sortylation approach independent of Ca^{2+} we envisioned to create a Ca^{2+}-independent variant of Srt2A. In analogy to the rational design of Srt7M we substituted D105 in Srt2A with a lysine residue, speculating that it might form a salt bridge with E171, thereby balancing the electrostatic repulsion, which might destabilize the closed conformation of the ß6/ß7 loop in the absence of Ca^{2+} [299]. Furthermore, we introduced an E108Q mutation to reduce the negative charge within this pocket (Figure S2.13). After cloning, expression and purification of the Srt2A variant bearing two point mutations D105K and E108Q (dubbed mSrt2A), we wanted to test its catalytic efficiency in our standard sortylation setup, a diUb formation assay. To guarantee Ca^{2+}-free conditions, we supplemented the reactions lacking $CaCl_2$ with 5 mM of Ca^{2+}-chelating agent ethylene glycol tetraacetic acid (EGTA). We combined UbK6GGK, Ub(LAT) and Srt2A/mSrt2A in the absence and presence of 5 mM $CaCl_2$ followed by SDS-PAGE analysis. To our delight, we could observe robust diUb formation for mSrt2A independent of Ca^{2+}, while Srt2A showed limited activity in the absence of Ca^{2+} (Figure 2.14).

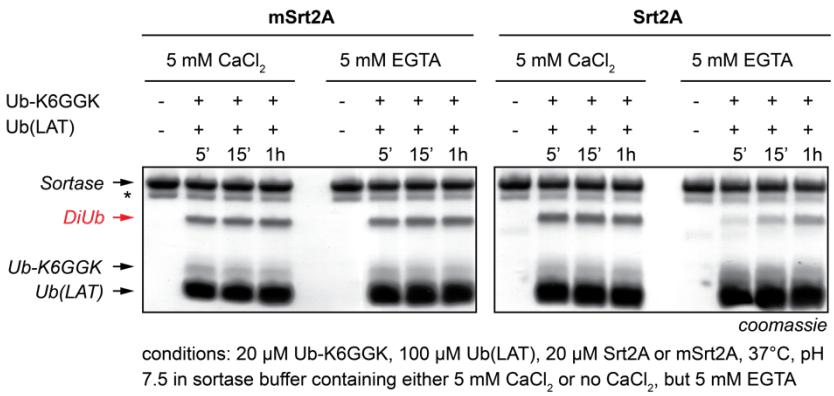
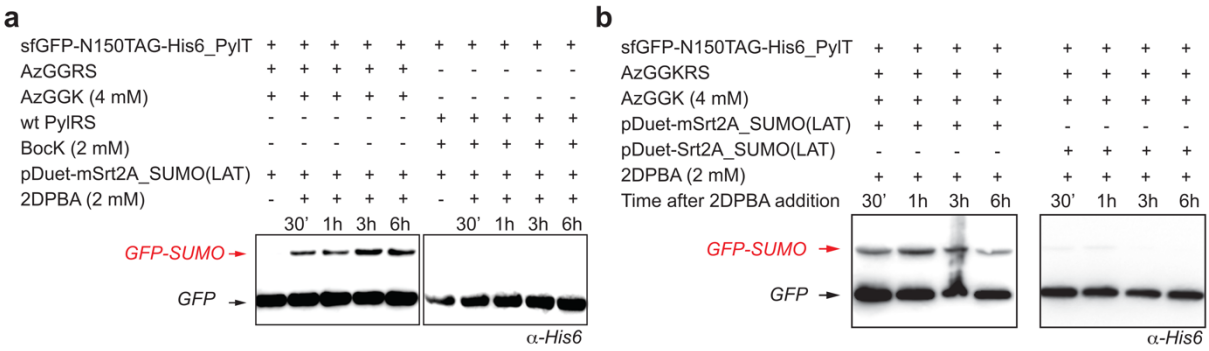


Figure 2.14 *In vitro* comparison of Srt2A and mSrt2A. Purified-Ub-K6GGK-His6 was incubated with an excess of Ub(LAT) in the presence of Srt2A or mSrt2A. The reactions were either supplemented with 5 mM $CaCl_2$ or 5 mM EGTA (no $CaCl_2$ was added). mSrt2A performed equally well in presence and absence of Ca^{2+}. Srt2A showed reduced efficiency in Ca^{2+}-free conditions. *denotes an impurity from Srt2A/mSrt2A

Our motivation for creating a Ca^{2+}-independent Srt2A mutant stemmed from our vision to expand sortylation towards the site-specific ubiquitylation and SUMOylation of proteins in living cells. In order to do so, we first set out to build Ub- and SUMO-conjugates in living bacteria. We designed and cloned a three-plasmid system, which allows the overexpression of alle needed components, namely the sfGFPN150TAG, AzGGKRS, PylT, Srt2A/mSrt2A and Ub(LAT) or SUMO(LAT). After triple transformation, we co-expressed all necessary components in *E. coli* for 24 hours followed by washing of the cells in order to remove residual AzGGK. Afterwards, we treated the cells with 2DPBA to induce reduction to sfGFPN150GGK and thereby trigger SUMOylation.



Figurer 2.15 Comparison of Srt2A and mSrt2A in *in vivo* SUMOylation experiments of GFP. *In vivo* SUMOylation of sfGFP-GGK in *E. coli.* sfGFP-AzGGK/BocK was co-expressed together with SUMO(LAT) and Srt2A or mSrt2A for 24 hours. After washing of cells to remove AzGGK, cells were treated with 2DPBA for indicated time points, washed again and analyzed by anti-His6 western blots. **a)** SUMOylation in living *E. coli* using mSrt2A is specific for sfGFP-GGK as analyzed by anti-His6 western blots. Time points refer to times after 2DPBA addition. **b)** Consistent with results for *in vitro* experiments conducted in the absence of Ca^{2+}, *in vivo* SUMOylation was much more effective with mSrt2A. Srt2A and mSrt2A-mediated SUMOylation in living *E. coli* was carried out as described in Chapter 4.3.

To our delight, we were able to observe formation of the sfGFP-SUMO(AT) conjugate already 30 min after 2DPBA addition in anti-His6 western blots. Importantly, SUMOylation did not take place when we expressed sfGFP-BocK instead of sfGFP-AzGGK or when we omitted 2DPBA, proving the specificity and inducibility of our approach (Figure 2.15 a).

In agreement with *in vitro* experiments, SUMOylation in living *E. coli* was only effective with mSrt2A, while we could barely detect SUMOylated GFP with Srt2A (Figure 2.15 b). By using the same setup, we were also able to attach Ub(LAT) and SUMO(LAT) to PCNA-K164AzGGK in a inducible fashion (Figure 2.16).

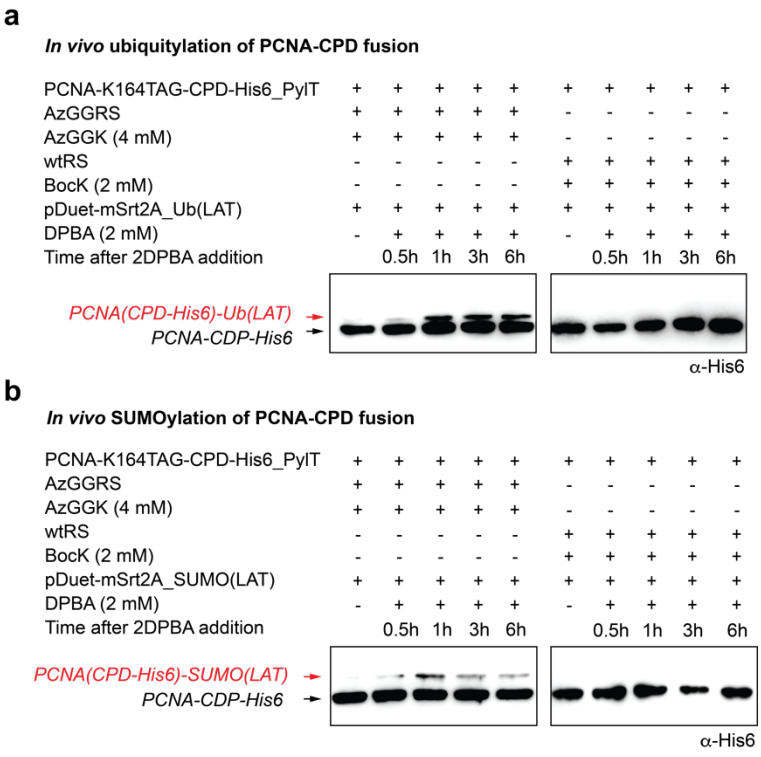


Figure 2.16 *In vivo* ubiquitylation and SUMOylation of PCNA. a) PCNA-K164AzGGK-CPD-His6 was co-expressed together with mSrt2A and Ub(LAT) for 24 hours. After washing of cells to remove AzGGK, cells were treated with 2DPBA for indicated time points, washed again and analyzed by anti-His6 western blots. Expression of PCNA-K164BocK-CPD-His6 shows that *in vivo* ubiquitylation is dependent on GGK-bearing proteins. b) same as (a) for SUMOylation. *In vivo* ubiquitylation/SUMOylation on PCNA-CPD-His6 in living *E. coli* was carried out as described in Chapter 4.3.

2.10 Inducible, site-specific ubiquitylation and SUMOylation of proteins in mammalian cells

Encouraged by successful sortase-mediated generation of Ub- and SUMO-conjugates in living *E. coli*, we set out to incorporate AzGGK into proteins in mammalian cells to test the possibility of ubiquitylating and SUMOylating proteins in living HEK293T in an inducible fashion. As sortase-mediated Ub-conjugates are resistant to isopeptidase activity of various DUBs, such an approach would provide an attractive tool for studying the effect of stable site-specific mono-ubiquitylation in physiological settings. In order to direct the site-specific incorporation of AzGGK in HEK293T cells, we transferred the mutations of AzGGKRS into a mammalian optimized *Mm*PylRS. Upon transient co-transfection of a plasmid encoding for sfGFPN150TAG/4xPylT and a plasmid encoding *Mm*AzGGKRS/4xPylT into HEK293T followed by cultivation in presence of 2 mM AzGGK, we were able to detect highly efficient incorporation of AzGGK into sfGFPN150TAG confirmed by anti His6 western blotting and fluorescence microscopy (Figure 2.17 a)[349].

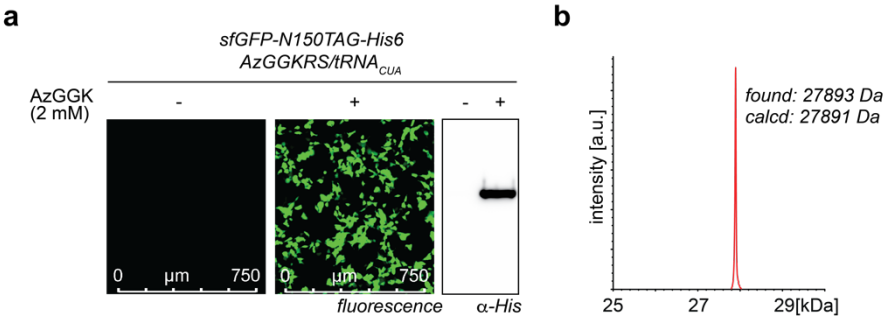


Figure 2.17 Incorporation of AzGGK into sfGFP in mammalian cells a) Fluorescence microscopy and anti-His6 western blot of HEK293T cells expressing sfGFP-N150TAG-His6 in the absence and presence of 2 mM AzGGK. **b)** MS-analysis of sfGFP-N150AzGGK-His6 purified from HEK193T cells.

Importantly, we could not observe misincorporation in the absence of AzGGK and LC-MS analysis of purified sfGFP-N150AzGGK-His confirmed its identity (Figure 2.17 b). We next tested if AzGGK-bearing proteins could be reduced with 2DPBA *in vivo*. 2DPBA and other benign triarylphosphine reagents have been used extensively for bioorthogonal Staudinger ligation/reduction in mammalian cells and 2DPBA shows good cell permeability[322,350]. Therefore, we expressed sfGFP-N150AzGGK-His in HEK293T cells for 48 hours, washed cells to remove AzGGK and treated them with 500 µM 2DPBA overnight followed by another washing step. Afterwards, cells were lysed in sortase reaction buffer by freeze-thaw and the lysates were treated with Ub(LAT) and different Ca^{2+}-dependent and independent sortase variants for one hour (Figure 2.18). Western blot analysis revealed specific formation of the sfGFP-Ub(LAT) conjugate but only when cells were treated with 2DPBA prior to lysis. Importantly, conjugate formation was visible for all sortase variants in presence of Ca^{2+}, while lysate supplemented with EGTA showed conjugate formation only for the Ca^{2+}-independent sortase variants mSrt2A and Srt7M (Figure 2.18).

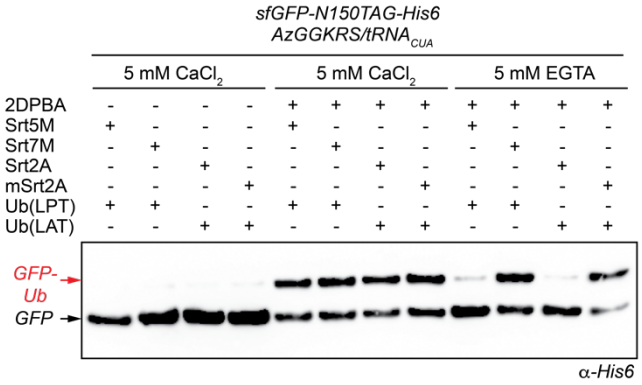


Figure 2.18 Sortase-mediated ubiquitylation and SUMOylation in mammalian cell lysates. Ubiquitylation and SUMOylation in HEK293T cell lysates: sfGFP-N150AzGGK-His6 was expressed for 48 hours in HEK293T cells. Cells were washed with AzGGK-free medium, incubated with 500 µM 2DPBA overnight, washed again and lysed by consecutive freeze-thaw cycles. Lysates were treated with 20 µM SrtA variant (either Srt5M, Srt7M, Srt2A or mSrt2A) and 100 µM Ub(LPT) or Ub(LAT) for one hour at 37 °C and analyzed by anti-His6 western blotting. In the absence of 2DPBA no ubiquitylation of sfGFP could be observed. All four tested SrtA variants were active in forming sfGFP-Ub conjugates in buffer containing 5 mM $CaCl_2$, while only the Ca^{2+}-independent mutant Srt7M and mSrt2A were able to ubiquitinate sfGFP-GGK in the presence of Ca^{2+}-chelating agent EGTA within one hour.

We next turned to the question of whether the sortase-mediated ubiquitylation and SUMOylation approach could be used in the cytosol of living HEK293T cells. In order to do so, we set up a four-plasmid system for transient expression in HEK293T cells. We designed plasmids for the expression of codon optimized variants of Ub(LAT), SUMO(LAT) and C-terminal Myc-tagged mSrt2A (all under CMV promoters). After transient transfection of the four plasmids (Ub(LAT) or SUMO(LAT), Srt2A, GFPN150TAG/4xPylT and AzGGKRS/4xPylT) into HEK293T cells and cultivation in presence of AzGGK for 36-48 hours, we washed the cells with AzGGK-free medium and treated them with 400 µM 2DPBA overnight. After additional washing steps to remove 2DPBA, cells were lysed and analyzed by anti-His6 western blotting. To our delight, for both Ub(LAT) and SUMO(LAT), distinct bands corresponding to ubiquitylated and SUMOylated sfGFP (with ca. 10 % conversion rate, respectively) were detected (Figure 2.19 a). Importantly, samples that were not treated with 2DPBA, did not yield the corresponding Ub- or SUMO-conjugates. Additionally, in cells where we overexpressed Ub(wt) and SUMO(wt) instead of the Ub/SUMO-mutants with sortase-compatible C-terminus, we were not able to detect any ubiquitylated and SUMOylated protein, proving the specificity and inducibility of our approach (Figure 2.19 a).

Interestingly, and in contrast to sortase-mediated transpeptidation in living *E. coli*, in HEK293T cells, mSrt2A and Srt2A lead to similar ubiquitylation and SUMOylation yields 16 hours after triggering the reaction through addition of 2DPBA (Figure 2.19 a). Next we set out to ubiquitylate/SUMOylate PCNA in live mammalian cells (Figure 2.19 b). Therefore, we co-expressed all components in HEK293T cells for 48 hours, and triggered ubiquitylation/SUMOylation *via* Staudinger reduction of PCNA-K164AzGGK with 2DPBA.

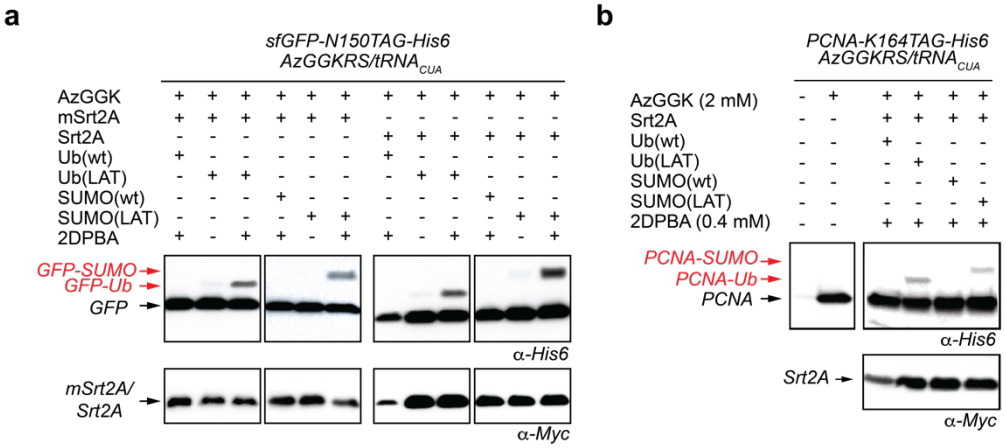
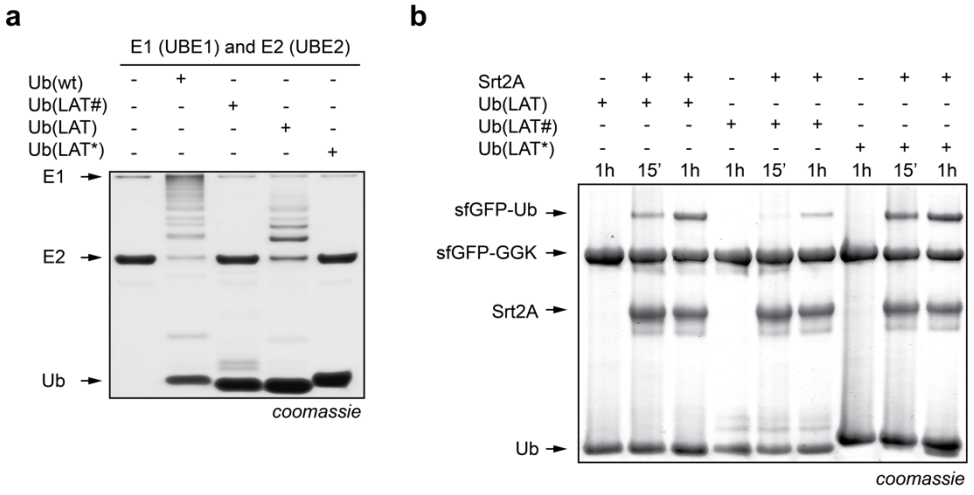


Figure 2.19 Sortase-mediated ubiquitylation and SUMOylation in mammalian cells. a) mSrt2A-mediated ubiquitylation and SUMOylation in live HEK293T cells shows specificity for Ub(LAT) and SUMO(LAT) as analyzed by anti-His6 western blots. Overexpression of SUMO(wt) and Ub(wt) did not lead to sfGFP-Ub or sfGFP-SUMO conjugates. Likewise, in the absence of 2DPBA (i.e. when sfGFP-AzGGK-His6 is not reduced to sfGFP-GGK-His6) mSrt2A-mediated ubiquitylation and SUMOylation did not take place, proving the specificity and inducibility of our approach. **b)** AzGGK-dependent expression of PCNA-K164TAG-His6 in HEK293T cells as analyzed by anti-His6 western blots. Srt2A-mediated ubiquitylation and SUMOylation of PCNA-K164AzGGK-His6 in the presence of 2DPBA is dependent on overexpression of Ub- and SUMO-variants bearing a sortagging-motif. Srt2A variants were tagged with a C-terminal Myc-tag. Experiments were carried out as described in Chapter 4.3. Sortylation yields were determined densitometrically using ImageJ software: Figure 2.14 b: sfGFP-Ub(LAT), Srt2A, 8%; sfGFP-SUMO(LAT), Srt2A, 9%; sfGFP-Ub(LAT), mSrt2A, 6%; sfGFP-SUMO(LAT), Srt2A, 7%.

Once more, western blot analysis 16 hours after addition of 2DPBA showed specific ubiquitylation (ca. 10 % conversion) and SUMOylation (ca 5 % conversion) of PCNA in the presence of Ub(LAT) or SUMO(LAT), but not when over-expressing Ub(wt) or SUMO(wt) (Figure 2.19 b). Collectively, these data establish sortylation as a powerful tool to generate site-specific isopeptide-linked Ub/SUMO-POI conjugates under physiological conditions in living mammalian cells. Importantly, the sortylation is triggerable through addition of a small molecule *via* a bioorthogonal Staudinger reaction conferring temporal resolution over the ubiquitylation/SUMOylation event and yields hydrolysis-resistant monoubiquitylated proteins.

Since the endogenous Ub conjugation (E1/E2/E3 enzymes) and deconjugation machinery is present in HEK293T cell we were interested to study if our sortylation approach interferes with endogenous processes. In order to validate the orthogonality of sortylation approach towards the endogenous ubiquitylation machinery, we subjected Ub(LAT) and Ub(wt) to an *in vitro* autoubiquitination assay using human E1 and E2 enzymes. We speculate that Ub(LAT) might still be a substrate for E1-activating enzymes. Therefore, we designed Ub variants containing the minimal LALTG sortase recognition motif, either directly as C-terminal sequence (dubbed Ub(LAT#) = Ub-LLALTG) or followed by a sequence of six amino acids that form a rigid helical linker[351] (dubbed Ub(LAT*) = Ub-LLALTGYEAAAK). We reasoned that by omitting glycine 76 and thereby mask the C-terminal GG motif of Ub, the Ub-variants should more unlikely serve as substrates of endogenous E1/E2/E3 enzymes. Interestingly, when we incubated E1 and E2 enzymes in the presence of Ub(wt) and to a lower extent also with Ub(LAT), we observed assembly of unattached polyubiquitin chains, while the newly created Ub variants Ub(LAT#) and Ub(LAT*) are completely orthogonal to E1 and E2 enzymes (Figure 2.20 a). With the two orthogonal Ub variants in hand we next set out to test them for their compability with sortylation. Sortase-mediated *in vitro* ubiquitylation assays of GGK-bearing sfGFP demonstrated that Ub(LAT*) is equally active as Ub(LAT), while Ub(LAT#) is not a good substrate for Srt2A and shows reduced formation of sfGFP-Ub conjugates (Figure 2.20 b).



Ub(wt) = Ub-LRLRGG; Ub(LAT) = Ub-LLALTGG; Ub(LAT#) = Ub-LLALTG; Ub(LAT) = Ub-LLALTGYEAAAK*

Figure 2.20 Autoubiquitylation and GFP ubiquitylation assay with different Ub variants. a) Autoubiquitylation assays with E1 and E2 enzymes using Ub variants with different C-termini shows that Ub(LAT*) and Ub(LAT#) are not substrates for E1 and E2 and thereby orthogonal to the endogenous ubiquitylation machinery. **b)** Sortase-mediated ubiquitylation of sfGFP-N150GGK using different Ub variants. While Ub(LAT*) showed similar activity as Ub(LAT) in forming sfGFP-Ub conjugates, Ub(LAT#) was not a good substrate of Srt2A. Reaction conditions: 20 µM sfGFP-N150GGK, 20 µM Srt2A, 100 µM Ub in sortase buffer, 37 °C for indicated time points. Experiments were carried out as described in Chapter 4.3.

Encouraged by our *in vitro* results, we next wanted to prove that Ub(LAT*) is selective and orthogonal also in living mammalian cells. To this end, we cloned the corresponding Ub constructs bearing N-terminal HA-epitope tags (HA-Ub(LAT*) and HA-Ub(wt)) and over-expressed them in HEK293T cells. While over-expression of HA-Ub(wt) lead to ubiquitylation of many endogenous proteins, ubiquitylation of the mammalian cell proteome using HA-Ub(LAT*) was barely detectable *via* anti-HA western blotting, both in the presence and absence of proteasomal inhibitor MG132 (Figure 2.21 a and Figure S2.14 a). This demonstrates that Ub(LAT*) is orthogonal towards the endogenous ubiquitylation machinery *in vivo*.

In order to investigate if Srt2A mediates unwanted off-target ubiquitylation by recognizing potential N-terminal glycine moieties within the cytoplasm, we co-expressed Srt2A together with HA-Ub(LAT*) or HA-Ub(wt) (Figure 2.21 a). Gratifyingly, we could not observe a difference between over-expressing HA-Ub(LAT*) alone or by co-expressing Srt2A, meaning that there are no suitable N-terminal glycine moieties present within the cytoplasm of HEK293T. Furthermore, we were interested if incorporation of AzGGK in response to endogenous amber codons results in the sortase-mediated ubiquitylation of endogenous proteins. Therefore, we co-expressed HA-Ub(LAT*), Srt2A and AzGGKRS/$tRNA_{CUA}$ in the presence of AzGGK, either with adding 2DPBA or not, in HEK293T cells. To our delight we could only observe low levels of proteome background ubiquitylation resulting from AzGGK incorporation compared to endogenous ubiquitylation events (Figure 2.21 b).

Lastly, we co-expressed all components needed for site-specific ubiquitylation of sfGFPN150TAG-His6 with our optimized HA-Ub(LAT*) construct. Anti HA western blotting revealed that the most prominent band corresponds to the GFP-Ub conjugate. In comparison with HA-Ub(LAT) our new HA-Ub(LAT*) construct was slightly more effective as in ubiquitylating GGK-bearing sfGFP, with up to 13% of the sfGFP (Figure 2.21 c and d).

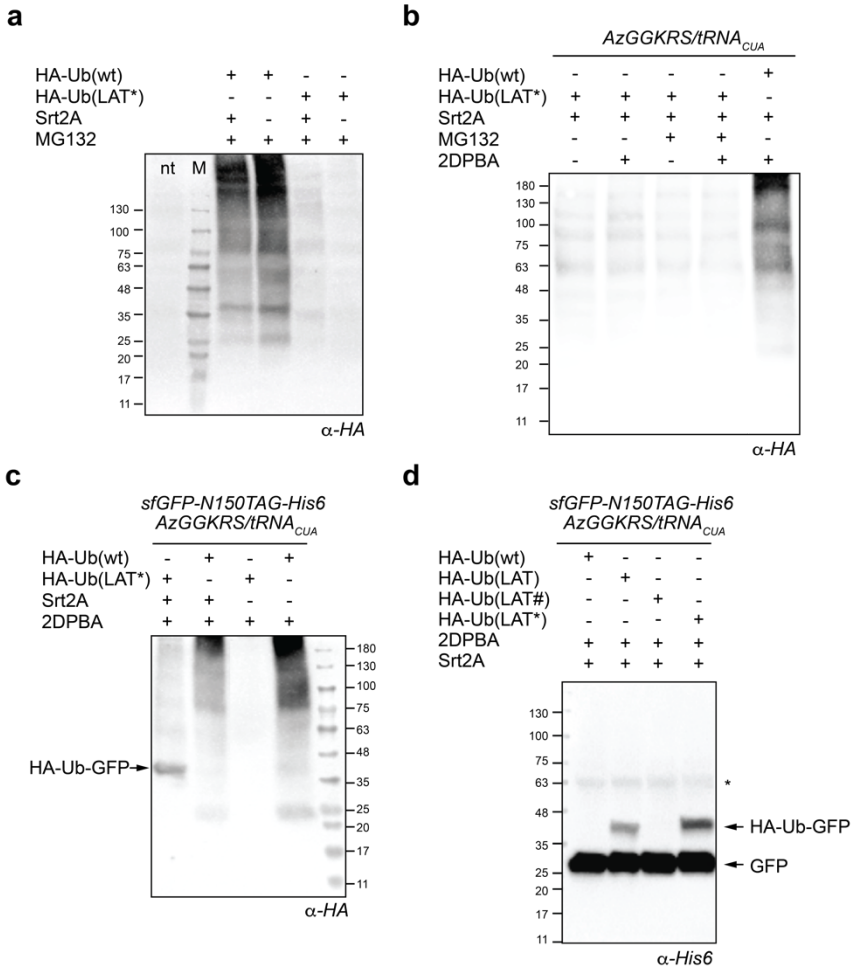


Figure 2.21 Background assays showing orthogonality and specificity of sortase-mediated ubiquitylation in living HEK293T cells. a) HEK293T cells were transfected with HA-Ub(wt) or HA-Ub(LAT*) in presence and absence of Srt2A and treated with MG132. Anti-HA western blotting confirmed orthogonality of the HA-Ub(LAT*) variant and Srt2A. **b)** HEK293T cells were transfected with AzGGKRS/tRNA$_{CUA}$, HA-Ub variant and Srt2A and treated with AzGGK and DPBA (or not, as indicated) to confirm lack of ubiquitylation of endogenous protein, when no amber-containing POI was over-expressed. **c)** Anti-HA western blotting confirming selective ubiquitylation of sfGFP-N150GGK with HA-Ub(LAT*). **d)** Anti-His6 western blotting confirmed selective and efficient ubiquitylation of sfGFP-N150GGK in the presence of Ub(LAT*), but not in the presence of Ub(LAT#). *denotes an unspecific band of the His-antibody. Experiments were carried out as described in Chapter 4.3.

Similarly, also for SUMOylation we could prove that omitting the C-terminal glycine and adding a further six amino acid linker to the C-terminus, thereby creating the SUMO variant SUMO(LAT*) (=SUMO-LLAQTGYEAAAK), resulted in selective SUMOylation of GGK-bearing proteins in an orthogonal fashion *in cellulo* (Figure S2.14 b and c). Taken together, these data establish and validate sortylation as a selective and orthogonal tool to site-specifically ubiquitylate and SUMOylate proteins in living mammalian cells. The majority of mammalian cell experiments was performed by Verena Bittl in the lab of Anja Bremm.

2.11 Discussion and Outlook

We have shown that site-specific incorporation of AzGGK into proteins in bacteria and mammalian cells *via* genetic code expansion and its reduction to GGK through a bioorthogonal Staudinger reaction allows sortase-mediated ubiquitylation and SUMOylation (Sortylation). To the best of our knowledge this represents the first approach where a site-specifically introduced UAA serves as a platform for a chemoenzymatic reaction, namely the sortase-mediated transpeptidation with Ub- and SUMO-mutants. The resulting sortase-generated Ub-conjugates display a native isopeptide-bond linking the C-terminal G76 to the ε-amino group of a chosen lysine in a target protein. To be recognized by sortase, we introduced two point mutations (R72A and R74T) into the unstructured Ub C-terminus. Importantly, amino acids L71 and L73 located at the C-terminus of Ub, that are essential for formation of the hydrophobic patch that centres around I36 (I36 patch) are preserved.

We demonstrate the generality of sortylation by producing differently linked diUbs. Even though sortase-generated diUbs are linked *via* a native isopeptide-bond, the introduced mutations (R72A and R74T) confer resistance to DUBs, providing a valuable tool to interrogate cell-signaling pathways and to assign Ub-specific proteins. Importantly, the introduced point mutations in the linker region of sortase-generated Ub-conjugates do not alter structural and dynamic properties of diUbs, as shown by MD simulations of K48- and K63-linked diUbs. Furthermore, we demonstrated that sortase-generated diUbs retain binding capabilities and affinities to their designated UBDs, thereby corroborating their physiological integrity and intact biological function in triggering various cellular signaling pathways.

Sortylation works under native, aqueous conditions and allows modification of complex, non-refoldable, multi-domain protein targets; an endeavour that is challenging using present chemical ubiquitylation approaches or protein targets where the corresponding E2/E3-enzymes are not known or show reduced activity/specificity *in vitro*. Additionally, we used sortylation to site-specifically ubiquitylate the homotrimeric DNA repair protein PCNA. Similarly to sortase-generated diUbs, the PCNA-Ub(AT) conjugate is refractory to cleavage by its specific DUB complex.

Furthermore, we have expanded sortylation towards SUMOylation by introducing point mutations into the unstructured C-terminus of SUMO1, namely Q92L and E93A. Besides ubiquitylation and SUMOylation, we envision that sortylation can potentially be expanded towards covalent modification of proteins with other Ubls (e.g. Nedd8, ISG15), processes that are much less understood than ubiquitylation and SUMOylation.

Importantly, we show that sortylation can be extended to site-specific mono-ubiquitylation and mono-SUMOylation under physiological conditions in living cells. The above described bacterial system might be advantageous for future applications to specifically produce ubiquitylated and SUMOylated eukaryotic proteins in the work horse *E. coli* in order to facilitate future crystallographic, biophysical and biochemical analyses of ubiquitylated and SUMOylated proteins[352]. Additionally, we envision, that our *in vivo* mammalian cell ubiquitylation and SUMOylation system can be expanded towards engineered cell lines that stably express sortase and/or modified Ub/SUMO. Additionally, also AzGGKRS/$tRNA_{CUA}$ may be stably incorporated into the mammalian cell genome to enable more efficient and homogenous incorporation of AzGGK into target proteins and their sortase-mediated ubiquitylation/SUMOylation in diverse mammalian cells[353].

As we have shown above, our approach relies on site-specific incorporation of AzGGK rather than GGK, ubiquitylation becomes triggerable by a small molecule, in principle enabling the study of temporal aspects of mono-ubiquitylation in live cells. We postulate, that the incorporation of a photocaged version of GGK would allow triggering monoubiquitylation in cells with high temporal and spatial resolution and would enable the study of ubiquitylation and its effect on substrate localization in real time.

In conclusion, we have described a chemoenzymatic approach to ubiquitylate and SUMOylate proteins *in vitro* and in live cells. Sortylation creates DUB-resistant Ub/SUMO conjugates with native isopeptide-linkages that retain structural and physiological integrity. We imagine this tool, which for the first time shows sortase-based ubiquitylation and SUMOylation in living cells, thereby creating an enzymatic approach that is orthogonal to highly specialized E1/E2/E3-enzymes, will have the potential to provide immediate impact to many Ub researchers.

CHAPTER 3

Expanding sortylation towards complex ubiquitin architectures using orthogonal sortases

The work in the following is based on the manuscript:

Fottner, M.; Weyh, M.; Schwarz, D.; Lang, K. Generating complex ubiquitin architectures by using genetic code expansion and orthogonal sortases. *To be submitted*

3. Expanding sortylation towards complex ubiquitin architectures using orthogonal sortases

3.1 Aim & Background

In the previous chapter we have described the development of a novel method for the site-specific ubiquitylation of target proteins both *in vitro* and in living cells (sortylation, Figure 3.1 a). With sortylation in hand we were able to (1) monoubiquitylate complex non-refoldable proteins and (2) recreate an orthogonal ubiquitylation system within mammalian cells. However, most of the information stored in the Ub code is encrypted in polymeric signals (as described in Chapter 1). Since sortylation is limited to monoubiquitylation of target proteins, we next set out to broaden its scope towards polymeric Ub architectures.

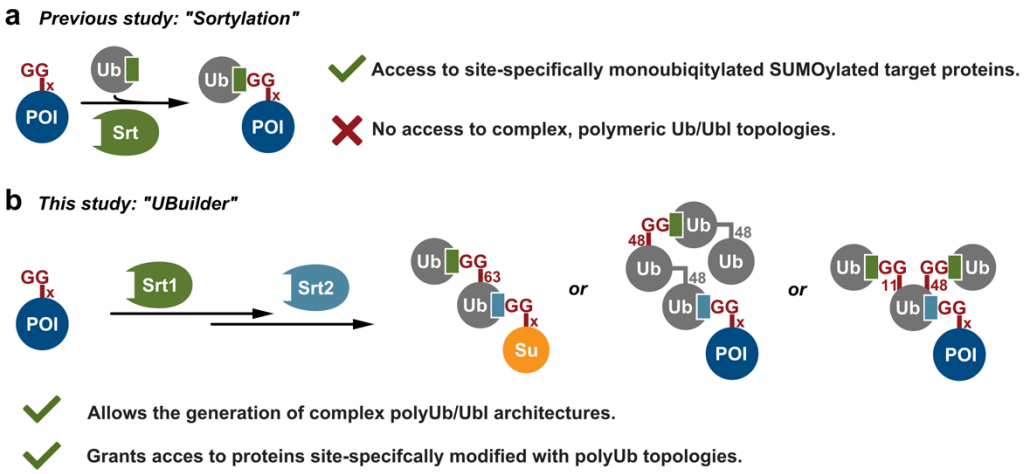


Figure 3.1 Schematic overview of Sortylation and UBuilder. a) Sortylation: The combination of genetic code expansion, sortase-mediated transpeptidation and Staudinger reduction has previously allowed the site-specific monoubiquitylation and monoSUMOylation of target proteins. **b)** UBuilder expands this concept by using a pair of orthogonal sortases which allows iterative use of sortase-mediated transpeptidation giving access to defined complex Ub/Ubl architectures that can be site-specifically installed on target proteins. Green and blue squares next to Ub indicate orthogonal sortase motifs at the C-terminus.

In this chapter, we propose 'UBuilder', a novel approach, which should be capable to expand sortylation towards complex polymeric Ub topologies (Figure 3.1 b). Since the transpeptidation reaction catalyzed by sortase is an inherently reversible reaction, as the sortase recognition motif is re-installed in the product, the sequential use of the same sortase for the creation of defined polymers impossible. To overcome this limitation, we envision to identify sortases that recognize different sortase motifs in an orthogonal manner (orthogonal sortases). With this orthogonal behavior, these sortases could be used one after another (iterative sortylation), thereby allowing the assembly of complex Ub architectures. Iterative sortylation should allow us to create Ub-SUMO2 hybrid chains displaying all feasible linkages as well as to investigate their interaction with the BRCA1 adaptor protein Rap80. Additionally, we envision to combine UBuilder with enzymatic Ub assembly to gain access to target proteins site-specifically modified with complex Ub architectures, such as K48-linked tetraUb and branched triUbs, signals well known to play fundamental roles in proteasomal degradation among other cellular signaling pathways[354,355]. We envision that the modular concept of UBuilder will be a valuable tool to further decipher the Ub code.

3. 2 Identification of orthogonal sortases suitable for sortylation

In order to broaden the scope of sortylation towards complex Ub topologies we set out to identify a pair of orthogonal sortase variants suitable for iterative sortylation. Over the last decade, SrtA and engineered variants thereof were extensively used in protein semisynthesis and protein labelling strategies *in vitro* and *in vivo*[287] (see Chapter 1). Sortase A enzymes with different substrate specificities, called orthogonal sortases, have facilitated dual labelling of protein N- and C-termini[356], circularization and functionalization of interferon[357], protein immobilization on gold surfaces[358] and modification of the capsids of bacteriophages[359]. These approaches, however, combine *S. aureus* and *S. pyogenes* SrtA enzymes whose orthogonality is based on the difference in acceptor nucleophiles: while *S. aureus* SrtA is specific for a polyglycine based acceptor nucleophile, *S. pyogenes* SrtA favours a polyalanine acceptor nucleophile. Consequently, *S. pyogenes* SrtA is not suitable for sortylation since site-specific incorporation of the unnatural amino acid AzGGK and its subsequent reduction to GGK yields a diglycine moiety as the acceptor nucleophile. As described above, directed evolution of *S. aureus* SrtA has recently resulted in variants with reprogrammed recognition motifs: Srt2A, which recognizes the LAXTG motif, and Srt4S, which recognizes the LPXSG motif [300]. Both of these variants are based on Srt5M which recognizes the archetypal LPXTG motif [297]. We therefore envisioned to implement Srt4S and Srt2A, which were reported to be orthogonal for their recognition motifs, into our approach.

In order to investigate orthogonality of sortases we designed a sortylation-tailored sortase orthogonality assay, which is based on the hydrolysis of diUb conjugates which are linked *via* different sortase motifs. In other words, incubation of a diUb, linked *via* a sortase motif, together with a sortase variant can have two different outcomes: (1) The diUb is hydrolysed by the sortase into the two monoUb moieties, meaning that the sortase is not orthogonal towards the sortase motif at the linkage site. (2) The diUb is stable upon prolonged exposition to the sortase variant, implying that the sortase is not able to recognize the sortase motif displayed between the two Ub moieties, meaning it is an orthogonal sortase. As we have described above[360], sortylation efficiency is highly increased when a single leucin is introduced as a spacer between the compact Ub β−grasp fold and the C-terminus bearing the sortase motif.

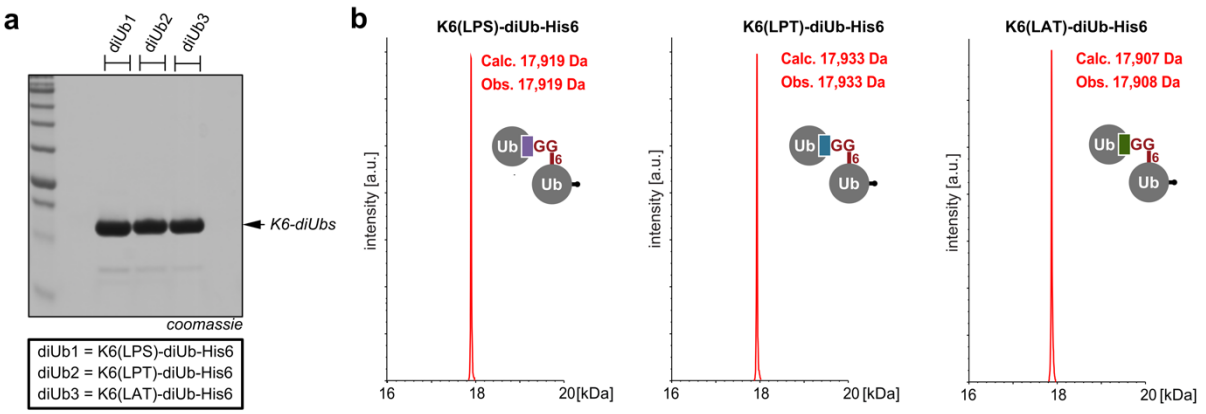


Figure 3.2 Overview of K6-diUbs linked *via* different sortase motifs. a) SDS-PAGE analysis of the purified K6-diUbs. **b)** LC-MS characterization of the different K6-diUbs.

In order to study the orthogonality of Srt5M, Srt4S and Srt2A we generated diUb conjugates linked *via* different sortase motifs for subsequent use in diUb hydrolysis assays. For increased sensitivity we included the leucin spacer at the linkage site that improves sortase accessibility and therefore also accelerates hydrolysis rates. K6-linked diUbs with different sortase motifs at their linkage site were created by incubating amber-suppression-derived UbK6GGK-His6 with Ub variants displaying different sortase motifs at their C-terminus (Ub-LLALTGG = Ub(LAT), Ub-LLPLSGG = Ub(LPS) and Ub-LLPLTGG = Ub(LPT)) in the presence of the corresponding sortase. Purification by affinity chromatography using Ni-NTA followed by SEC (Figure 3.2 a) yielded all three differently linked diUbs in milligram quantities and LC-MS analysis confirmed their identity (Figure 3.2 b).

Subsequently we performed diUb hydrolysis assays by incubating the three differently linked diUbs with the three different sortases (Figure 3.3). In these assays, incubation of the diUb together with the sortase which was used to create it serves as internal control, since each sortase should be able to hydrolyse its own motif at the linkage site (target motif hydrolysis). As expected, all sortase variants showed hydrolysis of the diUb with their own motif at the linkage site, which proves that our diUb hydrolysis assay works. Besides hydrolysis of the target motif by each corresponding sortase, we observed off-target motif hydrolysis when we incubated Srt5M with the diUb displaying the Srt4S motif and *vice versa*, meaning that these sortases are not orthogonal to each other.

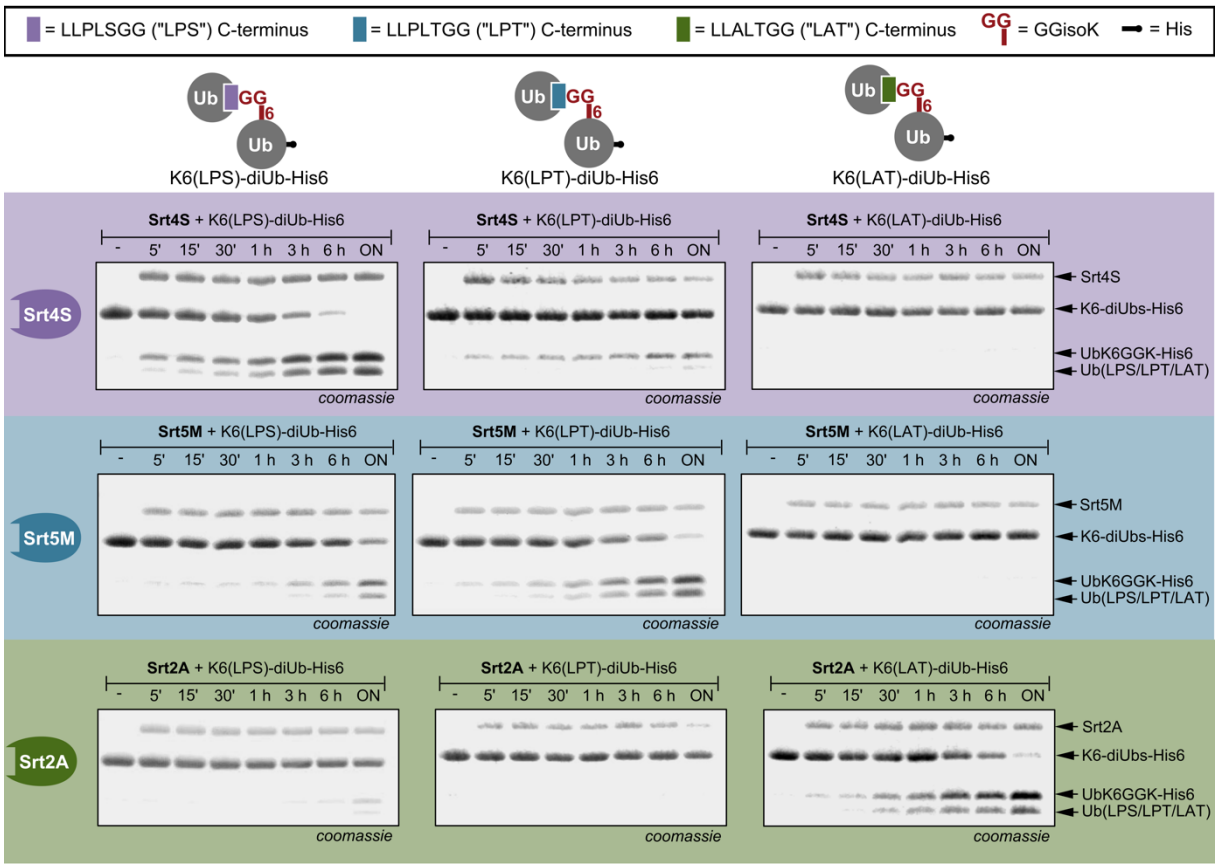


Figure 3.3 Identification of orthogonal sortases using diUb hydrolysis assays. Overview of the performed hydrolysis assays. Each sortase variant is incubated with all three differently linked K6-diUbs. Samples taken at the denoted time points were analyzed by SDS-PAGE. Target motif hydrolysis is already visible after 5 min for all variants (positive control). Srt4S shows off-target reactivity towards the LPLTG motif and is not able to hydrolyse the LALTG linked K6-diUb. Srt5M shows off-target reactivity towards the LPLSG motif and is not able to hydrolyse the LALTG motif. Srt2A shows weak off-target reactivity for the LPLSG motif and is not able to hydrolyse the LPLTG motif.

