



Technische Universität München

Fakultät für Chemie

Labor für Synthetische Biochemie

Genetic code expansion tools for studying protein-protein interactions and posttranslational modifications

Dissertation zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften von

Tuan-Anh Nguyen



Technische Universität München

Fakultät für Chemie

Labor für Synthetische Biochemie

Genetic code expansion tools for studying protein-protein interactions and posttranslational modifications

Tuan-Anh Nguyen

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

Die Dissertation wurde am 31.05.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 06.07.2021 angenommen.

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

Danksagung

Oft gibt es Momente im Leben, die man sich ersehnt und wenn sie dann eintreten, weiß man nicht, wie man damit umgehen soll. Die folgenden Worte zu schreiben ist so ein Moment für mich, aber hier ist mein Versuch.

Das Abenteuer Doktorarbeit wäre um ein Vielfaches schwieriger gewesen, vielleicht sogar unmöglich, hätte ich nicht all die tollen Menschen auf diesem Weg kennengelernt, die diese Zeit wesentlich erträglicher gemacht haben.

Zuallerst möchte ich mich natürlich bei der Person bedanken, ohne die es keine Doktorarbeit gegeben hätte. Vielen Dank Kathrin, dass du mir das Vertrauen gegeben hast, bei Dir und mit Dir promovieren zu dürfen. Bei Dir habe ich genau das gefunden, was ich für die Doktorarbeit gesucht habe. Danke für deine Offenheit und Unterstützung, um verschiedene, vielleicht auch verrückte Ideen ausprobieren zu dürfen. Letztendlich macht das doch den Reiz von Wissenschaft aus. Und natürlich ein großer Dank für die angenehme Atmosphäre, erheiternden wissenschaftlichen Diskussionen, deine ‚Open Door Policy‘ und dass du mir auch ab und zu in den Hintern getreten hast, wenn es mal nicht so rund lief.

Natürlich gibt es kein AK Lang ohne das entsprechenden ‚Bodenpersonal‘. Ich weiß nicht, wie ich es am besten beschreiben soll, außer dass es am Ende einfach cool war mit Euch. Die entspannte, verrückte, sich-oft-nicht-zu-ernst nehmende Atmosphäre war genau das richtige, um mal von den Untiefen namens Doktorarbeit wegzukommen. Jeder von Euch hatte ja seine/ihre Stärken, Schwächen, Macken, Besonderheiten, die Euch ja umso sympathischer machen. Die Mischung aus all dem war einfach viel zu gut und nach was Besserem musste man garnicht suchen. All die unzähligen Parties, Feierabendbiere, ‚Seminare‘, random gatherings und chitchats – die ganzen Erlebnisse hier aufzulisten würde schon allein viele Seiten füllen. Aber wenn wir uns in Jahren und Jahrzehnten noch daran erinnern und davon erzählen können, dann haben wir doch vieles richtig gemacht. Vielen Dank euch allen: Kristina (Azo, Alkoxide), Marie-Lena (Lenni, Lennox), Max (Mr. EtOH, Schnefski), Toni (Tonno, Mr. Correcto), Vera (Wanksta, Wänka Snänka) und nicht zu vergessen unsere Ehemaligen: Marko, Marie und Susanne.

Darüberhinaus möchte ich mich bei unseren Kurzzeit AK Lang Labmembers aka meinen Praktikanten und Bacheloranden/innen und deren Unterstützung bedanken: Tobi, Yannick, Maxi und Alex. Ihr wart alle toll und ich hoffe ich konnte euch irgendetwas beibringen.

Der nächste Absatz geht an das ‚charmante‘ Chemiedepartment und die Menschen, die es tagtäglich mit Leben füllen. Ohne die Geräte und Expertise der verschiedenen Arbeitsgruppen und Kollegen wäre ich mit Sicherheit nicht weit gekommen mit der Doktorarbeit. Das gilt natürlich auch für die unzähligen ‚Arbeitsessen und -drinks‘ und erheiterten Diskussionen. Also vielen Dank AK Buchner, AK Itzen, AK Gulder, AK Feige, AK Sieber und Ben natürlich, dass ihr mich habt ‚gewähren‘ lassen. Spezieller Dank geht hier an dieser Stelle an Prof. Dr. Stephan

Sieber und Thomas (Tommy G) für die Proteomicsmessungen. Ich wäre sonst wirklich aufgeschmissen.

Und natürlich megagroßer Dank an Michi (mein SPP Buddy) und Sebastian für die tolle Kollaboration, eure Offenheit und Hilfe, was Zellkultur betraf.

Glücklicherweise konnte ich auch auf Unterstützung von Menschen außerhalb des Labors zählen. Ein großer Dank geht an Jürgen, meinen TUM Mentor, der immer für mich da war und mich über meine Zukunftsoptionen beraten hat. Dein Input war und ist goldwert. Und natürlich darf ich Stefan und Matthias nicht vergessen, bei denen ich oft abhängen und abschalten durfte, sowie Robert, auch wenn paar km zwischen uns liegen. Mal nicht über Wissenschaft zu reden, sondern über die gute alte Zeit sinnieren und sich kritisch mit anderen Themen auseinandersetzen tut auch gut.

Special thanks an Adriana, Krisitina und Kathrin für das Proofreaden und Schleifen dieses ‚Rohdiamanten‘. Das war noch echt viel Arbeit gewesen, von v.1.0 bis zu dieser finalen Version und ohne Euren kritischen Input wäre der nachfolgende Text nicht entstanden.

Zuletzt geht mein Dank an meine Eltern und Schwester. Ihr wart immer da und habt mich immer unterstützt. Ohne Euch wäre ich nicht ansatzweise da, wo ich heute bin. Das war und ist unbezahlbar.

Vermutlich habe ich bis dahin einige Personen vergessen. Deshalb nochmals ein großes Dankeschön an all jene, die sich angesprochen fühlen. Auch an alle flüchtigen Bekanntschaften und Kurzzeitbegegnungen aus unzähligen, zufälligen Ereignissen. All das waren wichtige Puzzleteile, um mein Abenteuer namens Doktorarbeit zu vervollständigen.

List of Publications

Parts of this thesis have been published as papers, reviews, book chapter or have been presented as posters at scientific conferences.

Journal Publications (peer reviewed)

M. D. Bartoschek, E. Ugur, **T. A. Nguyen**, G. Rodschinka, M. Wierer, K. Lang, S. Bultmann; Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells, *Nucleic Acid Research*, 2021 gkab132, <https://doi.org/10.1093/nar/gkab132>

Y.G. Mideksa, M. Fottner, S. Braus, C. A. M. Weiss, **T. A. Nguyen**, S. Meier, K. Lang, M. J. Feige; Site-specific Protein Labeling with Fluorophores as a Tool To Monitor Protein Turnover, *ChemBioChem* 21 (2020), 1861-1867.

T. A. Nguyen, M. Cigler, K. Lang, Expanding the Genetic Code to Study Protein-Protein Interactions, *Angew. Chem. Int. Ed.* 57 (2018), 14350-14361.

Book chapter

M. Cigler, **T. A. Nguyen**, K. Lang; Genetic code expansion approaches to introduce artificial covalent bonds into protein in vivo. In *Oxidative folding of proteins: Basic principles, cellular recognition and engineering* (pp. 399-420), Matthias J. Feige (Ed), The Royal Society of Chemistry, 2018.

Manuscript in preparation

T. A. Nguyen, T. F. Gronauer, S. A. Sieber, K. Lang; Proteomic Profiling of hClpP substrates (working title)

Poster presentations (peer reviewed)

T. A. Nguyen, T. Gronauer, M. Lakemeyer, S. Sieber, K. Lang; Genetic encoding of a bifunctional photocrosslinking amino acid to study protein-protein interactions in mammalian cells; poster presentation; EMBO Conference: Chemical Biology 2018, 29.08-01.09, Heidelberg, Germany.

T. A. Nguyen, T. Gronauer, S. Sieber, K. Lang; Studying protein-protein interactions via genetic code expansion; 8th CPS Meeting 2019, 16.06-19.06; Berlin, Germany.

T. A. Nguyen, T. Gronauer, M. Lakemeyer, S. Sieber, K. Lang; Genetic encoding of a bifunctional photocrosslinking amino acid to study protein-protein interactions in mammalian cells; poster presentation; LMB-GGNB Student Symposium 2019, 10.07-12.07, Cambridge, UK.

Table of Contents

Abstract	X	
Zusammenfassung	XII	
1	Introduction	1
1.1	Natural methods of methods diversification	3
1.2	Strategies towards defined protein modifications	4
1.2.1	Constructing proteins by combining chemical synthesis with protein expressions	5
1.2.2	Selective reactions on natural amino acid functionalities	6
1.2.3	Sequence-guided modifications of proteins	8
1.3	Selective reactions on unnaturally occurring moieties of proteins	10
1.3.1	Expanding the canonical amino acid pool via residue-specific incorporation of unnatural amino acids	11
1.3.2	Genetic code expansion to site-specifically introduce UAAs into proteins	11
1.4	Application areas of amber suppression	13
1.4.1	Encoding protein-PTMs	13
1.4.2	Transformation of chemical handles of UAAs - Bioorthogonal reactions on UAAs	14
1.4.3	Modulation of protein function on demand	17
1.4.4	Capturing and locking protein interactions on demand	18
1.5	Modification and extension of the genetic code expansion technology	21
2	Substrate profiling of human caseinolytic protein P (hClpP) using a photocrosslinker UAA	25
2.1	Proteases and their conventional profiling tools	27
2.2	Genetic code expansion as a suitable tool for protease-substrate profiling	29
2.3	Aim of this work	29
2.4	hClpP biology	30
2.4.1	hClpP and its role in mitochondrial homeostasis	30
2.4.2	Identifying ClpP interactors - current status	31
2.4.3	Structural analysis of the hClpX/P binding	32
2.4.4	Determining suitable positions for photocrosslinker dK incorporation in hClpP	34
2.5	Amber suppression of dK in hClpP and <i>E. coli</i> and functionality tests	35
2.6	Expression of hClpP-dK variants in mammalian cells and photocrosslinking	36
2.7	Proteomic analysis of photocrosslinked hClpP	37
2.8	Target validation of putative ClpP substrates	44
2.9	Rotenone treatment of HEK293T cells expressing hClpP-dK	45
2.10	Conclusion and Outlook	47

3	Development and application of a bifunctional photocrosslinker UAA	49
3.1	Design of a bifunctional UAA	51
3.2	Current list of bifunctional UAAs	51
3.3	Aim of this work	53
3.4	Synthesis and incorporation of the novel bifunctional dKA	53
3.5	Application of the proposed workflow at proof-of-principle PPIs	54
3.6	Incorporation of dKA in mammalian cells	57
3.7	Conclusion and Outlook	59
4	Developing a general approach to access proteins bearing site-specific Lysine acylations	61
4.1	Utilising traceless bioorthogonal reactions to generate modified proteins with native linkages	63
4.2	Bioorthogonal amide bond reactions	64
4.3	Scope of this work	68
4.4	Synthesis and KAT reaction evaluation of the carbamate hydroxylamine derivatives	69
4.5	Synthetase selection for the incorporation of hydroxylamine lysine derivatives	71
4.6	Design, synthesis and KAT reaction of carboxyl hydroxylamine derivatives	73
4.7	Design, synthesis and KAT reaction of alkyl hydroxylamine derivatives	80
4.8	Probing KAT reactions with an aryl hydroxylamine lysine derivative	82
4.9	Conclusion and outlook	83
5	Experimental Part	89
5.1	General remarks – chemistry	91
5.2	General remarks – biochemistry and molecular biology	91
5.2.1	Organisms	91
5.2.2	Cloning and sequencing	92
5.2.3	DNA isolation and purification	93
5.2.4	Antibiotics - stock solutions and working concentrations	93
5.2.5	Preparation of chemically competent DH10 β <i>E. coli</i> cells	93
5.2.6	Preparation of electrocompetent competent DH10 β <i>E. coli</i> cells	94
5.2.7	Heat shock transformation of chemically competent <i>E. coli</i> cells	94
5.2.8	Electroporation of electrocompetent <i>E. coli</i> cells	94
5.2.9	Maintenance and passaging of HEK293T cells	94
5.2.10	Seeding of HEK 293T cells	95
5.2.11	Transfection of HEK293T cells	95
5.2.12	List of Plasmids	96
5.2.13	List of buffers and media	97

5.2.13.1	Buffers	97
5.2.13.2	Growth media	98
5.2.13.3	Stock solutions for protein expression in autoinduction media	99
5.2.13.4	Stocks for aaRS selections	99
5.3	General methods for Chapter 2	100
5.3.1	Chapter 2 - chemical synthesis	100
5.3.2	Chapter 2 - biological experiments	102
5.3.2.1	Cloning of hClpP constructs	102
5.3.2.2	Expression and purification of recombinant (amber suppressed) hClpP	102
5.3.2.3	Expression and purification of recombinant <i>E. coli</i> ClpX	103
5.3.2.4	Protein assembly analysis via size-exclusion chromatography	103
5.3.2.5	hClpP protease activity assay	103
5.3.2.6	Mammalian cell culturing and photocrosslinking	104
5.3.2.7	Western Blot	104
5.3.2.8	Rotenone treatment of HEK293T expressing hClpP	105
5.3.2.9	Experiments for proteomic analyses of hClpP-crosslinked proteins	105
5.4	General methods for Chapter 3	109
5.4.1	Chapter 3 - chemical synthesis	109
5.4.2	Chapter 3 - biological experiments	113
5.4.2.1	Protein expression and purification of sfGFP-N149TAG bearing dKA	113
5.4.2.2	General crosslink, then click approach	113
5.5	General methods for Chapter 4	116
5.5.1	Chapter 4 – chemical synthesis	116
5.5.2	General set-up for KAT reaction studies	128
5.5.3	Chapter 4 – biological experiments	129
5.5.3.1	Synthetase selection protocol	129
6	References	133
7	List of abbreviations	153
8	Appendix	159
8.1	Supporting information for chapter 2	159
8.2	Supporting information for chapter 3	171
8.3	Supporting information for chapter 4	173
9	Eidesstattliche Erklärung	187

Abstract

Proteins are the most prominent key-players in orchestrating multi-layered interactions and reactions within cells to maintain homeostasis and promote proliferation. Hence, investigating their functions is critical towards elucidating their role in health as well as disease states and efforts to develop novel approaches, which enable studies at enhanced resolution, are in dire need.

Genetic code expansion (GCE) enables to extend the side chain diversities and functionalities of proteins beyond what is provided with the canonical amino-acid set. This is achieved by manipulating the host-cell's protein translational machinery to site-specifically incorporate unnatural amino acids (UAAs) into the protein of interest. The structural and chemical variety of UAAs include amino acids bearing posttranslational modifications (PTMs) as well as biophysical and bioorthogonal handles. Along with the advancement of biophysical probe and bioorthogonal reaction developments, this now enables functional protein studies at great resolution. Specifically, usage of photo- and chemical crosslinker UAAs allows to isolate protein interaction partners under physiological conditions and with high spatiotemporal control via selective functional group activation and covalent bond formation between interacting proteins. Also, bioorthogonal chemistries allow the selective labelling of UAA-bearing proteins and thereby to trace, modulate and characterise their function under defined conditions and different physiological states.

This thesis applies two concepts, namely photocrosslinking and bioorthogonal chemistries in combination with GCE, to elucidate protein function and facilitate with a new level of molecular, temporal and spatial resolution. The first two chapters describe how GCE and photocrosslinking may be harnessed for protease substrate profiling and how this can be combined with bioorthogonal chemistries to further enhance the resolution in identifying protein-protein interactions. Within the third chapter, a general approach towards generation of post-translationally proteins via bioorthogonal chemistries is discussed.

In the first chapter of this thesis, the mitochondrial (mt) human caseinolytic peptidase P (hClpP) is investigated at greater detail. Since hClpP plays an important role in mt homeostasis and its dysfunction has been associated with various disease phenotypes such as infertility and cancer, we wanted to develop an approach that helps in characterising its general function in mitochondria via identification of its processing substrates. For this, we incorporated a photocrosslinking diazirine-lysine UAA at various sites in hClpP in mammalian cells, followed by UV-light dependent capture of multiple substrates and identifications via proteomic analyses. This approach also allowed us to monitor the changing substrate dynamics of hClpP in an unbiased manner, when cells were exposed to oxidative stress. This work was performed in collaboration with the Sieber group, TU Munich.

The next chapter describes the development of a bifunctional UAA, which aims to facilitate analyses of protein-protein interactions. This UAA bears an additional alkyne handle besides

a diazirine photocrosslinker, which provides the opportunity for postcrosslink Copper-catalysed azide-alkyne cycloadditions (CuAAC) with a fluorescent probe or an enrichment tag and thus an additional layer of probe analysis. The sequential 'crosslink, then CuAAC-reaction' workflow was tested with fluorogenic probes on various proof-of-principle protein-protein interactions in bacterial and mammalian cells.

Within the last part of this thesis, we envisioned the development of a general approach for the production of site-specifically lysine-acylated proteins. Proteins are subjected to a plethora of PTMs, including various lysine acylations. Despite direct access to some protein acyl-lysine PTMs via incorporation of the respective UAAs, it is unlikely that all currently known acylation statuses are readily accessible by GCE due to limited binding pocket-sizes of the amino-acyl tRNA synthetases and therefore lack of acceptance of larger PTMs. Thus, the goal was to circumvent the size-limitations posed by the translational machinery via synthesis of a universal lysine-based UAA precursor that would yield the respective PTMs through bioorthogonal amide bond-forming chemistries. Successful incorporation such an UAA precursor into the protein of interest followed by reaction with tailor-made PTM probes bearing the respective reactive group would open the avenue to generate target proteins bearing defined amide bond-linked PTMs of any type. KAT ligation, i.e. the reaction between hydroxylamines and potassium acyltrifluoroborates, was chosen as the bioorthogonal reaction and various hydroxylamine-lysine derivatives were synthesised in this regard. This chapter describes the scope and limitations of this UAA class when utilised for GCE purposes and also discusses potential alternatives towards establishing a general protein acylation protocol.

Zusammenfassung

Proteine sind die wichtigsten Schlüsselfiguren bei der Orchestrierung vielschichtiger Interaktionen und Reaktionen innerhalb von Zellen zur Aufrechterhaltung der Homöostase und Förderung der Proliferation. Daher ist die Untersuchung ihrer Funktionen von entscheidender Bedeutung für die Aufklärung ihrer Rolle in Gesundheits- und Krankheitszuständen. Die Entwicklung neuartiger Ansätze, die Proteinstudien mit erhöhter ermöglichen, ist daher dringend erforderlich.

Die genetische Codeerweiterung (GCE) ermöglicht es, die Seitenkettenvielfalt und die Funktionalitäten von Proteinen über das hinaus zu erweitern, was mit dem kanonischen Aminosäuresatz bereitgestellt wird. Dies wird erreicht, indem die Protein-Translationsmaschinerie der Wirtszelle manipuliert wird, um unnatürliche Aminosäuren (UAAs) ortsspezifisch in Zielproteine einzubauen. Die strukturelle und chemische Vielfalt umfasst dabei UAAs, die posttranslationale Modifikationen (PTMs) tragen, sowie biophysikalische und bioorthogonale Gruppen. In Kombination mit der Weiterentwicklung von biophysikalischen Sonden und bioorthogonalen Reaktionen ermöglichen derartige Ansätze nun funktionelle Proteinstudien mit hoher räumlicher, zeitlicher und molekularer Auflösung. Insbesondere die Verwendung von photo- und chemisch crosslinkenden UAAs erlaubt die Isolierung von Proteininteraktionspartnern unter physiologischen Bedingungen mit hoher räumlicher und zeitlicher Kontrolle durch selektive Aktivierung funktioneller Gruppen und kovalente Verknüpfungen zwischen den interagierenden Proteinen. Darüber hinaus ermöglicht die bioorthogonale Chemie die selektive Markierung von Proteinen, die die spezifische UAAs tragen, und damit die Nachverfolgung, Modulation und Charakterisierung ihrer Funktion unter definierten Bedingungen.

Diese Arbeit wendet zwei Konzepte an, das Photocrosslinking und die bioorthogonale Chemie in Kombination mit GCE, um die Proteinfunktionen aufzuklären und Proteinforschung mit hoher räumlicher, zeitlicher und molekularer Auflösung zu erleichtern. In den ersten beiden Kapiteln wird beschrieben, wie GCE und Photocrosslinking für die Erstellung von Protease-Substratprofilen genutzt werden können und wie dies mit bioorthogonaler Chemie kombiniert werden kann, um die Auflösung für die Identifikation von Protein-Protein Interaktionen zu verbessern. Im dritten Kapitel wird ein allgemeiner Ansatz zur Erzeugung von nativ modifizierten Proteinen mittels bioorthogonaler Chemie diskutiert.

Im ersten Kapitel dieser Arbeit wird die mitochondriale (mt) humane caseinolytische Peptidase P (hClpP) näher untersucht. Da hClpP eine wichtige Rolle in der mt-Homöostase spielt und eine Dysfunktion mit verschiedenen Krankheitsphänotypen wie Unfruchtbarkeit und Krebs in Verbindung gebracht wird, hatten wir es uns zum Ziel gesetzt, die allgemeine Funktion von hClpP durch Identifizierung hClpP-Prozessierungssubstrate zu charakterisieren und besser zu verstehen. Dies wurde durch den ortsspezifischen Einbau einer photocrosslinkenden Diazirin-Lysin UAA an verschiedenen Stellen in hClpP in humanen Zelllinien und das darauffolgende UV-Licht-abhängige Einfangen verschiedener Substrate und

deren Identifizierungen durch proteomische Analysen erreicht. Dieser Ansatz erlaubte es auch, die sich ändernde Substratdynamik von hClpP auf unvoreingenommene Weise nachzuverfolgen, wenn die Zellen beispielsweise oxidativem Stress ausgesetzt wurden. Diese Arbeit wurde in Zusammenarbeit mit der Sieber-Gruppe, TU München, durchgeführt.

Das nächste Kapitel beschreibt die Entwicklung einer bifunktionellen UAA, die Proteinanalysen erleichtern soll. Diese UAA trägt neben einem Diazirin-Photocrosslinker eine zusätzliche Alkingruppe, die die Möglichkeit der Kupfer-katalysierten Alkin-Azid Cycloaddition (CuAAC) mit einer Fluoreszenzsonde oder einem Anreicherungs-Tag nach erfolgter photo-Vernetzung bietet und damit eine zusätzliche Ebene der Sondenanalyse zur Analyse von Protein-Protein Interaktionen ermöglicht. Der sequentielle Arbeitsablauf von "Crosslink, dann CuAAC-Reaktion" wurde an verschiedenen Proof-of-Principle-Protein-Interaktionen in Bakterien- und Säugetierzellen mit fluorogenen Sonden getestet.

Im letzten Teil dieser Arbeit wurde die Entwicklung eines allgemeinen Ansatzes zur Herstellung von ortsspezifisch lysin-acylierten Proteinen angestrebt. Proteine unterliegen einer Fülle von PTMs, darunter auch verschiedenen Lysin-Acylierungen. Trotz des direkten Zugangs zu einigen Protein-Acyl-Lysin-PTMs über GCE ist es unwahrscheinlich, dass alle derzeit bekannten Acylierungszustände durch GCE verwirklicht werden können, da die Bindungstaschengrößen der Amino-Acyl-tRNA-Synthetasen begrenzt sind und deshalb größere PTMs nicht akzeptiert werden. Daher wurde versucht, die von der Translationsmaschinerie vorgegebenen Größenbeschränkungen durch die Synthese eines universellen Lysin-basierten UAA-Vorläufers zu umgehen. Diese UAA wurde so konzipiert, dass an ihr posttranslationale, bioorthogonale Amidbindungs-Chemie durchgeführt werden kann. Der erfolgreiche Einbau dieses UAA-Precursors in ein Zielprotein, gefolgt von der Reaktion mit maßgeschneiderten PTM-Sonden, die die entsprechende reaktive Gruppe tragen, würde den Weg für die Generierung beliebiger Proteine mit definierten amidverknüpften PTMs jeglicher Art eröffnen. Als bioorthogonale Reaktion wurde die KAT-Ligation, d.h. die Reaktion zwischen Hydroxylaminen und Kaliumacyltrifluorboraten, gewählt und verschiedene Hydroxylamin-Lysin-Derivate wurden in diesem Zusammenhang synthetisiert. Dieses Kapitel beschreibt den Umfang und die Grenzen dieser UAA-Klasse, wenn sie für GCE-Zwecke eingesetzt wird, und erörtert auch mögliche Alternativen zur Etablierung eines allgemeinen Protein-Acylierungsprotokolls.

Chapter 1

Introduction

1 Introduction

1.1 Natural methods of proteome diversification

Proliferation and maintenance of life is orchestrated by numerous complex biomolecular interaction networks. Its organisation ranges from single cells as the smallest self-sustaining unit to ensembles thereof with evolved specialised functionalities, which combined form a highly developed organism. The central dogma of molecular biology¹ specifies the procedure of how nature's building plan, embedded within each cell, is extracted. Information is stored in coding sections (genes) in DNA, temporarily transcribed into mRNA where it serves as the template to be translated into proteins. Since proteins are the final product of this gene transcription-RNA translation route they are assigned to crucial functions such as transcription and translation, catalysis, signalling and transport of various molecules.

Proteins are chemically speaking amide bond linked biopolymers consisting of the twenty proteinogenic amino acid building blocks in any combination and length. Whilst unreactive in aqueous environments, nature has evolved a sophisticated mechanism to condense amino and carboxyl groups of amino acids together. Central to this reaction is the ribosome, a multi-protein/rRNA complex, which recruits the mRNA and matches the proper aminoacylated-tRNA according to the triplet codon on the mRNA. The ribosome then slides to the next triplet and readily repeats the codon/tRNA-amino acid matching process. The resulting proper vicinal positioning of the preceding carboxyl group-activated tRNA-amino acid complex and the amino group of the proceeding amino acid (-tRNA complex) in the catalytic centre enables the nucleophilic attack to yield the amide bond. The catalytic cycle continues until the stop codon on the mRNA signals the end of chain elongation. Recruitment of release factors leads to disassembly of the mRNA/ribosomal complex and release of synthesised protein.²

Efforts have been made to quantify cellular maintenance and proliferation by counting and correlating the number of coding genes to the amount of proteins in cells. The rough estimation of 20000 protein-coding genes,³ however, poorly represents the actual number of proteins in cells. The observed variety of 'proteofoms' is explained with additional mechanisms occurring at co- and posttranslational level such as alternative splicing, RNA editing, usage of upstream alternative open reading frames, posttranslational protein modifications (PTMs), formation of multicellular protein complexes and localisation (Fig. 1.1).³⁻⁶ The proteome complexity is already illustrated by the number of proteins derived from alternative splice events,⁵ which is estimated to be >100000 and is expected to be even greater considering the other processes and combinations thereof. The diverse pathways are regarded as nature's elegant way to maximise information density within limited storage space and as a rapid respond mechanism to swiftly react to changing environmental conditions, respectively.⁶

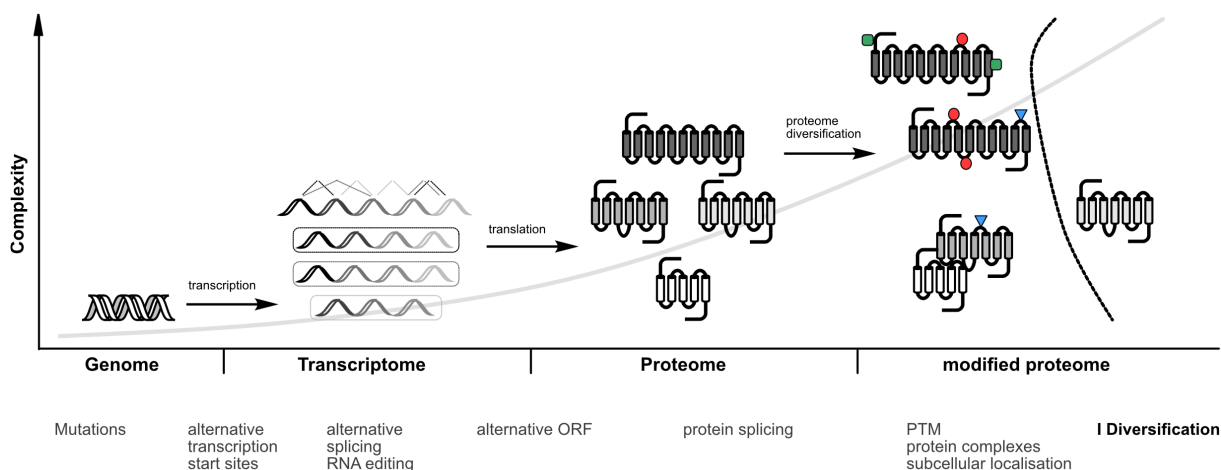


Fig. 1.1 | Routes towards proteome diversification. Any step of the DNA decoding-protein translation route is accompanied by programmed and spontaneous diversification possibilities. (Figure is partially based on ref. ⁶)

Additionally, unforced proteome diversity arises for instance from flaws in the readout process^{3, 7-8} during replication (error rate: 10^{-11} - 10^{-9} per base), transcription (error rate: 10^{-6} - 10^{-4}) and translation (10^{-5} - 10^{-4} per codon). Seemingly very low, mutations may accumulate very quickly considering large genome sizes, fast cell division rates and the absolute cell number of (some) organisms and may ultimately integrate into the DNA, if the resulting altered protein features are advantageous for survival and proliferation.

Given the pivotal role of the proteins as central hubs in numerous, critical cellular processes in both health and disease states, much interest lies in understanding their functions on isolated and global scale as well as developing tools thereof.

1.2 Strategies towards defined protein modifications

While molecular biological and biophysical methods are routinely applied since decades to express, isolate and study proteins in defined experimental conditions (*in vitro*), they are only partly capable of producing the majority of the various proteoforms and thus addressing the protein context at physiological conditions. Genetic alterations such as amino acid replacement as regulators of protein functionality (e.g. mutation of catalytically relevant residues or binding surfaces) and modification mimics (e.g. serine to aspartate as a phosphoserine mimic) or protein fusions (e.g. fluorescent protein tags) for labelling and localisation purposes in cells are cost-effective means to address the afore-mentioned drawbacks, but are to the detriment of appropriate biological and chemical representation and dynamic accuracy. The overall drawback is mainly the reliance on the in-host protein synthesis machinery, which has only access to the predetermined 20 proteinogenic amino-acid pool and is therefore difficult to overcome solely with biomolecular methods. Thus, novel functionalisation strategies are in dire need and organic chemists and chemical biologists are addressing this issue with the creation of novel reaction schemes (Fig. 1.2).^{4, 9-12}

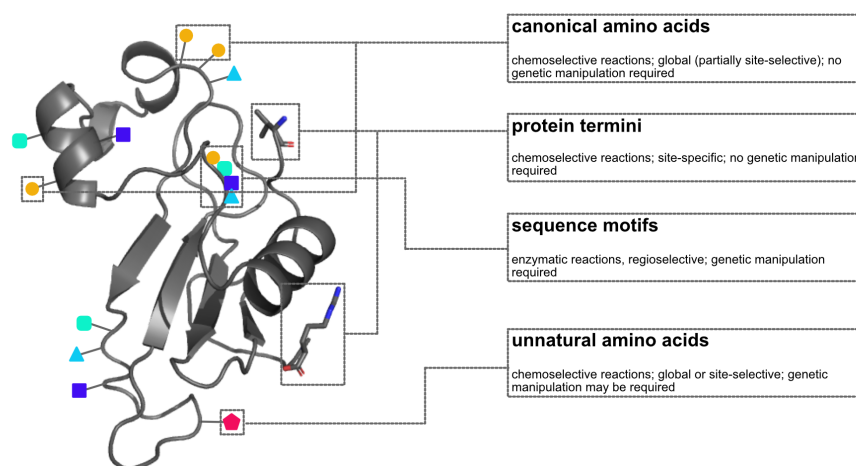


Fig. 1.2 | Putative protein modifications entry sites. Functional group availability on proteins determines the type of chemistries, which may be applied and thus cargo diversity and resolution of the modifications. In some cases, additional genetic manipulation may extend functional group variety and thus the type of reactions, which can be performed. The listed chemistries will be subsequently explained in the following chapters. (Figure is partially based on ref. ¹⁰)

1.2.1 Constructing proteins by combining chemical synthesis with protein expressions

A major breakthrough towards that direction came with the development of native chemical ligation (NCL).^{9,13-14} This methodology extends the inherent sequence length limitations of solid-phase peptide synthesis (SPPS) via fusions of various peptide segments among themselves or with recombinantly expressed proteins (N-terminal extension, Fig. 1.3a). In brief, a C-terminal thioester fragment is incubated with an N-terminal cysteine bearing fragment, which results into a transthioesterification reaction and subsequent S→N acyl shift to yield the final amide bond (with a Cys at the junction site). SPPS may provide both terminal fragments with site-specific amino-acid features, while N-terminal cysteines of recombinantly expressed proteins are mainly accessed through subsequent enzymatic processing. Continuous improvements include means to increase reaction kinetics and aqueous solubility as well as development of thiol-free methods (desulfurization or Cys-alkylation, Ser/Thr ligation, selenide-chemistry, traceless Staudinger ligation, KAHA and KAT ligation).^{4, 9, 15} The optimised ligation schemes nowadays enable the synthesis of proteins bearing multiple functionalities like madanin-like protein 2 bearing sulfotyrosines;¹⁶ SUMO-Ubc9 decorated with a photocrosslinker;¹⁷ histone 3 protein bearing phospho-Ser,¹⁸ polyacetylated-Lys (ref. ¹⁹) or seritonylated-Gln (ref. ²⁰) and phospho-Thr modified Huntingtin.²¹ Even complex topologies such as glycosylated EPO²² and tetra-Ubiquitin (Ub) α-globin,²³ with 53 kDa one of the largest known chemically synthesised proteins to date⁴ (Fig. 1.3b) were successfully generated.

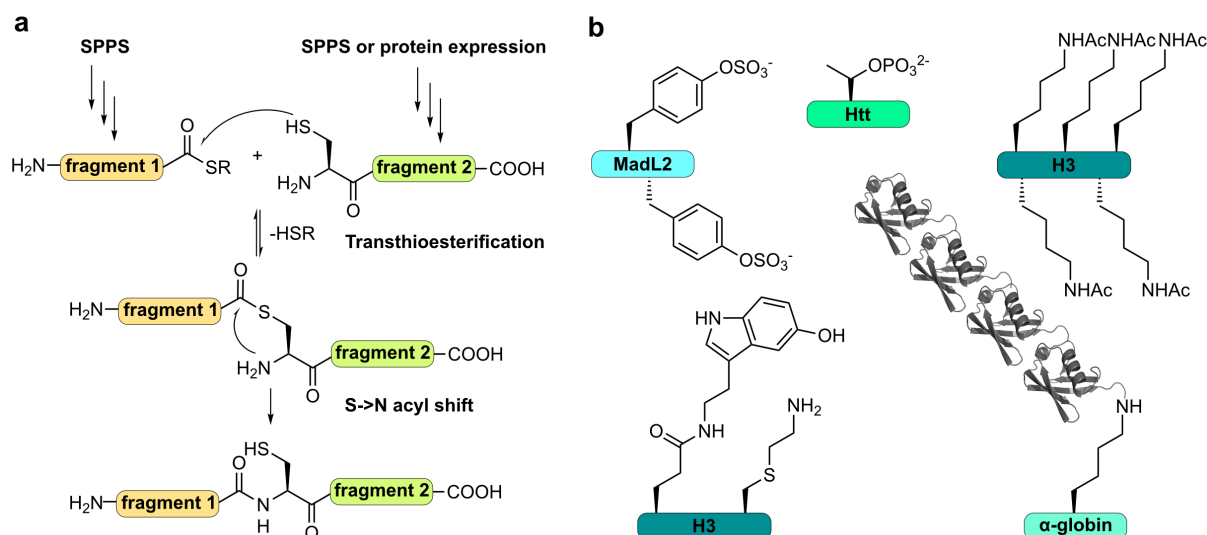


Fig. 1.3 | Concept and application of native chemical ligation. The general ligation scheme is depicted in **a**. As depicted in **b**, the approach led to the synthesis of PTM-proteins of various size and modification numbers. (based on ref. 4)

Other labelling strategies seek a more direct approach and lie in either targeted chemical and/or chemoenzymatic reactions on chemically distinct motifs or extension of the canonical amino-acid pool beyond the twenty side-chain diversity with novel reactive groups for selective chemistries.

1.2.2 Selective reactions on natural amino acid functionalities

Chemical transformations preferably appear at polarised sites (Fig. 1.4a). Thus, reactions aimed at the amino-acid side-chains mainly exclude hydrophobic residues from being chemically modified.^{10, 24-25} Since proteins harbour a plethora of functional groups, development of 'chemoselective' reactions, which exclusively target the 'unique within the unique', represents a great challenge. Further difficulties arise from challenging conditions (from the perspective of an organic chemist) under which such reactions have to perform.¹¹⁻¹² This includes reactions under aqueous conditions, ambient temperature and low concentrations. Additionally, target selectivity needs to be maintained, when performed in nucleophile-rich, *in vivo* environments as well as kinetic, thermodynamic and metabolic stability of the linkage ensured.

Some reactions on popular nucleophiles such as cysteines,²⁵⁻²⁶ lysines^{24-25, 27} or at the N-terminus²⁸⁻²⁹ exploit readily modifiable physicochemical side-chain properties (pK_a of thiol for Michael-additions and alkylations; amino groups for S_N2 -reactions) to react with tailor-made electrophiles and have become routine within the labelling community (Fig. 1.4b and c). Nonetheless, sample heterogeneity caused by multiple modifications events as well as possible disturbances of Cys-Cys bridges, which are important for protein secondary structure and as a redox-relay, should be kept in mind when such chemistries are pursued.

Other side-chains however, require more careful handling as the polarity differences are subtler. Functional group selectivity is often achieved by crafting suitable reactants, which aims to solely target the desired functional group of choice, often combined with

(pre)activations of masked co-reagents via non-conventional methods. This may include redox-, light or organometal-based activation. For instance, histidine labelling has been achieved by either targeting the electrophilic C²-centre³⁰ or nucleophilic N³-position³¹ of the imidazole side-chain (Fig. 1.4d). Alkylation of the former involves dihydropyridines (DHP; charged with various cargos) and blue LED light at elevated temperature and inert atmosphere. While yet not exclusively His-selective as reactions with Cys may occur, several peptides and natural products have been labelled *in vitro*. N³-nucleophile targeting is more straightforward, since the reaction with a suitable electrophile should be sufficient. Incubation of proteins with thiophosphoro alkyne dichloridate (TPAC) exclusively yielded modifications on His. Besides *in vitro* modifications of several model proteins (ribonuclease A, calmodulin, myoglobin), in cellulo proof-of-principle experiments included *in vitro* TPAC-alkyne labelling at the His-tag of purified GFP. Subsequent click reaction with an azide-containing cell penetrating poly-Arg₉ motif enabled delivery and detection of GFP inside mammalian cells. Tryptophan is the least abundant amino acid and thus resembles an attractive target for defined labelling. Similar to selective His-reactions, catalyst-free approaches include reactions at the C²- (ref. ³²) and C³- (ref. ³³) position of the indole ring (Fig. 4e). C² selectivity was achieved by exploiting the amino acids favourable photochemical properties, specifically via incubation with N-substituted pyridinium salts at 302 nm under redox controlled and slightly acidic conditions. Beyond successful labelling of various clinically relevant peptides at milligram scale with great yield (>86%), larger proteins such as lysozyme could be also labelled. C³ on the other hand is readily conjugated with nitric oxides. The organoradical reaction is chemoselective (vs. Ser, Lys, His, Tyr, Met, Cys-Cys) and has been successfully applied for *in vitro* reactions on various peptides, proteins and even antibodies.

Following similar ideas, several electrophiles have been designed which match the sensitive groups of Asp/Glu³⁴ (2H-azirines), Met³⁵⁻³⁶ (oxaziridine, λ^3 -iodanes), Tyr³⁷⁻³⁸ (e-Y-click, electrochemical activation of PTAD electrophile; allylic substrates, Pd catalysed) and even the C-terminus of proteins³⁹ (α,β -unsaturated carbonyls, via photoredox C-terminal decarboxylation reaction) (Fig 1.4f-i).

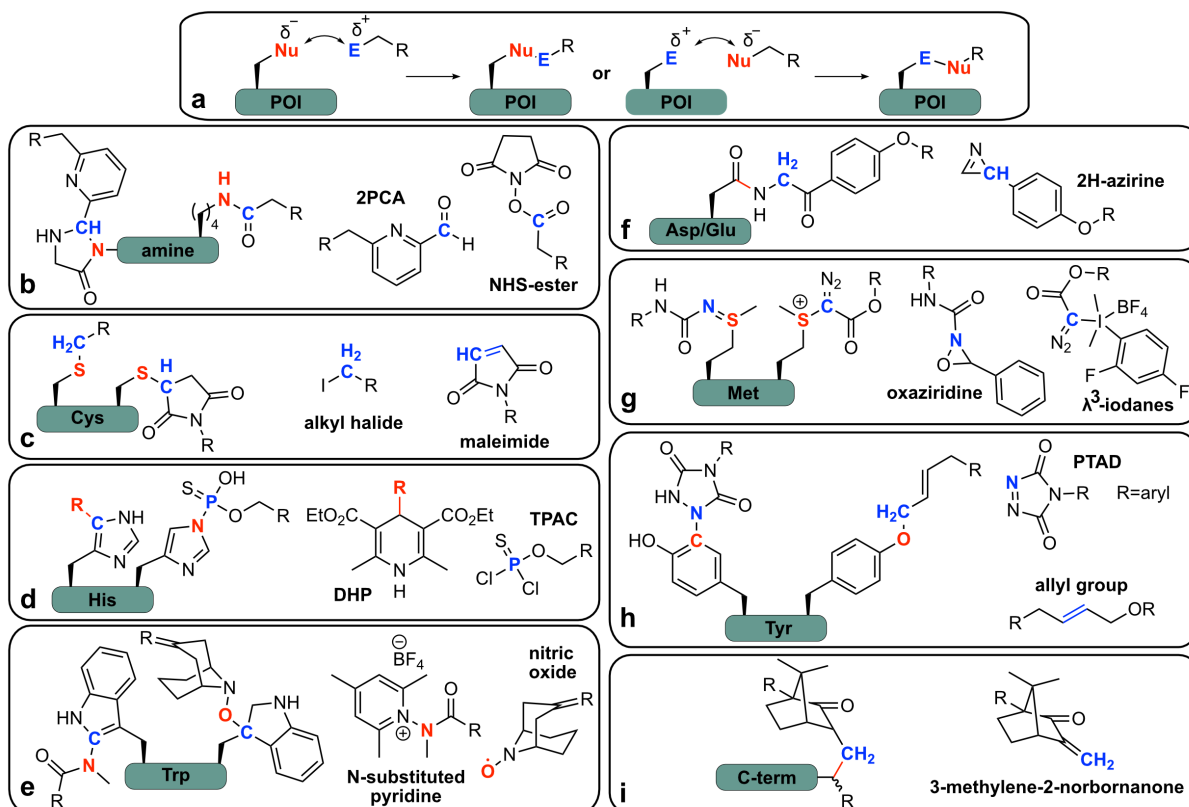


Fig. 1.4 | Modifications of natural amino acids. (a) Chemical reactions preferably occur at polarised sites. With this general principle in mind, a variety of amino acids with their functional groups (b-i) are readily modified with appropriate chemistries and tailor-made reactants, of which a selection is depicted in the boxes. Nu - nucleophile (red), E - electrophile (blue), POI - protein of interest; **a**) 2PCA - 2-pyridincarboxyaldehyde; **d**) - DHP - 4-alkyl-1,4-dihydropyridine, TPAC - thiophosphoro alkyl dichloridate; **h**) PTAD - phenyl-3H-1,2,4-triazole-3,5(4H)-diones.

Despite the primarily proof-of-concept nature with few examples *in vivo*, the combination of biochemistry with related disciplines (electrochemistry, photochemistry, organometallic chemistry) provide promising templates for future chemoselective without the need of much genetic manipulations. General concerns such as site-specificity and broader application *in vivo* still remain to be optimised.

1.2.3 Sequence-guided modifications of proteins

Biology frequently exploits recognition motifs as a distinguishing element to perform protein-selective transformations within a crowded environment of biomolecules. Repurposing those motifs thus adds powerful, biochemical-driven means to the labelling tool box.

Protein-splicing, a measure for posttranslational proteome diversification (*vide supra*), is an auto-catalytic rearrangement process of protein segments (exteins), which is guided by a separating protein element (intein)^{4, 9, 40} Fused to a protein segment of interest, the intein may catalyse the first half of the splicing reaction and trap the segment in a thioester intermediate between the C-terminus of the extein and the N-terminal cysteine of the intein. Subsequent thiolysis yields a protein thioester, which may be C-terminally extended when incubated with a defined, synthetic N-terminal cysteine peptide fragment bearing desired

modifications (often accessed via SPPS) in similar fashion as described for NCL (cf. Fig 1.3a – fragment 1 = thiolysed protein part, fragment 2 = SPPS substrate). Expressed protein ligation (EPL) in conjunction with NCL complements the protein semisynthesis toolbox at both terminal directions and was applied to generate well-defined proteins with an array of modifications at literally any position.^{9, 14, 41} Notable examples include various PTM bearing histones (K_{Ub} and pSer H2A.X (ref. ⁴²), K_{Ub} H2B (ref. ⁴³)), clinically relevant proteins like α -synuclein (pTyr,⁴⁴ poly-GlcNAc⁴⁵), Tau protein (pSer;⁴⁶ K_{Ac}, pTyr and pSer⁴⁷) and proteins involved in signalling pathways such as interleukins (IL6 and 13-glycosylated⁴⁸⁻⁴⁹), Rab GTPase Ypt1 (prenylated⁵⁰) and AKT1 (poly-pSer⁵¹).

Diverse in substrate scope with crucial contributions towards mechanistic understanding protein semisynthesis has proven to be an invaluable tool for protein research, although some bottlenecks remain. This includes on one hand an advanced understanding of synthesis design and execution. On the other hand, performance under denaturing conditions, which requires refolding steps thereafter thus restricts general application under *in vivo* conditions. Therefore, more complementary modification approaches were designed.^{4, 10}

Enzyme specificity is guided by recognition of certain amino acid sequences.⁵² Tags of various lengths (usually 7-15 amino acids) may be installed in protein regions of interest to promote motif-selective labelling.⁵³ Ligases (biotin⁵⁴ and lipoic acid⁵⁵ ligase) and transferases (phosphopantetheinyl⁵⁶ and phosphocholine⁵⁷ transferase) readily attach small molecule cargos bearing various biophysical probes to the designated motif side-chain. Several enzymes are also capable to ligate two larger fragments together. Transpeptidases such as sortase⁵⁸, butelase⁵⁹ and OaAEP1b⁶⁰ in principal catalyse the rearrangement of amide bonds between two fragments and only differ in the recognition sequence (sortase A – LPXTG, butelase – NHV, OaAEP1b – NXX; X denotes usually flexible amino acids, Fig. 1.5a). Those enzymes cut at a defined position of the signal sequence and trap the major protein part in a thioester intermediate. Incubation with a donor nucleophile carrying the complementary motif leads to enzyme-guided attack at the thioester to form a newly ligated product. This way, diverse protein assemblies with artificial payloads (fluorescent probes, biotin, PEG, protein segments) were successfully created. Additional optimisations focus on development of improved expression protocols for those enzymes, applicability at *in vivo* conditions as well as expanding the motif pool for more native sequence-alike ligations.^{9, 52} Through protein engineering, the catalytic cycle of proteases may be reverted,⁹ allowing for so-called peptiligases such as subtiligase⁶¹ and trypsiligase⁶² to perform similar catalysis as transpeptidases, when supplied with the acyl donor fragment and cargo-nucleophile (Fig. 1.5b). However, in either case, reaction yields are diminished due hydrolytic cleavage and reversibility of the reaction, since the signal sequence of the cleaved signal motif remains similar.⁹ The ligation yields may be improved by increasing the starting material input, altering the chemical intermediate structure or removal of the side-product via dialysis or subsequent chemical quenching.⁹ Nonetheless, the main bottlenecks which concerns the accessibility of the recognition motif, remaining of ligation scars and restricted conjugation options at only both protein termini sites remain.^{4, 9}

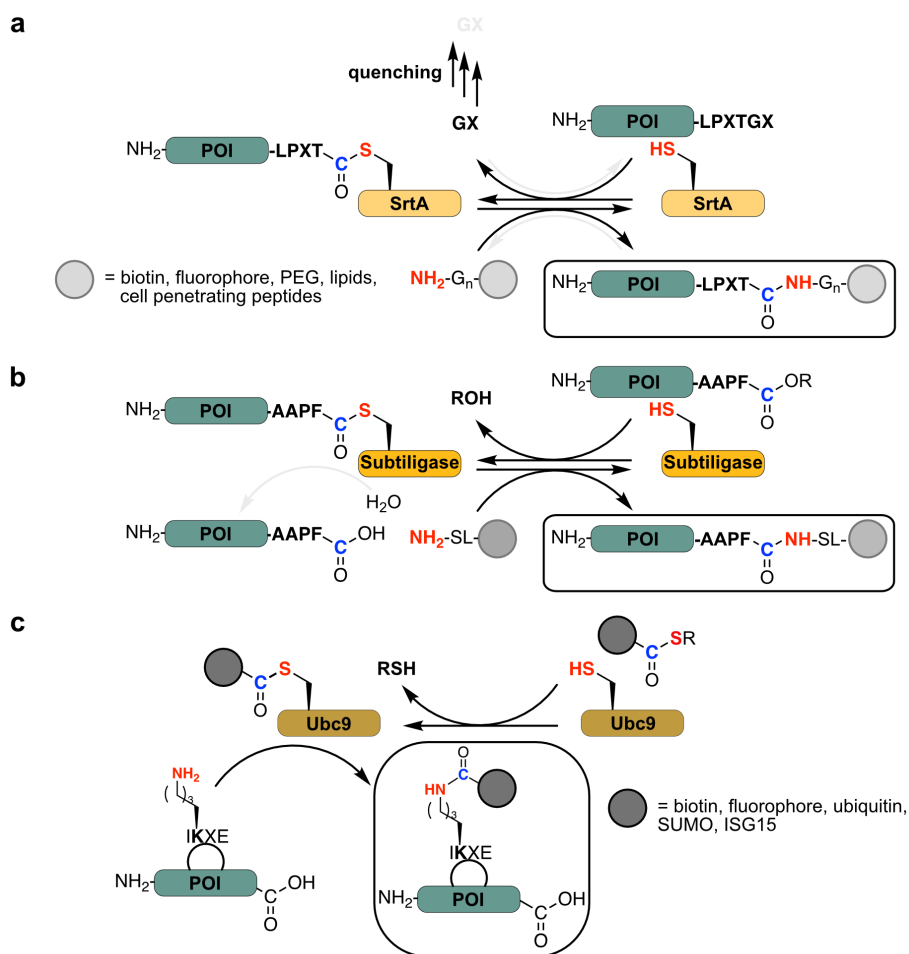


Fig. 1.5 | Chemoenzymatic protein modification approaches. In **a**, the catalytic cycle of transpeptidases, exemplarily shown for Sortase2A (SrtA) is shown leading to modified POI with different cargos (all final ligation products are highlighted in the box). Leaving group GX may be removed from the equilibrium via subsequent modification. Likewise, in **b** the catalytic cycle is shown for Subtiligase as an example for peptiligases. Hydrolysis of the POI-Subtiligase thioester intermediate may occur as a side-reaction (light grey arrow). Alternatively, side-chain extension in **c** is achieved via incubation of an IKXE-motif bearing POI with Ubc9 and a thioester conjugate of choice.

Intriguingly, a recent study expanded the enzyme-guided ligations beyond end-to-end modification (Fig. 1.5c).⁶³ Ubc9, an E2 ligase of the E1-E3 ubiquitin loading machinery, has been repurposed to transfer tailor-made C-terminal cargo-thioesters on the N ϵ -position of the lysine group. Site-specific labelling only required an IKXE-motif to be installed within the POI, rendering it the smallest tag known to-date. Cargo-sizes included whole proteins such as ubiquitin, SUMO and ISG15, which allows for POI-PTM studies with native linkages.

1.3 Selective reactions on unnaturally occurring moieties of proteins

Reliance on natural amino-acid side-chain diversity often restrains the scope of reaction optimisation to the remaining parameters and reactants. As depicted above (Fig. 1.4 and 1.5), selectivity and resolution issues of targeted yet global chemoselective reactions on single amino acids, limited enzyme pool, motif diversity and motif-scars are difficult to resolve, when dependent only on natural amino-acids.⁶⁴⁻⁶⁵ Thus, having unique moieties in proteins would

relieve some chemoselectivity issues because more parameters at the left side of the chemical equation may be altered. Introducing amino-acids with novel chemical entities would however require to extend the tRNA charging mechanism.⁶⁴ This is because the aminoacyl tRNA synthetases (aaRS), which catalyse the acylation of amino acids onto their respective tRNA, carefully scrutinise their catalytic action with sophisticated proof-reading and editing mechanism to ensure correct amino acid charging.⁶⁶

1.3.1 Expanding the canonical amino acid pool via residue-specific incorporation of unnatural amino acids

While discrimination amongst the proteinogenic amino acids is an essential feature for all natural aaRSs, this is not necessarily the case for unnatural amino acids (UAAs), since the synthetases were not subjected to selection against or for it.⁶⁷ This substrate promiscuity is subsequently exploited in selective pressure incorporation, where structural analogues of certain amino acids may be accepted by the respective aaRSs and incorporated according to the sense codon, respectively (Fig. 1.6a).^{11, 67-68} To avoid competition with natural substrates, protein expression is usually performed in defined media supplemented with the UAA and auxotrophic *E. coli* strains, which lack the biosynthetic pathway of the to-be-substituted amino acid. Noteworthy is that this also results in whole-proteome incorporation of the UAA. Useful amino acids to be incorporated include any with bioorthogonal handles (*vide infra*). For instance, azidohomoalanine⁶⁹ and homopropargylglycine⁷⁰ have been incorporated as methionine substitutes. In some instances, rational mutation of the aaRS editing domain⁷¹ or binding pocket⁷² may lead to greater substrate acceptance. This was for instance achieved with azidonorleucine⁷³ as a Lys analogue and *p*-azidophenylalanine and *p*-ethynylphenylalanine as Phe surrogates.⁷⁴ Despite the ability to introduce new chemical entities into proteins for whole proteome analyses (BONCAT⁷⁵ and FUNCAT⁷⁶), again performing protein- and site-selective reactions remain unsolved. Also, the substrate scope is limited due to the restricted UAA acceptance scope of the natural aaRS. Circumvention of this issue would require additional extensions of the above-mentioned bypassing strategy.