Additionally, we observed off-target motif hydrolysis for Srt2A towards the diUb bearing the Srt4S motif. Interestingly, Srt4S was not able to hydrolyse diUbs with the Srt2A motif, suggesting that Srt2A and Srt4S represent a monodirectional orthogonal sortase pair. This means, that one could use Srt2A and Srt4S in an orthogonal manner, as long as Srt2A is used before Srt4S. Excitingly, we did not observe any off-target hydrolysis for Srt2A towards diUbs displaying the Srt5M motif and *vice versa*, which means this pair is bidirectionally orthogonal. Taken together, our sortylation-tailored diUb hydrolysis assay enabled us to uncover a distinct sortase variant cross-reactivity network (Figure 3.4) which resulted in the identification of a monodirectional (Srt2A/Srt4S) and a bidirectional (Srt2A/Srt5M) pair of orthogonal sortases. Due to the superior orthogonality compared to the Srt2A/Srt4S pair, we decided to use the Srt2A/Srt5M pair for subsequent experiments.

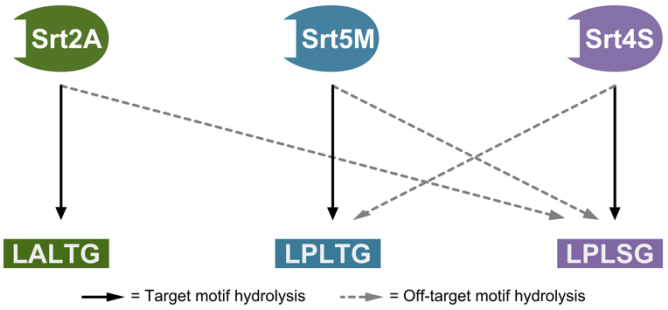


Figure 3.4 Overview of sortase orthogonality Schematic representation of the orthogonality network. All sortase variants used in this study exhibit on-target hydrolysis activity of a K6-linked diUb with their own motif at the linkage site (bold arrows). Srt2A as well as Srt5M show off-target reactivity for the LPLSG motif (dashed arrows) and Srt4S shows off-target reactivity towards the LPLTG motif. Taken together, we identified one bidirectional orthogonal pair (Srt2A+Srt5M) and one monodirectional orthogonal pair (Srt2A+Srt4S).

3.3 Generation of a mixed triUb

In a proof of concept experiment we tackled the challenge of creating a triUb with mixed linkages by using our newly identified pair of orthogonal sortases *via* iterative sortylation (Figure 3.5 a). Therefore, we first incubated a bifunctional Ub, which has GGK installed at K48 and an 'LPT' sortase motif at its C-terminus (UbK48GGK(LPT)-His6), together with Ub(LAT) in presence of Srt2A. As expected, this led to the formation of a K48-linked diUb displaying the 'LPT' sortase motif at the C-terminus of the proximal Ub moiety (K48(LAT)-diUb(LPT)-His6, Figure 3.5 b). Purification of the diUb *via* affinity chromatography using Ni-NTA and SEC resulted in milligram quantities of the diUb, whose identity was confirmed by LC-MS (Figure 3.5 c and d).

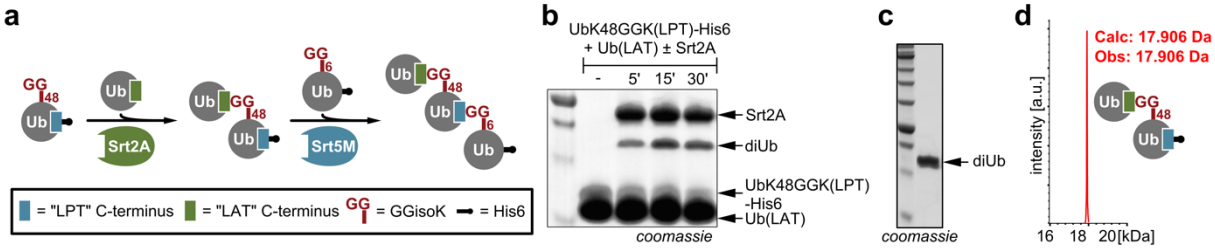


Figure 3.5 Generation of the diUb used for triUb formation. a) Scheme showing the use of orthogonal sortases to generate a mixed triUb. **b)** SDS-PAGE showing the generation of the diUb. Incubation of UbK48GGK(LPT)-His6 with Ub(LAT) in the presence of Srt2A leads to formation of K48(LAT)-diUb(LPT)-His6. **c)** SDS-PAGE displaying SEC purification of K48(LAT)-diUb(LPT)-His6. **d)** LC-MS characterization of K48(LAT)-diUb(LPT)-His6.

In order to generate the mixed triUb we set up a second transpeptidation reaction catalyzed by Srt5M. For this reason, we combined the purified K48(LAT)-diUb(LPT)-His6 with UbK6GGK in the presence of Srt5M. To our delight we observed time-dependent formation of the K48/K6-linked triUb (Figure 3.6 a).

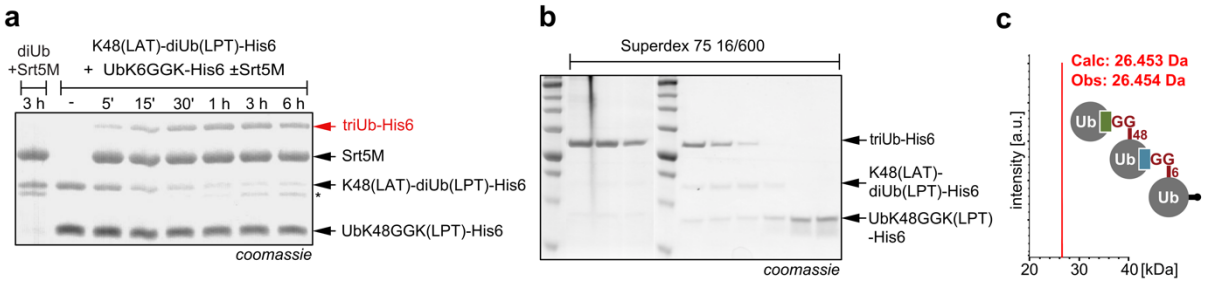


Figure 3.6 Generation and purification of the mixed triUb a) SDS-PAGE analysis of Srt5M-mediated triUb formation between K48(LAT)-diUb(LPT)-His6 and UbK6GGK-His6. The asterisk corresponds to diUb lacking the His6 tag. **b)** Characterization of the purified mixed triUb *via* SDS-PAGE c) Characterization of the purified mixed triUb *via* LC-MS.

Importantly, we could not detect any hydrolysis of the K48-linked diUb generated in the first sortase reaction with Srt2A upon incubation for 3 h with Srt5M, which confirms once more the orthogonal behaviour of Srt5M towards the 'LAT' linkage (Figure 3.6 a, first lane). Additionally, we observed a band appearing in the SDS-PAGE closely below the diUb, which results from Srt5M-dependent hydrolysis of the 'LPT' (as indicated by the asterisk, Figure 3.6 a). Finally, we were able to isolate milligram quantities of the mixed triUb from an upscaled reaction, by using affinity chromatography and SEC (Figure 3.6 b). Furthermore LC-MS analysis confirmed the identity of the mixed triUb (Figure 3.6 c).

3.4 Studying hybrid chains using UBuilder

Encouraged by the successful application of UBuilder to create a mixed triUb we next focused on biologically relevant hybrid chains. Ubiquitylation and SUMOylation play key roles in DNA damage repair. It is therefore of no surprise that the response to one of the most devastating types of DNA damage, double strand breaks (DSBs), is also mediated by Ub and SUMO. At DNA lesion sites, Ub- and SUMO-modifications of chromatin, presumably at histone H2AX, lead to the recruitment of Rap80, which is part of the BRCA1-A complex[361]. Rap80 contains both a tUIM and a SIM located at its N-terminus (Figure 3.7 a). While the tUIMs selectively recognize K63-linked Ub chains by closely binding the Ile44 patches of Ub[116] the interaction between the SIM and SUMO2 takes place in a defined binding groove between the α_1 helix and the β_2 strand of SUMO2 (Figure 3.7 b)[362]. The presence of a tUIM and a SIM in close proximity suggests the involvement of Ub/SUMO hybrid chains. Synthesis of Ub/SUMO hybrid chains is endogenously facilitated by a certain family of E3-ligases called SUMO-targeted-ubiquitin ligases (STUbls) that can charge Ub onto the N-terminus or onto one of the different lysine residues of SUMO. In case of DSB repair, the STUbl RNF4 is recruited to lesion sites by recognizing polySUMO2/3 chains on chromatin.

Mass spectrometry analysis has shown that RNF4 is capable of modifying the N-terminus and nearly all lysine residues of SUMO2 with Ub resulting in differently linked hybrid chains[363]. Studies with heterogeneously linked hybrid chains generated by incubating K63-diUb with RNF4 and SUMO2 revealed 80-fold tighter binding of hybrid chains to Rap80 (K_D

< 1 µM) than K63-diUb or SUMO2 alone (both ~ 20 µM)[140]. However, our understanding of hybrid K63-diUb-SUMO chains and their biological functions in Rap80-related DNA-damage repair is limited, since SUMO2 displays nine possible sites to which Ub can be conjugated (the N-terminus and eight lysine residues (K5, K7, K11, K21, K33, K35, K42 and K45) and it is not known which linkage leads to most efficient Rap80 binding.

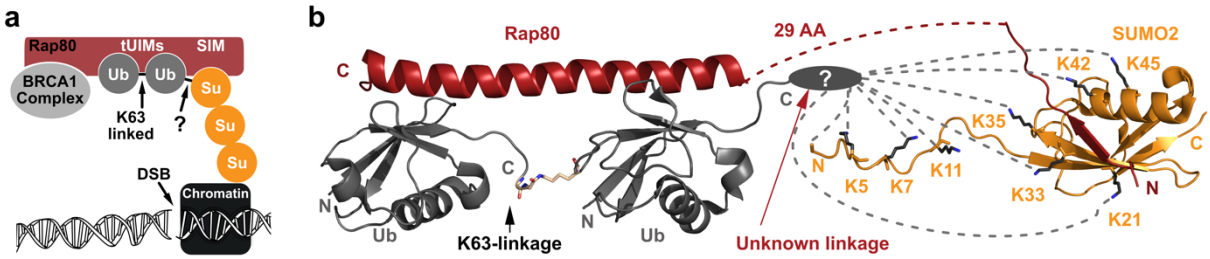


Figure 3.7 Hybrid chains – A structural insight. a) Model for hybrid chain-mediated recruitment of Rap80 to double strand breaks (DSBs). Chromatin, presumably histone H2Ax, is polySUMOylated close to DSBs which is followed by the monoubiquitylation of SUMO by SUMO-targeted ubiquitin ligases (STUbls). Extension of Ub to K63-linked polyUb chains leads to the recruitment of Rap80, mediated by the tandem Ub interacting motif (tUIM) and the SUMO interacting motif (SIM). Rap80 serves as an adaptor for subsequent recruitment of the BRCA1 complex. **b)** Structural insight into the interactions between Rap80 and a proposed hybrid chain. The tUIMs of Rap80 closely bind the Ile44 patches of the K63-linked diUb while the binding between SUMO2 and the SIM of Rap80 occurs in a defined binding grove of SUMO2. C-terminal tUIMs and N-terminal SIM are linked *via* a flexible unresolved region comprised of 29 aa. The linkage type between the C-terminus of the K63-linked diUb and SUMO2 is unknown. Grey doted lines depict possible linkage sites between the diUb and SUMO2. The scheme was created based on the NMR structure of the tUIMs of Rap80 bound to a K63-linked diUb (PDB: 2RR9)[116] and the NMR structure of SUMO2 bound to a peptide resembling the phosphorylated SIM of the Rap80 N-terminus (PDB: 2N9E)[362].

Accessing homogenously linked hybrid chains enzymatically *via* RNF4 is not possible due to its unspecific nature in an *in vitro* setting. Recent advances in complex protein total synthesis facilitated by iterative NCLs and assisted by removable solubilizing tags gained access to four differently linked hybrid chains (linked *via* the N-terminus, K11, K33 or K42)[364]. This strategy required however complex chemistries, mutation of the cysteine residues of SUMO2 and the resulting hybrid chains were not tested towards their binding abilities to Rap80.

We postulated that slight adaptions of the two-step procedure used for generating the mixed triUb will gain access to homogenously linked hybrid chains in a simple fashion (Figure 3.8 a). First, we incubated UbK63GGK(LPT)-His6 with Ub(LAT) in presence of Srt2A to produce a K63-linked diUb bearing the 'LAT' motif in the linker region between the two Ub moieties and the 'LPT'-motif at the C-terminus of the proximal Ub (K63(LAT)-diUb(LPT)-His6, Figure 3.8 b). After purification of the K63-diUb *via* SEC, we set up a second sortase reaction, catalyzed by Srt5M, with SUMO2 bearing an N-terminal diglycine moiety, which was obtained by TEV-cleavage of a N-terminally His6 tagged SUMO2. To our satisfaction, we observed robust hybrid chain formation within an hour of incubation (Figure 3.8 c). After upscaling and purification *via* SEC, SDS-PAGE and LC-MS analysis confirmed the purity and identity of the linearly linked K63-diUb-SUMO2 hybrid chain (Figure 3.8 d).

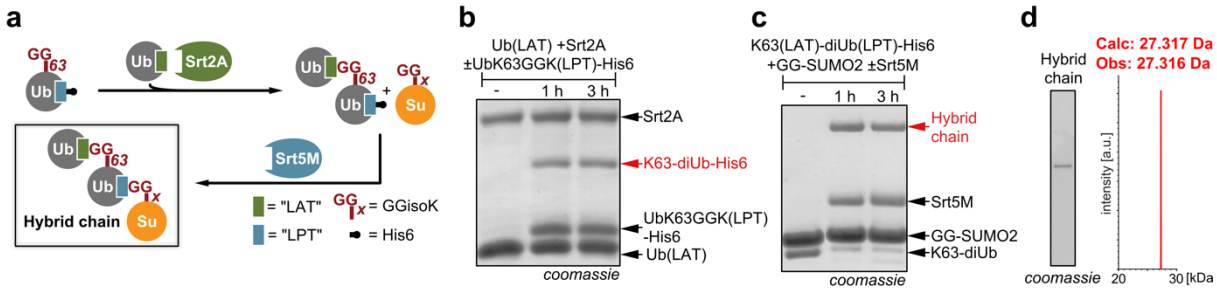


Figure 3.8 Generation of a linearly linked hybrid chain using UBuilder a) Schematic representation of the two-step strategy to generate hybrid chains using orthogonal sortases. **b)** SDS-PAGE showing the formation of the K63-linked diUb (K63(LAT)-diUb(LPT)-His6), which serves as basis for hybrid chain formation. **c)** Incubation of the K63(LAT)-diUb(LPT)-His6 with SUMO2 bearing a diglycine at its N-terminus in presence of Srt5M leads to the formation of a linear hybrid chain. **d)** Characterization of the purified linear hybrid chain *via* SDS-PAGE and LC-MS.

Next, we turned our interest into identifying the preferred linkage type between the K63-diUb and SUMO2 for Rap80 binding (Figure 3.7 b). By replacing each lysine codon of SUMO2 with a TAG codon, followed by site-specific incorporation of AzGGK at respective positions, protein purification and Staudinger reduction we gained access to SUMO2 bearing GGK either at K11, K21, K33, K35, K42 or K45 (TAG codons at positions K5 and K7 lead to insufficient yields of full-length SUMO2 and were therefore not included). Analogously to the generation of the linearly linked hybrid chain, we first generated a K63-linked diUb by combining UbK63GGK(PT) and Ub(AT) in presence of Srt2A. In this setup we omitted the leucine spacer in order to generate hybrid chains that mimic their natural counterparts as closely as possible. The resulting K63(AT)-diUb(PT)-His6 was then incubated with different SUMO2KxxGGK variants in the presence of Srt5M leading to the formation of all six different hybrid chains (Figure 3.9 a and Figure S3.1 b and c for complete gels). In order to prove the specificity of the sortase reaction and the isopeptidic nature of the derived hybrid chains we incubated the K63(AT)-diUb(PT)-His6 together with SUMO2 variants bearing the unreactive UAA BocK in place of GGK in the presence of Srt5M. Within this setup, we could not observe any hybrid chain formation but instead the hydrolysis of the sortase motif at the C-terminus of the K63-linked diUb due to the absence of an acceptor nucleophile (Figure 3.9 a and Figure S3.1 c for complete gels). After upscaling of the reaction mixtures and purification *via* SEC we were able to obtain seven differently linked hybrid chains, that only differ from their native counterparts by four mutations (Figure 3.9 b). These mutations, which we introduced to equip the C-termini of the Ub moieties with sortase motifs, reside within the flexible linker regions between the Ub-Ub and Ub-SUMO moieties.

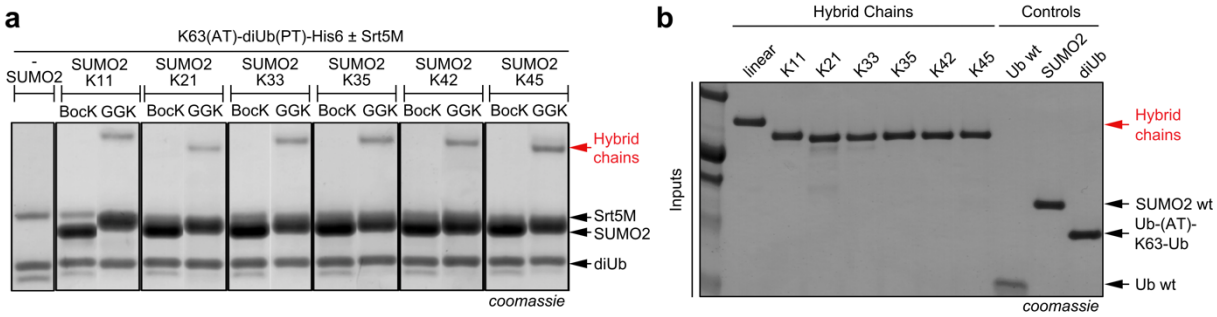


Figure 3.9 Generation of differently linked hybrid chains and pull-down experiments with Rap80. a) Incubating sortase-derived K63(AT)-diUb(PT)-His6 with SUMO2 (bearing GGK at depicted residues) in the presence of Srt5M leads to the formation of isopeptide-linked hybrid chains. Control experiments with SUMO2 bearing unreactive BocK at identical positions does not lead to hybrid chain formation. Incubation of the K63-linked diUb solely with Srt5M leads to the hydrolysis of the sortase recognition motif at the C-terminus of the diUb. **b)** Overview of proteins used for the pull-down experiments. 0.5 µg of each protein was applied to SDS-PAGE to verify equal concentration.

In order to evaluate the binding behaviour of our hybrid chains towards Rap80, we expressed and purified the Rap80 domain harbouring the tUIMs and the SIM (amino acids 1-137). The wt linker between the two UIMs was replaced by a seven-alanine linker, which is known to increase the affinity towards K63-diUb (see Chapter 2), and an N-terminal His6-tag was introduced. Affinity purification *via* Ni-NTA followed by SEC yielded milligram quantities of Rap80 (1-137) for pull-down (PD) experiments.

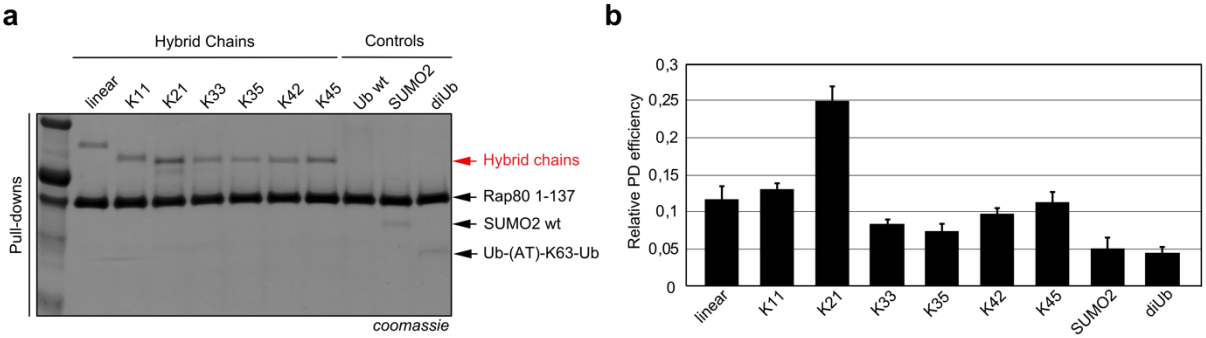


Figure 3.10 Pull-down assays for differently linked hybrid chains. a) Representative PD experiment. Ni-NTA charged with Rap80 was incubated with 3.5 µM of the denoted proteins, washed and applied to SDS-PAGE. Differently linked hybrid chains display different binding preferences towards Rap80. As expected, PDs with K63-diUb and SUMO2 as controls show reduced binding compared to hybrid chains, due to lack of avidity. **b)** Graph displaying the results of pull-down (PD) experiments using Rap80 (1-137) with a 7A linker immobilized on beads. After incubation of Rap80-beads with 3.5 µM of hybrid chains and different controls, SDS-PAGE analysis was performed. Gels of three independently performed pull-down experiments (n = 3) were analyzed using densitometry. Relative PD efficiencies were calculated by dividing the densitometry of hybrid chain PD by the densitometry of Rap80 (normalization against the amount of beads). Linear and K11-linked hybrid chains show similar Rap80 binding. Surprisingly, the K21-linked hybrid chain interacts twice as good with Rap80 compared to linear/K11-linked hybrid chains. K33-, K35- and K42-linked hybrid chains show decreased binding towards Rap80. The hybrid chain linked *via* K45 binds equally well as the linear/K11-linked chains. Controls with K63-linked diUb and SUMO2 show only weak interactions with Rap80. Center values are the median, error bars show s.d..

Incubation of Ni-NTA agarose-bead-bound Rap80 with 3.5 µM of each hybrid chain (and SUMO2 and K63(AT)-diUb as controls), followed by washing and SDS-PAGE analysis gave us a first insight into of different binding affinities of the hybrid chains (Figure 3.10 a). Densitometric analysis of three independent PD experiments (see all replicates Figure S.3.2) with the differently linked hybrid chains and the corresponding controls confirmed our first observation (Figure 3.10 b). Hybrid chains with the linkage site located in the flexible N-terminus of SUMO2, namely the linear- and the K11-linked hybrid chain, showed similar binding affinity towards Rap80. Most surprisingly, the hybrid chain where the K63-diUb is linked to SUMO2 *via* K21 shows two-fold increased affinity compared to the N-terminally linked one. These observations may suggest a "backside" binding mode where the 29 amino acid residue long unstructured flexible linker of Rap80 adopts a kinked conformation to allow binding of the closely tethered, sterically restrained diUb-SUMO2 conjugate (Figure 3.11).

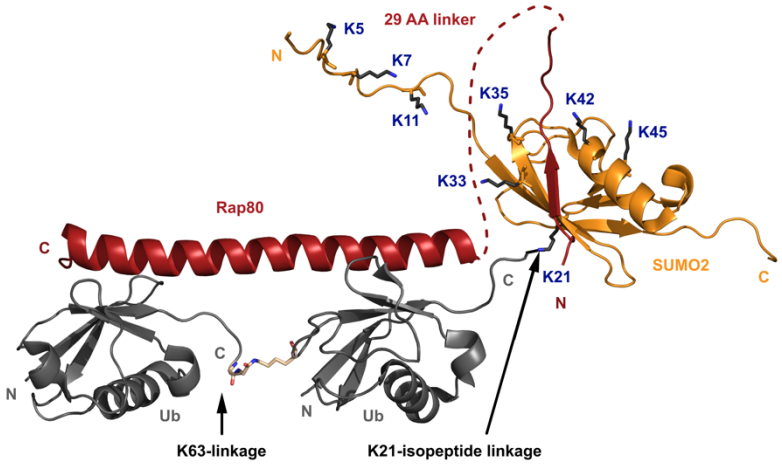


Figure 3.11 Proposed model of a K21-mediated backside binding mode. Schematic representation of a K21-linked hybrid chains bound to Rap80. The flexible 29-aa linker is kinked to facilitate the binding of a K21-linked isopeptide bond. The scheme was created based on the NMR structure of the tUIMs of Rap80 bound to a K63-linked diUb (PDB: 2RR9)[116] and the NMR structure of SUMO2 bound to a peptide resembling the phosphorylated SIM of the Rap80 N-terminus (PDB: 2N9E)[362].

As expected, the K33-, K35- and K42-linked hybrid chains show decreased binding compared to the linear hybrid chain since these residues are located either in the α_1 helix or in the β_2 strand of SUMO2 where the interaction of SUMO2 with the SIM of Rap80 takes place (Figure 3.12). Linkages at these positions most likely sterically impair binding of the SIM of Rap80 to SUMO2. This is in consent with NMR studies of the interaction between SUMO2 and a peptide resembling the phosphorylated SIM of Rap80 which shows that K33 (*via* electrostatic interactions), K35 (*via* interaction with Ser/pSer residues 44 and 46 of Rap80) and K42 (*via* interaction with D45 of Rap80) are essential for the binding of the SIM to SUMO2. In contrast, the K45-linked hybrid chain binds nearly as efficiently as the linear linked chain indicating that a linkage at this position no longer impairs the SIM binding. Control experiments with SUMO2 and K63(AT)-diUb showed only weak binding compared to hybrid chains, underlying the avidity effect.

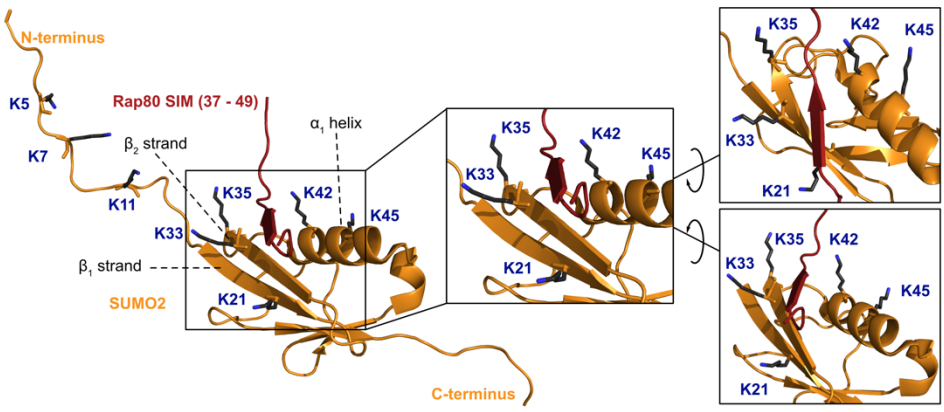


Figure 3.12 Structural insight into the Rap80-SUMO2 interaction Scheme focusing on the interaction between the N-terminal SIM of Rap80 and SUMO2. The flexible N-terminal region of SUMO2 harboring lysine residues K5, K7 and K11 is not involved in binding of the SIM of Rap80. Interaction mainly occurs in a defined hydrophobic binding groove located between the α_1 helix and the β_2 strand of the SUMO2 core. Lysine residues K33 (electrostatic interactions), K35 (interaction with Ser/pSer residues 44 and 46 of Rap80) and K42 (interaction with D45 of Rap80) were shown to be essential for the interaction with the SIM of Rap80. In contrast, lysine residues K21 (located in the β_1 strand) and K45 of the SUMO2 core do not seem to be involved in the binding of the SIM of Rap80. The scheme was created based on the NMR structure of the tUIMs of Rap80 bound to a K63-linked diUb (PDB: 2RR9)[116] and the NMR structure of SUMO2 bound to a peptide resembling the phosphorylated SIM of the Rap80 N-terminus (PDB: 2N9E)[362].

3.5 Using UBuilder to generate and charge complex Ub architectures onto target proteins

After exemplifying the ability of UBuilder to create homogenously linked hybrid chains in a straightforward, robust fashion and proving its convenience to tackle biological questions, we decided to expand the concept of UBuilder even further. As mentioned above, the Ub-code conveys information by using complex Ub architectures site-specifically installed on target proteins. The unarguably best studied example employing these topologies is the ubiquitin-proteasome system, where ubiquitylation marks proteins for degradation. The archetypal signal for proteasomal degradation is the attachment of a K48-linked tetraUb chain to a lysine residue of a target protein[354]. However, recent studies showed, that Ub chains branched *via* K11 and K48 can also facilitate proteasomal degradation[365]. Structural studies suggest that the K11/K48 branching offers a unique interdomain interface, which is recognized by the proteasomal subunit Rpn1[366]. The study of proteasomal recognition and degradation of these polymeric conjugates is of utmost importance, due to its involvement in diseases such as cancer, and requires access to proteins site-specifically and homogenously modified with defined Ub chains. Advances in protein total synthesis facilitated by the iterative use of NCLs resulted in the generation of a K48-linked tetraUb[277] and just recently in the first total synthesis of a tetraubiquitylated protein, being α-globulin[367]. Since total synthesis is limited to small, refoldable proteins, we envisioned to adapt UBuilder to gain access to tetraubiquitylated proteins without any limitation in size of the target protein or its refolding capability.

We decided to combine enzymatic Ub chain assembly and UBuilder to provide a flexible platform for generating and charging Ub chains site-specifically to target proteins. In order to create a POI site-specifically charged with a K48-linked tetraUb we designed an elaborated assembly scheme, where we placed native Ub linkages and linkages containing sortase motifs at the isopeptidic positions of the conjugate (Figure 3.13).

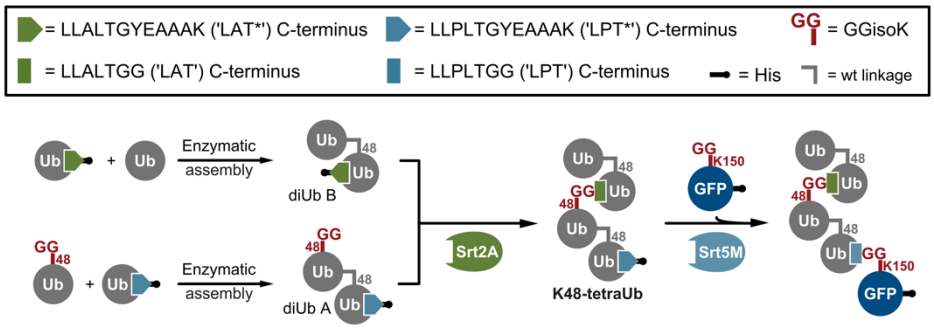


Figure 3.13 Schematic representation of the generation of a site-specifically K48-tetraubiquitylated GFP. Enzymatic assembly of carefully designed monoUb building blocks results in the K48-linked diUbs A and B. In a first sortase reaction employing Srt2A the K48-linked tetraUb is formed from the diUb building blocks. Subsequently, the K48-linked tetraUb is site-specifically charged onto GFP bearing GGK at position 150 by using Srt5M.

Our final scheme encompasses three steps: (1) In the first step we will use four carefully designed monoUb building blocks, which will be assembled in two enzymatic reactions into two natively linked K48-diUbs (diUb A and B, Figure 3.13). Each of the resulting diUbs displays a sortase motif at the C-terminus of the proximal Ub and diUb A additionally bears GGK at K48 of the distal moiety (Figure 3.13). (2) These diUb building blocks will then be assembled in a first sortase reaction mediated by Srt2A to form a tetraUb, which displays the

'LAT' motifs between Ub moiety two and three of the chain and the 'LPT' motif at the proximal Ub (Figure 3.13). (3) In the final step we will use Srt5M in a second transpeptidation reaction to charge the K48-linked tetraUb site-specifically onto a GFP bearing GGK at position 150.

Since Ub(LAT) is still recognized by the E1/E2/E3 machinery (see chapter 2.9) we had to adjust our monoUb building blocks for the enzymatic assembly of K48-linked diUb A and diUb B in order to prevent uncontrolled polymerization. Analogously to Chapter 2.9, we therefore used Ub variants which lack glycine 76 and display a rigid linker sequence (YEAAAK) at the C-terminus (Ub(LAT*) and Ub(LPT*)). We tested all four monoUb building blocks for their ability to form K48-linked diUbs in an E1/E2 assay (using the E1 enzyme UBE1 and the E2 enzyme CDC34, which specifically forms K48-linked chains). To our delight only Ub bearing the wt C-terminus is effectively polymerized, while Ub(LPT*) and Ub(LAT*) polymerization is inhibited *via* C-terminal blockage. UbK48GGK-His6 is also polymerization-incompetent since GGK is blocking the conjugation site K48 (Figure 3.14).

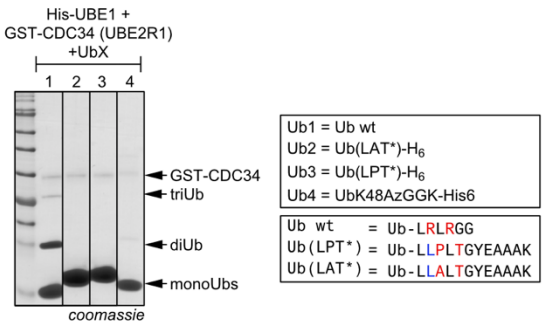


Figure 3.14 E1/E2 orthogonality assay for the generation of K48-linked diUbs. The ubiquitin activating enzyme (UBE1 (E1)) and the ubiquitin conjugating enzyme (CDC34 (E2)) are incubated with different Ub variants. Ub wt (1) shows di/tri-Ub formation because both the C-terminus and the lysine residues are conjugation competent. In contrast, Ub2/3 are not able to form a di/triUb because the C-terminus is blocked. Ub4 bears GGK at K48 making it conjugation incompetent for enzymatic assembly of di/triUbs *via* K48. Nevertheless, little diUb formation is observed for Ub4, which could stem from lysine misincorporation at K48 or lack of specificity of the E2 enzyme CDC34.

Enzymatic assembly of diUb A (a K48-linked diUb displaying GGK at K48 of the distal Ub and a 'LPT*' sortase recognition motif at the C-terminus of the proximal Ub) was achieved by incubating UbK48GGK-His6 with Ub(LPT*) in presence of UBE1 and CDC34 (Figure 3.15 a and b). After affinity chromatography using Ni-NTA and SEC purification (Figure 3.15 c) several milligrams of diUb A were obtained. However, diUb A showed a minor impurity which could be identified *via* LC-MS (Figure 3.15 d) as remaining's of Ub(LPT*). It is important to mention, that this impurity cannot interfere in the next step, the tetraUb assembly, since this transpeptidation reaction is catalyzed by Srt2A.

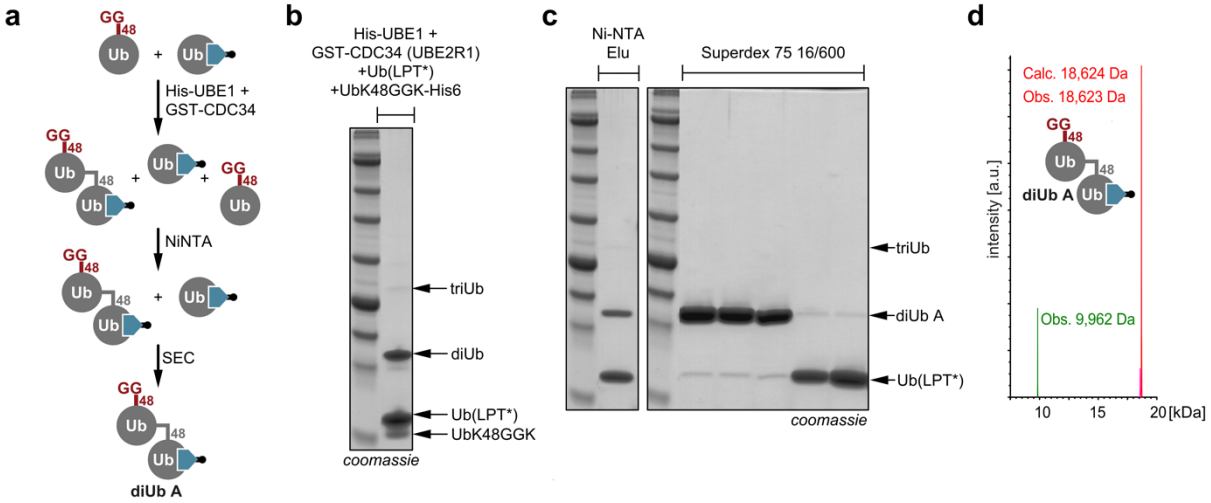


Figure 3.15 Enzymatic assembly of diUb A for tetraUb formation using E1/E2 enzymes. a) Schematic overview of enzymatic assembly and purification of diUb A. **b)** Overnight incubation of Ub(LPT)* and UbK48GGK in presence of E1- and E2-enzymes leads to formation of diUb A. **c)** Purification of diUb A using Ni-NTA chromatography followed by SEC. Final diUb A contains residual Ub(LPT)* as impurity, which is negligible since it cannot interfere in the subsequent sortase reaction using Srt2A. **d)** LC-MS characterization of diUb A.

DiUb B (a K48-linked diUb with an 'LAT*' sortase recognition motif at the C-terminus of the proximal Ub moiety) was created in a similar fashion using wildtype Ub and Ub-(LAT*) (Figure 3.16 a). In contrast to the assembly of diUb A, the assembly reaction of diUb B contains Ub wt, which is the only polymerization competent building block used in diUb assembly. The presence of Ub wt leads to two different diUb species as reaction products: (1) native K48-linked diUb and (2) K48-linked diUb with the 'LAT*' motif at the C-terminus of the proximal Ub. Both of these diUbs, in contrast to diUb A, bear a free K48 moiety which additionally leads to the formation of K48-linked tri and tetraUbs (Figure 3.16 b). For successful isolation of the target diUb (diUb B), we used affinity chromatography *via* Ni-NTA followed by SEC (Figure 3.16 c). To our delight we were able to obtain several milligrams of pure diUb B whose identity was also confirmed by LC-MS (Figure 3.16 d).

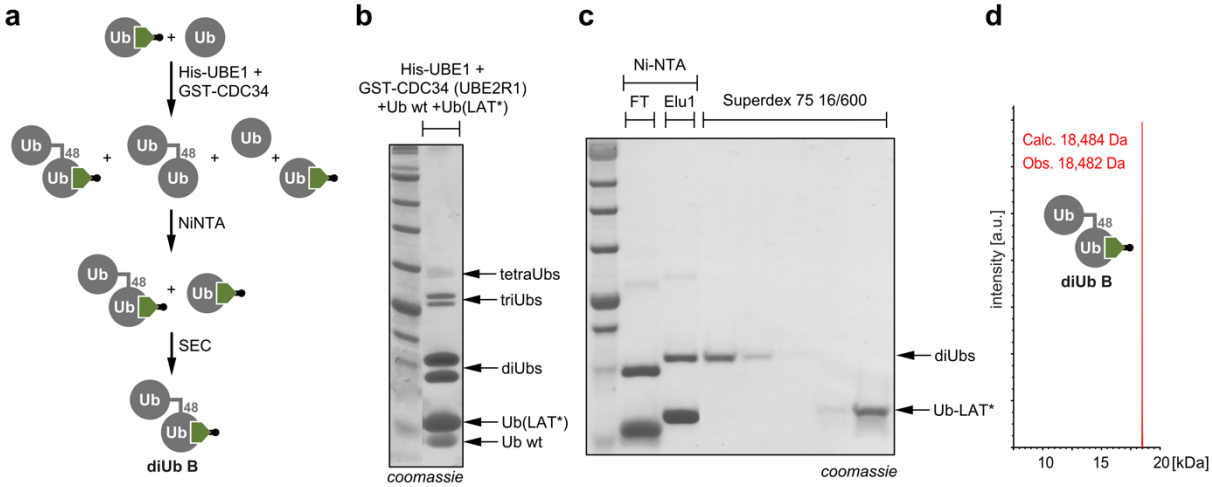


Figure 3.16 Enzymatic assembly of diUb B for tetraUb formation using E1/E2 enzymes a) Schematic overview of enzymatic assembly and purification of diUb B. **b)** Overnight incubation of Ub wt and Ub(LAT*) in presence of E1- and E2-enzymes leads to formation of K48-linked di- and triUb species. **c)** Purification of diUb B using Ni-NTA chromatography followed by size-exclusion chromatography (SEC). **d)** LC-MS characterization of diUb B.

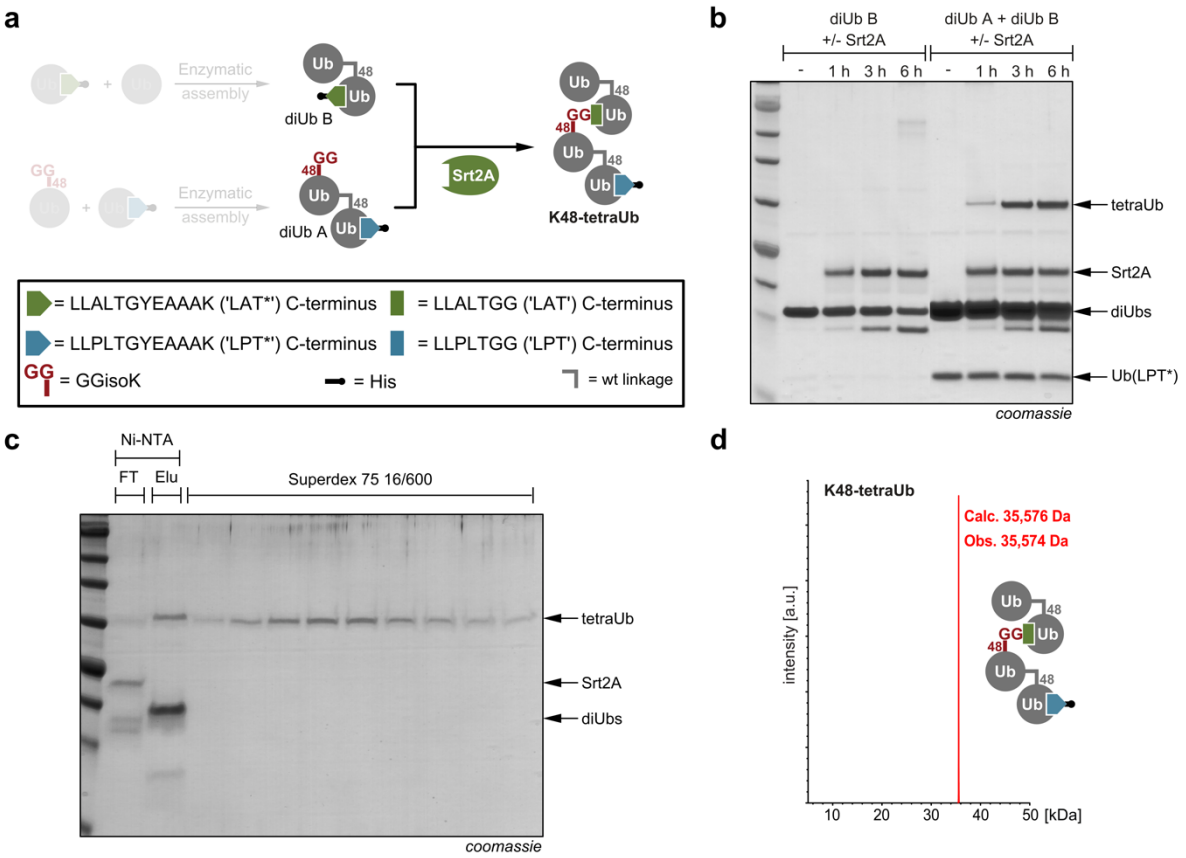


Figure 3.17 Sortase-mediated generation of K48-linked tetraUb. a) Schematic representation of sortase-mediated tetraUb formation using beforehand generated diUb A and diUb B. **b)** Incubation of diUb A and diUb B in presence of Srt2A leads to the formation of K48 linked tetraUb. No tetraUb formation is observed in the absence of diUb B. The band appearing below the diUb B band is the hydrolysis product of the sortase motif in the absence of an acceptor (e.g. diUb A). **c)** Purification of K48-linked tetraUb using Ni-NTA chromatography followed by size-exclusion chromatography (SEC). **d)** LC-MS characterization of K48-linked tetraUb.

The successful generation and characterization of diUb A and diUb B forms the basis of the second step of our assembly scheme, the first transpeptidation reaction. In order to generate the K48-tetraUb, we combined diUb A and diUb B with Srt2A. This led to the time-dependent formation the K48-tetraUb (Figure 3.17 a and b). Importantly, when we omitted the acceptor diUb (diUb A), tetraUb assembly was abolished, and we could only observe hydrolysis of the sortase motif at the C-terminus of the proximal Ub moiety of diUb B (Figure 3.17 b). Finally, we isolated the K48-linked tetraUb from an upscaled reaction mixture using affinity chromatography and SEC (Figure 3.17 c). LC-MS analysis confirmed the identity of K48-tetraUb (Figure 3.17 d).

With the K48-tetraUb in hand, we tackled the last step of our assembly scheme: The second transpeptidation reaction catalyzed by Srt5M (Figure 3.18 a). Therefore, we combined sfGFPN150GGK-His6 with the K48-tetraUb in presence of the orthogonal sortase Srt5M. To our delight SDS-PAGE analysis confirmed successful charging of the K48-tetraUb chain onto the POI sfGFP (Figure 3.18 b). Taken together, our multistep assembly scheme involving the enzymatic assembly of two diUbs followed by two transpeptidation reactions with our pair of orthogonal sortases led to the successful formation of a site-specifically K48-linked tetraubiquitylated GFP - an endeavour which has so far only been achieved with complex chemical protein synthesis.

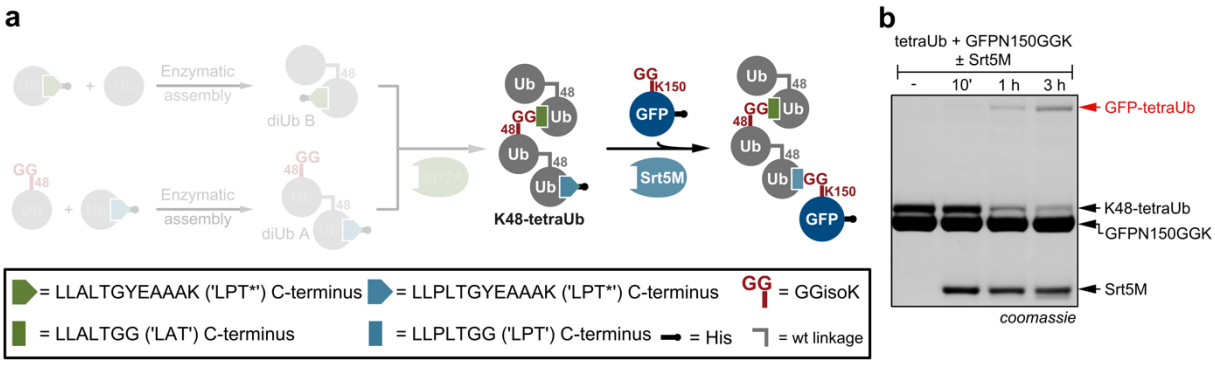


Figure 3.18 Site-specific tetraubiquitylation of sfGFP. a) Schematic representation of the second sortase-mediated transpeptidations to charge the K48-linked tetraUb onto sfGFP. **b)** SDS-PAGE showing the Srt5M catalyzed charging of the K48-linked tetraUb onto sfGFPN150GGK.

Next to the study of homotypic Ub chains (chains that are linked through the same linkage type), studies addressing heterotypic chains represent an emerging field in the Ub research community[355]. Heterotypic chains can be further categorized into mixed chains (chains that are linked through different lysine residues) and branched chains (chains in which one Ub serves as a platform for more than one Ub conjugation; e.g. K11/K48 branched chains)[100]. Branched chains add a new layer of complexity to the existing Ub code, and play important roles in endocytosis of host proteins by viruses (*via* K11/K63 branched chains)[127] or by enhancing proteasomal degradation during mitosis (*via* K11/K48 branched chains)[365].

However, approaches to generate heterotypic branched chains rely hitherto on the combination of different E2 enzymes with mutant Ubs[366,368], sophisticated chemical protein synthesis approaches[369] or on using click chemistry to produce non-natively triazole-linked branched chains[370]. Therefore, we once more adapted our general two-step UBuilder approach to gain access to target proteins site-specifically modified with branched Ub architectures (Figure 3.19 a and b). In a first proof-of-principle study, we incorporated AzGGK twice into one Ub monomer bearing a C-terminal His6-tag by concomitantly suppressing two amber codons at positions K48/K63, K6/K63, K6/K48, K11/K63, K11/K48 and K6/K11. After affinity chromatography using Ni-NTA and Staudinger reduction to convert AzGGK to GGK, the resulting double-GGK-bearing Ub variants were incubated with Ub(LAT) and Srt2A. To our delight, we observed successful formation of all six differently linked branched triUbs (Figure 3.19 a and Figure S3.3 for complete gels). Interestingly, we could also observe the various intermediate diUb species, which in dependence of their linkage type show a different migration behaviour in the SDS-PAGE.

In an analogous fashion, we incubated the double-GGK-bearing Ubs with diUb B in presence of Srt2A. SDS-PAGE analysis confirmed to the generation of all six different branched pentaUbs as well as the corresponding triUb intermediates (Figure 3.19 b and Figure S3.4 for complete gels). With the results of the branched triUbs and pentaUbs in hand we envisioned to charge a branched Ub architecture onto a target protein. Compared to triUb assembly, however, pentaUb assembly is rather low yielding and the preparation of diUb B is time consuming. We, therefore, decided to charge a branched triUb site-specifically onto a POI.

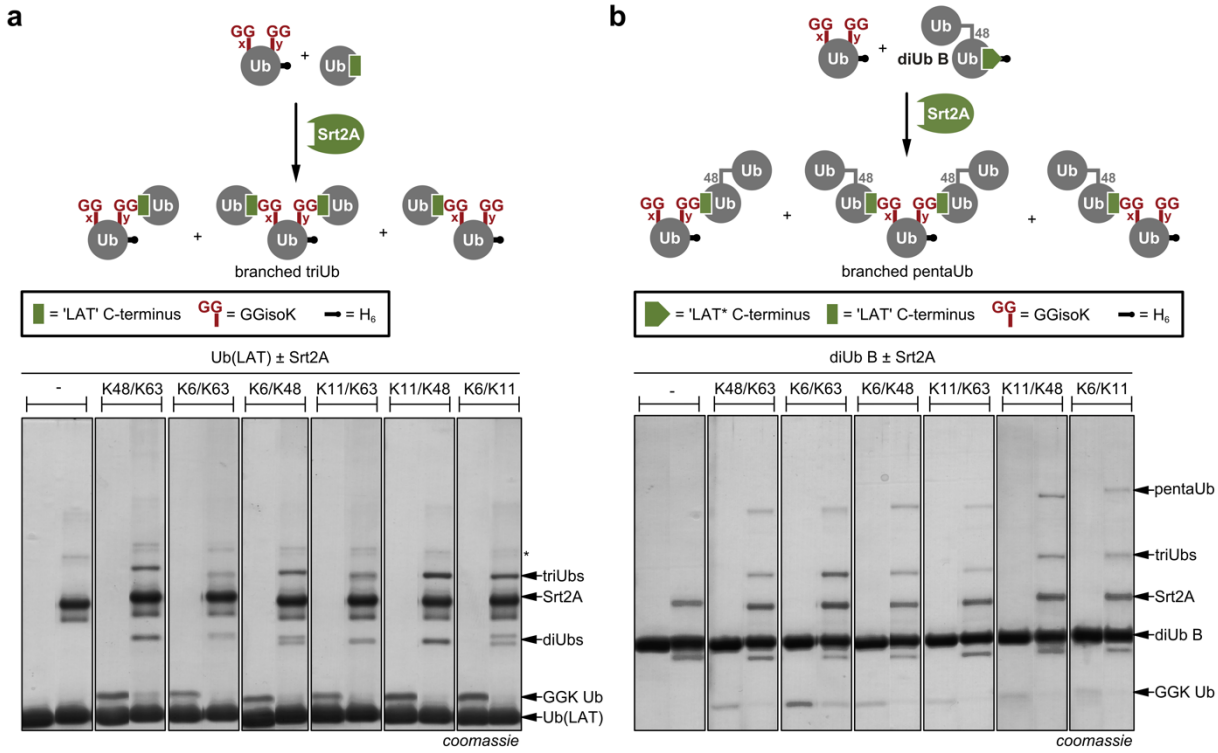


Figure 3.19 Generation of differently linked branched triUbs and pentaUbs. a) Top: Schematic representation of sortase-mediated branched triUb formation. Bottom: Incubation of different UbKxx/KyyGGK-His6 with Ub(LAT) in the presence of Srt2A leads to formation of the desired branched triUbs and both diUb intermediates. Asterix indicates Intermediate between sortase and Ub(LAT) **b)** Top: Schematic representation of sortase-mediated branched pentaUb formation. Bottom: Incubation of different UbKxx/KyyGGK-His6 with diUb B in the presence of Srt2A leads to formation of the desired branched pentaUbs and both triUb intermediate.

Due to the high yield of the branched triUb conjugate (as judged from SDS-PAGE analysis, figure 3.15 a) and the intriguing role of K11/K48-linked branched chains in proteasomal degradation[368], we decided from heron to focus on charging a K11/K48 branched triUb onto GFP (Figure 3.20 a). Therefore, we expressed and purified a Ub variant, which bears GGK at K11 and K48, and displays the 'LPT' motif at its C-terminus. In the first sortase reaction we incubated UbK11/K48GGK(LPT)-His6 with Ub(LAT) in the presence of Srt2A. This led to the time-dependent formation of the corresponding K11/K48 branched triUb bearing the 'LPT' motif at the C-terminus of the distal Ub (Figure 3.20 b). Afterwards, we purified the K11/K48-linked branched triUb *via* affinity chromatography using Ni-NTA followed by SEC (Figure 3.20 c). The identity of the K11/K48 branched triUb was furthermore confirmed using LC-MS (Figure 3.20 d). Finally, we set up a second Srt5M-catalyzed sortase reaction between sfGFPN150GGK-His6 and the K11/K48 branched triUb bearing the 'LPT' motif at the C-term of the proximal Ub. After already 30 minutes we observed robust formation of sfGFP site-specifically charged with a K11/K48 branched chain (Figure 3.20 e). As judged by SDS-PAGE the yield of the conjugate between GFP and the branched triUb is higher compared to the conjugate between GFP and the K48-tetraUb (see Figure 3.18 b), which might facilitate preparation of the conjugate in quantities useful for downstream application such as structural biology.

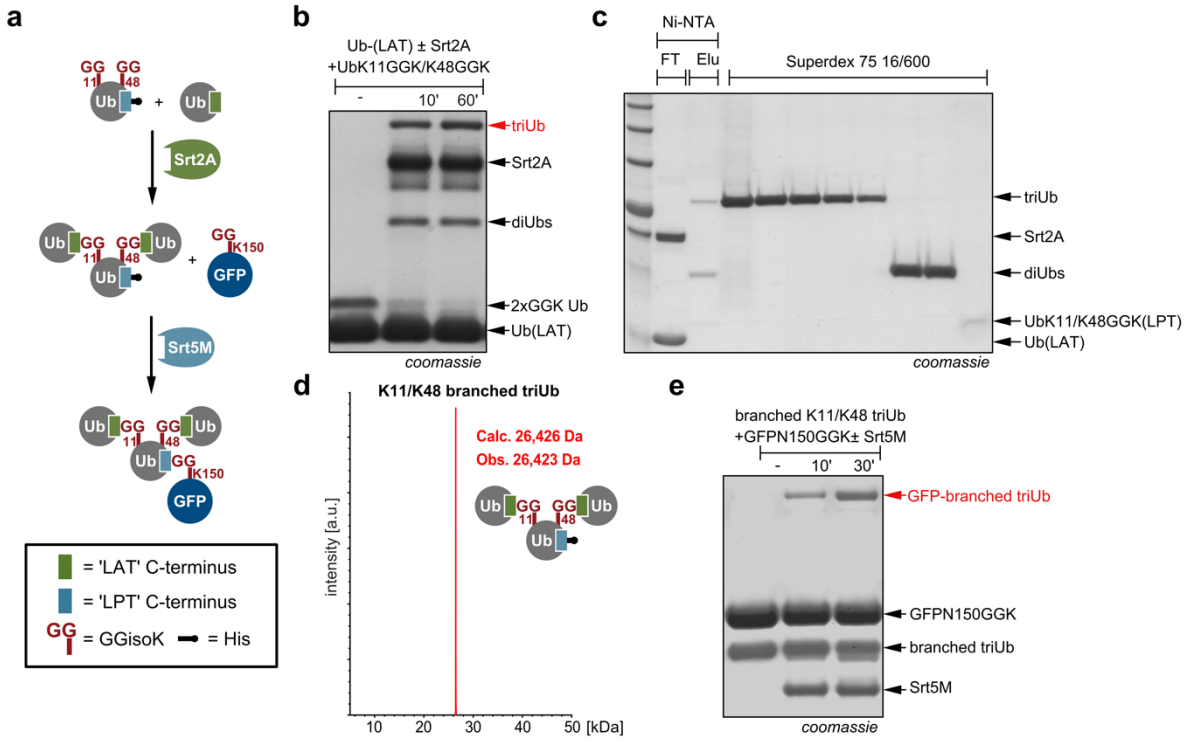


Figure 3.20 Formation and charging of a K11/K48 branched triUb onto GFP. a) Scheme displaying the formation and charging of a K11/K48-linked triUb onto GFP. **b)** A Ub bearing GGK at K11 and K48 is incubated with Ub(LAT) in presence of Srt2A leading to the formation of a branched triUb and the two diUb intermediates. **c)** Purification of the K11/K48 branched triUb vis affinity chromatography using Ni-NTA chromatography followed SEC. **d)** LC-MS characterization of K11/K48 branched triUb. **e)** Incubation of the branched triUb (bearing a "LPT" sortase recognition motif at the C-terminus of the proximal Ub) with sfGFPN150GGK in presence of Srt5M leads to the formation of GFP bearing a branched triUb.

3.6 Discussion and Outlook

In summary, we report UBuilder as a novel tool for the generation of complex Ub/Ubl architectures. UBuilder expands its predecessor "sortylation" by the introduction of a pair of orthogonal sortases which allow the iterative use of sortase-mediated transpeptidation. We have proven the straightforward workflow of UBuilder by the generation and purification of seven differently linked Ub/SUMO hybrid chains and showed in subsequent binding studies with Rap80 that UBuilder is able to give valuable insights into important biological questions. So far, access to target proteins site-specifically modified with defined complex Ub topologies has been restricted to chemical total synthesis, which is limited to small refoldable proteins. By combining UBuilder with enzymatic Ub assembly, we underlined its potential to solve this problem by charging a K48-linked tetraUb chain site-specifically on sfGFP. Lastly, we expanded UBuilder to the emerging field of heterotypic Ub chains. Therefore, we installed GGK twice into a Ub monomer which served as a platform for the generation of six differently linked triUbs and pentaUbs; an advancement which allowed us to create sfGFP site-specifically modified with a K11/K48 linked triUb.

We envision that UBuilder can be expanded by identifying additional orthogonal sortases with distinct recognition motifs, which would give access to even more complex topologies. Enzymatic assembly of Ub chains *via* the combination of different E2 enzymes is complementary to UBuilder and can be further expanded in its scope. Furthermore, we envisage

UBuilder to have great impact especially on the generation of various hybrid chains containing different Ubls with linkages which would otherwise be inaccessible.

Additionally, we anticipate that UBuilder might be highly useful for structural biology applications. Recent years have experienced a boom in structural biology, led by cryo-EM, focusing on how the proteasome engages ubiquitylated substrate proteins[371,372]. In these studies, the ubiquitylated substrates were, however, prepared by fusing domains displaying a single ubiquitylation site to substrates that do not have an endogenous ubiquitylation site on their own. Afterwards enzymatic assembly was used to generate Ub chains of random length on these artificial substrates. In this regard, we envisage that UBuilder might be a valuable tool to provide endogenous substrate proteins site-specifically modified with Ub chains of defined linkage and length. Additionally, with UBuilder we are in the unique position to place DUB resistant linkages with surgical precision, e.g. between the substrate protein and the Ub chain. Since these conjugates should not be processed by the proteasome, they might lead to the elucidation of early snapshots of proteasomal substrate-engagement.

In conclusion, we believe that UBuilder will have immediate impact on Ub researchers, due to its use of easily accessible building blocks, its simplicity and its adaptability to virtually any kind of Ub/Ubl architecture.

4. Experimental Part

4.1 General material and methods - Chemistry

All solvents and chemicals were purchased from Sigma Aldrich, Carbolution, Acros Organics, VWR, Iris Biotech or Fisher Scientific and were used without further purification unless otherwise stated. All air or moisture sensitive reactions were carried out under argon atmosphere. Thin-layer chromatography (TLC) was performed using Merck Millipore silica gel 60 F-254 plates. TLC plates were visualized using UV light (254 nm) and/or staining with ninhydrin.
NMR spectra were recorded on a Bruker AVANCE 500 UltraShield™ spectrometer (500 MHz for ^{1}H-NMR, 125 MHz for ^{13}C-NMR). Chemical shifts (δ), reported in ppm, are referenced to the residual proton solvent signals (DMSO-d_6 – 2.50 ppm for ^{1}H-NMR and 39.5 ppm for ^{13}C-NMR spectra). Coupling constants (J) are reported in Hertz (Hz) while peak multiplicities are descripted as follows: s (singlet), d (doublet), t (triplet), q (quartet), quint (quintet), dt (doublet of triplets), ddd (doublet of doublet of doublets), m (multiplet), br (broad signal).
LC-MS was carried out on an Agilent Technologies 1260 Infinity LC-MS system with a 6310 Quadrupole spectrometer. The solvent system consisted of 0.1 % formic acid in water as buffer A and 0.1 % formic acid in ACN as buffer B. Small molecule LC-MS was carried out on a Jupiter C18 5 µm (2 x 150 mm) capillary column. Protein LC-MS was carried out on a Jupiter C4 5 µm (2 x 150 mm) capillary column. Protein mass was calculated by deconvolution within the Chemstation software (Agilent Technologies). Theoretical masses of proteins were calculated using ProtParam and were manually corrected for the mass of the unnatural amino acid. The samples were analyzed in both positive and negative mode and followed by UV absorbance at 193, 254 and/or 280 nm.

4.1.1 Solid phase peptide synthesis (SPPS)

General synthetic procedures for solid phase peptide synthesis (SPPS)
All peptides (e.g. N^6-glycylglycyl-L-lysine (GGK), N^6-((2-azidoacetyl)glycyl)-L-lysine (AzGGK) and Fmoc-VLPLTGG) were synthesized via solid phase peptide synthesis (SPPS). SPPS was performed in a custom-made glass apparatus with a frit for larger amounts of resin or in plastic syringes with a frit for small amounts (< 1 g). Shaking was performed manually or by using a rotary unit. The equivalents (eq.) used were based on the maximal loading capacity of the CTC resin given by the supplier.

Loading and capping of CTC resin
SPPS was performed according to the Fmoc-strategy for solid phase synthesis using CTC-resin. Therefore, 1.2 eq. of Fmoc-protected amino acid and 2.5 eq. of DIPEA were dissolved in anhydrous DCM (10 mL/g resin) and then added to 1.0 eq. of CTC-resin (100-200 mesh, 1.0-1.6 mmol/g maximal loading capacity) in a syringe equipped with a frit. The mixture was allowed to shake for 1 h at RT. Capping of the remaining chlorotrityl-groups was performed by adding 3 eq. of MeOH and 2.5 eq DIPEA to the syringe followed by 15 minutes of shaking at RT. Subsequently, the resin was washed five times with DCM and five times with DMF. The

loading capacity of the resin was determined by mass gain after washing the resin with MeOH and drying in high vacuum.

On-resin Fmoc-deprotection

Removal of the Fmoc-protection group was performed by adding a solution of 20% piperidine in DMF to the resin in the syringe, followed by five minutes of shaking at RT. This procedure was repeated for 10 minutes. The resin was subsequently washed five times with DMF.

Standard on-resin coupling of amino acids

For the coupling of amino acids, a 0.1-0.2 M solution of the Fmoc-protected amino acid (2 eq.), HATU (2 eq.), HOAt (2 eq.) and DIPEA (5 eq.) was prepared in DMF and stirred for 5 min at RT. Afterwards, the syringe/reactor containing the Fmoc-deprotected resin-bound free amino acid was incubated with this solution at RT for 45 minutes. Subsequently, the resin was washed five times with DMF.

Cleavage from the CTC resin

The resin was washed five times with DMF and five times with DCM and then treated with a solution of 20% HFIP in DCM (v/v) at RT for 10 min. The filtrate was collected, and the procedure was repeated two more times. Afterwards, the resin was washed three times with DCM. The filtrates were combined, and the solvent was evaporated under reduced pressure.

Removal of acid-labile protection groups

For the removal of acid-labile protection groups the product was dissolved in a mixture of TFA/DCM/H_2O (90/5/5, v/v/v) and stirred at RT for 1h. Subsequently, the solvents were removed under reduced pressure followed by co-evaporation with toluene.

Precipitation with Et_2O

After removal of the solvents the product was added dropwise to a centrifugal tube with ice-cold Et_2O. After centrifugation, the precipitate was washed twice with ice-cold Et_2O and centrifuged again. The resulting precipitate was dissolved in H_2O and lyophilized.

4.2 General material and methods – Biology

4.2.1 Material

4.2.1.1 Organisms

Table 4.2.1.1.1 Bacterial strains

Strain	Species	Genotype	Purpose	Origin
BW25113 (K12)	*E. coli*	*lacI*$^{+}$*rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR*514 Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ *rph-1* Δ*(araB–D)567* Δ*(rhaD–B)568* Δ*lacZ4787*(::*rrnB-3*) *hsdR514 rph-1*	Protein expression	GE Healthcare
NEB 10-beta	*E. coli*	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14-ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrrhsdRMS-mcrBC)*	Cloning	NEB
BL21 (DE3)	*E. coli*	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	Protein expression	NEB
BL21 (DE3) Rosetta II	*E. coli*	*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (CamR)*	Protein expression	Sigma Aldrich

Table 4.2.1.1.2 Mammalian cell lines

Cell line	Species	Specification	Purpose	Origin
HEK293T	*Homo sapiens*	*Highly transfectable cell line derived from the primary embryonal kidney cell line 293. Transfected with sheared human adenovirus type 5. Express SV40 large T antigen.*	Transient transfection	Feige Lab, TUM (Germany)

4.2.1.2 Oligonucleotides

Oligonucleotides were designed using SnapGene software 5.06 (GSL Biotech LLC) and Tm Calculator v 1.12.0 (NEB). Oligonucleotides were purchased lyophilized and desalted from Sigma Aldrich.

4.2.1.3 Plasmids

The plasmids used in this work were either already available in the Lang lab, purchased at addgene (as noted in the plasmid description) or cloned during the course of the projects. New genes were ordered as strings supplied by Thermo and cloned into available backbones using restriction cloning or Gibson assembly[373]. Point mutations, deletions and small insertions were introduced using Site-directed ligase independent mutagenesis (SLIM)[374]. Tables 4.2.1.2.1 to 4.2.1.2.1.8 list all plasmids used in this work.

Table 4.2.1.3.1: Plasmids encoding Ub and Ubl variants for expression in bacterial cells as well as duet plasmids encoding Ub/Ubls an sortase variants for sortylation in living *E. coli*.

Identifier	Plasmid	Resistance
MF UbC1	pET17b Ub wt	Amp

MF UbC2	pET17b Ub-VLPLTGG	Amp
MF UbC9	pET17b Ub-VLLPLTGG	Amp
MF UbC19	pET17b Ub-VLALTGG	Amp
MF UbC20	pET17b Ub-VLLALTGG	Amp
MF UbC24	pET17b Ub-VLPLSGG	Amp
MF UbC25	pET17b Ub-VLLPLSGG	Amp
MF UbC33	pBAD Sortase7M-TEV-Strep and Ub-VLLPLTGG	Kan
MF UbC34	pBAD Sortase2A-TEV-Strep and Ub-VLLALTGG	Kan
MF UbC35	pBAD SortaseMF1-TEV-Strep and Ub-VLLALTGG	Kan
MF UbC36	pBAD Sortase7M-TEV-Strep C184A and Ub-VLLPLTGG	Kan
MF UbC37	pBAD Sortase2A-TEV-Strep C184A and Ub-VLLALTGG	Kan
MF UbC38	pBAD SortaseMF1-TEV-Strep C184A and Ub-VLLALTGG	Kan
MF UbC40	pBAD Sortase2A-TEV-Strep and SUMO1-LAQTG	Kan
MF UbC41	pBAD SortaseMF1-TEV-Strep and SUMO1-LAQTG	Kan
MF UbC43	pBAD Sortase2A-TEV-Strep and SUMO1-LLAQTG	Kan
MF UbC44	pBAD SortaseMF1-TEV-Strep and SUMO1-LLAQTG	Kan
MF UbC45	pBAD Sortase7M-TEV-Strep and Ub-VLPLTGG	Kan
MF UbC46	pBAD Sortase2A-TEV-Strep and Ub-VLALTGG	Kan
MF UbC47	pBAD SortaseMF1-TEV-Strep and Ub-VLALTGG	Kan
MF UbC60	pET17 Ub-LLALTGYEAAAK	Amp
MF UbC62	pET17 Ub-LALTG	Amp
MF UbC74	pET17 Ub-LLALTGYEAAAK-His	Amp
MF UbC76	pET17 Ub-LLALTGYEAAAK-His	Amp
MF UbC77	pET17 Ub-LPLTGYEAAAK-His	Amp
MF UbC78	pET17 Ub-LLPLTGYEAAAK-His	Amp
MF UbC82	pET17 Ub-VLALTGYEAAAK	Amp
MF UbC83	pET17 Ub-VLPLTGYEAAAK	Amp
MF UbC84	pET17 Ub-VLLPLTGYEAAAK	Amp
MF UbC86	pET17 Ub-VLALTGYEAAAK-His	Amp

Table 4.2.1.3.2: Plasmids encoding Ubl variants for expression in bacterial cells.

Identifier	Plasmid	Resistance
U2	pET17b His-TEV-Sumo1-VLAQTGG	Amp
U3	pET17b His-TEV-Sumo1-VLLAQTGG	Amp
U16	pET17 His-TEV-SUMO2 wt	Amp
U17	pET17 His-TEV-G-SUMO2	Amp

Table 4.2.1.3.3: Plasmids encoding POIs for genetic code expansion in bacterial cells.

Identifier	Plasmid	Resistance
OA6	pPylT sfGFP-His N150TAG MbPylT	Tet
OA11	pET11a PCNA-C80-Strep K164TAG MbPylT	Amp

OA12	pPylT sfGFP-His N150TAG V2P MbPylT	Tet
OA13	pPylT PCNA-C80-Strep K164TAG MbPylT	Tet
OA14	pPylT PCNA-C80-Strep wt MbPylT	Tet
OA15	pPylT PCNA-C80-His K164TAG MbPylT	Tet
OA16	pPylT PCNA-C80-His wt MbPylT	Tet
OA20	pPylT PCNA-His K164TAG MbPylT	Tet
OA32	pPylT PCNA-His wt MbPylT	Tet
OA45	pPylT SUMO2-His wt MbPylT	Tet
OA46	pPylT SUMO2-His K5TAG MbPylT	Tet
OA47	pPylT SUMO2-His K7TAG MbPylT	Tet
OA48	pPylT SUMO2-His K11TAG MbPylT	Tet
OA49	pPylT SUMO2-His K33TAG MbPylT	Tet
OA50	pPylT SUMO2-His K35TAG MbPylT	Tet
OA51	pPylT SUMO2-His K42TAG MbPylT	Tet
OA52	pPylT SUMO2-His K45TAG MbPylT	Tet
OA53	pPylT SUMO2-His K21 TAG MbPylT	Tet
OA82	pPylT SUMO2 K5TAG MbPylT	Tet
OA83	pPylT SUMO2 K7TAG MbPylT	Tet
OA108	pPylT Ub-His K48TAG, K63TAG MaPylT	Tet
OA109	pPylT Ub-His K6TAG, K63TAG MaPylT	Tet
OA110	pPylT Ub-His K6TAG, K48TAG MaPylT	Tet
OA111	pPylT Ub-His K11TAG, K63TAG MaPylT	Tet
OA112	pPylT Ub-His K11TAG, K48TAG MaPylT	Tet
OA113	pPylT Ub-His wt MaPylT	Tet
OA114	pPylT Ub-His K6TAG, K11TAG MaPylT	Tet
OA116	pPylT Ub-VLLPLTGG-His K48TAG, K63TAG MaPylT	Tet
OA117	pPylT Ub-VLLPLTGG-His K6TAG, K63TAG MaPylT	Tet
OA118	pPylT Ub-VLLPLTGG-His K6TAG, K48TAG MaPylT	Tet
OA119	pPylT Ub-VLLPLTGG-His K11TAG, K63TAG MaPylT	Tet
OA120	pPylT Ub-VLLPLTGG-His K11TAG, K48TAG MaPylT	Tet
OA121	pPylT Ub-VLLPLTGG-His K6TAG, K11TAG MaPylT	Tet

Table 4.2.1.3.4: Plasmids encoding Ub variants for genetic code expansion in bacterial cells.

Identifier	Plasmid	Resistance
MF UbA2	pPylT Ub-His K6TAG MbPylT	Tet
MF UbA3	pPylT Ub-His wt MbPylT	Tet
MF UbA4	pPylT Ub-His K11TAG MbPylT	Tet
MF UbA5	pPylT Ub-His K27TAG MbPylT	Tet
MF UbA6	pPylT Ub-His K33TAG MbPylT	Tet
MF UbA7	pPylT Ub-His K29TAG MbPylT	Tet
MF UbA8	pPylT Ub-His K48TAG MbPylT	Tet
MF UbA9	pPylT Ub-His K63TAG MbPylT	Tet
MF UbA11	pPylT Ub-VLLPLTGG-His K6TAG MbPylT	Tet

MF UbA12	pPylT Ub-VLLPLTGG His MbPylT	Tet
MF UbA13	pPylT Ub-VLLPLTGG-His K48TAG MbPylT	Tet
MF UbA25	pPylT Ub-VLLPLTGG-C80-His K48TAG MbPylT	Tet
MF UbA34	pPylT Ub-LALTGG-His K63TAG MbPylT	Tet
MF UbA35	pPylT Ub-LLALTGG-His K63TAG MbPylT	Tet
MF UbA36	pPylT Ub-LLPLTGG-His K63TAG MbPylT	Tet
MF UbA37	pPylT Ub-LPLTGG-His K63TAG MbPylT	Tet
MF UbA38	pPylT Ub-LLALTGG-His K48TAG MbPylT	Tet
MF UbA39	pPylT Ub-LALTGG-His K48TAG MbPylT	Tet
MF UbA40	pPylT Ub-LPLTGG-His K48TAG MbPylT	Tet

Table 4.2.1.3.5: Mammalian cell plasmids.

Identifier	Plasmid	Resistance
MA1	pIRES *Sa*Sortase2A-myc	Amp
MA2	pIRES *Sa*Sortase7M-myc	Amp
MA3	pcDNA Ub wt	Amp
MA4	pcDNA Ub-VLLALTGG	Amp
MA5	pcDNA Ub-VLLPLTGG	Amp
MA6	pcDNA HA-Ub wt	Amp
MA7	pcDNA HA-Ub-VLLALTGG	Amp
MA8	pcDNA HA-Ub-VLLPLTGG	Amp
MA9	pcDNA Ub-HA wt	Amp
MA10	pcDNA Ub-VLLALTGG-HA	Amp
MA11	pcDNA Ub-VLLPLTGG-HA	Amp
MA12	pIRES *Sa*SortaseMF1-myc	Amp
MA13	pcDNA SUMO1-LLAQTG	Amp
MA14	pcDNA SUMO1-LAQTG	Amp
MA15	pcDNA His-Ub-VLLALTGG-HA	Amp
MA16	pcDNA His-Ub-HA wt	Amp
MA17	pcDNA His-Ub-VLLPLTGG-HA	Amp
MA18	pcDNA HA-SUMO-LLAQTG	Amp
MA19	pcDNA HA-SUMO-LAQTG	Amp
MA20	PCNA K164TAG-His 4x PylT	Amp
MA21	pcDNA HA-SUMO wt	Amp
MA22	pcDNA SUMO wt	Amp
MA23	pIRES *Sa*Sortase2A-NLS-myc	Amp
MA24	pIRES *Sa*SortaseMF1-NLS-myc	Amp
MA47	pcDNA Ha-Ub VLLALTG-YEAAAK	Amp
MA48	pcDNA HA-SUMO-LLAQTG-YEAAAK	Amp
MA49	pIRES *Sa*Sortase2A-His	Amp
MA55	pcDNA Kozak HA-SUMO-wt	Amp
MA56	pcDNA Kozak HA-SUMO-LLAQTGG	Amp
MA57	pcDNA Kozak HA-SUMO-LLAQTGYEAAAK	Amp

MA58	pcDNA Kozak HA-Ub-wt	Amp
MA59	pcDNA Kozak HA-Ub-LLALTGG	Amp
MA60	pcDNA Kozak HA-Ub-LLALTGYEAAAK	Amp
MA61	pIRES Kozak *Sa*Sortase2A-myc	Amp
MA62	SE325 PCNA wt 4xPylT	Amp
MA66	pcDNA Kozak HA-SUMO-LLAQTG	Amp
MA67	pcDNA Kozak HA-Ub-LLALTG	Amp
MA68	pIRES Kozak Myc(NLS)-*Sa*Sortase 2A-myc	Amp
MA69	pIRES Kozak *Sa*Sortase2A-myc-SV40(NLS)	Amp
MA70	pIRES Kozak SV40(NLS)-*Sa*Sortase2A-myc	Amp
MA71	pIRES Kozak *Sa*Sortase2A-myc-Myc(NLS)	Amp
MA72	SE363 sfGFP-150TAG-His6-4xPylT	Amp
MA73	pEF1 MF18 AzGGKRS 4xPylT	Amp

Table 4.2.1.3.6: Plasmids encoding different sortases for expression in bacterial cells.

Identifier	Plasmid	Resistance
MF So1	pET29 *Sa*SortaseA-His 5M	Kan
MF So2	pET23b *Sa*SortaseA-His wt	Amp
MF So3	pET28a *Sp*SortaseA-His wt	Kan
MF So4	pET29 *Sa*SortaseA-TEV-His 5M	Kan
MF So5	pET29 *Sa*SortaseA-His 2A	Kan
MF So6	pET29 *Sa*SortaseA-His 4S	Kan
MF So7	pET29 *Sa*SortaseA-TEV-His 2A	Kan
MF So8	pET29 *Sa*SortaseA-TEV-His 4S	Kan
MF So9	pET29 *Sa*Strep-SortaseA-TEV-His 2A	Kan
MF So10	pET29 *Sa*SortaseA-TEV-His 7M	Kan
MF So11	pET29 *Sa*Sortase-TEV-Strep 2A	Kan
MF So12	pET29 *Sa*Sortase-TEV-His MF1	Kan

Table 4.2.1.3.7: Plasmids encoding PylRS variants used for genetic code expansion in bacteria.

Identifier	Plasmid	Resistance
MF RS1	pBK Mb PylRS wt	Amp
MF RS2	pBK Mb PylRS L274A N311Q C313S (MF18)	Amp
MF RS3	pBK Ma PylRS wt	Amp
MF RS4	pBK Ma PylRS M129A, N166Q V168S H227I Y228P (MF18 “IP”)	Amp

Table 4.2.1.3.8: Plasmids encoding proteins for expression in bacterial cells and helper plasmids for cloning.

Identifier	Plasmid	Resistance
OP1	pET28a His-Thr-USP2 254-606 wt	Kan
OP10	pET21d His-E1	Amp
OP11	pGEX6P1-UBE2N (Ubc13)	Amp
OP12	pGEX6P1-UBE2V1 (Uev1A)	Amp

OP13	pGEX6P1-UBE2R1 (CDC34)	Amp
OP14	pOPINK His-UBE2S delta C	Kan
OP25	pET28 Rap80(79-124) wt	Kan
OP26	pET28 Rap80(79-124) 7A linker	Kan
OP32	1124 pGEX-4T-1/NBR1(UBA, 876-966)	Amp
OP33	1125 pGEX-4T-1/p62(UBA, 376-440)	Amp
OP34	1322 hHR23A-UBA2-pGEX-4T1	Amp
OP35	1744 Cbl-b UBA-pGEX	Amp
OP36	pGEX GST-USP37 UIM23	Amp
OP37	pGEx GST-USP37 UIM123	Amp
OP38	pGEX GST-Tab2-NZF	Amp
OP39	pGEX GST-ANKRD13A-UIM34	Amp
OP40	pET17 GST-TEV-Rap80(1-137) SIM and UIMs	Amp
OP41	pGEX GST-Thrombin-Rap80(1-137) 7A	Amp
OP42	pET17 GST-TEV-Rap80(1-137) SIM and UIMs C70S	Amp
OP43	pET17 GST-TEV-Rap80(1-137) SIM and UIMs 7A linker	Amp
OP44	pET17 GST-TEV-Rap80(1-137) SIM and UIMs 7A linker C70S	Amp
OP45	pET28 GST USP2 C276A	Kan
OP46	pET17 GST	Amp
OP47	pET17 His-Rap80(1-137) SIM and UIMs 7A linker	Amp
OP48	pET17 His-Rap80(1-137) SIM and UIMs 7A linker C70S	Amp
OP50	pET17 His-Rap80(1-137) SIM and UIMs	Amp
OP51	pET17 His-Rap80(1-137) SIM and UIMs C70S	Amp
OP53	pET17 GFP-Rap80(1-137) wt	Amp
OP54	pET17 GST-GGGS-Rap80(1-137) wt	Amp
OP69	pET17 GST-GGGS-Rap80(1-137) SIM and UIMs 7A linker	Amp
OP70	pET17 GST-3xGGGS-Rap80(1-137) SIM and UIMs 7A linker	Amp
OP71	pET17 Strep-Rap80(1-137) SIM and UIMs 7A linker	Amp
OP73	pET17 His-Strep-Rap80(1-137) SIM and UIMs 7A Linker	Amp
OP74	pET17 Strep-2xGGGGS-Rap80(1-137) SIM and UIMs 7A linker	Amp
OP75	pET17 Strep-Linker(Rap80)-Rap80(1-137) SIM and UIMs 7A linker	Amp
OP76	pET17 Strep-His-Rap80(1-137) SIM and UIMs 7A Linker	Amp
OP77	pET17 Spottag-His-Rap80(1-137) SIM and UIMs 7A Linker	Amp

4.2.1.4 Buffers and media

Table 4.2.1.4.1: Compositions of buffers.

Buffer	Components	Concentration/quantity
5x Q5 reaction buffer	Tris base, pH 8.8	100 mM
	KCl	50 mM
	$MgCl_2$	5 mM
	$(NH_4)SO_4$	50 mM

	Triton X-100	0.5 % (v/v)
	BSA	0.5 mg/mL
10x DNA Loading buffer	Sucrose	40 % (w/v)
	Orange G	0.15 % (w/v)
	Xylen cyanol	0.05 % (w/v)
	Bromophenol blue	0.05 % (w/v)
His lysis buffer	Tris base, pH 8.0	20 mM
	NaCl	300 mM
	Imidazole	30 mM
	PMSF	500 µM
	DNAse I	
	cOmplete™ protease inhibitor	1 tablet/100 ml
His wash buffer	Tris base, pH 8.0	20 mM
	NaCl	300 mM
	Imidazole	30 mM
His elution buffer	Tris base, pH 8.0	20 mM
	NaCl	300 mM
	Imidazole	300 mM
His mammalian lysis buffer	Tris-HCl, pH 8.0	50 mM
	NaCl	150 mM
	Triton-X-100	1 % (v/v)
	Imidazole	20 mM
	Mammalian protease inhibitor (VWR)	1x
His mammalian wash buffer	Tris-HCl, pH 8.0	50 mM
	NaCl	150 mM
	Triton-X-100	1 % (v/v)
	Imidazole	40 mM
His mammalian elution buffer	Tris-HCl, pH 8.0	50 mM
	NaCl	150 mM
	Triton-X-100	1 % (v/v)
	Imidazole	300 mM
Ub lysis buffer	Tris-HCl, pH 8.0	50 mM
	$MgCl_2$	10 mM
	EDTA	1 mM
	NP-40	0.1 % (v/v)
Sortase lysis buffer	Tris-HCl, pH 8.0	50 mM
	NaCl	300 mM
	$CaCl_2$	1 mM
	PMSF	1 mM
	DNAse I	
	cOmplete™ protease inhibitor	1 tablet/100 ml
RIPA lysis buffer	Tris-HCl, pH 8.0	50 mM

	NaCl	150 mM
	EDTA	1 mM
	SDS	0.1 % (w/v)
	Sodium deoxycholate	0.5 % (w/v)
	Triton-X-100	1 % (w/v)
	Mammalian protease inhibitor (VWR)	1x
Sortase reaction buffer	Tri-HCl, pH 8.0	50 mM
	NaCl	150 mM
	$CaCl_2$	5 mM
DUB reaction buffer	Tris-HCl, pH 8.0	50 mM
	NaCl	150 mM
	$CaCl_2$	5 mM
	DTT	100 µM
4x SDS-PAGE resolving buffer	Tris-HCL, pH 8.8	1.5 M
	SDS	0.4 % (w/v)
4x SDS-PAGE stacking buffer	Tris-HCL, pH 6.8	500 mM
	SDS	0.4 % (w/v)
4x Lämmli loading buffer	Tris-HCL, pH 6.8	240 mM
	Glycerol	40 % (v/v)
	SDS	8 % (w/v)
	Bromophenol blue	0.4 % (w/v)
	B-mercaptoethanol	5 % (v/v)
MES SDS-PAGE running buffer	MES, pH 7.3	50 mM
	TRIS, pH 7.3	50 mM
	SDS	0.1 % (w/v)
	EDTA	1 mM
Immunoblot transfer buffer	Tris base, pH 8.9-9.4	48 mM
	Glycine	39 mM
	SDS	1.3 mM
	Methanol	20 % (v/v)
50x TAE buffer	Tris-HCl, pH 8.8	2 M
	Acetic acid	0.9 M
	EDTA	127 mM
10x TBS	Tris-HCl, pH 7.6	200 mM
	NaCl	1.5 M
TBS-T	TBS, pH 7.6	1x
	Tween-20	0.1 %
10x PBS pH 7.4	Na_2HPO_4	100 mM
	KH_2PO_4	18 mM
	NaCl	1.37 M
	KCl	27 mM

5x hybridisation buffer (Slim buffer)	Tris-HCl, pH 9.0	125 mM
	NaCl	750 mM
	EDTA	100 mM

Table 4.2.1.3.2: Composition of media.

Medium	Components	Concentration/quantity
Lysogeny Broth (LB), Lennox formulation	Tryptone	10 g/l
	Yeast extract	5 g/l
	NaCl	5 g/l
2x yeast extract/tryptone (2xYT) medium	Tryptone	16 g/l
	Yeast extract	10 g/l
	NaCl	5 g/l
Super optimal broth (SOB) medium	Tryptone	20 g/l
	Yeast extract	5 g/l
	$MgCl_2$	0,96 g/l
	NaCl	0.5 g/l
	KCl	0.186 g/l
Super optimal broth with catabolite repression (SOC) medium	Tryptone	20 g/l
	Yeast extract	5 g/l
	MgCl2	0,96 g/l
	NaCl	0.5 g/l
	KCl	0.186 g/l
	Glucose	20 mM
Autoinduction medium[375]	5% Aspartate, pH 7.5	50 ml/l
	10% Glycerol	50 ml/l
	25x 17 AA mix	40 ml/l
	50x M	20 ml/l
	Leucine 4 mg/ml, pH 7.5	10 ml/l
	20% Arabinose	2.5 ml/l
	$MgSO_4$ 1 M	2 ml/l
	40 % Glucose	1.25 ml/l
	5,000x trace metals	200 µl/l
	Nicotinamide 1 M	6 ml/l
Non-induction medium[375]	5% Aspartate, pH 7.5	50 ml/l
	25x 17 AA mix	40 ml/l
	50x M	20 ml/l
	Leucine 4 mg/ml, pH 7.5	10 ml/l
	$MgSO_4$ 1 M	2 ml/l
	40 % Glucose	1.25 ml/l
	5,000x trace metals	200 µl/l
GMML medium[376]	5x M9 salts	200 ml/l
	50% Glycerol	20 ml/l
	Leucine 4 mg/ml, pH 7.5	10 ml/l
	NaCl 850 mM	10 ml/l

	Fe_2SO_4 5 mM	1 ml/l
	$CaCl_2$ 100 mM	1 ml/l
	$MgSO_4$ 1 M	1 ml/l
	5,000x trace metals	200 µl/l
	Nicotinamide 1 M	6 ml/l
LB agar	Tryptone	10 g/l
	Yeast extract	5 g/l
	NaCl	5 g/l
	Agar-agar	15 g/l
4/3 buffered agar	TB buffer	1 ml/l
	10x PBS	1 ml/l
	Agar-agar	20 g/l
GMML agar[376]	5x M9 salts	200 ml/l
	50% Glycerol	20 ml/l
	Leucine 4 mg/ml, pH 7.5	10 ml/l
	NaCl 850 mM	10 ml/l
	Fe_2SO_4 5 mM	1 ml/l
	$CaCl_2$ 100 mM	1 ml/l
	$MgSO_4$ 1 M	1 ml/l
	5,000x trace metals	200 µl/l
	Nicotinamide 1 M	6 ml/l
	4/3 buffered agar	Fill up to 1 L
Autoinduction agar[375]	5% Aspartate, pH 7.5	50 ml/l
	10% Glycerol	50 ml/l
	25x 17 AA mix	40 ml/l
	50x M	20 ml/l
	Leucine 4 mg/ml, pH 7.5	10 ml/l
	20% Arabinose	2.5 ml/l
	$MgSO_4$ 1 M	2 ml/l
	40 % Glucose	250 µl/l
	5,000x trace metals	200 µl/l
	Nicotinamide 1 M	2 ml/l
	H_2O	73 ml/l
	4/3 buffered agar	Fill up to 1 L

4.2.1.5 Stock solutions

Table 4.2.1.5.1: Stock solutions

Stock	Components	Concentration/quantity
5% Aspartate pH 7.5	Aspartic acid (pH adjusted with NaOH)	5 % (w/v)
25x 17AA mix[330]	Glutamic acid sodium salt, aspartic acid, lysine-HCl, arginine-HCl,	5 g/l (each)

	histidine-HCl-H_2O, alanine, proline, glycine, threonine, serine, glutamine, asparagine·H2O, valine, leucine, isoleucine, tryptophan, methionine	
50x M	Na_2HPO_4	1.25 M
	KH_2PO_4	1.25 M
	NH_4Cl	2.5 M
	Na_2SO_4	0.25 M
5x M9 Salts	$Na_2HPO_4 \cdot 7$ H2O	64 g/l
	KH_2PO_4	15 g/l
	NH_4Cl	5 g/l
	NaCl	2.5 g/l
5000x Trace Metals[330]	$CaCl_2$	200 µM
	$MnCl_2$	100 µM
	$ZnSO_4$	100 µM
	$CoCl_2$	20 µM
	$CuCl_2$	20 µM
	$NiCl_2$	20 µM
	Na_2MoO_4	20 µM
	Na_2SeO_3	20 µM
	H_3BO_3	20 µM
	$FeCl_3$	500 µM

Antibiotics stock solutions and working concentrations

Antibiotics were dissolved, sterile filtered using a 0.2 µm membrane filter and stored at - 20 °C.

Table 4.2.1.5.2: Antibiotic stock solutions and working concentrations.

Antibiotic	Stock solution (mg/mL)	Working concentration (µg/mL)	Solvent
Ampicillin (Amp)	100	100	H_2O
Kanamycin (Kan)	50	50	H_2O
Tetracycline (Tet)	12.5	12.5	70 % (v/v) EtOH
Chloramphenicol (Cam)	50	50	96 % (v/v) EtOH

4.2.2 Molecular biology methods

4.2.2.1 DNA isolation and purification

Plasmid DNA was isolated from *E. coli* overnight cultures grown in 2x YT media (6 mL) or from *E. coli* wiped off from agar plates using the peqGOLD Plasmid Mini Prep Kit I (VWR) according to the instructions given by the supplier. Elution of the plasmid DNA from the columns was done with 20-50 µL nuclease free water. The DNA concentration and the purity were measured photometrically at 260 nm, 230 nm and 280 nm. For further use the plasmids were stored at -20 °C.

Double stranded DNA was purified from agarose gels via gel extraction using the Monarch® DNA gel extraction kit (NEB) according to the manufacturer's instructions. Elution of the plasmid DNA from the columns was done with 20-50 µL nuclease free water. The DNA concentration and the purity were measured photometrically at 260 nm, 230 nm and 280 nm. For further use the plasmids were stored at -20 °C.
PCR purification was performed using a PCR purification kit (Jena Biosciences) according to the instructions given by the supplier. Elution of the plasmid DNA from the columns was done with 20-50 µL nuclease free water. The DNA concentration and the purity were measured photometrically at 260 nm, 230 nm and 280 nm. For further use the plasmids were stored at -20 °C.

4.2.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis was used for the size-dependent separation of DNA samples. Therefore, 1 % agarose gels were casted by dissolving agarose in 1x TAE buffer using a microwave, followed by the addition of SERVA DNA Stain Clear G (Serva). DNA samples were mixed with 6x DNA loading dye (NEB). High range ladder (Jena Biosciences) was used as size standard. Electrophoreses were run at a constant voltage of 135 V for 30 min.