1.3.2 Genetic code expansion to site-specifically introduce UAAs into proteins

With rare exceptions of selenocysteine⁷⁷ and pyrrolysine⁷⁸ in some archaea, the same proteinogenic amino-acid set is utilised by all three kingdoms of life. Differences occur in the protein constitutions of the translation machinery rendering them potentially 'invisible' amongst the kingdoms of life (i.e. no interference of the imported elements within the host's protein translation process).^{11, 79-80} Importing tRNA/aaRS pairs into phylogenetically orthogonal organisms thus already provides the catalytic machinery for putative UAA loading. The last missing element, a blank codon at which the UAA is incorporated in response to, is usually provided with the amber stop codon, due to its rarest occurrence in genomes (*E. coli*: 8%, yeast: 23%, human: 24%) and low usage in terminating essential genes (Fig. 1.6b).⁸⁰⁻⁸¹ First pairs for the so-called amber suppression approach included the archaeal

Methanocaldococcus jannaschi tyrosyl-tRNA synthetase (*Mj*TyrRS)/tRNA_{CUA} pair⁸² in *E. coli* and the *E. coli* tyrosyl- (ref. ⁸³) and leucyl-tRNA synthetase/tRNA_{CUA} (ref. ⁸⁴) pair in yeast and mammalian cell.^{11, 79} However, the requirement to abolish the aaRS's natural amino acid affinity prior priming for the acceptance towards the UAA of choice limits their general application. Likewise, partial orthogonality in bacteria and mammalian cells requires usage of different pairs for similar applications. Nowadays, the pyrrolysyl-tRNA synthetase (PylRS)/tRNA_{CUA} pair from *Methanosarcina barkeri* (*Mb*) or *mazei* (*Mm*) is used instead due to being orthogonal in both bacterial and eukaryotic host systems. This obviates the necessity to destroy the affinity to its natural substrate pyrrolysine.^{11, 79} While wt PylRS already accepts a range of UAAs, a sophisticated, iteratively performed two-tiered selection scheme was designed to expand the UAA pool even further (Fig. 1.6c).^{11, 67, 79}

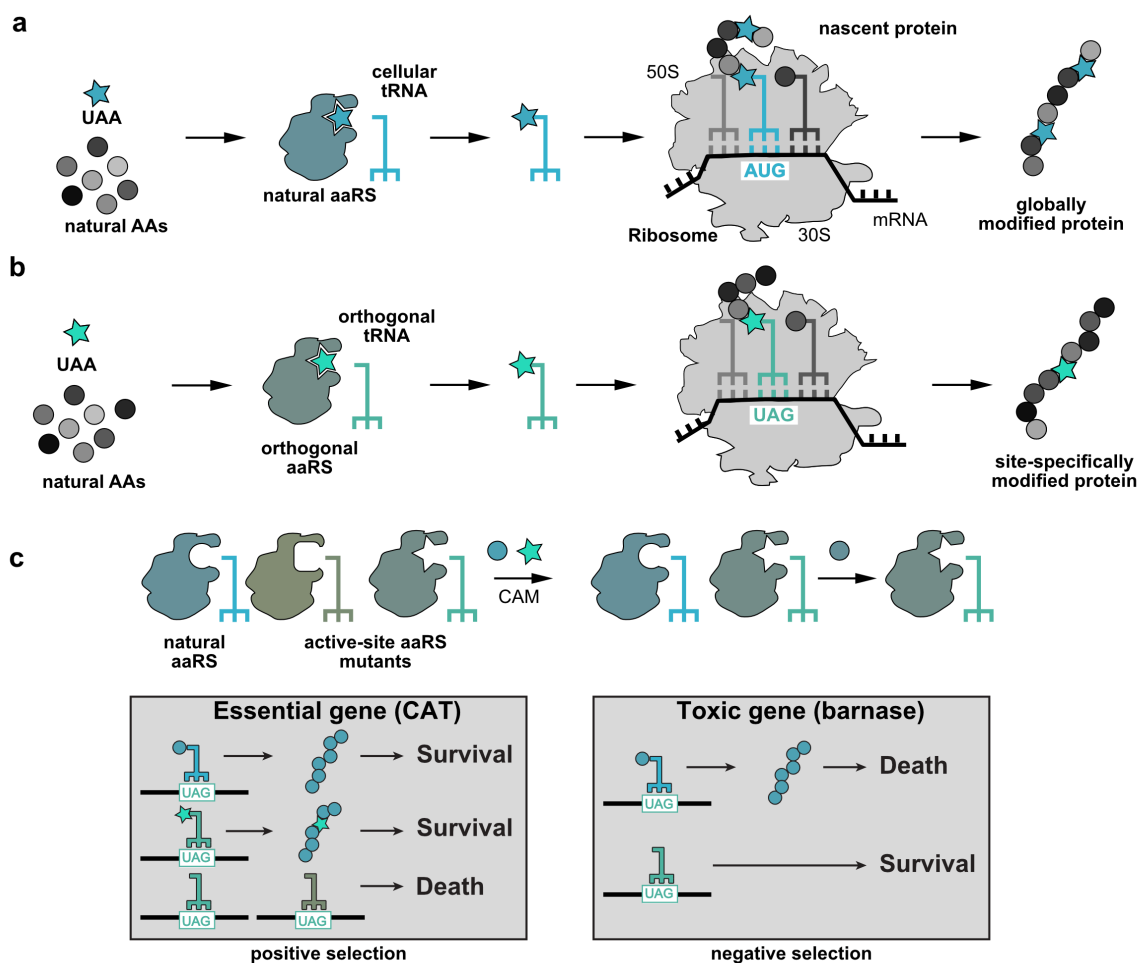


Fig. 1.6 | Decoration of proteins with unnatural amino acids. In a, the general principle of selection pressure incorporation is depicted. Some UAAs may be accepted by natural aaRSs, which in response to sense codons (here: AUG) would lead to globally modified proteins. This may be circumvented by applying genetic code expansion, as illustrated in b. Use of an orthogonal aaRS/tRNA pair and the UAG codon as the blank codon leads to site-specifically modified proteins. The process of selecting orthogonal pairs is shown in c, a two-tiered selection approach, by which the survival of cells is dependent on the proper amber codon readout in response to or lack of UAA and external selection markers. (partially adapted with permission from Maximilian Fottner, TUM).

The principle of this selection method, which is usually performed in *E. coli*, relies on the survival of cells either in presence or absence of the UAA and in response to an external selection marker. At first, *E. coli* cells are co-transformed with an active-site saturated PylRS library (diversity 10^7 - 10^8) and a chloramphenicol acetyltransferase (CAT) plasmid containing an in-frame amber codon (as well as tRNA_{CUA}). Cells, which are grown in presence of the UAA and chloramphenicol only survive this positive selection step, if the amber codon of CAT gene is successfully suppressed. In the following negative selection step, the isolated aaRS plasmid library from the positive selection step is co-transformed with a barnase plasmid containing two in-frame amber codons (as well as tRNA_{CUA}). In absence of the UAA, successful suppression of the amber codons leads to production of the toxic barnase and subsequent cell death. This ensures removal of any colonies and aaRS variants, which amino-acylate any endogenous amino acid. Hence, iterative performance of the selection steps will lead to surviving colonies with the aaRS variant(s), which specifically incorporates the UAA in response to the amber codon.

The robust and directed selection approach has provided synthetases for thus far over 250 different UAAs to be incorporated in various prokaryotic and eukaryotic cells (*vide infra*), multicellular organisms (*D. melanogaster*⁸⁵, *C. elegans*⁸⁶, *Arabidopsis thaliana*⁸⁷) and even complex tissues⁸⁸ (mouse brain⁸⁹, hematopoietic stem cells⁹⁰). Additionally, the site-specificity of amber suppression resolves most of the previously mentioned issues. The variety of UAA covers PTMs as well as biophysical and bioorthogonal handles for protein functionalisation and studies *in vitro* and *in vivo* with great spatiotemporal precision.

1.4 Application areas of amber suppression

1.4.1 Encoding protein-PTMs

Since in many cases, knowledge of the enzymes catalysing protein PTMs are unknown, genetic encoding of such modifications provides a major shortcut for the (large scale) production of homogeneously posttranslationally modified proteins.

Most of the polarised amino-acids are subjected to modifications, with lysine being the most prominent residue. The majority of the protein modifications are 'small' in size, for which an aaRS may be readily selected with appropriate aaRS libraries. This includes several lysine acylations such as formylation⁹¹, acetylation⁹², propionylation⁹³⁻⁹⁴, butyrylation⁹³⁻⁹⁴, crotonylation⁹³⁻⁹⁵ and hydroxyisobutyrylation⁹⁶ as well as small lysine alkyl modifications⁹⁷ (Fig. 1.7a). Negatively charged PTMs such as phosphorylation (Ser⁹⁸, Thr⁹⁹, Tyr¹⁰⁰) and sulfation (Tyr¹⁰¹) are also directly amenable to genetic code expansion (GCE) (Fig. 1.7b), despite previous concerns regarding bacterial cell wall penetration and cytosolic availability of the corresponding UAAs. More exotic modifications such as Tyr nitration¹⁰² and Arg citrullination¹⁰³, are also on the list of encodable PTM-bearing UAAs (Fig. 1.7c).

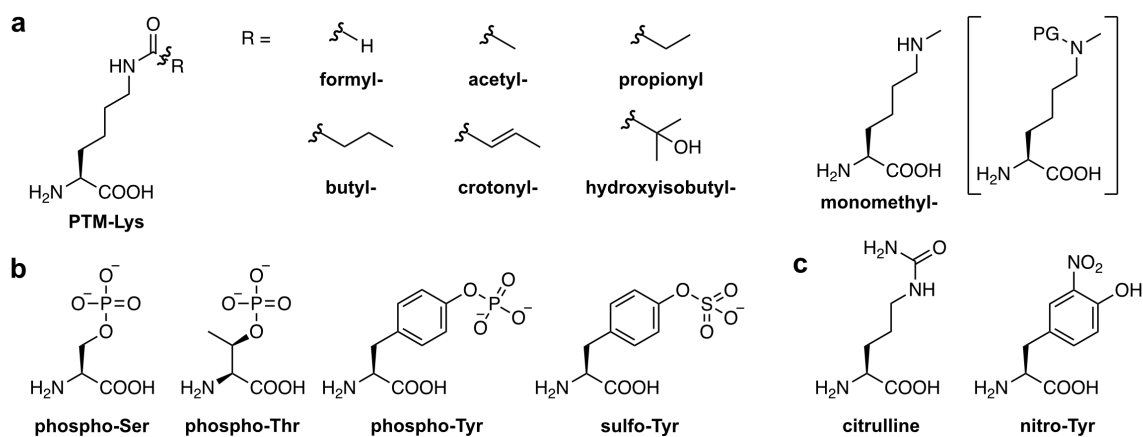


Fig. 1.7 | List of PTM-UAs. The list of PTM-UAs readily accessible via amber suppression includes **a** - various lysine acylations and alkylation (PG - protecting group, decaging is performed posttranslationally), **b** - negatively charged PTMs and **c** - rare PTMs.

The diversity of UAAs bearing posttranslational modifications is likely to be solely restricted by the limited size of the aaRS's recognition pocket. Larger modifications such as lysine lipidation, fatty acid modifications and complex topologies such as whole proteins (ubiquitination and Ubls) and branched glycosylations are up until now not directly accessible via GCE. Complementary approaches were devised to address such sophisticated modifications (Fig. 1.5c and *vide infra*).

1.4.2 Transformation of chemical handles of UAAs - Bioorthogonal reactions on UAAs

Besides the possibility to generate specific isoforms of POIs, UAAs bearing unique reaction handles may be introduced in POIs as well, which allows to perform site-specific chemistries for labelling studies or production of more complex protein conjugates. However, as highlighted above, finding the right reactivity scope of chemoselective reactions remains a major challenge. Additional critical aspects such as toxicity and fast kinetics need to be further considered, especially when performing real-time *in vivo* protein studies.^{11, 104-105}

One of the earliest reaction examples to be regarded 'bioorthogonal' is the oxime/hydrazone ligation^{11, 106}, a condensation reaction between carbonyl groups and strong α -nucleophiles like hydroxylamines and hydrazines (Fig. 1.8a). Performing best under slightly acidic conditions, the reaction however still suffers from sluggish reaction kinetics ($k_2 = 10^{-4}$ - $10^{-3} \text{ M}^{-1} \text{ s}^{-1}$). Small molecule catalysts¹⁰⁷⁻¹⁰⁸ such as aniline may boost the rate up to $10^3 \text{ M}^{-1} \text{ s}^{-1}$ at neutral pH. Nonetheless, the combined reaction conditions and natural carbonyl availability inside cells render this reaction type limitedly suitable, especially for *in vivo* protein studies. A notable example includes the cell-surface protein labelling with a BODIPY-fluorophore.¹⁰⁹ Combined with the incorporation of tailor-made carbonyl or aminoxy UAAs by means of GCE, site-specific generation of oxime-linked protein conjugates bearing for instance glyco-PTM mimics could be accessed.¹¹⁰ Another example includes the *in vitro* generation of di-ubiquitin (K6 and K48 linked) and SUMO-ubiquitin (K11 linked) complexes.¹¹¹ While one ubiquitin part

was amenable through amber suppression of an aminoxy-lysine precursor, the other part bearing the aldehyde counterpart was accessed via intein-technology (Fig. 1.8a). Throughout biophysical experiments, the oxime linkage has proven to be hydrolytically resistant, while the K48-diubiquitin maintained high affinity to deubiquitinase USP2.

Perhaps, the most well-known bioorthogonal reaction is the copper(I)-catalysed click reaction between azides and alkynes (CuAAC).¹¹²⁻¹¹³ Due to the robustness of the triazole linkage, ease of use as well as biological stability of both reactive groups, this 1,3-dipolar cycloaddition has become a standard procedure in research labs for diverse biomolecule labelling reactions. UAAs bearing either azides or alkynes are available for site-specific protein click reaction and have been for instance exploited in proof-of-principle small-molecule labelling of proteins¹¹⁴, building ubiquitin dimers¹¹⁵ in a similar fashion as mentioned above and functional TEV-protease fragment ligation¹¹⁶ (Fig. 1.8b). Toxicity issues related to Cu(I) may be minimised with usage of chelating ligands.¹¹⁷ Additionally, metal-free ligations have been developed as an alternative to circumvent toxicity issues. The major approach lies in elevation of alkyne reactivity through ring strain. Cyclooctynes readily react with azides,¹¹⁸ though at slower rates compared to CuAAC (approximately 10^{-3} vs. $10^2 \text{ M}^{-1} \text{ s}^{-1}$).¹¹ Improved cyclic variants however can push the rate up to $1-10 \text{ M}^{-1} \text{ s}^{-1}$.¹¹⁹ Also, conceptionally similar [3+2] cycloadditions^{105, 120} with different dipoles such as sydnone¹²¹ and nitrones¹²², which react with strained alkynes, have been shown to be potential alternatives, overall showcasing the versatility of this reaction type.

With the possibility to incorporate tetrazine- or strained alkene and alkyne groups via GCE, scientists in addition now have access to the fastest bioorthogonal reaction known to date (Fig. 1.8c).^{11, 123} Typical reaction rates of the [4+2] inverse electron demand Diels-Alder (iEDDA) cycloaddition between tetrazine and strained alkenes/alkynes lie in the range of $10^1-10^4 \text{ M}^{-1} \text{ s}^{-1}$, but may reach up to $\sim 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (ref. ¹²⁴) with carefully fine-tuned reactants. The fast rates enable minimal use of reactants, which becomes important for life-cell labelling, imaging¹²⁵ and kinetic studies.¹²⁶ An advantageous feature is the quenching effect of tetrazines when attached to fluorophores, which is suspended upon alkene/alkyne reaction (turn-on fluorescence, Fig. 1.8c).¹²⁷ On demand signal output was also achieved through installation of a photoprotective group on alkynes.¹²⁸ While the Lys-tetrazine derivative was genetically encoded, successful iEDDA-reaction only occurred when the cyclopropenone-caged bicyclononyne was deprotected with UV light. This not only enabled precise control of on-surface protein labelling and imaging of bacterial OmpC, but also principal execution of sequential labelling.

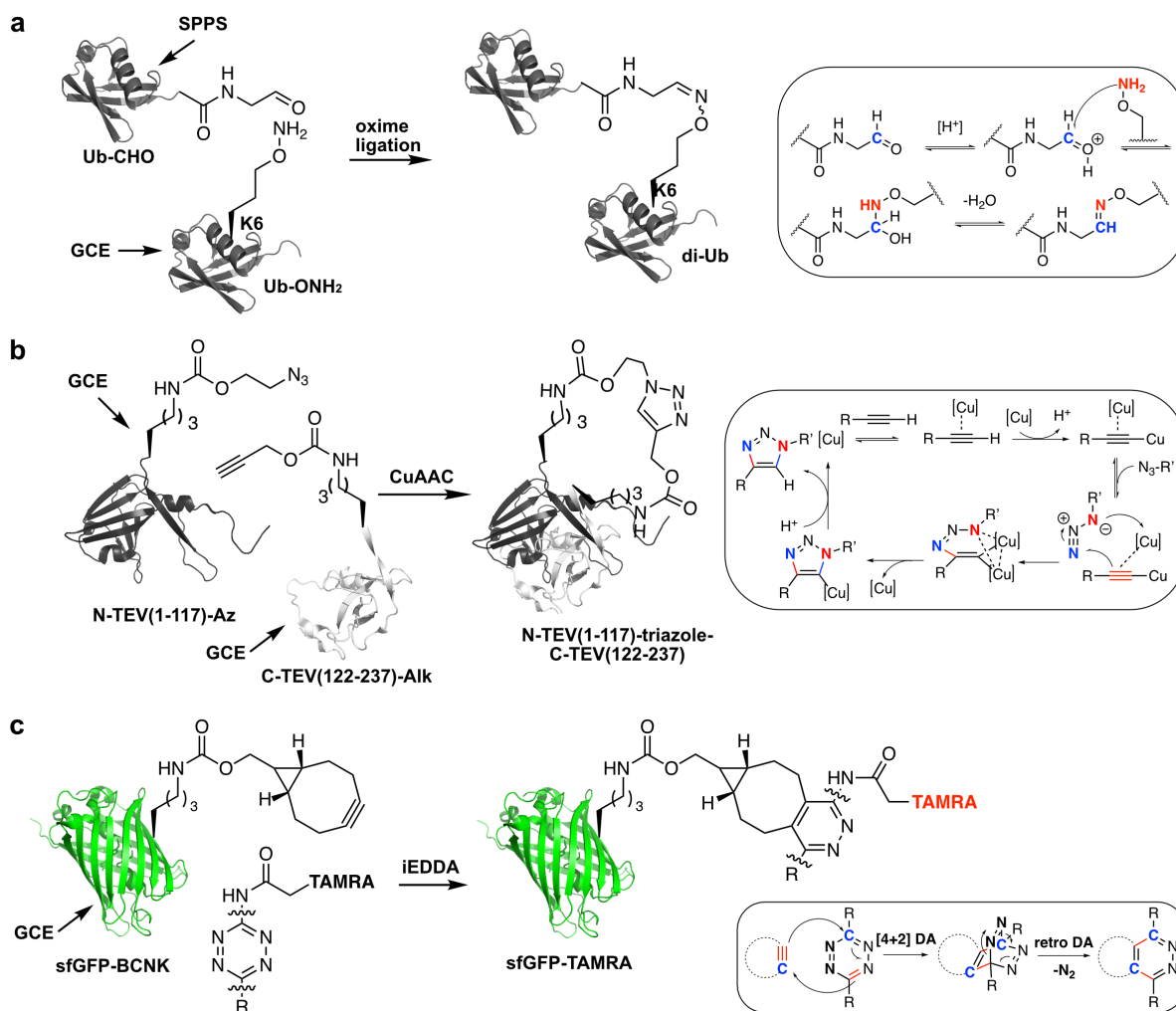


Fig. 1.8 | Bioorthogonal chemistries on amber suppressed proteins. The combination of both, bioorthogonal chemistries on amber suppressed proteins bearing suited UAA-handles, led to synthesis of well-defined protein conjugates. These include: **a** - synthesis of K6 di-Ub via GCE of K₆-aminoxy-Ub (Ub-ONH₂) and oxime ligation with an Ub-C-terminal aldehyde (Ub-CHO), which was synthesised via Intein technology; **b** - protein fragment fusion via click chemistry, exemplarily shown for TEV protease, where both terminal fragments with Lys-azide (N-TEV(1-117)-Az) and -alkyne (C-TEV-(122-237)-Alk) functionality were accessed via GCE; **c** - biophysical probe labelling via iEDDA reaction. BCNK-sfGFP was reacted with tetrazine-TAMRA fluorophore. In some cases, bioorthogonal reactions can be specifically harnessed for turn-on signal output (**b** - TEV activity, **c** - turn-on fluorescence). For each bioorthogonal reaction, the reaction mechanism is highlighted in the box to the right. DA - Diels-Alder.

Slightly altered variants to iEDDA reactions include different types of dienophile (vinylboronic acids¹²⁹) and dienes (triazines¹³⁰) or mode of reaction, such as shown for isonitriles/tetrazines conjugation¹³¹, which is considered a [4+1] cycloaddition.

Overall, the scope of bioorthogonal chemistry is highlighted by the fact that boronic acid/semicarbazide¹³², Pictet-Spengler¹³³ or cross-metathesis reaction¹³⁴ amongst others are additional options to choose from (in some circumstances also to be combined with amber suppression as well) and it is likely that the boundaries of bioorthogonal reactions will continue to be pushed forward. This also includes determining and developing 'orthogonal' bioorthogonal reaction routes, to allow for multiple yet selective protein labellings simultaneously.¹³⁵

To bear in mind is that nearly all bioorthogonal reactions developed to-date create unnatural linkages, which potentially perturb protein performances when compared to their wildtype constitutions. Thus, devising reaction schemes, which yield natively occurring linkages, represents a highly desired alternative. In this regard, a special subtype of bioorthogonal reactions, namely those yielding amide bonds, will be separately covered in chapter 4.

1.4.3 Modulation of protein function on demand

The flexible amber codon placement does not only allow for the direct (cf. chapter 1.4.1) or indirect (cf. chapter 1.4.2) production of various, previously inaccessible, modified proteins, but also for the introduction of UAAs at positions critical for protein function. This would subsequently allow to manipulate protein activity upon triggering functional groups by external cues. This renders GCE an attractive option for spatiotemporally controlled (*in vivo*) protein studies.¹³⁶⁻¹³⁷ Protein activation may be achieved through uncaging of protected UAAs. For instance, on demand Staudinger reduction of 2-azidobenzoyloxycarbonyl lysine UAA bearing proteins with phosphine reagents liberates the functionally critical lysine residue, which subsequently results in the activation of previously hibernated proteins (GFP fluorescence; luciferase, Cre recombinase and CRISPR/Cas9 activity, protein localisation; Fig. 1.9a top left).¹³⁸ Likewise, tetrazine-¹³⁹ and enzyme mediated lysine-decaging¹⁴⁰ were successfully applied for protein activation (Fig. 1.9a bottom left and right).

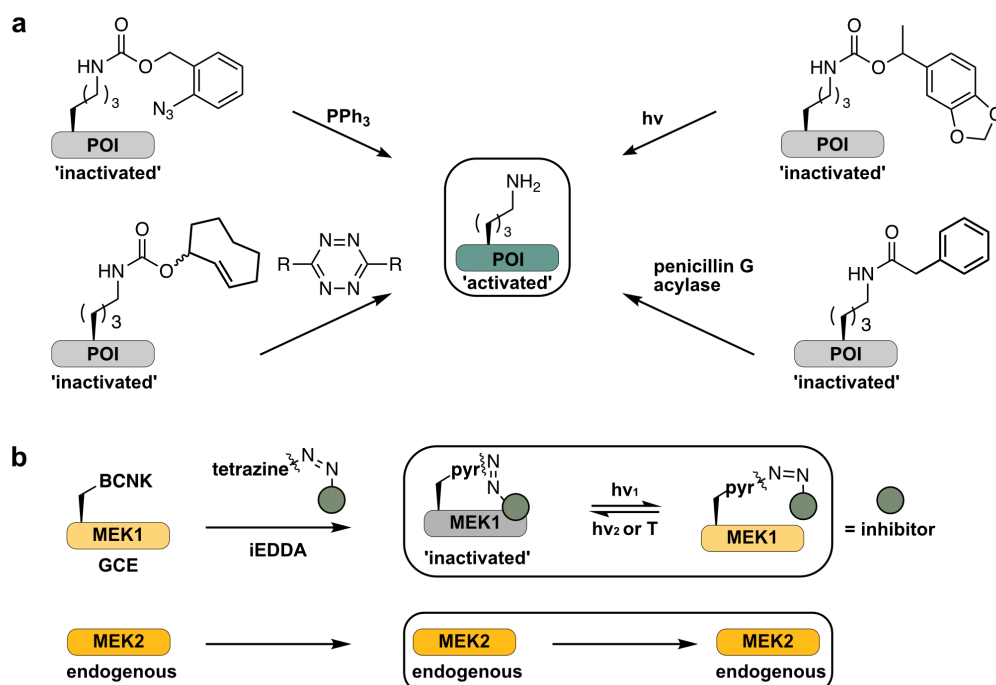


Fig. 1.9 | Modulation of protein function on demand. As depicted in **a**, proteins can be expressed in a ‘hibernated mode’ and become activated by external triggers (chemically or physically) when required. Also, activity of otherwise indistinguishable isozymes can be selectively and reversibly switched, as shown in **b** with the iBOLT approach. Pyr = pyridazine (BCNK-azobenzene-tetrazine DA adduct).

As briefly outlined above, physical triggers such as light provide an easy-to-handle, non-invasive alternative with great spatiotemporal precision. Cysteine-¹⁴¹, lysine-¹⁴² and tyrosine-¹⁴³⁻¹⁴⁴ UAAs bearing so-called photoprotection groups were incorporated into critical positions of proteins, which required proper wavelength for decaging and unlocking the full (catalytic) functionality (Fig. 1.9a top right). Depending on the functional group properties, light is also a powerful method to temporarily modulate protein conformational states and thus activity. This is most prominently exploited with azobenzene-UAAs, where cis/trans isomerism is readily light-controlled. Successful applications include regulation of bacterial transcription activator CaP¹⁴⁵, sensor protein CaM affinities¹⁴⁶ (via helix-switching) and catalytic activity of luciferase.¹⁴⁷ Ultimately, combinations of alkyne-amber suppression and iEDDA-chemistry with tetrazine-azobenzene-inhibitor constructs (bioorthogonal ligand tethering - iBOLT) enabled selective activity control of the otherwise indistinguishable MEK1/2 isozymes in mammalian cells (Fig. 1.9b).¹⁴⁸

1.4.4 Capturing and locking protein interactions on demand

UAAs may also be harnessed to capture protein interactions at a certain time and space. Unlike classic pull-downs of transient interactions, generation of covalent linkages preserve the interaction complex beyond processing steps with little loss of information. Such approaches become particularly attractive, when one wants to study transient and low-affinity bindings.¹⁴⁹⁻¹⁵⁰ Mainly two approaches are used for capturing of protein complexes. Chemically induced stabilisation is best performed by genetically encoding electrophiles, since proteins already bear a variety of nucleophiles. The electrophilicity of the UAA needs to be carefully fine-tuned to withstand side-reactions under physiological conditions and only crosslink, when the POI is in proximity to desired interaction partner.¹⁵¹ The general concept of stable electrophilic UAA design lies in the exploitation of the proximity effect (Fig. 1.10a and b). Upon protein interactions, the effective local concentrations of both, the protein bearing the UAA electrophile and a vicinal nucleophile at the interaction site of the other interaction partner are increased by orders of magnitude, which is sufficient to exceed the reactivity threshold, leading to a subsequent displacement reaction and creation of a covalent linkage.¹⁵²⁻¹⁵³ Several UAA-electrophiles have been successfully crafted to be only activated once this proximity effect is induced. The majority targets the thiol group of cysteines. The corresponding chemical crosslinker UAAs bear handles such as halides¹⁵⁴⁻¹⁵⁵ or Michael-acceptors,¹⁵⁶ although moieties which aim at other residues were also successfully employed (aryl isothiocyanates¹⁵⁷ for amines; aryl carbamates¹⁵⁸ for Lys, Cys, Tyr; aryl fluorosulfates¹⁵⁹ for Lys, His, Tyr). Beyond several proof-of-principle experiments, proximity-driven reactions were for instance employed to facilitate structural elucidation of the low-affinity ternary Rab1b:GDP:DrrA complex, which plays an important role in vesicle trafficking and Legionella pathogenicity.¹⁶⁰ However, difficulties arise when no suitable nucleophile in proximity to the installed UAA is available. Screening various amber sites and occasional mutations of the corresponding nucleophilic site (e.g. in *in vitro* experiments) are then additionally required, minimising the general applicability.

Photocrosslinking represents a suitable alternative for capturing transient interactions in an unbiased fashion and without any further manipulations.¹⁶¹⁻¹⁶³ The general concept for all chemically stable photocrosslinkers lies in the conversion into highly reactive species upon UV-light irradiation. Covalent linkages are generated when the intermediates become immediately quenched by proximal residues, ideally from the neighbouring interaction partner. Differences between the various photocrosslinkers mainly lie in the activation wavelength and overall physicochemical properties.^{161-162, 164}

Benzophenones (Fig. 1.10c, top left) are typically activated at wavelengths of 350-365 nm.¹⁵⁰ The generated diradical abstracts a hydrogen radical from C-H group of the amino acid backbone, followed by subsequent reaction of the newly created backbone radical with the benzophenone radical.¹⁶⁵ The relatively long-lived diradical species (up to 120 μ s) coupled with reversibility of the excitation may lead to comparably high crosslinking efficiencies. This renders benzophenones a still frequently used photocrosslinker, despite its bulkiness and rigidity, great hydrophobicity and potential side-reactions of the carbonyl functionality with amines.¹⁶⁶

Aryl azides (Fig. 1.10c, top right) are activated at lower wavelengths of 250 nm.¹⁵⁰ Similar to benzophenone diradicals, the generated nitrene readily inserts into C-H and heteroatom-H bonds. However, broader usage is hampered by the required low activation wavelength, which is damaging to cells as well as the conversion of the nitrene to the more stable dehydroazepine intermediate, when no suitable reactant is available. This intermediate is less reactive compared to the nitrene and only reacts with nearby nucleophiles, which decreases the overall crosslink scope.

The small size and benign activation wavelength of 350-365 nm renders diazirines (Fig. 1.10c, bottom left) nowadays a highly attractive and widely used photocrosslinker. The generated highly reactive carbene species (under loss of N₂) with typical half-lives of only nanoseconds swiftly inserts, alike nitrenes, into proximal C-H and heteroatom-H bonds.¹⁵⁰ The major side-reaction pathway includes carbene quenching by surrounding water molecules as well as rearrangement of the diazine into the more stable diazo intermediate. The diminished reactivity overall decreases the options of insertion sites and hence crosslink resolution. The degree of side-reactivity however may be reduced by introduction of a trifluoromethyl and aromatic group into the diazine scaffold.¹⁶⁷

Tetrazoles (Fig. 1.10c, bottom right) have a similar activation wavelength as diazirines. Upon UV-light irradiation and elimination of N₂, the reactive nitrile-imine intermediate is readily quenched by nucleophiles, preferably vicinal carboxylic acids.¹⁶⁸

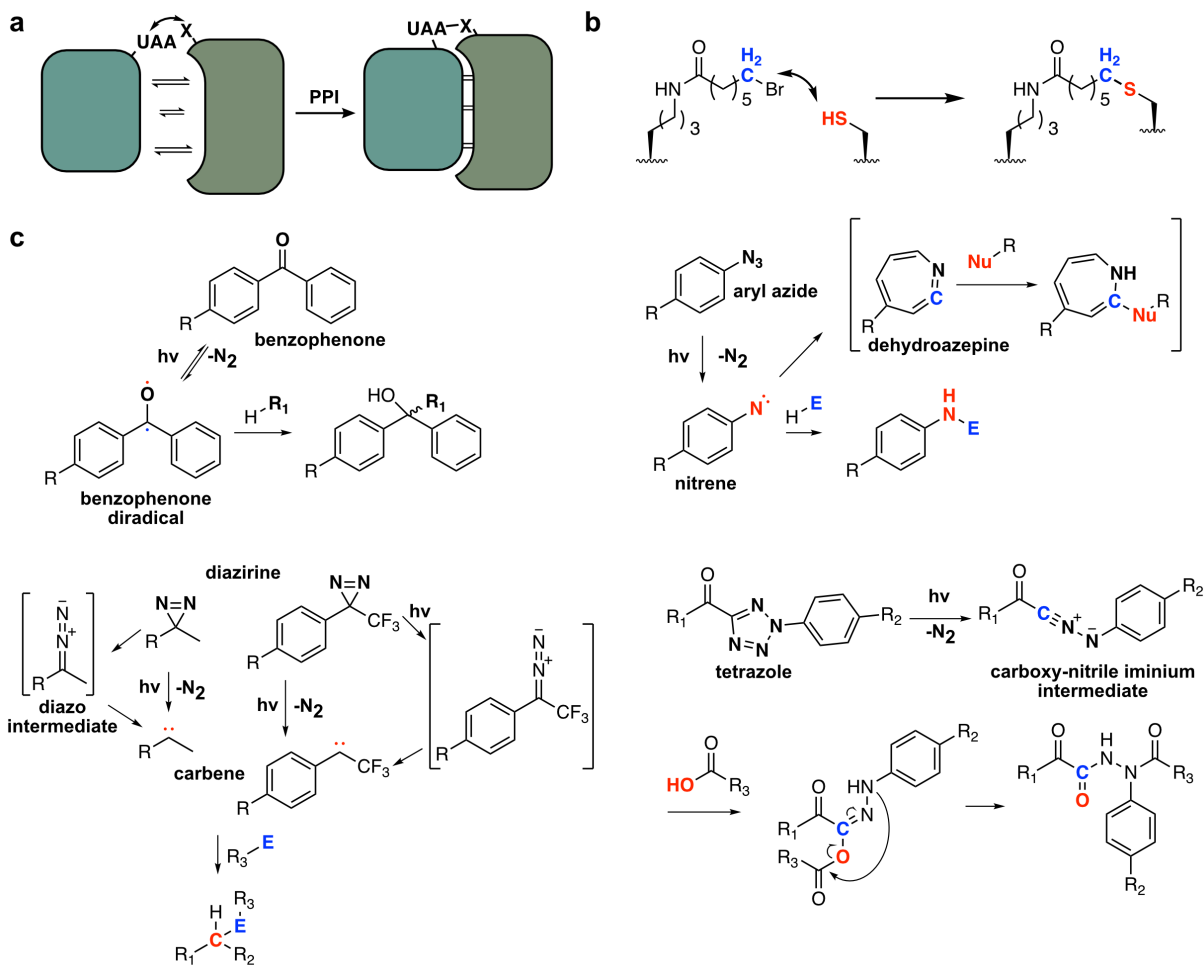


Fig. 1.10 | Proximity-driven locking of protein interactions. Attractive protein interactions (a) arrange the UAA (expressed in the POI) proximal to a neighbouring functional group of the other interactor (denoted as X), eventually trapping of the protein interaction. This may readily occur with UAA-electrophiles as exemplarily depicted in b or in dependence of light as an external trigger, as exemplarily shown for the various photocrosslinkers in c (top right – benzophenone, top left - aryl azide, bottom left - diazirine, bottom right – tetrazole). The highly reactive intermediate is immediately quenched by suitable groups or e.g. by surrounding water molecules.

In all cases, diazirine¹⁶⁹, benzophenone¹⁷⁰, aryl azide¹⁷¹ and tetrazole-bearing¹⁷² UAAs have been synthesised and are readily available options for the genetic encoding into proteins and various photocrosslinking studies. The post-crosslinking analysis is then further accompanied by western blot (WB) and proteomics studies to elucidate the interactome of the POI in great detail.

Overall, diazirines provide the greatest flexibility in terms of benign UV-activation wavelength, chemical stability, hydrophilicity, side-chain and crosslink (radii) flexibility and incorporation via the PyIRS system. Examples include substrate profiling of bacterial chaperone HdeA¹⁶⁹ under various stress conditions and elucidation of chromatin rearrangement dynamics in yeast.¹⁷³ Nonetheless, the other moieties have been used as well, especially in regions of high hydrophobicity such as inside barrels or protein surfaces.¹⁷⁴⁻¹⁷⁶

1.5 Modification and extension of the genetic code expansion technology

Despite the great advances in protein research, which were made amenable with genetic code expansion and bioorthogonal chemistry, certain obvious obstacles still remain.

Repurposing the amber codon for UAA incorporation inherently competes with natural stop codon interpretation and subsequent recruitment of release factor 1 (RF1) of host cells, which terminates protein translation and thus overall diminishes protein expression yields. To minimise the truncation effects during protein expressions, an *E. coli* strain,¹⁷⁷ which lacks RF1 and all genomic amber codons (replaced to TAA stop codons), was created. However, diminished proliferation rates resulted in limited use thus far. A more straightforward approach to improve the protein amber suppression yields lies in increasing the cellular availability of any elements required for amber suppression. This includes the development of modified plasmid systems¹⁷⁸ (e.g. promoter choice, aaRS and tRNA copy numbers on plasmids) for increased availability of the GCE elements, enhancing the catalytic efficiency of aaRSs with appropriate mutations¹⁷⁹ or raising UAA concentrations during protein expressions (Fig. 1.11a). Recent research in mammalian cells also successfully established an orthogonal translating organelle.¹⁸⁰ The recruitment of all amber suppression factors to that compartment not only effectively increases their local concentrations and thus amber codon suppression efficiency but also deprives any interactions with the cytosolic translation machinery with its accompanied side-effects of UAA misincorporations at other amber stop codon containing genes and early termination due to RF recruitments. Another recently published study also highlighted optimisations of the codon context up- and downstream the amber position as considerable factor for improving protein yields.¹⁸¹

Besides developing strategies to improve the amber suppression yield, additional research aims to further extend the side-chain diversity. This may be achieved with adequate combination of bioorthogonal chemistry and selective protein labelling (Fig. 1.11b). For instance, amber suppression of proteins bearing azide-GG-lysine UAA, followed by Staudinger-reduction produces proteins bearing a GG-lysine motif. Incubation with sortase and proteins bearing the other half of the sortase recognition motif, this enzyme mediated ligation enabled for the first time the generation of site-specific protein-PTM (ubiquitin, SUMO) on POIs such as ubiquitin and PCNA with minimal motif alterations (also cf. Fig. 1.5a).¹⁸² Exemplarily shown for sortase generated K63-diubiquitin, the sorting sequence only minimally influenced binding (approximately 4x lower) towards the ubiquitin-interacting motif of Rap80 when compared to enzymatically generated wt K63-diubiquitin. The sorting motif was also proven resistant to hydrolytic cleavages, when incubated with DUBs. Intriguingly, the dubbed 'sortylation' approach is also applicable at *in vivo* settings, enabling potential real-time Ub-pathway studies in mammalian cells.

The third major research area aims to develop strategies to extend the number of UAAs, which can be incorporated (Fig. 1.11c). Access to such multiple suppression schemes is highly desirable for the production of proteins bearing multiple bioorthogonal handles and biophysical probes, which would allow biochemical and biophysical characterisations in greater detail. For instance, proteins bearing multiple PTMs may be readily projected to

investigate effects on protein function and activity, especially their interdependency and counteracting effects (PTM crosstalk).^{4, 6} Accompanied with such scheme designs is the need to discover additional aaRS/tRNA pairs and blank codons. Following the stop codon recoding approach and limited number thereof (TAG, TAA, TGA), further stop codon repurposing would become increasingly challenging with rising concurrent suppression events. Additionally, all newly discovered and evolved aaRS/tRNA pairs need to exclusively recognise their cognate UAA and to be orthogonal not only to the host system but also amongst themselves. Double suppression was readily achieved utilising the mutually orthogonal *Mj*TyrRS and *Py*IRS in *E. coli* expressions with TAA and TAG as blank codons and their respective tRNAs.¹⁸³ To extend double suppressions to eukaryotic systems, recent genome mining and bioinformatic analyses of evolutionary distinct tRNA and aaRS variants revealed additional aaRS/tRNA pairs (e.g. *Methanomethylphilus alvus* *Py*IRS/tRNA pair) that are mutually orthogonal to *E. coli* aaRS/tRNA pairs, whilst bearing similar advantageous properties of the *Mb**Py*IRS/tRNA and *Mm**Py*IRS/tRNA pairs.¹⁸⁴⁻¹⁸⁶ The extension of the cellular codon capacity was achieved via the development of an orthogonal translation system along the advancement of amber suppression. Extension of the triplet codon system to quadruplet codons may provide up to 256 blank codons for UAA incorporations. A corresponding orthogonal ribosome ribo-Q1 was evolved to decode quadruplet codons of orthogonal mRNAs with extended anticodon tRNAs and respective aaRSs, enabling dual UAA incorporations into POIs in *E. coli* together with the amber codon.¹⁸⁷ Ultimately, the combination of this quadruplet translation system with the newly identified orthogonal aaRS/tRNA pairs led to successful triple UAA incorporation into proteins.¹⁸⁵

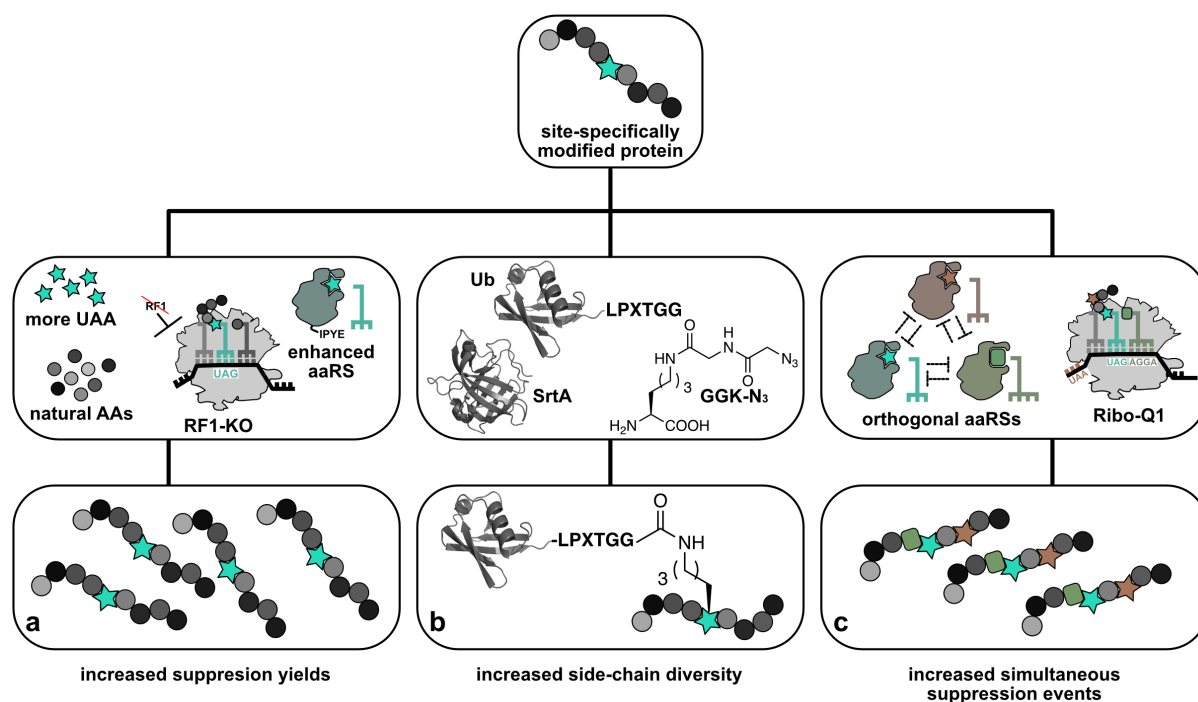


Fig. 1.11 | Strategies to expand GCE. Several approaches have been developed and are still under development to optimise GCE, of which a selection is highlighted in this figure. This regards (a) increased yields, (b) side-chain diversity via combinations of different labelling techniques and (c) development of reliable simultaneous multiple suppression schemes.

The recently published synthetic 'Syn61' *E. coli* strain provides a highly promising alternative to the standard stop codon and quadruplet suppression approaches. This strain bears a synthetically created genome and only relies on 59 sense codons and two stop codons for survival. The liberated TCG and TCA codons (which encodes Ser) along with the deletion of their cognate tRNAs, the TAG codon and RF1 thus offers the opportunity for sense codon repurposing and multiple UAA incorporations without implementing an orthogonal translation machinery and generation of truncated proteins.¹⁸⁸

Altogether, genetic code expansion along with the incorporation of UAAs bearing a plethora of different biorthogonal and biophysical probes provides a powerful approach for advanced *in vitro* and *in vivo* protein studies. Developments in all directions are ongoing to push protein research further for more native-like studies and to possibly enable the synthesis of artificial biopolymers for biotechnological and pharmaceutical purposes.

Chapter 2

Substrate profiling of human caseinolytic protein P (hClpP) using a photocrosslinker UAA

This chapter is based on a manuscript in preparation.

Contributions

Organic syntheses, genetic code expansion and photocrosslinking experiments as well as protein expressions and purifications were performed by Tuan-Anh Nguyen, Lab for Synthetic Biochemistry, TU Munich. Thomas F. Gronauer, Chair of Organic Chemistry II, TU Munich, performed enzyme activity assays, mass-spectrometry experiments and corresponding analyses.

2 Substrate profiling of hClpP

2.1 Proteases and their conventional profiling tools

As outlined in the previous chapter, proteins constitute the main executive units in cells to maintain overall homeostasis. Considering their pivotal roles, much research focuses on developing and applying methods to detect, monitor and understand their functions at greater detail. However, along with cellular protein productions, a complementary mechanism, which selectively processes proteins, is of equal importance for cellular homeostasis. Proteolysis, i.e. the cleavage of protein backbone is catalysed by so-called proteases, is considered a further crucial PTM for the regulation of various protein functions involved in cell signalling and proliferation, metabolic processes, immune response, cellular homeostasis and cell death.¹⁸⁹⁻¹⁹¹ The mode of catalysis is in principal similar for all proteases and involves nucleophilic attack of a water molecule at the amide bond leading to bond cleavage. Activation of the water molecule either occurs directly via Brønsted/Lewis acids (Asp/Glu proteases, matrix metalloproteases/MMPs) or indirectly via acylintermediate trapping, which requires a catalytically active amino acid set (catalytic dyad or triad; Ser, Thr, Cys proteases)¹⁸⁹. Functional importance and irreversibility of the reaction demand protease activity and selectivity^{189, 191} to be tightly regulated, as aberrant functionality is often related to pathophysiological phenotypes such as cancer, neurodegenerative disorder, inflammatory and cardiovascular disease. Likewise, many pathogens (e.g. HIV, *C. difficile*, *Y. pestis*, *L. pneumophila*, *P. falciparum*) need the activity of proteases for proliferation.¹⁹¹ Of the approximately 600-700 human enzymes annotated with known or putative proteolytic activity, only a minority is however characterised to date.¹⁹⁰⁻¹⁹² Thus, elucidating their physiological role, i.e. their substrate scope and specificity, remains crucial, also in respect to drug development processes to tackle various diseases.

Much of the protease characterisations are based on monitoring natural substrate turnover *ex vivo* (coupled to immunoblotting or mass spectrometry) or on determining sequence-selectivity and profiling cleavage-sequence preferences with tailor-made substrate-based probes *in vitro* (combined with fluorescence assays, cell-based imaging).¹⁹¹⁻¹⁹⁷ While these methods are readily scalable and offer flexible molecule design, assignment of proteolytic events to a certain (active) enzyme state remains difficult¹⁹¹. Activity-based probes (ABPs) are designed to selectively trap active proteases.^{162, 191} The probes are small to medium-sized molecules and equipped with a recognition motif with a closely attached electrophilic warhead and a reporter tag. The recognition motifs of ABPs are designed to resemble structural and functional groups of the natural substrates to fit within the catalytic pocket of enzymes. The induced proximity effect upon ABP binding leads to covalent trapping of the closely attached warhead with a vicinal nucleophile, while the reporter tag facilitates downstream identification of the ABP-linked enzyme. Thus, ABP trapping represent a direct measure of detecting the active portion of an entire protease pool.¹⁶² In cases, where no proximal nucleophile in the active site is available (e.g. in MMPs), covalent linkages may be induced via photocrosslinking instead (affinity-based probes, AfBP).¹⁶² In either case, reporter

tag affinity purification coupled to tandem-MS facilitates isolation and identification of the respective trapped proteases. However, while those probes were routinely applied for mechanistical studies, pharmacophore development and discovery of disease biomarkers, this approach insufficiently addresses the physiological substrate scope.¹⁶² This is largely due to the method design, where the input is and needs to be (to some degree) biased for detection purposes (Fig. 2.1, left). Reverting this principle would liberate the bias from the input and thus allow full substrate scope studies *in vivo*, but would require manipulations on the protease site for detection (Fig 2.1, right). Also, since the substrates are by virtue unknown, any protease alterations still need to be able to cover the entire substrate pool in an unbiased fashion.

Loss-of-function studies would allow for the comparison of proteomes between protease inactive (via KO or temporary gene silencing) and wt control cell samples, where gain-of-signal is possibly linked to the protease activity and specificity.¹⁹⁰ Likewise, gain-of-function assays, i.e. overexpression of proteases in cells, which ideally result in loss of signals, may be complementary pursued.¹⁹⁰ However, to bear in mind is that unexpected signal turn-out may be originated from functional redundancy, where e.g. loss of proteolysis is rescued by complementary processes.¹⁹⁰ Consequently, downstream experiments are required to validate putative candidates.

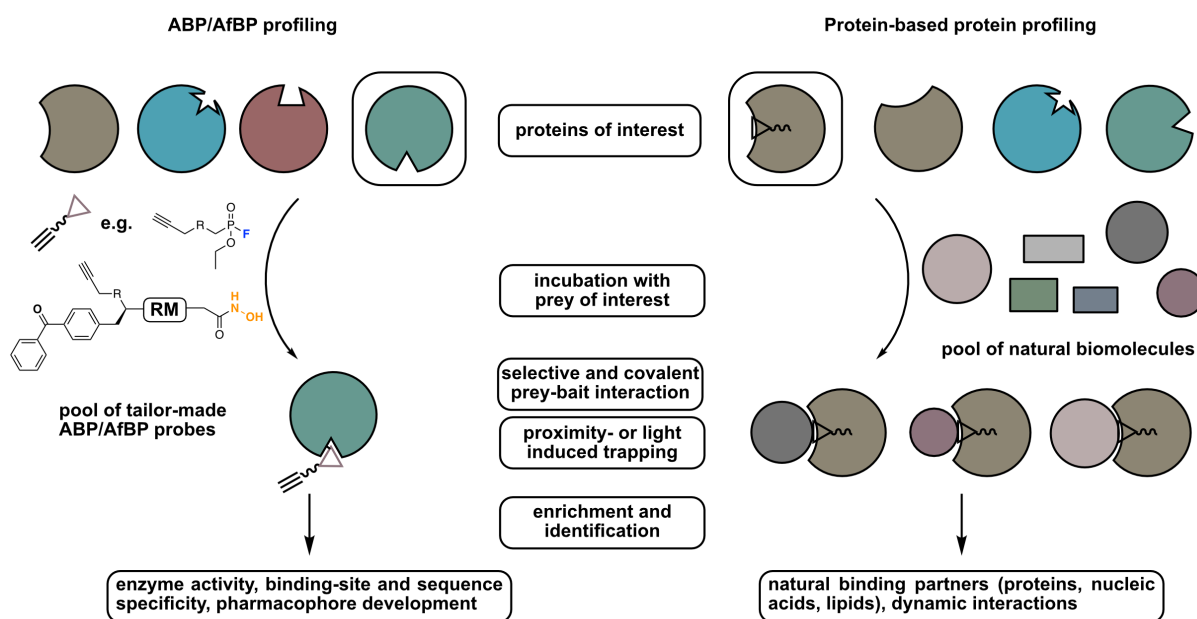


Fig 2.1 | Protease-profiling approaches. ABP/AfBP-profiling (left) utilises tailor-made small-molecule reporter groups (as indicated by coloured elements, equipped e.g. with an alkyne tag) to investigate biochemical features. Chemically speaking, reporter groups/warheads may include fluorophosphonates¹⁹⁸, which react with the catalytic Ser residue of Serine proteases or a construct bearing a hydroxamate head group (orange), which coordinates to Zn²⁺ in MMPs, while UV-light induction leads to activation, crosslink with the benzophenone (RM – recognition motif).¹⁹⁹ Switching the principle via installation of the profiling handle in the protease (right) would enable unbiased, dynamic screening of natural substrates *in vivo* and preferential turn-over.

Following the principal of the ABP-/AfBP-profiling approach, *in vivo* substrate crosslinks of a protease of choice and subsequent identification directly links targets to their processing

enzyme and thus should provide a first good estimation of substrate preferences (protein-based protein profiling). As argued above, this requires the installation of reactive probes into the protease of choice (Fig. 2.1, right).

2.2 Genetic code expansion as a suitable tool for protease-substrate profiling

Amber suppression would allow to investigate the natural substrate scope of proteases with great spatiotemporal control. Specifically, spatial control is achieved with the site-specific incorporation of UAAs bearing reactive warheads close to the catalytic centre. This would ensure that any captured substrates were destined to proteolysis. Because chemical crosslinkers require nucleophiles in close proximity, they are limitedly suited for unbiased trapping. Alternatively, the highly reactive intermediates of photocrosslinkers delivers the demanded reactivity profile for unbiased insertions (cf. chapter 1.4.4).¹⁶¹ Also, UV-light dependent photocrosslinker activation at defined time-points provides additional temporal control for protease profiling studies under defined conditions. Considering mechanistic details, physicochemical properties, effective crosslinking radii and the possibility of mammalian applications, diazirine UAA derivatives represents the most appropriate option (cf. chapter 1.4.4). While GCE and diazirine photocrosslinkings have been already exploited for various transient PPI studies on chaperones¹⁶⁹, adaptor proteins²⁰⁰⁻²⁰¹ and in chromatin biology²⁰²⁻²⁰³ and bacterial protease-substrate profiling,²⁰⁴⁻²⁰⁵ however, no applications of photocrosslinking studies for profiling mammalian proteases are known so far.

2.3 Aim of this work

In this chapter, we aimed to establish and prove the generality of the photocrosslinking approach for mammalian protease substrate profiling (Fig. 2.2). This was exemplarily shown with the site-specific incorporation of photocrosslinker UAA 4 (Fig. 2.2, Scheme 2.1; UAA subsequently denoted as dK) at selected positions of mitochondrial (mt) human caseinolytic protein P (hClpP) in HEK293T cells and identification of its target substrates/interactors. hClpP plays a crucial role in mt homeostasis via its proteolytic activity, which also includes removal of dysfunctional and oxidatively damaged proteins and dysfunction is associated with various diseases. Previous eukaryotic ClpP target trapping and identification studies relied on catalytically inactive ClpP, entire deletion of its gene or additional genetic and chemical manipulations within various host systems (see chapter 2.4). Site-specific incorporation of dK in proximity to the catalytic centre of hClpP followed by substrate trapping via UV-light induced photocrosslinking hence provides a complementary approach by directly linking protease activity with processed substrates. Target identification was achieved via selective enrichment of dK-containing hClpP bearing an HA-tag, followed by mass spectrometry measurements and analyses of the crosslinked proteins. The identified crosslinked proteins were compared with literature known substrates for mutual overlap as well as identification of novel substrates of hClpP (Fig. 2.2). Additionally, this approach allowed to investigate shifts

of substrate specificity under stress conditions. Specifically, HEK293T cells expressing hClpP-dK variants were challenged with rotenone, an inhibitor of complex I of the respiratory chain complex (RCC) to investigate the substrate scope hClpP under oxidative stress.