4.2.2.3 Polymerase chain reaction (PCR)

Polymerase chain reactions (PCRs) were used to amplify double stranded DNA as well as to generate random and saturation mutagenesis libraries. Polymerase chain reactions were typically carried out in a total volume of 50 µL containing 200 µM dNTP mix (NEB), 500 nM of forward and reverse primer, ca. 10 ng template, and 0.25 µL Q5 polymerase (NEB) in 1x Q5 reaction buffer. T_a was calculated using Tm Calculator v 1.12.0 (NEB).

Table 4.2.2.3.1: Protocol for PCR using Q5 polymerase.

Step	Temperature	time	# of cycles
Initial denaturation	98 °C	45 s	1
Denaturation	98 °C	10 s	32
Annealing	T_a [°C]	20 s	
Elongation	72 °C	30 s /kb	
Final elongation	72 °C	5 min	1
Storage	10 °C	∞	1

4.2.2.4 Sequencing

Sanger sequencing of plasmid DNA was performed to verify cloning. To this end, 5 µL DNA sample (50 – 100 ng/µL) was mixed with 5 µL of a 5 µM sequencing primer solution and sent to Genewiz (Genewiz) using the overnight premix service.

4.2.3 Microbiological Methods

4.2.3.1 Preparation of chemically competent *E. coli*

For the preparation of chemically competent *E. coli* a cryostock of the desired strain was inoculated into 50 mL 2x YT media. After overnight incubation at 37 °C (200 rpm), the pre-culture was diluted to an OD_{600} of 0.05 into 100 mL 2x YT followed by incubation at 37 °C (200 rpm) until an OD_{600} of 0.4 – 0.6 was reached. The culture was chilled on ice for 30 minutes and harvested by centrifugation (4000 g, 10 min, 4 °C). Thereafter, the resulting pellet was gently resuspended in 40 mL ice-cold 100 mM $CaCl_2$ and incubated on ice for 30 min. After incubation the pellet was harvested (4000 g, 10 min, 4 °C) and resuspended in 40 mL ice-cold 100 mM $CaCl_2$. After repeating this step twice, the cells were resuspended in 5 mL 100 mM $CaCl_2$, 15 % glycerol (v/v) and incubated on ice for 5 min. After incubation, the cell suspension was aliquoted into precooled 1.5 mL centrifugation tubes (100 µL/tube) and immediately flash frozen using liquid nitrogen. Competent *E. coli* cells were stored at -80 °C until further use. For the quality control of each batch of competent cells, the transformation efficiency (cfu/ug of transformed plasmid) was determined and the absence of contamination was assessed by cultivation of the batch on LB agar plates with/without different antibiotics.

4.2.3.2 Preparation of electrocompetent *E. coli*

Electrocompetent *E. coli* were prepared to reach a higher transformation efficiency which is sometimes needed for library cloning and transformation. For the preparation of electrocompetent *E. coli*, a cryostock of the desired strain was inoculated into 50 mL of 2xYT media. After overnight incubation at 37 °C (200 rpm), the pre-culture was diluted to an OD_{600} of 0.05 into 300 mL 2xYT followed by incubation at 37 °C (200 rpm) until an OD_{600} of 0.4 – 0.6 was reached. The culture was chilled on ice for 30 minutes and harvested by centrifugation (4000 g, 10 min, 4 °C). Thereafter, the resulting pellet was gently resuspended in 40 mL ice-cold 10 % glycerol (v/v) and incubated on ice for 30 min. After incubation the pellet was harvested (4000 g, 10 min, 4 °C) and resuspended in 40 mL ice-cold 10 % glycerol (v/v). After repeating this step twice, the cells were resuspended in 2 mL 10 % glycerol (v/v) and incubated on ice for 5 min. After incubation the cell suspension was immediately used for library electroporation or aliquoted into precooled 1.5 mL centrifugation tubes (100 µL/tube) and immediately flash frozen using liquid nitrogen. Competent *E. coli* cells were stored at -80 °C until further use. In the case of preparation of electrocompetent cells bearing a plasmid (e.g. for negative and positive selections) the growth medium contained the antibiotic (full strength) corresponding to the resistance gene of the plasmid. For the quality control of each batch of competent cells, the transformation efficiency (cfu/ug of transformed plasmid) was determined and the absence of contamination was assessed by cultivation of the batch on LB agar plates with/without different antibiotics.

4.2.3.3 Heat shock transformation of chemically competent *E. coli*

For heat shock transformation aliquots of chemically competent cells were thawed on ice followed by addition of 20 – 200 ng of the target plasmid DNA. In the case of co-transformations equimolar amounts of each plasmid DNA was added. For homogenization the tube was gently taped followed by incubation on ice for 30 min. Afterwards the cells were heat shocked for 45 sec in a water bath at 42 °C followed by incubation on ice for additional 5 min, before they were rescued with 900 µL of prewarmed SOC (or SOB) media. Thereafter, the cells

were incubated at 37 °C, 200 rpm for 1 h. Finally, cells were either inoculated into 2xYT supplemented with the corresponding antibiotic(s) or spun down, resuspended in 100 µL SOC and plated on LB agar plates supplemented with the corresponding antibiotic(s).

4.2.3.4 Transformation of electrocompetent *E. coli* via electroporation

For electroporation of *E. coli*, 200 – 2000 ng of the target plasmid DNA was added to freshly prepared competent cells or to an aliquot thawed on ice. For homogenization the tube was gently taped followed by incubation on ice for 30 min. Afterwards the cells were transferred to a precooled 2 mm gap width electroporation cuvette (Fisher Scientific). Afterwards electroporation was performed using an ECM399 electroporation system (BTX) employing an exponential decay wave pulse and 2.5 kV as settings. Immediately after the pulse, cells were rescued with 900 µL of prewarmed SOC (or SOB) media and transferred to a centrifugal tube. Thereafter, the cells were incubated at 37 °C, 200 rpm for 1 h. Finally, cells were inoculated into 2xYT supplemented with the corresponding antibiotic(s) and a dilution series was performed on LB agar plates supplemented with the corresponding antibiotics, in order to extrapolate the transformation efficiency

4.2.4 Mammalian cell culture

4.2.4.1 Maintenance of mammalian cells

HEK 293T cells (ATCC® CRL-3216™) were maintained in Dulbecco's Modified Medium – High Glucose (DMEM, D0822) (Sigma) with 10 % FBS (Gibco, U.S.) and 500 µg/ml PenStrep (Sigma Aldrich) at 37 °C in a humidified atmosphere with 5 % CO_2 in p100 dishes (Sarstedt). Cells were passaged every second (1:5 dilution) or every third (1:10 dilution) day. For this, the medium was carefully aspirated, and cells were washed with 10 ml sterile PBS. After 2 min incubation with 1 mL 1x Trypsin-EDTA (0.05%) solution (Gibco) at 37 °C, 5 % CO_2, cells were resuspended 10 ml pre-warmed medium and diluted 1:5 or 1:10 into new p100 dishes containing pre-warmed medium.

4.2.4.2 Seeding of cells

For transfection, a fixed cell number (500,000 for 6-well plates, 3,500,000 for p100 dishes) was seeded 24 h prior to transfection. To this end, cells were washed, and resuspended as described above (section Maintenance of mammalian cells). The number of cells was counted using a Countess II FL Automated Cell Counter system (Thermo Fisher) using the manufacturers protocol. Thereafter, cells were diluted in pre-warmed fresh medium and split into new dishes.

4.2.4.3 Transfection of cells

Transient transfection of HEK293T cells was performed using polyethyleneimine (PEI) (Sigma Aldrich)[377]. Therefore, all constructs to be transfected were diluted in Gibco™ Opti-MEM™ (Fisher Scientific) (see details table 4.1). After homogenization, PEI was added, and the mixture was vortexed for 2 s. Subsequently, samples were incubated at RT for 15 min. Meanwhile, the medium of the previously seeded cells was aspirated, and fresh medium supplemented with the

UAA was added. After incubation, the transfection mixtures were added dropwise to the fresh medium.

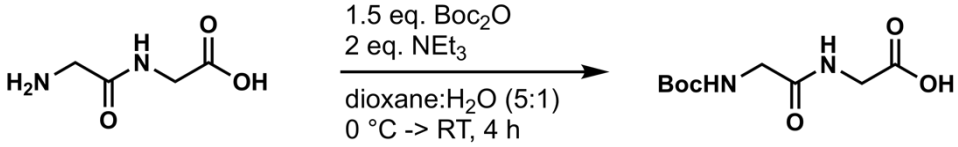
Table 4.2.4.3.1: PEI transfection conditions.

	6-well dish	p100 dish
Total transfection volume (µL)	200	1000
Total amount of plasmid DNA (µg)	2	8
PEI stock solution (1mg/mL)	9	27

4.3 Supporting information for Chapter 2

4.3.1 Chemical Synthesis

4.3.1.2 Synthesis of Boc-glycylglycine (Boc-Gly-Gly-OH)

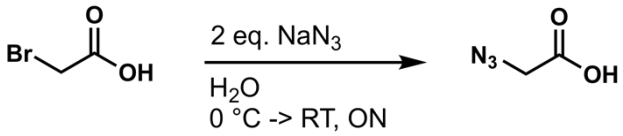


Diglycine (4.00 g, 0.030 mol) and Boc-anhydride (9.81 g, 0.045 mol) were dissolved in a mixture of dioxane:H_2O (5:1). Subsequently, triethylamine (8.30 ml, 0.060 mol) was added and the mixture was stirred for four hours in a melting ice-bath. Afterwards, 500 ml of H_2O were added and the mixture was acidified to pH 2 using 1 M HCl. The resulting solution was extracted three times with EtOAc. The organic layers were pooled, washed against brine, dried over Na_2SO_4 and filtered. Evaporation of the remaining solvent delivered Boc-diglycine as a white powder (6.57 g, 94 %). Boc-diglycine was stored at -20 °C until further use.

^{1}H-NMR (500 MHz, DMSO-d6): δ (ppm) 8.04 (t, J = 5.83 Hz, 1H, NH), 6.97 (t, J = 6.16 Hz, 1H, NH), 3.75 (d, J = 5.87 Hz , 2H, CH_2), 3.56 (d, J = 6.15 Hz, 2H, CH_2), 1.38 (s, 9H, 3 CH_3);

MS (ESI+) *m/z* 233.4 $[M+H]^+$
Calculated for $C_9H_{16}N_2O_5$: 233.1 $[M+H]^+$

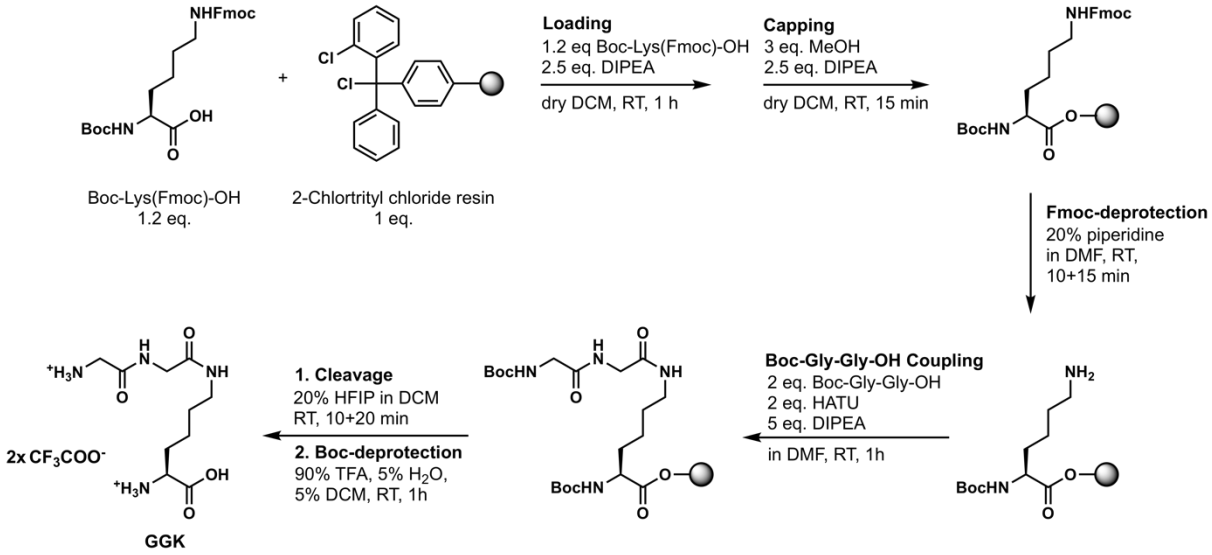
4.3.1.3 Synthesis of 2-azidoacetic acid[378]



Bromoacetic acid (17.25 g, 0.124 mol) was dissolved in 200 ml of H_2O and the solution was cooled to 0 °C in an ice bath while stirring. Afterwards, NaN_3 (16.12 g, 0.248 mol) was added to the solution and stirred overnight allowing the reaction mixture to reach RT. The reaction mixture was then acidified to pH 1 with 3 M HCl and extracted three times with Et_2O. The organic layers were pooled, washed with brine, dried over Na_2SO_4 and filtered. Evaporation of the remaining solvent delivered azidoacetic acid as a yellowish liquid (11.80 g, 95 %). Azidoacetic acid was stored at RT until further use.

^{1}H-NMR (500 MHz, $CDCl_3$): δ (ppm) 7.53 (br, 1H, COOH), 3.96 (s, 2H, CH_2)

4.3.1.4 Synthesis of N^6-glycylglycyl-L-lysine (GGK)



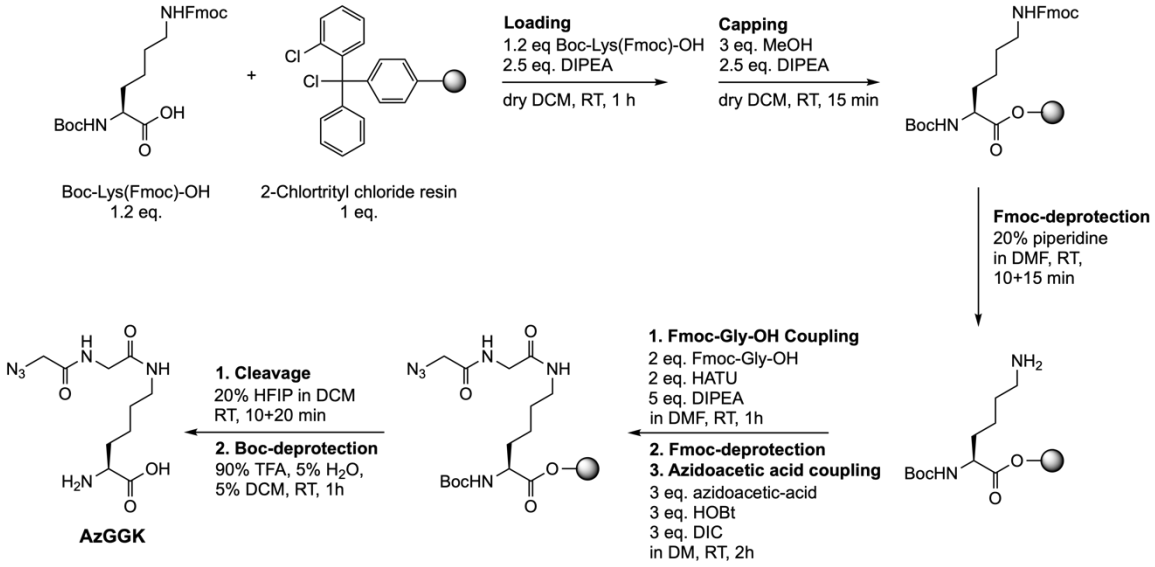
CTC-resin (10.0 g, 1 mmol/g maximal loading capacity) and Boc-Lys(Fmoc)-OH (5.616 g, 0.012 mol) were weighted into a customized glass apparatus with a frit. Subsequently, DIPEA (4.345 ml, 0.025 mol) in anhydrous DCM (100 ml) was added to the glass reactor followed by shaking at RT for 45 min. Afterwards, a solution of DIPEA (4.345 ml, 0.025 mol) and MeOH (1.215 ml, 0.030 mol) in anhydrous DCM (50 ml) was added to the glass reactor followed by shaking at RT for 15 min. The resin was then washed five times with DCM (10 ml/g resin) and five times with DMF (10 ml/g resin). For Fmoc-deprotection, 20% piperidine in DMF was added to the glass reactor followed by shaking at RT for 10 min. This procedure was repeated once for 15 min. Thereupon the glass reactor was washed five times with DMF (10 ml/g resin) and a solution of Boc-Gly-Gly-OH (4.640 g, 0.020 mol), HATU (7.600 g, 0.020 mol) and DIPEA (8.692 ml, 0.050 mol) in DMF (10 ml/g resin) was added. This mixture was allowed to shake at RT for 60 min followed by five times washing with DMF (10 ml/g resin) and five times washing with DCM (10 ml/g resin). For cleavage from the resin, 100 ml of 20 % HFIP in DCM were added to the glass reactor followed by shaking for 10 min at RT. This procedure was repeated, and the filtrates were combined in a round bottom flask. The solvent was evaporated under reduced pressure and the product was dissolved in a mixture of TFA/DCM/H_2O (90/5/5, v/v/v, 50 ml) for Boc-deprotection. The mixture was allowed to stir at RT for 1 h, after which the solvent was removed under reduced pressure to a small volume (< 10 ml). The product, which was still dissolved, was precipitated with 300 ml of ice-cold Et_2O in 50 ml centrifugal tubes. After centrifugation and washing of the precipitate with ice-cold Et_2O, the product was dissolved in H_2O and lyophilized, which afforded N^6-glycylglycyl-L-lysine (GGK) di(trifluoroacetate) as a yellowish solid (3.65 g, 75 %;calculated using the maximal loading capacity as 100 %). GGK was stored at -20 °C. Stock solutions were prepared by dissolving GGK in H_2O followed by neutralization with NaOH.

^{1}H-NMR (500 MHz, DMSO-d6): δ (ppm) 8.65 (t, *J* = 5.75 Hz, 1H, NH), 8.32 (s, 3H, NH_3), 8.14 (s, 3H, NH_3), 7.99 (t, *J* = 5.67 Hz, 1H, NH), 3.84 (dt, *J* = 4.8 Hz, 1H, CH), 3.74 (d, *J* =

5.73 Hz, 2H, CH_2), 3.61 (s, J = 4.58 Hz, 2H, CH_2), 3.04 (dt, J = 6.47 Hz, 2H, CH_2), 1.85 – 1.66 (m, 2H, CH_2), 1.48 – 1.34 (m, 2H, CH_2), 1.34 – 1.22 (m, 2H, CH_2);

MS (ESI+) *m/z* 261.3 $[M+H]^+$; Calculated for $C_{10}H_{21}N_4O_4^+$: 261.2 $[M+H]^+$

4.3.1.5 Synthesis of N^6-((2-azidoacetyl)glycyl)-L-lysine (AzGGK)



CTC-resin (10.0 g, 1 mmol/g maximal loading capacity) and Boc-Lys(Fmoc)-OH (0.012 mol, 5.616 g) were weighted into a customized glass apparatus with a frit. Subsequently DIPEA (0.025 mol, 4.345 ml) in anhydrous DCM (100 ml) was added to the glass reactor followed by shaking at RT for 45 min. Afterwards a solution of DIPEA (0.025 mol, 4.345 ml) and MeOH (0.030 mol, 1.215 ml) in anhydrous DCM (50 ml) was added to the glass reactor followed by shaking at RT for 15 min. Subsequently the resin was washed five times with DCM (10 ml/g resin) and five times with DMF (10 ml/g resin). For Fmoc-deprotection 20% piperidine in DMF (10 ml/g resin) was added to the glass reactor followed by shaking at RT for 10 min. This procedure was repeated once for 15 min. The glass reactor was washed five times with DMF (10 ml/g resin) and a solution of Fmoc-Gly-OH (0.020 mol, 5.94 g), HATU (0.020 mol, 7.60 g) and DIPEA (0.050 mol, 8.69 ml) in DMF (10 ml/g resin) was added. This mixture was shaken at RT for 1 h followed by five times washing with DMF (10 ml/g resin). After Fmoc-deprotection the glass reactor was washed five times with DMF (10 ml/g resin) and Azidoacetic acid (0.030 mol, 2.24 ml), HOBt (0.030 mol, 4.08 g) and DIC (0.030 mol, 4.64 ml) in DMF (10 ml/g resin) was added. This mixture was shaken at RT for 2 h followed by five times washing with DMF (10 ml/g resin) and five times with DCM (10 ml/g resin). For cleavage 100 ml of 20 % HFIP in DCM were added to the glass reactor followed by shaking for 10 min at RT. This procedure was repeated, and the filtrates were combined in a round bottom flask. The solvent was evaporated under reduced pressure and the product was dissolved in a mixture of TFA/DCM/H_2O (90/5/5, v/v/v, 50 ml) for Boc-deprotection. The mixture was stirred at RT for 1 h then the solvent was evaporated under reduced pressure to a small volume. The product was precipitated in 300 ml ice-cold Et_2O in 50 ml centrifugal tubes. After centrifugation and

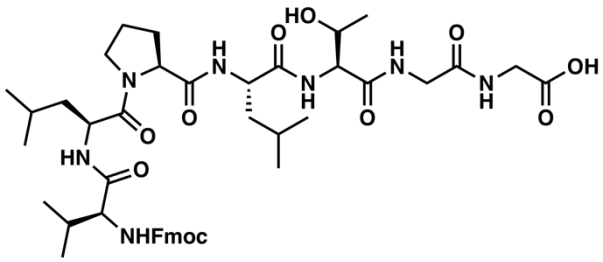
washing of the precipitate with ice-cold Et_2O the product was dissolved in H_2O and lyophilized, which delivered N^6-((2-azidoacetyl)glycyl)-L-lysine (AzGGK) di(trifluoroacetate) as a yellowish solid (3.45 g, 86 %; calculated using the maximal loading capacity as 100 %). AzGGK was stored at -20 °C. Stock solutions were prepared by dissolving AzGGK (100 mM) in H_2O, followed by neutralization with NaOH.

^{1}H-NMR (500 MHz, DMSO-d6): δ (ppm) 8.35 (t, *J* = 5.73 Hz, 1H, NH), 8.27 (s, 3H, NH_3), 7.94 (t, *J* = 5.63 Hz, 1H, NH), 3.88 (s, 2H, CH_2), 3.86 (bs, *J* = 4.8 Hz, 1H, CH), 3.70 (d, *J* = 5.75 Hz, 2H, CH_2), 3.05 (q, *J* = 6.40 Hz, 2H, CH_2), 1.84 – 1.66 (m, 2H, CH_2), 1.47 – 1.35 (m, 2H, CH_2), 1.34 – 1.22 (m, 2H, CH_2)

^{13}C-NMR (100 MHz, D_2O): δ (ppm) 175.0, 171.2, 170.9, 54.7, 51.8, 42.5, 38.8, 30.1, 28.0, 21.6

MS (ESI+) *m/z* 287.1 $[M+H]^+$; Calculated for $C_{10}H_{19}N_6O_4^+$: 287.1 $[M+H]^+$

4.3.1.6 Synthesis of (((9*H*-fluoren-9-yl)methoxy)carbonyl)-*L*-valyl-*L*-leucyl-*L*-prolyl-*L*-leucyl-*L*-threonylglycylglycine (Fmoc-VLPLTGG)



Fmoc-VLPLTGG

For the synthesis of the heptapeptide VLPLTGG, 0.166 g of CTC-resin (1.0 mmol/g maximal loading capacity) were used. The loading of the resin with Fmoc-Gly-OH and all the coupling steps with the according Fmoc-protected amino acids were performed as described in the standard Fmoc SPPS protocol above. After coupling of Fmoc-Val-OH, the product was cleaved from the resin by charging the syringe with a solution of 20% HFIP in DCM (v/v) followed by shaking at RT for 20 min. This procedure was repeated, and the filtrates were combined in a round bottom flask. The solvent was evaporated under reduced pressure and the product was dissolved in a mixture of TFA/DCM/H_2O (90/5/5, v/v/v) for *tert*-butyl ester deprotection. This solution was stirred at RT for 1 h then the solvent was evaporated under reduced pressure to a little volume. The remaining's were precipitated in ice-cold Et_2O in a centrifugal tube. After centrifugation and washing of the precipitate with ice-cold Et_2O the product was dissolved in H_2O and lyophilized, which delivered Fmoc-VLPLTGG as a white powder (0.101 g, 69 %; calculated using the maximal loading capacity as 100 %). Fmoc-VLPLTGG was stored at -20 °C and dissolved in DMSO prior to use.

MS (ESI+) *m/z* 878.6 $[M+H]^+$; Calculated for $C_{45}H_{64}N_7O_{11}^+$: 878.5 $[M+H]^+$

4.3.2 Library construction and directed evolution of AzGGKRS

4.3.2.1 Construction of a tailor-made library for AzGGK (lib-shuffle)

In order to create a new tailor-made PylRS library for AzGGK, a two-step process was conducted: Firstly, an error-prone PCR (epPCR) was performed to introduce random mutations in the C-lobe of the PylRS, and secondly, we shuffled 17 available PylRS variants that are known to accept bulky unnatural amino acids (see Table Table 4.3.2.1.1).

For the introduction of random mutations by epPCR into the 495 bp *M. barkeri* (*Mb*) wild type PylRS C-lobe we used the GeneMorph II kit (Agilent Technologies) with the Random_C-lobe_Megaprimer primer-pair (Table 4.3.2.1.2). Resulting megaprimer-amplicons were used for *Mb* wt PylRS amplification. Non-randomized plasmid copies were removed by DpnI digestion at 37 °C and 1,500 rpm for 2 hours, followed by chemical transformation into *E. coli* XL-1 gold cells as described in the GeneMorph II kit (Agilent Technologies). Transformants were grown overnight in 1 L of LB medium supplemented with kanamycin (50 µg/mL) at 37 °C while shaking at 200 rpm. Cells were harvested by centrifugation at 4,000 g and 20 °C for 10 minutes. Plasmids were isolated with a plasmid midi-prep kit (HiSpeed Plasmid Midi Kit, Qiagen) following manufacturer's instructions and stored at -20 °C.

Thereafter, seventeen carefully selected *Mb*PylRS derivatives (see Table Table 4.3.2.1.1) known to incorporate large and bulky unnatural amino acids, were subjected to DNA-shuffling in combination with the *Mb*PylRS error-prone PCR product. The 723 bp long *Mb*PylRS C-lobe fragment was amplified with the Shuffle_C-lobe primer pair (Table 4.3.2.1.2) and the Q5 High-Fidelity DNAPolymerase (NEB). 4 µg of the 18 equally mixed amplicon variants were digested with 0.25 U of DNAse I (Roche) for 5 minutes at 16 °C in 1x DNAase I reaction buffer (100 mM TrisHCl, 10 mM $MnCl_2$, pH 7.4). The reaction was stopped with 1x DNAse I stop solution (250 mM EDTA, 50% Glycerol, pH 8.2), followed by gel purification of the 50-500 bp DNA fragments with the QIAquick Gel Extraction kit (Qiagen). DNA fragments were reassembled by self-priming PCR followed by amplification with the Shuffle_PylRS_Clobe primer-pair (Table 4.3.2.1.2) and the Q5 High-Fidelity DNA Polymerase (NEB). The PCR product was cloned into the wt PylRS backbone by restriction digest with PstI and BsteII (NEB) restriction enzymes for 2 hours at 37 °C, followed by overnight ligation at 16 °C with T4 Ligase (NEB) and transformation into electrocompetent *E. coli* DH10β cells at 2.0 kV, 200 Ohm, and 25 µF (BioRad MicroPulser). Cells were recovered in 1 mL SOB medium for one hour at 37 °C and 200 rpm. Recovered cells were grown overnight in 1 L LB supplemented with kanamycin (50 µg/mL) at 37 °C and 200 rpm. Cells were harvested by centrifugation at 4,000 g and 20 °C while the plasmids were isolated with a plasmid midi-prep kit (HiSpeed Plasmid Midi Kit, Qiagen), and stored at -20°C. Library depth was calculated by dilution experiments and verified by sequencing, using the PylRS_seq primer-pair (Table 4.3.2.1.2).[328,379]

Table 4.3.2.1.1 *Mb* PylRS derivatives used for DNA-shuffling.

PylRS Variant #	A266	AA270	AA271	AA274	AA311	AA313	AA349	AA366	AA382
1			Y271A			C313V			
2		L270I	Y271F	L274A	N311A	C313V			
3		L270L	Y271L	L274F	N311A	C313M			
4		L270S	Y271G	L274S	N311D	C313A			
5		L270S	Y271G	L274L	N311G				
6	L266M	L270I	Y271F	L274A	C311F				
7		L270I	Y271F	L274G		C313F	Y349F		
8			Y271A	L274M		C313A			
9			Y271M	L274G		C313A	Y349W		
10						C313I			
11			Y271L	L274F	N311M	C313G			
12					N311M	C313Q		V366G	W382N
13					N311Q	C313S		V366G	W382N
14					N311Q	C313A		V366G	W382N
15			Y271M	L274A		C313A	Y349F		
16					N311M	C313Q	Y349F	V366G	W382N
17			Y271L	L274A		C313F			
18	Error-Prone PCR results of 495 base-pair PylRS C-lobe (G223 to F387)								

Table 4.3.2.1.2 Primer used for library construction and sequencing

Primer	Sequence (5' - 3')
Random_C-lobe_Megaprimer for	GATATCACCAAATTTTTTGTGGATCGC
Random_C-lobe_Megaprimer rev	CTTTCAGCAGACGTTCCAGGCC
Shuffle_Clobe for	GAACCGGAACTGGTGACCCGTCG
Shuffle_Clobe rev	CGTTTGAAACTGCAGTTACAGGTTCGTGC
Shuffle_PylRS_Clobe for	GAACCGGAACTGGTGACCCGTCG
Shuffle_PylRS_Clobe rev	CGTTTGAAACTGCAGTTACAGGTTCGTGC
PylRS_seq for	AGTTGAAGGATCCTCGG
PylRS_seq rev	CCTACAAAAGCACGCAAAC

4.3.1.2 Directed Evolution for AzGGKRS

M. barkeri PylRS evolution was performed by positive and negative selection,[329,380] followed by a second positive selection coupled to a reporter-sfGFP-150TAG-readout.[381,382] The positive selection plasmid pRep_PylT encodes a tetracycline resistance cassette and a constitutively expressed chloramphenicol acetyltransferase gene, bearing an amber codon at position 111 and a constitutively expressed Mb-pyrrolysyl-$tRNA_{CUA}$ (PylT). The dual-reporter plasmid psfGFP_150TAG_CAT_111TAG_PylT additionally encodes an sfGFP-150TAG. The negative selection plasmid pYOBB_PylT encodes a chloramphenicol resistance cassette, a constitutively expressed PylT and an L-arabinose-inducible barnase gene, interrupted by amber codons at positions 3 and 45. In the first positive selection step, 3 µg of lib-shuffle library were transformed by electroporation into 100 µL of freshly prepared *E. coli* DH10β cells containing the positive selection plasmid pRep_PylT with an electroporator (BioRad Micropulser) using a 0.2 cm Micropulser electroporator cuvette (Biorad) at 2.0 kV, 200 Ohm, and 25 µF and rescued with 1 mL of SOC medium for 1 hour at 37 °C and 200 rpm. The cell suspension was incubated overnight at 37 °C and 200 rpm in 1 L of 2xYT-medium with kanamycin (50 µg/mL) and tetracycline (12.5 µg/mL). Next morning, the cell suspension was diluted in 500 mL 2xYT-medium with kanamycin (25 µg/mL) and tetracycline (6.25 µg/mL) to reach an $OD_{600} < 0.1$, followed by incubation at 200 rpm and 37 °C until reaching an $OD_{600} < 0.3$. 10 mL of the cell suspension were transferred into a 50 mL centrifugation tube, 2 mM AzGGK were added to the cells and incubated for 4 hours at 37 °C. 600 µL of AzGGK-containing culture were plated on 24 cm x 24 cm plates containing 200 mL of GMML agar with chloramphenicol (60 µg/mL), tetracycline (12.5 µg/mL) and kanamycin (50 µg/mL), as well as 2 mM AzGGK. Plates were incubated for 36-48 h aiming for a single-clone distribution on the plates. Surviving bacterial clones were scraped off from the plates in the presence of 25 mL LB-medium with kanamycin (25 µg/mL) and tetracycline (6.25 µg/mL), followed by a two-hour incubation at 200 rpm and 37 °C to remove agar residuals and amplify low-abundant bacterial clones. The bacterial suspension was pelleted for 20 minutes at 3,000 g and 4 °C and subjected to plasmid isolation by HiSpeed Midiprep kit (Qiagen). 6 µg of the isolated plasmid DNA were separated by agarose gel electrophoresis while the PylRS-encoding plasmid was cut out and purified by gel extraction (Qiagen). For negative selection, 50 ng of the lib-shuffle-DNA obtained from the surviving positive selection clones were transformed into freshly prepared electrocompetent *E. coli* DH10β cells containing the negative selection plasmid pYOBB_PylT. 2 x 500 µL of the SOC-rescued culture were plated onto two square 24 cm x 24 cm plates containing 200 ml of LB-agar, kanamycin (50 µg/mL) and chloramphenicol (50 µg/mL), as well as 0.2 % L-arabinose. Plates were incubated for 36-48h at 37 °C, aiming for single-clone distribution on the plates. Surviving bacterial clones were scraped off from the plate in the presence of 25 mL LB-medium with kanamycin (25 µg/mL) and chloramphenicol (25 µg/mL), followed by a two-hour incubation at 200 rpm and 37 °C to remove agar residuals and amplify low-abundant bacterial clones. The bacterial suspension was pelleted for 20 minutes at 3,000 g and 4 °C and subjected to plasmid isolation by HiSpeed midiprep (Qiagen). 6 µg of the isolated plasmid-DNA were separated by agarose gel electrophoresis, the PylRS-encoding plasmid band was cut out and purified by gel extraction (Qiagen). For dual-positive selection and reporter-readout, 50 ng isolated lib-shuffle-DNA of the surviving clones from the negative selection were transformed into freshly prepared electrocompetent *E. coli* DH10β containing the dual-reporter plasmid

psfGFP150TAG_CAT111TAG_PylT. All co-transformants were plated on a single 24 cm x 24 cm autoinduction agar plate, containing 0.2 % L-arabinose, kanamycin (50 µg/mL), as well as tetracycline (12.5 µg/mL), chloramphenicol (25 µg/mL) and 2 mM of AzGGK. Plates were incubated for 48 hours at 37 °C. sfGFP-expressing colonies were picked into 96-deep-well plates containing 1 mL of non-inducing media with kanamycin (50 µg/mL) and tetracycline (12.5 µg/mL), followed by 48 hours incubation at 200 rpm and 37 °C. 50 µL of the cell suspension were transferred into two new 96-deep-well plates containing 1 mL of autoinducing medium with kanamycin (50 µg/mL) and tetracycline (12.5 µg/mL), as well as one of them containing 2 mM AzGGK, followed by 48 hours of incubation at 37 °C (200 rpm). The residual 96-deep-well plate with non-inducing medium was centrifuged at 4.000 g and 4 °C, followed by storage at -20 °C. 20 µl of the cell suspension were transferred into corning 96-well-plates with a clear bottom and mixed with 180 µL of 1x PBS. The fluorescence-intensity for sfGFP-150TAG expression was measured for both plates incubated with and without AzGGK in a Tecan plate reader (excitation: 480nm; emission: 527nm). The OD_{600} was used to normalize the sfGFP-150TAG expression within each well. PylRS variants from cells showing at least 5-fold fluorescence intensity increase (as compared to cells grown in the absence of AzGGK) were isolated by agarose gel electrophoresis, purified by gel extraction and sequenced. For further expressions, the AzGGRS was recloned into a pBK-backbone with ampicillin resistance.

4.3.3 Protein expression and purification

4.3.3.1 Expression and purification of sortase mutants

Chemical competent *E. coli* BL21(DE3) were transformed with pET29-sortase-His6 plasmid (Addgene). After recovery with 1 mL of SOC medium for one hour at 37 °C, the cells were cultured overnight in 50 mL of LB medium containing kanamycin (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 1 L of fresh 2xYT medium supplemented with kanamycin (25 µg/mL) and cultured at 37 °C while shaking (200 rpm) until an OD_{600} = 0.5-0.8 was reached. IPTG was added to a final concentration of 0.4 mM and protein expression was induced for three hours at 30 °C. The cells were harvested by centrifugation (4,000xg, 15 min, 4 °C) and re-suspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl supplemented with 1 mM $MgCl_2$, 0.1 mg/mL DNAse I, one cOmplete™ protease inhibitor tablet (Roche) and 0.175 mg/mL PMSF). Cells were lysed by sonication, centrifuged (15,000 xg, 40 min, 4 °C), and the cleared lysate was added to 2 mL Ni^{2+}-NTA slurry (Jena Bioscience) and the mixture was incubated with agitation for one hour at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV of wash buffer (20 mM Tris pH 8.0, 30 mM imidazole pH 8.0 and 300 mM NaCl). The protein was eluated in 1 mL fractions with wash buffer supplemented with 300 mM imidazole pH 8.0. The fractions containing the protein were pooled together, concentrated and rebuffered (20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$) with Amicon® Ultra-4 10K MWCO centrifugal filter units (Millipore). Enzyme concentration was calculated from the measured A280 absorption (extinction coefficients were calculated with ProtParam (https://web.expasy.org/protparam/)). All Sortase variants were stored at 4 °C for further use.

Sortase variants harboring a C-terminal TEV-His-tag were expressed and purified identically. In order to cleave off the His-Tag, the fractions containing the protein were pooled together and 200 µL of TEV protease (1.8 mg/mL) was added. The mixture was transferred to a dialysis tubing (Roth) and the dialysis bag was immersed in 2 L of cold dialysis buffer (25 mM Tris pH 8.0, 150 mM NaCl, 2 mM DTT) and stirred at 4 °C overnight. The protein mixture was recovered from the dialysis tubing and centrifuged (15,000xg, 10 min, 4 °C) in order to precipitate the TEV protease. 2 mL of Ni^{2+}-NTA slurry (Jena Bioscience) were added to the supernatant and the mixture was incubated with agitation for one hour at 4 °C. The mixture was then poured into an empty plastic column and the flow-through was collected. The Ni^{2+}-NTA beads were washed twice with 15 mL of wash buffer (20 mM Tris pH 8.0, 150 mM NaCl and 5 mM $CaCl_2$). Flow-through and wash fractions containing pure Sortase without His-Tag were pooled together, concentrated and rebuffered (20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$) with Amicon® Ultra-4 10K MWCO centrifugal filter units (Millipore). Enzyme concentration was calculated from the measured A280 absorption (extinction coefficients were calculated with ProtParam (https://web.expasy.org/protparam/)). All Sortase variants were stored at 4 °C for further use.

4.3.3.2 Expression and purification of wt Ub and Ub mutants

Chemical competent *E. coli* Rosetta2 (DE3) were transformed with pET17-Ub plasmid. After recovery with 1 mL of SOC medium for one hour at 37 °C, the cells were cultured overnight in 50 mL of LB medium containing ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 1 L of fresh 2 xYT medium supplemented with ampicillin (50 µg/mL) and chloramphenicol (25 µg/mL) and cultured at 37 °C with shaking (200 rpm) until OD_{600} = 0.8-1.0. IPTG was added to a final concentration of 1 mM and protein expression was induced for 4 hours at 37 °C. The cells were harvested by centrifugation (4,000xg, 15 min, 4 °C) and resuspended in lysis buffer (50 mM Tris pH 7.6, supplemented with 10 mM MgCl2, 1 mM EDTA, 0.1 % NP-40, 0.1 mg/mL DNAseI, one cOmplete™ protease inhibitor tablet and 0.175 mg/mL PMSF). Cells were lysed by sonication and centrifuged (15,000 xg, 40 min, 4 °C).
The cleared lysate was transferred into a glass beaker in an ice-bath that was placed on a magnetic stirrer. Precipitation was performed with 35 % perchloric acid until pH 4.0 – 4.5 was reached. After 5-minute incubation at 4 °C while stirring, the milky solution was centrifuged (15,000xg, 40 min, 4 °C) and the supernatant was transferred into a dialysis tubing with a MWCO of 2 kDa. Dialysis was performed over night at 4 °C with 50 mM Ammonium acetate buffer pH 4.5. The dialyzed solution was centrifuged (15,000xg, 40 min, 4 °C), filtered and purified via a HiTrap SP FF 5 mL cation exchange chromatography (GE, gradient 0 – 1 M NaCl). Fractions that were > 95 % purity, as judged by SDS-PAGE, were pooled and rebuffered (20 mM Tris pH 8.0, 150 mM NaCl, 5 mM $CaCl_2$) with Amicon® Ultra-15 3kDa MWCO centrifugal filter units (Millipore). Enzyme concentration was calculated from the measured A280 absorption (extinction coefficients were calculated with ProtParam (https://web.expasy.org/protparam/)). All purified Ub variants were stored at 4 °C for further use.

4.3.3.3 PCNA expression and purification

Chemical competent *E. coli* K12 cells were co-transformed with pPylT_PCNA-K164TAG-CPD-His6 (which encodes *Mb*tRNA$_{CUA}$ and a C-terminally His6-tagged PCNAK164TAG-CPD-His6 fusion protein. CPD is a cysteine protease domain of the *Vibrio cholerae* MARTX toxin)[383] and pBK_AzGGKRS (which encodes AzGGKRS) plasmids. After recovery with 1 mL of SOC medium for 1 h at 37 °C, the cells were cultured overnight in 50 mL of non-inducing medium containing full strength antibiotics (tetracycline and ampicillin), at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in autoinduction medium supplemented with full strength antibiotics and 4 mM AzGGK. After incubation overnight at 37 °C the cells were harvested by centrifugation (4000 xg, 15 min, 4 °C), flash frozen in liquid nitrogen and stored at -80 °C. The composition of non-inducing and autoinducing media can be found in the Appendix, Supplementary Tables S3-S7.
The pellet was thawed on ice and re-suspended in lysis buffer (20 mM Tris pH 7.5, 0.5 mM PMSF, 300 mM NaCl, 0.2 % NP-40, 0.2 mg/mL Lysozyme). The cell suspension was incubated on ice for 30 minutes and sonicated at 4 °C on ice. The lysed cells were centrifuged (15,000xg, 30 min, 4 °C), the cleared lysate added to Ni^{2+}-NTA slurry (Jena Bioscience) (0.2 mL of slurry per 100 mL of culture) and the mixture was incubated with agitation for 1 h at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV wash buffer (20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$). In order to induce auto cleavage of the CPD the slurry was incubated with 1 CV elution buffer 1 (20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$ and 1 mM IP6) for 1 h at 4 °C. Afterwards PCNA was eluated in 1 mL fractions with elution elution buffer 1. The column bound CPD-His6 was eluted afterwards with elution buffer 2 (20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$ and 300 mM imidazole). The fractions containing PCNA-K164AzGGK (identified via 15% SDS-PAGE) were pooled together, concentrated and rebuffered (20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$) with Amicon® Ultra-4 10K MWCO centrifugal filter units (Millipore). Protein concentration was calculated from the measured A280 absorption (extinction coefficients were calculated with ProtParam (https://web.expasy.org/protparam/)). PCNA was stored at 4 °C for further use.

4.3.3.4 Expression and Purification of His-TEV-SUMO1 variants and His-TEV-Ub variants

Chemical competent *E. coli* Rosetta2 (DE3) were transformed with pET17-His6-TEC-Ub/SUMO1 plasmid. After recovery with 1 mL of SOC medium for 1 hour at 37 °C, the cells were cultured overnight in 50 mL of LB medium containing ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 1 L of fresh 2 xYT medium supplemented with ampicillin (50 µg/mL) and chloramphenicol (25 µg/mL) and cultured at 37 °C, 200 rpm, until OD_{600} = 0.8-1.0. IPTG was added to a final concentration of 1 mM and protein expression was induced for 4 h at 37 °C. The cells were harvested by centrifugation (4,000xg, 15 min, 4 °C) and resuspended in lysis buffer (50 mM Tris pH 7.6, supplemented with 10 mM $MgCl_2$, 1 mM EDTA, 0.1 % NP-40, 0.1 mg/mL DNAseI, one cOmplete™ protease inhibitor tablet and 0.175 mg/mL PMSF). Cells were lysed by sonication and centrifuged (15,000xg, 40 min, 4 °C).

The obtained cell pellets were resuspended in 20 mL of lysis buffer (20 mM Tris pH 8.0, 30 mM imidazole, 300 mM NaCl, 0.175 mg/mL PMSF, 0.1 mg/mL DNase I and one cOmplete™ protease inhibitor tablet (Roche)). The cell suspension was incubated on ice for 30 minutes and sonicated at 4 °C on ice. The lysed cells were centrifuged (15,000xg, 40 min, 4 °C), the cleared lysate added to Ni^{2+}-NTA slurry (Jena Bioscience) (0.2 mL of slurry per 100 mL of culture) and the mixture was incubated with agitation for 1 h at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV of wash buffer (20 mM Tris pH 8.0, 30 mM imidazole pH 8.0 and 300 mM NaCl). The protein was eluated in 1 mL fractions with wash buffer supplemented with 300 mM imidazole pH 8.0. The fractions containing the protein were pooled together, concentrated and rebuffered (20 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$) with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). Purified proteins were analyzed by 15 % SDS-PAGE and mass spectrometry. Ub and SUMO1 variants were stored at 4 °C until further use.

4.3.3.5 Expression and purification of GST-Tab2-NZF and GST-hHR23A-UBA2

Chemical competent *E. coli* Rosetta2 (DE3) were transformed with pGEX-5X-1 GST-Tab2-NZF plasmid (which was kindly provided by the Vermeulen lab)[339] or pGEX-4T-1 hHR23A-UBA2 plasmid. After recovery with 1 mL of SOC medium for one hour at 37 °C, the cells were cultured overnight in 50 mL of LB medium containing ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 500 mL of fresh 2 xYT medium supplemented with ampicillin (50 µg/ml) and chloramphenicol (25 µg/mL) and cultured at 37 °C with shaking (200 rpm) until OD_{600} = 0.8. IPTG was added to a final concentration of 0.5 mM and protein expression was induced for 18 hours at 16 °C (In case of pGEX-5X-1 GST-Tab2-NZF expression 200 µM $ZnSO_4$ was added to the 2xYT medium). The cells were harvested by centrifugation (4,000xg, 15 min, 4 °C) and resuspended in lysis buffer (1x PBS supplemented with 1 mM DTT, 0.1 mg/mL DNAseI, one cOmplete™ protease inhibitor tablet and 0.175 mg/mL PMSF). The cell suspension was incubated on ice for 30 minutes and sonicated at 4 °C on ice. The lysed cells were centrifuged (15,000xg, 40 min, 4 °C), the cleared lysate added to Glutathione Sepharose 4B (GE Healthcare) (0.1 mL of slurry per 100 mL of culture) and the mixture was incubated with agitation for 1 h at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV of wash buffer (1x PBS supplemented with 1 mM DTT). The protein was eluated in 500 µl fractions with wash buffer supplemented with 20 mM glutathione. The fractions containing the protein were pooled together, concentrated and rebuffered (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT and 0.1% NP-40) with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). Purified proteins were analyzed by 15 % SDS-PAGE. GST-Tab2-NZF and GST-hHR23A-UBA2 were stored at -80 °C until further use.

4.3.3.6 Expression and purification of His-Rap80 7A

Rap80 7A tandem UIM peptide[340] synthesized by GeneArt (Thermo Fisher) was cloned in-frame between the XbaI and NotI sites of a pET28a vector. Chemical competent *E. coli* Rosetta2 (DE3) were transformed with pET28a-Rap80-7A plasmid. After recovery with 1 mL

of SOC medium for one hour at 37 °C, the cells were cultured overnight in 50 mL of LB medium containing kanamycin (50 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 500 mL of fresh 2 xYT medium supplemented with kanamycin (25 µg/mL) and chloramphenicol (25 µg/mL) and cultured at 37 °C with shaking (200 rpm) until OD_{600} = 0.8. IPTG was added to a final concentration of 0.5 mM and protein expression was induced for 4 hours at 37 °C. The cells were harvested by centrifugation (4,000xg, 15 min, 4 °C) and resuspended in lysis buffer (20 mM Tris pH 8.0, 30 mM imidazole, 300 mM NaCl, 0.175 mg/mL PMSF, 0.1 mg/mL DNase I and one cOmplete™ protease inhibitor tablet (Roche)). The cell suspension was incubated on ice for 30 minutes and sonicated with cooling in an ice-water bath. The lysed cells were centrifuged (15,000 x g, 40 min, 4 °C), the cleared lysate was added to Ni^{2+}-NTA slurry (Jena Bioscience) (0.2 mL of slurry per 100 mL of culture) and the mixture was incubated with agitation for one hour at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV wash buffer (20 mM Tris pH 8.0, 30 mM imidazole pH 8.0 and 300 mM NaCl). The protein was eluated in 1 mL fractions with elution buffer (wash buffer supplemented with 300 mM imidazole pH 8.0). The fractions containing Rap80 7A were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore) followed by size-exclusion chromatography (SEC) using a Superdex S75 16/600 (GE Healthcare) with 1x PBS. Fractions containing the Rap80 7A were pooled together and concentrated with Amicon® 3 kDa MWCO centrifugal filter units (Millipore). Rap80 7A was stored at -80 °C until further use.

4.3.3.7 Expression and purification of E1 enzyme His-UBE1

Chemical competent *E. coli* Rosetta2 (DE3) were transformed with pET21d His-UBE1.[384] After recovery with 1 mL of SOC medium for one hour at 37 °C, the cells were cultured overnight in 50 mL of LB medium containing ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 1 L of fresh 2xYT medium supplemented with ampicillin (50 µg/ml) and chloramphenicol (25 µg/mL) and cultured at 37 °C with shaking (200 rpm) until OD_{600} = 0.6. IPTG was added to a final concentration of 0.5 mM and protein expression was induced for 20 hours at 16 °C. The cells were harvested by centrifugation (4,000xg, 15 min, 4 °C) and resuspended in lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.1% (w/v) Triton X-100, 1 mM EDTA, 1 mM dithiothreitol (DTT), 0.175 mg/mL PMSF, 0.1 mg/mL DNase I and one cOmplete™ protease inhibitor tablet (Roche)). The cell suspension was incubated on ice for 30 minutes and sonicated with cooling in an ice-water bath. The lysed cells were centrifuged (15,000 xg, 40 min, 4 °C), the cleared lysate was added to Ni^{2+}-NTA slurry (Jena Bioscience) (0.2 mL of slurry per 100 mL of culture) and the mixture was incubated with agitation for one hour at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV wash buffer (20 mM Tris pH 8.0, 30 mM imidazole pH 8.0 and 300 mM NaCl). The protein was eluated in 1 mL fractions with wash buffer supplemented with 300 mM imidazole pH 8.0. The fractions containing His-UBE1 were pooled together, concentrated and rebuffered (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 1 mM DTT) with Amicon® with the corresponding MWCO centrifugal filter units (Millipore) followed by anion-exchange chromatography using a MonoQ (GE, gradient 0-500 mM NaCl). Fractions containing His-UBE1 were pooled together and concentrated followed

by size exclusion chromatography (SEC) using a Superdex 200 16/600 (GE Healthcare) with 20 mM Tris pH 8.0, 100 mM NaCl,1 mM EDTA, 1 mM DTT). The fractions containing pure His-UBE1 were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). His-UBE1 was stored at -80 °C until further use.

4.3.3.8 Expression and purification of E2 enzymes Ubc13, Uev1A and CDC34

Chemical competent *E. coli* Rosetta2 (DE3) were transformed with pGEX-6P-1-UBE2N, pGEX-6P-1-UBE2V1 or pGEX-6P-1-UBE2R1. After recovery with 1 mL of SOC medium for one hour at 37 °C, the cells were cultured overnight in 50 mL of LB medium containing ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 500 mL of fresh 2 xYT medium supplemented with ampicillin (50 µg/mL) and chloramphenicol (25 µg/mL) and cultured at 37 °C with shaking (200 rpm) until OD_{600} = 0.8. IPTG was added to a final concentration of 0.25 mM and protein expression was induced for 18 hours at 20 °C. The cells were harvested by centrifugation (4,000xg, 15 min, 4 °C) and resuspended in lysis buffer (50 mM TRIS pH 8.0, 270 mM sucrose, 50 mM NaF, 1 mM DTT, 0.1 mg/mL DNase I and one cOmplete™ protease inhibitor tablet (Roche)). The cell suspension was incubated on ice for 30 minutes and sonicated at 4 °C on ice. The lysed cells were centrifuged (15,000xg, 40 min, 4 °C), the cleared lysate added to Glutathione Sepharose 4B (GE Healthcare) (0.1 mL of slurry per 100 mL of culture) and the mixture was incubated with agitation for 1 h at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV of wash buffer (25 mM Tris pH 8.5, 400 mM NaCl, 5 mM DTT). For site-specific cleavage of the GST tag, immobilized fusion proteins were incubated with PreScission protease (GE Healthcare) over night. Cleaved proteins were eluted with wash buffer and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). Purified proteins were analyzed by 15 % SDS-PAGE. Ubc13, Uev1A and CDC34 were stored at -80 °C until further use.

4.3.4 Ubiquitylation and SUMOylation of GGK-bearing proteins

4.3.4.1 Preparation of diUbs

Ub-GGK acceptor ubiquitins with GGK at positions K6, K11, K33, K48 and K63 were diluted to 20 µM in sortase buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$). Afterwards 100 µM of the donor Ub (either Ub(AT) or Ub(LAT)) was added followed by the addition of 20 µM Srt2A (without His-Tag). Incubation was performed at 37 °C, 600 rpm, for one hour when Ub(LAT) was used and for 18 hours when Ub(AT) was used. Sortase-mediated transpeptidation was stopped by the addition of 200 µM phenylvinylsulfon and further incubation for 10 minutes at 37 °C, 600 rpm.
Afterwards Ni^{2+}-NTA slurry (Jena Bioscience) (0.1 mL/mg of acceptor Ub) was added to the reaction mixture and the mixture was incubated agitating for 1 h at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 40 CV of wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$, 30 mM imidazole) to remove Srt2A and the excess of donor Ub. The protein was eluated in 0.2 mL fractions with wash buffer supplemented with 300 mM imidazole pH 8.0. The fractions containing the mixture of diubiquitin and unreacted acceptor Ub were pooled together and concentrated with Amicon® with the

corresponding MWCO centrifugal filter units (Millipore). In order to remove unreacted acceptor Ub, size-exclusion chromatography (SEC) was performed using a Superdex S75 16/600 (GE Healthcare) with sortase buffer. Fractions containing the diubiquitn were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). Diubiquitins were stored at 4 °C until further use.

4.3.4.2 Preparation of ubiquitylated PCNA

PCNA-K164GGK was diluted to 5 µM in sortase buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$). Afterwards 100 µM of His-TEV-Ub(AT) and 20 µM Srt2A (without His-Tag) was added and the mixture incubated at 25 °C (600 rpm) for 44 hours. Sortase mediated transpeptidation was stopped by the addition of 200 µM phenylvinylsulfon and further incubation for 10 minutes at 25 °C, 600 rpm.
After incubation, the mixture was subjected to Ni-NTA affinity chromatography using a His-Trap FF Column (GE Life Sciences) to remove Srt2A and unreacted PCNA. The fractions containing ubiquitylated PCNA were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore).
In order to remove the excess of His-TEV-Ub(AT), size-exclusion chromatography (SEC) was performed using a Superdex S75 16/600 (GE Healthcare) with SEC-buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM TCEP, 5 % Glycerol (w/v)). Fractions containing pure ubiquitylated PCNA were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). Protein concentration was calculated from the measured A280 absorption (extinction coefficients were calculated with ProtParam (https://web.expasy.org/protparam/)). Ubiquitylated PCNA was stored at 4 °C until further use.

4.3.4.3 Peptide-based sortase Assays

Fmoc-VLPLTGG (20 mM stock solution in DMSO) was diluted in sortase buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$) to a final concentration of 1 mM followed by the addition of 10 mM AzGGK (50 mM stock in H_2O) or GGK (50 mM stock in H_2O). Subsequently Srt5M (300-800 µM stock in sortase buffer) was added to a final concentration of 20 µM. Incubation was performed at 37 °C (600 rpm). Samples were taken at the denoted time points by quenching the reaction mixture with 10 volumes of 0.5 % formic acid prior to HPLC-MS analysis. Typical reaction volumes were 50 µL.

4.3.5 Preparation of natively linked diubiquitins

4.3.5.1 Assembly of natively linked K48 and K63 Diubiquitins

Assembly of K48- and K63-linked diubiquitins was carried out in 1-mL reactions.[385] In order to assemble K63-linked diubiquitins the reaction mixture contained 250 nM His-UBE1, 4 µM Ubc13, 4 µM Uev1a, 1.4 mM Ub, 10 mM ATP, 40 mM Tris (pH 7.5), 10 mM MgCl2 and 0.6 mM DTT and was incubated at 37 °C for 5 h. For the assembly of natively linked K48 diubiquitin the reaction mixture contained 250 nM His-UBE1, 4 µM CDC34, 1.4 mM Ub, 1 mM ATP, 40 mM Tris (pH 7.5), 10 mM MgCl2 and 0.6 mM and was incubated at 37 °C for

5 h. After incubation, 20 mL of 50 mM ammonium acetate (pH 4.5) were added to the reaction mixture in order to precipitate enzymes. Afterwards the solution was filtered and the assembled diubiquitins were purified by cation exchange using a Resource S column (GE Healthcare, gradient 0-500 mM NaCl). Fractions containing pure diubiquitins were pooled together and concentrated with Amicon® 3 kDa MWCO centrifugal filter units (Millipore). Diubiquitins were stored at -80 °C until further use.

4.3.6 Enzymatic assays

4.3.6.1 Deubiquitylation assay of diubiquitins[129]

$USP2_{CD}$ (100 ng, Boston Biochem) was diluted into DUB dilution buffer (25 mM Tris pH 7.5, 150 mM NaCl, 10 mM DTT) and incubated at room temperature for 10 minutes to activate the enzyme. 2 µg of native diubiquitin (UbiQBio) or sortase-generated diubiquitin were added to 3 µL 10x DUB buffer (500 mM Tris pH 7.5, 500 mM NaCl, 50 mM DTT) and constituted to 20 µL with H_2O. Afterwards 10 µL of the activated DUB was added to diubiquitin samples followed by incubation at 37 °C. 6 µL samples were taken at the denoted time points and quenched by the addition of 4x SDS loading buffer and boiling at 95 °C for 10 minutes. Samples were loaded on SDS-PAGE gels and visualized by Coomassie staining.

4.3.6.2 Deubiquitylation assay of ubiquitylated PCNA

Natively ubiquitylated PCNA (a kind gift from Christian Biertümpfel, MPI Martinsried) or sortase-generated PCNA-Ub(AT) conjugate was diluted to 1 µM into DUB buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 1 mM EDTA). UAF1 and USP1 (Boston Biochem) were added in equimolar ratio to a final concentration of 100 nM to the ubiquitylated PCNA conjugates in DUB buffer. The mixture was incubated at 37 °C. 10 µL samples were taken at the denoted time points and quenched by the addition of 4x SDS loading buffer and boiling at 95 °C for 5 minutes. Samples were loaded on to SDS-PAGE gels and visualized by Coomassie staining.

4.3.6.3 E2 charging assay[129]

Human E1 enzyme (UBE1) and human E2 enzyme (UBE2S) were obtained from Boston Biochem (E-304 and E2-690). The charging assay was performed in reaction buffer (40 mM TRIS pH 7.5, 10 mM $MgCl_2$, 10 mM ATP and 0.6 mM DTT) and the reaction mixtures contained 250 nM E1 enzyme, 2.5 µM E2 enzyme and 19.5 µM Ub variant (Ub(wt), Ub-LLALTGG (=Ub(LAT), Ub-LLALTG (=Ub(LAT#) or Ub-LLALTGYEAAAK =Ub(LAT*)). Reaction mixtures were incubated at 37°C for 2 h. Subsequent SDS-PAGE analysis was performed by diluting 10 µL of reaction mixture into 2.5 µL of non-reducing Laemmli buffer.

4.3.7 Pull down assays

4.3.7.1 Pull down assays for GST-Tab2-NZF, GST-hHR23A-UBA2 and His-Rap80 7A

Pull down assays were performed as described previously.[336] In short, GST-Tab2-NZF or GST-hHR23A-UBA2 (20 µg) were incubated with 30 µl 50% slurry of glutathione sepharose 4B (GE Healthcare) equilibrated with pull down buffer (PDB, 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT and 0.1% NP-40) at 4 °C for 1 h and subsequently washed three times with PDB. Afterwards the beads were incubated with 6 µg K48/K63-linked diubiquitin variants in a total volume of 200 µl PDB at 4 °C for 16 h. Subsequently beads were washed six times with 100 µl PDB followed by addition of 50 µL 1x Laemmli buffer. After boiling at 95 °C for 10 min the samples were subjected to SDS-PAGE analysis. Western blotting was performed using either a mouse Ub antibody (Boston Biochem, A-104, 1:2000) or a rabbit linkage-specific K63 antibody (abcam, ab179434, 1:5000).d

4.3.8 Fluorescence methods

4.3.8.1 Fluorescent labelling of Rap80 7A

Rap80 7A was fluorescently labelled on the sulfhydryl group of C121. Atto488-maleimide (Atto-tec) was coupled to C121 according to the manufacturer's instructions and LC-MS analysis showed quantitative labelling of Rap80 7A with Atto488. Excess of fluorophore was removed by size exclusion chromatography (SEC) using a Superdex S75 10/300 (GE Healthcare) with 25 mM sodium phosphate pH 7.4, 150 mM NaCl, 5 mM β-mercaptoethanol, 1 mM EDTA and 0.005% Tween 20. Fractions containing the Atto488 labelled Rap80 7A were pooled together and concentrated with Amicon® 3 kDa MWCO centrifugal filter units (Millipore). Atto488 labelled Rap80 7A was stored at -80 °C until further use.

4.3.8.2 Fluorescence anisotropy binding assays

Ub and diubiquitin binding assays were performed as described by Joshua J. Sims and Robert E. Cohen.[340] Experiments were conducted on a Jasco Fluorescence Spectrometer FP-8500 equipped with polarizers (Jasco, Groß-Umstadt, Germany). Excitation and emission monochromators were set at 490 nm and 520 nm. Measurements were performed in 25 mM sodium phosphate pH 7.4, 150 mM NaCl, 5 mM β-mercaptoethanol, 1 mM EDTA and 0.005% Tween 20. Total cuvette volume was 60 µl and Atto488 labelled Rap80 7A concentration was held constant at 300 nM for all measurements.
All data processing was performed using Kaleidagraph software (Synergy Software, Reading, UK). K_D was determined from a hyperbolic fit and experiments were performed in triplicates.

4.3.9 Mammalian cell culture

4.3.9.1 Expression and purification of proteins containing AzGGK in mammalian cells

Human embryonic kidney 293T cells (HEK293T) were cultured in Dulbecco's modified Eagle's medium (DMEM, Sigma Aldrich) supplemented with 10 % (v/v) fetal bovine serum (FBS, Biochrom) and 1 % antibiotic-antimycotic solution (25 µg/mL amphotenicin B, 10 mg/mL streptomycin, and 10,000 units of penicillin, Sigma-Aldrich) at 37 °C in a humidified chamber with 5 % CO_2. HEK293T cells were seeded in Poly-L-lysine coated 6-well plates (Greiner) at $6x10^5$ cells/well for western blotting and in a 100 mm dish at $3x10^6$ cells/dish for protein purification followed by ESI-MS analysis. Fresh DMEM, supplemented with 2 mM AzGGK, was added to the cells prior to transfection. Transfection was performed using PEI[i] Transfection Reagent (Sigma Aldrich). A plasmid ratio of 3:1/pEF1-sfGFPN150TAG:pEF1-AzGGKRS was used for the co-transfection.[15] Cells were incubated for 24 to 48 hours prior to optional *in vivo* reduction. For protein purification of sfGFP-N150AzGGK-His6, cells were lysed with 500 µL lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 10 mM imidazole, 1 % Triton X-100, 1 mM PMSF, 1x protease inhibitor cocktail) for 30 minutes on ice. Cell lysate was clarified by centrifugation for 15 minutes at 14,000 xg and sfGFP was purified by Ni^{2+}-NTA affinity chromatography following the manufacturer's (Jena Bioscience) instructions. The purified protein was characterized by SDS-PAGE and ESI-MS analysis.

4.3.9.2 *In lysate* sortase-mediated ubiquitylation

By adding the cell-permeable reduction agent 2DPBA to the medium of cells expressing AzGGK-bearing protein, AzGGK-bearing proteins can be reduced to GGK-bearing proteins. The reduction was performed 24 to 48 hours post transfection as follows: cells were washed twice with PBS, and fresh, complete DMEM was added. Subsequently the cells were incubated for another 2-3 h to eliminate residual traces of AzGGK. After incubation, another washing step with PBS was carried out and 400 µM 2-DPBA containing complete DMEM was added. Cells were then incubated at 37 °C, 5 % CO_2 for another 4-18 h before harvesting them. Prior to harvesting two additional washing steps with PBS were conducted. For *in lysate* sortase-mediated ubiquitylation, 2DPBA-treated cells (growing in 6-well plates) were scraped in 500 µL PBS, transferred into reaction tubes, and pelleted via centrifugation (500xg, 5 min, 4 °C). The supernatant was discarded, and cells were re-suspended in 50-100 µL buffer of choice. After re-suspension, cells were flash frozen in liquid nitrogen and thawed on ice. This was repeated three times to break the membrane. Cell debris and nuclei were pelleted via centrifugation for 15 min at max. speed and 4 °C. The supernatant was removed and used for *in lysate* sortase-mediated ubiquitylation experiments. In order to work in calcium-free conditions, freeze thaw lysis was performed using calcium-depleting buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM EGTA) containing ethylene glycol tetraacetic acid (EGTA), which is a calcium chelating agent. For Ca^{2+}-containing conditions sortase buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$) was used for freeze-thaw lysis. 25 µL of supernatant were supplemented with 20 µM of the corresponding sortase variant, followed by ten minutes incubation at 37 °C. Afterwards 100 µM of Ub(LAT) or Ub(LPT) was added and the mixture was incubated for one hour at 37 °C, 200 rpm. 10 µL of reaction mixture were loaded onto a

SDS-PAGE after diluting and boiling with 4x SDS loading buffer. Consecutively, anti-His Western-Blot analysis was performed.

4.3.9.3 Sortylation background assays

In order to study sortase-mediated mis-ubiquitylation of endogenous targets, background assays were performed. For this, HEK293T cells were cultured in DMEM (Sigma Aldrich) supplemented with 10% (v/v) fetal bovine serum (FBS, Biochrom) and 1 % antibiotic-antimycotic solution (25 µg/mL amphotenicin B, 10 mg/mL streptomycin, and 10,000 units of penicillin, Sigma-Aldrich) at 37 °C in a humidified chamber with 5 % CO_2. On the day prior to transfection HEK293T cells were seeded in Poly-L-lysine coated 6-well plates (Greiner) at $6x10^5$ cells/well to reach 40-50 % confluence for transfection. Fresh DMEM was added followed by transfection with 1 µg pcDNA-HA-Ub (HA-Ub(wt) or HA-Ub(LAT*)), each in the presence or absence of 1 µg pIRES_Srt2A-myc. All transfections were performed using Lipofectamine2000 (Invitrogen) according to the manufacturer´s instructions. Cells were grown for 24 hours and then treated with 20 µM MG132 for three hours or remained untreated. Cell lysis was performed with RIPA buffer, followed by western blotting using α-HA antibody (Cell signaling, 2367) to detect possible ubiquitylation of endogenous proteins. Expression of Srt2A was detected via α-Myc (Cell Signaling, 2276) western blotting.

In order to study potential sortase-mediated and genetic code expansion-mediated mis-ubiquitylation of endogenous proteins (potential AzGGK incorporation in response to endogenous amber codons), background assays were performed. HEK293T cells were cultured in DMEM (Sigma Aldrich) supplemented with 10% (v/v) fetal bovine serum (FBS, Biochrom) and 1 % antibiotic-antimycotic solution (25 µg/mL amphotenicin B, 10 mg/mL streptomycin, and 10,000 units of penicillin, Sigma-Aldrich) at 37 °C in a humidified chamber with 5 % CO_2. On the day prior to transfection HEK293T cells were seeded in Poly-L-lysine coated 6-well plates (Greiner) at $6x10^5$ cells/well to reach 40-50 % confluence for transfection. Fresh DMEM, supplemented with 4 mM AzGGK was added followed by transfection using 1 µg of pIRES_Srt2A, 1 µg of pcDNA-HA-Ub(LAT*) and 0.3 µg of the pEF1-AzGGK-PylRS plasmid. All transfections were performed using Lipofectamine2000 (Invitrogen) according to the manufacturer´s instructions. Cells were grown for 35 hours in the presence of AzGGK (4 mM). To remove AzGGK, cells were washed twice with PBS, and fresh, complete DMEM was added. Cells were then incubated for another 2-3 hours to eliminate residual traces of AzGGK. After incubation, another washing step with PBS was carried out and 400 µM 2DPBA-containing complete DMEM was added. Cells were then incubated at 37 °C, 5 % CO_2 for 16-18 hours. Prior to cell lysis, both 2-DPBA reduced and non-reduced cells were treated with 20 µM MG132 for 3h or remained untreated. Cell lysis was performed with RIPA buffer followed by western blotting using α-His antibody (Cell Signaling, 2367) or α-HA antibody (Cell Signaling, 2367) to detect ubiquitylation and SUMOylation of proteins. Expression of Srt2A or mSrt2A was detected via α-Myc (Cell Signaling, 2276) western blotting.

4.3.10 MD-Simulations

The structure of Ub (PBD: 1UBQ), and Lys48- (PDB: 1AAR) and Lys63-linked diubiquitin (PDB: 2JF5) were used for initiating nine independent atomistic molecular dynamics (MD) simulations with native, AT, and LAT linkers. Each protein model was embedded in a TIP3P[386] water box with a 0.1 M NaCl concentration, and simulated for at least 0.5 μs at 310 K, using a timestep of 2 fs and the CHARMM36 force field.[387] Long range electrostatic interactions were treated using the Particle Mesh Ewald approach.[388] All simulations were performed using the GROMACS program suite[389] and VMD[390] was used for analysis in visualization.

4.4 Supporting information for Chapter 3

4.4.1 Protein expression and purification

4.4.1.1 Expression and purification of amber suppressed Ub, SUMO2 and GFP

Chemically competent *E. coli* K12 cells were co-transformed with pPylT_POI (which encodes *Mb* $tRNA_{CUA}$ and the C-terminally His tagged POI with one or two TAG codons at the denoted positions) and pBK_aaRS (which encodes either the wt *Mb* aaRS or *Mb* AzGGKRS) plasmids. After recovery with 1 mL of SOC medium for 1 h at 37 °C, the cells were cultured overnight in 50 mL of non-inducing medium[330] supplemented with tetracycline (17.5 µg/mL) and ampicillin (100 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in autoinduction medium[330] containing antibiotics (tetracycline (8.75 µg/mL) and ampicillin (50 µg/mL)) and the corresponding UAA (either 2 mM BocK or 4 mM AzGGK). After incubation overnight at 37 °C the cells were harvested by centrifugation (4000 xg, 20 min, 4 °C), flash frozen in liquid nitrogen and stored at -80 °C.
The obtained cell pellets were thawed on ice and re-suspended in lysis buffer (50 mM TRIS pH 8.0, 300 mM NaCl, 0.1 mg/mL DNase I (AppliChem), one cOmplete™ protease inhibitor tablet (Roche) and 0.175 mg/mL PMSF). Afterwards, the cell suspension was incubated on ice for 30 minutes and sonicated with cooling in an ice-water bath. The lysed cells were centrifuged (15,000 x g, 40 min, 4 °C) and the clear lysate was added to 1 mL Ni-NTA slurry / 1 L culture (Jena Bioscience) equilibrated with wash buffer (20 mM TRIS pH 8.0, 300 mM NaCl and 30 mM imidazole). Afterwards, the mixture was incubated with agitation for one hour at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV of wash buffer. The protein was eluted in 1 mL fractions with wash buffer supplemented with 300 mM imidazole pH 8.0. The fractions containing the POI (identified *via* 15% SDS-PAGE) were pooled together, concentrated and rebuffered (50 mM TRIS pH 7.5 and 150 mM NaCl) with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). Protein concentration was calculated from the measured A280 absorption (extinction coefficients were calculated with ProtParam (https://web.expasy.org/protparam/)). In case of Ub and SUMO the determination of protein concentration using the absorption at 280 nm is inaccurate (due to their low extinction coefficient (ε)), so BCA (Thermo Scientific™) and Bradford (Sigma Aldrich) assays were used to determine the accurate protein concentration. POIs with incorporated UAAs were flash frozen using liquid nitrogen and stored at -80 °C until further use.

4.4.1.2 Reduction of AzGGK-bearing POIs *via* Staudinger reduction

Reduction of the azide moiety of AzGGK to the amine (GGK) was performed on purified proteins by adding 2 eq. of 2-(diphenylphosphino)benzoic acid (2DPBA) or tris(2-carboxyethyl) phosphine (TCEP), followed by incubation for 1 h at room temperature and subsequent rebuffering to remove excess of 2DPBA/TCEP. The Staudinger reduction was monitored by LC-MS.

4.4.1.3 Cleavage of His-tag from SUMO2-(KxxGGK)-His variants using SENP2

After elution of SUMO2-KxxGGK-His from Ni-NTA the eluate (2 mL) was diluted in 20 mL of cleavage buffer (50 mM TRIS pH 7.5, 150 mM NaCl and 0.5 mM DTT), 100 µL of SUMO protease (SENP2) (with a His-tag, 1.4 mg/mL) were added followed by incubation at RT for 30 min. LC-MS was used to verify the completion of the H_6-tag cleavage from the C-terminus of SUMO2-KxxGGK-His. After successful cleavage, the reaction mixture was added to equilibrated Ni-NTA (200 µL/1 mg SUMO-KxxGGK) and incubated at 4 °C for 1 h with agitation. After incubation, the mixture was transferred to an empty plastic column, the flow-through was collected, concentrated and rebuffered (50 mM TRIS pH 7.5 and 150 mM NaCl) with an Amicon®. Protein concentration was determined using BCA (Thermo Scientific™) and Bradford (Sigma Aldrich) assays. Tagless SUMO-KxxGGK was flash frozen in liquid nitrogen and stored at -80 °C until further use.

4.4.1.4 Cleavage of His-tag from Ub-(KxxGGK)-His variants using USP2

His-tag cleavage from Ub-KxxGGK-His was performed analogous to above described His-tag cleavage from SUMO2-KxxGGK-His using a DUB (100 µL USP2, 2 mg/mL) instead of SUMO protease (SENP2).

4.4.1.5 Expression and purification of sortase mutants

Sortase mutants were expressed and purified as already described in Chapter 4.3.3.

4.4.1.6 Expression and purification of tagless Ub and Ub mutants

Ub and Ub mutants were expressed and purified as already described in Chapter 4.3.3.

4.4.1.7 Expression and purification of His-tagged Ub and SUMO variants

Expression and purification of His-tagged Ub and SUMO variants was performed as already described in Chapter 4.3.3.

4.4.1.8 Expression and purification of His-Rap80(1-137) 7A

Chemically competent *E. coli* Rosetta2 (DE3) were transformed with pET28a-Rap80(1-137)-7A-Linker plasmid. After recovery with 1 mL of SOC medium for one hour at 37 °C, the cells were cultured overnight in 50 mL of 2x YT medium containing kanamycin (50 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 500 mL of fresh 2x YT medium supplemented with kanamycin (25 µg/mL) and chloramphenicol (25 µg/mL) and cultured at 37 °C with shaking (200 rpm) until OD_{600} = 0.8. IPTG was added to a final concentration of 0.5 mM and protein expression was induced for 16 hours at 18 °C. The cells were harvested by centrifugation (4,000 xg, 20 min, 4 °C) and resuspended in lysis buffer (20 mM TRIS pH 8.0, 300 mM NaCl, 30 mM imidazole, 1 mM TCEP, 0.175 mg/mL PMSF, 0.1 mg/mL DNase I and one cOmplete™ protease inhibitor tablet (Roche)). The cell suspension was incubated on ice for 30 minutes and sonicated with cooling in an ice-water bath. The lysed cells were centrifuged (15,000 x g, 40 min, 4 °C) and the cleared

lysate was added to Ni-NTA slurry (Jena Bioscience) (1 mL of slurry per 1 L of culture) equilibrated with lysis buffer. Subsequently, the mixture was incubated with agitation for one hour at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV wash buffer (50 mM TRIS pH 8.0, 30 mM imidazole pH 8.0 and 300 mM NaCl). The protein was eluted in 1 mL fractions with elution buffer (wash buffer supplemented with 300 mM imidazole pH 8.0). The fractions containing Rap80 7A were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore) followed by size-exclusion chromatography (SEC) using a Superdex S75 10/300 (GE Healthcare) with SEC buffer (50 mM TRIS pH 7.5 and 150 mM NaCl). Fractions containing the Rap80 7A were pooled together and concentrated with Amicon® 3 kDa MWCO centrifugal filter units (Millipore). Protein concentration was determined using BCA (Thermo Scientific™) and Bradford (Sigma Aldrich) assays (since Rap80 1-137 7A Linker does not contain aromatic amino acids with absorbance at 280 nm). Rap80(1-137) 7A Linker was flash frozen using liquid nitrogen and stored at -80 °C until further use.

4.4.1.9 Expression and purification of GST-CDC34

Chemically competent *E. coli* Rosetta2 (DE3) were transformed with pGEX-6P-1-UBE2R1 (encoding for GST-CDC34). After recovery with 1 mL of SOC medium for one hour at 37 °C, the cells were cultured overnight in 50 mL of 2x YT medium containing ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, 200 rpm. The overnight culture was diluted to an OD_{600} of 0.05 in 1 L of fresh 2x YT medium supplemented with ampicillin (50 µg/mL) and chloramphenicol (25 µg/mL) and cultured at 37 °C with shaking (200 rpm) until OD_{600} = 0.8. IPTG was added to a final concentration of 0.25 mM and protein expression was induced for 18 hours at 20 °C. The cells were harvested by centrifugation (4,000 xg, 20 min, 4 °C) and resuspended in lysis buffer (50 mM TRIS pH 8.0, 270 mM sucrose, 50 mM NaF, 1 mM DTT, 0.1 mg/mL DNase I and one cOmplete™ protease inhibitor tablet (Roche)). The cell suspension was incubated on ice for 30 minutes and sonicated with cooling in an ice-water bath. The lysed cells were centrifuged (15,000 xg, 40 min, 4 °C), the cleared lysate added to Glutathione Sepharose 4B (GE Healthcare) (0.1 mL of slurry per 100 mL of culture) and the mixture was incubated with agitation for 1 h at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 10 CV of wash buffer (25 mM TRIS pH 8.5, 400 mM NaCl, 5 mM DTT). The protein was eluted in 1 mL fractions with elution buffer (wash buffer supplemented with 10 mM GSH pH 8.0). The fractions containing GST-CDC34 were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore) followed by size-exclusion chromatography (SEC) using a Superdex S75 16/600 (GE Healthcare) with SEC buffer (50 mM TRIS pH 7.5, 150 mM NaCl and 1 mM DTT). Fractions containing the GST-CDC34 were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). Protein concentration was calculated from the measured A280 absorption (extinction coefficients were calculated with ProtParam (https://web.expasy.org/protparam/). GST-CDC34 was flash frozen using liquid nitrogen and stored at -80 °C until further use.

4.4.1.10 Expression and purification of USP2 and SENP2

USP2[391] and SENP2[392] were expressed and purified as previously described.

4.4.2 Preparation of natively linked diUbs via lysine K48

Assembly of K48-linked diUbs was carried as previously described with slight adjustments.[385] In short, the assembly reactions contained 50 nM His-UBE1, 4.5 µM GST-CDC34 and variable concentrations of Ub mutants (as depicted in Figure S3.5). Reactions were incubated at 37 °C for typically 16 h in diUb reaction buffer (40 mM TRIS pH 7.5, 10 mM $MgCl_2$, 0.6 mM DTT and 10 mM ATP).
After incubation, 250 µL (per mL reaction volume) of Ni-NTA slurry (Jena Bioscience, equilibrated with wash buffer (50 mM TRIS pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$, 30 mM imidazole)) was added to the reaction and the mixture was incubated agitating for 1 h at 4 °C. After incubation, the mixture was transferred to an empty plastic column and washed with 40 CV of wash buffer to remove GST-CDC34, unreacted Ub wt and native K48 diUb. The proteins were eluted in 0.2 mL fractions with wash buffer supplemented with 300 mM imidazole pH 8.0. The fractions containing the mixture of K48-linked diUb and remaining acceptor Ub were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). In order to remove remaining acceptor Ub, size-exclusion chromatography (SEC) was performed using a Superdex S75 16/600 (GE Healthcare) with SEC Buffer (50 mM TRIS pH 7.5 and 150 mM NaCl). Fractions containing the desired K48-linked diUbs were pooled together and concentrated with Amicon® with the corresponding MWCO centrifugal filter units (Millipore). Protein concentration was determined using BCA (Thermo Scientific™) and Bradford (Sigma Aldrich) assays. DiUbs were flash frozen in liquid nitrogen and stored at -80 °C until further use.

4.4.3 Generation of complex Ub/Ubl topologies using sortase

4.4.3.1 Preparation of diUbs using sortase

GGK-bearing acceptor Ub (either with a native C-terminus (e.g. Ub-K6GGK) or with a C-terminus compatible for a subsequent reaction with an orthogonal sortase (e.g. Ub-K63GGK-LXLXGG) was diluted to 20 µM in sortase buffer (50 mM TRIS pH 7.5, 150 mM NaCl, 5 mM $CaCl_2$). Afterwards, 100 µM of the donor Ub (displaying the motif of the sortase used in this reaction at its C-terminus) was added, which was followed by the addition of 20 µM of the corresponding Srt mutant (without His-tag). Incubation was performed at 37 °C for 1 h when a donor Ub variant with spacer was used and 18 h when a donor Ub variant without spacer was used. Sortase-mediated transpeptidation was stopped by the addition of 200 µM phenyl vinyl sulfone and further incubation for 10 min at 37 °C. Afterwards, affinity purification using Ni-NTA slurry (Jena Bioscience) was performed as described above. Fractions containing the mixture of diUb and unreacted acceptor Ub were pooled together and concentrated using Amicon® with the corresponding MWCO centrifugal filter units (Millipore). To remove unreacted acceptor Ub, size-exclusion chromatography (SEC) was performed using a Superdex

S75 16/600 (GE Healthcare) with 50 mM TRIS pH 7.5, 150 mM NaCl. Fractions containing the diUb were pooled together and concentrated using Amicon® centrifugal filters. DiUbs were flash frozen in liquid nitrogen and stored at -80 °C until further use.

This protocol was used: (1) to generate all K6-linked diUbs to investigate sortase orthogonality in sortase hydrolysis assays (2) to assemble the K48-linked diUb used for the formation of triubiquitin and (3) to prepare K63-linked diUbs which serve as a basis for all hybrid chains.

4.4.3.2 Preparation of hybrid chains using sortase

The sortase reaction to generate hybrid chains was carried out similarly as described above for diUbs. In short, 20 µM of K63(AT)-diUb(PT)-His6 was used as donor and 100 µM SUMO2-KxxGGK (without C-terminal His-tag, which was removed using SENP2) was used as acceptor. Srt5M (20 µM, with His-tag) was added followed by incubation at 37 °C for 72 h. Ni-NTA was used to remove unreacted diUb and Srt5M. The concentrated Ni-NTA flow-through, containing the desired hybrid chain and the excess of SUMO2-KxxGGK, was applied to SEC (with 50 mM TRIS pH 7.5, 150 mM NaCl) for separation. The fractions containing pure hybrid chains (identified by 15% SDS-PAGE) were pooled and concentrated using Amicons®. Protein concentration was determined using BCA (Thermo Scientific) and Bradford (Sigma Aldrich) assays. Hybrid chains were flash frozen in liquid nitrogen and stored at -80 °C until further use. LC-MS analysis was used to verify the identity of hybrid chains.
Analytical assays for hybrid chain formation were performed analogously. 6 µL samples were taken at the denoted time points and quenched by the addition of 4× SDS loading buffer and boiling at 95 °C for 10 min. Samples were loaded on 15% SDS–PAGE gels and visualized by Coomassie staining.

4.4.3.3 Preparation of branched tri/pentaUbs using sortase

Generation of branched tri/pentaUbs followed the same protocol as above described sortase reactions. In short, 5/10 µM of UbKxxGGK/KyyGGK-His (in case of preparative assembly UbK11GGK/K48GG-(LPT)-His) was used as acceptor and 100 µM Ub-(LAT) (or 50 µM Ub-*isoK48*-Ub-(LAT)-His in case of pentaUb assembly) as donor. Srt2A (10/20 µM, without His-tag) was added followed by incubation at 37 °C for 1 h. Ni-NTA was used to remove unreacted Ub-(LAT) and Srt2A. The Ni-NTA eluate containing the desired branched triUb as well as the two diUb intermediates and the UbKxxGGK/KyyGGK-(LPT)-His was applied to SEC (with 50 mM TRIS pH 7.5, 150 mM NaCl) for separation. The fractions containing pure branched triUb (identified by 15% SDS-PAGE) chains were pooled and concentrated using Amicons®. Protein concentration was determined using BCA (Thermo Scientific™) and Bradford (Sigma Aldrich) assays. Branched triUb was flash frozen in liquid nitrogen and stored at -80 °C until further use. LC-MS analysis was used to verify the identity of the branched triUb.
Analytical assays for formation of branched triUbs/pentaUbs were performed analogously. 6 µL samples were taken at the denoted time points and quenched by the addition of 4× SDS loading buffer and boiling at 95 °C for 10 min. Samples were loaded on 15% SDS–PAGE gels and visualized by Coomassie staining.

4.4.3.4 Preparation of K48-linked tetraUb using sortase

Assembly of K48-linked tetraUb was performed analogous to above described sortase reactions. Taken together, 10 µM of Ub-*isoK48*-Ub-(LAT*)-His ("diUb B") was used as donor and 20 µM of UbK48GGK-*isoK48*-Ub-(LPT*)-His bearing GGK at K48 of the distal Ub ("diUb A") as acceptor (see Supplementary figures for clarification). Srt2A (5 µM, without His6-tag) was added followed by incubation at 37 °C for 3 h. Ni-NTA was used to remove Srt2A. The Ni-NTA eluate containing the desired K48-linked tetraUb as well as the two diUbs was applied to SEC (with 50 mM TRIS pH 7.5, 150 mM NaCl) for separation. The fractions containing pure K48-linked tetraUb (identified by 15% SDS-PAGE) chains were pooled and concentrated using Amicons®. Protein concentration was determined using BCA (Thermo Scientific) and Bradford (Sigma Aldrich) assays. K48-linked tetraUb was flash frozen in liquid nitrogen and stored at -80 °C until further use. LC-MS analysis was used to verify the identity of K48-linked tetraUb.
Analytical assays for formation of K48-linked tetraUb were performed analogously. 6 µL samples were taken at the denoted time points and quenched by the addition of 4× SDS loading buffer and boiling at 95 °C for 10 min. Samples were loaded on 15% SDS–PAGE gels and visualized by Coomassie staining.

4.4.3.5 Preparation of K48-, K6-linked triUb using sortase

The sortase reaction to generate the triUb was carried out as described above for diUbs. In short, 20 µM of Ub-(LAT)-*isoK48*-Ub-(LPT)-His was used as donor Ub and 100 µM Ub-K6GGK were used as acceptor Ub. Srt5M (20 µM, without His-tag) was added followed by incubation at 37 °C for 1 h. Affinity purification followed by SEC (with 50 mM TRIS pH 7.5, 150 mM NaCl) was used to purify the triUb. TriUb was flash frozen in liquid nitrogen and stored at -80 °C. LC-MS analysis was used to verify the triUb. Analytical assay of triUb formation was performed analogously. 6 µL samples were taken at the denoted time points and quenched by the addition of 4× SDS loading buffer and boiling at 95 °C for 10 min. Samples were loaded on 15% SDS–PAGE gels and visualized by Coomassie staining.

4.4.3.6 Charging of mono/di/tetraUb on GFP

Charging of Ub-(LAT), "diUb A" and K48-linked tetraUb was performed as similar as sortase reactions described above. 20 µM of sfGFPN150GGK-His6 was used as acceptor and 10 µM of mono/di/tetraUb-(LAT) was used as donor. Srt2A (5 µM, without His6-tag) was added followed by incubation at 37 °C. 6 µL samples were taken at the denoted time points and quenched by the addition of 4× SDS loading buffer and boiling at 95 °C for 10 min. Samples were loaded on 15% SDS–PAGE gels and visualized by Coomassie staining.

4.4.4 DiUb hydrolysis assays

K6-linked diUbs displaying different sortase motifs at their linkage site (LAT, LPT and LPS) were diluted into sortase buffer to 40 µM followed by addition of 10 µM of different Srt mutants. Reaction mixtures were incubated at 37 °C and 6 µL samples were taken at the denoted

time points and quenched by the addition of 4× SDS loading buffer. After boiling at 95 °C for 10 min and centrifugation (14,000 xg, 5 min) samples were loaded on 15% SDS–PAGE gels and visualized by Coomassie staining.

4.4.5 Pull down assays with His-Rap80(1-137) 7A Linker

His-Rap80(1-137) with 7A Linker (20 µg) was incubated with 100 µL 50% slurry of Ni-NTA agarose (Jena Bioscience) equilibrated with pull down buffer (PDB, 50 mM TRIS pH 7.5, 150 mM NaCl, 2 mM TCEP and 0.1% NP-40) at 4 °C for 1 h and subsequently washed three times with PDB (4000 xg, 2 min, 4°C). Afterwards 50% (v/v) slurry was generated using PDB. To 30 µL of Ni-NTA slurry charged with Rap80(1-137), 3.5 µM of differently linked hybrid chains and respective controls (Ub wt, K48(AT)-diUb and SUMO2) were added in a total volume of 50 µl. After incubation for 1 h at 4 °C beads were washed five times with 100 µl PDB followed by addition of 50 µl 1x Laemmli buffer. After boiling at 95 °C for 10 min and centrifugation (14,000 xg, 10 min) the samples were loaded on 15% SDS–PAGE gels and visualized by Coomassie staining.

5. References

(1) Berg, J. M.; Tymoczko, J. L.; Gatto, G. J.; Stryer, L. *Biochemistry*; Eighth edition. ed.; W.H. Freeman & Company, a Macmillan Education Imprint: New York, 2015.

(2) Crick, F. H. On protein synthesis. *Symp Soc Exp Biol* **1958,** *12*, 138.

(3) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.et al. The sequence of the human genome. *Science* **2001,** *291* (5507), 1304.

(4) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.et al. How many human proteoforms are there? *Nat Chem Biol* **2018,** *14* (3), 206.

(5) Beadle, G. W.; Tatum, E. L. Genetic Control of Biochemical Reactions in Neurospora. *Proceedings of the National Academy of Sciences* **1941,** *27* (11), 499.

(6) Smith, L. M.; Kelleher, N. L.; Consortium for Top Down, P. Proteoform: a single term describing protein complexity. *Nat Methods* **2013,** *10* (3), 186.

(7) Mohler, K.; Ibba, M. Translational fidelity and mistranslation in the cellular response to stress. *Nat Microbiol* **2017,** *2*, 17117.

(8) Drummond, D. A.; Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet* **2009,** *10* (10), 715.

(9) Virág, D.; Dalmadi-Kiss, B.; Vékey, K.; Drahos, L.; Klebovich, I.; Antal, I.; Ludányi, K. Current Trends in the Analysis of Post-translational Modifications. *Chromatographia* **2019,** *83* (1), 1.

(10) Roy, B.; Haupt, L. M.; Griffiths, L. R. Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity. *Curr Genomics* **2013,** *14* (3), 182.

(11) Picardi, E.; D'Erchia, A. M.; Lo Giudice, C.; Pesole, G. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res* **2017,** *45* (D1), D750.

(12) Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med* **2012,** *4* (6), 565.

(13) Venne, A. S.; Kollipara, L.; Zahedi, R. P. The next level of complexity: crosstalk of posttranslational modifications. *Proteomics* **2014,** *14* (4-5), 513.

(14) Garcia, B. A.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Pervasive combinatorial modification of histone H3 in human cells. *Nat Methods* **2007,** *4* (6), 487.

(15) Xu, Y. M.; Du, J. Y.; Lau, A. T. Posttranslational modifications of human histone H3: an update. *Proteomics* **2014,** *14* (17-18), 2047.

(16) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **2005,** *44* (45), 7342.

(17) Muller, M. M. Post-Translational Modifications of Protein Backbones: Unique Functions, Mechanisms, and Challenges. *Biochemistry* **2018,** *57* (2), 177.

(18) Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **1992,** *111* (2), 229.

(19) Yokoyama, H.; Mizutani, R.; Noguchi, S.; Hayashida, N. Structural and biochemical basis of the formation of isoaspartate in the complementarity-determining region of antibody 64M-5 Fab. *Sci Rep* **2019,** *9* (1), 18494.

(20) Eschenburg, S.; Schonbrunn, E. Comparative X-ray analysis of the un-liganded fosfomycin-target murA. *Proteins* **2000,** *40* (2), 290.