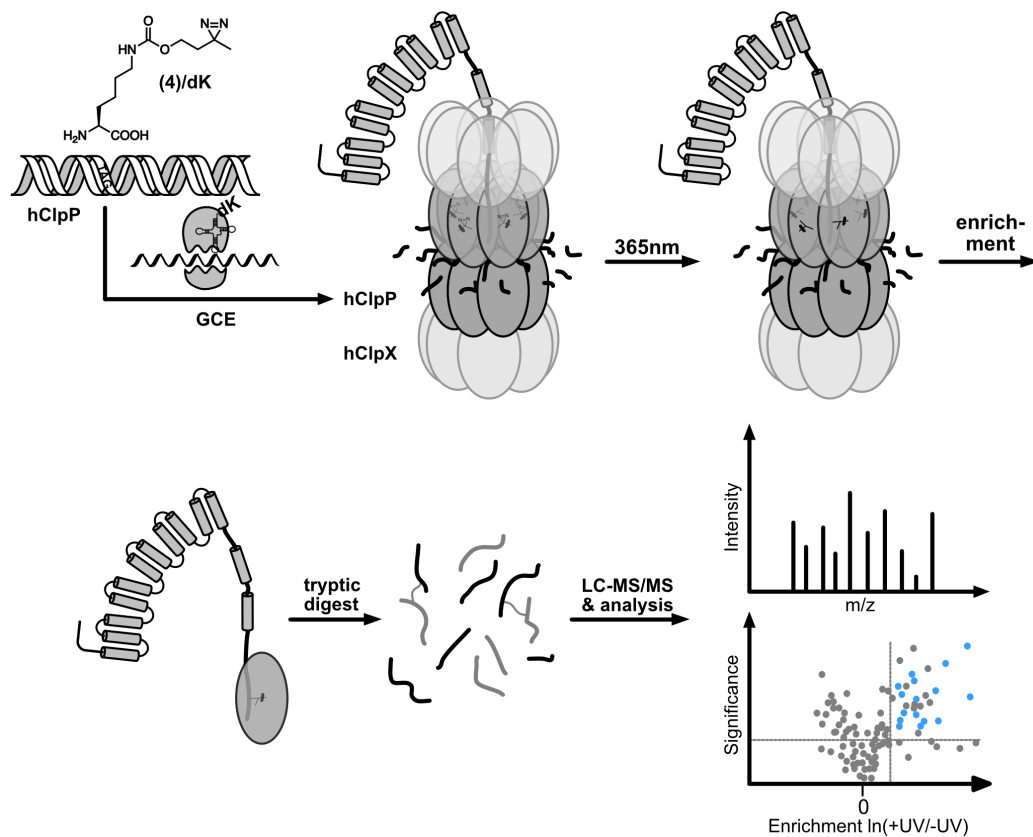


Figure 2.2 | General protease-profiling approach for hClpP substrate determination. After incorporation of dK at selected positions of hClpP, the hClpX/P system becomes fully assembled to perform its proteolytic activity. Irradiation at 365 nm leads to diazirine activation and subsequent crosslink with the proteolytic substrates. This covalent crosslink is preserved during selective hClpP enrichment via its encoded HA-tag. Tryptic digest followed by proteomic analysis of the digested peptide fragments leads to identification of the trapped substrate.

2.4 hClpP biology

2.4.1 hClpP and its role in mitochondrial homeostasis

The mt proteome comprises more than 1200 proteins, of which only 13 proteins, all essential subunits of the RCC, are encoded in the mt genome, while the remaining mt members are shuttled from nucleus and cytosol.²⁰⁶⁻²⁰⁹ Maintenance of the separate organelle's proteome hence requires an independent machinery, including a protein recycling system.^{206, 209} Eukaryotic ClpP, in conjunction with ClpX, belongs to the mt ATP-dependent protease system and plays an active role in its protein quality control.^{206, 210-211} While ClpX has ATP-dependent unfolding capability (AAA+ chaperone), ClpP possesses the actual proteolytic activity. Residing in the mt matrix in proximity to the RCC, a primary function is the degradation of oxidatively damaged proteins and dysfunctional proteins. Also, maintenance of mt protein homeostasis includes constant processing of metabolic enzymes, members from the RCC, proteins involved in mt transcription and translation as well as chaperones. For instance, ERAL1, a putative 12S

rRNA chaperone is targeted by ClpP.²¹² While required for assembly of ribosomal 28S units, degradation is necessary to ensure complete mitoribosome formation. Overall reduced levels and thus mt translation rates were observed in ClpP^{KO} mice, which can be considered as a positive feedback mechanism for the limited proteolytic capacity, when hClpP is absent. Research in *C. elegans* furthermore suggests a role in the mt unfolded protein response (mtUPR),²¹³ a mitochondria-to-nucleus stress signalling pathway, which overall downregulates mt translation and alters metabolism as a counter-measure towards the stress inducer. To what extent this applies to mammals, remains to be studied. Research hints to a minor role of ClpP in the mtUPR.²¹⁴

Additionally, cellular malfunctions in form of underrepresented or overrepresented ClpP levels has been linked to various diseases.²¹⁰⁻²¹¹ Double ClpP^{KO} mouse models have shown phenotypes related to Perrault Syndrome in humans, a rare genetic disease which manifests in loss of hearing and infertility.^{212, 215} ClpP is however upregulated in various cancer types²¹⁶, likely to compensate for increased oxidatively damaged proteins caused by higher energy consumption and upregulated oxidative phosphorylation.²¹⁰⁻²¹¹ Hence, metabolic boosting renders cancer cells particularly dependent on and susceptible towards ClpP activity. Indeed, KO of ClpP in various cancer cell lines resulted in decreased viability, proliferation and metastasis.²¹⁶⁻²¹⁷ Protease inhibition has been proposed as a promising approach for various cancer treatments, especially as non-malignant cells seem to be minimally affected by inhibition²¹⁸. For instance, β -lactones²¹⁶ and phenyl esters²¹⁹⁻²²⁰ have been identified as putative inhibitor classes in this regard. Interestingly, while inhibition increases the oxidative burden resulting in cell stress and death, hyperactivation shows similar results. This is mainly due to overconsumption of various substrate proteins, including members of the RCC resulting in dysfunctional mitochondria. Acyldepsipeptides²²¹ and imipridones^{218, 222} are promising activators, which bind in the ClpXP interface and force the protease in a permanent active state resulting in uncontrolled protein degradations.

2.4.2 Identifying ClpP interactors – current status

The current knowledge of eukaryotic ClpP interactors and substrates stems from heterogeneous enrichment protocols performed in various species and tissues. Strategies included either genetic alterations of ClpP or chemical manipulations within the target system. For instance, ClpP^{KO} in *Podospora anserina*, a fungal model organism, was rescued with hClpP^{wt} or hClpP^{S153A} (catalytically inactive) to study its substrate scope in native environment.²²³ While the *P. anserina* ClpX still performed substrate unfolding in presence of hClpP, the S153A mutant was unable to process proteins, thus providing whole ClpP-substrate complexes. Isolation and comparative MS/MS analysis yielded enrichment of protein clusters (and thus potential ClpP targets) belonging to several metabolic pathways, namely members of the tricarboxylic acid (TCA) cycle, pyruvate dehydrogenase (PDH) and RCC. Likewise, N-terminome profiling²²⁴ and label-free proteomics^{212, 214} of ClpP^{KO} or ClpP^{S153A} vs. ClpP^{wt} in mouse embryonic fibroblasts and heart lysates revealed similarly enriched clusters, notably comprising metabolism, RCC and mt translation (*vide supra*). Subsequent WB analysis of

selected targets in ClpP^{wt} and hClpP^{KO} cells confirmed the substrate identities via increased signals in KO samples.²²⁴

ClpP analyses in human context are scarce thus far, but likely display similarly enriched patterns, although specific protein identities may be altered due to genomic and proteomic heterogeneity between species. For instance, BioID-MS, i.e. biotinylation of proteins proximal to the ClpP-BirA* biotin ligase fusion protein (R118G mutation results in promiscuous ligase activity) in acute myeloid leukemia cells followed by enrichment and MS/MS revealed a profound protein interaction network of similar scale as determined above (RCC, metabolism).^{216, 218} Likewise, proximity labelling was performed with global, on-lysine reactive chemical crosslinkers in K562 and HepG2 cells.²²⁵ The results after pull-down against hClpP and proteomics revealed again a similar functional protein network as described above. However, to note is that in both cases, protein proximity-labelling potentially results in loss of spatial resolution since interactors/proteolytic substrates and ClpP unrelated proteins may not be clearly distinguished.

2.4.3 Structural analysis of the hClpX/P binding

Biochemical analysis has shown hClpP to assemble into a ring of seven monomers.²²⁶ However, the heptamer has only limited peptidase activity, due to improper arrangement of the catalytic triad (S153, H178, D227) and missing protein unfolding capability. Full catalytic potential is achieved with the assembly of two hClpP heptamers to a barrel-shaped tetradecamer (Fig. 2.3) accompanied by the interaction with at least one ring-shaped hClpX hexamer on the apical surface of the tetradecameric hClpP barrel.²²⁷⁻²²⁸ While ClpX is responsible for substrate recognition, ATP-dependent protein unfolding and direction into the proteolytic ClpP chamber, the mode of interaction with ClpP is not fully resolved. Docking of ClpX onto ClpP is ATP-dependent and believed to be mediated by contacts of 1) the IGF-loop of ClpX with the hydrophobic pocket of the apical surface of ClpP and 2) pore-2 loop of ClpX with the N-terminal loop of ClpP (Fig. 2.3a).²²⁹⁻²³⁰ Interestingly, species specificity of ClpX seems to be loose for hClp activation (*vide supra* with *P. anserina* ClpX, also with *E. coli* ClpX), while the opposite arrangement (e.g. hClpX with *EcClpP*) remains inactive.²²⁷ Likewise, the ClpP heptamer subunits are mainly stabilised by inter-domain interactions,^{228, 231} including packaging of $\alpha 2/3$ -helices against the $\beta 1$ -sheet of the next subunit, contacts of $\alpha 2$ with $\alpha 1$ of the next subunit as well as $\alpha 5$ with $\beta 2/\alpha 7$ of the other subunit (Fig. 2.3b and Fig. 2.4). The unique and flexible C-terminal amino-acid extension (249-277) of hClpP is further stabilised by hydrophobic interactions (Fig. 2.3c).^{228, 231} For instance, P248 contacts Y138 and L140 of an adjacent monomer, while P249 forms an intramolecular hydrophobic core at the base of the C-terminal loop.²²⁸ Intriguingly, this C-terminal extension seems to negatively affect ClpX binding, as evident from *in vitro* binding studies and was further suggested to serve as a docking station for additional adaptor proteins possibly important for ClpP structure and function.²²⁸

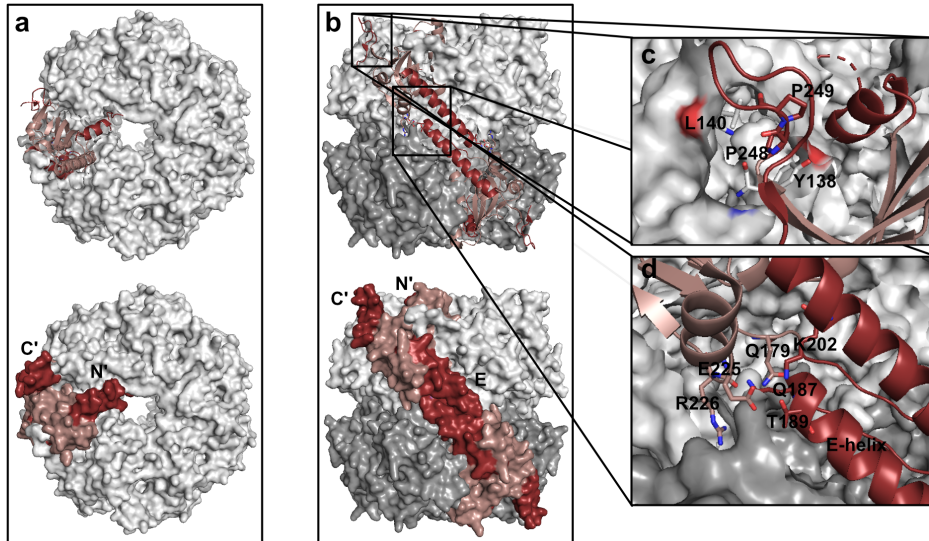


Fig. 2.3 | Crystal structure of the barrel-shaped hClpP tetradecamer (PDB: 1TG6). hClpP forms a dimer of heptamers, with a monomer highlighted in red (**a** and **b**). The top view in **a** illustrates the apical surface with both protein termini (in dark red) and the hydrophobic interaction area of N' for ClpX docking. The flexible C'-loop lies in proximity to the surface. The side view in **b** highlights intermolecular contacts of the opposing monomers (highlighted in dark red, E - E-helix). In **(c)**, inter- and intramolecular stabilisations of the C-terminal extension is highlighted, while **(d)** depicts the close view of E-helix residues from opposing monomers, which are involved in heptamer-heptamer stabilisations. For reasons of clarity, only interactions of E-helix residues from one monomer with the other opposing monomer are shown.

In addition, residues of the E-helix such as Q187 and T189 intercalate with residues of the opposing monomer such as Q179, Q194, K202, E225, R226 (oligomerisation sensor) and stabilise catalytic triad arrangement and heptamer-heptamer contact by favourable H-bonds and salt bridge networks (Fig. 2.3d).^{226, 228} Interestingly, the origin and purpose of the symmetry mismatch between ClpX/P is not fully understood yet, but thought to be important for the complex flexibility to facilitate protein unfolding, translocation, proteolysis and fragment release.^{228, 231}

2.4.4 Determining suitable positions for photocrosslinker dK incorporation in hClpP

Structural homology analyses and *in vitro* substrate peptide profiling²³² identified two major binding pockets of hClpP: The S1 pocket, which shows preferences for hydrophobic residues, is comprised of M154, L157, P180, L205, Y206, Y209, and M224 and the S3 pocket (affinity for aromatic residues), which includes R174, G182, G183, Q194, D197, I198 and L201 (Fig. 2.4a left). Thus, taking the critical residues required for multimer assembly and substrate recognition into consideration, seven positions were elected for amber suppression and substrate crosslinking. M88 (β 1-sheet) and D92 (α 2-helix) lie between the substrate entry pore and the catalytic triad, while the G123/G124/V125 junction loop between β 2-sheet and α 3-helix is in close proximity to the catalytic S153. Furthermore, S181 resides on the β 7-sheet between S1 and S3 proximal to the H178 (part of the catalytic triad) (Fig. 2.4a and b). By introduction of dK at these positions, we aim to crosslink various fragment sizes of proteolyzed proteins from multiple angles in order to diversify crosslinking scope and maximise yield. K261, which lies at the flexible C-terminal loop and outside the barrel, was additionally chosen to potentially account for adaptor proteins of hClpP.

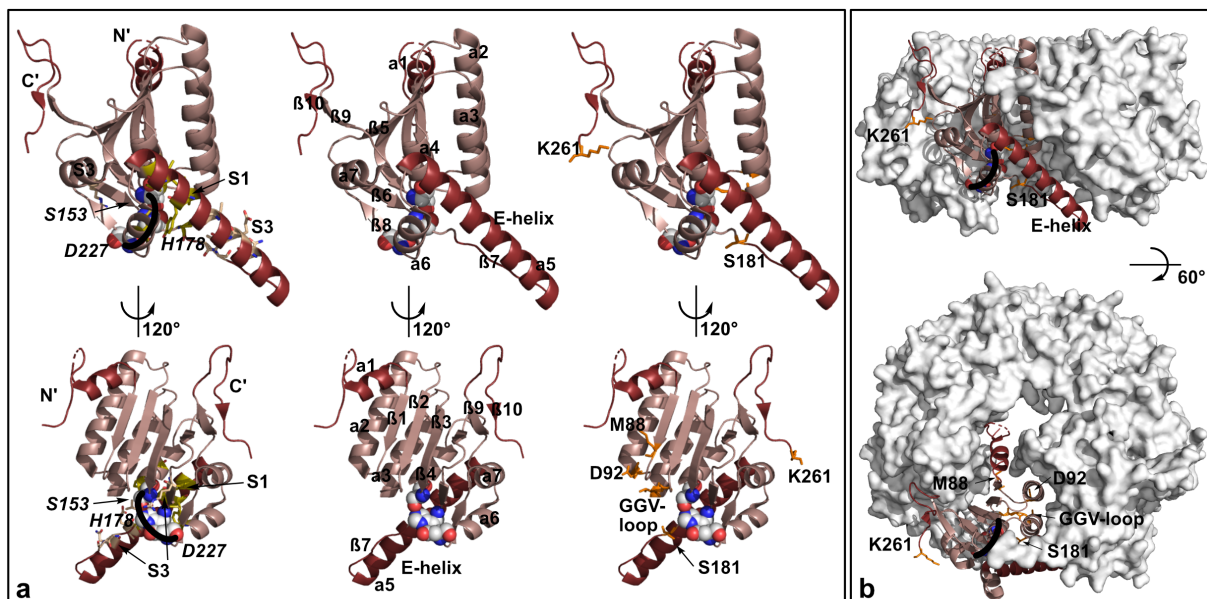
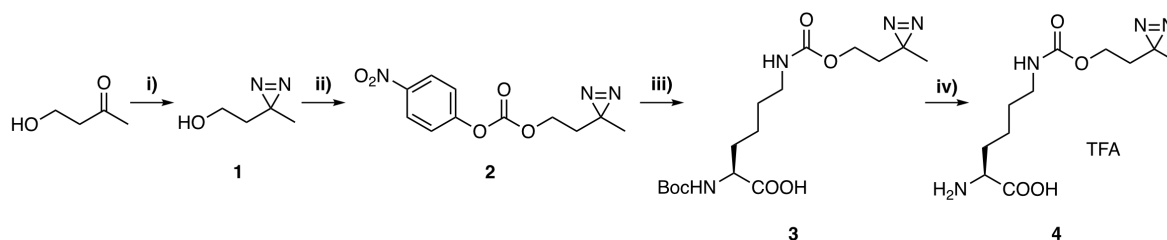


Fig. 2.4 | Structural analysis for amber site introduction. The hClpP monomer with the secondary structural elements (numbered and denoted as: 'a' for α -helices and ' β ' for β -sheets), substrate binding pockets (S1 and S3) and aligned catalytic triad are depicted in **a**, left and middle. Based on the required elements for hClpP functionality, seven positions (highlighted in orange) were chosen for amber suppression (**a**, right). Six positions lie within the catalytic pocket and K261 lies outside. In **(b)**, the structural orientation of the chosen amber sites in within heptamer (side and top view) is illustrated.

2.5 Amber suppression of dK in hClpP and *E. coli* and functionality tests

To prove, whether dK may be incorporated in hClpP while protease functionality is still retained, all amber mutants were expressed in *E. coli*, purified and tested for activity first. The synthesis of dK was performed according to a published procedure²³³ (Scheme 2.1) and was already shown that it may be incorporated into proteins in *E. coli* and mammalian cells.²³³⁻²³⁴



Scheme 2.1 | Synthesis of UAA 4/dK. **i)** a - 7 M NH₃ in MeOH, hydroxylamine-*O*-sulfonic acid (1.1 eq.), 0°C → r.t., o.n., b - Et₃N, I₂, 0°C, 3h, 42%; **ii)** *p*NO₂-phenyl chloroformate (1.2 eq.), pyridine (1.2 eq.), 0°C → r.t., o.n., 62%; **iii)** Boc-Lysine-OH (1.2 eq.), Et₃N (2.0 eq.), r.t., o.n., 60%; **iv)** TFA, 0°C, 3 h, 96%.

However, screening a panel of in-house available PylRS variants led to the identification of a highly efficient synthetase (*Mb* PylRS: Y271M, L271A, C313A; subsequently denoted as dK-RS) for dK incorporation (data not shown). All hClpP-TAG variants were cloned into the well-established pPylt-vector, which was optimised for bacterial amber suppression. Additionally, all hClpP variants contained a C-terminal His₆-tag and lack the N-terminal 56 amino acid mt guiding sequence,²³¹ which would be normally processed in the mitochondria. As shown in Fig. 2.5a, amber suppression experiments in *E. coli* DH10β revealed the general possibility to incorporate dK at any of the seven TAG positions. The G124TAG mutant seemed to exhibit slightly lower expression yields compared to the remaining variants, indicating slight protein instabilities when dK is specifically placed at this position of the loop region. All proteins were subsequently purified and subjected to functional analyses, with purified wt hClpP-His₆ serving as the reference. Analytical size-exclusion chromatography (SEC) provided the first control line in this regard. Expressed wt hClpP in *E. coli* has been reported to elute in a heptameric form on SEC.²²⁶ As shown in Fig. 2.5b, all mutants exhibit similar retention times as the wt hClpP indicating no major influence of the UAA in the hClpP monomer structure and multimer assembly. The functionality of hClpP was further tested with protease activity measurements. This established assay relies on proper ClpX/P complex formation to unfold and degrade model substrate eGFP under ATP consumption.²³⁵ Specifically, *Ec*ClpX, which is known to interact and form a fully functional proteolytic complex with hClpP, was used as the ATP-dependent chaperone. While the *ssrA*-tag on eGFP serves as a degron to be recognised by *Ec*ClpX, hClpP readily degrades eGFP. The resulting loss of fluorescence should thus provide a good estimation of the hClpP activity *in vivo*. As shown in Fig. 2.5c, the activity of hClpP-dK mutants remained similar to wt hClpP. Only in cases, where G123 or G124 was replaced with dK, diminished activities to approximately 40% of wt hClpP were observed. This is likely attributed to the replacement of the achiral glycine with the chiral dK within the loop region,

which leads to local distortion of the β 2-sheet/ α 3-helix structure and orientation of the close residing S153 residue without affecting global assembly (also cf. Fig. 2.4). Interestingly, incorporation of dK at position V125TAG within the loop region did not affect protease activity. The side-chain of V125 turns away from S153 and more towards the inner chamber, implicating a similar orientation of the dK side-chain and less interference with the catalytic serine residue. Overall, despite reduced catalytic activity, the positions G123TAG and G124TAG were included in further crosslinking experiments and proteomic analyses to broaden the coverage scope.

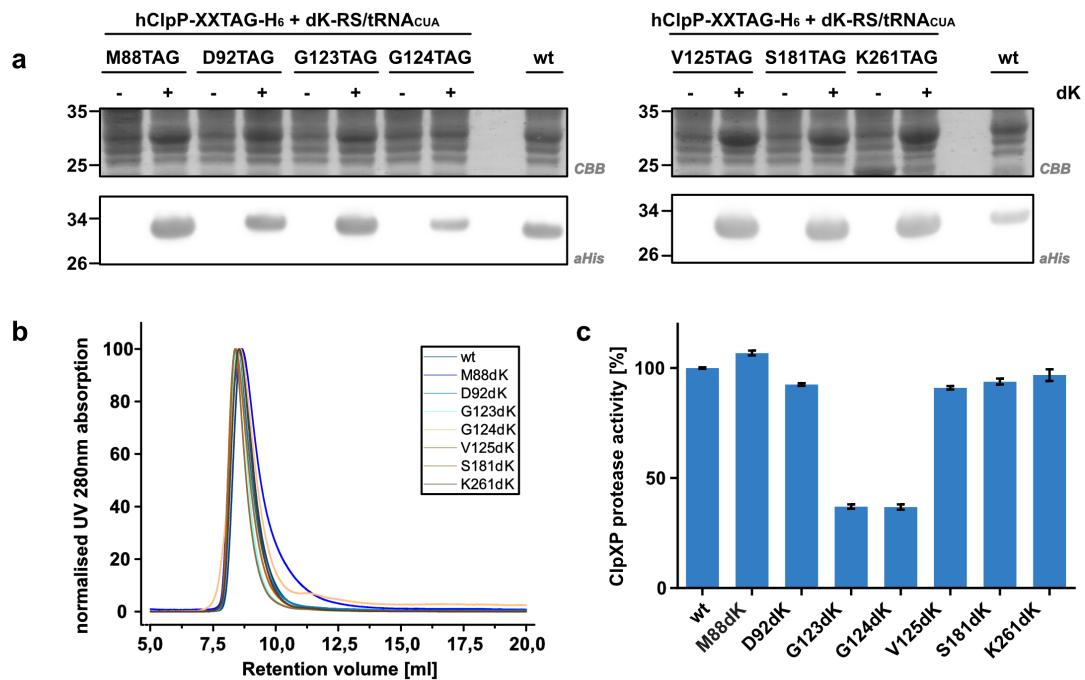


Fig. 2.5 | Bacterial expression and activity test of various hClpP-dK amber mutants. (a) Amber suppression of hClpP only occurs in presence of dK (2 mM). Purified amber mutants exhibited similar SEC profiles (b) and retained wt protease activity (c), apart from both glycine replacements. CBB – Coomassie brilliant blue.

2.6 Expression of hClpP-dK variants in mammalian cells and photocrosslinking

Amber suppression in HEK293T cells was performed with an optimised pEF1 vector system bearing a strong EF1 α promoter and multiple copies of the PyltRNA_{CUA} (under a U6 promoter) on both plasmids coding for hClpP and dK-RS.²³⁶ To facilitate protein enrichment and proteomic analysis, hClpP amber variants (now containing the 56 amino acid mt localisation sequence) were cloned with a C-terminal HA-tag. Brief screening of optimal POI:Mm wtRS plasmid ratios with Bock revealed a 3:1 ratio as optimal in terms of signal output (Fig. SI 2.1). Again, expression of hClpP only occurred in presence of dK, highlighting the efficient amber suppression plasmid system (Fig. 2.6a). As previously observed in *E. coli* expressions, hClpP G124dK exhibited lowest expression amongst all the mutants. The additional WB band running slightly above the main band represents unprocessed hClpP, which still bear the 56 amino acid mt guiding sequence. The amount of processed and unprocessed hClpP between

the different dK mutants and among the same mutants (experiments performed on different days) appears to be highly variable, despite same vector mass input. This is likely attributed to deviations inherent to the transient transfection protocol and conditions.²³⁷

Next, photocrosslinking experiments were performed and detected via WB to gain a first impression of the photocrosslinking efficiency. As depicted in Fig. 2.6b, activation of the diazirine moiety at 365 nm selectively led to occurrence of newly formed, higher running crosslink bands. The broad, partially smeary bandwidths indicate diverse substrate crosslinks of different molecular weights as well as capture of heterogenous fragment sizes of the same protein. Variation of signal intensity among the mutants reflects the various orientation of the diazirine moiety in the substrate chamber resulting into different contact efficiencies and thus meaningful crosslinks. The major band signal in all mutants occurs between 46-58 kDa, which may hint to covalent trapping of two hClpP monomers. Alternatively, since the baseline intensities of hClpP between +/- UV are almost equal (normalised via BCA assay), this may also originate from a highly processed substrate(s).

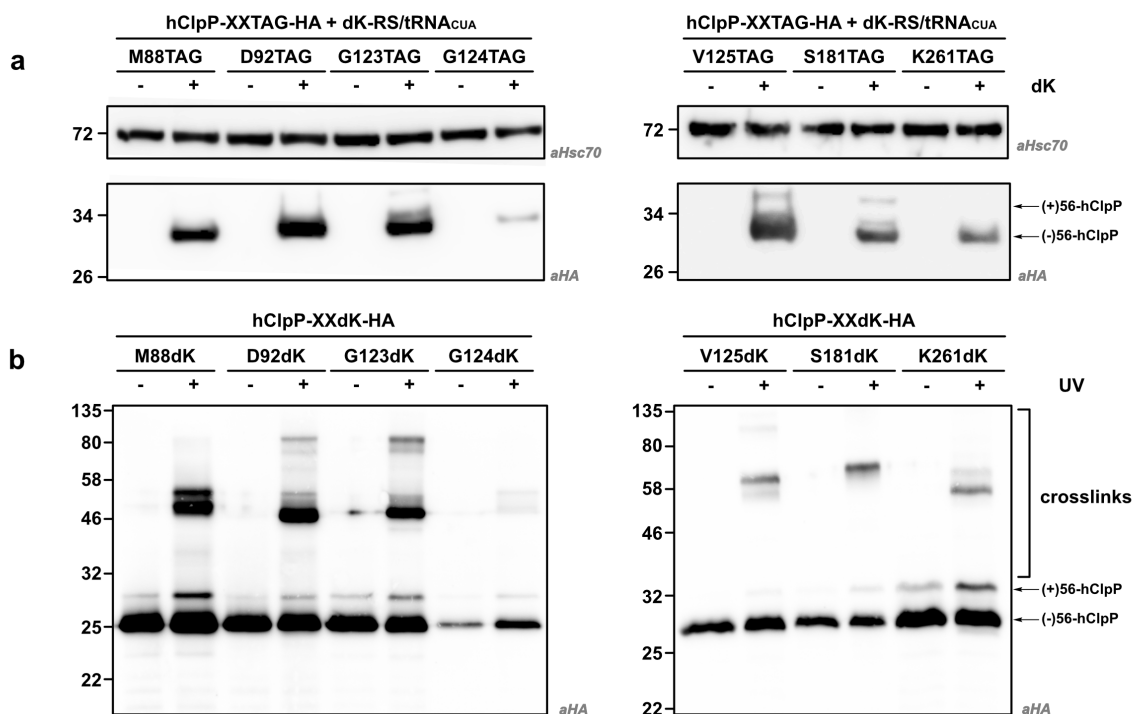


Fig. 2.6 | Amber suppression and photocrosslinking of hClpP-dK variants in HEK293T. (a) Amber suppression of hClpP-dK in HEK293T is only efficient in presence of the UAA. **(b)** Photocrosslinking revealed new, higher running bands, indicating covalent hClpP-adducts.

2.7 Proteomic analysis of photocrosslinked hClpP

Having confirmed that UV-light irradiation led to covalent crosslinks of proteins with hClpP, their identities were analysed next. For this, proteomes between +/- UV treated hClpP-dK variants were compared (n=3). UV-light treated and untreated hClpP-dK variants were enriched via the HA-tag, followed by trypsination and LC-MS/MS analyses. The proteomic profiles of all amber mutants are depicted in Fig. 2.7 as a function of ln(+UV/-UV) of enriched

protein vs. $-\log_{10}(\text{p-value})$ of significance. The detection threshold of signal fidelity was set at $\ln(+UV/-UV) = 1$, i.e. a minimum two-fold enrichment with a minimum significance of $-\log_{10}(\text{p-value}) = 1.3$ ($p < 0.05$). Based on the set criteria, the majority of enriched proteins in the respective quadrant were identified to be of mitochondrial origin, indicating that the approach is generally suitable for protease-substrate trapping (Fig. 2.7). Furthermore, partial overlap with published proteins together with previously not annotated mitochondrial substrates highlights the complementarity of this approach in finding new hClpP interactors (Fig. 2.7). The highest number of crosslinked proteins appeared with hClpP D92dK ($\Sigma = 85$, Table SI 2.2) and K261dK ($\Sigma = 79$, Table SI 2.7), while the remaining mutants contained considerably less enriched proteins (Fig. 2.7, Table SI 2.1 and 2.3-6). Notably, all hClpP-dK mutants in the GGV loop exhibited lowest numbers ($\Sigma = 12-13$), which is in line with observed diminished enzyme activity and probably suboptimal orientation of the photocrosslinker within the barrel.

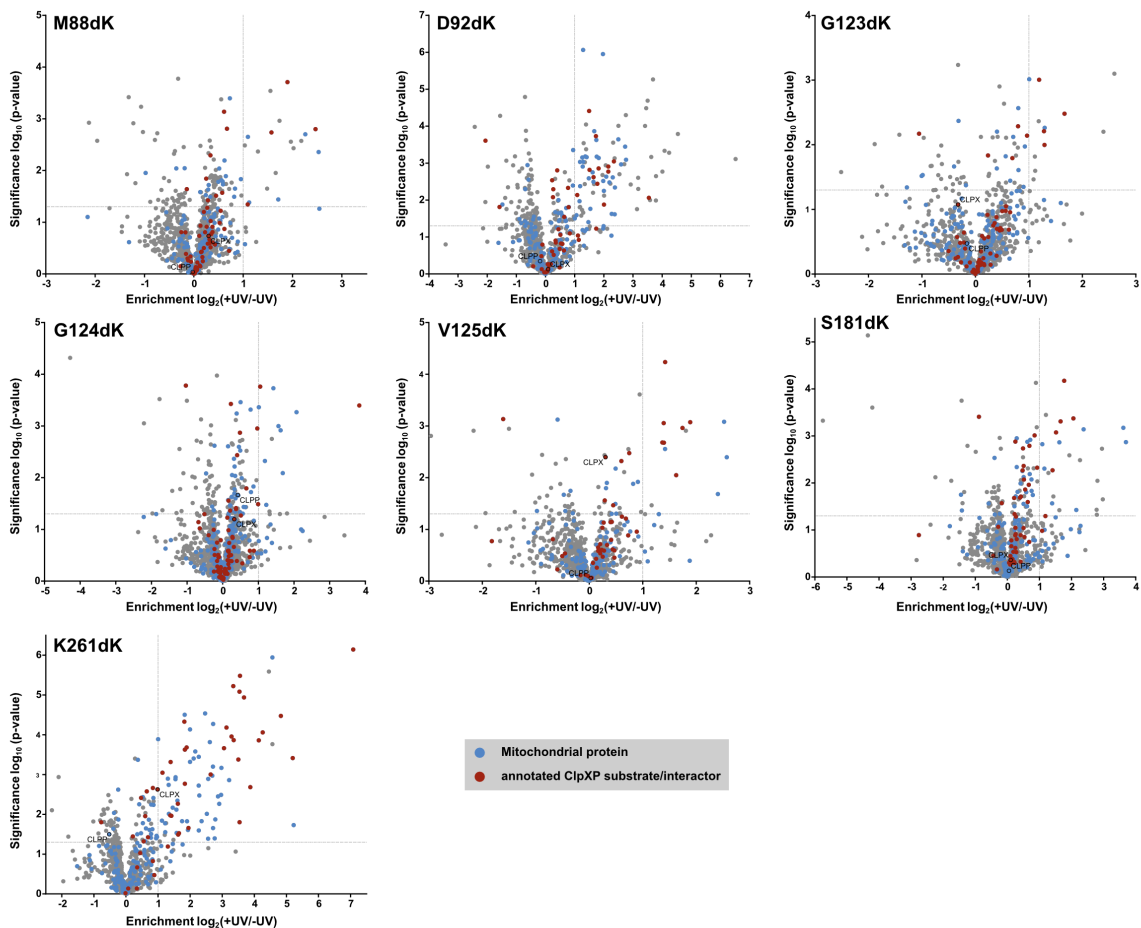


Fig. 2.7 | Volcano plots of all hClpP-dK mutants. All enriched proteins (experiments with $n = 3$, cut-offs see main text) were categorised as mt proteins (blue), already annotated hClpP interactors (red) or neither/nor. *Reprinted with permission of Thomas F. Gronauer, TUM.*

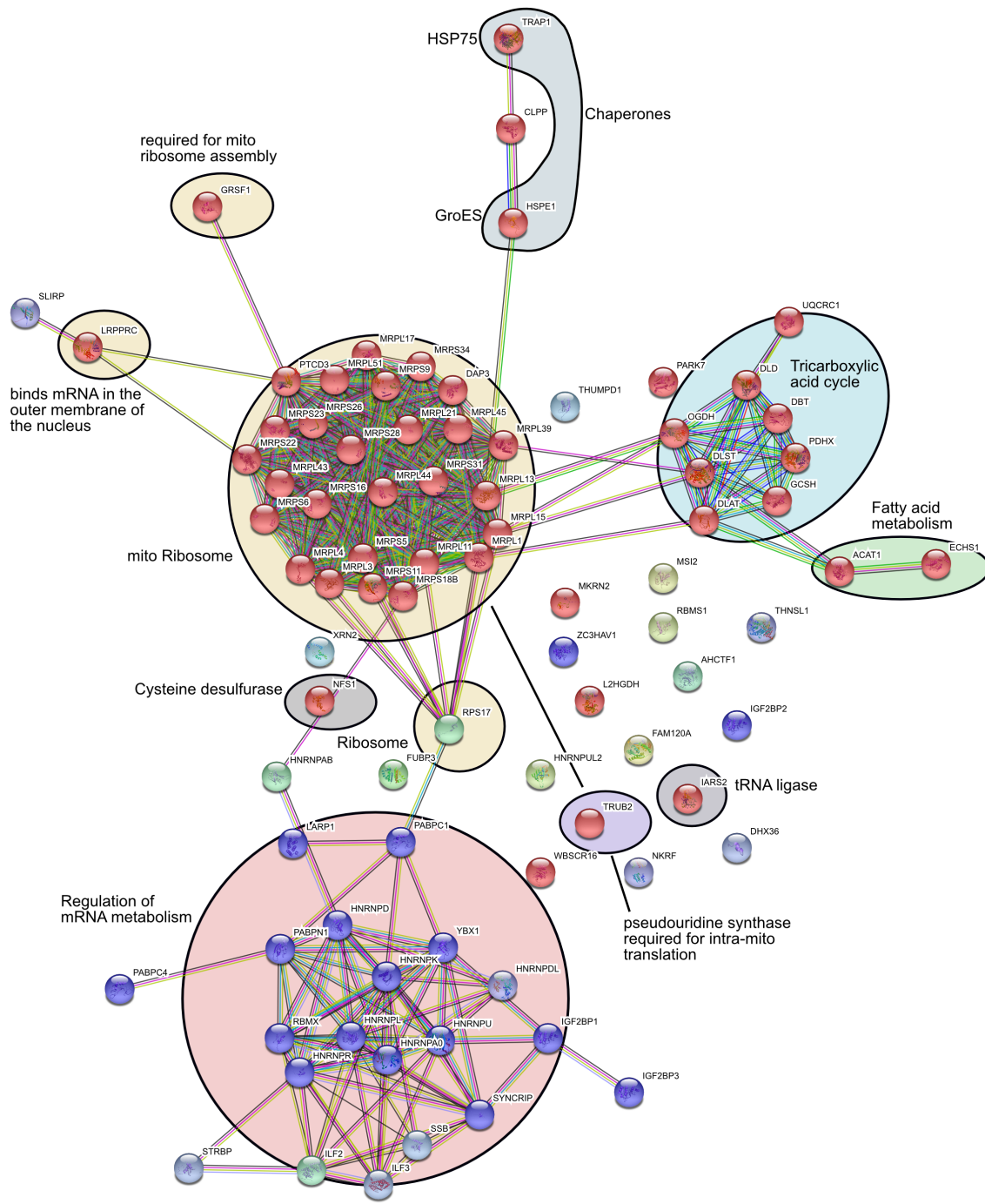
All enriched proteins were clustered according to their function using the STRING v11 database²³⁸ and full lists are available in the Table SI 2.1-7. For a better overview, identified proteins were annotated as either already known or unknown according to current literature (only eukaryotic proteins considered).

The two major pathway clusters occurring in hClpP D92dK are (associated) members of the mt ribosome ($\Sigma=27$ and LRPPRC, SLIRP, GRSF1) and metabolism (TCA cycle: $\Sigma=8$, fatty acid metabolism: $\Sigma=2$), which is in accordance with previous findings conducted in different species and tissues (Fig. 2.8, cf. chapter 2.3.2).^{210, 212, 224-225} The great number of isolated proteins involved in mt protein synthesis reflects heavy participation of hClpP in respiratory homeostasis, as all mt genes encode for proteins of the RCC. These do not only include direct members of the mitoribosome, but also factors important for mt transcription and translation. For instance, G-rich sequence factor 1 (GRSF1) regulates mt protein synthesis via mtRNA processing²³⁹ as well as mt ribosome assembly,²⁴⁰ while LRPPRC/SLIRP (leucine-rich pentatricopeptide repeat containing protein/SRA stem-loop-interacting RNA binding protein) forms a ribonucleoprotein complex to stabilise mt mRNAs via binding on the poly(A) tails.²⁴¹ Interestingly, this pathway was not found when Bio-ID mediated proximity labelling was performed,^{216, 218} highlighting superior spatial precision of this approach when the detecting functionality, i.e. the photocrosslinker is directly attached within or just onto the hClpP barrel. Notable examples of the mt metabolism cluster include PDHX, GCSH, ECHS1 as well as UQCRC1,²²⁴ which was already validated as a substrate in mouse tissues. A small cluster belongs to the chaperone family (HSPE1 and TRAP1), which indicates mutual crosstalk between protein folding and proteolysis. Altogether, within the hClpP-D92dK enrichments, 40 out of the 85 enriched proteins were annotated as known ClpP substrates and interactors (Table SI 2.2). Of the remaining 45 proteins, only a minor portion of nine proteins were found to be of mt origin.

The majority of the 45 proteins belongs to the family of heterogeneous ribonucleoproteins (hnRNPs).²⁴²⁻²⁴³ hnRNPs, which are involved in mRNA processing, are also shuttled between nucleus and cytosol to facilitate the transition from transcription to translation, specifically by functioning as translational suppressors/enhancers and being important parts of the 40S ribosome. Hence, observed crosslinks with this protein class are surprising, since hnRNPs are of non-mt origin and should not be contacted by hClpP. A possible explanation for crosslinks with hClpP-D92dK may be: The hClpP-D92TAG mRNA is continuously translated, even during photocrosslinking experiments. While the peptide chain is elongated and dK is incorporated, the diazirine moiety orients towards the ribosome and hnRNP adaptor proteins and crosslinks to respective proteins upon UV-light irradiation, while the ribosome completes full-length protein hClpP synthesis with a C-terminal HA-tag for subsequent pull-down. Further evidence arises from the crosslink with RPS17, the actual 40S subunit itself. Because the diazirine moiety is not fully embedded within the tetradecamer barrel during nascent hClpP synthesis, crosslinks of exposed dK with unrelated cytosolic proteins might occur and cannot be fully prevented. However, incorporation of dK at position D92 of hClpP seems to be an optimal position for unrelated, non-mt protein crosslinks, as no clusters (of such magnitude) were found in the other mutants (*vide infra*).

Pathway analysis of hClpP-K261dK enrichment revealed a huge intertwined network, indicating a more complex functionality profile beyond sole proteolysis (Fig. 2.9, Table SI 2.7). This is not surprising, considering the positioning of dK at K261 and on the hClpP surface instead of the inner chamber. As ClpX is known to act as a hub for several adaptor proteins,^{211,}

²²⁹ it is conceivable that hClpP may exhibit similar properties, considering its unique C-terminal extension. Initial evidence arises from unique crosslink with MIPEP for instance, a peptidase responsible for hClpP maturation²⁴⁴ and hence by virtue requires binding. Also, newly emerged clusters (mito import, ATP synthase, PDH, tetrahydrofolate interconversion, amino acid metabolism and RCC maturation), which are mainly absent in the remaining mutants, further hint towards hClpP's role as a hub for various mt protein interactions. The decreased mitoribosome cluster size (Fig. 2.9) compared to the hClpP-D92dK cluster also affirms hClpP's tight regulation of mt translation via proteolytic processing thereof. While several pathways are reoccurring (chaperones, TCA cycle, RNA processing) compared to hClpP-D92dK enrichment, some increased in absolute number (fatty acid metabolism, tRNA ligase) hinting to additional regulatory mechanism. In total, 17 novel mt proteins are added to the 57 already annotated hClpP (putative) interactors. However, due to partial overlap with hClpP-D92dK cluster, further experiments are required to differentiate interactors and substrates in detail. The remaining mutants (hClpP-M88dK, -G123dK, -G124dK, -V125dK and -S181dK) in generally reveal similar clusters (TCA cycle, chaperone, mitoribosome, fatty acid metabolism) compared to D92dK and K261dK enrichments, though limited in size (Fig. 2.10, Table SI 2.1 and 2.3-6). Additionally, a subset of the identified mt proteins belong to the mt import proteins. However, to what extent TOMM import proteins, which reside on the outer mt membrane, are related to hClpP remains to be seen. A plausible explanation might be crosslink and pull-down, while hClpP is shuttled into the mt matrix.



Known interactors	Predicted Interactors	Others
From curated database	Gene neighborhood	Text-mining
Experimentally determined	Gene fusions	Co-expression
	Gene co-occurrence	Protein homology

Fig. 2.8 | Network analysis of hClpP-D92dK enriched proteins via String v11 database. Proteins are clustered according to their overall function and pathways and connected (high confidence score) based on literature data indicated in the legend. Red spheres mark mt proteins, blue spheres proteins involved in various mRNA processing. *Reprinted with permission of Thomas F. Gronauer, TUM.*

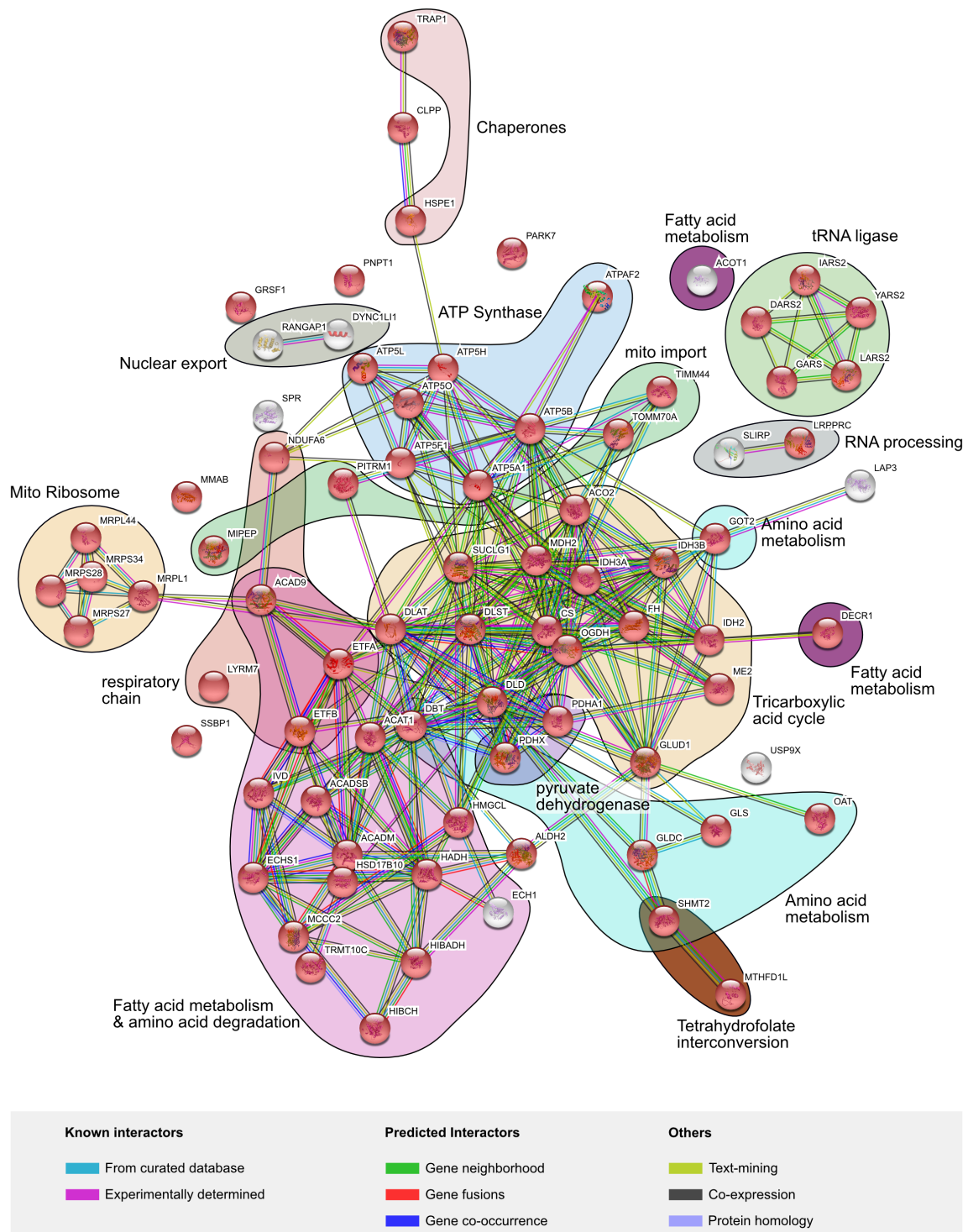


Fig. 2.9 | Network analysis of hClpP-K261dK enriched proteins via String v11 database. Proteins are clustered according to their overall function and pathways and connected (high confidence score) based on literature data indicated in the legend. Red spheres mark mt proteins. *Reprinted with permission of Thomas F. Gronauer, TUM.*

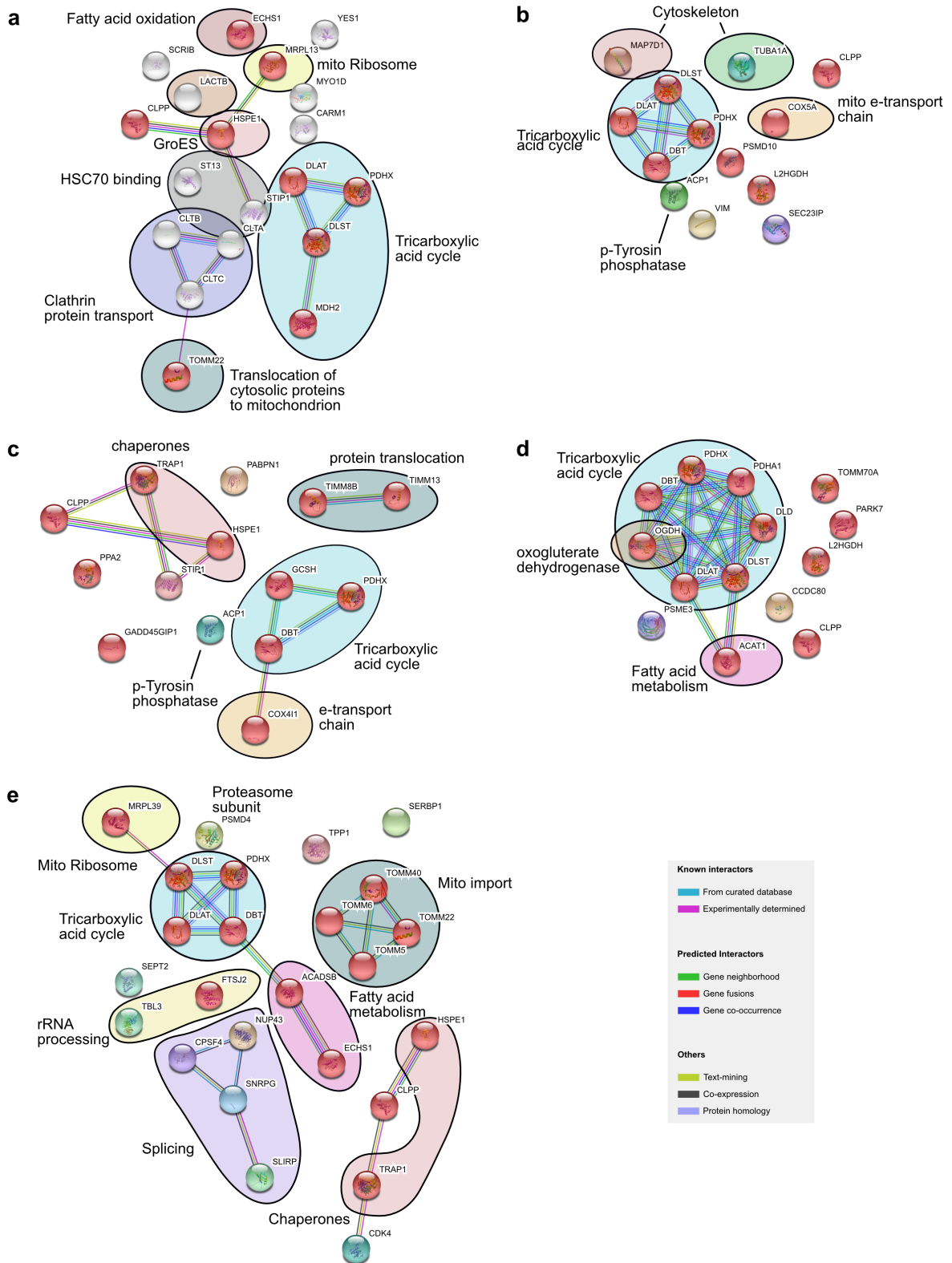


Fig. 2.10 | Network analysis of hClpP a) -M88dK, b) -G123dK, c) -G124dK, d) -V125dK and e) -S181dK enriched proteins via String v11 database. Proteins are clustered according to their overall function and pathways and connected (high confidence score) based on literature data indicated in the legend. Red spheres mark mt proteins. Reprinted with permission of Thomas F. Gronauer, TUM.