(21) Wold, F. In vivo chemical modification of proteins (post-translational modification). *Annu Rev Biochem* **1981,** *50*, 783.
(22) Creighton, T. E. *Proteins : structures and molecular properties*; 2nd ed.; W.H. Freeman: New York, 1993.
(23) Bradshaw, R. A.; Brickey, W. W.; Walker, K. W. N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families. *Trends Biochem Sci* **1998,** *23* (7), 263.
(24) Wagner, T.; Kahnt, J.; Ermler, U.; Shima, S. Didehydroaspartate Modification in Methyl-Coenzyme M Reductase Catalyzing Methane Formation. *Angew Chem Int Ed Engl* **2016,** *55* (36), 10630.
(25) Kahnt, J.; Buchenau, B.; Mahlert, F.; Kruger, M.; Shima, S.; Thauer, R. K. Post-translational modifications in the active site region of methyl-coenzyme M reductase from methanogenic and methanotrophic archaea. *FEBS J* **2007,** *274* (18), 4913.
(26) Nayak, D. D.; Mahanta, N.; Mitchell, D. A.; Metcalf, W. W. Post-translational thioamidation of methyl-coenzyme M reductase, a key enzyme in methanogenic and methanotrophic Archaea. *Elife* **2017,** *6*.
(27) Varland, S.; Osberg, C.; Arnesen, T. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* **2015,** *15* (14), 2385.
(28) Ree, R.; Varland, S.; Arnesen, T. Spotlight on protein N-terminal acetylation. *Exp Mol Med* **2018,** *50* (7), 1.
(29) Udenwobele, D. I.; Su, R. C.; Good, S. V.; Ball, T. B.; Varma Shrivastav, S.; Shrivastav, A. Myristoylation: An Important Protein Modification in the Immune Response. *Front Immunol* **2017,** *8*, 751.
(30) Drazic, A.; Myklebust, L. M.; Ree, R.; Arnesen, T. The world of protein acetylation. *Biochim Biophys Acta* **2016,** *1864* (10), 1372.
(31) Aksnes, H.; Drazic, A.; Marie, M.; Arnesen, T. First Things First: Vital Protein Marks by N-Terminal Acetyltransferases. *Trends Biochem Sci* **2016,** *41* (9), 746.
(32) Bradbury, A. F.; Smyth, D. G. Peptide amidation. *Trends Biochem Sci* **1991,** *16* (3), 112.
(33) Lange, P. F.; Overall, C. M. Protein TAILS: when termini tell tales of proteolysis and function. *Curr Opin Chem Biol* **2013,** *17* (1), 73.
(34) Ichimura, Y.; Kirisako, T.; Takao, T.; Satomi, Y.; Shimonishi, Y.; Ishihara, N.; Mizushima, N.; Tanida, I.; Kominami, E.; Ohsumi, M.et al. A ubiquitin-like system mediates protein lipidation. *Nature* **2000,** *408* (6811), 488.
(35) Paulick, M. G.; Bertozzi, C. R. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry* **2008,** *47* (27), 6991.
(36) Paulus, H. Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* **2000,** *69*, 447.
(37) Shah, N. H.; Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chem Sci* **2014,** *5* (1), 446.
(38) Vila-Perello, M.; Muir, T. W. Biological applications of protein splicing. *Cell* **2010,** *143* (2), 191.
(39) Weiss, M. S.; Jabs, A.; Hilgenfeld, R. Peptide bonds revisited. *Nat Struct Biol* **1998,** *5* (8), 676.
(40) Exarchos, K. P.; Papaloukas, C.; Exarchos, T. P.; Troganis, A. N.; Fotiadis, D. I. Prediction of cis/trans isomerization using feature selection and support vector machines. *J Biomed Inform* **2009,** *42* (1), 140.
(41) Lu, K. P.; Finn, G.; Lee, T. H.; Nicholson, L. K. Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* **2007,** *3* (10), 619.

(42) Yaffe, M. B.; Schutkowski, M.; Shen, M.; Zhou, X. Z.; Stukenberg, P. T.; Rahfeld, J. U.; Xu, J.; Kuang, J.; Kirschner, M. W.; Fischer, G.et al. Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. *Science* **1997,** *278* (5345), 1957.
(43) Cohen, P. The origins of protein phosphorylation. *Nat Cell Biol* **2002,** *4* (5), E127.
(44) Wilson, L. J.; Linley, A.; Hammond, D. E.; Hood, F. E.; Coulson, J. M.; MacEwan, D. J.; Ross, S. J.; Slupsky, J. R.; Smith, P. D.; Eyers, P. A.et al. New Perspectives, Opportunities, and Challenges in Exploring the Human Protein Kinome. *Cancer Res* **2018,** *78* (1), 15.
(45) Hardman, G.; Perkins, S.; Brownridge, P. J.; Clarke, C. J.; Byrne, D. P.; Campbell, A. E.; Kalyuzhnyy, A.; Myall, A.; Eyers, P. A.; Jones, A. R.et al. Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation. *EMBO J* **2019,** *38* (21), e100847.
(46) Fuhs, S. R.; Hunter, T. pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. *Curr Opin Cell Biol* **2017,** *45*, 8.
(47) Yaffe, M. B.; Smerdon, S. J. PhosphoSerine/threonine binding domains: you can't pSERious? *Structure* **2001,** *9* (3), R33.
(48) Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med* **2017,** *40* (2), 271.
(49) Ferguson, F. M.; Gray, N. S. Kinase inhibitors: the road ahead. *Nat Rev Drug Discov* **2018,** *17* (5), 353.
(50) Hansen, B. K.; Gupta, R.; Baldus, L.; Lyon, D.; Narita, T.; Lammers, M.; Choudhary, C.; Weinert, B. T. Analysis of human acetylation stoichiometry defines mechanistic constraints on protein regulation. *Nat Commun* **2019,** *10* (1), 1055.
(51) Verdone, L.; Agricola, E.; Caserta, M.; Di Mauro, E. Histone acetylation in gene regulation. *Brief Funct Genomic Proteomic* **2006,** *5* (3), 209.
(52) Narita, T.; Weinert, B. T.; Choudhary, C. Functions and mechanisms of non-histone protein acetylation. *Nat Rev Mol Cell Biol* **2019,** *20* (3), 156.
(53) Gong, F.; Chiu, L. Y.; Miller, K. M. Acetylation Reader Proteins: Linking Acetylation Signaling to Genome Maintenance and Cancer. *PLoS Genet* **2016,** *12* (9), e1006272.
(54) Resh, M. D. Fatty acylation of proteins: The long and the short of it. *Prog Lipid Res* **2016,** *63*, 120.
(55) Stevenson, F. T.; Bursten, S. L.; Locksley, R. M.; Lovett, D. H. Myristyl acylation of the tumor necrosis factor alpha precursor on specific lysine residues. *J Exp Med* **1992,** *176* (4), 1053.
(56) Stevenson, F. T.; Bursten, S. L.; Fanton, C.; Locksley, R. M.; Lovett, D. H. The 31-kDa precursor of interleukin 1 alpha is myristoylated on specific lysines within the 16-kDa N-terminal propiece. *Proc Natl Acad Sci U S A* **1993,** *90* (15), 7245.
(57) Wisniewski, J. R.; Zougman, A.; Mann, M. Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic Acids Res* **2008,** *36* (2), 570.
(58) Liu, B.; Lin, Y.; Darwanto, A.; Song, X.; Xu, G.; Zhang, K. Identification and characterization of propionylation at histone H3 lysine 23 in mammalian cells. *J Biol Chem* **2009,** *284* (47), 32288.
(59) Chen, Y.; Sprung, R.; Tang, Y.; Ball, H.; Sangras, B.; Kim, S. C.; Falck, J. R.; Peng, J.; Gu, W.; Zhao, Y. Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol Cell Proteomics* **2007,** *6* (5), 812.
(60) Dai, L.; Peng, C.; Montellier, E.; Lu, Z.; Chen, Y.; Ishii, H.; Debernardi, A.; Buchou, T.; Rousseaux, S.; Jin, F.et al. Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nature Chemical Biology* **2014,** *10* (5), 365.

(61) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **2011,** *146* (6), 1016.
(62) Xie, Z.; Dai, J.; Dai, L.; Tan, M.; Cheng, Z.; Wu, Y.; Boeke, J. D.; Zhao, Y. Lysine succinylation and lysine malonylation in histones. *Mol Cell Proteomics* **2012,** *11* (5), 100.
(63) Tan, M.; Peng, C.; Anderson, K. A.; Chhoy, P.; Xie, Z.; Dai, L.; Park, J.; Chen, Y.; Huang, H.; Zhang, Y.et al. Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metab* **2014,** *19* (4), 605.
(64) Mukherjee, S.; Hao, Y. H.; Orth, K. A newly discovered post-translational modification--the acetylation of serine and threonine residues. *Trends Biochem Sci* **2007,** *32* (5), 210.
(65) Britton, L. M.; Newhart, A.; Bhanu, N. V.; Sridharan, R.; Gonzales-Cope, M.; Plath, K.; Janicki, S. M.; Garcia, B. A. Initial characterization of histone H3 serine 10 O-acetylation. *Epigenetics* **2013,** *8* (10), 1101.
(66) Thinon, E.; Hang, H. C. Chemical reporters for exploring protein acylation. *Biochem Soc Trans* **2015,** *43* (2), 253.
(67) Biggar, K. K.; Li, S. S. Non-histone protein methylation as a regulator of cellular signalling and function. *Nat Rev Mol Cell Biol* **2015,** *16* (1), 5.
(68) Wagner, T.; Robaa, D.; Sippl, W.; Jung, M. Mind the methyl: methyllysine binding proteins in epigenetic regulation. *ChemMedChem* **2014,** *9* (3), 466.
(69) Chen, C.; Nott, T. J.; Jin, J.; Pawson, T. Deciphering arginine methylation: Tudor tells the tale. *Nat Rev Mol Cell Biol* **2011,** *12* (10), 629.
(70) Zhang, J.; Jing, L.; Li, M.; He, L.; Guo, Z. Regulation of histone arginine methylation/demethylation by methylase and demethylase (Review). *Mol Med Rep* **2019,** *19* (5), 3963.
(71) Jenuwein, T.; Allis, C. D. Translating the histone code. *Science* **2001,** *293* (5532), 1074.
(72) Lanouette, S.; Mongeon, V.; Figeys, D.; Couture, J. F. The functional diversity of protein lysine methylation. *Mol Syst Biol* **2014,** *10*, 724.
(73) Blanc, R. S.; Richard, S. Arginine Methylation: The Coming of Age. *Mol Cell* **2017,** *65* (1), 8.
(74) Novelli, G.; D'Apice, M. R. Protein farnesylation and disease. *J Inherit Metab Dis* **2012,** *35* (5), 917.
(75) Berndt, N.; Sebti, S. M. Measurement of protein farnesylation and geranylgeranylation in vitro, in cultured cells and in biopsies, and the effects of prenyl transferase inhibitors. *Nat Protoc* **2011,** *6* (11), 1775.
(76) Leung, K. F.; Baron, R.; Ali, B. R.; Magee, A. I.; Seabra, M. C. Rab GTPases containing a CAAX motif are processed post-geranylgeranylation by proteolysis and methylation. *J Biol Chem* **2007,** *282* (2), 1487.
(77) Gorres, K. L.; Raines, R. T. Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol* **2010,** *45* (2), 106.
(78) Vermeer, C. Gamma-carboxyglutamate-containing proteins and the vitamin K-dependent carboxylase. *Biochem J* **1990,** *266* (3), 625.
(79) Cummins, P. M.; O'Connor, B. Pyroglutamyl peptidase: an overview of the three known enzymatic forms. *Biochim Biophys Acta* **1998,** *1429* (1), 1.
(80) Farrelly, L. A.; Thompson, R. E.; Zhao, S.; Lepack, A. E.; Lyu, Y.; Bhanu, N. V.; Zhang, B.; Loh, Y. E.; Ramakrishnan, A.; Vadodaria, K. C.et al. Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **2019,** *567* (7749), 535.

(81) Sakurai, S.; Inai, Y.; Minakata, S.; Manabe, S.; Ito, Y.; Ihara, Y. A novel assay for detection and quantification of C-mannosyl tryptophan in normal or diabetic mice. *Sci Rep* **2019,** *9* (1), 4675.
(82) Lyons, J. J.; Milner, J. D.; Rosenzweig, S. D. Glycans Instructing Immunity: The Emerging Role of Altered Glycosylation in Clinical Immunology. *Front Pediatr* **2015,** *3*, 54.
(83) Chauhan, J. S.; Rao, A.; Raghava, G. P. In silico platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences. *PLoS One* **2013,** *8* (6), e67008.
(84) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in health and disease. *Nat Rev Nephrol* **2019,** *15* (6), 346.
(85) Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **2012,** *13* (7), 448.
(86) Berner, N.; Reutter, K. R.; Wolf, D. H. Protein Quality Control of the Endoplasmic Reticulum and Ubiquitin-Proteasome-Triggered Degradation of Aberrant Proteins: Yeast Pioneers the Path. *Annu Rev Biochem* **2018,** *87*, 751.
(87) Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **2012,** *22* (6), 736.
(88) Yang, X.; Qian, K. Protein O-GlcNAcylation: emerging mechanisms and functions. *Nat Rev Mol Cell Biol* **2017,** *18* (7), 452.
(89) Varki, A. *Essentials of glycobiology*; 2nd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2009.
(90) Ihara, Y.; Inai, Y.; Ikezaki, M.; Matsui, I.-S. L.; Manabe, S.; Ito, Y. In *Glycoscience: Biology and Medicine*; Taniguchi, N.;Endo, T.;Hart, G. W.;Seeberger, P. H.;Wong, C.-H., Eds.; Springer Japan: Tokyo, 2015, DOI:10.1007/978-4-431-54841-6_67 10.1007/978-4-431-54841-6_67.
(91) Vliegenthart, J. F. The complexity of glycoprotein-derived glycans. *Proc Jpn Acad Ser B Phys Biol Sci* **2017,** *93* (2), 64.
(92) Oman, T. J.; Boettcher, J. M.; Wang, H.; Okalibe, X. N.; van der Donk, W. A. Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat Chem Biol* **2011,** *7* (2), 78.
(93) Liu, Q.; Florea, B. I.; Filippov, D. V. ADP-Ribosylation Goes Normal: Serine as the Major Site of the Modification. *Cell Chem Biol* **2017,** *24* (4), 431.
(94) Leung, A. K. L. Poly(ADP-ribose): A Dynamic Trigger for Biomolecular Condensate Formation. *Trends Cell Biol* **2020,** *30* (5), 370.
(95) Luscher, B.; Butepage, M.; Eckei, L.; Krieg, S.; Verheugd, P.; Shilton, B. H. ADP-Ribosylation, a Multifaceted Posttranslational Modification Involved in the Control of Cell Physiology in Health and Disease. *Chem Rev* **2018,** *118* (3), 1092.
(96) Alemasova, E. E.; Lavrik, O. I. Poly(ADP-ribosyl)ation by PARP1: reaction mechanism and regulatory proteins. *Nucleic Acids Res* **2019,** *47* (8), 3811.
(97) Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu Rev Biochem* **1998,** *67*, 425.
(98) Kim, W.; Bennett, E. J.; Huttlin, E. L.; Guo, A.; Li, J.; Possemato, A.; Sowa, M. E.; Rad, R.; Rush, J.; Comb, M. J.et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* **2011,** *44* (2), 325.
(99) Wagner, S. A.; Beli, P.; Weinert, B. T.; Nielsen, M. L.; Cox, J.; Mann, M.; Choudhary, C. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol Cell Proteomics* **2011,** *10* (10), M111 013284.
(100) Swatek, K. N.; Komander, D. Ubiquitin modifications. *Cell Res* **2016,** *26* (4), 399.
(101) Husnjak, K.; Dikic, I. Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu Rev Biochem* **2012,** *81*, 291.

(102) Rape, M. Ubiquitylation at the crossroads of development and disease. *Nat Rev Mol Cell Biol* **2018,** *19* (1), 59.
(103) Zheng, N.; Shabek, N. Ubiquitin Ligases: Structure, Function, and Regulation. *Annu Rev Biochem* **2017,** *86*, 129.
(104) Grou, C. P.; Pinto, M. P.; Mendes, A. V.; Domingues, P.; Azevedo, J. E. The de novo synthesis of ubiquitin: identification of deubiquitinases acting on ubiquitin precursors. *Sci Rep* **2015,** *5*, 12836.
(105) Metzger, M. B.; Hristova, V. A.; Weissman, A. M. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci* **2012,** *125* (Pt 3), 531.
(106) Walden, H.; Rittinger, K. RBR ligase-mediated ubiquitin transfer: a tale with many twists and turns. *Nat Struct Mol Biol* **2018,** *25* (6), 440.
(107) Pao, K. C.; Wood, N. T.; Knebel, A.; Rafie, K.; Stanley, M.; Mabbitt, P. D.; Sundaramoorthy, R.; Hofmann, K.; van Aalten, D. M. F.; Virdee, S. Activity-based E3 ligase profiling uncovers an E3 ligase with esterification activity. *Nature* **2018,** *556* (7701), 381.
(108) McClellan, A. J.; Laugesen, S. H.; Ellgaard, L. Cellular functions and molecular mechanisms of non-lysine ubiquitination. *Open Biol* **2019,** *9* (9), 190147.
(109) Pisano, A.; Albano, F.; Vecchio, E.; Renna, M.; Scala, G.; Quinto, I.; Fiume, G. Revisiting Bacterial Ubiquitin Ligase Effectors: Weapons for Host Exploitation. *Int J Mol Sci* **2018,** *19* (11).
(110) Calistri, A.; Munegato, D.; Carli, I.; Parolin, C.; Palu, G. The ubiquitin-conjugating system: multiple roles in viral replication and infection. *Cells* **2014,** *3* (2), 386.
(111) Scott, D.; Oldham, N. J.; Strachan, J.; Searle, M. S.; Layfield, R. Ubiquitin-binding domains: mechanisms of ubiquitin recognition and use as tools to investigate ubiquitin-modified proteomes. *Proteomics* **2015,** *15* (5-6), 844.
(112) Berk, J. M.; Lim, C.; Ronau, J. A.; Chaudhuri, A.; Chen, H.; Beckmann, J. F.; Loria, J. P.; Xiong, Y.; Hochstrasser, M. A deubiquitylase with an unusually high-affinity ubiquitin-binding domain from the scrub typhus pathogen Orientia tsutsugamushi. *Nat Commun* **2020,** *11* (1), 2343.
(113) Lai, M. Y.; Zhang, D.; Laronde-Leblanc, N.; Fushman, D. Structural and biochemical studies of the open state of Lys48-linked diubiquitin. *Biochim Biophys Acta* **2012,** *1823* (11), 2046.
(114) Ye, Y.; Blaser, G.; Horrocks, M. H.; Ruedas-Rama, M. J.; Ibrahim, S.; Zhukov, A. A.; Orte, A.; Klenerman, D.; Jackson, S. E.; Komander, D. Ubiquitin chain conformation regulates recognition and activity of interacting proteins. *Nature* **2012,** *492* (7428), 266.
(115) Varadan, R.; Assfalg, M.; Raasi, S.; Pickart, C.; Fushman, D. Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Mol Cell* **2005,** *18* (6), 687.
(116) Sekiyama, N.; Jee, J.; Isogai, S.; Akagi, K.; Huang, T. H.; Ariyoshi, M.; Tochio, H.; Shirakawa, M. NMR analysis of Lys63-linked polyubiquitin recognition by the tandem ubiquitin-interacting motifs of Rap80. *J Biomol NMR* **2012,** *52* (4), 339.
(117) Mevissen, T. E. T.; Komander, D. Mechanisms of Deubiquitinase Specificity and Regulation. *Annu Rev Biochem* **2017,** *86*, 159.
(118) Fox, J. T.; Lee, K. Y.; Myung, K. Dynamic regulation of PCNA ubiquitylation/deubiquitylation. *FEBS Lett* **2011,** *585* (18), 2780.
(119) 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* **2018,** *115* (7), E1401.
(120) Chen, Z. J. Ubiquitin signalling in the NF-κB pathway. *Nature Cell Biology* **2005,** *7* (8), 758.

(121) Lee, B. L.; Singh, A.; Mark Glover, J. N.; Hendzel, M. J.; Spyracopoulos, L. Molecular Basis for K63-Linked Ubiquitination Processes in Double-Strand DNA Break Repair: A Focus on Kinetics and Dynamics. *J Mol Biol* **2017,** *429* (22), 3409.
(122) Mendes, M. L.; Fougeras, M. R.; Dittmar, G. Analysis of ubiquitin signaling and chain topology cross-talk. *J Proteomics* **2020,** *215*, 103634.
(123) Palicharla, V. R.; Maddika, S. HACE1 mediated K27 ubiquitin linkage leads to YB-1 protein secretion. *Cell Signal* **2015,** *27* (12), 2355.
(124) Gregori, L.; Poosch, M. S.; Cousins, G.; Chau, V. A uniform isopeptide-linked multiubiquitin chain is sufficient to target substrate for degradation in ubiquitin-mediated proteolysis. *J Biol Chem* **1990,** *265* (15), 8354.
(125) Ohtake, F.; Saeki, Y.; Ishido, S.; Kanno, J.; Tanaka, K. The K48-K63 Branched Ubiquitin Chain Regulates NF-kappaB Signaling. *Mol Cell* **2016,** *64* (2), 251.
(126) Goto, E.; Yamanaka, Y.; Ishikawa, A.; Aoki-Kawasumi, M.; Mito-Yoshida, M.; Ohmura-Hoshino, M.; Matsuki, Y.; Kajikawa, M.; Hirano, H.; Ishido, S. Contribution of lysine 11-linked ubiquitination to MIR2-mediated major histocompatibility complex class I internalization. *J Biol Chem* **2010,** *285* (46), 35311.
(127) Boname, J. M.; Thomas, M.; Stagg, H. R.; Xu, P.; Peng, J.; Lehner, P. J. Efficient internalization of MHC I requires lysine-11 and lysine-63 mixed linkage polyubiquitin chains. *Traffic* **2010,** *11* (2), 210.
(128) Komander, D.; Clague, M. J.; Urbe, S. Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol* **2009,** *10* (8), 550.
(129) 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* **2010,** *17* (8), 939.
(130) Mevissen, T. E.; Hospenthal, M. K.; Geurink, P. P.; Elliott, P. R.; Akutsu, M.; Arnaudo, N.; Ekkebus, R.; Kulathu, Y.; Wauer, T.; El Oualid, F.et al. OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. *Cell* **2013,** *154* (1), 169.
(131) He, M.; Zhou, Z.; Shah, A. A.; Zou, H.; Tao, J.; Chen, Q.; Wan, Y. The emerging role of deubiquitinating enzymes in genomic integrity, diseases, and therapeutics. *Cell Biosci* **2016,** *6*, 62.
(132) Reyes-Turcu, F. E.; Ventii, K. H.; Wilkinson, K. D. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu Rev Biochem* **2009,** *78*, 363.
(133) Clague, M. J.; Urbe, S.; Komander, D. Breaking the chains: deubiquitylating enzyme specificity begets function. *Nat Rev Mol Cell Biol* **2019,** *20* (6), 338.
(134) D'Arcy, P.; Wang, X.; Linder, S. Deubiquitinase inhibition as a cancer therapeutic strategy. *Pharmacol Ther* **2015,** *147*, 32.
(135) Komander, D.; Rape, M. The ubiquitin code. *Annual review of biochemistry* **2012,** *81*, 203.
(136) Cappadocia, L.; Lima, C. D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem Rev* **2018,** *118* (3), 889.
(137) Impens, F.; Radoshevich, L.; Cossart, P.; Ribet, D. Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli. *Proc Natl Acad Sci U S A* **2014,** *111* (34), 12432.
(138) Hay, R. T. SUMO: a history of modification. *Mol Cell* **2005,** *18* (1), 1.
(139) Pérez Berrocal, D. A.; Witting, K. F.; Ovaa, H.; Mulder, M. P. C. Hybrid Chains: A Collaboration of Ubiquitin and Ubiquitin-Like Modifiers Introducing Cross-Functionality to the Ubiquitin Code. *Frontiers in Chemistry* **2020,** *7* (931).
(140) Guzzo, C. M.; Berndsen, C. E.; Zhu, J.; Gupta, V.; Datta, A.; Greenberg, R. A.; Wolberger, C.; Matunis, M. J. RNF4-dependent hybrid SUMO-ubiquitin chains are

signals for RAP80 and thereby mediate the recruitment of BRCA1 to sites of DNA damage. *Sci Signal* **2012,** *5* (253), ra88.
(141) Ohtake, F.; Saeki, Y.; Sakamoto, K.; Ohtake, K.; Nishikawa, H.; Tsuchiya, H.; Ohta, T.; Tanaka, K.; Kanno, J. Ubiquitin acetylation inhibits polyubiquitin chain elongation. *EMBO Rep* **2015,** *16* (2), 192.
(142) Wauer, T.; Swatek, K. N.; Wagstaff, J. L.; Gladkova, C.; Pruneda, J. N.; Michel, M. A.; Gersch, M.; Johnson, C. M.; Freund, S. M.; Komander, D. Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. *EMBO J* **2015,** *34* (3), 307.
(143) Cui, J.; Yao, Q.; Li, S.; Ding, X.; Lu, Q.; Mao, H.; Liu, L.; Zheng, N.; Chen, S.; Shao, F. Glutamine deamidation and dysfunction of ubiquitin/NEDD8 induced by a bacterial effector family. *Science* **2010,** *329* (5996), 1215.
(144) Yang, C. S.; Jividen, K.; Spencer, A.; Dworak, N.; Ni, L.; Oostdyk, L. T.; Chatterjee, M.; Kusmider, B.; Reon, B.; Parlak, M.et al. Ubiquitin Modification by the E3 Ligase/ADP-Ribosyltransferase Dtx3L/Parp9. *Mol Cell* **2017,** *66* (4), 503.
(145) Bhogaraju, S.; Kalayil, S.; Liu, Y.; Bonn, F.; Colby, T.; Matic, I.; Dikic, I. Phosphoribosylation of Ubiquitin Promotes Serine Ubiquitination and Impairs Conventional Ubiquitination. *Cell* **2016,** *167* (6), 1636.
(146) Koyano, F.; Okatsu, K.; Kosako, H.; Tamura, Y.; Go, E.; Kimura, M.; Kimura, Y.; Tsuchiya, H.; Yoshihara, H.; Hirokawa, T.et al. Ubiquitin is phosphorylated by PINK1 to activate parkin. *Nature* **2014,** *510* (7503), 162.
(147) Wauer, T.; Simicek, M.; Schubert, A.; Komander, D. Mechanism of phospho-ubiquitin-induced PARKIN activation. *Nature* **2015,** *524* (7565), 370.
(148) Nguyen, T. N.; Padman, B. S.; Lazarou, M. Deciphering the Molecular Signals of PINK1/Parkin Mitophagy. *Trends Cell Biol* **2016,** *26* (10), 733.
(149) Morimoto, D.; Shirakawa, M. The evolving world of ubiquitin: transformed polyubiquitin chains. *Biomol Concepts* **2016,** *7* (3), 157.
(150) Lacoursiere, R. E.; O'Donoghue, P.; Shaw, G. S. Programmed ubiquitin acetylation using genetic code expansion reveals altered ubiquitination patterns. *FEBS Lett* **2019**, DOI:10.1002/1873-3468.13702 10.1002/1873-3468.13702.
(151) Ullmann, R.; Chien, C. D.; Avantaggiati, M. L.; Muller, S. An acetylation switch regulates SUMO-dependent protein interaction networks. *Mol Cell* **2012,** *46* (6), 759.
(152) Gartner, A.; Wagner, K.; Holper, S.; Kunz, K.; Rodriguez, M. S.; Muller, S. Acetylation of SUMO2 at lysine 11 favors the formation of non-canonical SUMO chains. *EMBO Rep* **2018,** *19* (11).
(153) Yan, F.; Huang, C.; Wang, X.; Tan, J.; Cheng, S.; Wan, M.; Wang, Z.; Wang, S.; Luo, S.; Li, A.et al. Threonine ADP-Ribosylation of Ubiquitin by a Bacterial Effector Family Blocks Host Ubiquitination. *Mol Cell* **2020,** *78* (4), 641.
(154) Mali, S. M.; Singh, S. K.; Eid, E.; Brik, A. Ubiquitin Signaling: Chemistry Comes to the Rescue. *J Am Chem Soc* **2017,** *139* (14), 4971.
(155) Bock, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. Selenocysteine: the 21st amino acid. *Mol Microbiol* **1991,** *5* (3), 515.
(156) Srinivasan, G.; James, C. M.; Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **2002,** *296* (5572), 1459.
(157) Wang, L.; Xie, J.; Schultz, P. G. Expanding the genetic code. *Annu Rev Biophys Biomol Struct* **2006,** *35*, 225.
(158) Kimmerlin, T.; Seebach, D. '100 years of peptide synthesis': ligation methods for peptide and protein synthesis with applications to β-peptide assemblies*. *The Journal of Peptide Research* **2005,** *65* (2), 229.
(159) Dawson, P. E.; Kent, S. B. H. Synthesis of Native Proteins by Chemical Ligation. *Annual Review of Biochemistry* **2000,** *69* (1), 923.

(160) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **1989,** *244* (4901), 182.
(161) Lee, N.; Bessho, Y.; Wei, K.; Szostak, J. W.; Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nature Structural Biology* **2000,** *7* (1), 28.
(162) Kent, S. B. Chemical synthesis of peptides and proteins. *Annu Rev Biochem* **1988,** *57*, 957.
(163) Mendel, D.; Cornish, V. W.; Schultz, P. G. Site-directed mutagenesis with an expanded genetic code. *Annu Rev Biophys Biomol Struct* **1995,** *24*, 435.
(164) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Current Opinion in Chemical Biology* **2010,** *14* (6), 774.
(165) Ngo, J. T.; Tirrell, D. A. Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis. *Accounts of Chemical Research* **2011,** *44* (9), 677.
(166) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annual Review of Biochemistry* **2010,** *79* (1), 413.
(167) Davis, L.; Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nature Reviews Molecular Cell Biology* **2012,** *13*, 168.
(168) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* **2014,** *114* (9), 4764.
(169) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr Opin Chem Biol* **2010,** *14* (6), 774.
(170) Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **1990,** *9* (5), 1665.
(171) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc Natl Acad Sci U S A* **2002,** *99* (1), 19.
(172) Tang, Y.; Wang, P.; Van Deventer, J. A.; Link, A. J.; Tirrell, D. A. Introduction of an aliphatic ketone into recombinant proteins in a bacterial strain that overexpresses an editing-impaired leucyl-tRNA synthetase. *Chembiochem* **2009,** *10* (13), 2188.
(173) Kast, P.; Hennecke, H. Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. *J Mol Biol* **1991,** *222* (1), 99.
(174) Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. Cell-selective metabolic labeling of proteins. *Nat Chem Biol* **2009,** *5* (10), 715.
(175) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A., 3rd; Tirrell, D. A. Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. *Proc Natl Acad Sci U S A* **2009,** *106* (36), 15285.
(176) Carrico, I. S.; Maskarinec, S. A.; Heilshorn, S. C.; Mock, M. L.; Liu, J. C.; Nowatzki, P. J.; Franck, C.; Ravichandran, G.; Tirrell, D. A. Lithographic patterning of photoreactive cell-adhesive proteins. *J Am Chem Soc* **2007,** *129* (16), 4874.
(177) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc Natl Acad Sci U S A* **2006,** *103* (25), 9482.

(178) Dieterich, D. C.; Hodas, J. J.; Gouzer, G.; Shadrin, I. Y.; Ngo, J. T.; Triller, A.; Tirrell, D. A.; Schuman, E. M. In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat Neurosci* **2010,** *13* (7), 897.
(179) Arranz-Gibert, P.; Patel, J. R.; Isaacs, F. J. The Role of Orthogonality in Genetic Code Expansion. *Life (Basel)* **2019,** *9* (3).
(180) Nakamura, Y.; Gojobori, T.; Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* **2000,** *28* (1), 292.
(181) Xie, J.; Schultz, P. G. An expanding genetic code. *Methods* **2005,** *36* (3), 227.
(182) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of Escherichia coli. *Science* **2001,** *292* (5516), 498.
(183) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nature Biotechnology* **2002,** *20* (10), 1044.
(184) Liu, D. R.; Schultz, P. G. Progress toward the evolution of an organism with an expanded genetic code. *Proceedings of the National Academy of Sciences* **1999,** *96* (9), 4780.
(185) Wang, L.; Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chemistry & Biology* **2001,** *8* (9), 883.
(186) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2014,** *1844* (6), 1059.
(187) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. An Expanded Eukaryotic Genetic Code. *Science* **2003,** *301* (5635), 964.
(188) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochemical and Biophysical Research Communications* **2008,** *371* (4), 818.
(189) Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W. Genetically Encoded Photocontrol of Protein Localization in Mammalian Cells. *Journal of the American Chemical Society* **2010,** *132* (12), 4086.
(190) Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nature Chemistry* **2012,** *4*, 298.
(191) Li, F.; Zhang, H.; Sun, Y.; Pan, Y.; Zhou, J.; Wang, J. Expanding the Genetic Code for Photoclick Chemistry in E. coli, Mammalian Cells, and A. thaliana. *Angewandte Chemie International Edition* **2013,** *52* (37), 9700.
(192) Greiss, S.; Chin, J. W. Expanding the Genetic Code of an Animal. *Journal of the American Chemical Society* **2011,** *133* (36), 14196.
(193) Bianco, A.; Townsley, F. M.; Greiss, S.; Lang, K.; Chin, J. W. Expanding the genetic code of Drosophila melanogaster. *Nature Chemical Biology* **2012,** *8*, 748.
(194) Ernst, R. J.; Krogager, T. P.; Maywood, E. S.; Zanchi, R.; Beránek, V.; Elliott, T. S.; Barry, N. P.; Hastings, M. H.; Chin, J. W. Genetic code expansion in the mouse brain. *Nature Chemical Biology* **2016,** *12* (10), 776.
(195) Zheng, Y.; Lewis, T. L.; Igo, P.; Polleux, F.; Chatterjee, A. Virus-Enabled Optimization and Delivery of the Genetic Machinery for Efficient Unnatural Amino Acid Mutagenesis in Mammalian Cells and Tissues. *ACS Synthetic Biology* **2017,** *6* (1), 13.
(196) Gautier, A.; Deiters, A.; Chin, J. W. Light-Activated Kinases Enable Temporal Dissection of Signaling Networks in Living Cells. *Journal of the American Chemical Society* **2011,** *133* (7), 2124.

(197) Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nature Chemical Biology* **2007,** *3*, 769.
(198) Coin, I.; Katritch, V.; Sun, T.; Xiang, Z.; Siu, Fai Y.; Beyermann, M.; Stevens, Raymond C.; Wang, L. Genetically Encoded Chemical Probes in Cells Reveal the Binding Path of Urocortin-I to CRF Class B GPCR. *Cell* **2013,** *155* (6), 1258.
(199) Nguyen, T.-A.; Cigler, M.; Lang, K. Expanding the Genetic Code to Study Protein–Protein Interactions. *Angewandte Chemie International Edition* **2018,** *57* (44), 14350.
(200) Wang, L. Genetically encoding new bioreactivity. *New Biotechnology* **2017,** *38*, 16.
(201) Cigler, M.; Müller, T. G.; Horn-Ghetko, D.; von Wrisberg, M.-K.; Fottner, M.; Goody, R. S.; Itzen, A.; Müller, M. P.; Lang, K. Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes In Vitro and In Vivo. *Angewandte Chemie International Edition* **2017,** *56* (49), 15737.
(202) Yang, B.; Tang, S.; Ma, C.; Li, S.-T.; Shao, G.-C.; Dang, B.; DeGrado, W. F.; Dong, M.-Q.; Wang, P. G.; Ding, S.et al. Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions. *Nature Communications* **2017,** *8* (1), 2240.
(203) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **2008,** *4* (4), 232.
(204) Kim, C. H.; Kang, M.; Kim, H. J.; Chatterjee, A.; Schultz, P. G. Site-specific incorporation of epsilon-N-crotonyllysine into histones. *Angew Chem Int Ed Engl* **2012,** *51* (29), 7246.
(205) Gattner, M. J.; Vrabel, M.; Carell, T. Synthesis of epsilon-N-propionyl-, epsilon-N-butyryl-, and epsilon-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem Commun (Camb)* **2013,** *49* (4), 379.
(206) Xiao, H.; Xuan, W.; Shao, S.; Liu, T.; Schultz, P. G. Genetic Incorporation of epsilon-N-2-Hydroxyisobutyryl-lysine into Recombinant Histones. *ACS Chem Biol* **2015,** *10* (7), 1599.
(207) Groff, D.; Chen, P. R.; Peters, F. B.; Schultz, P. G. A genetically encoded epsilon-N-methyl lysine in mammalian cells. *Chembiochem* **2010,** *11* (8), 1066.
(208) Wang, Y. S.; Wu, B.; Wang, Z.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Moe, Y. N.; Russell, D. H.; Liu, W. R. A genetically encoded photocaged Nepsilon-methyl-L-lysine. *Mol Biosyst* **2010,** *6* (9), 1557.
(209) Ai, H. W.; Lee, J. W.; Schultz, P. G. A method to site-specifically introduce methyllysine into proteins in E. coli. *Chem Commun (Camb)* **2010,** *46* (30), 5506.
(210) Park, H. S.; Hohn, M. J.; Umehara, T.; Guo, L. T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Soll, D. Expanding the genetic code of Escherichia coli with phosphoserine. *Science* **2011,** *333* (6046), 1151.
(211) Zhang, M. S.; Brunner, S. F.; Huguenin-Dezot, N.; Liang, A. D.; Schmied, W. H.; Rogerson, D. T.; Chin, J. W. Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat Methods* **2017,** *14* (7), 729.
(212) Fan, C.; Ip, K.; Soll, D. Expanding the genetic code of Escherichia coli with phosphotyrosine. *FEBS Lett* **2016,** *590* (17), 3040.
(213) Neumann, H.; Hazen, J. L.; Weinstein, J.; Mehl, R. A.; Chin, J. W. Genetically encoding protein oxidative damage. *J Am Chem Soc* **2008,** *130* (12), 4028.
(214) Liu, C. C.; Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat Biotechnol* **2006,** *24* (11), 1436.
(215) Mali, S. M.; Singh, S. K.; Eid, E.; Brik, A. Ubiquitin Signaling: Chemistry Comes to the Rescue. *Journal of the American Chemical Society* **2017**, DOI:10.1021/jacs.7b00089 10.1021/jacs.7b00089.

(216) Weller, C. E.; Pilkerton, M. E.; Chatterjee, C. Chemical strategies to understand the language of ubiquitin signaling. *Biopolymers* **2014,** *101* (2), 144.
(217) Hemantha, H. P.; Brik, A. Non-enzymatic synthesis of ubiquitin chains: where chemistry makes a difference. *Bioorg Med Chem* **2013,** *21* (12), 3411.
(218) Trang, V. H.; Valkevich, E. M.; Minami, S.; Chen, Y. C.; Ge, Y.; Strieter, E. R. Nonenzymatic polymerization of ubiquitin: single-step synthesis and isolation of discrete ubiquitin oligomers. *Angewandte Chemie* **2012,** *51* (52), 13085.
(219) Chen, J.; Ai, Y.; Wang, J.; Haracska, L.; Zhuang, Z. Chemically ubiquitylated PCNA as a probe for eukaryotic translesion DNA synthesis. *Nat Chem Biol* **2010,** *6* (4), 270.
(220) Jung, J. E.; Wollscheid, H. P.; Marquardt, A.; Manea, M.; Scheffner, M.; Przybylski, M. Functional ubiquitin conjugates with lysine-epsilon-amino-specific linkage by thioether ligation of cysteinyl-ubiquitin peptide building blocks. *Bioconjugate chemistry* **2009,** *20* (6), 1152.
(221) Weikart, N. D.; Mootz, H. D. Generation of site-specific and enzymatically stable conjugates of recombinant proteins with ubiquitin-like modifiers by the Cu(I)-catalyzed azide-alkyne cycloaddition. *Chembiochem* **2010,** *11* (6), 774.
(222) Li, X.; Fekner, T.; Ottesen, J. J.; Chan, M. K. A pyrrolysine analogue for site-specific protein ubiquitination. *Angew Chem Int Ed Engl* **2009,** *48* (48), 9184.
(223) Virdee, S.; Kapadnis, P. B.; Elliott, T.; Lang, K.; Madrzak, J.; Nguyen, D. P.; Riechmann, L.; Chin, J. W. Traceless and site-specific ubiquitination of recombinant proteins. *J Am Chem Soc* **2011,** *133* (28), 10708.
(224) Virdee, S.; Ye, Y.; Nguyen, D. P.; Komander, D.; Chin, J. W. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nature chemical biology* **2010,** *6* (10), 750.
(225) Bi, X.; Pasunooti, K. K.; Tareq, A. H.; Takyi-Williams, J.; Liu, C.-F. Genetic incorporation of 1,2-aminothiol functionality for site-specific protein modification via thiazolidine formation. *Organic & Biomolecular Chemistry* **2016,** *14* (23), 5282.
(226) Stanley, M.; Virdee, S. Genetically Directed Production of Recombinant, Isosteric and Nonhydrolysable Ubiquitin Conjugates. *Chembiochem* **2016,** *17* (15), 1472.
(227) Eger, S.; Scheffner, M.; Marx, A.; Rubini, M. Synthesis of defined ubiquitin dimers. *J Am Chem Soc* **2010,** *132* (46), 16337.
(228) Sommer, S.; Weikart, N. D.; Brockmeyer, A.; Janning, P.; Mootz, H. D. Expanded click conjugation of recombinant proteins with ubiquitin-like modifiers reveals altered substrate preference of SUMO2-modified Ubc9. *Angew Chem Int Ed Engl* **2011,** *50* (42), 9888.
(229) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* **2009,** *48* (38), 6974.
(230) Lang, K.; Chin, J. W. Bioorthogonal reactions for labeling proteins. *ACS Chem Biol* **2014,** *9* (1), 16.
(231) Thirumurugan, P.; Matosiuk, D.; Jozwiak, K. Click chemistry for drug development and diverse chemical-biology applications. *Chem Rev* **2013,** *113* (7), 4905.
(232) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *J Am Chem Soc* **2009,** *131* (25), 8720.
(233) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J Am Chem Soc* **2002,** *124* (31), 9026.
(234) Schneider, T.; Schneider, D.; Rosner, D.; Malhotra, S.; Mortensen, F.; Mayer, T. U.; Scheffner, M.; Marx, A. Dissecting ubiquitin signaling with linkage-defined and protease resistant ubiquitin chains. *Angew Chem Int Ed Engl* **2014,** *53* (47), 12925.

(235) Choe, K. N.; Moldovan, G. L. Forging Ahead through Darkness: PCNA, Still the Principal Conductor at the Replication Fork. *Mol Cell* **2017,** *65* (3), 380.
(236) Eger, S.; Castrec, B.; Hubscher, U.; Scheffner, M.; Rubini, M.; Marx, A. Generation of a mono-ubiquitinated PCNA mimic by click chemistry. *Chembiochem* **2011,** *12* (18), 2807.
(237) Rosner, D.; Schneider, T.; Schneider, D.; Scheffner, M.; Marx, A. Click chemistry for targeted protein ubiquitylation and ubiquitin chain formation. *Nat Protoc* **2015,** *10* (10), 1594.
(238) Conibear, A. C.; Watson, E. E.; Payne, R. J.; Becker, C. F. W. Native chemical ligation in protein synthesis and semi-synthesis. *Chem Soc Rev* **2018,** *47* (24), 9046.
(239) Weikart, N. D.; Sommer, S.; Mootz, H. D. Click synthesis of ubiquitin dimer analogs to interrogate linkage-specific UBA domain binding. *Chem Commun (Camb)* **2012,** *48* (2), 296.
(240) Dresselhaus, T.; Weikart, N. D.; Mootz, H. D.; Waller, M. P. Naturally and synthetically linked lys48 diubiquitin: a QM/MM study. *RSC Advances* **2013,** *3* (36), 16122.
(241) Hagmann, V.; Sommer, S.; Fabian, P.; Bierlmeier, J.; van Treel, N.; Mootz, H. D.; Schwarzer, D.; Azevedo, J. E.; Dodt, G. Chemically monoubiquitinated PEX5 binds to the components of the peroxisomal docking and export machinery. *Sci Rep* **2018,** *8* (1), 16014.
(242) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Ed Engl* **2009,** *48* (52), 9879.
(243) Hackenberger, C. P.; Schwarzer, D. Chemoselective ligation and modification strategies for peptides and proteins. *Angew Chem Int Ed Engl* **2008,** *47* (52), 10030.
(244) Ulrich, S.; Boturyn, D.; Marra, A.; Renaudet, O.; Dumy, P. Oxime ligation: a chemoselective click-type reaction for accessing multifunctional biomolecular constructs. *Chemistry* **2014,** *20* (1), 34.
(245) Shanmugham, A.; Fish, A.; Luna-Vargas, M. P.; Faesen, A. C.; El Oualid, F.; Sixma, T. K.; Ovaa, H. Nonhydrolyzable ubiquitin-isopeptide isosteres as deubiquitinating enzyme probes. *J Am Chem Soc* **2010,** *132* (26), 8834.
(246) Wilkinson, K. D.; Gan-Erdene, T.; Kolli, N. Derivitization of the C-terminus of ubiquitin and ubiquitin-like proteins using intein chemistry: methods and uses. *Methods Enzymol* **2005,** *399*, 37.
(247) Wang, Z.-P. A.; Tian, C.-L.; Zheng, J.-S. The recent developments and applications of the traceless-Staudinger reaction in chemical biology study. *RSC Advances* **2015,** *5* (130), 107192.
(248) Dumas, A. M.; Molander, G. A.; Bode, J. W. Amide-forming ligation of acyltrifluoroborates and hydroxylamines in water. *Angew Chem Int Ed Engl* **2012,** *51* (23), 5683.
(249) Aimoto, S. Polypeptide synthesis by the thioester method. *Biopolymers* **1999,** *51* (4), 247.
(250) Tan, Z.; Shang, S.; Halkina, T.; Yuan, Y.; Danishefsky, S. J. Toward homogeneous erythropoietin: non-NCL-based chemical synthesis of the Gln78-Arg166 glycopeptide domain. *J Am Chem Soc* **2009,** *131* (15), 5424.
(251) Tam, J. P.; Heath, W. F.; Merrifield, R. B. Mechanisms for the removal of benzyl protecting groups in synthetic peptides by trifluoromethanesulfonic acid-trifluoroacetic acid-dimethyl sulfide. *Journal of the American Chemical Society* **1986,** *108* (17), 5242.
(252) Licchesi, J. D.; Mieszczanek, J.; Mevissen, T. E.; Rutherford, T. J.; Akutsu, M.; Virdee, S.; El Oualid, F.; Chin, J. W.; Ovaa, H.; Bienz, M.et al. An ankyrin-repeat

ubiquitin-binding domain determines TRABID's specificity for atypical ubiquitin chains. *Nat Struct Mol Biol* **2011,** *19* (1), 62.
(253) Akutsu, M.; Ye, Y.; Virdee, S.; Chin, J. W.; Komander, D. Molecular basis for ubiquitin and ISG15 cross-reactivity in viral ovarian tumor domains. *Proc Natl Acad Sci U S A* **2011,** *108* (6), 2228.
(254) Ritorto, M. S.; Ewan, R.; Perez-Oliva, A. B.; Knebel, A.; Buhrlage, S. J.; Wightman, M.; Kelly, S. M.; Wood, N. T.; Virdee, S.; Gray, N. S.et al. Screening of DUB activity and specificity by MALDI-TOF mass spectrometry. *Nat Commun* **2014,** *5*, 4763.
(255) Bett, J. S.; Ritorto, M. S.; Ewan, R.; Jaffray, E. G.; Virdee, S.; Chin, J. W.; Knebel, A.; Kurz, T.; Trost, M.; Tatham, M. H.et al. Ubiquitin C-terminal hydrolases cleave isopeptide- and peptide-linked ubiquitin from structured proteins but do not edit ubiquitin homopolymers. *Biochem J* **2015,** *466* (3), 489.
(256) Castaneda, C.; Liu, J.; Chaturvedi, A.; Nowicka, U.; Cropp, T. A.; Fushman, D. Nonenzymatic assembly of natural polyubiquitin chains of any linkage composition and isotopic labeling scheme. *J Am Chem Soc* **2011,** *133* (44), 17855.
(257) Castaneda, C. A.; Dixon, E. K.; Walker, O.; Chaturvedi, A.; Nakasone, M. A.; Curtis, J. E.; Reed, M. R.; Krueger, S.; Cropp, T. A.; Fushman, D. Linkage via K27 Bestows Ubiquitin Chains with Unique Properties among Polyubiquitins. *Structure* **2016,** *24* (3), 423.
(258) Castaneda, C. A.; Kashyap, T. R.; Nakasone, M. A.; Krueger, S.; Fushman, D. Unique structural, dynamical, and functional properties of k11-linked polyubiquitin chains. *Structure* **2013,** *21* (7), 1168.
(259) Castaneda, C. A.; Chaturvedi, A.; Camara, C. M.; Curtis, J. E.; Krueger, S.; Fushman, D. Linkage-specific conformational ensembles of non-canonical polyubiquitin chains. *Phys Chem Chem Phys* **2016,** *18* (8), 5771.
(260) Dixon, E. K.; Castaneda, C. A.; Kashyap, T. R.; Wang, Y.; Fushman, D. Nonenzymatic assembly of branched polyubiquitin chains for structural and biochemical studies. *Bioorg Med Chem* **2013,** *21* (12), 3421.
(261) Lee, A. E.; Geis-Asteggiante, L.; Dixon, E. K.; Kim, Y.; Kashyap, T. R.; Wang, Y.; Fushman, D.; Fenselau, C. Preparing to read the ubiquitin code: characterization of ubiquitin trimers by top-down mass spectrometry. *J Mass Spectrom* **2016,** *51* (4), 315.
(262) Lee, A. E.; Geis-Asteggiante, L.; Dixon, E. K.; Miller, M.; Wang, Y.; Fushman, D.; Fenselau, C. Preparing to read the ubiquitin code: top-down analysis of unanchored ubiquitin tetramers. *J Mass Spectrom* **2016,** *51* (8), 629.
(263) Enchev, R. I.; Schulman, B. A.; Peter, M. Protein neddylation: beyond cullin-RING ligases. *Nat Rev Mol Cell Biol* **2015,** *16* (1), 30.
(264) Singh, R. K.; Sundar, A.; Fushman, D. Nonenzymatic rubylation and ubiquitination of proteins for structural and functional studies. *Angew Chem Int Ed Engl* **2014,** *53* (24), 6120.
(265) Madrzak, J.; Fiedler, M.; Johnson, C. M.; Ewan, R.; Knebel, A.; Bienz, M.; Chin, J. W. Ubiquitination of the Dishevelled DIX domain blocks its head-to-tail polymerization. *Nat Commun* **2015,** *6*, 6718.
(266) Dawson, P. E.; Kent, S. B. Synthesis of native proteins by chemical ligation. *Annu Rev Biochem* **2000,** *69*, 923.
(267) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoet, M.; Monbaliu, J. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem Rev* **2019,** *119* (12), 7328.
(268) Wan, Q.; Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* **2007,** *46* (48), 9248.

(269) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *Journal of the American Chemical Society* **2001,** *123* (4), 526.
(270) Chatterjee, C.; McGinty, R. K.; Pellois, J. P.; Muir, T. W. Auxiliary-mediated site-specific peptide ubiquitylation. *Angew Chem Int Ed Engl* **2007,** *46* (16), 2814.
(271) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008,** *453* (7196), 812.
(272) Ajish Kumar, K. S.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A. Highly efficient and chemoselective peptide ubiquitylation. *Angew Chem Int Ed Engl* **2009,** *48* (43), 8090.
(273) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. Dual native chemical ligation at lysine. *J Am Chem Soc* **2009,** *131* (38), 13592.
(274) Bavikar, S. N.; Spasser, L.; Haj-Yahya, M.; Karthikeyan, S. V.; Moyal, T.; Kumar, K. S.; Brik, A. Chemical synthesis of ubiquitinated peptides with varying lengths and types of ubiquitin chains to explore the activity of deubiquitinases. *Angew Chem Int Ed Engl* **2012,** *51* (3), 758.
(275) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. *Chem Commun (Camb)* **2010,** *46* (38), 7199.
(276) Kumar, K. S.; Spasser, L.; Erlich, L. A.; Bavikar, S. N.; Brik, A. Total chemical synthesis of di-ubiquitin chains. *Angew Chem Int Ed Engl* **2010,** *49* (48), 9126.
(277) Kumar, K. S.; Bavikar, S. N.; Spasser, L.; Moyal, T.; Ohayon, S.; Brik, A. Total chemical synthesis of a 304 amino acid K48-linked tetraubiquitin protein. *Angew Chem Int Ed Engl* **2011,** *50* (27), 6137.
(278) Haj-Yahya, M.; Fauvet, B.; Herman-Bachinsky, Y.; Hejjaoui, M.; Bavikar, S. N.; Karthikeyan, S. V.; Ciechanover, A.; Lashuel, H. A.; Brik, A. Synthetic polyubiquitinated alpha-Synuclein reveals important insights into the roles of the ubiquitin chain in regulating its pathophysiology. *Proc Natl Acad Sci U S A* **2013,** *110* (44), 17726.
(279) Sun, H.; Meledin, R.; Mali, S. M.; Brik, A. Total chemical synthesis of ester-linked ubiquitinated proteins unravels their behavior with deubiquitinases. *Chem Sci* **2018,** *9* (6), 1661.
(280) Thompson, R. E.; Muir, T. W. Chemoenzymatic Semisynthesis of Proteins. *Chem Rev* **2020,** *120* (6), 3051.
(281) Li, Y. Split-inteins and their bioapplications. *Biotechnol Lett* **2015,** *37* (11), 2121.
(282) Mazmanian, S. K.; Liu, G.; Ton-That, H.; Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **1999,** *285* (5428), 760.
(283) Spirig, T.; Weiner, E. M.; Clubb, R. T. Sortase enzymes in Gram-positive bacteria. *Mol Microbiol* **2011,** *82* (5), 1044.
(284) Malik, A.; Kim, S. B. A comprehensive in silico analysis of sortase superfamily. *J Microbiol* **2019,** *57* (6), 431.
(285) Huang, X.; Aulabaugh, A.; Ding, W.; Kapoor, B.; Alksne, L.; Tabei, K.; Ellestad, G. Kinetic mechanism of Staphylococcus aureus sortase SrtA. *Biochemistry* **2003,** *42* (38), 11307.
(286) Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* **2004,** *126* (9), 2670.
(287) Pishesha, N.; Ingram, J. R.; Ploegh, H. L. Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annu Rev Cell Dev Biol* **2018,** *34*, 163.

(288) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **2007,** *3* (11), 707.
(289) Popp, M. W.; Ploegh, H. L. Making and breaking peptide bonds: protein engineering using sortase. *Angew Chem Int Ed Engl* **2011,** *50* (22), 5024.
(290) Kobashigawa, Y.; Kumeta, H.; Ogura, K.; Inagaki, F. Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. *J Biomol NMR* **2009,** *43* (3), 145.
(291) Freiburger, L.; Sonntag, M.; Hennig, J.; Li, J.; Zou, P.; Sattler, M. Efficient segmental isotope labeling of multi-domain proteins using Sortase A. *J Biomol NMR* **2015,** *63* (1), 1.
(292) Yamamura, Y.; Hirakawa, H.; Yamaguchi, S.; Nagamune, T. Enhancement of sortase A-mediated protein ligation by inducing a beta-hairpin structure around the ligation site. *Chem Commun (Camb)* **2011,** *47* (16), 4742.
(293) Williamson, D. J.; Fascione, M. A.; Webb, M. E.; Turnbull, W. B. Efficient N-terminal labeling of proteins by use of sortase. *Angew Chem Int Ed Engl* **2012,** *51* (37), 9377.
(294) David Row, R.; Roark, T. J.; Philip, M. C.; Perkins, L. L.; Antos, J. M. Enhancing the efficiency of sortase-mediated ligations through nickel-peptide complex formation. *Chem Commun (Camb)* **2015,** *51* (63), 12548.
(295) Reed, S. A.; Brzovic, D. A.; Takasaki, S. S.; Boyko, K. V.; Antos, J. M. Efficient Sortase-Mediated Ligation Using a Common C-Terminal Fusion Tag. *Bioconjug Chem* **2020,** *31* (5), 1463.
(296) Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc Natl Acad Sci U S A* **1999,** *96* (22), 12424.
(297) Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* **2011,** *108* (28), 11399.
(298) Hirakawa, H.; Ishikawa, S.; Nagamune, T. Ca2+ -independent sortase-A exhibits high selective protein ligation activity in the cytoplasm of Escherichia coli. *Biotechnol J* **2015,** *10* (9), 1487.
(299) Hirakawa, H.; Ishikawa, S.; Nagamune, T. Design of Ca2+-independent Staphylococcus aureus sortase A mutants. *Biotechnol Bioeng* **2012,** *109* (12), 2955.
(300) Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R. Reprogramming the specificity of sortase enzymes. *Proc Natl Acad Sci U S A* **2014,** *111* (37), 13343.
(301) Suree, N.; Liew, C. K.; Villareal, V. A.; Thieu, W.; Fadeev, E. A.; Clemens, J. J.; Jung, M. E.; Clubb, R. T. The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J Biol Chem* **2009,** *284* (36), 24465.
(302) Piotukh, K.; Geltinger, B.; Heinrich, N.; Gerth, F.; Beyermann, M.; Freund, C.; Schwarzer, D. Directed evolution of sortase A mutants with altered substrate selectivity profiles. *J Am Chem Soc* **2011,** *133* (44), 17536.
(303) Bentley, M. L.; Gaweska, H.; Kielec, J. M.; McCafferty, D. G. Engineering the substrate specificity of Staphylococcus aureus Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J Biol Chem* **2007,** *282* (9), 6571.
(304) Gianella, P.; Snapp, E. L.; Levy, M. An in vitro compartmentalization-based method for the selection of bond-forming enzymes from large libraries. *Biotechnol Bioeng* **2016,** *113* (8), 1647.
(305) Zou, Z.; Alibiglou, H.; Mate, D. M.; Davari, M. D.; Jakob, F.; Schwaneberg, U. Directed sortase A evolution for efficient site-specific bioconjugations in organic co-solvents. *Chem Commun (Camb)* **2018,** *54* (81), 11467.

(306) Zou, Z.; Mate, D. M.; Rubsam, K.; Jakob, F.; Schwaneberg, U. Sortase-Mediated High-Throughput Screening Platform for Directed Enzyme Evolution. *ACS Comb Sci* **2018,** *20* (4), 203.
(307) Jeong, H. J.; Abhiraman, G. C.; Story, C. M.; Ingram, J. R.; Dougan, S. K. Generation of Ca2+-independent sortase A mutants with enhanced activity for protein and cell surface labeling. *PLoS One* **2017,** *12* (12), e0189068.
(308) Witte, M. D.; Wu, T.; Guimaraes, C. P.; Theile, C. S.; Blom, A. E.; Ingram, J. R.; Li, Z.; Kundrat, L.; Goldberg, S. D.; Ploegh, H. L. Site-specific protein modification using immobilized sortase in batch and continuous-flow systems. *Nat Protoc* **2015,** *10* (3), 508.
(309) Wu, Q.; Ploegh, H. L.; Truttmann, M. C. Hepta-Mutant Staphylococcus aureus Sortase A (SrtA7m) as a Tool for in Vivo Protein Labeling in Caenorhabditis elegans. *ACS Chem Biol* **2017,** *12* (3), 664.
(310) Nguyen, G. K.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat Chem Biol* **2014,** *10* (9), 732.
(311) Harris, K. S.; Durek, T.; Kaas, Q.; Poth, A. G.; Gilding, E. K.; Conlan, B. F.; Saska, I.; Daly, N. L.; van der Weerden, N. L.; Craik, D. J.et al. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat Commun* **2015,** *6*, 10199.
(312) Mikula, K. M.; Tascon, I.; Tommila, J. J.; Iwai, H. Segmental isotopic labeling of a single-domain globular protein without any refolding step by an asparaginyl endopeptidase. *FEBS Lett* **2017,** *591* (9), 1285.
(313) Yang, R.; Wong, Y. H.; Nguyen, G. K. T.; Tam, J. P.; Lescar, J.; Wu, B. Engineering a Catalytically Efficient Recombinant Protein Ligase. *J Am Chem Soc* **2017,** *139* (15), 5351.
(314) Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A. Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **1991,** *30* (17), 4151.
(315) Chang, T. K.; Jackson, D. Y.; Burnier, J. P.; Wells, J. A. Subtiligase: a tool for semisynthesis of proteins. *Proc Natl Acad Sci U S A* **1994,** *91* (26), 12544.
(316) Tan, X.; Yang, R.; Liu, C. F. Facilitating Subtiligase-Catalyzed Peptide Ligation Reactions by Using Peptide Thioester Substrates. *Org Lett* **2018,** *20* (21), 6691.
(317) Weeks, A. M.; Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat Chem Biol* **2018,** *14* (1), 50.
(318) Mahrus, S.; Trinidad, J. C.; Barkan, D. T.; Sali, A.; Burlingame, A. L.; Wells, J. A. Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **2008,** *134* (5), 866.
(319) Henager, S. H.; Chu, N.; Chen, Z.; Bolduc, D.; Dempsey, D. R.; Hwang, Y.; Wells, J.; Cole, P. A. Enzyme-catalyzed expressed protein ligation. *Nat Methods* **2016,** *13* (11), 925.
(320) Toplak, A.; Nuijens, T.; Quaedflieg, P. J. L. M.; Wu, B.; Janssen, D. B. Peptiligase, an Enzyme for Efficient Chemoenzymatic Peptide Synthesis and Cyclization in Water. *Advanced Synthesis & Catalysis* **2016,** *358* (13), 2140.
(321) Schmidt, M.; Toplak, A.; Quaedflieg, P. J. L. M.; Ippel, H.; Richelle, G. J. J.; Hackeng, T. M.; van Maarseveen, J. H.; Nuijens, T. Omniligase-1: A Powerful Tool for Peptide Head-to-Tail Cyclization. *Advanced Synthesis & Catalysis* **2017,** *359* (12), 2050.
(322) Luo, J.; Liu, Q.; Morihiro, K.; Deiters, A. Small-molecule control of protein function through Staudinger reduction. *Nat Chem* **2016,** *8* (11), 1027.

(323) Lang, K.; Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem Rev* **2014,** *114* (9), 4764.
(324) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annual Review of Biochemistry, Vol 79* **2010,** *79*, 413.
(325) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing logical codon reassignment - Expanding the chemistry in biology. *Chem Sci* **2015,** *6* (1), 50.
(326) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim Biophys Acta* **2014,** *1844* (6), 1059.
(327) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (27), 11268.
(328) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498.
(329) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically encoding Nε-acetyllysine in recombinant proteins. *Nature chemical biology* **2008,** *4* (4), 232.
(330) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat Protoc* **2007,** *2* (10), 2601.
(331) Stanley, M.; Virdee, S. Chemical ubiquitination for decrypting a cellular code. *Biochem J* **2016,** *473* (10), 1297.
(332) Bekes, M.; Okamoto, K.; Crist, S. B.; Jones, M. J.; Chapman, J. R.; Brasher, B. B.; Melandri, F. D.; Ueberheide, B. M.; Denchi, E. L.; Huang, T. T. DUB-resistant ubiquitin to survey ubiquitination switches in mammalian cells. *Cell reports* **2013,** *5* (3), 826.
(333) Dikic, I.; Wakatsuki, S.; Walters, K. J. Ubiquitin-binding domains - from structures to functions. *Nature reviews. Molecular cell biology* **2009,** *10* (10), 659.
(334) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 A resolution. *Journal of molecular biology* **1987,** *194* (3), 531.
(335) Cook, W. J.; Jeffrey, L. C.; Carson, M.; Chen, Z.; Pickart, C. M. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *The Journal of biological chemistry* **1992,** *267* (23), 16467.
(336) Komander, D.; Reyes-Turcu, F.; Licchesi, J. D.; Odenwaelder, P.; Wilkinson, K. D.; Barford, D. Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO reports* **2009,** *10* (5), 466.
(337) Chen, Z. J.; Sun, L. J. Nonproteolytic functions of ubiquitin in cell signaling. *Molecular cell* **2009,** *33* (3), 275.
(338) Kulathu, Y.; Akutsu, M.; Bremm, A.; Hofmann, K.; Komander, D. Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain. *Nat Struct Mol Biol* **2009,** *16* (12), 1328.
(339) Zhang, X.; Smits, A. H.; van Tilburg, G. B.; Jansen, P. W.; Makowski, M. M.; Ovaa, H.; Vermeulen, M. An Interaction Landscape of Ubiquitin Signaling. *Molecular cell* **2017,** *65* (5), 941.
(340) Sims, J. J.; Cohen, R. E. Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of rap80. *Mol Cell* **2009,** *33* (6), 775.
(341) Sato, Y.; Yoshikawa, A.; Mimura, H.; Yamashita, M.; Yamagata, A.; Fukai, S. Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *EMBO J* **2009,** *28* (16), 2461.
(342) Bravo, R.; Frank, R.; Blundell, P. A.; Macdonald-Bravo, H. Cyclin/PCNA is the auxiliary protein of DNA polymerase-delta. *Nature* **1987,** *326* (6112), 515.

(343) Hoege, C.; Pfander, B.; Moldovan, G. L.; Pyrowolakis, G.; Jentsch, S. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **2002,** *419* (6903), 135.
(344) Chen, J.; Bozza, W.; Zhuang, Z. Ubiquitination of PCNA and its essential role in eukaryotic translesion synthesis. *Cell biochemistry and biophysics* **2011,** *60* (1-2), 47.
(345) Cohn, M. A.; Kowal, P.; Yang, K.; Haas, W.; Huang, T. T.; Gygi, S. P.; D'Andrea, A. D. A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway. *Molecular cell* **2007,** *28* (5), 786.
(346) Flotho, A.; Melchior, F. Sumoylation: a regulatory protein modification in health and disease. *Annual review of biochemistry* **2013,** *82*, 357.
(347) Race, P. R.; Bentley, M. L.; Melvin, J. A.; Crow, A.; Hughes, R. K.; Smith, W. D.; Sessions, R. B.; Kehoe, M. A.; McCafferty, D. G.; Banfield, M. J. Crystal structure of Streptococcus pyogenes sortase A: implications for sortase mechanism. *The Journal of biological chemistry* **2009,** *284* (11), 6924.
(348) Weiner, E. M.; Robson, S.; Marohn, M.; Clubb, R. T. The Sortase A enzyme that attaches proteins to the cell wall of Bacillus anthracis contains an unusual active site architecture. *The Journal of biological chemistry* **2010,** *285* (30), 23433.
(349) Schmied, W. H.; Elsasser, S. J.; Uttamapinant, C.; Chin, J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *Journal of the American Chemical Society* **2014,** *136* (44), 15577.
(350) Sletten, E. M.; Bertozzi, C. R. From mechanism to mouse: a tale of two bioorthogonal reactions. *Accounts of chemical research* **2011,** *44* (9), 666.
(351) Chen, X.; Zaro, J. L.; Shen, W. C. Fusion protein linkers: property, design and functionality. *Advanced drug delivery reviews* **2013,** *65* (10), 1357.
(352) Keren-Kaplan, T.; Attali, I.; Motamedchaboki, K.; Davis, B. A.; Tanner, N.; Reshef, Y.; Laudon, E.; Kolot, M.; Levin-Kravets, O.; Kleifeld, O.et al. Synthetic biology approach to reconstituting the ubiquitylation cascade in bacteria. *EMBO J* **2012,** *31* (2), 378.
(353) Elsasser, S. J.; Ernst, R. J.; Walker, O. S.; Chin, J. W. Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nature methods* **2016,** *13* (2), 158.
(354) Ciechanover, A.; Stanhill, A. The complexity of recognition of ubiquitinated substrates by the 26S proteasome. *Biochim Biophys Acta* **2014,** *1843* (1), 86.
(355) Haakonsen, D. L.; Rape, M. Branching Out: Improved Signaling by Heterotypic Ubiquitin Chains. *Trends Cell Biol* **2019,** *29* (9), 704.
(356) Antos, J. M.; Chew, G. L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W.; Ploegh, H. L. Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J Am Chem Soc* **2009,** *131* (31), 10800.
(357) Popp, M. W.; Dougan, S. K.; Chuang, T. Y.; Spooner, E.; Ploegh, H. L. Sortase-catalyzed transformations that improve the properties of cytokines. *Proc Natl Acad Sci U S A* **2011,** *108* (8), 3169.
(358) Raeeszadeh-Sarmazdeh, M.; Parthasarathy, R.; Boder, E. T. Site-specific immobilization of protein layers on gold surfaces via orthogonal sortases. *Colloids Surf B Biointerfaces* **2015,** *128*, 457.
(359) Hess, G. T.; Cragnolini, J. J.; Popp, M. W.; Allen, M. A.; Dougan, S. K.; Spooner, E.; Ploegh, H. L.; Belcher, A. M.; Guimaraes, C. P. M13 bacteriophage display framework that allows sortase-mediated modification of surface-accessible phage proteins. *Bioconjug Chem* **2012,** *23* (7), 1478.

(360) Fottner, M.; Brunner, A. D.; Bittl, V.; Horn-Ghetko, D.; Jussupow, A.; Kaila, V. R. I.; Bremm, A.; Lang, K. Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat Chem Biol* **2019,** *15* (3), 276.
(361) Lombardi, P. M.; Matunis, M. J.; Wolberger, C. RAP80, ubiquitin and SUMO in the DNA damage response. *J Mol Med (Berl)* **2017,** *95* (8), 799.
(362) Anamika; Spyracopoulos, L. Molecular Basis for Phosphorylation-dependent SUMO Recognition by the DNA Repair Protein RAP80. *J Biol Chem* **2016,** *291* (9), 4417.
(363) Tatham, M. H.; Geoffroy, M. C.; Shen, L.; Plechanovova, A.; Hattersley, N.; Jaffray, E. G.; Palvimo, J. J.; Hay, R. T. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat Cell Biol* **2008,** *10* (5), 538.
(364) Bondalapati, S.; Eid, E.; Mali, S. M.; Wolberger, C.; Brik, A. Total chemical synthesis of SUMO-2-Lys63-linked diubiquitin hybrid chains assisted by removable solubilizing tags. *Chem Sci* **2017,** *8* (5), 4027.
(365) Meyer, H. J.; Rape, M. Enhanced protein degradation by branched ubiquitin chains. *Cell* **2014,** *157* (4), 910.
(366) Boughton, A. J.; Krueger, S.; Fushman, D. Branching via K11 and K48 Bestows Ubiquitin Chains with a Unique Interdomain Interface and Enhanced Affinity for Proteasomal Subunit Rpn1. *Structure* **2020,** *28* (1), 29.
(367) Sun, H.; Mali, S. M.; Singh, S. K.; Meledin, R.; Brik, A.; Kwon, Y. T.; Kravtsova-Ivantsiv, Y.; Bercovich, B.; Ciechanover, A. Diverse fate of ubiquitin chain moieties: The proximal is degraded with the target, and the distal protects the proximal from removal and recycles. *Proc Natl Acad Sci U S A* **2019,** *116* (16), 7805.
(368) Yau, R. G.; Doerner, K.; Castellanos, E. R.; Haakonsen, D. L.; Werner, A.; Wang, N.; Yang, X. W.; Martinez-Martin, N.; Matsumoto, M. L.; Dixit, V. M.et al. Assembly and Function of Heterotypic Ubiquitin Chains in Cell-Cycle and Protein Quality Control. *Cell* **2017,** *171* (4), 918.
(369) Tang, S.; Liang, L. J.; Si, Y. Y.; Gao, S.; Wang, J. X.; Liang, J.; Mei, Z.; Zheng, J. S.; Liu, L. Practical Chemical Synthesis of Atypical Ubiquitin Chains by Using an Isopeptide-Linked Ub Isomer. *Angew Chem Int Ed Engl* **2017,** *56* (43), 13333.
(370) Zhao, X.; Scheffner, M.; Marx, A. Assembly of branched ubiquitin oligomers by click chemistry. *Chem Commun (Camb)* **2019,** *55* (87), 13093.
(371) de la Pena, A. H.; Goodall, E. A.; Gates, S. N.; Lander, G. C.; Martin, A. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. *Science* **2018,** *362* (6418).
(372) Dong, Y.; Zhang, S.; Wu, Z.; Li, X.; Wang, W. L.; Zhu, Y.; Stoilova-McPhie, S.; Lu, Y.; Finley, D.; Mao, Y. Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **2019,** *565* (7737), 49.
(373) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A., 3rd; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **2009,** *6* (5), 343.
(374) Chiu, J.; March, P. E.; Lee, R.; Tillett, D. Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res* **2004,** *32* (21), e174.
(375) Seitchik, J. L.; Peeler, J. C.; Taylor, M. T.; Blackman, M. L.; Rhoads, T. W.; Cooley, R. B.; Refakis, C.; Fox, J. M.; Mehl, R. A. Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes. *Journal of the American Chemical Society* **2012,** *134* (6), 2898.
(376) Schmidt, M. J.; Summerer, D. Directed Evolution of Orthogonal Pyrrolysyl-tRNA Synthetases in Escherichia coli for the Genetic Encoding of Noncanonical Amino Acids. *Methods Mol Biol* **2018,** *1728*, 97.

(377) Boussif, O.; Lezoualc'h, F.; Zanta, M. A.; Mergny, M. D.; Scherman, D.; Demeneix, B.; Behr, J.-P. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proceedings of the National Academy of Sciences* **1995,** *92* (16), 7297.
(378) Brabez, N.; Lynch, R. M.; Xu, L.; Gillies, R. J.; Chassaing, G.; Lavielle, S.; Hruby, V. J. Design, synthesis, and biological studies of efficient multivalent melanotropin ligands: tools toward melanoma diagnosis and treatment. *Journal of medicinal chemistry* **2011,** *54* (20), 7375.
(379) Stemmer, W. P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **1994,** *91* (22), 10747.
(380) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat Biotechnol* **2002,** *20* (10), 1044.
(381) Seitchik, J. L.; Peeler, J. C.; Taylor, M. T.; Blackman, M. L.; Rhoads, T. W.; Cooley, R. B.; Refakis, C.; Fox, J. M.; Mehl, R. A. Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes. *Journal of the American Chemical Society* **2012,** *134* (6), 2898.
(382) Blizzard, R. J.; Backus, D. R.; Brown, W.; Bazewicz, C. G.; Li, Y.; Mehl, R. A. Ideal Bioorthogonal Reactions Using A Site-Specifically Encoded Tetrazine Amino Acid. *Journal of the American Chemical Society* **2015,** *137* (32), 10044.
(383) Shen, A.; Lupardus, P. J.; Morell, M.; Ponder, E. L.; Sadaghiani, A. M.; Garcia, K. C.; Bogyo, M. Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. *PloS one* **2009,** *4* (12), e8119.
(384) Berndsen, C. E.; Wolberger, C. A spectrophotometric assay for conjugation of ubiquitin and ubiquitin-like proteins. *Analytical biochemistry* **2011,** *418* (1), 102.
(385) Komander, D.; Lord, C. J.; Scheel, H.; Swift, S.; Hofmann, K.; Ashworth, A.; Barford, D. The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B box module. *Molecular cell* **2008,** *29* (4), 451.
(386) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water *The Journal of Chemical Physics* **1983,** *79*, 926.
(387) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *Journal of chemical theory and computation* **2012,** *8* (9), 3257.
(388) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98*, 10089.
(389) Abraham, M. J.; Murtola, T.; Schulz, R.; Pálla, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Software X* **2015,** *1-2*, 19.
(390) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **1996,** *14* (1), 33.
(391) Renatus, M.; Parrado, S. G.; D'Arcy, A.; Eidhoff, U.; Gerhartz, B.; Hassiepen, U.; Pierrat, B.; Riedl, R.; Vinzenz, D.; Worpenberg, S.et al. Structural basis of ubiquitin recognition by the deubiquitinating protease USP2. *Structure* **2006,** *14* (8), 1293.
(392) Mikolajczyk, J.; Drag, M.; Bekes, M.; Cao, J. T.; Ronai, Z.; Salvesen, G. S. Small ubiquitin-related modifier (SUMO)-specific proteases: profiling the specificities and activities of human SENPs. *J Biol Chem* **2007,** *282* (36), 26217.

6. List of abbreviations

Abbreviations of unnatural amino acids

AcK	N_6-acetyl-L-lysine
AzF	*p*-azido-L-phenylalanine
Aha	Azidohomolanine
Anl	Azidonorleucine
AzGGK	N_6-((2-azidoacetyl)glycyl)-L-lysine
BocK	N_6-(tert-butoxycarbonyl)-L-lysine
δThioK	δ-thiolysine
GGK	N_6-glycylglycyl-L-lysine
Hpg	Homopropargylglycine
PlK	Nε-(propargyloxycarbonyl)-L-lysine
Se-Met	Selenomethionine

Abbreviations of proteins

aaRS	aminoacyl-tRNA-synthetase
DUB	Deubiquitylating enzyme
Dvl2	dishevelled 2
GFP	Green fluorescent protein
HECT	Homologous to E6AP C-terminus
ISG15	Interferon-stimulated gene 15
KAT	Lysine acetyl transferade
KDAC	Lysine deacetylases
KMT	Lysine methyl transferas
Mud1	U1 small nuclear ribonucleoprotein A
MurA	UDP-N-acetylglucosamine enolpyruvyl transferase
MAP	Methionine aminopeptidase
MINDY	Motif interacting with ubiquitin (MIU)-containing novel DUB family
MIU	Motif interacting with ubiquitin (MIU)
Nedd8	neural-precursor-cell-expressed developmentally down-regulated 8)
OTU	Ovarian tumor protease
PCNA	Proliferating cell nuclear antigen
PINK1	PTEN-induced kinase 1
PTEN	Phosphatase and tensin homolog
PHD	plant homeodomain
Pol δ	DNA polymerase delta
Pol η	DNA polymerase eta
PRMT	Arginine methyl transferase
RBR	RING-between-RING-RING
RanGAP1	Ran GTPase-activating protein 1
RING	Really interesting new gene

Ub	Ubiquitin
UBA	Ubiquitin-associated domains
Ubl	Ubiquitin-like protein
UCH	Ubiquitin carboxy-terminal hydrolase
UCHL3	Ubiquitin carboxyl-terminal hydrolase isozyme L3
USP	Ubiquitin-specific protease
SIM	SUMO interacting motif
SUMO	Small ubiquitin-related modifier
TRABID	TRAF-binding protein domain
TNF	Tumor necrosis factor
TRAF	TNF receptor associated factors
ZUP1	Zinc Finger Containing Ubiquitin Peptidase 1

Abbreviations of units

Å	angstrom
Da	Dalton
d	day
Hz	Hertz
h	hour(s)
g	gram(s)
g	acceleration of gravity
kDa	kilodalton
kV	kilovolt
L	liter
M	molar
mA	milliampere(s)
min	minute(s)
MHz	megahertz
mL	milliliter(s)
mM	millimolar
nm	nanometer(s)
nM	nanomolar
nmol	nanomole(s)
ppm	parts per million
rpm	revolutions per minute
s	second(s)
µM	micromolar
µmol	micromole(s)
V	volt
(v/v)	volume per volume
(w/v)	weight per volume
(w/w)	weight per weight

General abbreviations

AI	auto-induction
ADPr	ADP-ribose
Alloc	allyloxycarbonyl
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
bp	base pair
Boc	*tert*-butoxycarbonyl
BONCAT	Bioorthogonal non-canonical amino acid tagging
Cbz	benzyloxycarbonyl
Cbz-OSu	*N*-(benzyloxycarbonyloxy)succinimide
CCHFV	Crimean-Congo hemorrhagic fever orthonairovirus
cfu	colony forming units
CV	column volume
CDR	complementary-determining region
CoA	coenzyme A
CuAAC	Copper(I)-catalyzed alkyne-azide cycloaddition
diUb	diubiquitin
DMEM	Dulbeccos Modified Eagle Medium
eq	equivalent
ER	endoplasmic reticulum
ESI	electrospray ionization
FBS	fetal bovine serum
FUNCAT	Fluorescent non-canonical amino acid tagging
GCE	genetic code expansion
Gal	Galactose
GalNAc	N-acetylgalactosamin
Glc	Glucose
GlcNAc	N-acetylglucosamine
GOPAL	Genetically encoded orthogonal protection and activated ligation
HPLC	high performance liquid chromatography
LB	lysogeny broth
LC-MS	liquid chromatography mass spectrometry
MALDI	Matrix Assisted Laser Desorption/Ionization
Man	Mannose
MARylation	mono-ADP-ribosylation
MS	mass spectrometry
MWCO	molecular weight cut-off
NAD^+	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced
NCL	native chemical ligation
NMR	nuclear magnetic resonance
OD	optical density
OTS	Orthogonal translation system

PARylation	poly-ADP ribosylation
PARP	Poly-(ADP ribose)polymerases
PBS	phosphate buffered saline
PCR	polymerase chain reaction
POI	protein of interest
PPI	protein-protein interaction
PTM	post-translation modification
QM/MM	quantum mechanical and molecular mechanical
rt	room temperature
SAM	S-adenosylmethionine
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SLIM	site-directed ligase-independent mutagenesis
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
SPI	Selective pressure incorporation
SPPS	solide-phase peptide synthesis
TBS	Tris buffered saline
TFA	trifluoroacetic acid
TOF	Time-of-flight
tUIM	tandem ubiquitin interacting motif
UAA	unnatural amino acid
UBA	Ubiquitin associated domain
UBD	Ubiquitin-binding domain
UBZ	Ub-binding zinc finger
UIM	Ubiquitin interacting motif
UV	ultraviolet
wb	western blot

7. Appendix

7.1 Supporting Figures – Chapter 2

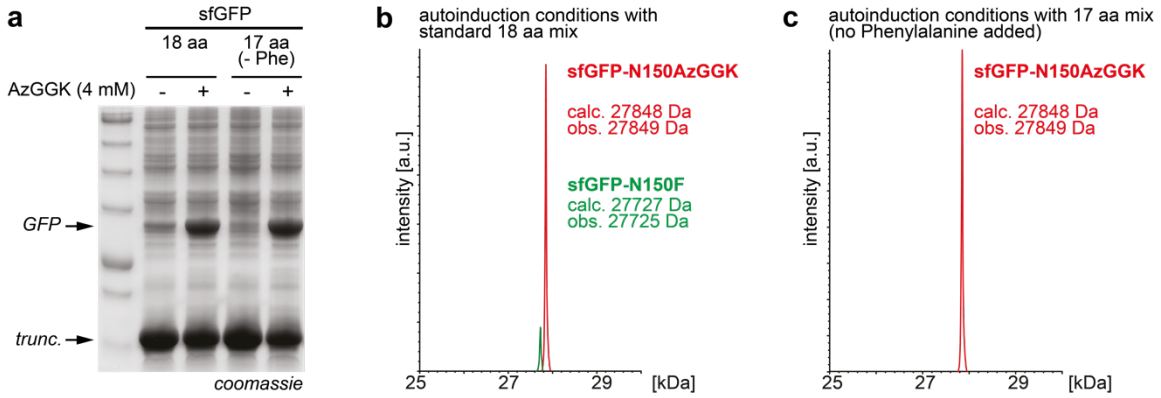


Figure S2.1 Site-specific incorporation of AzGGK into sfGFP in *E. coli.* a) Expression of sGFP-N150TAG-His6 in the presence of AzGGKRS/tRNACUA and AzGGK (4 mM) under auto-induction conditions[330,381,382] lead to AzGGK-dependent synthesis of full-length sfGFP-His6, **b)** as confirmed by LC-MS. Apart from the mass peak corresponding to sfGFP-AzGGK a further small peak (green) was observed in the ESI-MS analysis, which could be assigned to misincorporation of phenylalanine. **c)** Omitting phenylalanine from the amino acid mix lead to clean expression of sfGFP-AzGGK.

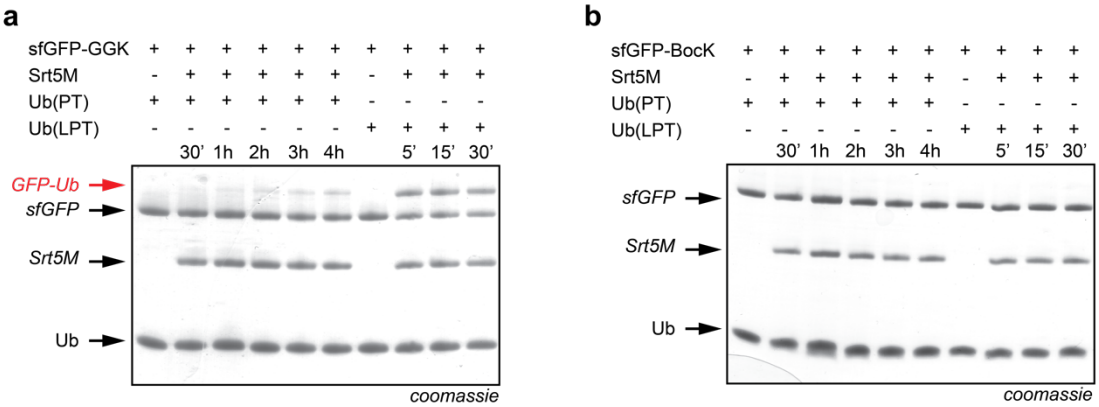


Figure S2.2 Srt5M-mediated ubiquitylation of sfGFP-GGK. a) Time course of Srt5M mediated ubiquitylation of sfGFP-GGK with Ub(PT) or Ub(LPT). sfGFP-GGK (20 µM) was incubated at 37 °C with 100 µM Ub(PT) or Ub(LPT) in the presence of Srt5M (20 µM) in sortase reaction buffer. Reactions were quenched at indicated time points with 4x Laemmli buffer (and heated to 95 °C) and analyzed *via* SDS-PAGE. **b)** Incubation of sfGFP-BocK under the same conditions did not lead to ubiquitylated product. Yields of sortylation were determined densitometrically using ImageJ software: sfGFP-Ub(PT), 3h: 12%; sfGFP-Ub(PT), 4h: 17%; sfGFP-Ub(LPT), 5 min: 49%.

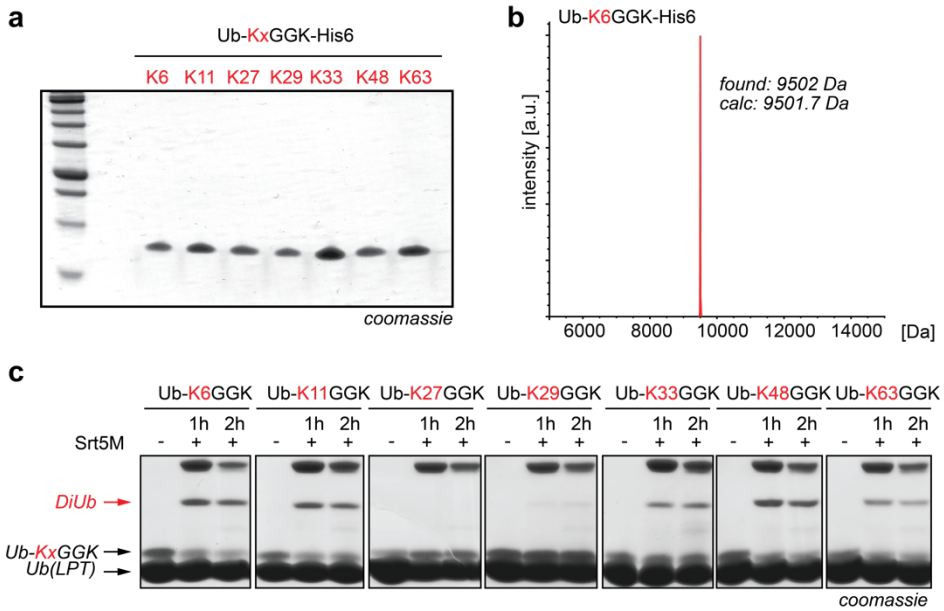


Figure S2.3 Srt5M-catalyzed formation of diubiquitins. a) SDS-PAGE analysis of purified C-terminal His6-tagged ubiquitins bearing GGK at the respective lysine positions (Ub-K6GGK-His6, Ub-K11GGK-His6, Ub-K27GGK-His6, Ub-K29GGK-His6, Ub-K33GGK-His6, Ub-K48GGK-His6, Ub-K63GGK-His6). **b)** LC-MS analysis of Ub-K6GGK-His6. **c)** SDS-PAGE analysis of differently linked diubiquitins formed by Srt5M-catalyzed transpeptidation between Ub-KxGGK and Ub(LPT). Yields of sortylation were determined densitometrically using ImageJ software from Supplementary Figure 5c: K6-DiUb(LPT), 1h, 62%; K11-DiUb(LPT), 1h, 60%; K33-DiUb(LPT), 1h, 43%; K48-DiUb(LPT), 1h, 59%; K63-DiUb(LPT), 1h, 44%.

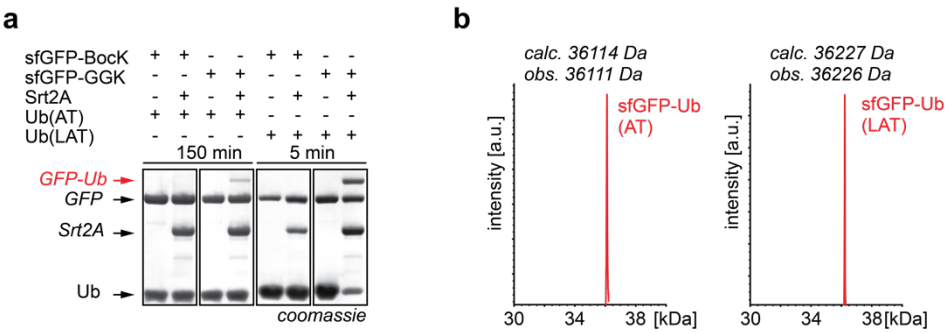


Figure S2.4 Srt2A-mediated ubiquitylation of GGK-bearing proteins. a) Incubation of sfGFP-GGK with Ub(AT) and Ub(LAT) in the presence of Srt2A leads to specific formation of sfGFP-Ub(AT) and sfGFP-Ub(LAT) conjugates, while sfGFP-BocK is unreactive under the same reaction conditions (100 µM Ub, 20 µM sfGFP-N150GGK or sfGFP-N150BocK, 20 µM Srt2A in sortase reaction buffer, 37° C). **b)** LC-MS-analysis of sfGFP-Ub(AT) and sfGFP-Ub(LAT) conjugates.

Ubiquitylation yields were determined densitometrically using ImageJ software: sfGFP-Ub(AT), 150 min, 10%, sfGFP-Ub(LAT), 5 min, 45%

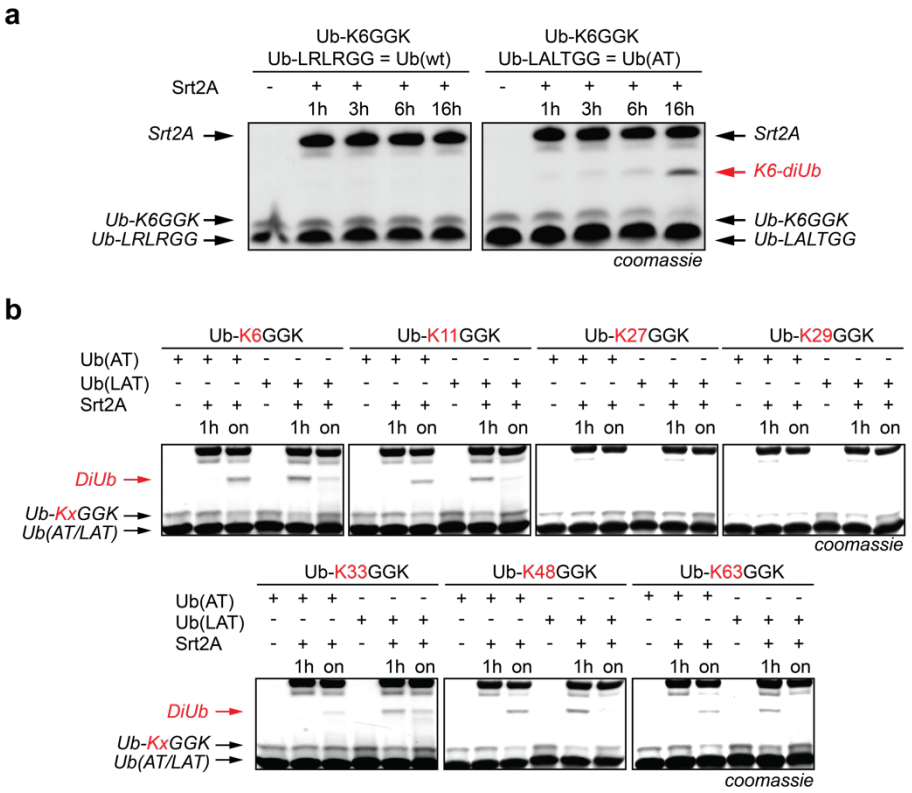


Figure S2.5 Srt2A-mediated ubiquitylation of GGK-bearing proteins. a) While incubation of Ub-K6GGK with Ub(AT) leads to efficient formation of the corresponding K6-linked diUb in the presence of Srt2A, addition of Ub(wt) instead of Ub(AT) did not lead to K6-diUb formation, showing that Srt2A cannot recognize Ub(wt). **b)** Srt2A-catalyed formation of differently linked diubiquitins by incubation of Ub-KxGGK-His6 with Ub(AT) or Ub(LAT) in the presence of Srt2A. K27- and K29 linked diubiquitins were not accessible through Srt2A-mediated transpeptidation, the K33-diubiquitin was only obtained in good yields when using the Ub(LAT) donor ubiquitin, displaying the more exposed sortagging motif.

Sortylation yields were determined densitometrically using ImageJ software:

Supporting Figure 2.5 a: K6-diUb(AT), 16h, 72%

Supporting Figure 2.5 b: K6-diUb(AT), on; 54%, K6-diUb(LAT), 1h, 69%; K11-diUb(AT), on, 41%; K11-diUb(LAT), 1h, 50%; K33-diUb(AT), on, 8%; K33-diUb(LAT), 1h, 38%; K48-diUb(AT), on, 64%; K48-diUb(LAT), 1h, 70%; K63-diUb(AT), on, 35%; K63-diUb(LAT), 1h, 31%.