Overall, the crosslink and pulldown results emphasize screening of various amber mutants to account for diverse photocrosslinking orientation and thus efficiency. In total, 182 unique proteins were enriched via the photocrosslinking approach, with 119 (65%) being of mt origin (Fig. 2.11a). The remaining 63 proteins are non-mt. However, to note is that a major portion resembles the mRNA metabolism cluster from D92dK enrichment. Of the 119 mt proteins, 86 have been previously associated to ClpP biology, leaving 33 proteins as novel hClpP substrates (Fig. 2.11b). The diversity of protein functionality reflects the central role of hClpP in mitochondria as one of the main mt matrix residing protease (apart from LONP1).²⁰⁶

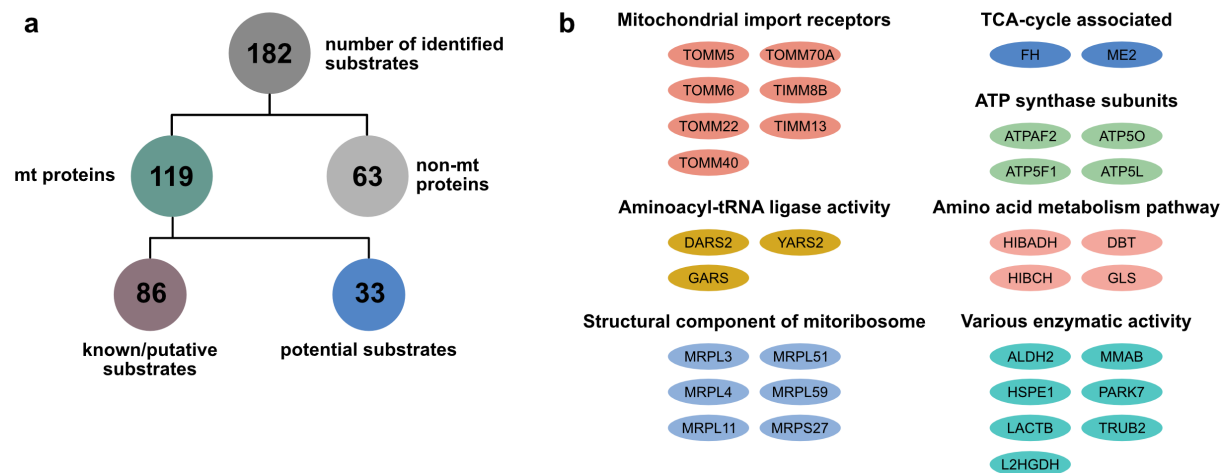


Fig. 2.11 | Overview of hClpP-dK enriched proteins. (a) Numerical summary of the hClpP-dK enrichment approach. Of 119 mt proteins, 33 were novel in regard to hClpP. The protein identities with general function are listed in **(b)**. *Partially adapted with permission of Thomas F. Gronauer, TUM.*

2.8 Target validation of putative ClpP substrates

Validation of potential proteolytic targets was performed via a top-down approach. Target protein levels in HEK293T Clp^{wt} and Clp^{KO} cells were analysed via WB. Ideally, signal intensity is increased in KO cells, if the proteins are proteolytic substrates of hClpP.^{212, 224} DBT and HSPE1 were selected as proof-of principle target. DBT, also known as the lipoamide acyltransferase component of branched-chain α -keto acid dehydrogenase (BCKD) complex, plays a crucial role in Ile, Val and Leu catabolism and was found in 6/7 enrichments. Similarly, the chaperonin HSPE1 was found in 5/7 samples. The high frequencies in both enrichments provide a good indication of actual proteolytic processing. WB analyses showed slightly increased levels of DBT and HSPE1 levels by 29% and 40% in hClp^{KO} cells, respectively (Fig. 2.12a and b). Further indirect evidence as a potential proteolytic target arises from constant HSPE1 levels detected in LonP1^{KO} cells, the other major mt-matrix residing ATP-dependent protease. Nonetheless, redundant and alternative degradation pathways, e.g. via LonP1, activated in absence of hClpP cannot be completely ruled out. This renders target identification via WB difficult and suboptimal, as observed from multiple attempts to reproduce the WB data. Additional methods, which may detect subtle differences at greater

precision and resolution, include MS-based approaches such as selected reaction monitoring.²⁴⁵

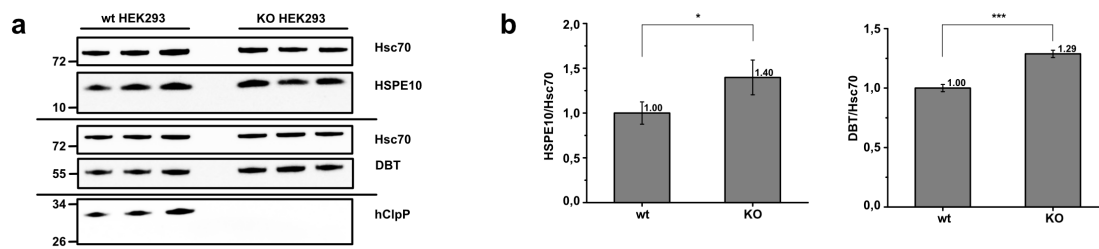


Fig. 2.12 | Semi-quantitative WB analysis of putative hClpP substrate. DBT and HSPE1 levels were analysed in HEK293T Clp^{wt} and Clp^{KO} (a, n = 3) and (b) normalised and compared to the respective the Hsc70 lanes; *, p < 0.05; ***, p < 0.001.

2.9 Rotenone treatment of HEK293T cells expressing hClpP-dK

Reactive oxygen species (ROS) are inevitable by-products of the RCC, especially of complex I, whilst generating a proton gradient under electron and oxygen consumption. During basal conditions oxidatively damaged proteins are readily removed by the protease system,²⁴⁶ while steady ROS exposure and accumulations of oxidatively damaged proteins eventually manifest in pathological phenotypes such as various neurodegenerative diseases.^{208, 247} Interestingly, hClpP was found not only to process oxidatively damaged proteins, but also to act upstream by degrading complex I proteins as a mean to counteract ROS overproduction.²⁴⁸

To gain additional understanding of hClpP's role in oxidative mitostress, HEK293T cells expressing hClpP-dK variants were challenged with rotenone, a complex I inhibitor, which inhibits electron flow ultimately resulting in ROS accumulation.²⁴⁹ Specifically, cells expressing hClpP D92dK and K261dK, which both previously showed highest number of enriched proteins, were incubated for 6 h with rotenone (1 μ M) prior work-up (subsequently denoted as hClpP^{rot}-D92dK and hClpP^{rot}-K261dK). The general proteomics work-flow remained similar as previously performed (Fig. 2.13a and b). Rotenone treated and untreated samples sets were compared for emergence of unique, inhibitor-dependent proteins. Of the total 56 enriched protein in hClpP^{rot}-D92dK, 18 were unique with eleven proteins of mt origin. Likewise, 28 proteins were in total enriched in hClpP^{rot}-K261dK, with ten being unique and three proteins of mt origin (Fig. 2.13c).

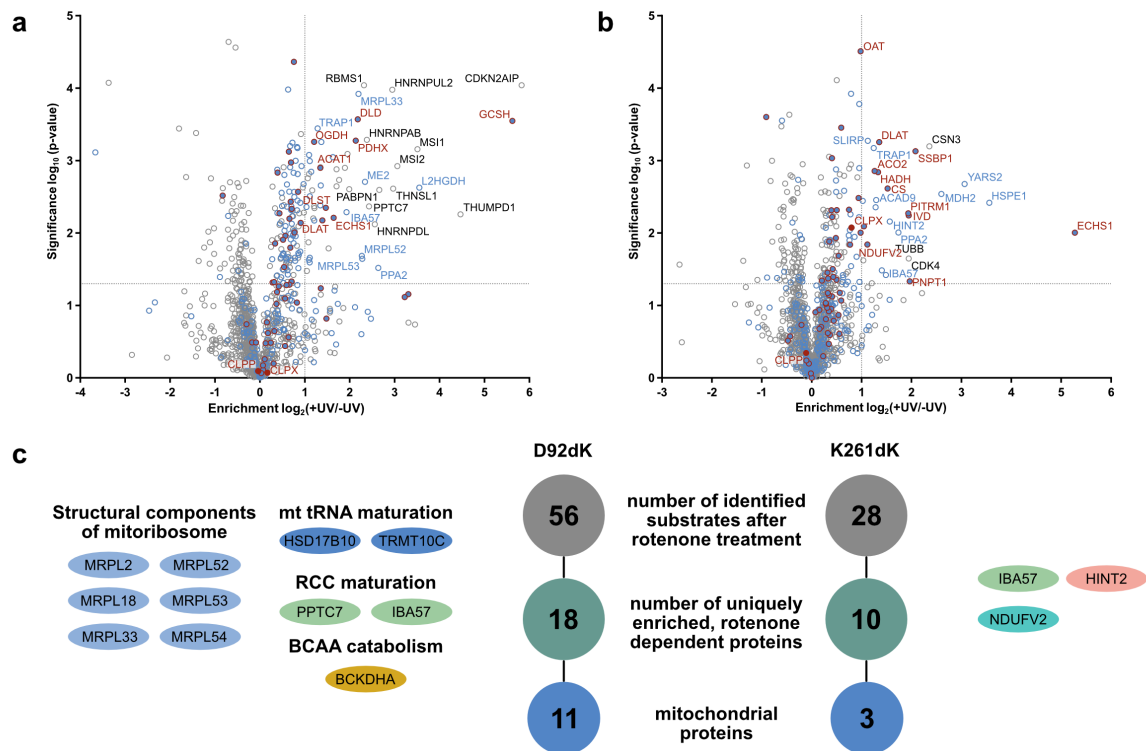


Fig. 2.13 | Volcano plots of rotenone treated HEK293T cells expressing hClpP-D92dK and K261dK. The volcano plots of enriched proteins for cells expressing (a) hClpP^{rot}-D92dK and (b) -K261dK are shown. (c) Comparison with enrichments from untreated samples led to the identification of novel, rotenone-treatment dependent proteins. Reprinted with permission of Thomas F. Gronauer, TUM.

The protein identities in hClpP-D92dK again included structural components of the mitoribosome. HSD17B10 has been reported with multiple functionalities.²⁵⁰ Besides involvement in fatty acid, branched-chain amino acid (i.e. Leu, Ile, Val; BCAA) and steroid metabolism,²⁵¹ HSD17B10 comprises together with TRMT10C, which was also enriched in hClpP^{rot}-D92dK samples, the MRPP1/MRPP2 complex.²⁵² This complex is crucial for mt 5'-tRNA maturation and 3'-CCA addition. Thus, enhanced proteolysis establishes a direct feedback response to reduced RCC protein levels by depletion of mt-tRNA and reduction of mt gene translation, which all encode RCC proteins. Processing of PPTC7 (ref. ²⁵³) and IBA57 (ref. ²⁵⁴), both involved in biosynthesis of Coenzyme Q₁₀ and maturation of Fe/S-proteins respectively, further depletes the number of functional RCC proteins and illustrates the versatile targeting points of hClpP in order to decrease ROS generation. Indirect involvement in ROS reduction is possibly highlighted by proteolytic processing of BCKDHA. This protein is an integral part of the multienzyme BCKD complex (together with DBT) and catalyses the second step of the BCAA catabolism. Specifically, the branched-chain α -keto acids, which are previously formed by transamination of BCAAs, are decarboxylated leading to production of C₅- and C₃-intermediates as well as C₄- and acetyl-CoA synthons.²⁵⁵ Stalling this synthetic pathway may lead to 1) depletion of CoA substrates and 2) in conjunction with processing of HSD17B10, accumulation of BCAAs and branched-chain α -keto acids, which inhibit various enzymes of the TCA cycle. This altogether may reduce hydride production and consequently input into the RCC (Fig. 2.14).

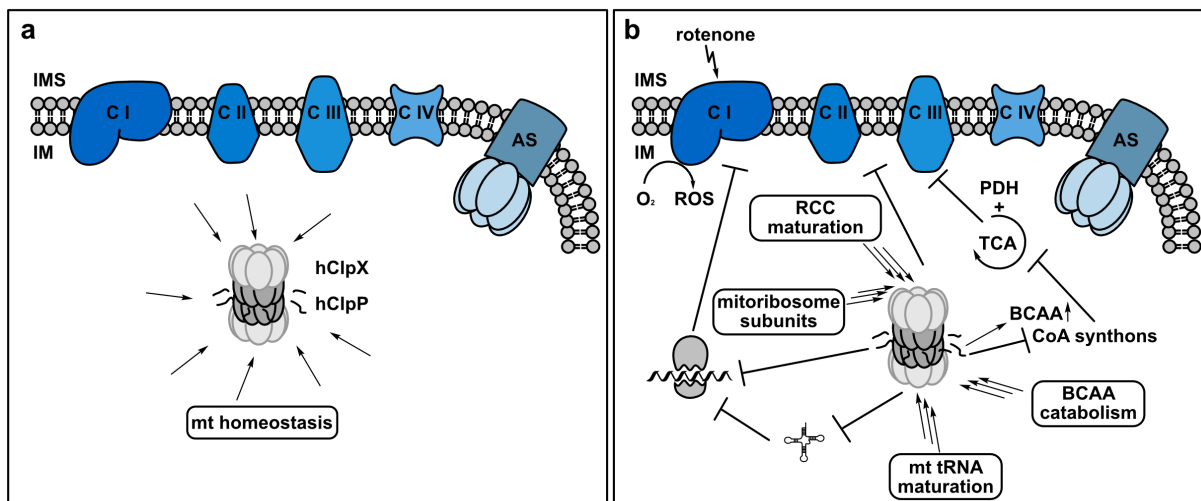


Fig. 2.14 | Potential role of hClpP in minimising ROS production. While under basal condition (a), hClpP fulfils its role in mt homeostasis, under oxidative stress (b), higher turnover of certain functional protein classes aims to reduce of ROS mediated stress at multiple levels. IMS – intermembrane space; IM – inner membrane; C I-IV - complex I-IV of RCC; AS - ATP synthase, ROS – reactive oxygen species, PDH – pyruvate dehydrogenase, TCA – tricarboxylic acid, BCAA – branched chain amino acid.

The number of the hClpP K261dK enriched proteins is greatly reduced. Amongst the three identified proteins, IBA57 is a shared substrate with hClpP D92dK. While NDUFV2 (ref. ^{248, 256}), a subunit of complex I, has been previously found to be a ClpP substrate and possibly requires association with ClpP, the exact role of HINT2 (histidine triad nucleotide-binding protein 2) remains to be determined. Several studies have hinted towards a regulatory role in RCC function.²⁵⁷⁻²⁵⁸

2.10 Conclusion and Outlook

About 2% of the human genome encodes for a single protein class, the proteases, whose designated task is the cleavage of proteins.¹⁹² Given the functional importance of proteases, detailed enzymatic characterisation is of crucial importance, since many pathophysiological phenotypes are associated with aberrant functionalities.

Protease profiling via various chemical and genetic manipulations coupled to high-throughput methods such as proteomics provides a valid workflow to gain detailed pictures of substrate preferences and mode of action.

In this regard, site-specific incorporation of photocrosslinkers by means of GCE represents a suitable route for protease characterisation. The incorporation of photoactivable moieties into proteins may be regarded as a complementary approach to small molecule ABPs/AfBPs and ideally suited for unbiased, on-demand substrate profiling *in vivo*. Photocrosslinker UAA dK was readily incorporated in human caseinolytic peptidase P at seven positions in HEK293T cells, of which six positions aimed at the catalytic chamber, while one position resides outside directed towards the mt matrix. HEK293T cells expressing various hClpP-dK variants were treated with UV-light and subsequently subjected to proteomic analyses for hClpP target identification. The option to readily scan different protein segments highlights the advantages of this approach for capturing substrates from diverse orientations to maximise crosslink

output. Indeed, each hClpP-dK variant showed preferentially enriched proteome clusters, which led to the compiled identification of 119 mt proteins. Partial substrate (86) and cluster overlaps with previous findings emphasises the complementary nature of this approach, while 33 novel proteins were identified as well. The diverse functional scope of the identified protein clusters underlines the important involvement of hClpP in mt homeostasis (Fig. 2.14a).

Furthermore, oxidative challenging of HEK293T cells expressing hClpP-D92dK and hClpP-K261dK with rotenone, a complex I inhibitor, revealed inhibitor specific protein enrichments, which are all directly or indirectly involved in RCC biology. Proteolytic processing of the identified proteins by hClpP appear to overall reduce ROS burden at three levels (Fig. 2.14b), namely by decreasing: 1) *de novo* mt RCC protein biosynthesis via processing of mt ribosome subunits and enzymes, which catalyse mt tRNA maturation; 2) enzymes involved in the RCC maturation and 3) CoA-catabolic pathways and levels to minimise hydride generation in the TCA-cycle. In this regard, additional pull-downs with further hClpP TAG-mutants could possibly provide an improved picture of inhibitor specific enrichments and mechanistic details of hClpP and its specific role during oxidative stress. Overall, the results hint towards a simultaneous, multi-layered regulation of hClpP in order to tackle ROS overproduction. Further downstream biochemical analyses may reveal the exact role and pathways of the identified proteins in oxidative stress response.

Other interesting aspects, which may be readily investigated with this photocrosslinking approach include putative occurrence of tissue specific substrate preferences of hClpP as well as additional substrate identifications, when cells are induced with additional forms of mitostress.

Chapter 3

Development and application of a bifunctional photocrosslinker UAA

3 Development and application of a bifunctional photocrosslinker UAA

3.1 Design of a bifunctional UAA

As depicted in Fig. 2.1, the principal architecture of affinity probes is tripartite. While the recognition motif (small molecule ligand or POI interface) initiates the interactions of interest, a reactive warhead or photocrosslinker that covalently reacts with the POI, preserves the bait-prey contact post workup. Usually, an additional tag facilitates detection, enrichment and analysis. ABP/AfBP molecules commonly contain an alkyne handle for CuAAC reactions with biotin-azide probes.¹⁶² This allows for either WB detection or stringent washing of unspecific binders during affinity enrichment with streptavidin beads prior to proteomic studies. Alternatively, strained alkenes with the option for iEDDA reactions have emerged as an attractive tag option, especially for simultaneous live-cell imaging studies.²⁵⁹ Despite the high functional group density within the probes, natural binding affinities towards the target of choice still needs to be preserved. This is usually achieved via extensive structure-activity optimisations and proper assembly of all three units, typically on a linker scaffold.

Protein-based profiling bears the advantage that the recognition motif is already provided by the POI, while amber suppression ensures the site-specific installation of the crosslinker of choice. Standard detection and enrichment tags²⁶⁰ (e.g. His₆-, HA-, Strep- or FLAG-) need to be additionally cloned into the POI and expressed along amber suppression (cf. chapter 2). However, since tag enrichments relies on weak non-covalent interactions, stringent washing protocols similar to biotin/streptavidin enrichment may lead to loss of information.²⁶⁰⁻²⁶¹ Conversely, mild washing procedures may carry false positive hits to proteomic analysis, which need to be carefully sorted out thereafter. Hence, installation of an additional affinity-tag on the UAA besides a photoaffinity tag would allow for more stringent enrichments procedures (after photocrosslink) beyond what may be achieved with biochemical tags. However, to bear in mind is that the general scope of attaching multiple functional groups on UAAs is limited due to the restricted aaRS pocket space and unlikely acceptance thereof. Thus, only small handles may be added to create bifunctional UAAs, which will be still accepted by an aaRS. The alkyne moiety in this regard bears the dual advantage of being small in size with the option to perform CuAAC chemistry, e.g. with a biotin-azide handle or fluorophore-azide probes, which allows for high-affinity streptavidin enrichments or fluorescence detection.

3.2 Current list of bifunctional UAAs

Generally, only a small number of bifunctional UAAs have been reported thus far (Fig. 3.1). Those include UAAs, which allow for photoinduced switching of protein tertiary structure after intramolecular crosslink (UAA 5)²⁶², capture of erasers of PTM modified proteins after photocrosslinking (UAA 6)²⁶³ or building of protein ternary structures with dual, sequential chemical labelling (UAA 7)²⁶⁴.

UAAs 8-12 were designed to facilitate post-photocrosslink analyses of captured interactors. The backbone of Se-UAAs 8 (ref. ²⁶⁵) and 9 (ref. ²⁰⁵) may be oxidatively cleaved after crosslink formation, leaving a unique MS-identifiable label on the prey protein. The newly emerged functional groups on the MS-identifiable label additionally offer a further layer of sample analysis and control. For instance, the liberated Se-OH moiety after UAA 8 cleavage allows for selective Se-OH (UAA 8) reaction with alkyne 3,5-dimethoxyaniline and subsequent CuAAC reactions with detection probes bearing the respective azide handle.²⁶⁵ Similarly, the emerging acrylamide group upon UAA 9 cleavage enables photoclick chemistry with tetrazole-bearing fluorogenic probes, which may be detected via in-gel fluorescence analysis. Both UAAs have been mainly utilised in proof-of-concept substrate and surface interaction profiling of bacterial chaperone HdeA.

The other subgroup comprising UAAs 10-12 aims to facilitate proteome analysis via the additional alkyne tag, which may be used for streptavidin enrichment after CuAAC reaction with a biotin-azide probe. Benzophenone based UAA 10 (ref. ²⁶⁶) has been utilised for characterisation of yeast transcriptional Gal4-Gal80 interactions. The advantage of this UAA was shown in successful pull-down experiments after biotin-azide CuAAC reaction, as evident in WB analysis. Diazirine-alkyne UAA 11 was introduced for proteome-wide CuAAC labelling with 5-Carboxytetramethylrhodamine-azide (TAMRA-azide) in *E. coli* lysates²⁶⁷ as well as attempted protein/RNA binding profiling.²⁶⁸ Specifically, dK and UAA 11 were incorporated in RNA binding human iron regulatory protein 1 (IRP1). While photocrosslinking with model RNA substrates was successful with dK bearing IRP1, detection of crosslink bands was absent when UAA 11 was placed at the same position. Failure of crosslinks is possibly due to the shifted diazirine positioning by one methylene unit compared to dK, which results in unfavourable orientation and distancing of this moiety towards the target area. Se-based diazirine alkyne UAA 12 (ref. ²⁰⁴) acts in a similar fashion as described for the previous Se-UAAs and was used for comparative proteome profiling to identify new DegP targets after biotin-azide CuAAC reaction and streptavidin enrichment.

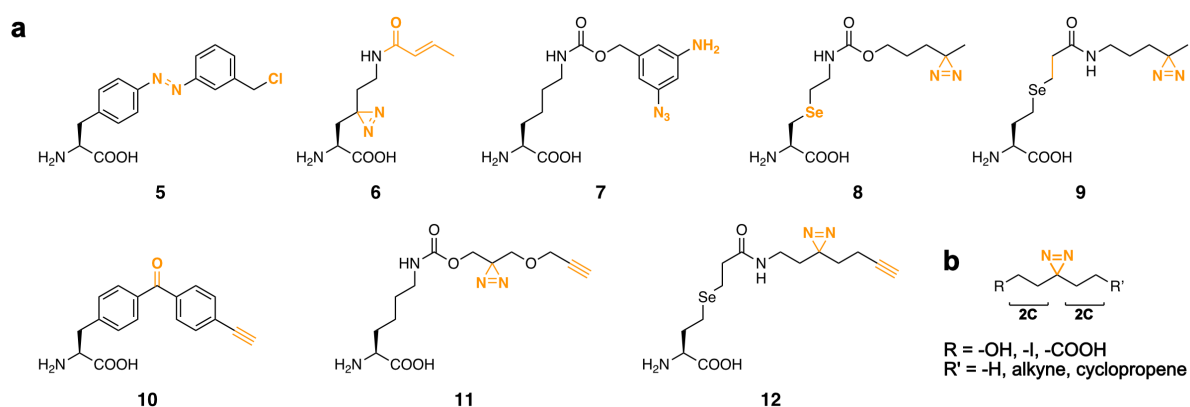


Fig. 3.1 | Bifunctional UAAs. Depicted in **a** are all GCE accessible bifunctional UAAs. The functional groups of each UAA are highlighted in orange. The minimal photocrosslinker scaffold in **b** represents the starting point for design of a novel, bifunctional UAA.

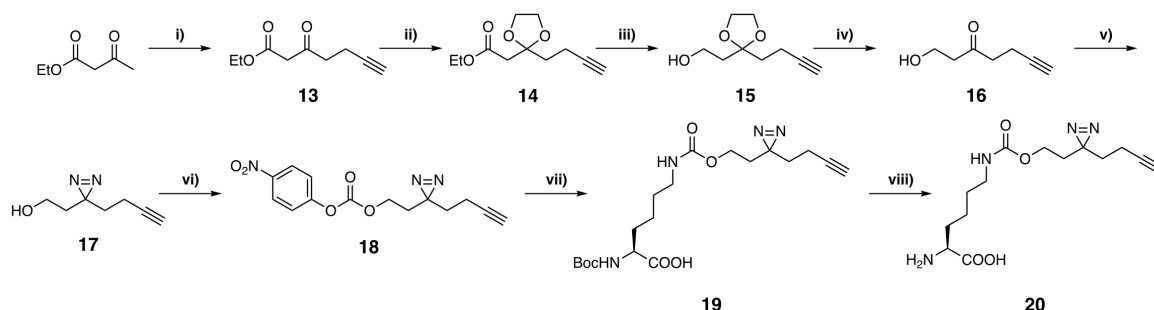
Despite the functional group variety of available bifunctional photocrosslinker UAAs, each of them bears disadvantages including knowledge to perform complex, multi-step synthesis steps (8, 9 or 12), limited target systems or lack of mammalian applications (UAA 10) and insufficient crosslinking performance (UAA 11).

3.3 Aim of this work

Employing bifunctional UAAs would greatly extend the possibilities for detailed protein studies. This is readily exemplified by UAAs bearing a photocrosslinker and an alkyne tag for detection and improved enrichment purposes of crosslinked protein complexes. However, bifunctional UAAs are accompanied with some disadvantages as mentioned in the previous section. For instance, the different crosslink performances between dK and UAA 11, despite its option for high-affinity enrichment highlight the importance of equal photocrosslinking positioning for comparative photocrosslinking studies. Thus, inspired by the alkyne-diazirine scaffold as reported in ref. ^{259, 269} (Fig. 3.1b) and the equal positioning of the diazirine moiety compared to dK, the novel bifunctional UAA 20 (subsequently denoted as dKA, Scheme 3.1 in the following chapter) was designed and synthesised. The additional propargyl extension was proven sufficiently small to be accepted by a PyIRS. The crosslinking performance was compared to dK in various model PPIs and subsequently proven for the general possibility to perform post-crosslink CuAAC reactions in bacteria as well as mammalian cells.

3.4 Synthesis and incorporation of the novel bifunctional dKA

The synthesis of the main scaffold 17 (Scheme 3.1) was performed as previously published.²⁶⁹ The last three steps concerned the attachment of 17 onto the lysine backbone. This included the carbonate functionalisation of the hydroxy group to yield intermediate 18, followed by carbamate functionalisation of the N ϵ -amino group of N α -Boc protected dKA (19) and acidic Boc deprotection to obtain the final dKA (20).



Scheme 3.1 | Synthesis route towards UAA 20. i) LDA (2.0 eq.), propargyl bromide (1.0 eq.), -16°C, 2h, 36%; ii) ethylene glycol (2.0 eq.), pTsOH (0.1 eq.), reflux, 6h, 46%; iii) LiAlH₄, 0°C, 3h, 67%; iv) pTsOH (0.25 eq.), r.t., 64%; v) a - 7M NH₃ in MeOH, hydroxylamine-O-sulfonic acid (1.1 eq.), 0°C->r.t., o.n., b - Et₃N, I₂, 0°C, 1h, 34%; vi) pNO₂-phenyl chloroformate (1.2 eq.), pyridine (1.2 eq.), 0°C-> r.t., o.n., 97%; vii) Boc-Lysine-OH (1.2 eq.), Et₃N (2.0 eq.), r.t., o.n., 87%; viii) TFA, 3h, r.t., 75%.

The synthetic effort however may be greatly shortened since intermediates 16 (CAS: 1450754-40-1) and 17 (CAS: 1450754-41-2) are nowadays commercially available.

With the new dKA UAA in hand, suitable PyIRSs variants for site-specific incorporation were screened next. As already observed in the previous chapter, PyIRSs variants may tolerate different substrates to a certain degree (polyspecificity). Screening of in-house available PyIRS variants for selective incorporation at position N149TAG of sfGFP revealed some selective variants (SI Fig. 3.1, subsequently denoted as dKA-RS). Expression and purification of dKA containing sfGFP with dKA-RS bearing the mutations Y271M, L274G and C313A (*Mb* numbering) provided the expected mass (Fig. 3.2a and b), confirming successful incorporation and integrity of the respective UAA with its functional groups during expression conditions. Next, the general workflow of ‘crosslink, then CuAAC reaction’ was tested for proof-of-principle PPI systems.

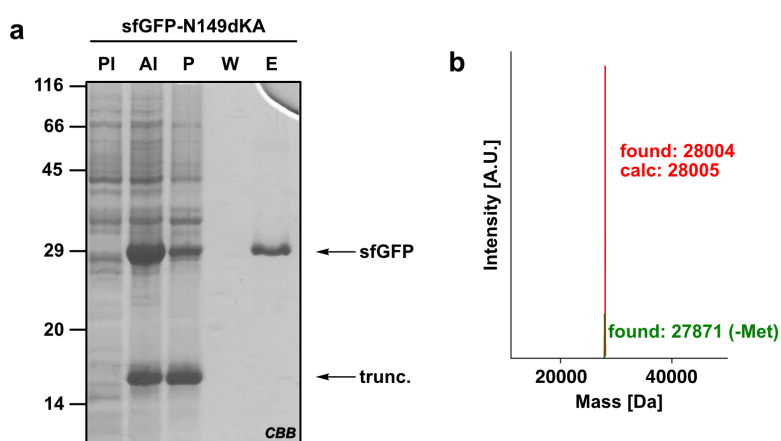


Fig. 3.2 | Expression, purification and MS analysis of sfGFP-N149dKA. (a) Arabinose induction lead to sfGFP-N149dKA expression with *Mb* dKA-RS/tRNA_{CUA}. Pure protein was obtained after His-purification. (b) MS analysis of the purified protein confirmed the incorporation of dKA. PI - preinduction, AI - after induction, P - pellet after lysis/cell debris, W - wash, E - elution, trunc. - truncated.

3.5 Application of the proposed workflow at proof-of-principle PPIs

Three different PPIs of various interaction distances (~ 6 -14 Å, Fig. SI 3.2) were chosen as proof-of-principle targets: sfGFP and GST dimerization and the Rab1b/DrrA interaction. Despite the exact diazirine positioning of dKA compared to dK, altered photocrosslinking performance may occur due to the additional propargyl group. Thus, all crosslinks with dKA were benchmarked to dK in order to estimate the influence of the alkyne linker on the interaction surface. Additionally, Bock was incorporated to serve as a UV-independent non-crosslink control (Fig. 3.3a).

As depicted in Fig. 3.3b, position N149 points towards the GFP dimerization interface. Previous successful capture of (sfGFP)₂ via installation of a chemical crosslinker indicates the presence of suitable nucleophile(s) in proximity to N149.¹⁶⁰ sfGFP-N149TAG variants bearing a C-terminal His₆-tag to detect full-length protein were expressed in presence of Bock, dK or dKA and their respective PyIRSs in *E. coli*. After induction and o.n. expression, the number of cells were adjusted according to their OD₆₀₀ values, followed by life-cell photocrosslinking, lysis and

WB analysis. As illustrated in Fig. 3.3c, dimeric crosslink bands only occurred in presence of a photocrosslinker and UV-light. The dimer signal intensity of sfGFP-N149dKA is slightly lower compared to sfGFP-N149dK suggesting partial influence of the propargyl group in photocrosslinker orientation and crosslinking efficiency. The overall low crosslink signal compared to baseline intensity is likely due to the weak dimerization tendency of GFP ($K_d = 0.11$ mM, ref. ²⁷⁰) as well as competitive quenching of the activated carbene, presumably by surrounding water molecules. Also, in all UV-exposed samples minor bands below the MW of the crosslinked dimer occurred. This may be sfGFP-dependent, unspecific crosslinks as such bands are not observed in the other PPI systems (*vide infra*).

Having confirmed the diazirine functionality for capture of sfGFP dimer, the functionality of the alkyne group was tested next via CuAAC reaction with TAMRA-azide. Gel fluorescence analysis showed only dKA dependent protein signals, proving that only the presence of the alkyne group results in fluorophore conjugation (Fig. 3.3d). Additionally, the dimer crosslink band was only detected, when sfGFP-N149dKA was treated with UV light.

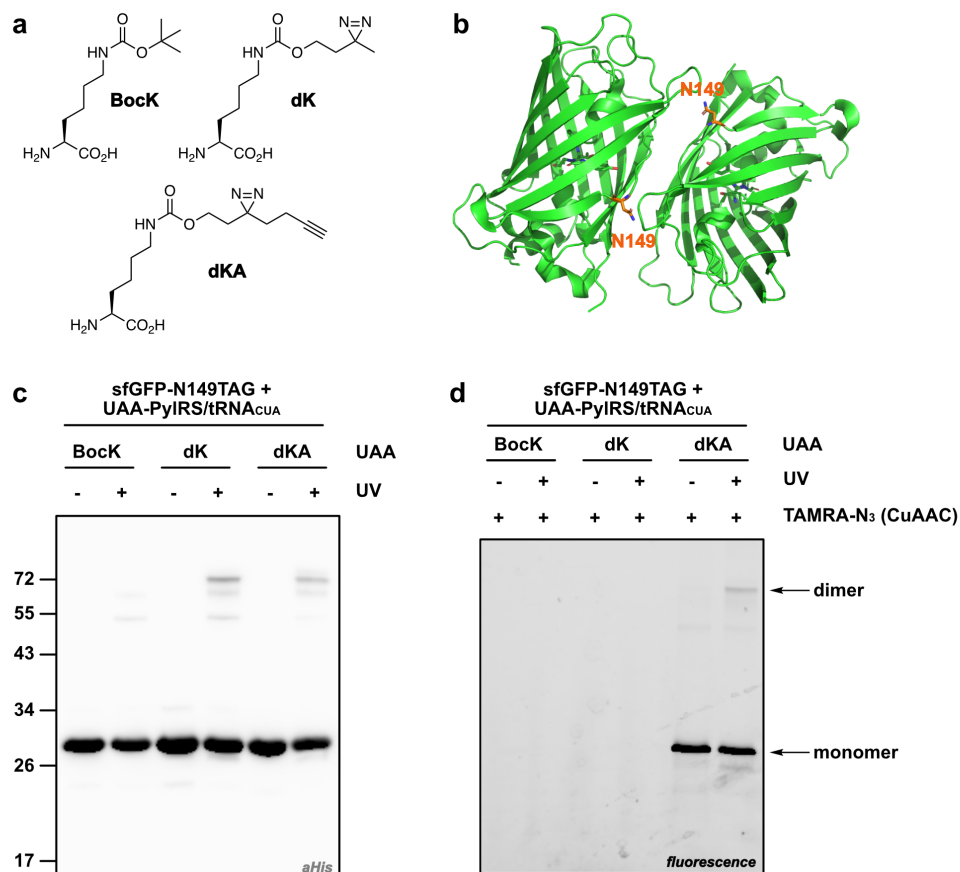


Fig. 3.3 | Photocrosslink, then click approach in sfGFP. (a) List of UAAs, of which all subsequent experiments were performed with. (b) Crystal structure of the GFP dimer (PDB: 5NHN). Position N149 of GFP points towards the interaction interface, indicating putative capture of the dimer via photocrosslinking. (c) WB and (d) gel fluorescence analysis after UV-dependent treatment of BocK, dK and dKA containing sfGFP. Reagents and procedures for CuAAC reactions are listed in 5.4.2.2 in detail. All proteins were expressed with their respective *Mb* PyIRS.

Next, this 'crosslink, then click approach' was tested on GST, which is known to dimerise as well. GST is ubiquitous amongst all living species due to its crucial role as part of the general,

cellular detoxification system.²⁷¹ Position F52 was chosen to be substituted with BocK, dK and dKA, since previous photocrosslinking experiments with dK proved successful for capturing the dimeric complex (Fig. 3.4a).²³³ Experiments were performed as described above with a pPylt_GST-F52TAG construct bearing a C-terminal His₆-tag. Again, crosslink bands only occurred in presence of the diazirine moieties (Fig. 3.4b). However, band intensity of GST-F52dK dimer is significantly weaker compared to published results. This is possibly a reflection of altered cellular protein concentrations and applied UV light intensities (not reported in ref. ²³³). As observed above, GST-F52dKA dimer intensity is slightly decreased compared to GST-F52dK dimer. However, selective CuAAC reaction with TAMRA-azide and clean signal output of monomer and dimer highlight the advantages of bifunctionalized dKA (Fig. 3.4c).

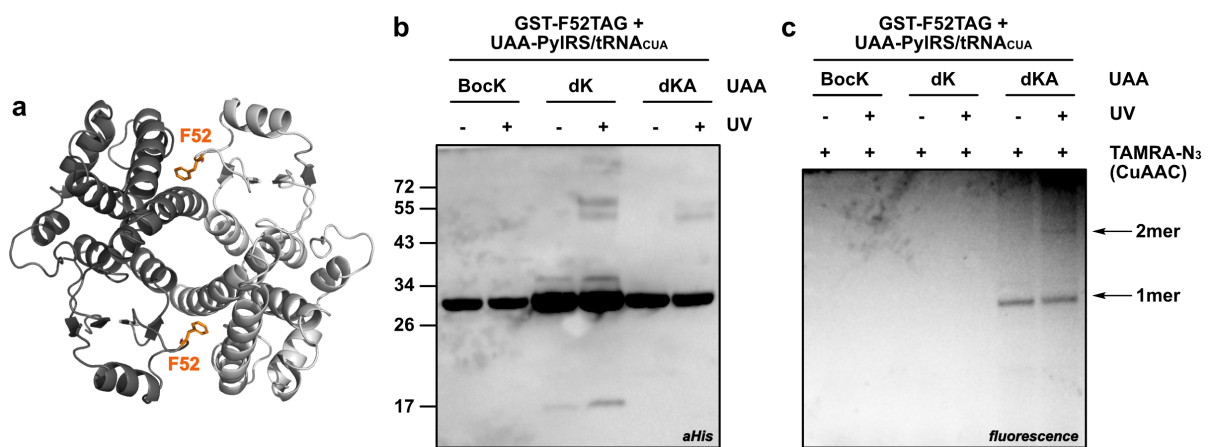


Fig. 3.4 | Photocrosslink, then click approach in GST. (a) Crystal structure of the GST dimer (PDB: 1GSD). Position F52 of GST points towards the interaction interface, indicating putative capture of the dimer via photocrosslinking. (b) WB and (c) gel fluorescence analysis after UV-dependent treatment of various UAA suppressed GST. Unspecific bands in GST-F52dK additionally occurred during high exposure of the WB to detect crosslinks in GST-F52dKA samples. All proteins were expressed with their respective *Mb* PylRS.

Lastly, this approach was tested on the Rab1b/DrrA PPI. Rab1b is a small GTPase, which is involved in the regulation of intracellular vesicle trafficking between ER and Golgi apparatus. To proliferate, pathogens such as *Legionella pneumophila* bypass this pathway by inactivating Rab1b via various effector proteins.²⁷² Based on the crystal structure of Rab with its effector proteins DrrA (16-352, ATase domain, Fig. 3.5d), several chemical crosslinkers have been previously installed in the Rab protein for complex stabilisation. Especially position R69 of Rab1b provided a suitable site for UAA incorporation due to its potential, multiple interactions with residues of DrrA.²⁷² Based on these results, *E. coli* DH10 β were co-transformed with a duet vector bearing both proteins, Rab1bQ67A-R69TAG-His₆ and StreptII-DrrA₁₆₋₃₅₂ and the respective PylRS variants for incorporation of BocK, dK and dKA (Fig. 3.5). As depicted in (Fig. 3.5a and b), the WB results reflect the previous observations of diazirine-dependent and UV-dependent complex formation. Again, the alkyne linker seems to minimally perturb binding architecture, but its utility is compensated by selective CuAAC reaction with TAMRA-azide (Fig. 3.5c).

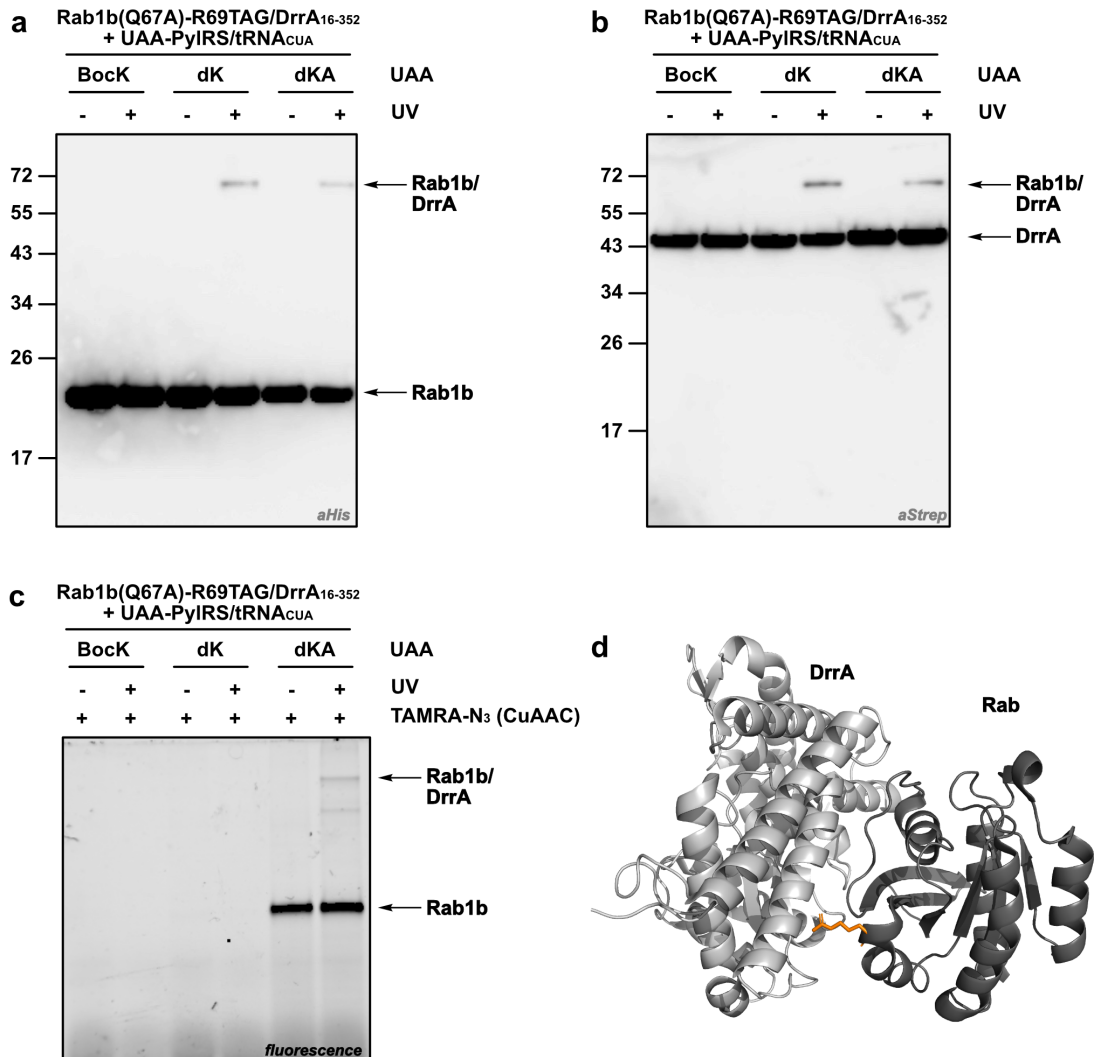


Fig. 3.5 | Photocrosslink, then click approach in Rab1b/DrrA. (a) and (b) WB and (c) gel fluorescence analysis after UV-dependent treatment of various UAA suppressed Rab1b. All proteins were expressed with their respective aaRS. Q67A was introduced to reduce hydrolytic activity of Rab1b. All proteins were expressed with their respective *Mb* PylRS (BocK - wtRS, dK - dK-RS, dKA - Y271G, C313V). (d) Crystal structure of the Rab/DrrA complex (PDB: 6XY5). Position R69 of Rab points towards the DrrA interaction interface, indicating putative capture of the complex via photocrosslinking

3.6 Incorporation of dKA in mammalian cells

Accessibility of the ‘crosslink, then click’ approach in mammalian cells would greatly expand the applicability of PPIs to be studied. Transfer of identified *Mb* PylRSs mutations into the appropriate plasmid system (cf. chapter 2) is readily accessible, therefore facilitating application of amber suppression in HEK293T. Incorporation efficiency of dKA was tested with sfGFP-N149TAG serving as a model protein system. As shown in Fig 3.6a, only in presence of dKA, sfGFP-N149TAG was expressed. Titration of increasing dKA concentrations suggests fluorescence signal saturation at around 1-2 mM. WB analysis confirms the identity of the protein, also highlighting the robustness of amber suppression as no full-length protein was detected without dKA (Fig. 3.6b). Purification and LC-MS analysis revealing the expected protein mass furthermore confirms the general integrity of both, diazirine and alkyne functionalities during the entire mammalian expression and purification process (Fig. 3.6c).

Overall, this generally opens the opportunity to perform similar ‘crosslink, then click’ studies as described above (Fig. 3.6d).

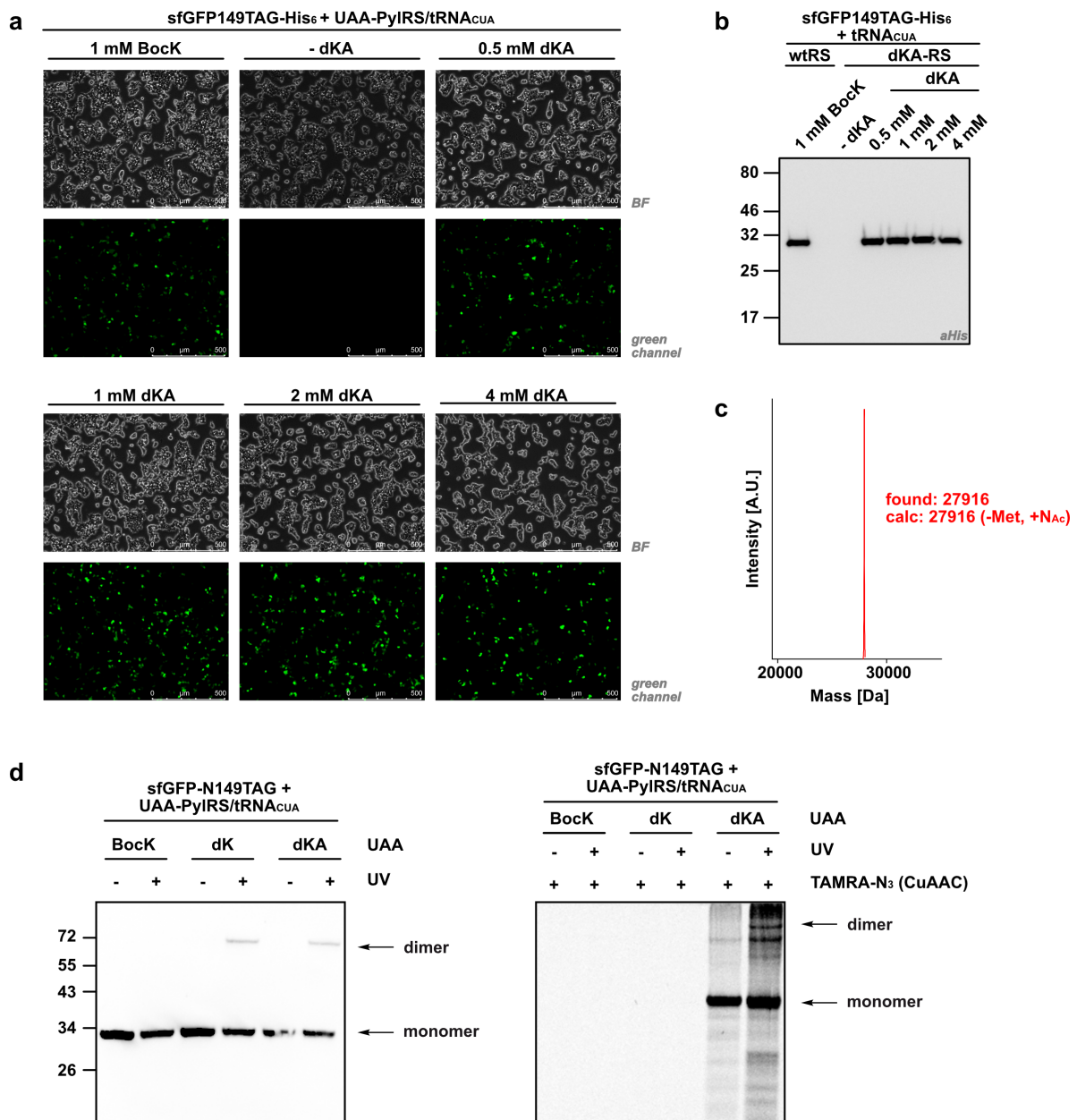


Fig. 3.6 | Incorporation of dKA in sfGFP-N149TAG in HEK293T. (a) sfGFP fluorescence only occur in presence of dKA. Screening of different dKA concentrations suggest protein expression saturation at 1-2 mM. **(b)** WB analysis of sfGFP-N149BocK or -N149dKA confirm the protein identities of **(a)**. Total absence of signal when sfGFP is expressed without UAA highlight the robust amber suppression approach. **(c)** MS analysis of purified sfGFP confirms the integrity of both functional groups of dKA. **(d)** The overall approach is similar to *E. coli* experiments and performed as described above. Selective TAMRA-labelling were only observed with sfGFP bearing dKA at position 149. Higher background is due to background amber suppression of endogenous TAG codons within the HEK293T genome. *Mm* PylRS with the following mutations was used for expressions with dKA: Y306M, L309A, C348A. BF - bright field.

3.7 Conclusion and Outlook

Altogether, the functionalities of the dKA in a 'crosslink, then click approach' has proven to be robust for a selection of PPIs. Both photocrosslinkers (dK and dKA) are able to capture PPIs of various distances (6-14 Å). The slightly reduced crosslink performance of dKA compared to dK is compensated by its alkyne functionality and the option for CuAAC chemistry related downstream applications and analyses such as fluorescence labelling or biotin enrichment. UAA dKA has been shown to be stable and applicable in bacterial and mammalian contexts, allowing for various PPI studies in appropriate contexts. Both UAAs or precursor thereof are readily accessible or nowadays commercially available, allowing for rapid set-up and screening of appropriate conditions without the need of advanced chemical synthesis expertise. A general workflow may involve investigation and benchmarking of PPI interfaces with dK and utilization of dKA, when the additional option of a click handle is considered advantageous.

Chapter 4

**Developing a general approach to access proteins
bearing site-specific Lysine acylations**

4 Developing a general approach to access proteins bearing site-specific Lysine acylations

4.1 Utilising traceless bioorthogonal reactions to generate modified proteins with native linkages

A further advantage of GCE besides the installation of orthogonal and inducible platforms for targeted protein studies (cf. chapter 2 and 3) lies in the homogeneous production of selected proteoforms. This allows for biophysical and -chemical characterisations under *in vitro* and even native conditions at great detail. Especially small PTM-proteins as a major subgroup are readily projected by UAAs and GCE (cf. Fig. 1.7, Fig. 4.1a).

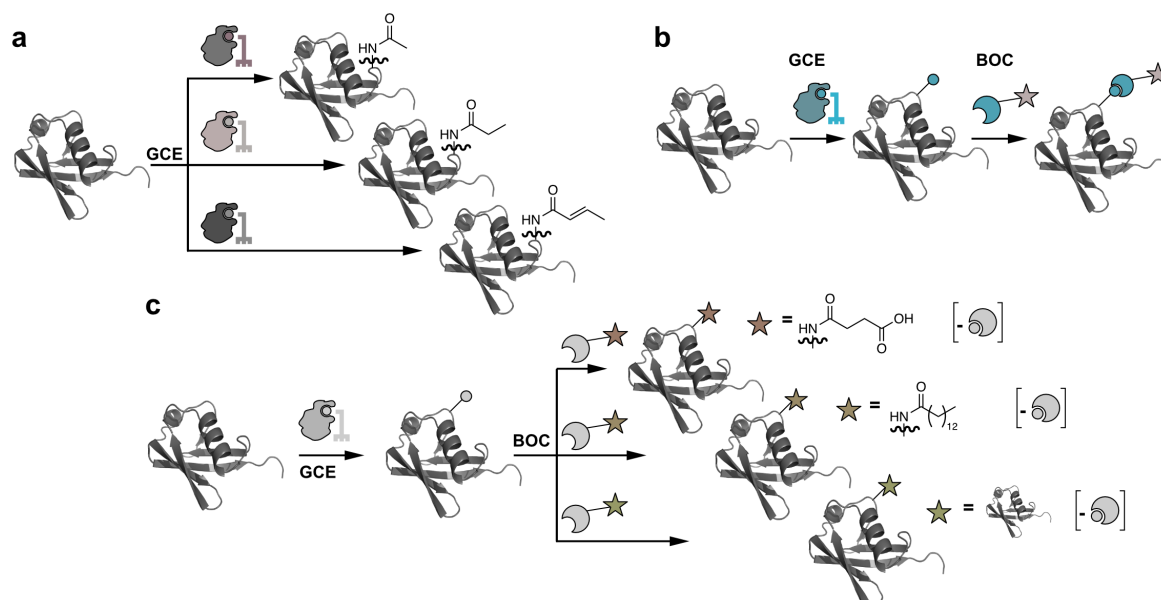


Fig. 4.1 | Routes towards site-specifically modified protein. Amber suppression allows site-specific installation UAAs bearing various handles. Obtaining natural proteoforms (a) may be achieved via direct installation of UAAs with small PTM groups, as exemplified with acetylation, propionylation and crotonylation. Alternatively, performing bioorthogonal chemistries (b) may yield larger modifications, however with unnatural linkages. Thus, a general, modular approach (c) may contain introduction of universal UAA precursor, at which bioorthogonal chemistries with diverse substrates and under excision of the leading groups may be performed to yield native PTMs, such as succinylation, myristoylation and Ubiquitination. BOC - bioorthogonal chemistry.

This represents a powerful bypass route, especially when the identities of the PTM generating enzymes and reaction cascades are unknown. Even combination with a crosslinking moiety provides an additional promising route towards deconvolution of the enzymatic cascades (e.g. cf. Fig. 3.1a, UAA 6). However, finite aaRS pocket size overall limits UAA side-chain diversity and thus direct access to larger, native protein PTMs. This has been partially circumvented through bioorthogonal and/or chemoenzymatic attachment of larger PTMs or even protein fragments onto POIs (cf. Fig. 1.8, Fig. 4.1b). The resulting unnatural motifs can be advantageous due to chemical and enzymatic stability, thereby providing constant signal output during analyses. However, additional control experiments need to ensure that the

newly created connections minimally perturb protein performances compared to wt constitutions and that any significant results are not solely derived therefrom. Nonetheless, deviations from 'wt results' are expected and inevitable (e.g. cf. ref. ¹⁸²). Thus, devising strategies, which enable the chemical installation of modifications post GCE with native connections would represent an ideal solution towards generation of natural proteoforms of any size (Fig. 4.1c).

4.2 Bioorthogonal amide bond reactions

Despite great sequence and functional heterogeneity, the backbone of all proteins is connected via the same bond type. Amide bonds are not only a frequently occurring motif in proteins and natural products but also a heavily favoured element in organic and medicinal chemistry due to their high stability and polarity.²⁷³⁻²⁷⁶ The common (bio)synthetic strategy follows a condensation pathway,^{273-274, 276} at which the electrophilic carbonyl centre of one building block is attacked by the amine of the other building block to yield the amide bond under loss of water. However, overcoming the high reaction barrier requires activation of either functional group. Usually, increasing the electrophilicity of the carboxyl group is more amenable than altering the nucleophilicity of amines. Indeed, a sophisticated ribosomal protein synthesis machinery has been evolved only for the condensation of amino acids via carboxyl group activation and nucleophilic attack of a preceding amine of the amino acid. While the manifold cofactors of biological peptide bond synthesis are required to ensure proper, genome dependent assembly of amino acid building blocks, reactions in organic chemistry may be performed under more controlled conditions. Nonetheless, the standard mode of carboxy group activation remains similar.

Amide bonds additionally occur as lysine side-chain PTMs on proteins (Fig. 4.2). The great substituent variety reflects heavy participation in critical cellular functions such as dynamic regulation of gene and chromatin structure, enzyme activities, protein interactions and localisations.²⁷⁷⁻²⁷⁸ Whole enzyme sets are only concerned with dynamic regulation of acyl transfer (writers), signal recognition (readers) and PTM removal (erasers - KDACs) of lysine acylated proteins depending on frequently changing external cues. Aberrant functionalities at any point of this cycle are often associated with various diseases.²⁷⁷⁻²⁷⁸

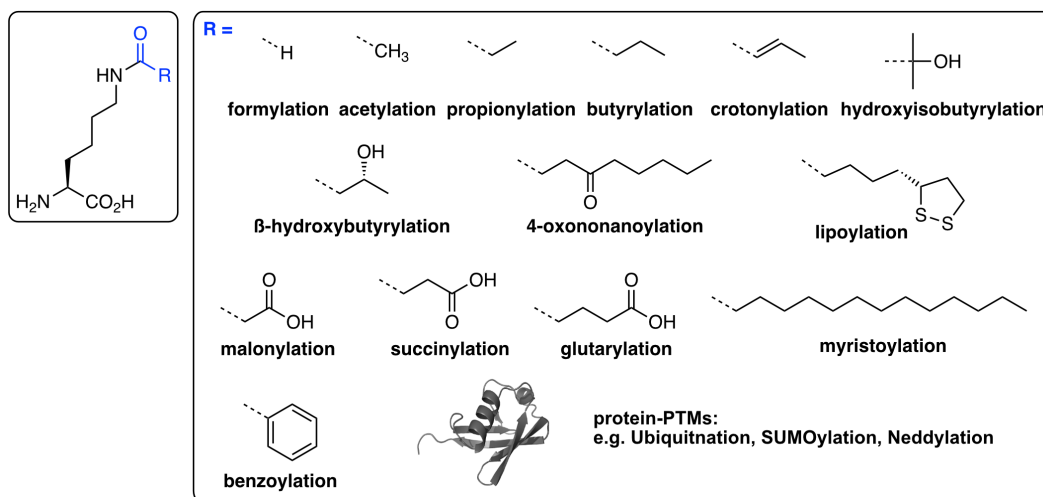


Fig. 4.2 | List of reported lysine PTMs. Lysine residues on proteins are subjected to numerous PTMs with diverse outcomes in their functions. The list is an extension of the lysine-PTMs, which are readily incorporated into proteins by GCE (cf. Fig. 1.7a). Additional PTMs include small polar side-chains such as β -hydroxybutyrylation²⁷⁹ and negatively charged groups like malonylation,²⁸⁰⁻²⁸¹ succinylation²⁸¹⁻²⁸² and glutarylation.²⁸¹ 4-oxononanoylation,²⁸³ lipoylation²⁸⁴ and myristoylation²⁸⁰⁻²⁸¹ may be regarded as long-chain fatty acid modifications. Also, even whole proteins²⁸⁵ such as Ubiquitin, SUMO or NEDD8 are also found on lysine residues of proteins. Intriguingly, benzoylation²⁸⁶ which has no natural source was found to be a histone mark with regulation in various gene expression. The likely source for this novel PTM was found in sodium benzoate, a common food preservative, thus linking nutrition and individual lifestyle with chromatin remodelling.

The current shortcut routes towards production of homogenous protein lysine acyl-PTMs by means of GCE generally follows a ‘one-UAA-one-aaRS’ approach (cf. Fig. 4.1a). While many (small) lysine acyl-UAAs (cf. Fig. 1.7) may be accepted by (shared) aaRSs, some PTMs require novel, but yet unidentified aaRS sequences (e.g. malonylation, succinylation²⁸¹⁻²⁸², glutarylation²⁸¹, lipoylation²⁸⁴) or likely will never be directly accessible via amber suppression (e.g. long chain fatty acid modification^{283, 287-288}, whole proteins²⁸⁵: Ubiquitination, SUMOylation, Neddylation). Alike bioorthogonal handles, incorporation of a universal lysine precursor at which selective amide bond yielding chemistries may be performed would allow to evade the pocket-size limitation of aaRSs and in principle enable the introduction of any (also larger) natural acyl modifications of choice (cf. Fig. 4.1c).

Development of improved amide forming reactions is an ongoing process in organic chemistry, especially with focus on optimising economic and ecological inefficiencies. Nowadays, a range of alternative amide bond yielding chemistries are available.^{273-274, 276} However, utilising such reactions for protein acyl transfers under bioorthogonal reaction conditions require masking of both functional groups to avoid side-reactions with endogenously occurring complementary groups.²⁸⁹ In this regard, three different reactions have been shown to provide amides under the proposed criteria.

Traceless Staudinger ligation is a modified version of the well-known Staudinger reduction (Fig. 4.3a1).²⁹⁰ The initial attack of the phosphine at the azide (amine surrogate), which leads to an aza-ylide intermediate under release of N₂, remains similar. But in contrast to the reductive pathway, at which the aza-ylide is becoming hydrolysed to reveal the amino group, the phosphine acyl-substrate (-CO₂H surrogate) offers an electrophilic carbonyl centre to be

attacked by the nucleophilic aza-ylide intermediate resulting in the formation of the amide bond.

This chemistry combined with the proposed approach as depicted in Fig. 4.1c has been previously published.²⁹¹ In detail, azidonorleucine UAA (Fig. 4.3a2) was incorporated in ubiquitin and histone protein and proteins were incubated with acetyl or succinyl-phosphinothioesters, which subsequently yield the respective post-translationally modified proteins. However, despite the flexible generation of acylated proteins, the main drawback lies in the competitive hydrolytic cleavage pathway, where the azide is becoming simultaneously reduced to the amine. This consequently results in heterogeneous mixtures of proteins bearing the free lysine and acetylated or succinylated lysine, which are impossible to separate. Thus, alternative ligation strategies need to ensure full stability and maximal conversion when the acylation reaction is executed.

The Bode research group, ETH Zurich, established with KAHA (**k**etoacid-**h**ydroxylamine) ligation a novel amide bond forming reaction. This follows mechanistically speaking a slightly different pathway (Fig. 4.3b).²⁹² The α -nucleophilicity of hydroxylamines enable initial attack at the carbonyl centre of α -ketoacids, forming a hemiaminal intermediate. The dual electrophilic and nucleophilic nature of both functional groups subsequently leads to effective elimination of both leaving groups (hydroxylamine substituent and CO₂) under formation of the kinetically stable amide bond. Depending on the hydroxylamine surrogate, the reaction either occurs via a three-membered cyclic transition state under loss of water (Type I, Fig. 4.3b.1) or directly via loss of the hydroxylamine benzoic acid leaving group (Type II, Fig. 4.3b.2).²⁹³ The ligation has been demonstrated suitable for multiple, unprotected SPPS segment condensations, providing peptides and proteins of various lengths such as GLP-1,²⁹⁴ UFM1,²⁹⁵ NEDD8,²⁹⁶ and SUMO.²⁹⁷ Though in principle compatible in aqueous buffers at 37°C, reactions are performed best when organic solvents (DMF, NMP, DMSO), a low pH (< 3) and elevated temperatures (37-60°C) are included.²⁸⁹ This renders general application on proteins as envisioned for site-specific installation of acyl-PTMs unsuitable.

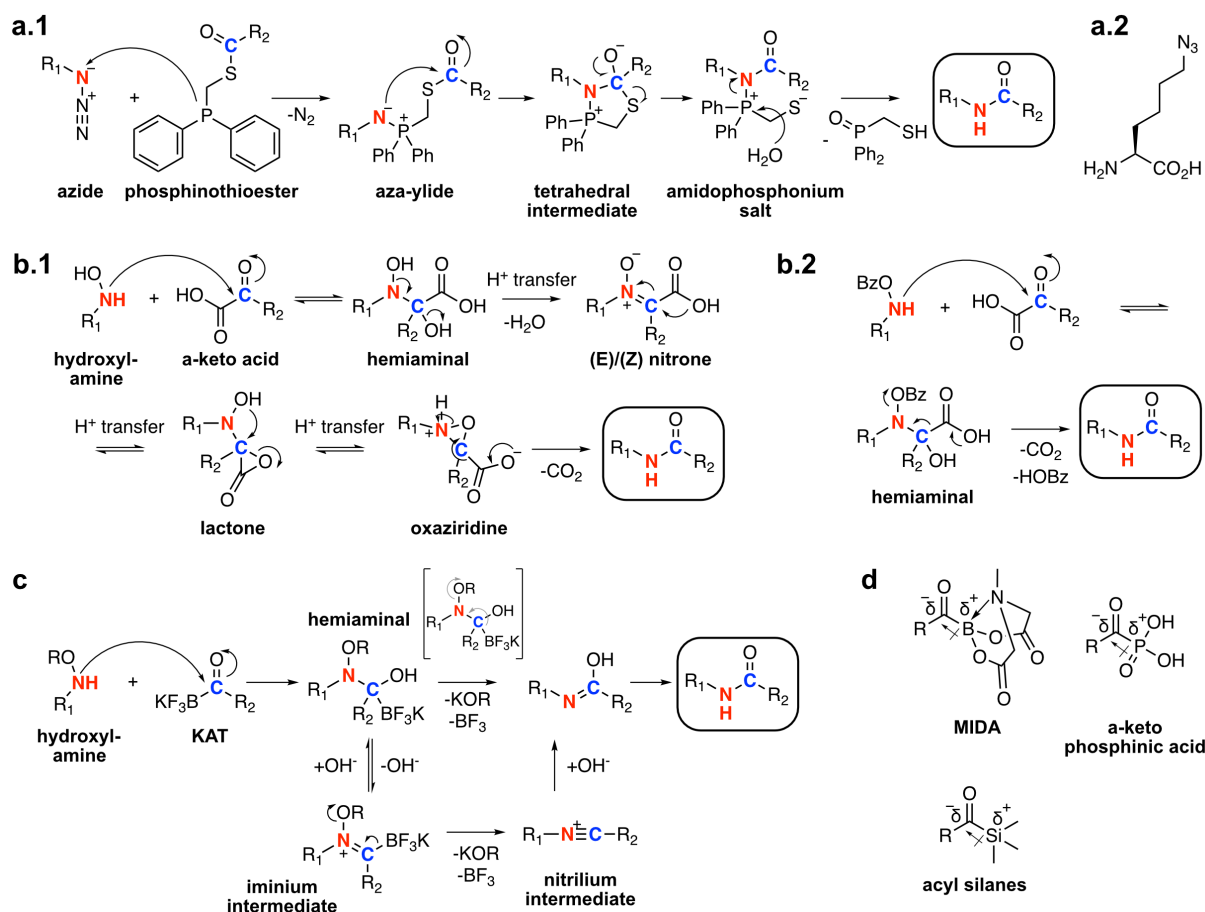


Fig. 4.3 | Routes towards amides under bioorthogonal conditions. (a) Traceless Staudinger ligation is a modified version of the Staudinger reduction, in which the aza-ylide is becoming trapped by a proximal acyl ester, yielding the amide bond. (b) KAHA ligation is mechanistically speaking slightly different, as both reactants with their embedded leaving groups are expelled in a concerted fashion. Depending on the amine substituents, amide bond formation may occur under two different pathways (**b.1** and **b.2**). Modification of the carboxylic acid surrogate resulted in the development of KAT ligation (**c**). The ketoacid-boronate replacement facilitates leaving group release due to the introduced electronegativity switch. Various KAT derivatives (**d**) bear this feature, yielding amide bonds with various hydroxylamine variants under a similar mechanism.

Subsequently, the same research group modified this reaction, which involved substitution of the ketoacid CO_2 leaving group with a boron derivative (Fig. 4.3c).²⁹⁸ The resulting electronegativity switch and labile $C-B$ bond facilitates the leaving group liberation of the carbonyl surrogate. Combined with ongoing modifications of the hydroxylamine side-chain structure, which finally led to the identification of carbamate moiety as an effective amine leaving group, the new amide bond forming reaction was found to react with accelerated rates of up to $20 M^{-1} s^{-1}$. This allowed for peptide labelling reactions at equimolar and bioorthogonal compatible conditions with great yield (>80%, equimolar conditions) and within a reasonable amount of time (minutes to o.n.).²⁹⁹ Protein labelling was also achieved via Cys-selective attachment of a small construct bearing a Cys-reactive group and the hydroxylamine moiety. Subsequent KAT ligations with constructs bearing various cargos were achieved with slight excess under slightly acidic conditions. Overall, prominent proof-of-principle examples of KAT (potassium/K acyltrifluoroborates) ligations include synthesis and labelling of GLP-1,²⁹⁹ sfGFP,³⁰⁰⁻³⁰¹ insulin,³⁰² Ubc9,³⁰⁰ and human thymidine phosphorylase³⁰⁰ with various cargos such as ^{18}F , PEG, biotin, fluorophores, fatty acid side-chains and dyes.

Reactivities may be furthermore adjusted with substitutions on both leaving groups. For instance, all KAT mimics bear the same feature of having an electropositive heteroatom placed next to the carbonyl centre, which preserves the facilitated lone electron pair rearrangements (Fig. 4.3d).³⁰³⁻³⁰⁵ The resulting KAT diversity enables flexible hydroxylamine design. While various acyl hydroxylamine derivatives readily undergo KAT ligation, alkyl hydroxylamine become trapped in an iminium intermediate, which is only slowly resolved. Incubation of the KAT surrogate *N*-methyliminodiacetyl acylboronates (MIDA) however leads to amide bond completion with the alkyl hydroxylamines.³⁰³

4.3 Scope of this work

The Bode research group has shown that KAT reactions may be performed on proteins under chemoselective conditions. The ligation strategy thus far included Cys-selective labelling of a protein with a bifunctional molecule bearing a cysteine-reactive moiety and a diethyl carbamate hydroxylamine containing molecule separated by a linker (Fig. 4.4a).

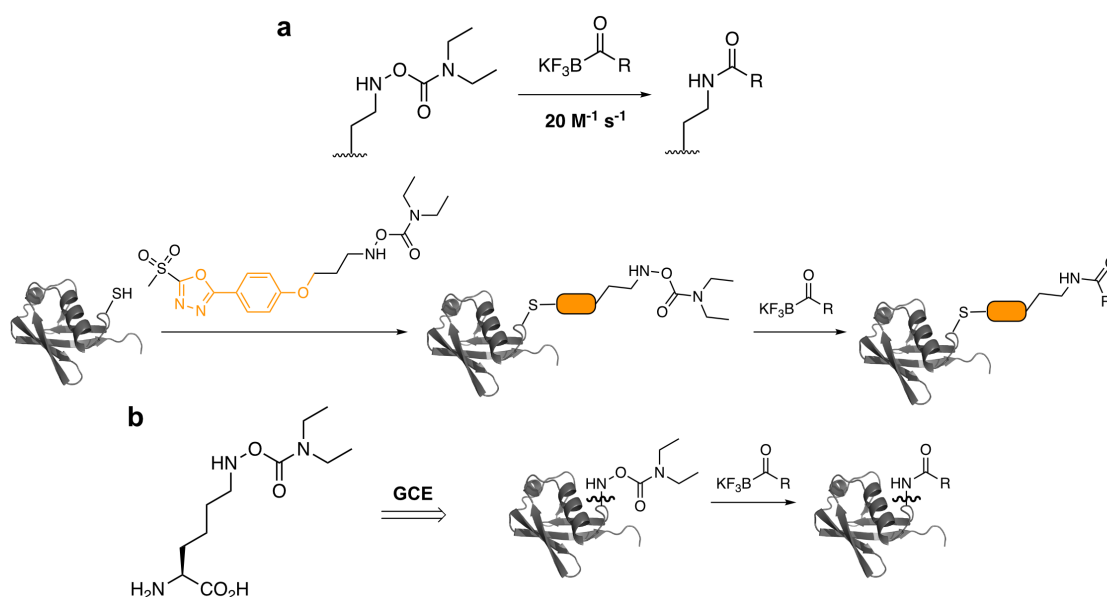


Fig. 4.4 | Utilising amide bond forming KAT ligation for the generation of defined lysine acylated proteins. (a) Diethyl carbamate hydroxylamine was previously shown to react selectively with KATs on peptide and protein levels. This moiety was introduced via a bifunctional construct (orange molecule part, depicted as orange box thereafter) bearing a Cys-selective handle and the hydroxylamine, followed by incubation with tailor-made KATs bearing various cargos such as PEG-linkers. **(b)** Based on this head group, lysine derivatives may be synthesised, which may be incorporated by means of GCE with an appropriate PyIRS. This would allow for subsequent incubation with tailor-made KATs bearing e.g. various PTM side-chains and thus the generation of well-defined lysine acyl-PTM proteins.

Subsequently, the hydroxylamine modified protein was reacted with slight excess of molecules bearing the KAT group and different cargos under slightly acidic conditions. Based on previous reports regarding stability, fast reaction rates and the possibility to perform conjugations on proteins, we speculated, that the carbamate hydroxylamine headgroup may be transferred onto a lysine backbone followed by site-specific incorporation into proteins by means of GCE with an appropriate PyIRS. Expressed and purified proteins bearing this UAA

may subsequently undergo selective reactions with tailor-made KATs bearing different lysine acyl-PTM side-chains to create well-defined acylated proteins with native lysine acyl-PTMs in a general fashion (Fig. 4.4b, also cf. Fig. 4.1c and Fig. 4.2b).

Hence, implementing such an approach would allow to generate any lysine acylated POI on demand and enable the biochemical and -physical characterisation of hitherto inaccessible lysine acyl-PTM bearing proteins. This chapter describes the synthesis of lysine-hydroxylamine derivatives, the reaction behaviour towards a model KAT substrate as well as the scope and limitations of this compound class, when combined with the genetic code expansion methodology with the final aim of producing acylated proteins.

4.4 Synthesis and KAT reaction evaluation of the carbamate hydroxylamine derivatives

The carbamate hydroxylamine moiety as depicted in Fig. 4.4 served as a starting point for design and synthesis of lysine hydroxylamine derivatives, which shall be site-specifically incorporated into proteins via GCE using suitable PylRS variants, followed by versatile protein acylation using tailor-made KATs.

Hydroxylamines may be directly accessed through oxidation of amines,³⁰⁶ which allows for a direct synthesis start on the lysine scaffold. The developed synthesis scheme is depicted in Fig. 4.5a. Because of the reported stability in acids, all protection groups were chosen to be removed with acidic treatment in a simultaneous and final deprotection step once the protected UAA precursor was synthesised. Starting from N ϵ -Cbz-protected N α -Boc-lysine, the carboxylic acid of the amino acid was protected as a tert-butyl ester first (21). After reductive deprotection of the Cbz-group to obtain free amine intermediate 22, the liberated amine was reacted with dibenzoyl peroxide to yield the protected hydroxylamine precursor 23. The higher nucleophilicity of the hydroxylamine nitrogen compared to the hydroxy group once it becomes deprotected for late-stage functionalisation would lead to various side-products in the subsequent late-stage functionalisation steps. Thus, this position was Boc-protected to obtain intermediate 24 prior basic deprotection of the hydroxyl group to obtain UAA precursor 25. Overall, the synthesis procedure until this point enabled large-scale production of the main precursor in overall good yields.

The final carbamate-linked UAAs 27a-d were readily synthesised when reacted with commercially available carbamyl chlorides followed by simultaneous acidic deprotection of all three protection groups (Fig. 4.5b).

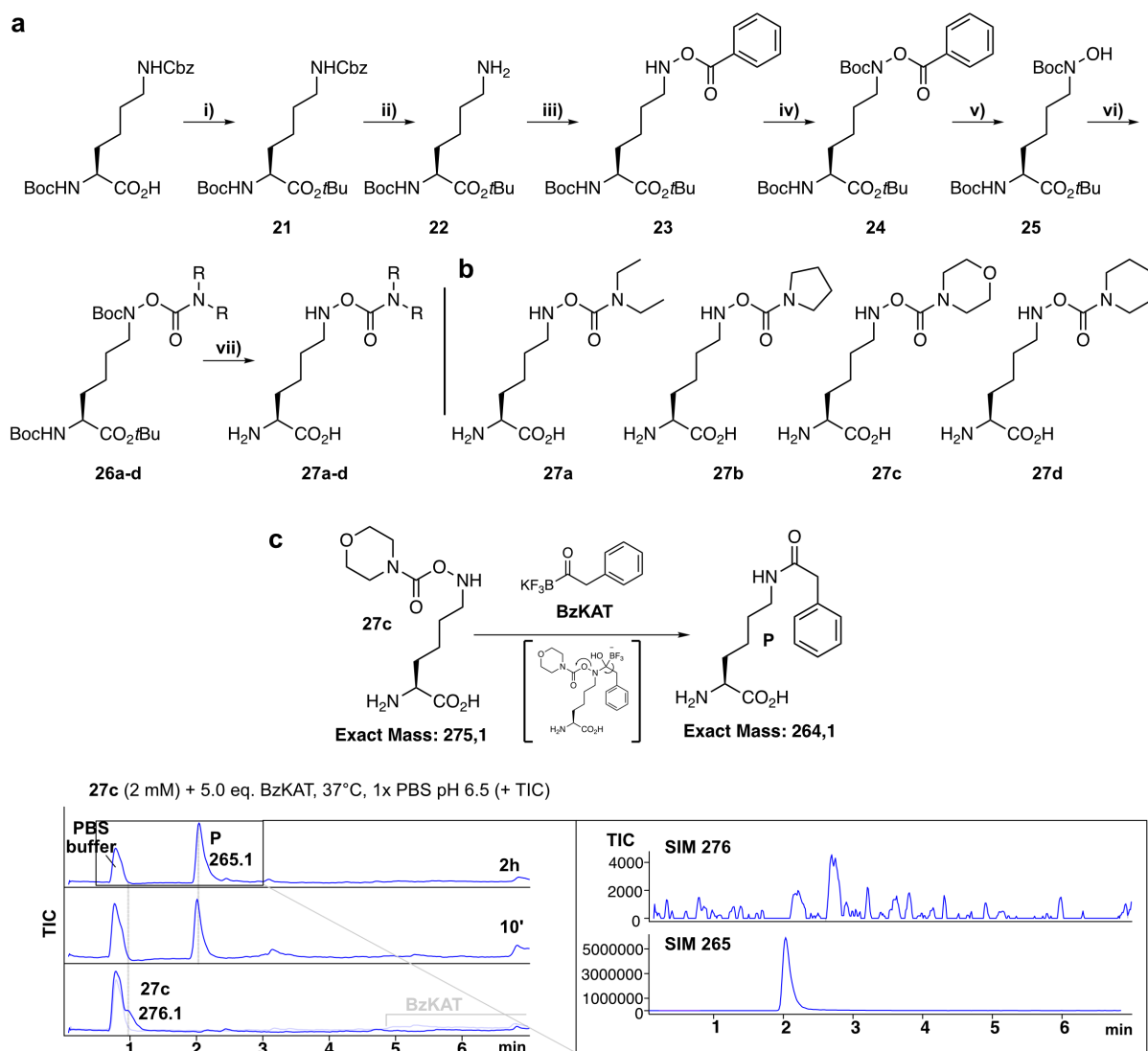


Fig. 4.5 I Design, synthesis and reactivity of carbamate hydroxylamine lysines. (a) Synthesis of carbamate hydroxylamine lysines, **i)** Boc₂O (1.40 eq.), DMAP (0.40 eq.), *t*BuOH, 30°C, o.n., 94%; **ii)** Et₃SiH (10.0 eq.), Pd/C (10% wt), 0°C → r.t., o.n., 98%; **iii)** dibenzoylperoxide (1.5 eq.), pH 10.5 buffer/DCM, 4h, 0°C, 61%; **iv)** Boc₂O (20.0 eq.), 50°C, o.n., 96%; **v)** LiOH (10.0 eq.), r.t., 40min, 77%; **vi)** carbamoyl chloride (2.2 eq.), DIPEA (2.5 eq.), DMAP (0.10 eq.), 60°C, o.n.; **vii)** HCl/dioxane, r.t.; (b) Designed and synthesised carbamate UAAs according to the procedure described in (a). (c) Time-dependent BzKAT reaction profile of UAA 27c. For general detection of both, starting material and product, the positive ion mode of this reactions is depicted. Due to partial overlap of 27c with PBS buffer peak, selected ion mode (SIM), where individual *m/z* values may be searched, was applied to detect UAA and product. No sign of UAA 27c was detected in contrast to product P after 2h reaction time, indicating absence of 27c and full conversion to product P. TIC - total ion count.

Incubation of UAA 27c with a model benzyl KAT substrate (BzKAT, Fig. 4.5c) in 1x PBS buffer and 5x excess of BzKAT led to reaction completion after 2h, confirming the previously reported swift reactivity of this hydroxylamine carbamate class towards formation of amide bonds. Additionally, incubation in PBS buffer of various pH over prolonged time at 37°C showed no sign of degradations (Fig. SI 4.1). UAAs were also screened with several, in-house available PyIRS variants, however, no positive expressions of the sfGFP-N149TAG reporter protein were detected. This is unsurprising, considering the unprecedented structure, where most PyIRS mutants were evolved to incorporate lysine derivatives bearing amide, carbamate or entirely different groups at this position. The Y349W *Mb* PyIRS mutation,¹¹¹ which stabilises the

structurally related Boc-protected aminoxylysine analogue, did also not yield any sfGFP protein. Thus, UAAs were subjected to aaRS selections as generally outlined in Fig 1.6c. To note is, that prior synthetase selections, LC-MS analyses already indicated (partial) degradation of UAA 27a, 27b and 27d in their stock solutions (100 mM in water) over time (Fig. SI 4.1). UAA 27c appeared stable over prolonged time and stock solutions were frequently checked prior performance of individual aaRS selection steps.

4.5 Synthetase selection for the incorporation of hydroxylamine lysine derivatives

The crystal structure of the *Mm* PylRS (PDB: 2Q7H) in complex with pyrrolysine hints towards several critical residues for stabilisation of the pyrrolysine substrate³⁰⁷ (Fig. 4.6a and b) and residues, which possibly need to be changed in order to accept the hydroxylamine lysine derivative: The pyrroline group is buried in a hydrophobic core consisting of residues (*Mb* numbering in brackets) L305 (270), C348 (313), Y384 (349), V401 (366), W417 (382). Additionally, the interactions of N346 (311) with the primary amine through an additional water molecule and directly with the N ϵ -amide bond as well as Y384 (349), which contacts the pyrroline nitrogen accounts for pyrrolysine substrate specificity. Thus, structural comparison with the carbamate hydroxylamine 27c suggests mainly those residues to be adjusted for accommodation of the new bond type and side-chain (Fig. 4.6c). Initial RS selections were performed with libraries randomised at following positions (libraries were generated by A.-D. Brunner, former AK Lang group member; *Mb* numbering): AB1 with L270, Y271, N311, C313, Y349 and AB2 with L270, Y271, N311, C313, Y349, with A267T fixed.

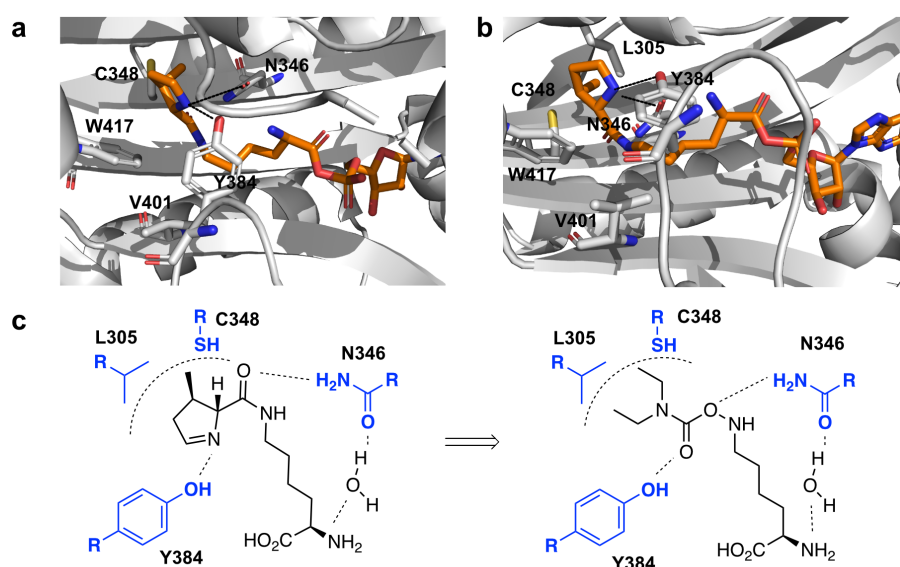


Fig. 4.6 | Crystal structure of the *Mm* wtPylRS with its pyrrolysine substrate. Top (a) and side view (b) of the binding pocket with pyrrolysine-ADP and the critical residues involved in amino acid recognition. (c) Graphical representation of the pyrrolysine binding mode and hypothesised positions, which need to be mutated to accommodate hydroxylamine UAAs. Dashed lines indicate H-bonding of Y384 and N346 with the pyrroline nitrogen and N ϵ -amide bond.

Synthetase selection with UAA 27c and library AB1 yielded 26 colonies after two rounds of negative and positive selections, while selections with library AB2 only provided unspecific clones. The 26 clones, which still contain the CAM-TAG reporter gene, were picked and plated out on agar plates containing increasing CAM concentrations in presence and absence of 27c. 17 clones showed considerably larger growth in presence of UAA than without, although little growth of various clones in absence of UAA was still observed. Synthetase plasmids were subsequently isolated and transformed again with the sfGFP-N149TAG reporter gene to screen for protein expression and mass identity in presence of UAA. No production of proteins either in presence or absence of UAA could be detected on SDS-PAGE (Fig. SI 4.2). However, cell pellets containing the UAA exhibited weak green fluorescence under UV light indicating suppression of the amber codon, though at low levels. Thus, all pellets were subsequently His-tag purified and analysed via LC-MS to identify the suppressed protein with its amino acid (Table 4.1, Fig. SI 4.3).

Clone	Observed MS	Clone	Observed MS
A2	27856	A12	27845
A3	27850	B3	27844
A4	No MS observed	B5	27857
A5	27858	B9	27855
A6	27853	B10	27856
A7	27860	B11	27855
A9	27855	B12	27855
A10	27853	C12	27852
A11	27854	MS expected	27970/27839 (-Met)

Table 4.1 | Obtained sfGFP-27c masses after RS selection with library AB1.

None of the observed MS values corresponded to the calculated mass bearing UAA 27c ($m/z_{\text{calc}} = 27970$ Da, $m/z_{\text{calc-Met}} = 27839$ Da). Phenylalanine misincorporation ($m/z_{149F} = 27859$ Da) is a frequently observed phenomenon from previous synthetase selections within the lab and the calculated mass lies indeed in the range of the observed m/z values. The observed sfGFP mass would also correspond to sfGFP containing a hydrolysed hydroxylamine lysine moiety with a calculated m/z value of 27857 Da. Considering that measured LRMS protein masses may differ several Da from the expected value, purified sfGFP proteins may contain either amino acid. Despite the observed stability of UAA 27c in PBS buffer, overall degradation in *E. coli* and complex media over prolonged time becomes a likely option. The hydrolytic pathway seems plausible given the previously observed degradations of structurally similar UAAs in their stock solutions. A detailed look at the structure identifies a highly electrophilic carbonyl centre with a good leaving group as a weak point for UAA degradation. Attack of a nucleophile, for instance a water molecule would split the UAA into its hydroxylamine part and the aminocarbonate (= leaving group), which further dissociates into the corresponding secondary amine and CO_2 (Fig. 4.7).

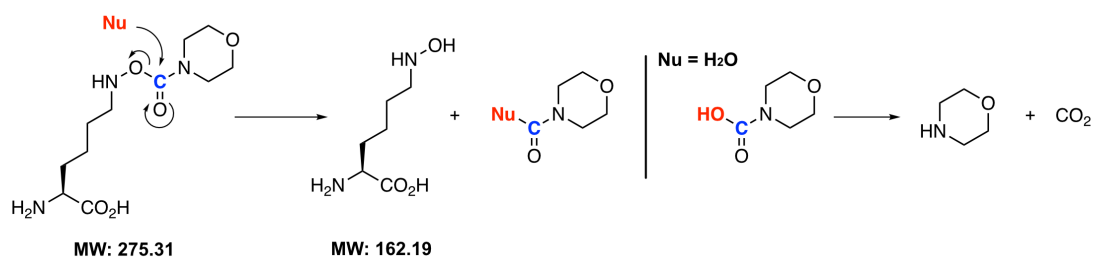


Fig. 4.7 | Hypothetical Nu-degradation pathway of the 1st generation hydroxylamine UAA. Nu - nucleophile.

Thus, the aaRS selection approach possibly enriched for sequences accepting either the hydrolysed UAA or for the original amino acid 27c, which becomes degraded co- and/or posttranslationally. A further option includes phenylalanine as the major misincorporation pathway. Overall, while the hydroxylamine carbamates appeared stable for aqueous peptide ligations within short time periods as previously reported, the carbamate hydroxylamine moiety appeared to be unstable, when exposed to cytosolic environments at prolonged time and temperature of 37°C. Hence, improvements of the UAA design concerned preventing the hydrolytic pathway, which is discussed in the next section.

4.6 Design, synthesis and KAT reaction of carboxyl hydroxylamine derivatives

The key principle of fast KAT reactions and hence hydroxylamine molecule design lies in the subsequent stabilising propensity of the leaving groups embedded in both reactants. While the preferential cleavage direction of KATs is already ‘encoded’ with their BF₃K group and labile C←B bond, the side-chain of hydroxylamines derivatives may be variably designed for optimal reaction performance. Specifically, the elimination of morpholinocarbonate of UAA 27c (Fig. 4.7), when incubated with BzKAT, appears to be highly favoured due to subsequent resonance stabilisation of the -CO₂H group combined with the beneficial entropic contribution via slow dissociation of CO₂. This renders the hydroxylamine carbamate class as efficient, but most likely unstable surrogate for KAT reactions. Generally, the quality of leaving groups is correlated with their pK_a values, i.e. the lower the value, the more stabilised is the conjugated base and the ‘better’ is the leaving group, respectively. Thus, replacement of the carbamate group with a carboxylic acid is hypothesised to prevent the Nu-cleavage pathway whilst retaining KAT reactivity due to expected similar pK_a values (~3-4) of both leaving groups (Fig. 4.8). However, one needs to bear in mind, that this likely comes with an enhanced electrophilicity of the carbonyl centre due to lack of the e⁻-donating +M-effect of the carbamate hydroxylamine side-chain.

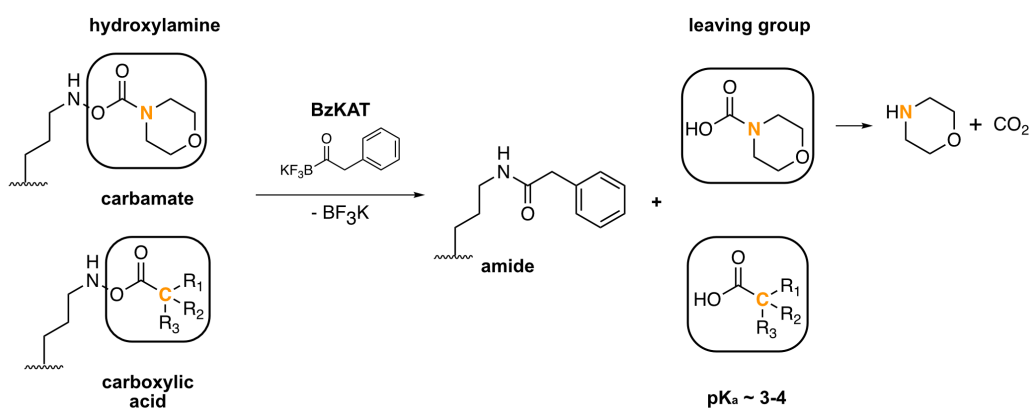


Fig. 4.8 | Design of the next generation of hydroxylamine derivatives. Based on the experiences of the carbamate hydroxylamine derivatives, the UAAs of the 2nd generation were designed bearing a carboxylic acid residue. This was hypothesised to have greater stability while retaining similar KAT reactivity due to expected similar pK_a values.

Similar to the previous UAA design, the synthesis of the new carboxy hydroxylamines derivatives may be continued with precursor 25, when reacted with appropriate acyl chlorides. Despite the switch to the carboxylic acid leaving group, the electrophilic carbonyl centre still remains a potential point for attack. Hence, three different UAAs with diverse side-chains were synthesised (Fig. 4.9a) with the main criteria of bearing a tertiary carbon centre next to the carbonyl group. It was hypothesised that the additional bulk further protects this centre and minimises putative side-reactions. Again, test incubation of 29a with 5.0 eq. BzKAT in 1x PBS buffer confirmed the possibility to perform KAT ligations with the modified hydroxylamine class (Fig. 4.9b). While starting material was readily consumed by BzKAT to form product P within 10', a minor subspecies (P*) additionally occurred, which only slowly decreased over time. The positive ionisation mode of the LC-MS pattern (+ TIC) showed a similar mass as the product, while the corresponding negative TIC revealed a mass of 332.2 (-H⁺). This might represent intermediate P* (Fig. 4.9b), where the amino leaving group is first expelled, followed by slow elimination of the BF₃ group over time (24 h), which eventually collapsed into the amide product P with only residual P* left.

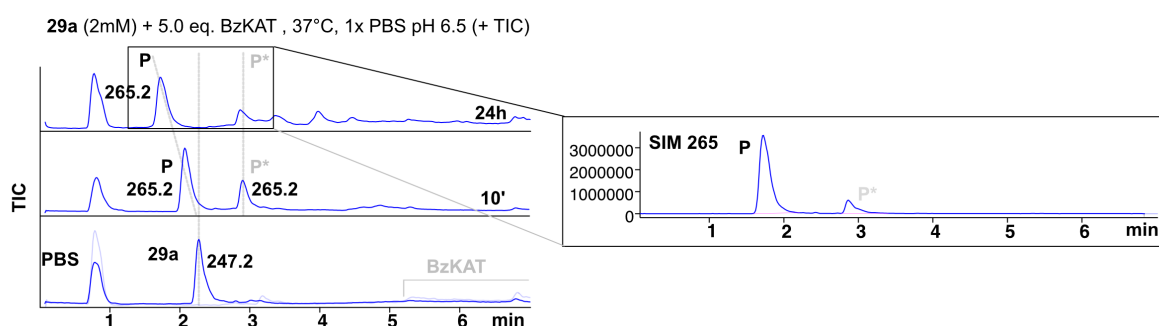
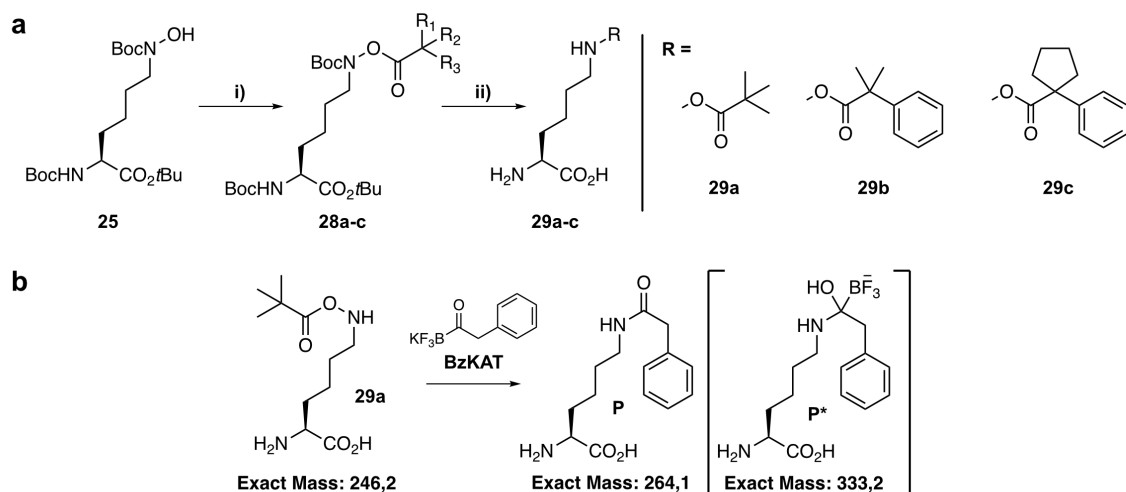


Fig. 4.9 | Design, synthesis and KAT reactivity of carboxyl hydroxylamine lysines. (a) Synthesis of acyl hydroxylamine lysines, i) acyl chloride (2.2 eq.), DIPEA (2.5 eq.), DMAP (0.01 eq.), 60°C, o.n.; ii) HCl/dioxane, r.t., o.n.; four different derivatives (**29a-c**) were obtained according to this synthesis procedure; (b) Time-dependent BzKAT reaction profile of UAA **29a**. Depicted are the positive ion mode. During the measurement of the 24h sample, the LC-MS calibration was slightly off resulting in a shifted P retention time. MS analysis confirmed the species to be P. TIC - total ion count.

Also, UAA **29a** appeared stable in PBS buffer over various pH ranges at 37°C over prolonged time, although little hydrolysis was already observed when **29a** was incubated solely in water at 37°C (SI Fig. 4.4). Due to the general possibility to perform KAT ligations with this modified carboxyl hydroxylamine lysine UAA class, aaRS selections were still attempted. Library choice remained the same due to similar side-chain resemblance compared to the previous UAA design. 23 clones survived the aaRS selection with **29a** and library AB1. To gain a better understanding of the selection procedure and putative sequence convergence, nine clones were additionally sequenced along with three unspecific mutants, which additionally exhibited survival on CAM plates in absence of UAA (Table 4.2).

AB1 library	L270	Y271	N311	C313	Y349
RS1	M	A	H	V	F
RS2	M	G	H	T	W
RS3	M	G	H	A	W
RS5*	V	G	H	V	F
RS9	M	G	H	V	F
RS13*	V	G	H	V	F
RS14	G	L	S	V	F
RS20*	V	G	H	V	F
RS23*	V	G	H	V	F
Unspecific					
RS7	L	M	M	A	F
RS11	M	G	T	F	F
RS19	M	C	S	L	-

Table 4.2 | Sequence analysis of the evolved specific and unspecific aaRSs after selection with UAA 29a and library AB1. * denotes similar sequences.

Overall, the evolved PyIRS variants exhibited strong convergence to the same sequence or positional similarities. Four out of nine sequences were identical, while 8/9 contained the N311H mutation, 7/9 Y349F, 7/9 Y271G, 7/9 C313V, 4/9 L270M and L270V. Importantly, no sequence correlations were found with the unspecific clones. Interestingly, the rather unusual and rare N311H amino acid substitution, an amino acid which is usually involved in catalytic reactions, evolved as the most constant mutation, while the hydrophobic replacements possibly account for proper tert-butyl group accommodation. The synthetase plasmids were isolated and reintroduced again in *E. coli* together with the sfGFP-N149TAG reporter gene and cells were grown in presence or absence of UAA as described above. WB analyses initially indicated selective protein expression mainly in presence of the UAA confirming the general selection procedure, although faint bands in absence of 29a were observed as well (Fig. 4.10a). The unspecific PyIRSs 7, 11 and 19 also showed misincorporation in absence of UAA 29a (Fig. 4.10a).

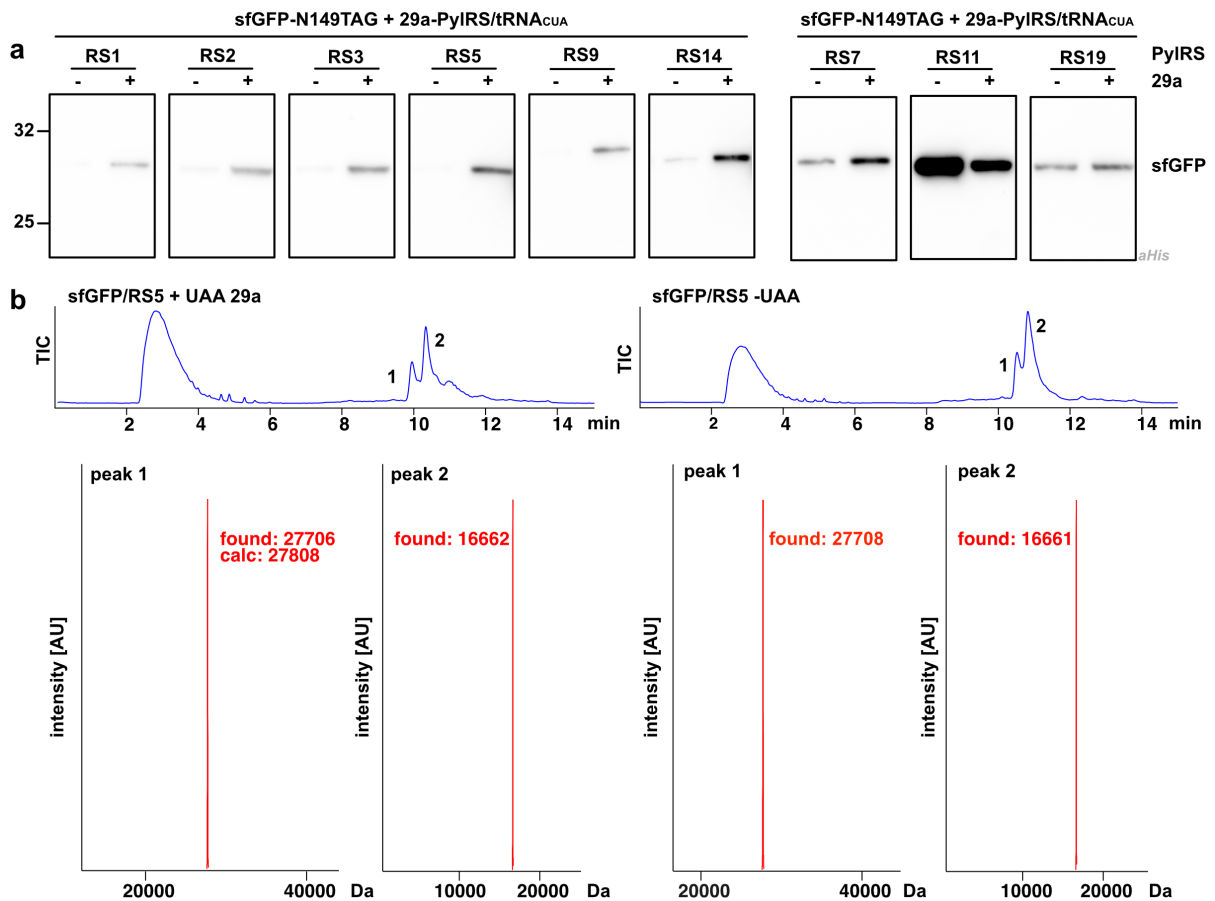


Fig. 4.10 | Amber suppression analysis of the evolved 29a-aaRSs. (a) WB analysis - Sequences, which were considered UAA 29a selective on the CAM plates (Table 4.2), appeared selective for suppression of the sfGFP149TAG gene in presence of 29a with minimal misincorporation in absence of the UAA. Unspecific aaRSs also showed unspecific suppression of the TAG codon in absence of 29a. **(b)** However, purification of the RS5/sfGFP149TAG with or without UAA lead to production of sfGFP with similar MS result (-M, 2P), corresponding to lysine misincorporation (peak 1). A second protein fragment was copurified with a MS result corresponding to apparently truncated sfGFP (position 2-148).

However, pellets exhibited only weak green fluorescence under UV light, as previously observed with the UAA 27c selections and expressions. Purifications with the RS5 variant as the most frequent PylRS from the selection did not reveal the expected $m/z = 27808$ Da. The observed mass was 100 Da lower, which matches the loss of the pivalic acid residue and sfGFP bearing a free lysine (Fig. 4.10b). To rule out fragmentation of the UAA head group during LC-MS measurement, which for instance was occasionally observed with Bock bearing proteins, sfGFP-29a was additionally incubated with BzKAT during purification. However, the mass remained similar, indicating no presence of intact UAA post protein translation (Fig. SI 4.5a). Intriguingly, purification of the -UAA pellet revealed the same mass, hinting towards misincorporation of lysine. Also, a second protein mass appeared along all protein purifications, independent of the presence or absence of the UAA. The m/z of 16661 Da corresponds to the truncated sequence of sfGFP (amino acids 2-148). Due to purification procedure via a C-terminal His₆-tag, it was first speculated that protein 'self-cleaving mechanism' has appeared after elution in dependence of preceding H148 and position 149, either bearing the UAA or lysine. However, sfGFP bearing N149K, H148A/N149K or H148A/N149Bock showed no sign of such fragmentations indicating that the combination

H148/K149 has no autocatalytic cleavage propensity (Fig. SI 4.5b). Additionally, sfGFP-H148A/N149TAG in presence as well as absence of 29a produced 149K bearing protein, though fluorescence was barely visible. Together with the result in Fig 4.10b, this indicates that the evolved synthetase likely misincorporates lysine instead of UAA. Indeed, repeating the WB experiments revealed now only unspecific suppression. This could not be rescued by growth in 17 amino acid auto-induction medium, which lacks lysine as a supplement, suggesting that endogenously produced lysine in bacteria is sufficient. Also, Δ hcp cells, which lack the enzyme catalysing hydroxylamine reduction, did not result into selective suppression (Fig. SI 4.6).

Overall, a possible explanation might be again enrichment of aaRSs sequences, which either accept lysine or UAA 29a along with co- and/or posttranslational decay. Unlike the degradation via the Nu-pathway (Fig. 4.7), this type of hydroxylamine lysine might undergo an E-type degradation pathway (Fig. 4.11). The enhanced electron-withdrawing effect of the carboxylic acids compared to the carbamate renders the N-O bond more labile to cleavages, which would lead to lysine bearing sfGFP when 29a was incorporated and E equals a H^+ (Fig. 4.11). The mechanism under which the truncated sfGFP sequence was generated and co-purified, remains however unknown.

PyIRS selection with UAA 29b and library AB1 and D3 (randomised positions: L270, Y271, L274, N311, C313) revealed less conserved sequences (Table 4.3) compared to previous selections with UAA 29a. First, six clones of the AB1 selection were sequenced, followed by three additional clones, which appeared after five days of incubation at 37°C. Three of the six initial clones showed identical sequences with N311S being the most prominent converged residue, while no clear trend for the remaining positions was deduced. The three additional PyIRS variants showed no sign of sequence similarity, indicating unspecific incorporations which presumably started to occur with decreasing CAM concentration and thus selection pressure over time. Library D3 revealed only three clones after the last positive selection step with an additional colony after five days of incubation. Two identical sequences with clustering of rather unusual, basic amino acids emerged. However, in all cases no green fluorescence of sfGFP-N149TAG was observed on auto-induction plates, irrespective of the presence or absence of the UAA. This suggests again decomposition of UAA 29b along with baseline activity of the evolved PyIRS variants towards incorporation of natural amino acids, which was sufficient to carry the clones through the selection procedures without substantial protein production.

AB1 library	L270	Y271	L274	N311	C313
RS1*	H	-	-	S	S
RS2	A	R	-	S	M
RS3	M	N	-	S	V
RS4	C	L	-	V	A
RS5*	H	-	-	S	S
RS6*	H	-	-	S	S
Clones after 5d					
RS7	-	-	-	E	I
RS8	F	W	-	A	L
RS9	Q	Q	-	S	H
D3 library					
RS1'	K	H	K	I	G
RS2''	-	-	-	Q	H
RS3'	K	H	K	I	G
Clone after 5d					
RS4''	-	-	-	Q	H

Table 4.3 | Sequence analysis of the evolved specific and unspecific aaRSs after selection with UAA 29b and library AB1 or D3. *, ', '' denotes similar sequences.

UAA 29c degraded in its stock solution over time, highlighting the fragile stability of this molecule class in aqueous environment.

Overall, the newly designed hydroxylamine derivatives exhibited sufficient reactivity towards KATs. The pK_a of the carboxyl acid is expected to be in a similar order as that of carbamates. This proved to be sufficiently acidic for KAT ligations to proceed within reasonable amount of time. However, the synthesised UAAs still remained prone to degradation and unstable in complex media and/or bacteria as evident from degradation in stock solutions and unsuccessful PyIRS selections. Despite the N→C change of the hydroxylamine side-chain and assumed prevention of Nu-degradation pathway, the E-route with carboxylic acids as an excellent leaving group likely became the major side-reaction pathway (Fig. 4.11a). Hence, keeping the overall acyl-like side-chain architecture is likely accompanied with secondary degradation pathways over time. Therefore, further modification of the UAA design focussed on replacement of the electrophilic carbonyl centre with a methylene group, which was hypothesised to substantially increase the stability of the UAA (Fig. 4.11b).

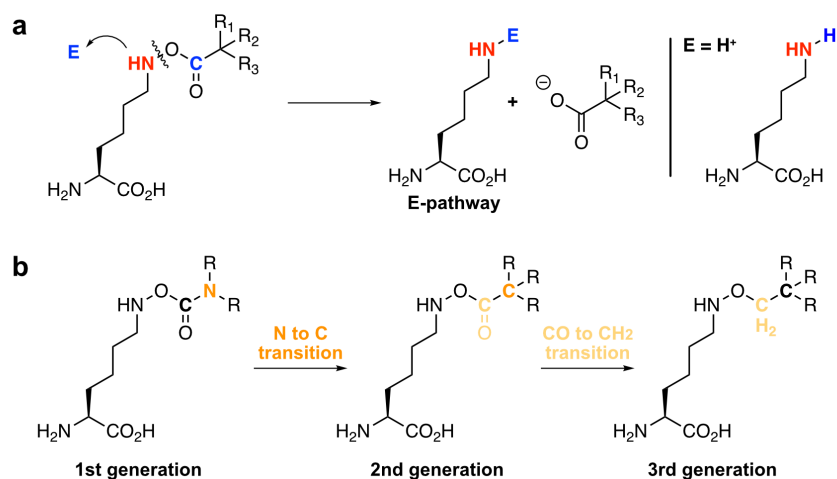


Fig. 4.11 | Hypothetical E-degradation pathway of the carboxylic acid hydroxylamine UAAs (a) and subsequent scaffold design (b). E - electrophile.

4.7 Design, synthesis and KAT reaction of alkyl hydroxylamine derivatives

Switching the carbonyl group to a methylene group increases the basicity of the resulting alcohol leaving group by several orders of magnitudes. A previous KAT reaction study with methyl hydroxylamine highlighted iminium intermediate trapping and incomplete amide bond formation due to stability of the zwitterionic state and weak leaving group propensity of the methoxide residue (Fig. 4.12).³⁰³

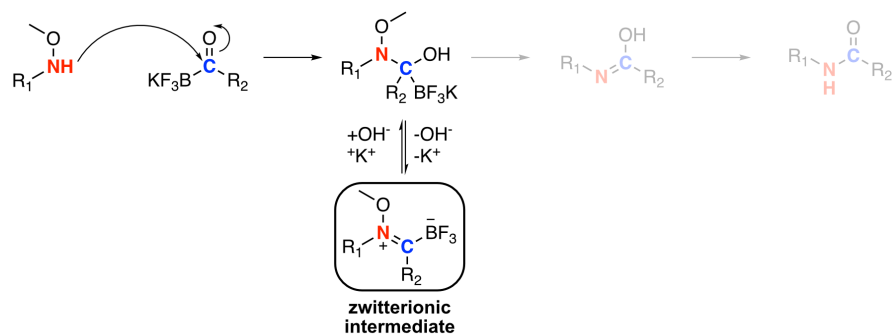


Fig. 4.12 | Intermediate trapping after KAT reaction with methyl hydroxylamine.