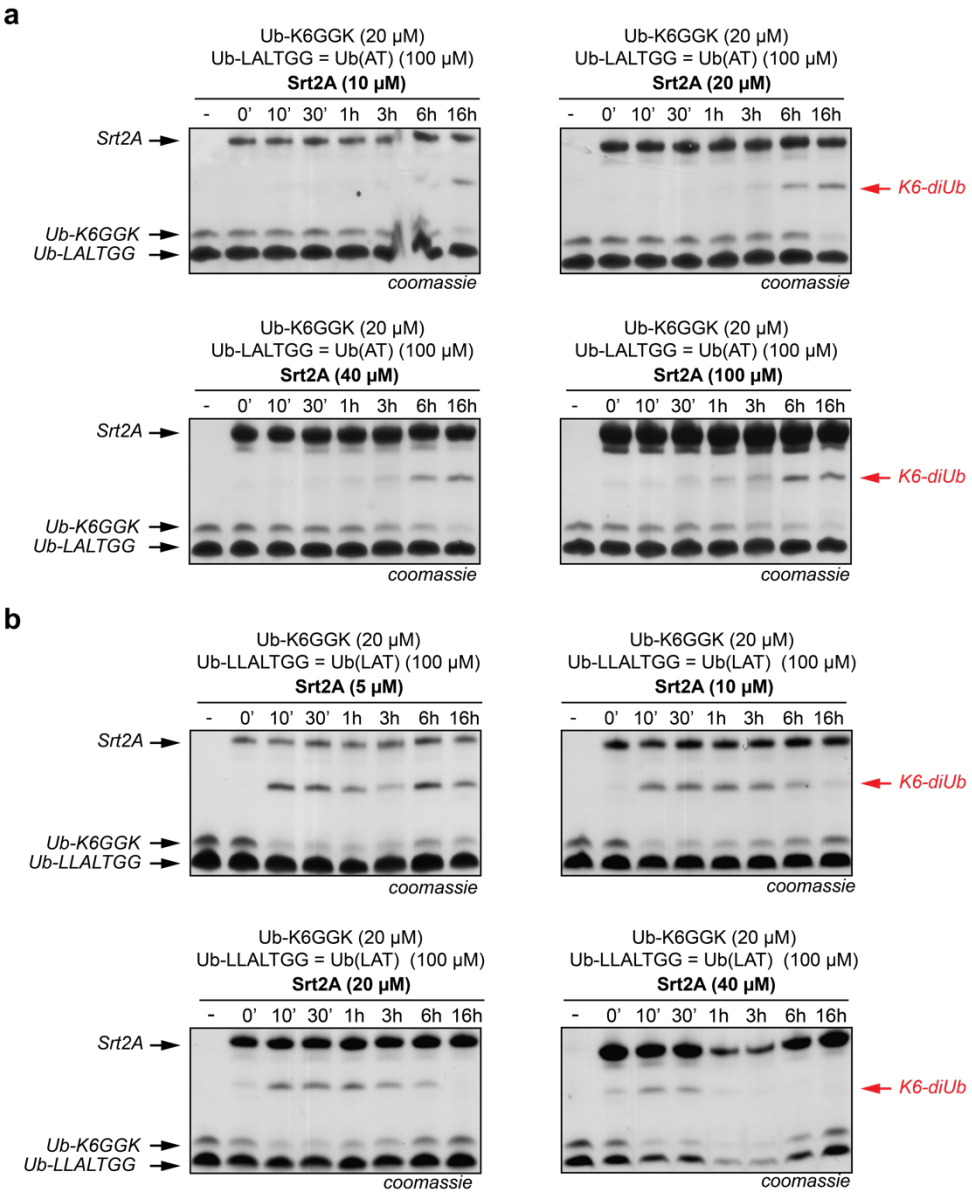


Figure S2.6 Optimizing conditions for Srt2A-mediated K6-diUb-formation. a) Time course of K6-diUb formation between Ub-K6GGK-His6 and Ub(AT) analyzed by SDS-PAGE (in sortase reaction buffer, 37 °C). Upon sortase-mediated transpeptidation between GGK-bearing acceptor ubiquitin and donor ubiquitin the sortagging motif LALTG is re-installed in the generated diubiquitin molecule, making sortylation in principle a reversible approach. Using the Ub(AT) variant with less accessible sortagging motif, however, the formed diubiquitins were stable over a period of 16 hours in the presence of high concentrations of Srt2A. **b)** Time course of K6-diUb formation between Ub-K6GGK-His6 and Ub(LAT) analyzed by SDS-PAGE. The donor ubiquitin with a leucine spacer introduced into its C-terminus (Ub(LAT)) that displays a more accessible sortagging motif, showed diubiquitin formation already within 10 minutes, but the formed ubiquitin dimers hydrolyzed over a time course of 16 hours, especially in the presence of high Srt2A concentrations.

Sortylation yields were determined densitometrically using ImageJ software:

Supporting Figure 2.6 a: K6-diUb(AT), 20µM Srt2A, 16h, 76%

Supporting Figure 2.6 b: K6-diUb(LAT), 10µM Srt2A, 10 min, 71%

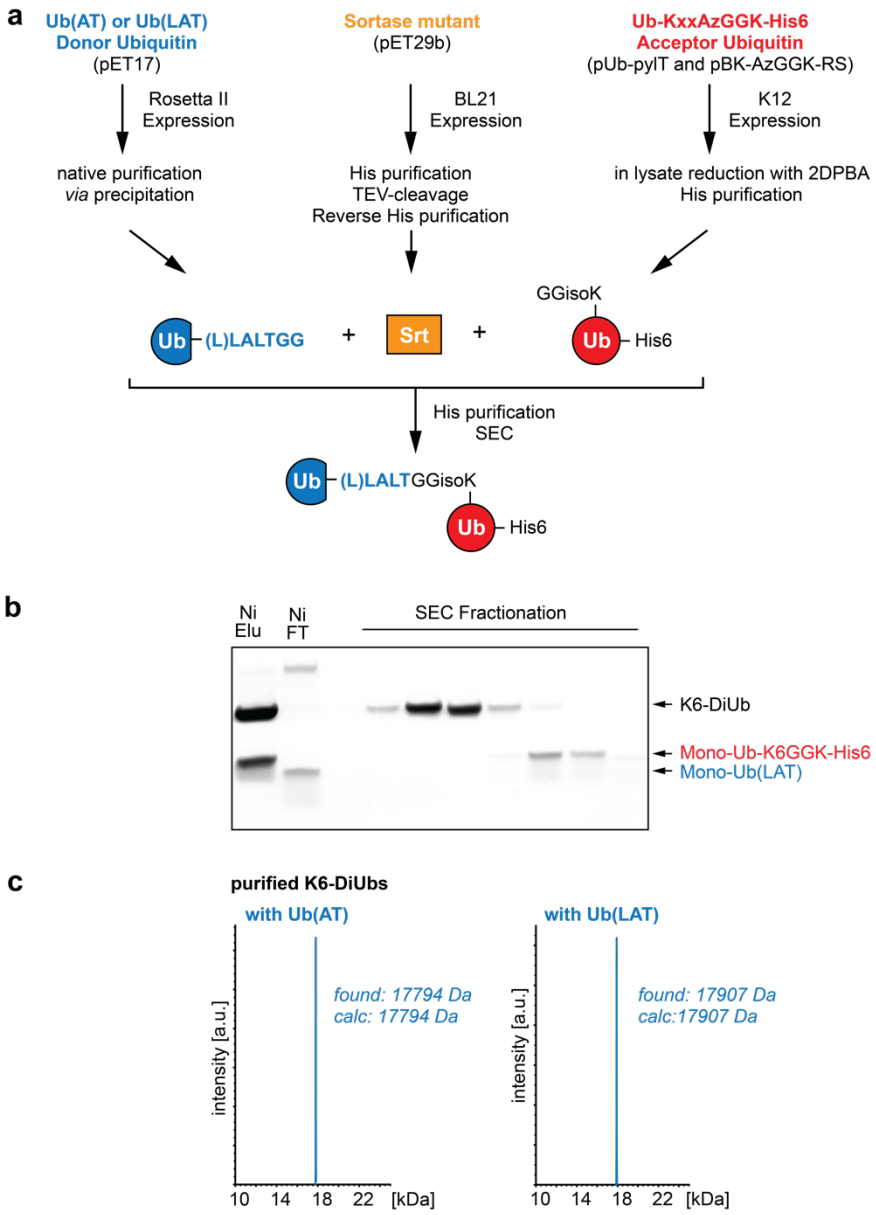


Figure S2.7 Preparative formation of Srt2A-generated diUbs. a) Overview of expression and purification of components needed for assembly of diubiquitins. Sortase-mediated reactions were typically stopped by adding the cysteine protease inhibitor phenylvinylsulfone. B) SDS-PAGE showing a representative diUb purification *via* Ni-NTA chromatography, followed by size-exclusion chromatography (SEC). **b)** All diubiquitin mutants were obtained in milligram scale and were characterized by LC-MS.

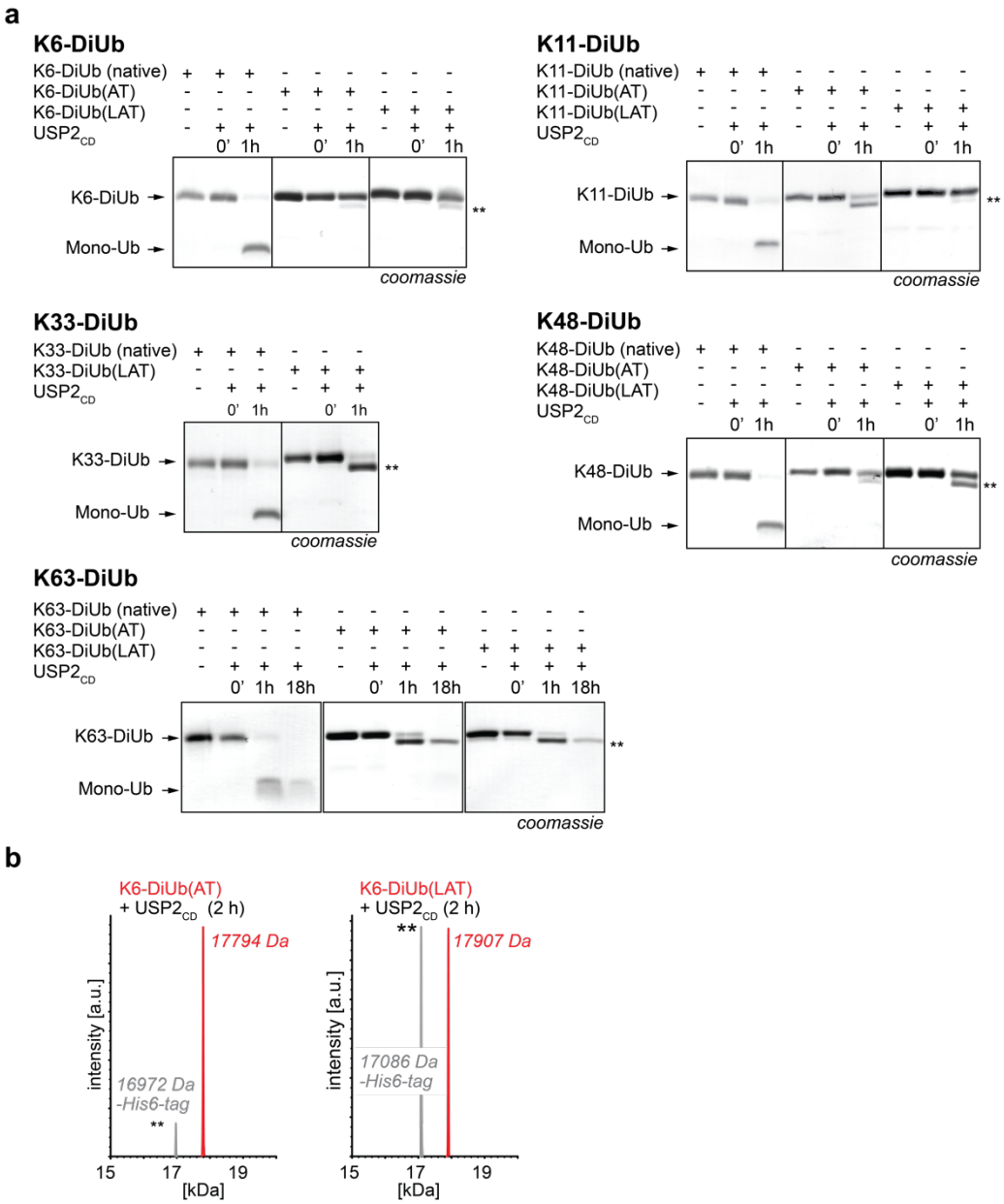


Figure S2.8 Deubiquitylation assays. a) Incubation of Srt2A-generated diUbs and native diUbs with the promiscuous DUB USP2CD showed that Srt2A-generated diubiquitins are stable against isopeptidase activity of the DUB, while all the native diubiquitins were cleaved within an hour at 37 °C. **b)** MS-analysis of DUB-cleavage. **denotes bands and MS-peaks that correspond to DiUbs where the C-terminal His6-tag of the proximal ubiquitin has been cleaved off. Deubiquitylation assays were carried out as described chapter 4.1.

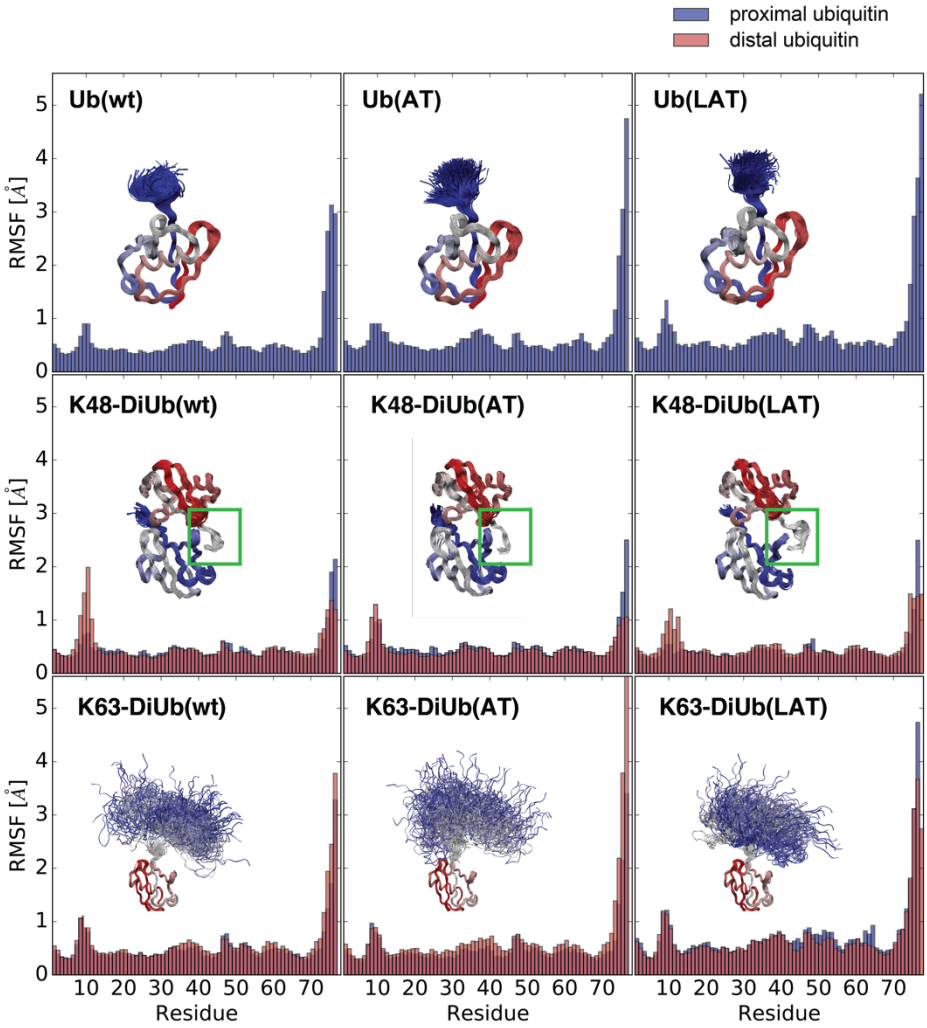


Figure S2.9. Molecular Dynamic Simulations of Ub and diUb variants. Effect of different linkers (wt, AT, LAT) on the dynamics of mono and diubiquitin. The root-mean-square fluctuations (rmsf) per residue are similar for Ub(wt) and monoubiquitins bearing the sortase recognition motif in the C-terminus (Ub(AT) and Ub(LAT)) as well as for differently K48- and K63-linked diUbs. The insets show an overlap of structures from the simulation trajectory aligned to the distal ubiquitin. K48-linked diubiquitin samples a compact conformation, whereas K63-linked diubiquitin shows flexible dynamics during the 0.5 µs MD simulations. Green Boxes indicated the linker region around K48.

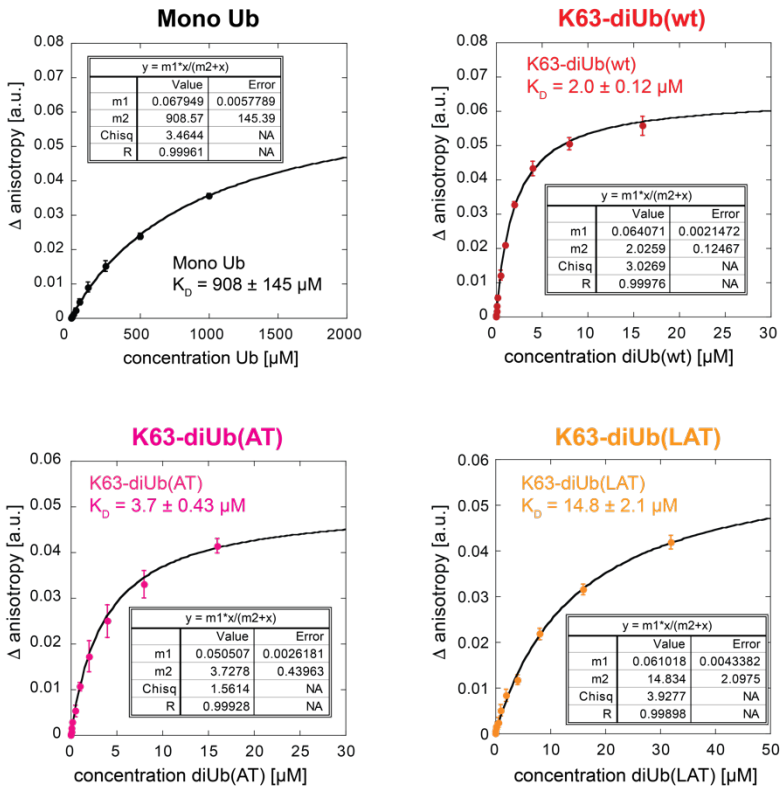


Figure S2.10. Interaction of sortase-generated K63-diUbs with Rap80-tUIM. To validate binding affinity of differently linked K63-diubiquitins towards Rap80-7A-tUIM we performed anisotropy measurements with fluorescently labelled Rap80-7A-tUIM to determine the K_D of the interaction.[340] Experimental details can be found in Chapter 4.1.

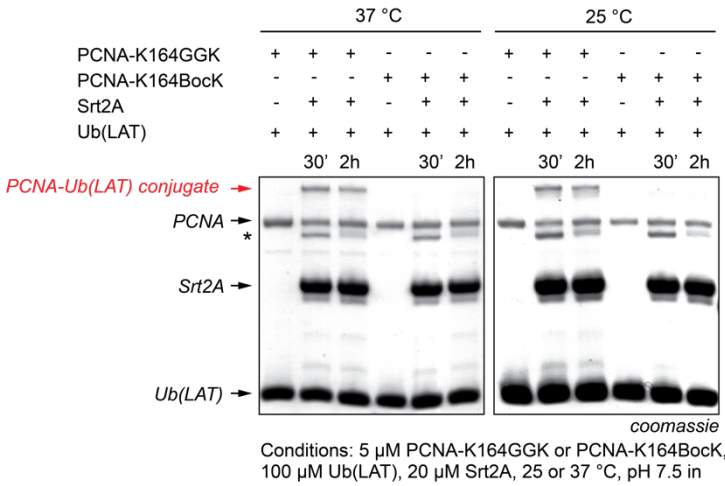


Figure S2.11 Ubiquitylation of PCNA-K164GGK. PCNA-K164GGK was purified from PCNA-K164AzGGK-CPD-His6 and treated with Ub(LAT) in the presence of Srt2A at 37 °C or 25 °C. PCNA-Ub(LAT) conjugate formation was observed within 30 minutes. *denotes thioester formed between Srt2A and Ub(LAT). The formation of PCNA-Ub(LAT) conjugate is specific for PCNA-bearing GGK. No ubiquitylated PCNA is detected when using PCNA-K164BocK.

Sortylation yields were determined densitometrically using ImageJ software:

PCNA-Ub(LAT), 37 °C, 30 min, 44%; PCNA-Ub(LAT), 25 °C, 30 min, 53%;

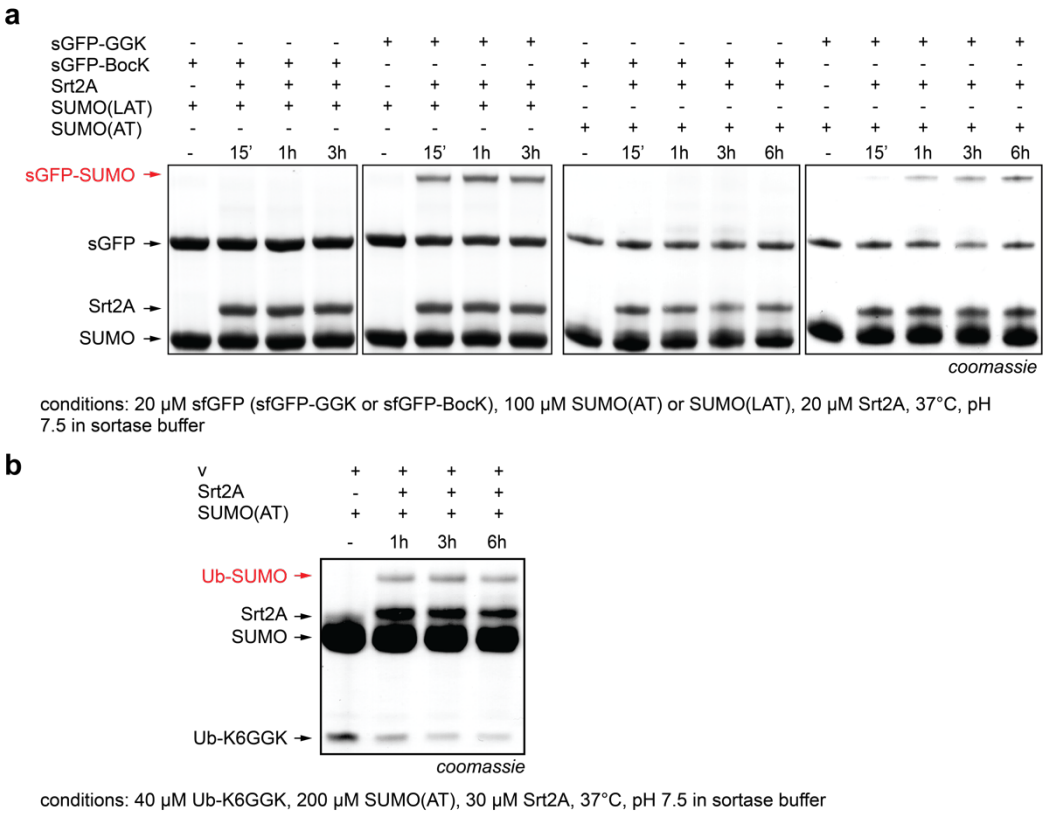


Figure S2.12 Srt2A-mediated SUMOylation of GGK-bearing proteins. a) Time course for formation of SUMO-sfGFP conjugates. sfGFP-GGK (20 µM) was incubated in the presence of Srt2A (20 µM) with an excess of SUMO(AT) or SUMO(LAT) (100 µM) at 37 °C and product formation was analyzed by SDS-PAGE at different time points. As seen for ubiquitylation, SUMOylation proceeded more rapidly when using the SUMO(LAT) variant that bears a leucine spacer preceding the sortagging motif. Incubation of sfGFP-BocK under the same reaction conditions confirmed specificity of Srt2A-mediated SUMOylation. **b)** Time course for Srt2A-catalyzed formation of a Ub-SUMO conjugate analyzed by SDS-PAGE.

Sortylation yields were determined densitometrically using ImageJ software:

Supporting Figure S2.12 a: sfGFP-SUMO(AT), 6 h, 34%; sfGFP-SUMO(LAT), 15 min, 36%

Supporting Figure S2.12 b: UbK6-SUMO(AT), 6 h, 72%

Figure S2.13 Comparison of different sortase A enzymes. Sequence alignment of SrtA-enzymes used in this study. In analogy to the Ca^{2+}-independent Srt7M mutant,[7] we introduced K105 and Q108 mutations into the ß3/ß4-loop of Srt2A, generating mSrt2A.

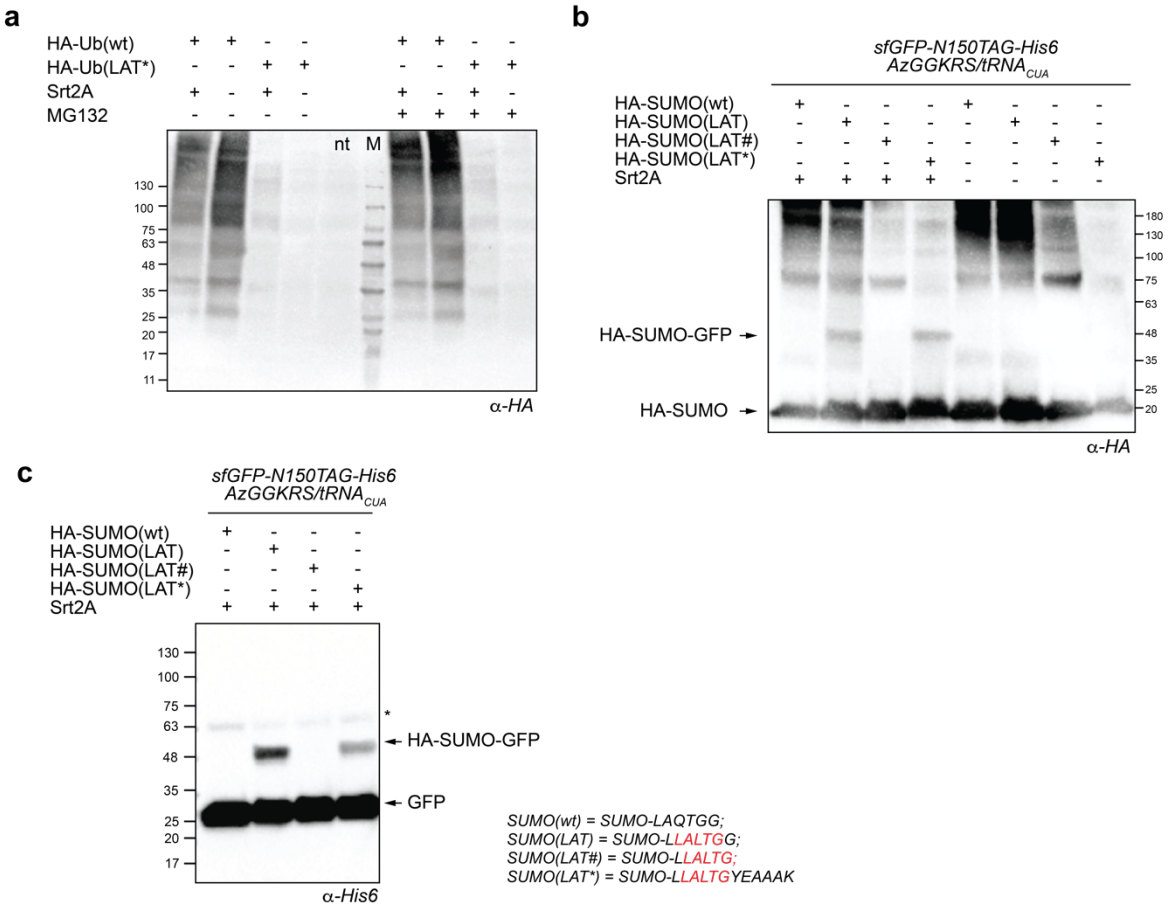


Figure S2.14 Background assays to study orthogonality and specificity of sortase-mediated ubiquitylation and SUMOylation in living HEK293T cells. a) HEK293T cells were transfected with HA-Ub(wt) or HA-Ub(LAT*) in presence and absence of Srt2A and treated with MG132 were indicated. Anti-HA western blotting confirmed orthogonality of the HA-Ub(LAT*) variant and Srt2A. **b)** HEK293T cells were transfected with the indicated HA-SUMO-variant, Srt2A, sfGFP-His6 bearing an amber codon at position 150 and AzGGKRS/tRNA$_{CUA}$ and were grown in the presence of AzGGK. After washing and treatment with 2DPBA, cells were lysed and analyzed by anti-HA western blotting. Anti-HA western blotting confirms that HA-SUMO(LAT*) is unlike the other SUMO-variants a bad substrate for the endogenous SUMOylation machinery. **c)** Anti-His6 western blotting show that both HA-SUMO(LAT) and HA-SUMO(LAT*) are able to SUMOylate sfGFP-N150GGK, while no sfGFP-SUMO conjugate could be detected in the presence of SUMO(LAT#), as observed for Ub variants.

7.2 Supporting Figures – Chapter 3

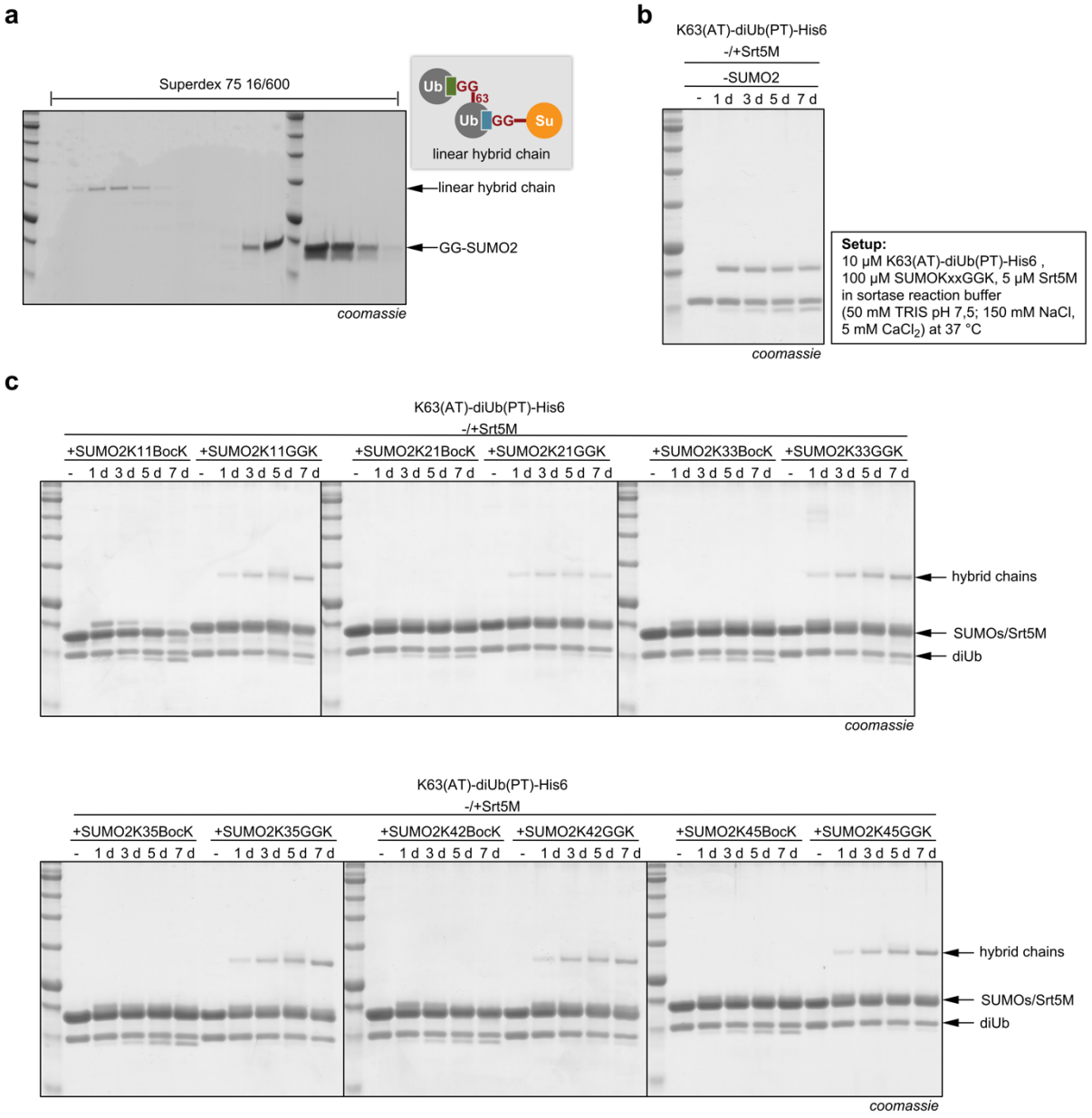


Figure S3.1 Generation of differently linked diUb-SUMO hybrid chains using orthogonal sortases. a) Exemplary SEC purification of a hybrid chain (linear hybrid chain; Ub-(LAT)-isoK63-Ub-(LPT)-SUMO2). **b/c)** Overview of the formation of differently linked hybrid chains. The K63-linked diUb (Ub-(AT)-isoK63-Ub-(PT)-H6) is incubated with SUMO2, bearing GGK/BocK at different positions, in presence of Srt5M. Hybrid chain formation is dependent on the presence of SUMO2 and selective for SUMO2 bearing GGK which proves the isopeptide nature of the linkage.

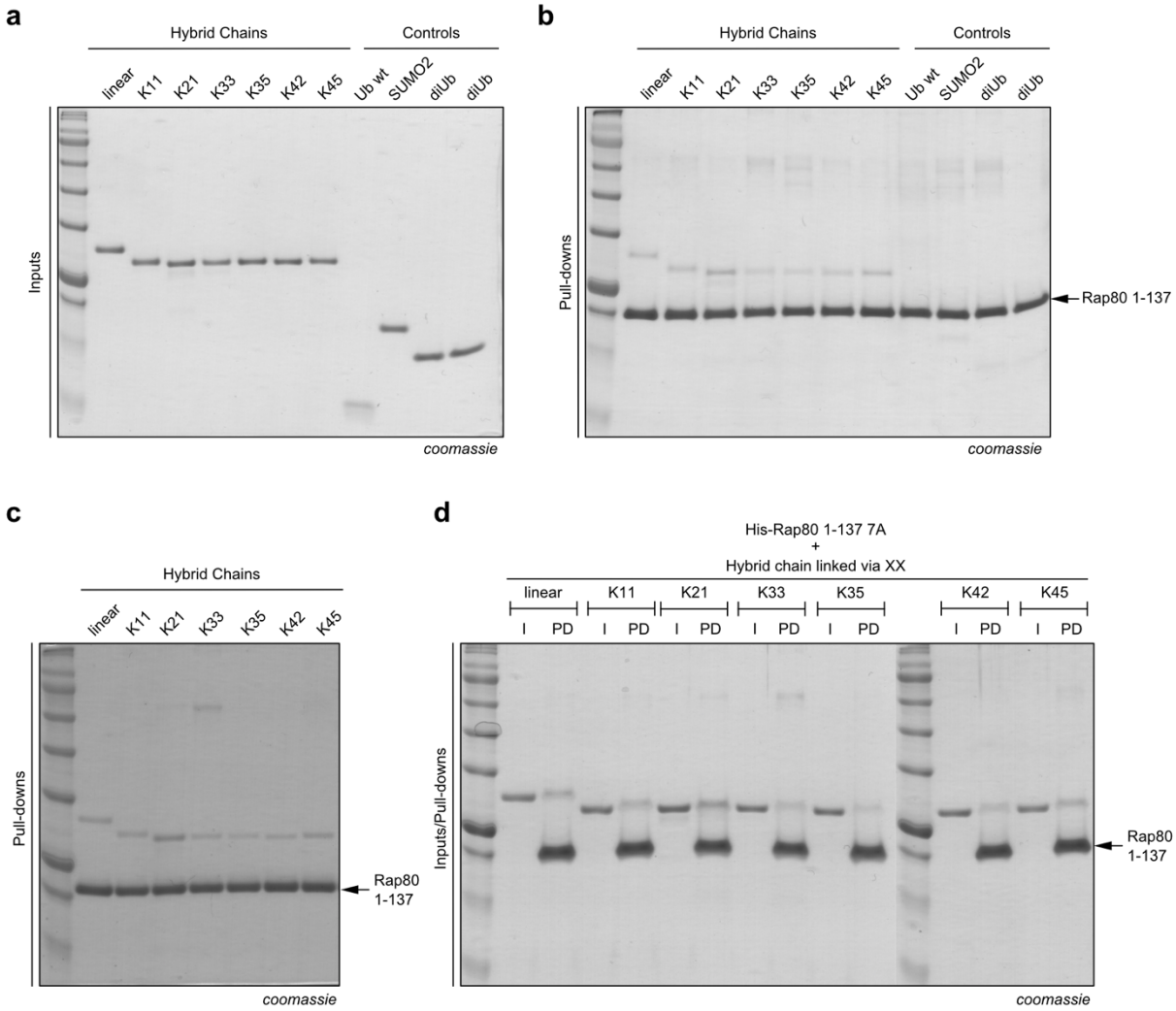


Figure S3.2 Replicates of pull-down experiments. a) 0.5 µg of used proteins as input for pull-downs of b) and c). **b)** Pull-down replicate 1. **c)** Pull-down replicate 2. **d)** Pull-down replicate 3.

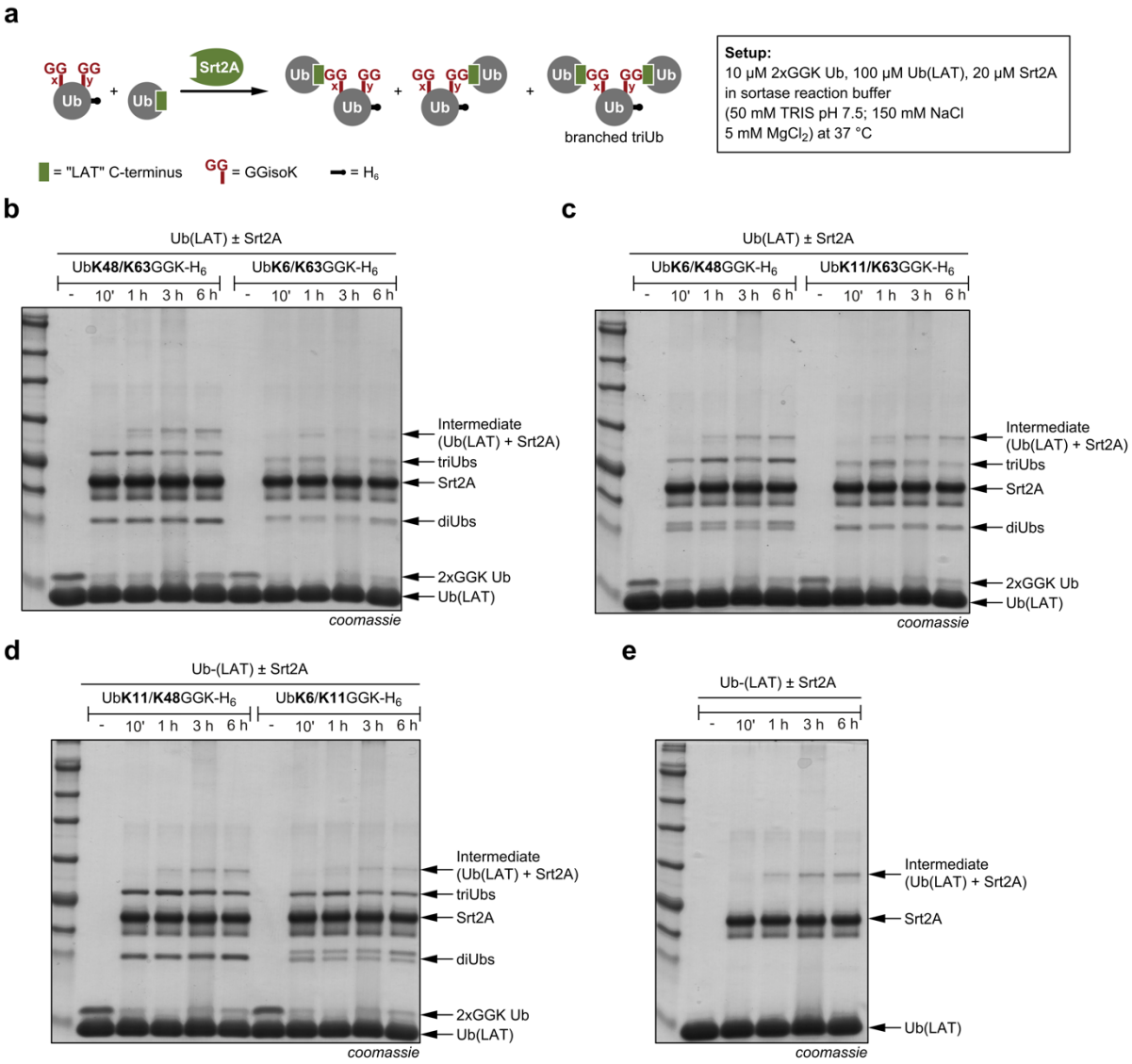


Figure 3.3 Generation of differently linked branched triUbs with spacer. a) Schematic representation of sortase-mediated branched triUb formation. **b)** Incubation of UbK48/K63GGK-H_6 / UbK6/K63GGK-H_6 with Ub(LAT) in the presence of Srt2A leads to formation of the desired branched triUb and both diUb intermediates. **c)** Incubation of UbK6/K48GGK-H_6 / UbK11/K63GGK-H_6 with Ub(LAT) in the presence of Srt2A leads to formation of the desired branched triUb and both diUb intermediates. **d)** Incubation of UbK11/K48GGK-H_6 / UbK6/K11GGK-H_6 with Ub(LAT) in the presence of Srt2A leads to formation of the desired branched triUb and both diUb intermediates. **e)** Incubation of Ub(LAT) with Srt2A does not lead to di/triUb formation since no GGK-acceptor Ub is available. Formation of the Ub(LAT)-Srt2A thioester intermediate can be observed instead.

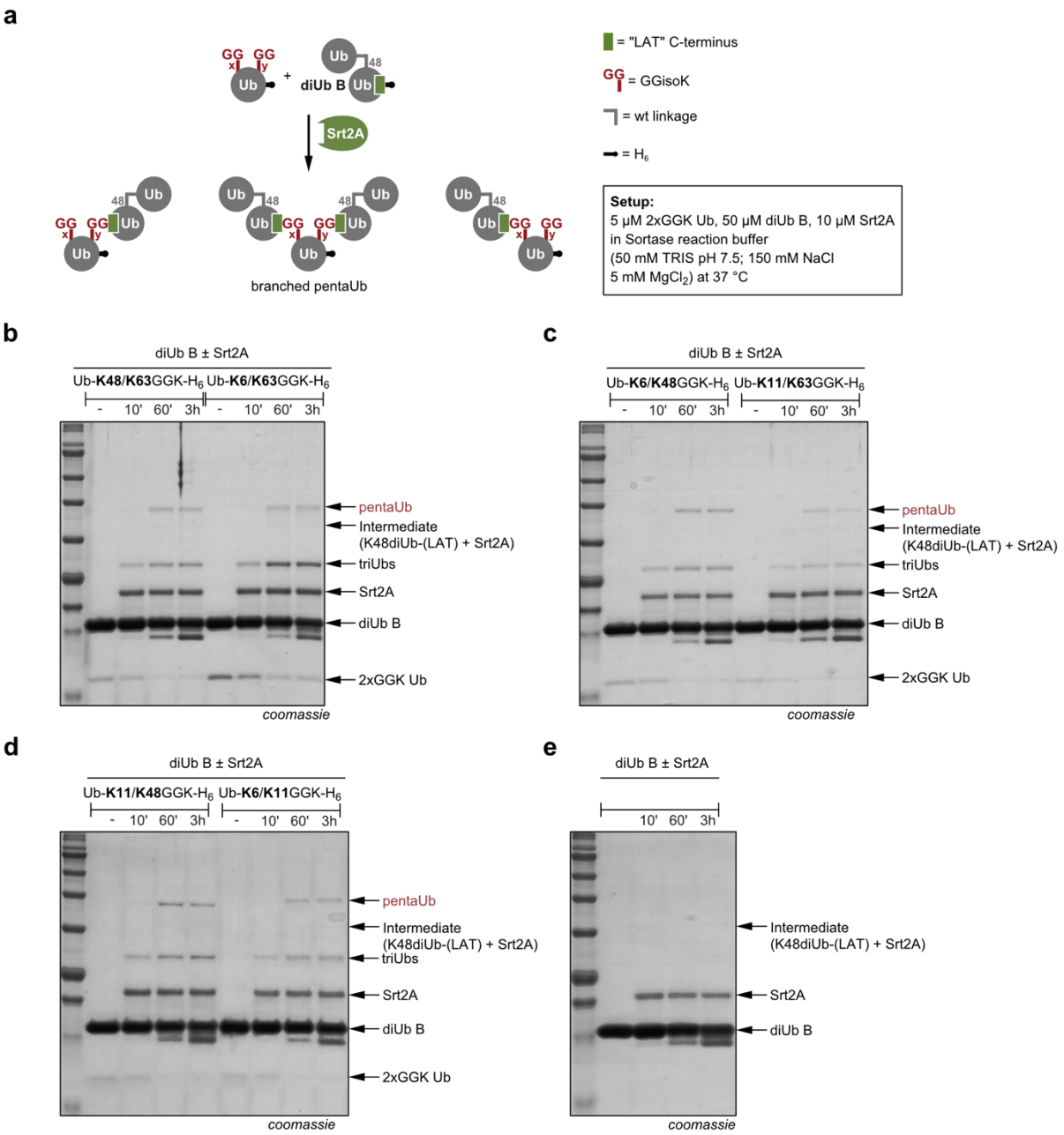


Figure S3.4 Generation of differently linked branched pentaUbs. a) Schematic representation of sortase-mediated branched pentaUb formation. **b)** Incubation of UbK48/K63GGK-H_6 / UbK6/K63GGK-H_6 with diUb B in the presence of Srt2A leads to formation of the desired branched pentaUb and both triUb intermediates. **c)** Incubation of UbK6/K48GGK-H_6 / UbK11/K63GGK-H_6 with diUb B in the presence of Srt2A leads to formation of the desired branched pentaUb and both triUb intermediates. **d)** Incubation of UbK11/K48GGK-H_6 / UbK6/K11GGK-H_6 with diUb B in the presence of Srt2A leads to formation of the desired branched pentaUb and both triUb intermediates. **e)** Incubation of diUb B with Srt2A does not lead to tri/pentaUb formation since no GGK-acceptor Ub is available.

8. Eidesstattliche Erklärung

Hiermit erkläre ich eidesstattlich, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die angegebenen Quellen oder Hilfsmittel verwendet habe. Alle in dieser Arbeit sinngemäß oder wortwörtlich übernommenen Stellen habe ich gekennzeichnet. Diese Arbeit wurde für keinen anderen akademischen Grad eingereicht wie angegeben.

....

Ort, Datum Unterschrift