Thus, the boundary for meaningful KAT reactions is set with the pK_a value of methanol (~15.5) as the hydroxylamine leaving group. Given the required CH₂-stability spacer, enhanced acidity of alcohols may be mainly achieved via introduction of electron-withdrawing groups (EWG). Prominent groups for instance include halides, carbonyls, nitriles and nitro groups. To gain an initial estimation of KAT reactivity with alkylhydroxylamines, UAA precursor 25 was reacted with *o*-nitro benzyl bromide to yield UAA 31 after acidic deprotection (Fig. 4.13a). The electron withdrawing nitro group renders *o*-nitrobenzyl alcohol (pK_a of 13.3, calculated with ChemDraw v.19.1) approximately 158x more acidic than methanol. UAA 31 (2 mM) was incubated with various excess ratios of KAT in PBS buffer o.n. at 37°C and reactions were monitored again via LC-MS.

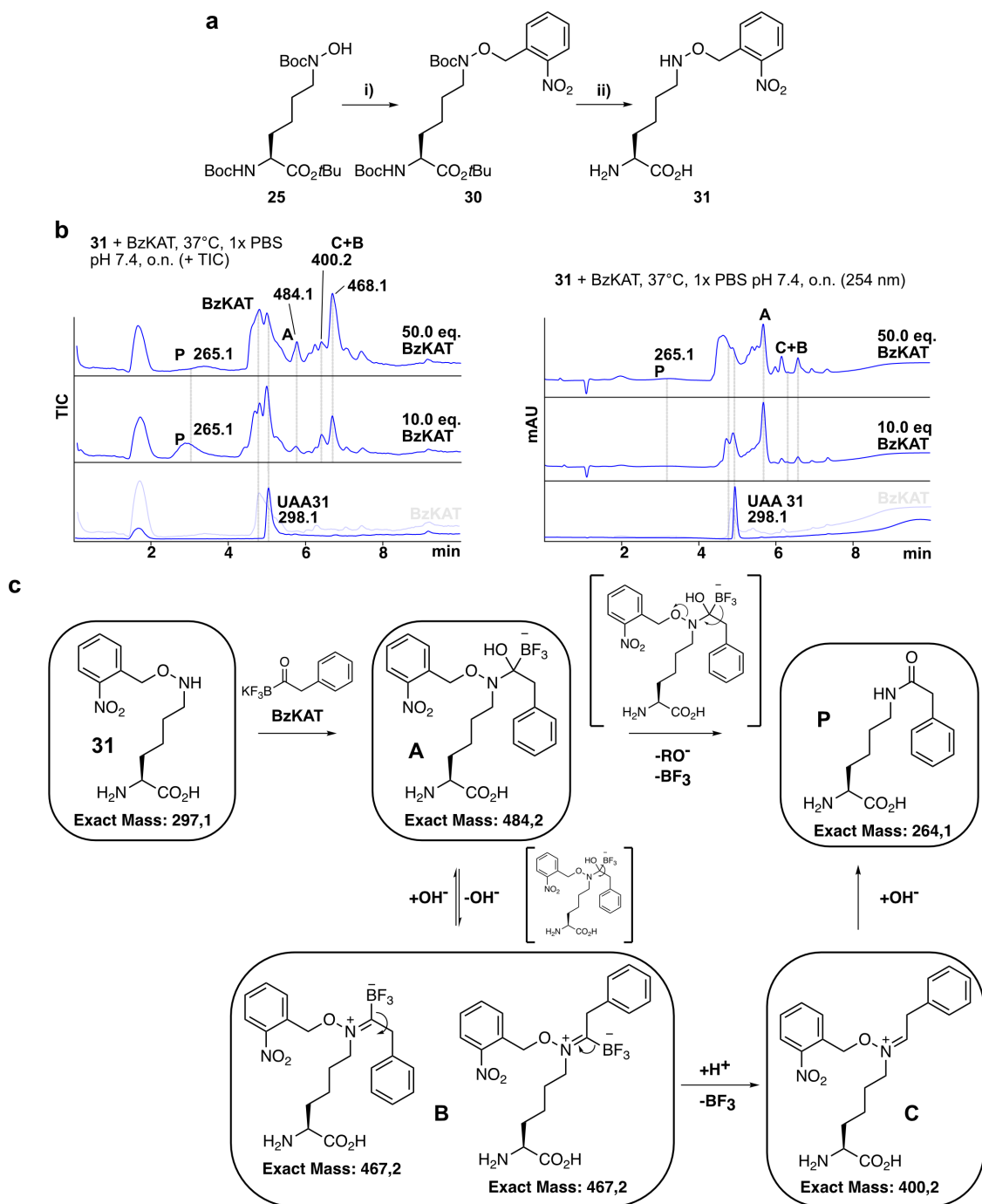


Fig. 4.13 | Design, synthesis and KAT reactivity of *o*-nitro benzyl hydroxylamine lysine. (a) Synthesis route of *o*-nitrobenzyl hydroxylamine lysine, **i)** *o*-nitrobenzyl bromide (1.2 eq.), K_2CO_3 (2.0 eq.), 60°C, 4d, 89%; **ii)** 2 M HCl/dioxane, r.t., o.n., quant.; **(b)** Concentration-dependent KAT reaction profile of UAA 31 (2 mM) with various excess of KAT. Depicted are starting material traces as well as the traces of the KAT ligation at the indicated conditions. The positive ion mode as well as the corresponding 254 nm spectrum of the KAT reactions are shown. Identified intermediates were denoted according to the intermediates as depicted in **(c)** with the observed *m/z* values. **(c)** Hypothetical reaction pathway with the stable intermediates according to observed MS values. TIC - total ion count.

As depicted in Fig. 4.13b, the reaction resulted in a complex profile and observed peaks could only be partially identified to the various intermediates of the KAT ligation (Fig. 4.13c). Overall, starting material mostly remained unreacted when incubated with 10x excess of BzKAT, indicating sluggish reaction kinetics. Increasing the KAT equivalents to 50x resulted in a similar

picture with incomplete conversion to product. Peak A possibly represents the initial hemiaminal, which appeared to be highly stable over prolonged incubation time, while peak B may refer to the $\text{H}_2\text{O}/\text{OH}^-$ eliminated nitron intermediate as already proposed in ref.³⁰³ The subspecies in peak C, which has a slightly shorter retention time compared to B possibly represents the other nitron isomer, where the BF_3 group becomes preferably cleaved in the ESI-MS. Alternatively, it represents the directly BF_3 eliminated iminium intermediate with the *o*-nitrobenzyl group still attached. In either case, nucleophilic addition of $\text{H}_2\text{O}/\text{OH}^-$ eliminates the remaining hydroxylamine leaving group, which is however a slow process as only little product P was detected over time. Reactions under diluted conditions (0.2 mM of UAA 31), higher KAT concentrations (up to 500x), slightly acidic or basic pH (6.4, 6.9 and 8.0) as well as prolonged reaction times (4 d) did not result into improved product yields (Fig. SI 4.7). In all cases, the various intermediates could still be detected besides little product P, highlighting the enhanced stabilities of the transition states. Overall, the leaving group propensity of the *o*-nitrobenzyl side-chain still remains too weak to effectively complete amide bond formation.

4.8 Probing KAT reactions with an aryl hydroxylamine lysine derivative

Given the sluggish KAT kinetics with UAA 31, the pK_a of the amine leaving group required further adjustments below 13.8. Generally, there are only few options to increase the acidity of alcohols. Consecutive attachment of EWGs along the alcohol carbon backbone may enhance the acidity via additive pull-effects, although this is limited with increasing distance. For instance, the predicted pK_a value of trifluoroethanol, pentafluoro-1-propanol and heptafluoro-1-butanol slightly decreases with 12.11, 11.89 and 11.72, due to equalising pull effects of the directly neighbouring fluorine atom and the distal fluorine atoms of the $-\text{CF}_{2x}$ group. Exceptions are represented by phenol derivatives, where the negative charge of the deprotonated hydroxyl group after proton release is stabilised by resonance, which results in a markedly increased acidity by several orders of magnitudes (phenol: pK_a 10 vs. alkyl alcohols: typically 15-16). While a previous study reported *N*-alkyl phenyl hydroxylamines to be unstable in aqueous environment,³⁰⁸ electron-pulling *p*-nitrophenol (pK_a ~7.15) hydroxylamine derivatives have been successfully synthesised with enhanced stability.³⁰⁹ The reported synthesis procedure, which involves oxime formation of the aryl hydroxylamine with a carbonyl followed by reduction of the oxime to yield the final hydroxylamine was altered due to unavailability of a carbonyl group on the lysine scaffold. Alternatively, the N_ϵ -group was transformed into a hydroxy group via oxidation at the δ - CH_2 position with RuO_2 hydrate powder and NaBrO_3 , followed by reduction with NaBH_4 (Fig. 4.14a).³¹⁰ The protected amino acid alcohol 34 was subsequently reacted with *p*-nitrophenyl Boc-hydroxylamine 35 (ref. ³¹¹) via a Mitsunobu reaction to yield the triple protected UAA derivative 36. Acidic deprotection with HCl however turned out to be challenging, as cleavage of the N-O bond was observed over time as judged by LC-MS whilst not all protection groups were removed. TFA or Lewis acid (ZnBr_2) treatment resulted in similar degradation patterns. Also, side-chain variation via introduction of an ortho-methyl group on the *p*- NO_2 phenol ring led to an even accelerated decay of the N-O bond indicating that positive inductive effects further destabilise

the N-O bond. UAA 37 was additionally purified by RP-HPLC to obtain a few milligrams, which was sufficient to perform initial reaction studies. The UAA appeared to be stable in various aqueous environments, such as PBS and even in *E. coli* culture at 37°C over two days (Fig. SI 4.8). However, amide formation with even 100x excess of KAT o.n. at 37°C remained slow as only starting material besides product was still detected. (Fig. 4.14b).

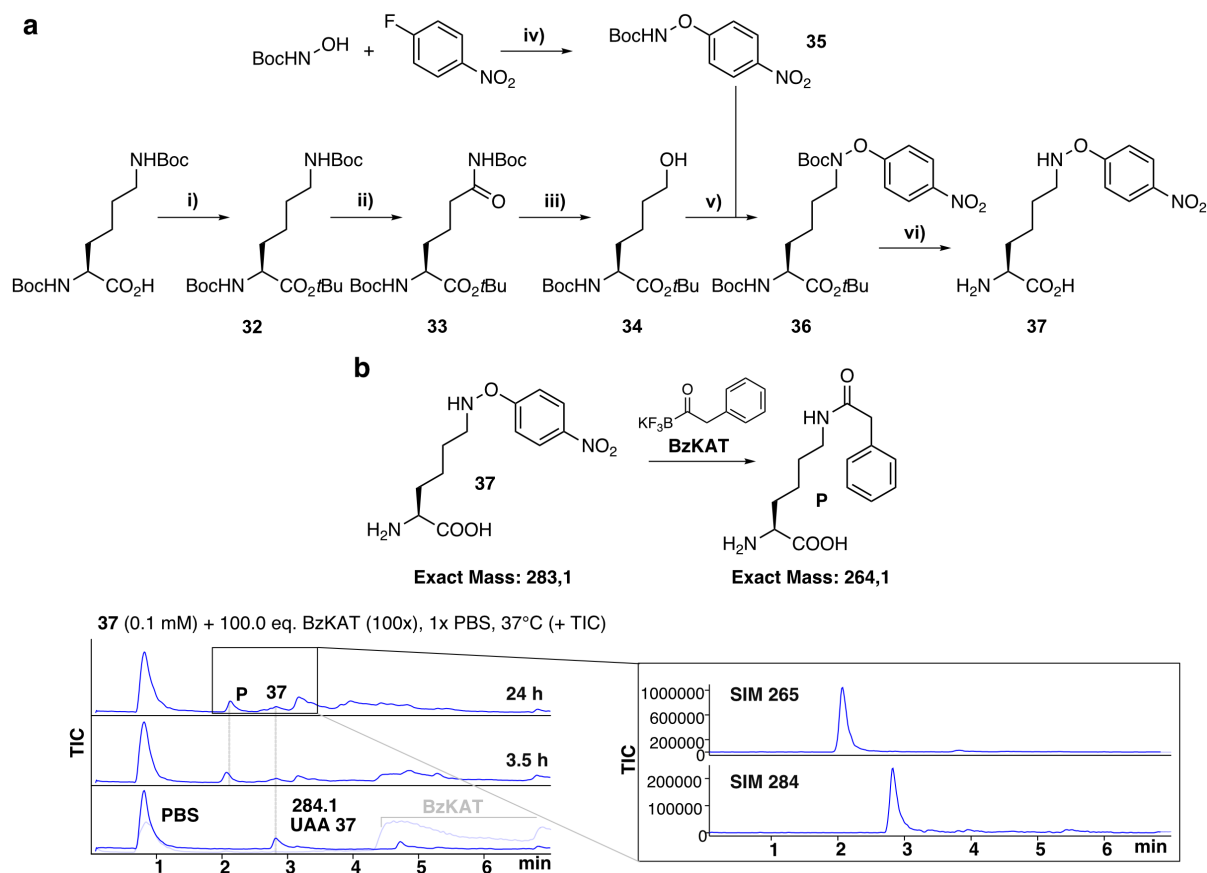


Fig. 4.14 | Synthesis and KAT reactivity of *p*-nitrophenoxyl lysine. (a) Synthesis route, i) Boc₂O (1.4 eq.), DMAP (0.40 eq.), *t*BuOH, 30°C, o.n., 86%; ii) NaBrO₃ (2.5 eq.), RuO₂ hydrate (~10 % wt), 35°C, 39 h, 47%; iii) NaBH₄ (1.6 eq.), *i*PrOH/H₂O, o.n., r.t., 74%; iv) KOH (1.1 eq.), 60°C, 3d, 75%; v) DIAD (1.1 eq.), PPh₃ (1.1 eq.), 35 (1.1 eq.), Ar atm, 0°C→r.t., 5h, 74%; vi) HCl/dioxane, RP-HPLC, 7%. (b) Time-dependent BzKAT (100.0 eq.) reaction profile of UAA 37 (0.1 mM). Depicted are the positive ion mode as well as SIM of the starting material and product after 24 h reaction.

The ongoing sluggish reaction kinetics might imply that the KAT ligation threshold lies just between the pK_a values of *p*-nitrophenol and that of carboxylic acids or specific influence of the aryl side-chain in the transition state process, which decelerates the amide bond forming process via an unknown mechanism. Due to the challenging deprotection step combined with the slow KAT ligation kinetics, this class of UAA was not examined any further.

4.9 Conclusion and outlook

Site-specific introduction of bioorthogonal functionalities into POIs represents a powerful route towards defined protein studies (cf. chapter 1-3). Specifically, generation of stable, yet unnatural junctions, induced via external physical triggers or by means of chemoselective

reactions facilitate detection and isolation of protein conjugates within their dynamic and native environment.

However, devising bioorthogonal chemistries which generate native motifs and/or PTMs on proteins in a general fashion would greatly enhance protein productions and studies in their native constitutions without the knowledge of their conjugating enzymes and need of additional control experiments. As outlined in chapter 4.2, chemistries fulfilling this criterion already exist in form of chemoselective amide bond forming reactions. The overall favourable amide bond features are reflected in the bulk amount present in nature, far beyond from solely being the connecting element of amino acid building blocks. Besides the N-terminus of proteins, lysine side-chains remain the only other primary amine source and thus are mainly subjected to additional amidation reactions, when incubated with appropriate acyl donors and conjugating enzymes. Carboxylic acid activation is a common theme in various metabolic pathways, thus providing readily activated synthons for lysine acylations. Indeed, many catalytic enzymes become frequently acylated with substrates from respective pathways they are involved in.³¹²⁻³¹³ This may be regarded as the most direct form of an activity-controlled feedback mechanism where the various acylation statuses act as metabolic sensors to varying metabolite concentrations. Histones represent another heavily acylated protein class. Gene expressions, which correlate with the chromatin compaction state, by virtue need to be controlled in a highly dynamic fashion in order to respond to swiftly changing environments (such as metabolic fluctuations).^{278, 313-314} The dynamic chemistry behind attachment and removal of versatile signal motifs on amines are unmatched and highly suited to the demanded criteria in this regard. While amides are chemically difficult to break, erasers readily provide the proper environment in their catalytic centre for electrophilic amide bond activation and cleavage via (indirect) attack of a water molecule (Zn²⁺ or NAD⁺ dependent).²⁷⁷ Despite the various chemical properties of the acyl side-chains (small hydrophobic and hydrophilic, charged, whole proteins), all acyl PTMs have the removal of the positively charged amino group of lysine in common. In an extreme form, modifications may also lead to complete charge inversion as exemplified with succinylation on core histone regions. This is associated with relaxation of the chromatin compaction state due to induced charge-charge repulsion with the acidic DNA backbone and facilitated access of gene regions for transcription.³¹⁵

Amide bond formation is a thermodynamically unfavoured process (Fig 4.15a). In the standard reaction pathway, the additional energy input to overcome the ligation barrier is unilaterally placed in the acyl fragment (Fig. 4.15b). This however comes with the cost of hydrolytic instability of the activated carboxylic acids as frequently experienced in chemical reactions. In chemoselective amide bond reactions, this issue is solved via bipartite distribution of the additional ligation energy at both ligating groups (Fig. 4.15c-f). For instance, some portion of the activation energy is stored in hydroxylamines, which is expressed in higher reactivity compared to amines (α -nucleophilicity). This allows for the reduction of the energy level in KATs compared to activated acyls, which results in greater aqueous stability as frequently observed throughout KAT ligation experiments. Nonetheless, the total energy sum of both reactants remains sufficient to overcome the activation barrier for amide bond formation to

proceed under ambient conditions. However, in contrast to organic chemical reaction development, where on demand regulation of reaction kinetics allows for flexible functional group design, bioorthogonal surrogates intrinsically need to bear not only sufficient energy levels but also good stability profiles to withstand constant thermal (up to 37°C), metabolite and biocatalytic influx from cells, at least during aaRS selections and protein expressions.

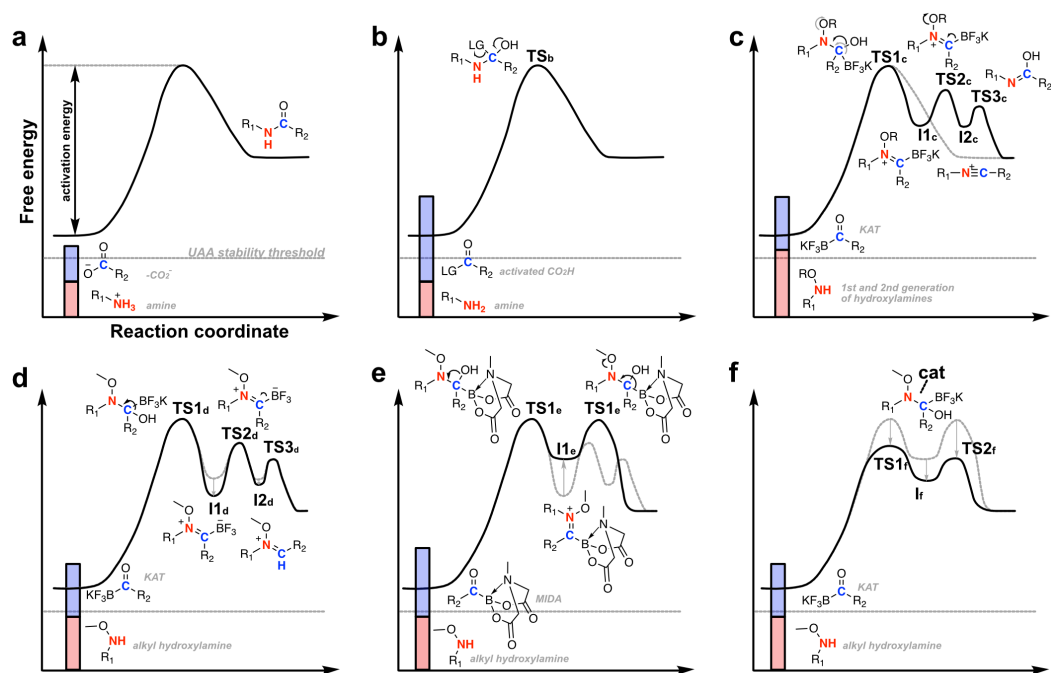


Fig. 4.15 | Proposed reaction profile of the various amide bond forming reactions. (a) The energetic contributions of amines (red bar) and carboxylic acids (blue bar) are insufficient to promote the thermodynamically unfavoured amide bond reaction. In the ‘classic’ pathway **(b)**, the total energetic contribution towards bond formation is placed in the acyl part (as illustrated by a higher blue bar). In standard KAT reactions **(c)**, bipartite division of the required ligation energy leads to higher energy levels of the amine surrogate, while the acyl mimic has reduced baseline intrinsic energy compared to **(b)**. **TS1_c** may undergo, as previously proposed,²⁹⁸ two pathways (highlighted by differently coloured arrows): The direct amidation pathway involves concerted release of both leaving groups (light grey arrows), while water/OH⁻ is eliminated in the iminium pathway first (**I1_c**). Collapse of the BF₃K group (**TS2_c**) leads to a nitrilium species **I2_c**, which under addition of OH⁻ leads to the amide bond formation. The energy profile of the 1st and 2nd generation hydroxylamine UAAs exceed the maximal UAA stability threshold, resulting in decomposition over time. This may be resolved via substitution to an alkyl residue **(d)**. Test reaction with KAT (cf. Fig. 4.13b and c) resulted in species **I1_d**, which appeared highly stable over time. This is due to the stabilising zwitterionic nature of **I1_d**, which results in a decreased energy level compared to **I1_c** (cf. ref. ³⁰³) and greater energy level difference compared to **TS2_d**. Amide bond formation possibly needs to overcome **I2_d**, as this species could be detected (cf. Fig. 4.13c). Reaction of alkyl hydroxylamines with MIDAs **(e)** destabilises the ionic intermediate **I1_d** (-> **I1_e**) via elimination of the zwitterion. The stability of MIDAs in chemoselective reaction environments however needs to be determined in greater detail. Alternatively, KAT reaction with alkyl hydroxylamines may occur via suitable catalysts due to reduction of the activation energies of the hypothetical **TS1/2_f**. This would allow the application of either stable reactants.

It appears that the ‘carbamate’ and ‘carboxylic acid’ hydroxylamines, besides being reactive towards KATs via the proposed pathways (Fig. 4.15c), exceed the critical stability threshold, which leads to decomposition in bacteria over time. Alkyl hydroxylamines in turn remain stable under the above-mentioned conditions, but too unreactive to overcome the transition states (Fig. 4.15d). KAT reaction and ionic iminium intermediate **I1_d** stabilisation reduces its intermediate energy level, which results in an apparent increase of the energy barrier

difference towards the transition state TS1_d or TS2_d and slow and incomplete amide bond reaction, respectively. This is readily solved when alkyl hydroxylamines are reacted with MIDAs, which sufficiently destabilise the intermediate to overcome the TS1_e state barrier (Fig. 4.15e).³⁰³ Furthermore, reported minor instabilities in acidic aqueous buffer hint towards a higher energy level compared to KATs, which possibly influences kinetics of the first half of the amide bond ligation as well.

Alike enzymes, small molecule catalysts accelerate reactions via reduction of the activation energy barrier (Fig. 4.15f). Thus, implementing a three-component amide ligation scheme with a suited catalyst would allow to introduce biologically stable amine surrogates during aaRS selections and protein expressions. For instance, while azides and KATs are both inert in organic and aqueous environments and unreactive towards each other, addition of HBF₄·Et₂O as a Lewis acid catalyst in MeCN readily promotes amide bond formation with moderate yields.³¹⁶ It was speculated, that similar reactions may occur when alkyl hydroxylamines and KATs are incubated with alike small molecule catalysts. Initial ligations of methylhydroxylamine UAA 39 and BzKAT with water stable Lewis acids Sc(OTf)₃ and Yb(OTf)₃ (ref. ³¹⁷) however showed no rate acceleration compared to samples without catalysts (Fig. SI 4.9).

Overall, the main bottleneck of translating 'batch' chemistries towards biological applications lies in functional group stabilities in cellular milieu, whilst still exhibiting meaningful reactivity. This was readily challenged with the attempted implementation of 'carbamate' and 'carboxylic acid' hydroxylamines, which appeared highly reactive towards KATs but unstable in bacteria. Alkyl and aryl sidechains of hydroxylamines derivatives with their respective pK_a values with their incomplete reaction profile represent the limits for KAT reactions to be performed.

Thus, future UAA designs need to match the energetic 'sweet spot' of maximal reactivity whilst not exceeding the lability threshold. A good indication for reactivity modulation is provided with the pK_a value of amine leaving groups. Acidity of alkyl alcohols may be enhanced with EWGs and UAA synthesis is mainly determined by starting material availability within an appropriate synthesis scheme. For instance, while glycolonitrile exhibits a strong -I effect as apparent from the pK_a value of 11.8 and an appropriate cyanomethyl bromide precursor is commercially available, acid lability of the nitrile prevents UAA synthesis when followed by the established route. Likewise, steric hindrance of starting materials such as HFIP (pK_a = 9.3) may require advanced reaction condition screening and further adjustments of the implemented synthesis scheme.

The other approach regards reactivity modulation on the acyl surrogate site. As already outlined above, utilising the MIDA ligation strategy allows to reduce the energy level of the amine surrogate and to install robust alkyl hydroxylamines into proteins. To what extent MIDAs or alike structures are stable in bioorthogonal reaction set-ups remains to be determined. However, short exposure exclusively during the ligation event suggests greater chemical flexibility compared to the amine part.

Likewise, implementation of a three-component reaction scheme would enable introduction of both, stable amine and carboxylic acid mimics. Finding an appropriate, water stable catalyst

within the proper reaction framework remains nonetheless challenging. Initial screen of inorganic Lewis acids did not provide promising results, urging for broader screening of catalyst (classes) and ligation conditions.

Once the optimal amide forming ligation system with an appropriate aaRS is found, diverse native-like protein acylation statuses may be readily produced via GCE and KAT/MIDA ligation. Ongoing reports regarding diverse and improved KAT and alike surrogate syntheses schemes³¹⁸⁻³²⁴ suggest that the generation of tailor-made acyl-PTM residues is a feasible task.

Chapter 5

Experimental Part

5 Experimental Part

5.1 General remarks - chemistry

All chemicals and solvents were obtained from commercial suppliers (Acros Organics, Carbolution, Fisher Scientific, Iris Biotech, Sigma Aldrich, VWR) and used without further purification unless otherwise stated. Technical grade pentane and DCM were distilled prior use. Thin-layer chromatography (TLC) was performed on Merck Millipore silica gel 60 F-254 plates. The developed silica plates were visualized by UV light (254 nm) and/or staining with ninhydrin or potassium permanganate. Flash column chromatography used for product purification was performed on silica gel 60 (230-400 mesh).

NMR spectra were recorded on Bruker AVHD300 (300 MHz for ^1H -NMR, 75 MHz for ^{13}C -NMR), Bruker AVHD400 (400 MHz for ^1H -NMR, 101 MHz for ^{13}C -NMR) or Bruker AVANCE 500 UltraShield™ spectrometer (500 MHz for ^1H -NMR, 125 MHz for ^{13}C -NMR).

Chemical shifts (δ), reported in ppm, are referenced to the residual proton solvent signals as reported in ref. ³²⁵. Coupling constants (J) are reported in Hertz (Hz) while peak multiplicities are described as follows: s (singlet), d (doublet), t (triplet), bs (broad singlet), p (pentet), dd (doublet of doublets), td (triplet of doublets), m (multiplet).

Small Molecule LC-MS were performed on an Agilent Technologies 1260 Infinity LC-MS system with a Phenomenex Aeris™ Peptide XB-C18 column (100 x 2.1 mm, 3.6 μm) coupled to a 6310 Quadrupole spectrometer. The solvent system consisted of MQ H₂O + 0.1% FA as buffer A and MeCN + 0.1% FA as buffer B. Protein LC-MS was carried out on a Jupiter C4 column (2 x 150 mm, 5 μm). Protein mass was calculated by deconvolution within the Chemstation software (Agilent Technologies). Theoretical protein masses were calculated using ProtParam and were manually corrected for the mass of the UAA. Samples were analysed by UV absorbance at 193, 254 and 280 nm followed by both positive and negative ESI-mode.

Reverse phase HPLC purification was carried out on a Shimadzu LC-20AT Prominence system with a Phenomenex Luna C18, 5 μm (4.6 x 250 mm) column. The solvent system consists of buffer A and buffer B, which were used without filtration.

5.2 General remarks – biochemistry and molecular biology

5.2.1 Organisms

All bacterial experiments were performed in *E. coli* DH10 β strain (origin: NEB) with the following genotype: $\Delta(\text{ara-leu})$ 7697 *araD139 fhuA Δ lacX74 galK16 galE15 e14- ϕ 80dlacZ Δ M15 recA1 relA1 endA1 nupG rpsL (Str^R) rph spoT1 Δ (mrr-hsdRMS-mcrBC). *E. coli* Δ hcp (KEIO collection) was a gift from the Jung Lab, LMU Munich.*

All mammalian experiments were performed in HEK293T cells (origin: homo sapiens, kidney tissue; provided by Feige Lab, TUM, Germany) with the following specification: highly transfectable human cell line derived from HEK293, which expresses a mutant version of the SV40 antigen.

5.2.2 Cloning and sequencing

Plasmids (list, cf. 5.2.12) were either already available in the Lang lab or provided by the Sieber lab (TU Munich). Mutations, insertions or deletions were cloned via Gibson assembly or site-directed ligase independent mutagenesis (SLIM)³²⁶ following recommended protocols, buffers and enzymes (NEB). Sequences were cloned into respective backbones via standard restriction cloning protocols or Gibson assembly (NEB, enzymes and protocols used as recommended). The required oligonucleotides were designed using Snapgene software (v.4.2.6) and NEB T_m calculator (online tool). Primers were purchased lyophilized and desalted from Sigma Aldrich. Polymerase chain reactions were carried out according to Table 5.1 and 5.2:

input	Volume [μ l]
plasmid (appr. 20-100 ng/ μ l)	1
dNTP mix (10 mM stock, NEB)	1
primers (100 μ M stock)	2x 2.5
5x Q ₅ reaction buffer	10
Q ₅ polymerase (NEB)	0.5
ddH ₂ O	32.5
	$\Sigma = 50$

Table 5.1 | Pipetting scheme for general PCR using Q₅ polymerase.

Annealing temperatures (T_a) were calculated with NEB T_m calculator. Usually, T_a was set 1°C below the recommended value.

Step	Temperature	Time	Number of cycles
Initial denaturation	98°C	60s	1
Denaturation	98°C	10s	32
Annealing	$T_a - 1^\circ\text{C}$	30s	
Elongation	72°C	30s/kb	
Final elongation	72°C	60s	1
storage	16°C	∞	1

Table 5.2 | Protocol for general PCR using Q₅ polymerase.

Cloned plasmids were verified by DNA sequencing (Genewiz, samples prepared as recommended by the company with appropriate sequencing primers).

Cloning analyses were performed via size-separating agarose gel electrophoresis. DNA samples were mixed with 6x DNA loading dye (NEB) and loaded onto a 1% w/v agarose gel (in 1x TAE buffer, dissolution by heating in microwave, followed by addition of SERVA DNA Stain Clear G; Serva). Electrophoreses were run in 1x TAE buffer at constant voltage (typically 120-

135 V) for 20-40 min, depending on the expected DNA sizes. High range DNA ladder (Jena Biosciences) was added to estimate various DNA fragment sizes.

5.2.3 DNA isolation and purification

Transformed *E. coli* DH10 β cells (cf. 5.2.7) or picked single colony from agar plates were grown in LB (6 ml, o.n., 37°C, 1x appropriate antibiotics concentration) and plasmids were isolated using the peqGOLD Plasmid Mini Prep KIT (VWR) according to the manufacturer's protocol. Elution of the DNA from the columns was performed with ddH₂O (20-55 μ l). Concentration was measured photometrically (230, 260, 280 nm) and plasmids were stored at -20°C until further use.

5.2.4 Antibiotics - stock solutions and working concentrations

Antibiotic (1000x stock solutions) were dissolved, sterile filtered (0.2 μ m), aliquoted and stored at -20°C until further use (Table 5.3). Aliquots of antibiotics were thawed prior dilution into expression media to final 1x working concentrations.

Antibiotic	c [mg/ml] stock solution	solvent	c [μ g/ml] working concentration
Ampicillin (Amp)	100	H ₂ O	100
Chloramphenicol (Cam)	50	96% (v/v) EtOH	50
Kanamycin (Kan)	50	H ₂ O	50
Tetracycline (Tet)	17	70% (v/v) in EtOH	17

Table 5.3 | List of antibiotic stock solutions and working concentrations.

5.2.5 Preparation of chemically competent DH10 β *E. coli* cells

E. coli DH10 β cryostock (NEB) was inoculated in LB and grown o.n. at 37°C (200 rpm). The culture was diluted into LB (100 ml) to an OD₆₀₀ of approximately 0.05 and grown at 37°C (200 rpm). At an OD₆₀₀ of ~0.5-0.6, the culture was chilled on ice for 30 min followed by harvesting at 4000 rpm, 4°C, 10 min. The pellet was carefully resuspended with cold 100 mM CaCl₂ (50 ml, 4°C) and chilled on ice for 30 min. The cultures were pelletized again, supernatant discarded, carefully resuspended with 100 mM CaCl₂ (10 ml), followed by incubation on ice for 30 min. After centrifugation, the pellet was resuspended with 100 mM CaCl₂ (10 ml) and glycerol (2 ml) was added. The suspension was gently mixed, chilled on ice for 5 min and aliquoted (50-100 μ l) into precooled (4°C) 1.5 ml Eppendorf tubes. Cells were flash frozen in liquid N₂ and stored at -80°C until further use.

5.2.6 Preparation of electrocompetent competent DH10 β *E. coli* cells

E. coli DH10 β cryostock (NEB) was inoculated in LB and grown o.n. at 37°C (200 rpm). The culture was diluted in 200 ml LB to an OD₆₀₀ of ~0.05 and grown at 37°C (200 rpm). At an OD₆₀₀ ~0.5-0.6, the culture was chilled on ice for 30 min, followed by harvesting at 4000 rpm, 4°C, 10 min (split into 4x 50ml). The pellet was carefully resuspended with 4x cold 10% glycerol (40 ml each, 4°C). The cultures were pelleted again, supernatant discarded and carefully resuspended again (cold 10% glycerol, 4x 10ml, merge into 2x 20ml). The harvesting step was repeated and pellets resuspended in cold 10% glycerol (2x 10ml). After obtaining the pellets, cells were resuspended in 2x 1ml cold 10% glycerol, the suspensions merged and aliquoted (~100 μ l) into precooled (4°C) 1.5 ml Eppendorf tubes. Cells were flash frozen in liquid N₂ and stored at -80°C until further use.

All freshly made competent cells were checked for contamination by streaking on LB-agar plates bearing all standard antibiotics as well as on a LB-agar without antibiotics. Plates were incubated o.n. at 37°C, followed by check for survival.

5.2.7 Heat shock transformation of chemically competent *E. coli* cells

Aliquot(s) of chemically competent *E. coli* cells were thawed on ice, followed by addition of target plasmid(s) of choice (typically 20-200 ng). Cells were chilled on ice for 5 min again and then incubated at 42°C for 45s (heat shock). Cells were immediately chilled on ice for 5min followed by addition of SOC (final volume of 1 ml). Afterwards, transformed cells were incubated for 1 h at 37°C (200 rpm), followed by inoculation in growth media (LB or 2YT) supplemented with the respective 1x plasmid resistance(s) and growth o.n. at 37°C. Alternatively, 50-100 μ l were plated out on LB-agar plates and incubated o.n. at 37°C.

5.2.8 Electroporation of electrocompetent *E. coli* cells

Aliquot(s) of electrocompetent *E. coli* cells were thawed on ice, followed by addition of target plasmid(s) of choice (typically 100-300 ng). Cells were chilled on ice for 5 min, followed by transfer into an electroporation cuvette (2 mm gap, Fisher scientific). Electroporation was performed (ECM399 electroporation system BTX) employing an exponential decay wave pulse with a 2.5 kV setting. Cells were immediately rescued with SOC or SOB (final volume of 1 ml) and incubated at 37°C for 1 h (200 rpm). Transformed cells were inoculated in growth media (LB or 2YT) supplemented with the respective 1x plasmid resistances and grown at 37°C, o.n.

5.2.9 Maintenance and passaging of HEK293T cells

HEK293T cells were cultivated in p100 dishes (d = 10 cm, total volume: 10ml) and high glucose Dulbecco's Modified Medium (DMEM, Gibco™) supplemented with 10% FBS (Gibco™) and

500 µg/ml PenStrep (Sigma) at 37°C in a humidified incubator with 5% CO₂. Cells were passaged either every 2nd (1:5 dilution) or every 3rd day (1:10 dilution). For this, medium was removed and adherent cells washed with 10 ml Dulbecco's phosphate buffered saline (Sigma). PBS was removed and cells were treated with 1 ml 1x Trypsin-EDTA solution (0.05%, Gibco™) at 37°C for 1 min in the cell culture incubator. Cells were detached by gentle shaking of the dish, followed by resuspension with 9 ml pre-warmed DMEM and diluted accordingly (1:5 or 1:10, 10 ml total volume) into a new p100 dish with fresh medium.

5.2.10 Seeding of HEK 293T cells

Depending on the well type (6 well plate: 5x10⁵/2ml per well; p100 dish: 3.5x10⁶/10ml per dish), defined numbers of cells were seeded 24 h prior transfections. Cells were detached and harvested as described in the previous section. Cell numbers were determined using a Countess II FL Automated Cell counter system (ThermoFisher) and according to the manufacturer's recommendation. Appropriate cells numbers were calculated, diluted with fresh DMEM and evenly distributed to the well(s).

5.2.11 Transfection of HEK293T cells

Transient transfections of HEK293T were performed using polyethyleneimine (PEI, 1mg/ml), Gibco™ OptiMEM™ and appropriate plasmids in defined ratios according to the table 5.4. All components were mixed and incubated for 15' at r.t. prior dropwise addition to the well(s). During plasmid mix incubation, media in wells were replaced with fresh DMEM supplemented with the UAA at desired concentration (pH was neutralized with equal amount of NaOH or TFA).

Type of dish	Total amount of DNA per well [µg]	OPTIMEM per well [µl]	PEI (1 mg/ml) per well [µl]
6 well	2	200	9
10cm dish	10	1000	30

Table 5.4 | Transfection table for protein expression in HEK293T.

Approximately 40-44h post-transfection, cells were washed and harvested for further analyses as described in the appropriate sections.

5.2.12 List of Plasmids

	Code	Plasmid	Resistance	Usage/comment
Chapter 1		pET301_hClpP		from Sieber lab, TUM; cloning
		pRK5SV40_hClpP		from Sieber lab, TUM; cloning
	TAN-BP-hClpP-SH1-7	pPylt_(-)56aa-hClpP_TAG-His ₆	Tet ^R	expression in <i>E. coli</i> DH10β, plasmids cloned for all hClpP TAG variants (M88; D92; G123; G124; V125; S181; K261 = SH1-7)
	D171	pBK_Mb_D171	Kan ^R	PylRS; Y271M, L274A, C313A; for dK incorporation in <i>E. coli</i> DH10β
	TAN-MP-hClpP-SH1-7	pEF1_(+)56aa-hClpP-TAG-HA_4xPylt	Amp ^R	expression in HEK293T, plasmids cloned for all hClpP TAG variants (SH1-7)
	TAN-MP-D171	pEF1_Mm_D171_4xPylt	Amp ^R	PylRS; Y306M, L309A, C348A; for dK incorporation in HEK293T
Chapter 2	TAN-BP-sfGFPN149TAG-H ₆	pPylt_sfGFP-N149TAG-His ₆	Tet ^R	cloning; expression in <i>E. coli</i> DH10β
	TAN-BP-GSTF52TAG-H ₆	pPylt_GST-F52TAG-His ₆	Tet ^R	expression in <i>E. coli</i> DH10β
	MW-P145	pBAD-Duet_Rab1b-Q67A-R69TAG-His ₆ -StrepII-TEV-DrrA(16-352)	Amp ^R	expression in <i>E. coli</i> DH10β
	D1	pBK_Mb_D1	Amp ^R	PylRS; Y271M, L274G, C313A; for dKA incorporation in <i>E. coli</i> DH10β
	D4	pBK_Mb_wtRS	Kan ^R	PylRS; for Bock incorporation in <i>E. coli</i> DH10β
	MF3	pBK_Mb_MF3	Amp ^R	PylRS; Y271G, C313V; for dKA incorporation in <i>E. coli</i> DH10β
	MW-S59	pEVOL_Mm_wtRS_pPylt	Cam ^R	PylRS; for Bock incorporation in <i>E. coli</i> DH10β
	MW-S60	pEVOL_Mm_TEMPOH-I_pPylt	Cam ^R	PylRS; Y306M, L309A, C348A; for dK incorporation DH10β
	MW-S65	pEVOL_Mb_MF3_pPylt	Cam ^R	PylRS; Y271G, C313V; for dKA incorporation in <i>E. coli</i> DH10β
	TAN-MP-sfGFPN149TAG-H ₆	pEF1_sfGFP-N149TAG-His ₆ _4xPylt	Amp ^R	cloning; expression in HEK293T
	TAN-MP-D4	pEF1_Mm_wtRS_4xPylt	Amp ^R	wtPylRS; for Bock incorporation in HEK293T
	TAN-MP-MF3	pEF1_Mm_MF3_4xPylt	Amp ^R	PylRS; Y306G, C348V; for dKA incorporation in HEK293T
Chapter 3	D7	D7_L_pYOBB2_extended_Pylt	Cam ^R	PylRS negative selection
	D65	D65_pREP_Pylt	Tet ^R	PylRS positive selection

Table 5.5 I List of plasmids used within this thesis.

5.2.13 List of buffers and media

5.2.13.1 Buffers

buffer	components	concentration
5x Q5 reaction buffer	Tris base, pH 8.8 KCl MgCl ₂ (NH ₄)SO ₄ Triton X-100 BSA	100 mM 50 mM 5 mM 50 mM 0.5% (v/v) 0.5 mg/ml
His lysis buffer	Tris base, pH 8.0 NaCl Imidazole PMSF DNase I complete™ protease inhibitor	20 mM 300 mM 30 mM 500 μM spatula tip 1 tablet/100 ml
His wash buffer	Tris base pH 8.0 NaCl Imidazole	20 mM 300 mM 30 mM
His elution buffer	Tris base, pH 8.0 NaCl Imidazole	20 mM 300 mM 300 mM
His mammalian lysis buffer	Tris-HCl, pH 8.0 NaCl Triton-X-100 Imidazole Mammalian protease inhibitor (VWR)	50 mM 150 mM 1% (v/v) 20 mM 1x
His mammalian elution buffer	Tris-HCl, pH 8.0 NaCl Triton-X-100 Imidazole	50 mM 150 mM 1% (v/v) 300 mM
4x SDS-PAGE resolving buffer	Tris-HCL, pH 8.8 SDS	1.5 M 0.4 % (w/v)
4x SDS-PAGE stacking buffer	Tris-HCL, pH 6.8 SDS	500 mM 0.4% (w/v)
4x Laemmli loading buffer	Tris-HCL, pH 6.8 Glycerol SDS Bromophenol blue β-mercaptoethanol	240 mM 40% (v/v) 8% (w/v) 0.4% (w/v) 5% (v/v)
20 x MES SDS-PAGE running buffer	MES, pH 7.3 TRIS, pH 7.3 SDS	1 M 1 M 2% (w/v)
Immunoblot transfer buffer	Tris base, Glycine SDS Methanol	48 mM 39 mM 1.3 mM 20% (v/v)
50x TAE buffer	Tris-HCl, pH 8.8 Acetic acid EDTA	2M 0.9 M 127 mM
10x PBS	NaCl KCl Na ₂ HPO ₄	1.37 M 27 mM 100 mM

	KH ₂ PO ₄	18 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%
5x hybridisation buffer (SLIM buffer)	Tris-HCl, pH 9.0	125 mM
	NaCl	750 mM
	EDTA	100 mM

Table 5.6 | List of buffers and their compositions.

5.2.13.2 Growth media

Medium	Components	Concentration
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
	MgCl ₂	0,96 g/l
	NaCl	0.5 g/l
	KCl	0.186 g/l
	glucose	20 mM
Autoinduction medium (single recipes, see table below)	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
	1 M MgSO ₄	2 ml/l
	40 % Glucose	1.25 ml/l
	5,000x trace metals	200 µl/l
	1M Nicotinamide	6 ml/l
LB agar	Tryptone	10 g/l
	Yeast extract	5 g/l
	NaCl	5 g/l
	Agar-agar	15 g/l

Table 5.7 | List of growth media and their compositions.

5.2.13.3 Stock solutions for protein expression in autoinduction media³²⁷

stock	components	concentration
5% Aspartate pH 7.5	aspartic acid (pH adjusted with NaOH)	5% (w/v)
25x 17AA mix ³²⁷	glutamic acid sodium salt, aspartic acid, lysine-HCl, arginine-HCl, histidine-HCl-H ₂ O, alanine, proline, glycine, threonine, serine, glutamine, asparagine-H ₂ O, valine, leucine, isoleucine, tryptophan, methionine	5 g/l (each)
50x M	Na ₂ HPO ₄ KH ₂ PO ₄ NH ₄ Cl Na ₂ SO ₄	1.25 M 1.25 M 2.5 M 0.25 M
5x M9 Salts	Na ₂ HPO ₄ · 7 H ₂ O KH ₂ PO ₄ NH ₄ Cl NaCl	64 g/l 15 g/l 5 g/l 2.5 g/l
5000x Trace Metals ³²⁷	CaCl ₂ MnCl ₂ ZnSO ₄ CoCl ₂ CuCl ₂ NiCl ₂ Na ₂ MoO ₄ Na ₂ SeO ₃ H ₃ BO ₃ FeCl ₃	200 μM 100 μM 100 μM 20 μM 20 μM 20 μM 20 μM 20 μM 20 μM 500 μM

Table 5.8 | List of stock solutions for autoinduction media and their compositions.

5.2.13.4 Stocks for aaRS selections³²⁸

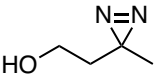
stock	components	concentration*
1 M MgSO ₄	MgSO ₄ x 7 H ₂ O	61.6 g/250 ml
50% (v/v) glycerol		250 ml/250 ml
100 mM CaCl ₂	CaCl ₂ x 2 H ₂ O	3.68 g/250 ml
850 mM NaCl		24.9 g/500 ml
30 mM L-leucine		0.99 g/250 ml
5 mM FeSO ₄	FeSO ₄ x 7 H ₂ O	69.5 mg/50 ml
TB-buffer	KH ₂ PO ₄ K ₂ HPO ₄	5.78 g/250 ml 31.4 g/250 ml
100 mM L-Phenylalanine		0.66 g/40 ml
Agar	agar-agar 4x PBS TB-buffer	3.75 g/250 ml 0.25 ml/250 ml 0.25 ml/250 ml

Table 5.9 | List of stock solutions for aaRS selection and their compositions. *All ingredients were dissolved in MQ H₂O and sterilised by autoclaving.

5.3 General methods for Chapter 2

5.3.1 Chapter 2 - chemical synthesis

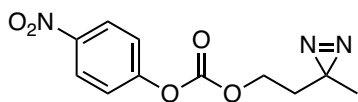
Synthesis of 2-(3-methyl-3H-diazirin-3-yl)ethan-1-ol (1)

 To 4-Hydroxy-2-butanone (4 g, 45 mmol, 1.0 eq.) on ice was added NH₃ (7 M in MeOH, 35 ml). After 3 h, hydroxylamine-O-sulfonic acid (5.65 g, 50 mmol, 1.1 eq.) was added and the reaction was stirred overnight, gradually allowed to warm up to r.t. The mixture was filtered through Celite and the filter cake was washed with MeOH. The solvent was removed and the crude intermediate was taken up in MeOH (40 ml) and Et₃N (8 ml) and stirred on ice. I₂ was added portion-wise until the reaction mixture maintained a dark yellow/brown colour. After 3 h, the solvent was removed and taken up in Et₂O (150 ml). The organic phase was washed with 1 M HCl (80 ml). The aqueous phase was extracted again with Et₂O (150 ml). The combined organic phase was washed with 20% w/v Na₂S₂O₃ (100 ml) and brine (100 ml). The organic phase was dried over Na₂SO₄, filtered and removed under reduced pressure to yield a dark yellow oil (1.88 g, 18.78 mmol, 42%). The crude product **1** was used without further purification.

¹H NMR (300 MHz, CDCl₃) δ = 1.07 (s, 3H), 1.63 (t, J = 6.3 Hz, 2H), 3.52 (t, J = 6.3 Hz, 2H).

The chemical shifts are according to literature.²³³

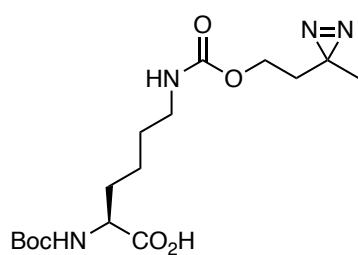
Synthesis of 2-(3-Methyldiazirin-3-yl)-ethyl (4-nitrophenyl) carbonate (2)

 To a solution of **1** (1.88 g, 18.78 mmol, 1.0 eq.) in DCM (100 ml) on ice was added 4-nitrophenyl chloroformate (4.54 g, 22.54 mmol, 1.2 eq.) and pyridine (1.82 ml, 22.54 mmol, 1.2 eq.) and the reaction was stirred overnight, gradually allowed to warm up to r.t. The reaction mixture was concentrated under reduced pressure and purified by flash chromatography (10→15% EtOAc in pentane) to obtain the final product **2** as a yellow oil (3.08 g, 11.61 mmol, 62%).

¹H NMR (300 MHz, CDCl₃): δ = 1.12 (s, 3H), 1.80 (t, J = 6.4 Hz, 2H), 4.25 (t, J = 6.4 Hz, 2H), 7.40 (d, J = 9.2 Hz, 2H), 8.29 (d, J = 9.2 Hz, 2H).

¹³C NMR (75 MHz, CDCl₃): δ = 20.0, 23.7, 33.9, 64.4, 122.0, 125.5, 145.6, 152.5, 155.6.

Synthesis of N²-(tert-butoxycarbonyl)-N⁶-((2-(3-methyl-3H-diazirin-3-yl)ethoxy)carbonyl)-L-lysine (3)



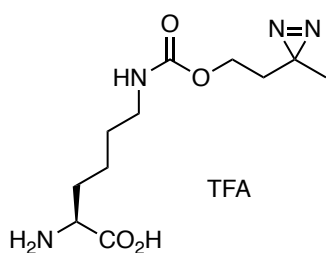
BocHN
CO₂H

To a solution of **2** (3.08 g, 11.61 mmol, 1.0 eq.) in dioxane (30 ml) was added N α -Boc-Lysine (3.43 g, 13.93 mmol, 1.2 eq.) and Et₃N (3.22 ml, 23.22 mmol, 2.0 eq.) and the reaction was stirred overnight at r.t. The reaction mixture was concentrated under reduced and purified by flash chromatography (2% MeOH in DCM \rightarrow 5% MeOH in DCM +0.5% AcOH) to yield the final product **3** as a colourless oil (2.58 g, 6.93 mmol, 60%).

The chemical shifts are according to literature.²³³

LC-MS (m/z): calcd. for C₁₆H₂₈N₄O₆ [M-H⁺] 371.2; found: 371.1.

Synthesis of N⁶-((2-(3-methyl-3H-diazirin-3-yl)ethoxy)carbonyl)-L-lysine TFA salt (4)



H₂N
CO₂H

TFA

To a stirred solution of **3** (4.84 g, 13.00 mmol, 1.0 eq.) in DCM (28 ml) on ice was added TFA (7 ml) and H₂O (1 ml). After 3 h, the solvent was removed, the crude product precipitated in ice-cold Et₂O and pelleted by centrifugation. This was repeated twice and the final product was lyophilised to obtain the final product **4** as a white powder (4.63 g, 13.54 mmol, 96%, TFA salt).

The chemical shifts are according to literature.²³³

LC-MS (m/z): calcd. for. C₁₁H₂₀N₄O₄ [M+H⁺] 273.2; found: 273.2.

UAA stock solutions (100 mM) were stored in 100 mM TFA in H₂O, filtered (0.2 μ M) and stored at -20°C prior use.

5.3.2 Chapter 2 - biological experiments

5.3.2.1 Cloning of hClpP constructs

Plasmids for bacterial expressions: hClpP without the N-terminal 56 amino acid signal sequence and bearing a C-terminal StrepII-tag was amplified from pET301_hClpP vector (provided by Sieber Lab, TU Munich) and cloned into the pPylt backbone (containing *Mb*_tRNA_{CUA} and an arabinose-inducible promoter) with a C-terminal His₆-tag using Gibson cloning protocol as recommended (NEB). Respective amber codon (TAG) positions were introduced via SLIM.

Plasmids for mammalian expression: hClpP bearing the N-terminal 56 amino acid mitochondrial signal recognition motif and a C-terminal FLAG-tag was amplified from pRK5SV40_hClpP vector (provided by Sieber Lab, TU Munich) and subcloned into a pET17 helper plasmid. FLAG-tag was replaced with an HA-tag via SLIM and respective TAG positions were introduced thereafter via SLIM. hClpP(TAG)-HA constructs were amplified with primers bearing XbaI and BamHI restriction sites prior and after the gene and cloned into pEF1_POI_4xPylt backbone²³⁶ using standard restriction cloning protocol.

5.3.2.2 Expression and purification of recombinant (amber suppressed) hClpP

Expression of hClpP-TAG-His₆ were performed as follows: *E. coli* DH10 β cells were chemically co-transformed with pBK_*Mb*_PylRS_D171 (which encodes the synthetase for dK, mutations: Y271M, L274A, C313A, Kan^R) and pPylt_(-)56aa-hClpP-TAG-His₆ (Tet^R), SOC rescued (1h, 37°C) and cultured in LB medium o.n. at 37°C with 1x antibiotic strengths. The o.n. culture was diluted into 2xYT to an OD₆₀₀ of ~0.05 with 1x antibiotics and incubated at 37°C, 200 rpm. At approximately OD₆₀₀ of ~0.3, dK was added to a final concentration of 1 mM (for amber suppressed proteins). At approximately OD₆₀₀ of ~0.6, protein expression was induced with arabinose (0.02% w/v final concentration). After 3 h (for wt hClpP) or 18 h (for amber suppressed hClpP), cells were harvested by centrifugation (4200 rpm, 15', 4°C). For SDS-PAGE analysis, 1 ml of culture was pelleted, resuspended in 1x Laemmli loading buffer (normalised according to OD₆₀₀ values; per OD = 1, 100 μ l of 1x loading buffer), boiled at 95°C for 10 min, pelleted again (max. speed, 10 min) and supernatant were analysed via 15% SDS-PAGE.

Protein purification was based on ref ²²⁷. In detail, pellets were resuspended in hClpP lysis buffer (50 mM Tris-HCl, pH 7.5 (rt), 100 mM KCl, 10% glycerol) and sonicated on ice. The lysate was cleared by centrifugation (13500 g, 35', 4°C) and supernatant was subjected to HisTrap purification on GE Äkta system (hClpP lysis buffer supplemented with imidazole, 20mM (A) and 300 mM (B); gradient A→B). Appropriate fractions were analysed via 15% SDS-PAGE, pooled, concentrated/rebuffered with an Amicon Ultra-4 10K MWCO centrifugal filter units (Millipore) and SEC buffer (20 mM HEPES, pH 7 (6°C), 100 mM NaCl). Samples were aliquoted, flash frozen and stored at -80°C until further use.

5.3.2.3 Expression and purification of recombinant *E. coli* ClpX

This part was performed by Thomas F. Gronauer, Sieber Lab, TU Munich and the protocol was adapted from the respective PhD Thesis (2021) with permission.

E. coli ClpX was expressed as described before.³²⁹ In brief, *E. coli* ClpX was overexpressed in *E. coli* (DE3) Rosetta 2 with a N-terminal His₆-TEV construct. For this, LB-Medium was inoculated with overnight cultures, incubated at 37 °C with constant shaking and overexpression was induced at OD₆₀₀ = 0.8 by addition of IPTG (1:2000 dilution, 0.5 M stock solution in H₂O). After incubation for 20 h at 25 °C at 200 rpm bacteria were harvested by centrifugation (6000 g, 10 min, 4 °C), washed with 30 mL PBS buffer (140 mM NaCl, 10 mM Na₂HPO₄, 2.7 mM KCl, 1.8 mM KH₂PO₄, pH 7.6) and again centrifuged at 15000 g for 10 min at 4 °C. Lysis was performed in *EcClpX* lysis buffer I (50 mM HEPES, 300 mM KCl, 1 mM DTT, 10 mM imidazole, 5 mM MgCl₂, 15% (v/v) glycerol, pH 7.6) by sonication. Lysate was cleared by centrifugation (38700 g, 30 min, 4 °C) and loaded on a pre-equilibrated HisTrap™ HP affinity column. After washing with 50 mL *E. coli* ClpX lysis buffer II (50 mM HEPES, 300 mM KCl, 1 mM DTT, 40 mM imidazole, 15% (v/v) glycerol, pH = 7.6) protein was eluted with 20 mL elution buffer (50 mM HEPES, 300 mM KCl, 1 mM DTT, 300 mM imidazole, 15% (v/v) glycerol, pH 7.6). Fractions containing protein were pooled and 1 mM EDTA and 2 mg/mL TEV-protease were added and incubated overnight at 10 °C under constant agitation. Completion of His₆-tag cleavage was confirmed by LC-MS and crude protein was purified by size exclusion chromatography in *EcClpX* lysis buffer I without imidazole. Fractions containing *EcClpX* were pooled and concentrated using a centrifugal filter with a cut-off of 30 kDa.

5.3.2.4 Protein assembly analysis via size-exclusion chromatography

His-purified protein samples (wt-hClpP-H₆ and all amber mutants) were subjected to SEC using a Superdex™ 75 10/300 (GE Healthcare) with the SEC buffer. Raw data of the elution profiles were processed using Excel and OriginPro16G. Peak intensities (280 nm absorption) from 5-20 min retention time were considered and all signals were normalised to the respective highest signal intensity.

5.3.2.5 hClpP protease activity assay

This part was performed by Thomas F. Gronauer, Sieber Lab, TU Munich and the protocol was adapted from the respective PhD Thesis (2021) with permission.

E. coli ClpX recognizes peptides C-terminally tagged with a short amino acid sequence and forms a functional complex with *H. Sapiens* ClpP. It is therefore suitable as a *H. Sapiens* ClpX substitute for *in vitro* protease activity assays. To determine residual protease activity of mutated ClpXP, cleavage of *ssrA*-tagged GFP was monitored by the decrease of fluorescence signal. For this, 59 µL of enzyme buffer mix (0.2 µM hClpP₁₄, 0.4 µM eClpX₆; 10x ATP-

regeneration mix: 40 mM ATP, 160 mM creatine phosphate, 200 U/mL creatine phosphokinase in 25 mM HEPES, 200 mM KCl, 1 mM DTT, 5 mM MgCl₂, 10% (v/v) glycerol, pH 7.6) was aliquoted in triplicates in a transparent 96-well plate and incubated for 15 min at 37 °C. Fluorescence measurement was started after addition of 1 μL ssrA-eGFP (final concentration: 0.4 μM) on a *Tecan Infinite F200 Pro* (excitation wavelength: 485 nm, emission wavelength: 535 nm). The slope of decreasing fluorescence signal in the time interval $t = 0 - 100$ s was determined via linear regression with *Microsoft Excel*. Residual proteolytic activity was defined as the fraction of activity of mutant ClpXP and wild type ClpXP.

5.3.2.6 Mammalian cell culturing and photocrosslinking

HEK293T cell culturing was performed as previously described.¹⁶⁰ Transfections were either performed in 6-well plates or in 10 cm dishes for proteomics experiments (triplicates, +/-UV). 24 h after seeding, media was replaced with fresh DMEM containing 2 mM dK (pH was neutralised with 1 M NaOH). Transfection mixtures with plasmids pEF1_(+)56aa_hClpP_TAG_HA_4xPylt and pEF1_Mm_D171_4xPylt (plasmid ratio 3:1) were prepared as described above. 40-44 h post-transfection, cells were washed 3x with 1x PBS (1x well volume). The plates or dishes (in 1x PBS, 0.5x of the usual well volume) were either placed on a cooling pad and irradiated for 15min with a 15 W, 365 nm UV lamp (Vilber, VL-215.L, +UV) prior scraping or harvested (-UV) immediately. Cells were pelleted at 700 g, 4°C, 15 min, supernatant discarded, flash frozen in liquid nitrogen and stored at -20°C until further use.

5.3.2.7 Western Blot

Qualitative western blots: Samples (-/+ UV) were lysed by freeze/thaw cycles (3x). Lysates were cleared and supernatant fractions were kept for further analysis. BCA assays (Pierce™ BCA Protein Assay Kit, Thermo Scientific) were performed to ensure equal loading prior SDS-PAGE loading. Western blot was performed via standard semidry blot procedure (with immublot transfer buffer) on nitrocellulose membranes (0.2 μm, Amersham™, Protran™, GE Healthcare Life Sciences). Membrane was blocked with 5% skim milk powder in 1x TBS-T (0.1%, 1h, r.t.), followed by incubation with primary anti-HA antibody (rabbit, 1:5000, provided by Itzen Lab, UKE Hamburg) in 1% skim milk powder in 1x TBS-T (0.1%), o.n. at 4°C. Membranes were washed 5x with 1x TBS-T, followed by incubation with the secondary goat anti-rabbit-IgG-HRP antibody (1:40000, provided by Itzen Lab, UKE Hamburg) for 1 h at r.t. After washing with 1x TBS-T (3x), blots were developed using Amersham ECI™ Prime Western Blotting Detection Reagent (GE Healthcare).

Semi-quantitative Western blot: wt^{ClpP} HEK293 and KO^{ClpP} HEK293T cells (gift from Trifunovic Lab, University of Cologne) were harvested at 60-70 % confluency, washed twice with ice-cold 1x PBS and lysed in RIPA-buffer (Sigma) supplemented with 1x protease inhibitor cocktail tablet (Roche) on ice for 30'. Supernatant was collected, BCA assays were performed and 6 μg proteome were loaded onto NuPAGE™ 4-10%, Bis-Tris Protein-Gel (ThermoFisher™).

Western blot was performed using iBlot2 Dry Blotting System (ThermoFisher™; NC membrane, programme P0). Membrane was blocked with 5% skim milk powder in 1x TBS-T (0.05%), 1h, r.t. followed by incubation of primary-AB in 1% skim milk powder in 1x TBS-T (0.05%), o.n. at 4°C. Membrane was washed 3x with 1x TBS-T (0.05%) followed by incubation of secondary-AB (1:10000) for 1 h at r.t. After additional washing (3x) blots were divided according to various MWs of desired protein bands and imaged as described above. Band intensity analysis was performed using ImageJ, referenced to their respective Hsc70 signal intensities and statistical analysis was done using OriginPro16G. Primary antibodies: Hsc70 (mouse, 1:5000, Santa Cruz Biotechnology), hClpP (rabbit, 1:5000, Abcam), DBT (mouse, 1:2000, Santa Cruz Biotechnology), HSPE1 (mouse, 1:5000, Santa Cruz Biotechnology); Secondary antibodies: anti-mouse or rabbit-IgG (1:10000, Santa Cruz Biotechnology).

5.3.2.8 Rotenone treatment of HEK293T expressing hClpP

HEK293T cells in 10cm dishes with dK containing DMEM were transfected with pEF1_(+)56aa_hClpP_D92TAG_HA_4xPylt or pEF1_(+)56aa_hClpP_K261TAG_HA_4xPylt and pEF1_Mm_D171_4xPylt as described in 5.3.2.5. 40 h post-transfection, medium was replaced with fresh DMEM supplemented with rotenone (final concentration: 1 µM). After 6 h, cells were harvested and processed for +/- UV-light treatment and proteomic measurements as described in 5.3.2.6.

5.3.2.9 Experiments for proteomic analyses of hClpP-crosslinked proteins

This part was performed by Thomas F. Gronauer, Sieber Lab, TU Munich and the protocols were adapted from the respective PhD Thesis (2021) with permission.

HA-Antibody Enrichment

HEK cell pellets were thawed one ice, resuspended in 1 mL lysis buffer (50 mM Tris/HCl, 150 mM NaCl, 1 mM MgCl₂, 1% (v/v) 4-Nonylphenyl-polyethylene glycol (NP-40), 5% (v/v) glycerol, pH = 7.4 at 4 °C) and incubated for 30 min at 4 °C. Afterwards membranes and cell debris were separated by centrifugation (21,100 g, 20 min, 4 °C). For protein enrichment, 30 µL HA-antibody agarose beads suspension (A2095, Merck) per replicate in LoBind Eppendorf tubes were equilibrated with 1 mL wash buffer (50 mM Tris/HCl, 150 mM NaCl, 0.05% (v/v) NP-40, 5% (v/v) glycerol, pH 7.4 at 4 °C) and centrifuged at 1,000 g for 1 min at 4 °C. 500 µL cytosolic fraction were incubated with equilibrated beads for 3 h at 4 °C on a rotating wheel. Afterwards beads were centrifuged at 1,000 g for 1 min at 4 °C. The supernatant was discarded and beads were washed two times each with 1 mL wash buffer and two times each with 1 mL basic buffer (50 mM Tris/HCl, 150 mM NaCl, pH 7.4 at 4 °C) to remove unspecifically bound proteins.

Digestion and reduction of enriched proteins was performed by addition of 25 μ L digestion buffer I (5 ng/ μ L trypsin (in 50 mM acetic acid), 50 mM Tris/HCl, 2 M urea, 1 mM dithiothreitol (DTT, freshly prepared and diluted 1:1000 from 1 M stock) pH 8.0) and incubated for 30 min. Afterwards, alkylation of free cysteines and further digestion was conducted by addition of 100 μ L digestion buffer II (5.5 mM iodoacetamide (IAA, freshly prepared and diluted 1:100 from 550 mM stock), 50 mM Tris/HCl, 2 M urea, pH 8.0, following incubation for 16 – 19 h at 25 °C with continuous mixing at 650 rpm in a thermoshaker.

Digestion was stopped by adjusting the pH to 2 - 3 via addition of 17.5 μ L 10% (v/v) formic acid in water. Peptide solutions were desalted by stage tips with two-layered C18 material (*SDC-XC, 3M*) according to a published protocol.³³⁰ In brief, per replicate two layers of C18 material were packed into a 200 μ L pipet tip and washed with 70 μ L MeOH, 70 μ L 80% (v/v) MeCN, 0.5% (v/v) FA and 3x 70 μ L 0.5% (v/v) FA (centrifugation: 1,000 g, 1-2 min, RT). Samples were loaded and centrifuged at 1,000 g for 1-2 min at r.t. Beads were washed with 70 μ L 0.5% (v/v) FA and the washing solution was also loaded onto the C18 material. Samples were desalted by washing two times with 70 μ L 0.5% (v/v) FA. Elution of peptides was conducted with 2x 30 μ L 80% (v/v) MeCN, 0.5% (v/v) FA into fresh LoBind Eppendorf tubes followed by speedvac assisted solvent removal. Samples were stored at -80 °C until further use.

In situ trapping and enrichment

Cells were thawed on ice, resuspended in 1 mL lysis buffer (50 mM Tris/HCl, 150 mM NaCl, 1 mM MgCl₂ · 6H₂O, 1% (v/v) 4-Nonylphenyl-polyethylene glycol (NP-40), 5% (v/v) glycerol, pH = 7.4 at 4 °C) and incubated for 30 min at 4 °C. Afterwards membranes and cell debris were separated by centrifugation (21,100 g, 20 min, 4 °C). For protein enrichment, 30 μ L monoclonal HA-antibody agarose beads suspension (A2095, isotype IgG1, Merck) per replicate in LoBind Eppendorf tubes were equilibrated with 1 mL wash buffer (50 mM Tris/HCl, 150 mM NaCl, 0.05% (v/v) NP-40, 5% (v/v) glycerol, pH = 7.4 at 4 °C) and centrifuged at 1,000 g for 1 min at 4 °C. 500 μ L cytosolic fraction were incubated with equilibrated beads for 3 h at 4 °C on a rotating wheel. Afterwards beads were centrifuged at 1,000 g for 1 min at 4 °C. The supernatant was discarded and beads were washed two times each with 1 mL wash buffer and two times each with 1 mL basic buffer (50 mM Tris/HCl, 150 mM NaCl, pH = 7.4 at 4 °C) to remove unspecifically bound proteins.

Digestion, reduction, alkylation

Digestion and reduction of enriched proteins was performed by addition of 25 μ L digestion buffer I (5 ng/ μ L trypsin (in 50 mM acetic acid), 50 mM Tris/HCl, 2 M urea, 1 mM DTT (freshly prepared and diluted 1:1,000 from 1 M stock) pH = 8.0) and incubated for 30 min. Afterwards, alkylation of free cysteines and further digestion was conducted by addition of 100 μ L digestion buffer II (5.5 mM IAA (freshly prepared and diluted 1:100 from 550 mM stock), 50

mM Tris/HCl, 2 M urea, pH = 8.0, following incubation for 16 – 19 h at 25 °C with continuous mixing at 650 rpm in a thermoshaker (Thermomixer comfort 5355, Eppendorf).

Desalting and sample preparation

Digestion was stopped by adjusting the pH to 2 - 3 via addition of 17.5 µL 10% (v/v) FA in water. Peptide solutions were desalted by stage tips with two-layered C18 material (SDC-XC, 3M) according to a published protocol.³³⁰ In brief, per replicate two layers of C18 material were packed into a 200 µL pipet tip and washed with 70 µL MeOH, 70 µL 80% (v/v) MeCN, 0.5% (v/v) FA and 3x 70 µL 0.5% (v/v) FA (centrifugation: 1,000 g, 1-2 min, r.t.). Samples were loaded and centrifuged at 1,000 g for 1-2 min at RT. Beads were washed with 70 µL 0.5% (v/v) FA and the washing solution was also loaded onto the C18 material. Samples were desalted by washing two times with 70 µL 0.5% (v/v) FA. Elution of peptides was conducted with 2x 30 µL 80% (v/v) MeCN, 0.5% (v/v) FA into fresh LoBind Eppendorf tubes followed by speedvac assisted solvent removal. Samples were stored at -80 °C until further use. For LC-MS/MS measurement sample peptides were dissolved in 25 µL 1% (v/v) aqueous FA and sonicated 3x 5 min with centrifugation for 1 min at 17,000 g in between. Dissolved peptides were filtered using 0.22 µm Ultrafree-MC® centrifugal filters (UFC30GVNB, Merck) pre-equilibrated with 300 µL 1% (v/v) FA. Filtrates were transferred to MS-vials prior to LC-MS/MS analysis.

Data acquisition on Orbitrap Fusion

Peptide samples from substrate enrichment experiments were analyzed with an UltiMate 3000 nano HPLC system (Dionex) using an Acclaim C18 PepMap100 (75 µm ID x 2 cm) trap and an Acclaim PepMap RSLC C18 (75 µm ID x 50 cm) separation column in EASY-spray setting coupled to an Orbitrap Fusion (Thermo Fisher Scientific Inc.). 1-5 µL peptide sample (see Table 5.10) were loaded on the trap and washed with 0.1% (v/v) TFA, then transferred to the analytical column (buffer A: H₂O with 0.1% (v/v) FA, buffer B: MeCN with 0.1% (v/v) FA, flow as indicated in Table 5.10, gradient: to 5% buffer B in 7 min, from 5% to 22% buffer B in 105 min, then to 32% buffer B in 10 min, to 90% buffer B in 10 min and hold at 90% buffer B for 10 min, then to 5% buffer B in 0.1 min and hold 5% buffer B for 9.9 min) and ionized by nanospray ionization (NSI) with capillary temperature of 275 °C. Spray voltage was applied as indicated in Table 5.6. Orbitrap Fusion was operated in a TOP speed data dependent mode. Master scan acquisition was carried out in the orbitrap at a resolution of R = 120,000, an AGC target of 2.0e5 in a scan range of 300 - 1500 m/z and a maximum injection time of 50 ms. Monoisotopic Peak Determination was set to "Peptide" and dynamic exclusion was enabled with dynamic exclusion duration set to 60 s with a mass tolerance (low/high) of 10 ppm. Precursors with a charge state of 2 - 7 and intensities greater than 5.0e3 were submitted to fragmentation by higher collisional dissociation (HCD). Isolation of precursors was performed in the Quadrupole with an isolation window of 1.6 m/z. Detection was carried out in the ion trap to an AGC target of 1.0e4 with first mass set to 120 m/z, Ion Trap Scan Rate set to "Rapid" and "Inject Ions for

All Available Parallelizable Time” set to true. Maximum Ion Injection Time was set to 100 ms and peptide fragments were generated by HCD with a collision energy of 30%.

Sample name	Injection volume [μ l]	flow [μ l/min]	Spray voltage [kV]
SH1 dk (M88TAG)	5	0.3	1.87
SH2 dk (D92TAG)	5	0.4	1.77
SH3 dk (G123TAG)	5	0.4	1.77
SH4 dk (G124 TAG)	5	0.4	1.77
SH5 dk (V125TAG)	5	0.4	1.77
SH6 dk (S181TAG)	5	0.4	1.77
SH7 dk (K261TAG)	5	0.4	1.77
SH2 dk rot	1	0.4	1.77
SH7 dk rot	1	0.4	1.77

Table 5.10 | Table Mass spectrometry instrument parameters for enriched samples.

Data processing

MS-data were processed using Andromeda search engine of MaxQuant (MQ) Software (version 1.6.0.1). For identification of peptides, MS/MS spectra were searched against a Uniprot reference proteome (taxon identifier 9606, canonical version, without isoforms, downloaded 2017/07/18). MaxQuant settings were largely set on default with Label-free quantification (LFQ) enabled and Trypsin as digestion enzyme with a maximum of 2 missed cleavages and a minimum peptide length of 7 amino acids. Methionine oxidation and N-terminal acetylation were set as variable modifications with a maximum number of 5 modifications per peptide and carbamidomethylation of cysteines was used as fixed modification. Second peptide search was enabled as well as match between runs with a matching time window of 0.7 min and an alignment time window of 20 min. For identification false discovery rate on the PSM, protein and site level were set to 0.01.

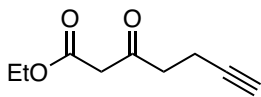
Statistical MS-data analysis

MaxQuant data were further statistically processed with Perseus software (version 1.6.2.3). ProteinGroups table was loaded into the program with LFQ intensities of all replicates as main column. Rows were filtered based on categorical columns omitting values that met the criterium of “reverse”, “potential contaminant” and “only identified by site”. After log₂ transformation, matrices were further filtered based on valid values. Proteins with less than 50% of valid values in LFQ intensity compared to all replicates were excluded. Remaining missing values were inserted by imputation from a normal distribution (width: 0.3, down shift: 1.2 - 1.8). Categorical annotation of UV-irradiated (+UV) and non-treated (-UV) samples was performed before data were analyzed by two-sided student’s t-test with -UV as single control group with Benjamini-Hochberg false discovery rate correction (FDR 0.05). Additionally, GO-term annotations with GOBP, GOMF and GOCC, downloaded from Uniprot were added.

5.4 General methods for Chapter 3

5.4.1 Chapter 3 - chemical synthesis

Synthesis of ethyl 3-oxohept-6-ynoate (13)

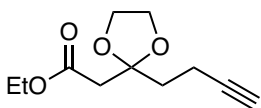


To a solution of lithium diisopropylamide (30.7 ml from 2M in THF/heptane/benzene, 61.5 mmol, 2.0 eq.) on an ice/salt bath (-16°C) was added ethylacetoacetate (4.0 g, 30.7 mmol, 1.0 eq.) and reaction was stirred on the cooling bath for 30 min. Propargyl bromide (80% in toluol, 3.41 ml, 30.7 mmol, 1.0 eq.) was added and the mixture was stirred for 2h, while maintaining the cooling bath temperature. Reaction was quenched with acetic acid (3.51 ml, 61.5 mmol, 2.0 eq.), Et₂O (50 ml) and H₂O were added and organic phase isolated. The aqueous phase was extracted twice with Et₂O (180 ml each), the combined organic phases washed with brine (80 ml), dried over Na₂SO₄, filtered and concentrated under reduced pressure. The crude mixture was purified by flash column chromatography (15% Et₂O in pentane) to yield product **13** as a pale-yellow oil (3.7 g, 22.0 mmol, 36%).

¹H-NMR (400MHz, CDCl₃): δ = 1.26 (t, *J* = 7.1 Hz, 3H), 1.95 (t, *J* = 2.7 Hz, 1H), 2.42-2.50 (m, 2H), 2.80 (t, *J* = 7.2 Hz, 2H), 3.45 (s, 2H), 4.18 (q, *J* = 7.1 Hz, 2H).

The chemical shifts are according to literature.²⁶⁹

Synthesis of ethyl 2-(2-(but-3-yn-1-yl)-1,3-dioxolan-2-yl)acetate (14)

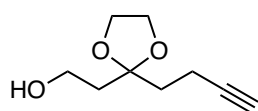


To a solution of **13** (3.7 g, 22.0 mmol, 1.0 eq.) in toluene (200 ml) were added ethyleneglycol (2.73 g, 44.0 mmol, 2.0 eq.) and p-toluenesulfonic acid (419 mg, 2.2 mmol, 0.1 eq.) and the reaction was stirred for 6 h under reflux on a Dean-Stark-apparatus. The solvent was removed and the crude product purified (10→15% Et₂O in pentane) to obtain product **14** as a colourless oil (2.15 g, 10.1 mmol, 46%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.26 (t, *J* = 7.2 Hz, 3H), 1.92 (t, *J* = 2.7 Hz, 1H), 2.08-2.16 (m, 2H), 2.25-2.35 (m, 2H), 2.65 (s, 2H), 3.95-4.02 (m, 4H), 4.15 (q, *J* = 7.2 Hz, 2H).

The chemical shifts are according to literature.²⁶⁹

Synthesis of 2-(2-(but-3-yn-1-yl)-1,3-dioxolan-2-yl)ethan-1-ol (15)

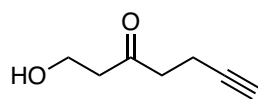


To a solution of **14** (2.14g, 10.1 mmol, 1.0 eq.) in Et₂O (50 ml) at 0°C was added LiAlH₄ (10.1 ml from 1M solution in THF, 1.01 mmol, 1.0 eq.) and the reaction was stirred for 3h at 0°C. H₂O and 10% NaOH_{aq.} (10 ml each) were added and crude reaction was filtered through Celite. Et₂O (150 ml) was added and phases separated. The aqueous phase was extracted twice with Et₂O (150 ml each) and the combined organic phases was washed with brine (100 ml), dried over Na₂SO₄, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography (40% Et₂O in pentane→Et₂O) to obtain product **15** as a colourless oil (1.15 g, 6.77 mmol, 67%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.90-1.97 (m, 5H), 2.23-2.31 (m, 2H), 2.42 (bs, 1H), 3.72-3.79 (m, 2H), 3.96-4.03 (m, 4H).

The chemical shifts are according to literature.²⁶⁹

Synthesis of 1-hydroxyhept-6-yn-3-one (16)

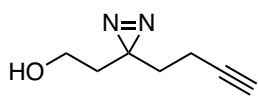


To a solution of **15** (1.15 g, 6.76 mmol, 1.0 eq.) in acetone (20 ml) at r.t. was added *p*-toloulsulfonic acid monohydrate (322 mg, 1.69 mmol, 0.25 eq.) and reaction was stirred for 4.5 h at r.t. NaHCO_{3satd.} (50 ml) and Et₂O were added and phases separated. The aqueous phase was extracted twice with Et₂O (70 ml each) and combined organic phases was washed with brine (70 ml). The organic phase was dried over Na₂SO₄, filtered, concentrated under reduced pressure and purified by flash column chromatography (20% Et₂O in pentane→Et₂O) to obtain **16** as a colourless oil (544 mg, 4.31 mmol, 64%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.96 (t, *J* = 2.7 Hz), 2.15 (bs, 1H), 2.47 (td, *J* = 7.4 Hz, *J* = 2.7 Hz, 2H), 2.70 (m, 4H), 3.87 (t, *J* = 5.7 Hz, 2H).

The chemical shifts are according to literature.²⁶⁹

Synthesis of 2-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)ethan-1-ol (17)



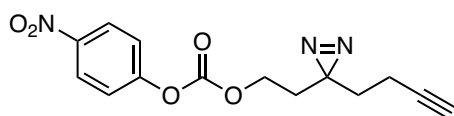
Starting material **16** (540 mg, 4.28 mmol, 1.0 eq.) was stirred in NH₃ (7M in MeOH, 25 ml) for 3 h on ice. Hydroxylamine-O-sulfonic acid (533 mg, 4.71 mmol, 1.1 eq.) was added and the reaction was stirred o.n., allowing to warm up to r.t. The suspension was filtered through Celite, solvent removed and taken up in MeOH (75 ml). Et₃N (1.5 ml) was added followed by I₂ on ice until the reaction solution maintained a brown colour. After 1h at 0°C, the solvent was removed and taken up in Et₂O (80 ml). The organic phase was washed with 1 M HCl_{aq} (80 ml). The aqueous phase was

extracted 3x with Et₂O (80 ml) and combined organic phases was washed with Na₂S₂O₃satd. (80 ml) and brine (80 ml). The organic phase was dried over Na₂SO₄, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography (Et₂O) to obtain product **17** as a yellow oil (199 mg, 1.44 mmol, 34%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.65-1.75 (m, 5H), 1.98-2.10 (m, 3H), 3.49 (t, *J* = 6.1 Hz, 2H).

The chemical shifts are according to literature.²⁶⁹

Synthesis of 2-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)ethyl (4-nitrophenyl) carbonate (18)

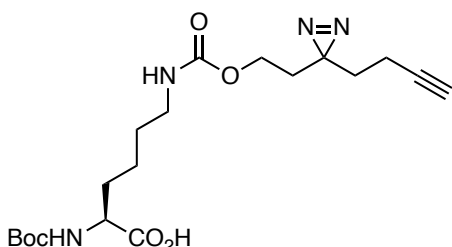


To a solution of **17** (197 mg, 1.43 mmol, 1.0 eq.) in DCM (20 ml) at 0°C was added *p*-nitrophenyl chloroformate (347 mg, 1.72 mmol, 1.2 eq.) and pyridine (139 μl, 1.72 mmol, 1.2 eq.) and reaction was stirred o.n., allowing to warm up to r.t. The crude reaction mixture was concentrated and purified by flash column chromatography (15% EtOAc in pentane) to obtain product **18** as a yellow oil (419 mg, 1.38 mmol, 97%).

¹H-NMR (400 MHz, CDCl₃): δ = 1.72 (t, 2H, *J* = 7.3 Hz), 1.90 (t, 2H, *J* = 6.3 Hz), 2.01 (t, 1H, *J* = 2.6 Hz), 2.06 (td, 2H, *J* = 7.3 Hz, *J* = 2.6 Hz), 4.20 (t, 2H, *J* = 6.3 Hz), 7.37-7.43 (m, 2H), 8.27-8.32 (m, 2H).

¹³C-NMR (101 MHz, CDCl₃): δ = 13.4, 26.2, 32.4, 32.5, 64.0, 69.6, 82.6, 122.0, 125.5, 145.7, 152.4, 155.6.

Synthesis of N⁶-((2-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)ethoxy)carbonyl)-N²-(tert-butoxycarbonyl)-L-lysine (19)



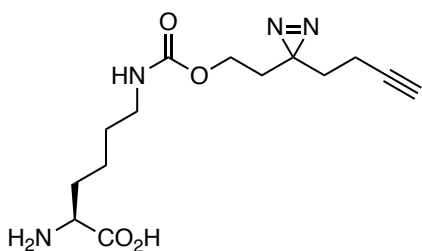
To a solution of **18** (419 mg, 1.38 mmol, 1.0 eq.) in DMF/dioxane (1/1, 8 ml each) were added Boc-lysine-OH (403 mg, 1.66 mmol, 1.2 eq.) and Et₃N (383 μl, 2.76 mmol, 2.0 eq.) and reaction was stirred o.n. at r.t. The solvent was removed and the crude reaction mixture was purified by flash column chromatography (2% MeOH in DCM → 2% MeOH in DCM + 0.5% AcOH) to obtain product **19** as a yellow oil (491 mg, 1.20 mmol, 87%).

¹H-NMR (400 MHz, CD₃OD): δ = 1.26-1.43 (m, 2H), 1.45 (s, 9H), 1.47-1.59 (m, 2H), 1.65 (t, *J* = 7.5 Hz, 2H), 1.68-1.75 (m, 2H), 1.75-1.90 (m, 2H), 2.03 (td, *J* = 2.7 Hz, *J* = 7.5 Hz, 2H), 2.27 (t, *J* = 7.5 Hz, 1H), 3.11 (t, *J* = 6.7 Hz, 2H), 3.93 (t, *J* = 6.3 Hz, 2H), 4.03-4.11 (m, 1H).

¹³C-NMR (101 MHz, CD₃OD): δ = 13.8, 24.1, 27.5, 28.7, 30.5, 32.5, 33.5, 41.4, 54.9, 60.6, 70.3, 80.5, 83.6, 158.2, 158.7, 176.3.

LC-MS (m/z): calcd. for. C₁₉H₃₀N₄O₆ [M+Na⁺] 433.2; found: 433.2.

Synthesis of N⁶-((2-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)ethoxy)carbonyl)-L-lysine (20)



To a solution of **19** in DCM (16 ml) and H₂O (1 ml) was added TFA (4 ml) and reaction was stirred for 3 h at r.t. The solvent was removed, H₂O was added and lyophilised. The obtained crude product was precipitated in Et₂O, solvent discarded and lyophilised again to obtain product **20** as a TFA salt. (380 mg, 0.99 mmol, 75%).

¹H-NMR (400 MHz, CD₃OD): δ = 1.40-1.50 (m, 2H), 1.50-1.60 (m, 2H), 1.65 (t, 2H, *J* = 7.4 Hz), 1.72 (t, *J* = 6.3 Hz, 2H), 1.76-1.95 (m, 2H), 2.04 (td, *J* = 2.6 Hz, *J* = 7.4 Hz, 2H), 2.28 (t, *J* = 2.6 Hz, 1H), 3.12 (t, *J* = 6.9 Hz, 2H), 3.53 (dd, *J* = 5.1 Hz, *J* = 7.1 Hz, 1H) 3.93 (t, *J* = 6.3 Hz, 2H).

¹³C-NMR (101 MHz, CD₃OD): δ = 13.8, 23.5, 27.5, 30.6, 32.0, 33.5, 33.7, 41.4, 56.2, 60.6, 70.3, 83.6, 158.7, 174.4.

LC-MS (m/z): calcd. for. C₁₄H₂₂N₄O₄ [M+H⁺] 311.2; found: 311.2.

UAA stock solutions (50 mM) were dissolved in 6/6/1 of DMSO/H₂O/1M NaOH, filtered (0.2 μM) and stored at -20°C prior use.

5.4.2 Chapter 3 - biological experiments

5.4.2.1 Protein expression and purification of sfGFP-N149TAG bearing dKA

In *E. coli*: *E. coli* DH10 β were chemically co-transformed with pPylt-sfGFP-N149TAG-His₆ and the respective dKA-RS plasmid (pBK_Mb_D1). After heat shock and SOC rescue, cells were grown o.n. at 37°C (200 rpm) in LB (6 ml) with 1x antibiotics (Kan^R, Tet^R). The preculture was diluted in LB (V_Σ = 20 ml) to an OD₆₀₀ ~0.05 and grown at 37°C (200 rpm) until the OD reached ~0.3. UAA 20/dKA was added to a final concentration of 1 mM. At an OD of 0.6, protein expression was induced by addition of arabinose (final concentration 0.02%) and culture were shaken o.n. (appr. 17 h) at 37°C (200 rpm). Cells were harvested by centrifugation (4200 rpm, 10 min, 4°C), supernatant discarded and taken up in His lysis buffer (18 ml). Cells were disrupted by sonication on ice and lysate was cleared by centrifugation (13500 g, 4°C, 30 min). The supernatant was incubated with 300 μ l NiNTA slurry (Jena Bioscience) on ice for 1 h with agitation. The slurry was transferred to a plastic column and washed with His washing buffer (5-10 CV). The protein was eluted with elution buffer (3 x 400 μ l) and subjected to further analyses (SDS-PAGE and LC-MS).

In HEK293T: 24 h after seeding of HEK293T cells in a 10 cm dish, the medium was replaced with fresh DMEM and dKA (1 mM, neutralised with equal amount of 100 mM TFA) and cells were transfected with pEF1_sfGFP-N149TAG-His₆-4xPylt and pEF1_MF3_4xPylT (which encodes the aaRS for dKA incorporation; Y306G, C348V) according to a 3:1 POI:aaRS ratio and transfection amounts listed in Table 5.4. 48 h post-transfection, cells were washed 1x PBS (2x 10 ml), followed by addition of 1 ml 1x PBS to the plate. Cells were scraped, pelleted (500 g, 4°C, 2x 2.5 min), supernatant discarded, resuspended with His mammalian lysis buffer (500 μ l) and incubated on ice for 30 min with vortexing after every 10 min. Supernatant was collected by centrifugation (max. speed, 4°C, 15 min), followed by incubation on NiNTA slurry (500 μ l) for 1 h on ice. Slurry was washed with His washing buffer (5-10 CV) followed by elution with His mammalian elution buffer. Protein fraction was collected and subjected to further analysis.

5.4.2.2 General crosslink, then click approach

In *E. coli*: sfGFP-N149TAG and GST-F52TAG amber mutants were cloned with a C-terminal His₆-tag in the pPylt vector (Tet^R) and combined with the following pBK_Mb_PylRSs plasmids for amber suppression with the following UAAs: wtRS (Kan^R) for Bock (100 mM stock solution in 100 mM NaOH), D171 PylRS (Y271M, L274A, C313A; Kan^R) for UAA 4/dK and MF3 PylRS (Y271G, C313V; Amp^R) for UAA 20/dKA. Rab1b(Q67A)-R69TAG/DrrA₁₆₋₃₅₂ constructs were cloned into a pBAD_Duet vector with Rab1b bearing a C-terminal His₆-tag and DrrA bearing a N-terminal StrepII-tag (Amp^R). The duet vector was combined with the following pEVOL_PylRS plasmids for amber suppression with the following UAAs: pEVOL_Mm_wtRS_Pylt for Bock,

pEVOL_Mm_TEMPOH-I_Pylt for UAA 4/dK and pEVOL_Mb_MF3_Pylt for UAA 20/dKA (all Cam^R). These constructs were cloned by Marie. v. Wrisberg (former Lang Lab, TU Munich; now Schneider Lab, LMU).

Co-transformation, inoculation and dilution into expression media were performed as described in chapter 5.3.2.2. After o.n. protein expression (~ 16 h) in autoinduction media (15 ml) with respective POIs/UAA aaRS plasmid combination and UAA (1 mM final concentration each) at 37°C, 5 ml of culture were pelleted and resuspended in 1x PBS (5 ml). OD₆₀₀ values of all cultures were measured and adjusted with 1x PBS to the same value (usually OD₆₀₀ of ~1) and volume (5 ml). 500 µl of the OD₆₀₀ adjusted cell suspension were transferred to a 6-well plate (2x on two wells for each UAA type) and irradiated with UV light for 15', while placed on a cooling pad (cf. chapter 5.3.2.6). UV-irradiated samples, together with UV-light untreated cells (2x 500 µl) were pelleted and stored at -20°C until further use. For WB analyses, one pellet of the +/-UV treated samples were boiled in 1x Laemmli buffer (80 µl) for 10 min at 95°C, centrifuged max. speed for 10' and loaded onto a 15% SDS-PAGE (13 µl). Qualitative WBs were performed as described in chapter 5.3.2.7 with immediate imaging after incubation of the primary antibody-HRP conjugate (1 h, r.t. or o.n. at 4°C, constant rolling) and washing with 1x TBS-T (0.1 %, 3x 10 ml, 5 min each). The following antibodies were used for detection of respective protein bands: His₆-tag - anti-His₆-peroxidase, 1:5000, Roche Applied Science; StrepII-tag – StrepMAB-classic, HRP conjugate, 1:5000, IBA GmbH, Göttingen, Germany.

For fluorophore labelling via CuAAC chemistry, +/-UV treated cell pellets were resuspended in 2/3 MeOH/H₂O (V_Σ=400 µl each) and lysed with glass beads (~50 mg, SigmaAldrich, G4649, acid-washed, <106 µm) under vigorous vortexing for 15 min. After centrifugation at full speed for 15 min at 4°C, the supernatants were carefully collected. To 81 µl of each supernatant were added: TAMRA-PEG-azide (1 µl of 5 mM stock solution in DMSO, 50 µM final concentration; SigmaAldrich, 760757), CuSO₄ (4µl of 50 mM stock solution in H₂O, 2 mM final concentration), sodium ascorbate (4 µl of 50 mM stock solution, 2 mM final concentration, always freshly prepared) and THPTA (10 µl of 10 mM stock solution in H₂O, 1 mM final concentration), vortexed and left reacting at r.t. for 2 h. Acetone (300 µl) was added and reaction mixtures were stored o.n. at -20°C. Precipitates were collected by centrifugation (max. speed, 4°C, 15 min; supernatant discarded), resuspended in MeOH (300 µl) and incubated for 1 h at -20°C. Precipitates were collected again (max. speed, 4°C, 15 min; supernatant discarded) and dried at 55°C until residual MeOH was evaporated. Laemmli buffer (1x, 20 µl) were added, samples boiled for 10 min at 95°C and 15 µl loaded for SDS-PAGE analysis. Fluorescence were detected on an ImageQuant LAS 4000 using the appropriate fluorescence filter.

In HEK293T: Cells were seeded in a 6-well plate 24 h prior transfection. Transfections were performed as described in chapter 5.2.11 and media was removed and replaced with fresh DMEM containing the UAA of interest (1 mM final concentration, pH neutralized with TFA or NaOH). The following plasmid mixtures were used: pEF1_sfGFP-N149TAG-His₆_4xPylt with (1) pEF1_Mm_wtRS_4xPylT for the incorporation of Bock; (2) pEF1_Mm_D171_4xPylT for the

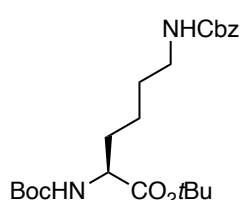
incorporation of UAA 4/dK; and pEF1_Mm_MF3_4xPyIT for the incorporation of UAA 20/dKA. Two wells were transfected for each UAA. 40-44 h post-transfection, the medium was removed and washed twice with 1x PBS (1 ml). Cells were placed in 1x PBS (500 μ l) and one well per UAA was scraped to serve as the -UV control. The remaining wells were irradiated with UV light for 15' on an ice pad and harvested thereafter. The cells were pelleted at 700 rpm for 15' at 4°C and supernatant removed. Pellets were flash frozen with liquid N₂ and stored at -20°C until further use. For WB analysis and click chemistry, pellets were resuspended in 1x PBS (200 μ l) and lysed by freeze/thaw lysis (3x). Supernatants were collected after 15' centrifugation at max. speed. BCA assay was performed to ensure equal loading. 5 μ g proteome (in 1x Laemmli buffer) were used for WB analysis, performed as described in 5.3.2.7.

For CuAAC chemistry, 100 μ g proteome were mixed with CuSO₄, sodium ascorbate, TAMRA-PEG-azide and THPTA with concentrations and amounts as described for bacterial labelling in the previous page. Work-up was performed as described above. 50 μ l of 1x Laemmli buffer was added to each probe, boiled and 15 μ l loaded for SDS-PAGE and gel fluorescence analysis.

5.5 General methods for Chapter 4

5.5.1 Chapter 4 – chemical synthesis

Synthesis of *tert*-butyl *N*⁶-((benzyloxy)carbonyl)-*N*²-(*tert*-butoxycarbonyl)-*L*-lysinate (**21**)



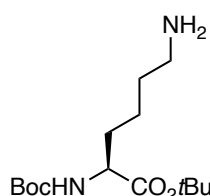
To a solution of *N*⁶-((benzyloxy)carbonyl)-*N*²-(*tert*-butoxycarbonyl)-*L*-lysine (Boc-Lys(Z)-OH, 400 mg, 1.05 mmol, 1.0 eq.) in *t*BuOH (15 ml) were added Boc₂O (0.34 ml, 1.47 mmol, 1.4 eq.) and DMAP (51 mg, 0.42 mmol, 0.4 eq.) and reaction was stirred o.n. at 30°C. The solvent was removed and the crude product was purified by flash column chromatography (20% EtOAc in pentane) to obtain product **21** as a colourless oil (0.43 g, 0.99 mmol, 94%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.34 - 1.60 (m, 23H), 1.68 - 1.85 (m, 1H), 3.14 - 3.24 (m, 2H), 4.14 (m, 1H), 4.80-5.10 (m, 2H), 5.09 (s, 2H), 7.28 - 7.37 (m, 5H).

LC-MS (m/z): calcd. for C₂₃H₃₆N₂O₆ [M+Na⁺] 459.3, found 459.2.

The chemical shifts are according to literature.³³¹

Synthesis of *tert*-butyl (*tert*-butoxycarbonyl)-*L*-lysinate (**22**)



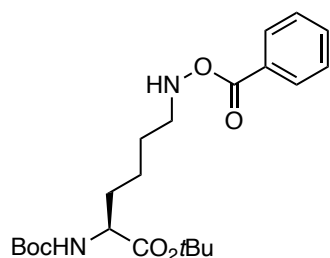
To a solution of **21** (3.64 g, 8.33 mmol, 1.0 eq.) in MeOH (60 ml) on ice were added Pd/C (0.36 g, 10% wt) and Et₃SiH (13.30 ml, 83.25 mmol, 10.0 eq.) and reaction was stirred o.n., allowing to warm up to r.t. The crude reaction mixture was filtered and residue washed with MeOH, the solvent was removed and the crude product was purified by flash column chromatography (2→10% MeOH in DCM +0.5% Et₃N) to obtain product **22** as slight yellow oil (2.47 g, 8.16 mmol, 98%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.29 - 1.41 (m, 1H), 1.44 (s, 9H), 1.46 (s, 9H), 1.48 - 1.68 (m, 4H), 1.68 - 1.87 (m, 1H), 2.73 (t, *J* = 6.9 Hz, 2H), 4.09 - 4.23 (m, 1H), 5.08 (d, *J* = 8.3 Hz, 1H).

LC-MS (m/z): calcd. for C₁₅H₃₀N₂O₄ [M+H⁺] 303.2, found 303.2.

The chemical shifts are according to literature.³³¹

Synthesis of *tert*-Butyl-*N*⁶-(benzyloxy)-*N*²-(*tert*-butoxycarbonyl)-(L)-lysinate (**23**)



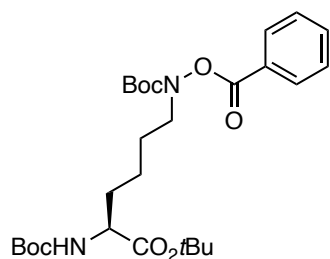
To a solution of **22** (4.15 g, 13.72 mmol, 1.0 eq.) in DCM (20 ml) and NaHCO₃/NaOH (100 ml, pH 10.5, ref. ³⁰⁶) on ice was added dibenzoyl peroxide (75% wt in H₂O, 6.65 g, 20.57 mmol, 1.5 eq.) in DCM (80ml) and reaction was stirred under vigorous stirring for 4 h at 0°C. The phases were separated and aqueous phase was extracted twice with DCM (100 ml each). The combined organic phases were washed with brine (100 ml), dried over Na₂SO₄, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography (10→20% EtOAc in pentane) to obtain product **23** as a colourless oil (3.56 g, 8.43 mmol, 61%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.43 (s, 9H), 1.45 (s, 9H), 1.50 - 1.72 (m, 5H), 1.74 - 1.89 (m, 1H), 3.12 (t, *J* = 7.0 Hz, 2H), 4.10 - 4.30 (m, 1H), 4.98 - 5.12 (m, 1H), 7.40 - 7.49 (m, 2H), 7.53 - 7.61 (m, 1H), 7.97 - 8.04 (m, 2H).

¹³C-NMR (75 MHz, CDCl₃): δ = 22.8, 27.1, 28.1, 28.4, 28.8, 32.9, 52.4, 53.9, 79.7, 81.9, 128.5, 128.6, 129.5, 133.4, 155.5, 167.0.

LC-MS (m/z): calcd. for C₂₂H₃₄N₂O₆ [M+H⁺] 423.3, found 423.2.

Synthesis of *tert*-butyl *N*⁶-(benzyloxy)-*N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-L-lysinate (**24**)



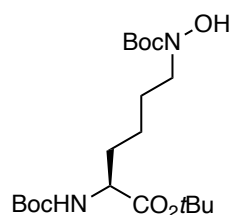
To a solution of **23** (3.47 g, 8.21 mmol, 1.0 eq.) in DCM (150 ml) was added Boc₂O (35.15 ml, 164.25 mmol, 20.0 eq.) and reaction was stirred o.n. at 50°C (reflux). The crude reaction mixture was concentrated under reduced pressure and purified by flash column chromatography (2→20% EtOAc in pentane) to obtain product **24** as a colourless oil (4.12 g, 7.88 mmol, 96%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.42 (s, 9H), 1.45 (s, 9H), 1.45 (s, 9H), 1.46 - 1.88 (m, 6H), 3.68 (t, *J* = 7.0 Hz, 2H), 4.09 - 4.20 (m, 1H), 4.92 - 5.11 (m, 1H), 7.42 - 7.51 (m, 2H), 7.55 - 7.65 (m, 1H), 8.02 - 8.10 (m, 2H).

¹³C-NMR (75 MHz, CDCl₃): δ = 22.5, 27.1, 28.1, 28.3, 28.5, 32.7, 50.5, 54.0, 77.4, 79.7, 81.9, 82.5, 127.8, 128.7, 130.0, 133.9, 154.9, 155.5, 164.8.

LC-MS (m/z): calcd. for C₂₇H₄₂N₂O₈ [M+H⁺] 523.3, found 523.3.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-hydroxy-*L*-lysinate (**25**)



Synthesis was performed according to ref.³³² To a solution of **24** (1.94 g, 3.72 mmol, 1.0 eq.) in a mixture of THF/H₂O/MeOH (1/1/2, V_Σ = 64 ml) was added LiOH·H₂O (1.56 g, 37.17 mmol, 10.0 eq.) and reaction was stirred at r.t. for 40 min. 1 M HCl (34 ml) and EtOAc (100 ml) were added, phases separated and the aqueous phase was extracted with EtOAc (100 ml). The combined organic phases were washed with brine (100 ml), dried over Na₂SO₄, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography (10→50% Et₂O in pentane, + 0.5% Et₃N) to obtain product **25** as a slight yellow oil (1.20 g, 2.88 mmol, 77%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.45 (s, 9H), 1.46 (s, 9H), 1.48 (s, 9H), 1.50 - 1.54 (m, 2H), 1.60–1.70 (m, 2H), 1.70–1.87 (m, 2H), 3.37–3.44 (m, 1H), 3.52–3.60 (m, 1H), 4.20 (s, 1H), 5.09 (d, *J* = 8.4 Hz, 1H), 6.60 (s, 1H).

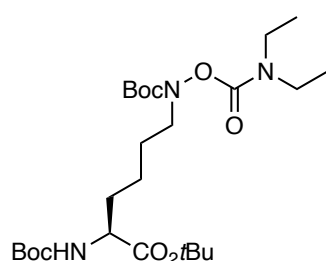
¹³C-NMR (75 MHz, CDCl₃): δ = 22.1, 26.2, 28.1, 28.5, 28.5, 33.1, 49.8, 53.5, 80.0, 81.5, 82.0, 155.9, 157.2, 172.2.

LC-MS (*m/z*): calcd. for C₂₀H₃₈N₂O₇ [*M*+H⁺] 441.3, found 441.2.

General protocol (GP-1) for the synthesis of protected carbamyl/carbonyl hydroxylamine lysines **26a-d**

To **25** (1.0 eq.) in DCM were added DIPEA (2.5 eq.), DMAP (0.1 eq.) and acylchloride (2.2 eq.). The reaction was stirred o.n. under reflux (60°C). The solvent was removed and the crude product was purified by flash column chromatography to obtain product **26a-d**.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-((diethylcarbamoyl)oxy)-*L*-lysinate (**26a**)

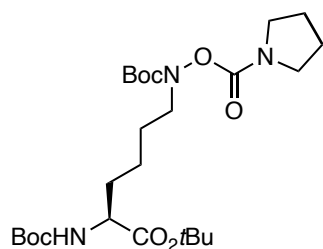


26a was synthesized according to GP-1 (0.09 g, 0.22 mmol of **25**) with *N,N*-diethylcarbamyl chloride and purified by flash column chromatography (20% EtOAc in pentane) to obtain the product as a colourless oil (0.10 g, 0.20 mmol, 89%).

¹H-NMR (500 MHz, CDCl₃): δ = 1.13-1.23 (m, 6H), 1.43-1.48 (m, 27H), 1.55-1.68 (m, 5H), 1.73-1.84 (m, 1H), 3.31 (bs, 4H), 3.60 (bs, 2H), 4.12-4.20 (m, 1H), 5.03 (d, *J* = 8.0 Hz, 1H).

LC-MS (*m/z*): calcd. for C₂₅H₄₇N₃O₈ [*M*+Na⁺] 540.3, found 540.3.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-((pyrrolidine-1-carbonyl)oxy)-*L*-lysinate (26b**)**

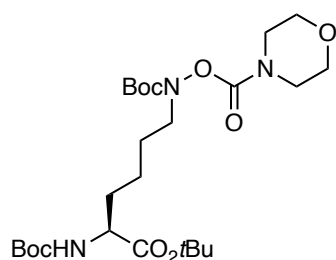


26b was synthesized according to GP-1 (0.56 g, 1.35 mmol of **25**) with 1-pyrrolidinecarbonyl chloride and purified by flash column chromatography (20→30% EtOAc in pentane) to obtain the product as a colourless oil (0.65 mg, 1.26 mmol, 93%).

¹H-NMR (500 MHz, CDCl₃): δ = 1.42-1.48 (m, 28H), 1.57-1.67 (m, 4H), 1.74-1.83 (m, 1H), 1.84-1.95 (m, 4H), 3.40-3.46 (m, 4H), 3.60 (s, 2H), 4.12-4.19 (m, 2H), 5.02 (d, *J* = 7.8 Hz, 1H).

LC-MS (*m/z*): calcd for C₂₅H₄₅N₃O₈ [*M*+Na⁺] 538.3, found 538.3.

Synthesis of *tert*-Butyl-*N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-((morpholine-4-carbonyl)oxy)-(*L*)-lysinate (26c**)**



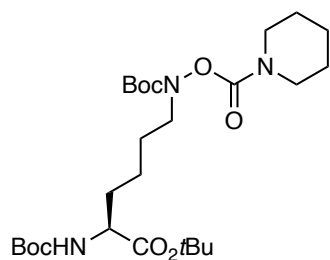
26c was synthesized according to GP-1 (0.54 g, 1.30 mmol of **25**) with morpholine-4-carbonyl chloride and purified by flash column chromatography (10→40% EtOAc in pentane) to obtain the product as a colourless oil (0.67 g, 1.26 mmol, 97%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.35 - 1.82 (m, 33H), 3.48 - 3.54 (m, 4H), 3.58 (t, *J* = 6.9 Hz, 2H), 3.65 - 3.72 (m, 4H), 4.03 - 4.23 (m, 1H), 4.93 - 5.10 (m, 1H).

¹³C-NMR (75 MHz, CDCl₃): δ = 22.5, 27.1, 28.1, 28.4, 28.5, 32.7, 44.7, 50.4, 54.0, 66.6, 77.4, 79.7, 81.9, 82.1, 153.9, 155.0, 155.5.

LC-MS (*m/z*): calcd for C₂₅H₄₅N₃O₉ [*M*+Na⁺] 554.3, found 554.3.

Synthesis of *tert*-Butyl-*N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-((piperidine-1-carbonyl)oxy)-(L)-lysinate (**26d**)



26d was synthesized according to GP-1 (0.53 g, 1.27 mmol of **25**) with 1-piperidinecarbonyl chloride and purified by flash column chromatography (10→20% EtOAc in pentane) to obtain the product as a colourless oil (0.64 g, 1.20 mmol, 95%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.36 - 1.69 (m, 39H), 3.39 - 3.50 (m, 4H), 3.52 - 3.63 (m, 2H), 4.02 - 4.26 (m, 1H), 4.95 - 5.05 (m, 1H).

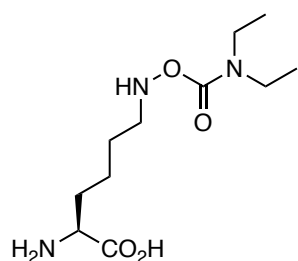
¹³C-NMR (75 MHz, CDCl₃): δ = 22.5, 24.3, 25.8, 27.1, 28.1, 28.4, 28.5, 32.7, 50.2, 53.5, 54.0, 77.4, 79.7, 81.8, 81.8, 153.9, 155.1, 155.5.

LC-MS (m/z): calcd. for C₂₆H₄₇N₃O₈ [M+H⁺] 530.3, found 530.3.

General protocol (GP-2) for the synthesis of carbamyl hydroxylamine lysines **27a-d**

To **26a-d** (1.0 eq.) DCM were added H₂O (~10 vol %) and 4 M HCl in dioxane (~2 M final concentration). The reaction was stirred at o.n. at r.t. until LC-MS analysis indicated full conversion. The reaction mixture was transferred into a 50 ml Falcon™ tube, the solvent removed, dissolved in a mixture of H₂O/MeCN, flash frozen and lyophilised.

Synthesis of *N*⁶-((diethylcarbamoyl)oxy)-L-lysine (**27a**)

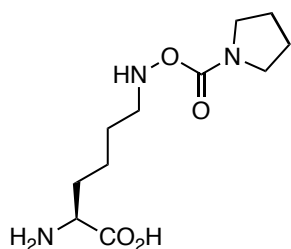


27a was synthesised according to GP-2 (0.38 g, 0.74 mmol of **26a**) to obtain a white powder after lyophilisation (HCl salt, 0.25 g, 0.85 mmol, 114% → possibly due to inaccuracies during weight measurements).

¹H-NMR (500 MHz, CD₃OD): δ = 1.19 (t, *J* = 7.1 Hz, 6H), 1.54-1.70 (m, 2H), 1.80 (p, *J* = 7.7 Hz, 2H), 1.89-2.05 (m, 2H), 3.31-3.45 (m, 6H), 4.00 (t, *J* = 6.4 Hz, 1H).

LC-MS (m/z): calcd. for C₁₁H₂₃N₃O₄ [M+H⁺] 262.3, found 262.3.

Synthesis of *N*⁶-((pyrrolidine-1-carbonyl)oxy)-*L*-lysine (**27b**)

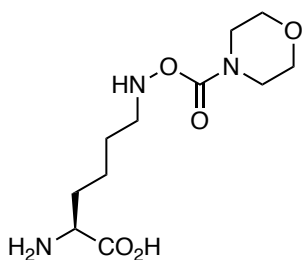


27b was synthesised according to GP-2 (0.65 g, 1.26 mmol of **26a**) to obtain a white powder after lyophilisation (HCl salt, 0.43 g, 1.46 mmol, 116% → possibly due to inaccuracies during weight measurements).

¹H-NMR (500 MHz, CD₃OD): δ = 1.55-1.73 (m, 2H), 1.85 (p, *J* = 7.7 Hz, 2H), 1.92-2.07 (m, 6H), 3.45 (t, *J* = 7.5 Hz, 3H), 3.49 (t, *J* = 6.7 Hz, 3H), 4.03 (t, *J* = 6.4 Hz, 1H).

LC-MS (*m/z*): calcd. for C₁₁H₂₁N₃O₄ [M+H⁺] 260.3, found 260.2.

Synthesis of *N*⁶-((morpholine-4-carbonyl)oxy)-*L*-lysine (**27c**)

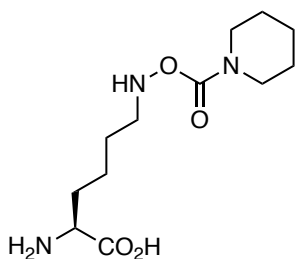


27c was synthesised according to GP-2 (0.52 g, 0.98 mmol of **26a**) to obtain a white powder after lyophilisation (HCl salt, 0.35 g, 1.14 mmol, 116% → possibly due to inaccuracies during weight measurements).

¹H-NMR (500 MHz, CD₃OD): δ = 1.51-1.69 (m, 2H), 1.80 (p, *J* = 7.7 Hz, 2H), 1.88-2.05 (m, 2H), 3.38-3.44 (m, 2H), 3.47-3.53 (m, 4H), 3.63-3.74 (m, 4H), 3.99 (t, *J* = 6.4 Hz, 1H).

LC-MS (*m/z*): calcd for C₁₁H₂₁N₃O₅ [M+H⁺] 276.3, found 276.1.

Synthesis of *N*⁶-((piperidine-1-carbonyl)oxy)-*L*-lysine (**27d**)

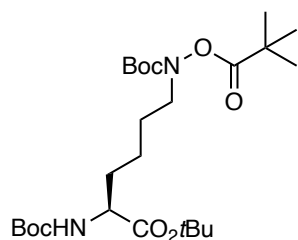


27d was synthesised according to GP-2 (0.60 g, 1.14 mmol of **26a**) to obtain a white powder after lyophilisation (HCl salt, 0.36 g, 1.04 mmol, 92%).

¹H-NMR (300 MHz, CD₃OD): δ = 1.32 - 1.58 (m, 9H), 1.61 - 1.95 (m, 4H), 3.27 - 3.37 (m, 6H), 3.39 - 3.56 (m, 2H), 3.85 (t, *J* = 6.3 Hz, 1H).

LC-MS (*m/z*): calcd. for C₁₂H₂₃N₃O₄ [M+Na⁺] 296.2, found 296.2.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-(pivaloyloxy)-*L*-lysinate (**28a**)



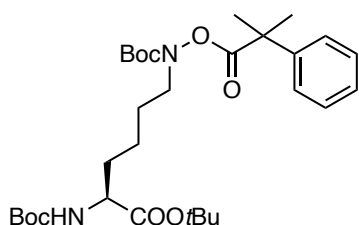
28a was synthesized according to GP-1 (0.22 g, 0.52 mmol of **25**) with pivaloyl chloride and purified by flash column chromatography (20% Et₂O in pentane) to obtain the product as a colourless oil (0.25 g, 0.49 mmol, 95%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.28 (s, 9H), 1.44 (s, 9H), 1.46 (s, 9H), 1.46 (s, 9H), 1.48 – 1.56 (m, 3H), 1.59 – 1.68 (m, 2H), 1.69 – 1.86 (m, 1H), 3.54 (t, *J* = 6.9 Hz, 2H), 4.11 - 4.19 (m, 1H), 5.02 (d, *J* = 8.5 Hz, 1 H).

¹³C-NMR (75 MHz, CDCl₃) δ = 22.4, 27.2, 28.2, 28.3, 28.5, 29.8, 30.5, 32.7, 38.3, 49.9, 79.7, 81.9, 82.1, 154.9, 155.5, 172.0, 175.9.

LC-MS (*m/z*): calcd. for C₂₅H₄₆N₂O₈ [M-Boc+H⁺] 403.3, found 403.2.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-((2-phenylpropan-2-yl)oxy)-*L*-lysinate (**28b**)



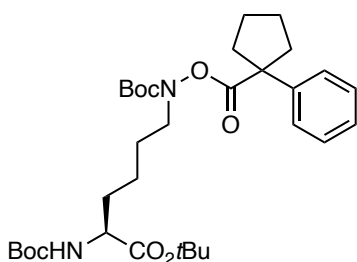
28b was synthesized according to GP-1 (0.28 g, 0.67 mmol of **25**) with 2-methyl-2-phenylpropanoyl chloride. The acyl chloride was freshly prepared by reaction of 2-methyl-2-phenylpropionic acid (1.0 eq.) with oxalyl chloride (1.2 eq.) in DCM on ice for 5 h. Solvent was removed and the crude mixture was immediately used for acylation of **25**.

The crude reaction mixture was concentrated under reduced pressure and purified by flash column chromatography (10% EtOAc in pentane) to obtain the product as a colourless oil (0.29 g, 0.51 mmol, 76%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.18-1.34 (m, 6H), 1.40-1.47 (m, 29H), 1.66 (s, 6H), 3.42 (t, *J* = 6.5 Hz, 1H), 4.04-4.13 (m, 1H), 4.96 (d, *J* = 9.1 Hz, 1H), 7.27-7.42 (m, 5H).

LC-MS (*m/z*): calcd. for C₃₀H₄₈N₂O₈ [M-Boc+H⁺] 465.3, found 465.3.

Synthesis of tert-butyl N^2,N^6 -bis(tert-butoxycarbonyl)- N^6 -((1-phenylcyclopentane-1-carbonyloxy)-L-lysinate (28c)



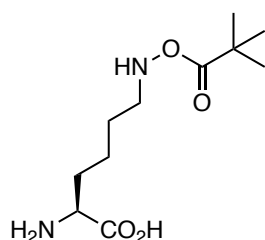
28c was synthesized according to GP-1 (0.35 g, 0.84 mmol of **25**) with 1-phenylcyclopentanecarboxylic acid chloride. The acyl chloride was freshly prepared by reaction of 1-phenylcyclopentanecarboxylic (1.0 eq.) with oxalyl chloride (1.2 eq.) in DCM on ice for 2 h. Solvent was removed and the crude mixture was immediately used for acylation of **25**.

The crude reaction mixture was concentrated under reduced pressure and purified by flash column chromatography (10% EtOAc in pentane) to obtain the product as a colourless oil (0.30 g, 0.50 mmol, 59%).

$^1\text{H-NMR}$ (300 MHz, CDCl_3): δ = 1.13-1.31 (m, 4 H), 1.39 (s, 9H), 1.43-1.46 (m, 18H), 1.74-1.82 (m, 4H), 1.93-2.03 (m, 2H), 2.66-2.79 (m, 2H), 3.28-3.39 (m, 2H), 4.01-4.10 (m, 1H), 4.95 (d, J = 7.9 Hz, 1H), 7.21-7.41 (m, 5H).

LC-MS (m/z): calcd. for $\text{C}_{32}\text{H}_{50}\text{N}_2\text{O}_8$ [$\text{M-Boc}+\text{H}^+$] 491.3, found 491.3.

Synthesis of N^6 -(pivaloyloxy)-L-lysine (29a)



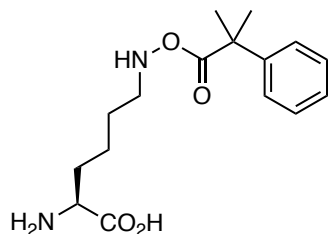
To a solution of **28a** (0.47 g, 0.91 mmol, 1.0 eq.) in 1,4-dioxane (6 ml) were added Et_3SiH (0.6 ml, 3.76 mmol, 4.18 eq.) and 4 M HCl in dioxane (2M final concentration) and reaction was stirred o.n. at r.t. The crude product was precipitated and washed in cold Et_2O and collected by centrifugation to obtain product **29a** as a white solid (HCl salt, 0.25 g, 0.87 mmol, 93%).

$^1\text{H-NMR}$ (300 MHz, CD_3OD): δ = 1.19 (s, 2H), 1.30 (s, 7H), 1.56 - 1.68 (m, 2H), 1.68 - 1.87 (m, 2H), 1.98-1.99 (m, 2H), 3.21-3.29 (m, 1H), 3.32-3.37 (m, 1H), 3.95 - 4.05 (m, 1H).

$^{13}\text{C-NMR}$ (75 MHz, CD_3OD): δ = 23.3, 24.2, 25.0, 27.1, 27.6, 31.0, 39.6, 50.5, 53.6, 171.5, 176.3.

LC-MS (m/z): calcd. for $\text{C}_{11}\text{H}_{22}\text{N}_2\text{O}_4$ [$\text{M}+\text{H}^+$] 247.2, found 247.2.

Synthesis of *N*⁶-((2-methyl-2-phenylpropanoyl)oxy)-*L*-lysine (**29b**)

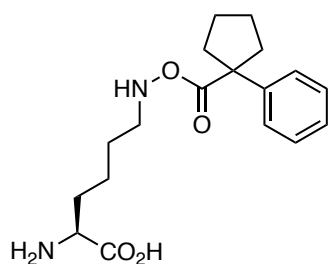


To a solution of **28b** (0.29 g, 0.51 mmol, 1.0 eq.) in 1,4-dioxane (3 ml) were added Et₃SiH (0.32 ml, 2.04 mmol, 4.0 eq.) and 4 M HCl in dioxane (4 ml, 2.29 M final concentration) and reaction was stirred o.n. at r.t. The crude product was precipitated and washed in cold Et₂O and collected by centrifugation to obtain product **29b** as a white solid (HCl salt, 0.16 g, 0.46 mmol, 90%).

¹H-NMR (300 MHz, CD₃OD): δ = 1.42-1.59 (m, 4H), 1.66 (s, 6H), 1.79-1.96 (m, 2H), 3.09 (t, *J* = 6.6 Hz, 2H), 3.94 (t, *J* = 6.3 Hz, 1H), 7.24-7.48 (m, 5H).

LC-MS (*m/z*): calcd. for C₁₆H₂₄N₂O₄ [*M* + H⁺] 309.2, found 309.2.

Synthesis of *N*⁶-((1-phenylcyclopentane-1-carbonyl)oxy)-*L*-lysine (**29c**)

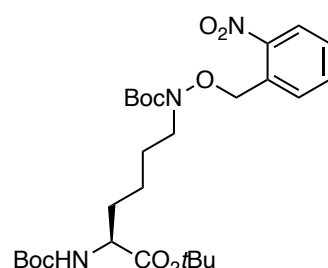


To a solution of **28c** (0.30 g, 0.5 mmol, 1.0 eq.) in 1,4-dioxane (2 ml) were added Et₃SiH (0.32 ml, 2.00 mmol, 4.0 eq.) and 4 M HCl in dioxane (3 ml, 2.4 M final concentration) and reaction was stirred o.n. at r.t. The crude product was precipitated and washed in cold Et₂O and collected by centrifugation to obtain product **29c** as a white solid (HCl salt, 0.11 g, 0.32 mmol, 65%).

¹H-NMR (300 MHz, CD₃OD): δ = 1.38-1.54 (m, 4H), 1.75-1.91 (m, 6H), 1.99-2.12 (m, 2H), 2.64-2.75 (m, 2H), 3.06-3.16 (m, 2H), 3.93 (t, *J* = 6.3 Hz, 1H), 7.24-7.47 (m, 5H).

LC-MS (*m/z*): calcd. for C₁₈H₂₆N₂O₄ [*M* + H⁺] 335.2, found 335.2.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-((2-nitrobenzyl)oxy)-*L*-lysinate (**30**)

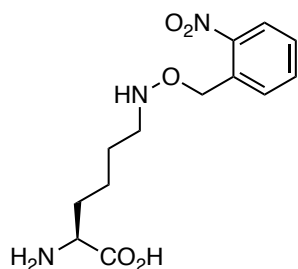


To **25** (0.29 g, 0.70 mmol, 1.0 eq) in MeCN (15 ml) were added K₂CO₃ (0.18 g, 0.84 mmol, 1.2 eq.) and 2-nitrobenzyl bromide (0.19 mg, 1.4 mmol, 2.0 eq.) and reaction was stirred for 4 d at 60°C. The crude mixture was concentrated under reduced pressure and taken up in DCM (50 ml) and H₂O (20 ml). Phases were separated and aqueous phase extracted with DCM (50 ml). The combined organic phases were washed with brine (30 ml), dried over Na₂SO₄, filtered and evaporated. The crude product was purified by flash column chromatography (5→15% EtOAc in pentane) to obtain product **30** as a yellow oil (0.31 g, 0.56 mmol, 80%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.42-1.49 (m, 27H), 1.58-1.67 (m, 4H), 1.68-1.85 (m, 2H), 3.45 (t, *J* = 7.2 Hz, 2H), 4.12-4.20 (m, 1H), 5.00 (d, *J* = 8.0 Hz, 1H), 5.22 (s, 2H), 7.43-7.53 (m, 1H), 7.64 (td, *J* = 7.6 Hz, *J* = 1.3 Hz, 1H), 7.72-7.79 (m, 1H), 8.03 (dd, *J* = 8.2 Hz, *J* = 1.2 Hz, 1H)

LC-MS (m/z): calcd. for C₂₇H₄₃N₃O₉ [M+Na⁺] 576.3, found 576.2.

Synthesis of *N*⁶-((2-nitrobenzyl)oxy)-*L*-lysine (31)

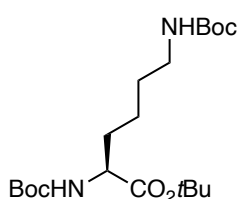


To a solution of **30** (0.30 g, 0.54 mmol, 1.0 eq.) in 1,4-dioxane (4 ml) were added Et₃SiH (0.34 ml, 2.16 mmol, 4.0 eq.) and 4 M HCl in dioxane (6 ml, 2.5 M final concentration) and reaction was stirred for 2 d at r.t. The crude product was precipitated and washed in cold Et₂O and collected by centrifugation to obtain product **29c** as a beige solid (HCl salt, 0.12 g, 0.36 mmol, 67%).

¹H-NMR (300 MHz, CD₃OD): δ = 1.54-1.73 (m, 2H), 1.85 (p, *J* = 7.7 Hz, 2H), 1.92-2.09 (m, 2H), 3.39-3.44 (m, 2H), 4.02 (t, *J* = 6.4 Hz, 1H), 5.53 (s, 2H), 7.68-7.85 (m, 3H), 8.16-8.20 (m, 1H).

LC-MS (m/z): calcd. for C₁₃H₁₉N₃O₅ [M+H⁺] 298.2, found 298.2.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*L*-lysinate (32)



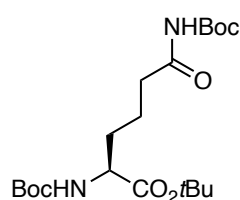
To a solution of *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*L*-lysine (3.82 g, 11.03 mmol, 1.0 eq.) in *t*BuOH (30 ml) were added Boc₂O (3.60 ml, 15.68 mmol, 1.2 eq.) and DMAP (0.54 g, 4.39 mmol, 0.4 eq.) and reaction was stirred at o.n. at 30°C. The reaction was concentrated under reduced pressure and purified by flash column chromatography (20→30% Et₂O in pentane) to obtain product **32** as a white solid (3.822 g, 9.50 mmol, 86%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.44 (s, 9H), 1.44 (s, 9H), 1.46 (s, 9H), 1.47 - 1.58 (m, 4H), 1.60–1.69 (m, 1H), 1.69–1.84 (m 1H), 3.04–3.17 (m, 2H), 4.07–4.22 (m, 1H), 4.55 (s, 1H), 5.05 (d, *J* = 8.3 Hz, 1H).

¹³C-NMR (75 MHz, CDCl₃) δ = 22.5, 28.2, 28.5, 28.6, 32.3, 32.8, 40.4, 53.9, 79.2, 79.8, 81.9, 155.6, 156.2, 172.1.

LC-MS (m/z): calcd. for C₂₀H₃₈N₂O₆ [M+Na⁺] 425.3, found 425.2.

Synthesis of *tert*-butyl (S)-2,6-bis((*tert*-butoxycarbonyl)amino)-6-oxohexanoate (33)



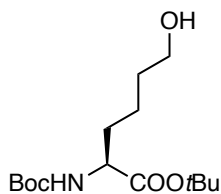
To a solution of **32** (1.05 g, 2.62 mmol, 1.0 eq.) in EtOAc/H₂O (3/4, V_Σ = 8.74 ml, 0.3 M of **32**) were added RuO₂ hydrate powder (35 mg, appr. 0.26 mmol, 0.10 mmol) and NaBrO₃ (0.99 g, 6.54 mmol, 2.50 eq.) and reaction was stirred at 35°C for 39 h. EtOAc and H₂O were added (50 ml each) and reaction was filtered. The phases were separated and the aqueous phase extracted with EtOAc (50 ml). The combined organic phases were washed with brine (50 ml), dried over Na₂SO₄, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography (15→30% Et₂O in pentane) to obtain product **33** as a yellow oil (0.51 g, 1.23 mmol, 47%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.44 (s, 9H), 1.46 (s, 9H), 1.48 (s, 9H), 1.62 - 1.77 (m, 3H), 1.77–1.89 (m, 1H), 2.68–2.86 (m, 2H), 4.17 (s, 1H), 5.10 (d, *J* = 8.2 Hz, 1H), 7.34 (s, 1H).

LC-MS (m/z): calcd. for C₂₀H₃₆N₂O₇ [M+Na⁺] 439.2, found 439.2.

Chemical shifts are according to literature.³³³

Synthesis of *tert*-butyl (S)-2-((*tert*-butoxycarbonyl)amino)-6-hydroxyhexanoate (34)



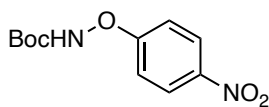
To a solution of **33** (0.51 g, 1.23 mmol, 1.0 eq.) in *i*PrOH (11.0 ml) was added NaBH₄ and reaction was stirred o.n. at r.t. The solvent was removed and the crude product was purified by flash column chromatography (20→50% Et₂O in pentane) to obtain product **34** as a colourless oil (0.28 g, 0.91 mmol, 74%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.44 (s, 9H), 1.46 (s, 9H), 1.54 – 1.67 (m, 6H), 1.71–1.86 (m, 1H), 3.64 (td, *J* = 1.4, 6.3 Hz, 2H), 4.11 – 4.22 (m, 1H), 5.04 (d, *J* = 8.3 Hz, 1H).

LC-MS (m/z): calcd. for C₁₅H₂₉NO₅ [M+Na⁺] 326.2, found 326.2.

Chemical shifts are according to literature.³³⁴

Synthesis of *tert*-butyl (4-nitrophenoxy)carbamate (35)



To 1-fluoro-4-nitrobenzene (897 μl, 8.45 mmol, 1.0 eq.) and *t*Bu-*N*-hydroxycarbamate (1.35 g, 10.14 mmol, 1.2 eq.) was added KOH (0.52 g, 9.30 mmol, 1.1 eq.) and reaction was stirred for 3 d at 60°C. The reaction mixture was concentrated under reduced pressure, taken up in EtOAc (60 ml) and washed with satd. NH₄Cl_{aq.} (3x 15 ml) and brine (15 ml). The organic phase was dried over Na₂SO₄, filtered, concentrated under reduced pressure and purified by flash column

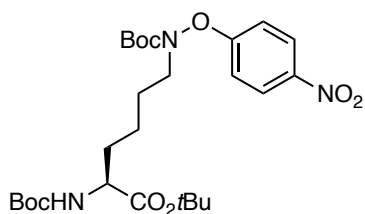
chromatography (10→20% EtOAc in pentane) to obtain product **35** as a beige solid (1.92 g, 7.56 mmol, 89%; 7% *p*-nitrophenol according to NMR analysis)

¹H-NMR (300 MHz, CDCl₃): δ = 1.50 (s, 9H), 7.17 - 7.24 (m, 2H), 7.68 (s, 1H), 8.16 - 8.26 (m, 2H).

¹³C-NMR (75 MHz, CDCl₃) δ = 28.2, 84.0, 113.6, 125.9, 143.2, 156.1, 164.8.

LC-MS (m/z): calcd. for C₁₁H₂₄N₂O₅ [M+H⁺] 255.1, found 255.1.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-(4-nitrophenoxy)-*L*-lysinate (**36**)



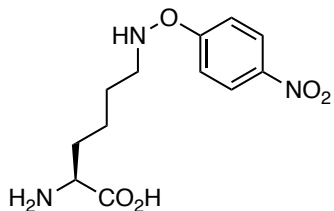
To a solution of **34** (0.27 g, 0.90 mmol, 1.0 eq.) and **35** (0.25 g, 1.0 mmol, 1.10 eq.) in THF_{dry} (7 ml) under Ar atm on ice were added DIAD (0.21 ml, 1.12 mmol, 1.2 eq.) and PPh₃ (0.28g, 1.1 mmol, 1.2 eq.) and reaction was stirred at r.t. After 5.5 h, reaction was concentrated under reduced pressure and purified by flash column chromatography (5→30% Et₂O in pentane) to obtain product **36** as a yellow oil. (0.41 g, 0.76 mmol, 84%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.43 (s, 9H), 1.44 (s, 9H), 1.45 (s, 9H), 1.46 - 1.55 (m, 2H), 1.57 - 1.70 (m, 3H), 1.72 - 1.93 (m, 1H), 3.61 (t, *J* = 7.3 Hz, 2H), 4.01 - 4.25 (m, 1H), 5.00 (d, *J* = 8.2 Hz, 1H), 7.06 - 7.17 (m, 2H), 8.16 - 8.27 (m, 2H).

¹³C-NMR (75 MHz, CDCl₃) δ = 22.7, 26.6, 28.1, 28.2, 28.5, 32.8, 51.0, 53.9, 79.8, 82.2, 83.3, 113.5, 126.0, 143.0, 155.5, 156.3, 164.6, 171.9.

LC-MS (m/z): calcd. for C₂₆H₄₁N₃O₉ [M+Na⁺] 562.3, found 562.3.

Synthesis of *N*⁶-(4-nitrophenoxy)-*L*-lysine (**37**)

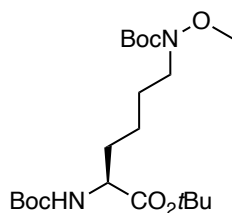


To a solution of **36** (0.17 g, 0.31 mmol, 1.0 eq.) in 1,4-dioxane (2 ml) were added Et₃SiH (0.20 ml, 1.24 mmol, 4.0 eq.) and 4 M HCl in dioxane (3 ml, 2.4 M final concentration) and reaction was stirred o.n. at r.t. The crude reaction was concentrated under reduced pressure and additionally purified by reverse phase column chromatography (gradient 10→55% MeCN in H₂O, + 0.1% formic acid). Appropriate fractions were collected and lyophilised to obtain product **37** (3 mg, 0.01 mmol, 3%).

¹H-NMR (300 MHz, CD₃OD): δ = 1.50-1.70 (m, 4H), 1.77-1.98 (m, 2H), 3.14 (t, *J* = 6.7 Hz, 2H), 3.53-3.61 (m, 1H), 7.25-7.33 (m, 2H), 8.17-8.25 (m, 2H).

LC-MS (m/z): calcd. for C₁₂H₁₇N₃O₅ [M+H⁺] 284.1, found 284.1.

Synthesis of *tert*-butyl *N*²,*N*⁶-bis(*tert*-butoxycarbonyl)-*N*⁶-methoxy-*L*-lysinate (**38**)



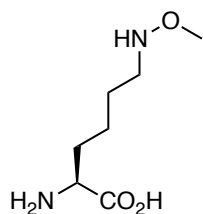
To a solution of **25** (0.11 g, 0.25 mmol, 1.0 eq.) in DMF (6.0 ml) were added K₂CO₃ (0.07 g, 0.49 mmol, 2.0 eq.) and MeI (17.2 μl, 0.28 mmol, 1.1 eq.) and reaction was stirred for 2 h at r.t., followed by o.n. reaction at 60°C. The reaction mixture was concentrated under reduced pressure and purified by flash column chromatography (20% Et₂O in pentane) to obtain product **38** as a colourless oil (0.06 g, 0.14 mmol, 56%).

¹H-NMR (300 MHz, CDCl₃): δ = 1.43 (s, 9H), 1.45 (s, 9H), 1.49 (s, 9H), 1.52 - 1.68 (m, 5H), 1.71 - 1.87 (m, 1H), 3.41 (t, *J* = 7.1 Hz, 2H), 3.66 (s, 3H), 4.11 - 4.19 (m, 1H), 5.01 (d, *J* = 8.3 Hz, 1H).

¹³C-NMR (75 MHz, CDCl₃) δ = 22.6, 27.0, 28.2, 28.5, 30.5, 32.7, 48.9, 54.1, 62.4, 79.7, 81.3, 81.9, 154.9, 156.5, 172.8.

LC-MS (m/z): calcd. for C₂₁H₄₀N₂O₇ [M+Na⁺] 455.3, found 455.2.

Synthesis of *N*⁶-methoxy-*L*-lysine (**39**)



To a solution of **38** (0.08 g, 0.18 mmol, 1.0 eq.) in 1,4-dioxane (4 ml) were added Et₃SiH (0.12 ml, 0.75 mmol, 4.2 eq.) and 4 M HCl in 1,4-dioxane (4 ml, 2 M final concentration) and reaction was stirred o.n. at r.t. The reaction was concentrated under reduced pressure, precipitated and washed in cold Et₂O, collected by centrifugation and lyophilised to obtain product **39** as a white solid (0.04 g, 0.18 mmol, quant.).

¹H-NMR (300 MHz, CD₃OD): δ = 1.49 - 1.71 (m, 2H), 1.73 - 1.86 (m, 2H), 1.87 - 2.10 (m, 2H), 3.32 - 3.38 (m, 2H), 3.94 (s, 3H), 4.00 (t, *J* = 6.4 Hz, 1H).

¹³C-NMR (75 MHz, CD₃OD): δ = 23.3, 24.2, 31.0, 49.7, 53.6, 62.3, 171.5.

LC-MS (m/z): calcd. for C₇H₁₆N₂O₃ [M+H⁺] 177.1, found 177.1.

5.5.2 General set-up for KAT reaction studies

Stock solutions of UAAs (usually 100 mM) and potassium 2-phenylacetyl)trifluoroborate (BzKAT, CAS: 1229039-50-2, SigmaAldrich; 100 mM or 1M in 95/5 MeCN/H₂O) were prepared and mixed together with the appropriate buffers (1x PBS or H₂O) in an 1.5 ml Eppendorf tube to final concentrations as indicated in the main text (V_Σ = 100 μl). KAT reactions were incubated at indicated duration and temperature. 10 μl of the reaction mixture were taken

and diluted into 90 μ l of a 1/1 mixture of H₂O/MeCN or H₂O and immediately subjected to LC-MS analyses.

Sc(OTf)₃ and Yb(OTf)₃ stock solutions (20 mM, in H₂O) were prepared and added to the KAT ligation reaction mixture to final concentration as indicated in the main text and SI.

5.5.3 Chapter 4 – biological experiments

5.5.3.1 Synthetase selection protocol

Plasmids

The following plasmids were used for positive and negative selections: D65_pREP_PylT plasmid (10.3 kb, Tet^R) was used for positive selections and encodes a constitutively expressed chloramphenicol acetyltransferase gene bearing an in-frame amber codon at position 111 and a constitutively expressed *Mb*-pyrrolysyl-tRNA_{CUA} (Pylt). D7_L_pYOBB2_Extended_Pylt (7.4 kb, Cam^R) was used for negative selections and contains an L-arabinose inducible barnase gene with two in-frame amber codons at position 3 and 45 as well as a constitutively expressed *Mb* Pylt.

The following NNK aaRS libraries (size: 3.3 kb) were used for selections: AB1 (randomised positions: L270, Y271, N311, C313, Y349; *Mb* numbering) and AB2 (randomised positions: L270, Y271, N311, C313, Y349, with A267T fixed; *Mb* numbering). Both plasmids contain a Kan^R and the library was designed and prepared by Andreas D. Brunner (former Lang Lab; now Mann Lab, MPI Biochemistry, Martinsried). D3 (randomised positions: L270, Y271, L274, N311, C313; *Mb* numbering) was provided by the Lang lab. This plasmid contains an Amp^R.

Preparation of GMMML selection plate and medium

To following ingredients were mixed and filled up with hand warm GMMML agar to a final volume of 20ml: 400 μ l 50% glycerol, 4 ml 5x M9 salts, 200 μ l 850 mM NaCl, 1x antibiotics and UAA stock solution with a final concentration of choice. This volume is sufficient to cover a 15 cm petri dish. GMMML medium was prepared accordingly with H₂O instead of GMMML agar.

Synthetase selection protocol

Mb synthetase selections were performed according to ref. ^{328, 335} under a slightly modified procedure. Overall, the procedure started with the negative selections first, followed by the positive selection step. This was repeated once again prior detailed analysis of the evolved sequences.

Electrocompetent cells *E. coli* DH10 β cells bearing D7_L and D65 selection plasmids (after plasmid transformation) were prepared as described in 5.2.8.

An aliquot of D7_L cells was transformed with the aaRS library of choice (100-200 ng) by electroporation, SOB rescued and inoculated into 400 ml LB with 1x appropriate antibiotics

(Cam^R and library^R) and incubated o.n. at 37°C (200 rpm). For calculation of the transformation efficiencies, 100 µl of the inoculated culture was taken and dilution series (with LB, V_Σ = 1 ml) with final dilutions of 1/(4x10⁵), 1/(4x10⁶), 1/(4x10⁷) were prepared. 100 µl of each dilution series (final dilutions: 1/(4x10⁶), 1/(4x10⁷), 1/(4x10⁸)) were plated on LB agar plates (with 1x Cam^R and library^R) and plates were incubated o.n. at 37°C. Colonies were counted, adjusted for the dilution factor and transformation efficiency was calculated. If the number of clones exceeded the number of library diversity with the required confidence of >95% (usually >10⁸ clones), 3 ml of the o.n. culture was diluted into 400 ml LB (with 1x antibiotic resistances) and grown until OD₆₀₀ has reached a value of ~0.6. In the meantime, a 24x24 cm LB-agar plate (V_Σ = 200 ml) was prepared with final concentrations of: 1x Cam^R/library^R, 0.2% arabinose, 1.5 mM phenylalanine. 150 µl of the culture was plated and incubated for 12-15 h at 37°C. LB-agar plates were scraped with 2x 25 ml LB, 10 ml LB was additionally added to the cell suspension and cells were incubated for 3.5 h at 37°C (200 rpm) with 1x antibiotics. DNA was isolated (peqGold Plasmid Miniprep Kit, VWR) and 30 µg of the plasmid was additionally purified by agarose gel electrophoresis. The aaRS gel band was isolated, gel slice was removed (Monarch DNA Gel extraction kit, NEB) and plasmid purified (PCR purification kit, Jena Biosciences). Approximately 100 ng of the aaRS library plasmid was transformed into an aliquot of D65 cells by electroporation. After SOC rescue and incubation, cells were inoculated into 100 ml LB with 0.5x antibiotics (Tet^R and library^R) and incubated at 37°C (200 rpm) until OD₆₀₀ of ~0.6. In the meantime, two GMML-agar plates (15 cm dish, V_Σ = 20 ml for each plate) were prepared, additionally supplemented with 1x Tet^R/library^R, 1.2x Cam and 1-2 mM UAA (depending on desired selection strengths). 150 µl and 250 µl of the LB culture were harvested (1000 rcf, 5 min), supernatant discarded and cell pellet washed with 150 µl GMML medium. Cells were pelleted again, resuspended in 150 µl GMML medium, plated out on positive selection plates and incubated for 40-45 h at 37°C.

GMML-agar plates were scraped with 2x 10 ml LB each plate, 10 ml LB was additionally added to the cell suspension and incubated with 1x antibiotics for 3.5 h at 37°C (200 rpm). Synthetase DNA library was isolated as described for the first negative selection step.

For the second negative selection step, approximately 100 ng of isolated aaRS DNA was transformed into an aliquot of D7_L cells by electroporation. After SOB rescue and incubation, cells were inoculated into 100 ml LB with 0.5x Cam/library resistance and incubated at 37°C (200 rpm) until OD₆₀₀ of ~0.6. In the meantime, two LB-agar plates (15 cm dish, V_Σ = 20 ml for each plate) with final concentrations of: 1x Cam^R/library^R, 0.2% arabinose, 2.0 mM phenylalanine. 50 µl (with 100 µl LB) and approximately 10⁵ cells (100 µl of a 10⁶-diluted sample, V_Σ = 1 ml) were plated and incubated for 18-20 h at 37°C. The plates were scraped with 2x 10 ml LB each plate, 10 ml LB was additionally added to the cell suspension and incubated with 1x antibiotics for 3.5 h at 37°C (200 rpm). DNA library was isolated as described for the first negative selection step.

For the second positive selection step, approximately 100 ng of isolated aaRS DNA was transformed into an aliquot of D65 cells by electroporation. The following steps are identical to the first positive selection step. One GMML agar plate (15 cm dish, V_Σ = 20 ml) was prepared, additionally supplemented with 1x Tet^R/library^R, 1.6x Cam and 1-2 mM UAA

(depending on desired selection strengths). 15-50 μ l of the culture (with 100 μ l LB) was plated out on the positive selection plate and incubated for 40-50 h at 37°C.

Individual colonies were picked, inoculated 96-deep-well plates containing LB (with the appropriate antibiotic combinations) and incubated o.n. at 37°C (200 rpm).

Five GMMML-agar selection plates (15 cm dish, V_{Σ} = 20 ml each plate) were prepared, each additionally supplemented with: 1x Tet^R/library^R and either 1) -UAA, -Cam; 2) -UAA, 1x Cam; 3) -UAA, 2.2x Cam; 4) +UAA, 1x Cam or 5) +UAA, 2.2x Cam; (1-2 mM UAA concentration). 30 μ l of the o.n. culture from the 96-deep-well cultures were diluted in 1x PBS and carefully stamped on the plates. Selection plates were incubated for 40-50 h at 37°C. Clones, which showed visible growth differences between +UAA/-UAA plates were considered for further studies, which include aaRS plasmid isolation for sequence analysis and protein expression.

sfGFP protein expressions and purifications with isolated aaRSs from various selections

Co-transformation of isolated aaRS plasmids with pPylt_sfGFP-N149TAG-His₆ was performed as previously described (cf. chapter 5.4.2.1). The o.n. cultures were diluted into AI media (OD₆₀₀ of ~0.05) containing appropriate antibiotics (1x) and UAA (if necessary, 2 mM final concentration) and incubated o.n. at 37°C. Cells were harvested and protein purifications were performed as described in chapter 5.4.2.1. 17 amino acid mixture was prepared analogously to the 18 amino acid mixture without supplementation of lysine (cf. Table 5.8).

6 References

1. Crick, F. H., On protein synthesis. *Symp Soc Exp Biol* **1958**, *12*, 138-63.
2. Kapp, L. D.; Lorsch, J. R., The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* **2004**, *73*, 657-704.
3. 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.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schluter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlen, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschlagler, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B., How many human proteoforms are there? *Nat Chem Biol* **2018**, *14* (3), 206-214.
4. Conibear, A. C., Deciphering protein post-translational modifications using chemical biology tools. *Nat Chem Rev* **2020**.
5. Harper, J. W.; Bennett, E. J., Proteome complexity and the forces that drive proteome imbalance. *Nature* **2016**, *537* (7620), 328-38.
6. Bludau, I.; Aebersold, R., Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nat Rev Mol Cell Biol* **2020**, *21* (6), 327-340.
7. Mohler, K.; Ibba, M., Translational fidelity and mistranslation in the cellular response to stress. *Nat Microbiol* **2017**, *2*, 17117.
8. <http://book.bionumbers.org/what-is-the-error-rate-in-transcription-and-translation/>. as of 05.02.21.
9. Thompson, R. E.; Muir, T. W., Chemoenzymatic Semisynthesis of Proteins. *Chem Rev* **2020**, *120* (6), 3051-3126.
10. Hoyt, E.; Cal, P.; Oliveira, B.; Bernardes, G., Contemporary approaches to site-selective protein modification. *Nature Reviews Chemistry* **2019**, *3*.
11. Lang, K.; Chin, J. W., Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem Rev* **2014**, *114* (9), 4764-806.
12. Tamura, T.; Hamachi, I., Chemistry for Covalent Modification of Endogenous/Native Proteins: From Test Tubes to Complex Biological Systems. *J Am Chem Soc* **2019**, *141* (7), 2782-2799.
13. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994**, *266* (5186), 776-9.
14. 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), 9046-9068.
15. 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), 7328-7443.
16. Watson, E. E.; Ripoll-Rozada, J.; Lee, A. C.; Wu, M. C. L.; Franck, C.; Pasch, T.; Premdjee, B.; Sayers, J.; Pinto, M. F.; Martins, P. M.; Jackson, S. P.; Pereira, P. J. B.; Payne, R. J., Rapid assembly and profiling of an anticoagulant sulfoprotein library. *Proc Natl Acad Sci U S A* **2019**, *116* (28), 13873-13878.
17. Zhang, Y.; Hirota, T.; Kuwata, K.; Oishi, S.; Gramani, S. G.; Bode, J. W., Chemical Synthesis of Atomically Tailored SUMO E2 Conjugating Enzymes for the Formation of Covalently Linked SUMO-E2-E3 Ligase Ternary Complexes. *J Am Chem Soc* **2019**, *141* (37), 14742-14751.

18. Shogren-Knaak, M. A.; Fry, C. J.; Peterson, C. L., A native peptide ligation strategy for deciphering nucleosomal histone modifications. *J Biol Chem* **2003**, *278* (18), 15744-8.
19. He, S.; Bauman, D.; Davis, J. S.; Loyola, A.; Nishioka, K.; Gronlund, J. L.; Reinberg, D.; Meng, F.; Kelleher, N.; McCafferty, D. G., Facile synthesis of site-specifically acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. *Proc Natl Acad Sci U S A* **2003**, *100* (21), 12033-8.
20. 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.; Heard, K. J.; Erikson, G.; Nakadai, T.; Bastle, R. M.; Lukasak, B. J.; Zebroski, H., 3rd; Alenina, N.; Bader, M.; Berton, O.; Roeder, R. G.; Molina, H.; Gage, F. H.; Shen, L.; Garcia, B. A.; Li, H.; Muir, T. W.; Maze, I., Histone serotonylation is a permissive modification that enhances TFIID binding to H3K4me3. *Nature* **2019**, *567* (7749), 535-539.
21. Ansaloni, A.; Wang, Z. M.; Jeong, J. S.; Ruggeri, F. S.; Dietler, G.; Lashuel, H. A., One-pot semisynthesis of exon 1 of the Huntingtin protein: new tools for elucidating the role of posttranslational modifications in the pathogenesis of Huntington's disease. *Angew Chem Int Ed Engl* **2014**, *53* (7), 1928-33.
22. Wang, P.; Dong, S.; Shieh, J. H.; Peguero, E.; Hendrickson, R.; Moore, M. A. S.; Danishefsky, S. J., Erythropoietin derived by chemical synthesis. *Science* **2013**, *342* (6164), 1357-1360.
23. 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), 7805-7812.
24. Koniev, O.; Wagner, A., Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem Soc Rev* **2015**, *44* (15), 5495-551.
25. deGruyter, J. N.; Malins, L. R.; Baran, P. S., Residue-Specific Peptide Modification: A Chemist's Guide. *Biochemistry* **2017**, *56* (30), 3863-3873.
26. Gunnoo, S. B.; Madder, A., Chemical Protein Modification through Cysteine. *ChemBiochem* **2016**, *17* (7), 529-53.
27. Matos, M. J.; Oliveira, B. L.; Martinez-Saez, N.; Guerreiro, A.; Cal, P.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jimenez-Oses, G.; Bernardes, G. J. L., Chemo- and Regioselective Lysine Modification on Native Proteins. *J Am Chem Soc* **2018**, *140* (11), 4004-4017.
28. Rosen, C. B.; Francis, M. B., Targeting the N terminus for site-selective protein modification. *Nat Chem Biol* **2017**, *13* (7), 697-705.
29. Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H., Selective N-terminal functionalization of native peptides and proteins. *Chem Sci* **2017**, *8* (4), 2717-2722.
30. Chen, X.; Ye, F.; Luo, X.; Liu, X.; Zhao, J.; Wang, S.; Zhou, Q.; Chen, G.; Wang, P., Histidine-Specific Peptide Modification via Visible-Light-Promoted C-H Alkylation. *J Am Chem Soc* **2019**, *141* (45), 18230-18237.
31. Jia, S.; He, D.; Chang, C. J., Bioinspired Thiophosphorodichloridate Reagents for Chemoselective Histidine Bioconjugation. *J Am Chem Soc* **2019**, *141* (18), 7294-7301.
32. Tower, S. J.; Hetcher, W. J.; Myers, T. E.; Kuehl, N. J.; Taylor, M. T., Selective Modification of Tryptophan Residues in Peptides and Proteins Using a Biomimetic Electron Transfer Process. *J Am Chem Soc* **2020**, *142* (20), 9112-9118.

33. Seki, Y.; Ishiyama, T.; Sasaki, D.; Abe, J.; Sohma, Y.; Oisaki, K.; Kanai, M., Transition Metal-Free Tryptophan-Selective Bioconjugation of Proteins. *J Am Chem Soc* **2016**, *138* (34), 10798-801.
34. Ma, N.; Hu, J.; Zhang, Z. M.; Liu, W.; Huang, M.; Fan, Y.; Yin, X.; Wang, J.; Ding, K.; Ye, W.; Li, Z., 2H-Azirine-Based Reagents for Chemoselective Bioconjugation at Carboxyl Residues Inside Live Cells. *J Am Chem Soc* **2020**, *142* (13), 6051-6059.
35. Elledge, S. K.; Tran, H. L.; Christian, A. H.; Steri, V.; Hann, B.; Toste, F. D.; Chang, C. J.; Wells, J. A., Systematic identification of engineered methionines and oxaziridines for efficient, stable, and site-specific antibody bioconjugation. *Proc Natl Acad Sci U S A* **2020**, *117* (11), 5733-5740.
36. Taylor, M. T.; Nelson, J. E.; Suero, M. G.; Gaunt, M. J., A protein functionalization platform based on selective reactions at methionine residues. *Nature* **2018**, *562* (7728), 563-568.
37. Alvarez-Dorta, D.; Thobie-Gautier, C.; Croyal, M.; Bouzelha, M.; Mevel, M.; Deniaud, D.; Boujtita, M.; Gouin, S. G., Electrochemically Promoted Tyrosine-Click-Chemistry for Protein Labeling. *J Am Chem Soc* **2018**, *140* (49), 17120-17126.
38. Tilley, S. D.; Francis, M. B., Tyrosine-selective protein alkylation using pi-allylpalladium complexes. *J Am Chem Soc* **2006**, *128* (4), 1080-1.
39. Bloom, S.; Liu, C.; Kolmel, D. K.; Qiao, J. X.; Zhang, Y.; Poss, M. A.; Ewing, W. R.; MacMillan, D. W. C., Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat Chem* **2018**, *10* (2), 205-211.
40. Paulus, H., Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* **2000**, *69*, 447-96.
41. Vila-Perello, M.; Muir, T. W., Biological applications of protein splicing. *Cell* **2010**, *143* (2), 191-200.
42. Kilic, S.; Boichenko, I.; Lechner, C. C.; Fierz, B., A bi-terminal protein ligation strategy to probe chromatin structure during DNA damage. *Chem Sci* **2018**, *9* (15), 3704-3709.
43. 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), 812-6.
44. Hejjaoui, M.; Butterfield, S.; Fauvet, B.; Vercruyse, F.; Cui, J.; Dikiy, I.; Prudent, M.; Olschewski, D.; Zhang, Y.; Eliezer, D.; Lashuel, H. A., Elucidating the role of C-terminal post-translational modifications using protein semisynthesis strategies: alpha-synuclein phosphorylation at tyrosine 125. *J Am Chem Soc* **2012**, *134* (11), 5196-210.
45. Levine, P. M.; Galesic, A.; Balana, A. T.; Mahul-Mellier, A. L.; Navarro, M. X.; De Leon, C. A.; Lashuel, H. A.; Pratt, M. R., alpha-Synuclein O-GlcNAcylation alters aggregation and toxicity, revealing certain residues as potential inhibitors of Parkinson's disease. *Proc Natl Acad Sci U S A* **2019**, *116* (5), 1511-1519.
46. Broncel, M.; Krause, E.; Schwarzer, D.; Hackenberger, C. P., The Alzheimer's disease related tau protein as a new target for chemical protein engineering. *Chemistry* **2012**, *18* (9), 2488-92.
47. Haj-Yahya, M.; Lashuel, H. A., Protein Semisynthesis Provides Access to Tau Disease-Associated Post-translational Modifications (PTMs) and Paves the Way to Deciphering the Tau PTM Code in Health and Diseased States. *J Am Chem Soc* **2018**, *140* (21), 6611-6621.
48. Reif, A.; Siebenhaar, S.; Troster, A.; Schmalzlein, M.; Lechner, C.; Velisetty, P.; Gottwald, K.; Pohner, C.; Boos, I.; Schubert, V.; Rose-John, S.; Unverzagt, C., Semisynthesis of biologically active glycoforms of the human cytokine interleukin 6. *Angew Chem Int Ed Engl* **2014**, *53* (45), 12125-31.

49. Okamoto, R.; Kimura, M.; Ishimizu, T.; Izumi, M.; Kajihara, Y., Semisynthesis of a post-translationally modified protein by using chemical cleavage and activation of an expressed fusion polypeptide. *Chemistry* **2014**, *20* (33), 10425-30.
50. Alexandrov, K.; Heinemann, I.; Durek, T.; Sidorovitch, V.; Goody, R. S.; Waldmann, H., Intein-mediated synthesis of geranylgeranylated Rab7 protein in vitro. *J Am Chem Soc* **2002**, *124* (20), 5648-9.
51. Chu, N.; Salguero, A. L.; Liu, A. Z.; Chen, Z.; Dempsey, D. R.; Ficarro, S. B.; Alexander, W. M.; Marto, J. A.; Li, Y.; Amzel, L. M.; Gabelli, S. B.; Cole, P. A., Akt Kinase Activation Mechanisms Revealed Using Protein Semisynthesis. *Cell* **2018**, *174* (4), 897-907 e14.
52. Zhang, Y.; Park, K. Y.; Suazo, K. F.; Distefano, M. D., Recent progress in enzymatic protein labelling techniques and their applications. *Chem Soc Rev* **2018**, *47* (24), 9106-9136.
53. Fottner, M.; Lang, K., Decorating proteins with LACE. *Nat Chem* **2020**, *12* (11), 980-982.
54. Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y., Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2005**, *2* (2), 99-104.
55. Fernandez-Suarez, M.; Baruah, H.; Martinez-Hernandez, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y., Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* **2007**, *25* (12), 1483-7.
56. Zhou, Z.; Cironi, P.; Lin, A. J.; Xu, Y.; Hrvatin, S.; Golan, D. E.; Silver, P. A.; Walsh, C. T.; Yin, J., Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem Biol* **2007**, *2* (5), 337-46.
57. Ochtrop, P.; Ernst, S.; Itzen, A.; Hedberg, C., Exploring the Substrate Scope of the Bacterial Phosphocholine Transferase AnkX for Versatile Protein Functionalization. *Chembiochem* **2019**, *20* (18), 2336-2340.
58. Popp, M. W.; Ploegh, H. L., Making and breaking peptide bonds: protein engineering using sortase. *Angew Chem Int Ed Engl* **2011**, *50* (22), 5024-32.
59. Nguyen, G. K.; Cao, Y.; Wang, W.; Liu, C. F.; Tam, J. P., Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide. *Angew Chem Int Ed Engl* **2015**, *54* (52), 15694-8.
60. 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), 1285-1294.
61. Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A., A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* **1994**, *266* (5183), 243-7.
62. Liebscher, S.; Kornberger, P.; Fink, G.; Trost-Gross, E. M.; Hoss, E.; Skerra, A.; Bordusa, F., Derivatization of antibody Fab fragments: a designer enzyme for native protein modification. *Chembiochem* **2014**, *15* (8), 1096-100.
63. Hofmann, R.; Akimoto, G.; Wucherpennig, T. G.; Zeymer, C.; Bode, J. W., Lysine acylation using conjugating enzymes for site-specific modification and ubiquitination of recombinant proteins. *Nat Chem* **2020**, *12* (11), 1008-1015.
64. Soye, B. J. D.; Patel, J. R.; Isaacs, F. J.; Jewett, M. C., Repurposing the translation apparatus for synthetic biology. *Curr Opin Chem Biol* **2015**, *28*, 83-90.
65. Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G., Designing logical codon reassignment - Expanding the chemistry in biology. *Chem Sci* **2015**, *6* (1), 50-69.
66. Ibba, M.; Soll, D., Quality control mechanisms during translation. *Science* **1999**, *286* (5446), 1893-7.
67. Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annu Rev Biochem* **2010**, *79*, 413-44.

68. 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), 774-80.
69. 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), 19-24.
70. van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A., Efficient Incorporation of Unsaturated Methionine Analogues into Proteins in Vivo. *Journal of the American Chemical Society* **2000**, *122* (7), 1282-1288.
71. Tang, Y.; Tirrell, D. A., Attenuation of the editing activity of the Escherichia coli leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* **2002**, *41* (34), 10635-45.
72. Kast, P.; Hennecke, H., Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. *J Mol Biol* **1991**, *222* (1), 99-124.
73. 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), 15285-90.
74. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A., Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues. *ChemBiochem* **2002**, *3* (2-3), 235-7.
75. 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), 9482-7.
76. Beatty, K. E.; Liu, J. C.; Xie, F.; Dieterich, D. C.; Schuman, E. M.; Wang, Q.; Tirrell, D. A., Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew Chem Int Ed Engl* **2006**, *45* (44), 7364-7.
77. Johansson, L.; Gafvelin, G.; Arner, E. S., Selenocysteine in proteins-properties and biotechnological use. *Biochim Biophys Acta* **2005**, *1726* (1), 1-13.
78. Rother, M.; Krzycki, J. A., Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea* **2010**, *2010*.
79. Chin, J. W., Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem* **2014**, *83*, 379-408.
80. Xie, J.; Schultz, P. G., An expanding genetic code. *Methods* **2005**, *36* (3), 227-38.
81. Nakamura, Y.; Gojobori, T.; Ikemura, T., Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **2000**, *28* (1), 292.
82. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001**, *292* (5516), 498-500.
83. 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), 964-7.
84. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G., A genetically encoded photocaged amino acid. *J Am Chem Soc* **2004**, *126* (44), 14306-7.
85. Bianco, A.; Townsley, F. M.; Greiss, S.; Lang, K.; Chin, J. W., Expanding the genetic code of Drosophila melanogaster. *Nat Chem Biol* **2012**, *8* (9), 748-50.
86. Davis, L.; Greiss, S., Genetic Encoding of Unnatural Amino Acids in C. elegans. *Methods Mol Biol* **2018**, *1728*, 389-408.
87. 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. *Angew Chem Int Ed Engl* **2013**, *52* (37), 9700-4.

88. Brown, W.; Liu, J.; Deiters, A., Genetic Code Expansion in Animals. *ACS Chem Biol* **2018**, *13* (9), 2375-2386.
89. Ernst, R. J.; Krogager, T. P.; Maywood, E. S.; Zanchi, R.; Beranek, V.; Elliott, T. S.; Barry, N. P.; Hastings, M. H.; Chin, J. W., Genetic code expansion in the mouse brain. *Nat Chem Biol* **2016**, *12* (10), 776-778.
90. Shao, S.; Koh, M.; Schultz, P. G., Expanding the genetic code of the human hematopoietic system. *Proceedings of the National Academy of Sciences* **2020**, *117* (16), 8845.
91. Wang, T.; Zhou, Q.; Li, F.; Yu, Y.; Yin, X.; Wang, J., Genetic Incorporation of N(epsilon)-Formyllysine, a New Histone Post-translational Modification. *Chembiochem* **2015**, *16* (10), 1440-2.
92. Neumann, H.; Peak-Chew, S. Y.; Chin, J. W., Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **2008**, *4* (4), 232-4.
93. Wilkins, B. J.; Hahn, L. E.; Heitmuller, S.; Frauendorf, H.; Valerius, O.; Braus, G. H.; Neumann, H., Genetically encoding lysine modifications on histone H4. *ACS Chem Biol* **2015**, *10* (4), 939-44.
94. Gattner, M. J.; Vrabel, M.; Carell, T., Synthesis of epsilon-N-propionyl-, epsilon-N-butryl-, and epsilon-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem Commun (Camb)* **2013**, *49* (4), 379-81.
95. Kim, C. H.; Kang, M.; Kim, H. J.; Chatterjee, A.; Schultz, P. G., Site-Specific Incorporation of ϵ -N-Crotonyllysine into Histones. *Angewandte Chemie International Edition* **2012**, *51* (29), 7246-7249.
96. Xiao, H.; Xuan, W.; Shao, S.; Liu, T.; Schultz, P. G., Genetic Incorporation of epsilon-N-2-Hydroxyisobutryl-lysine into Recombinant Histones. *ACS Chem Biol* **2015**, *10* (7), 1599-603.
97. 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), 1557-60.
98. Rogerson, D. T.; Sachdeva, A.; Wang, K.; Haq, T.; Kazlauskaitė, A.; Hancock, S. M.; Huguenin-Dezot, N.; Muqit, M. M.; Fry, A. M.; Bayliss, R.; Chin, J. W., Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat Chem Biol* **2015**, *11* (7), 496-503.
99. 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), 729-736.
100. Hoppmann, C.; Wong, A.; Yang, B.; Li, S.; Hunter, T.; Shokat, K. M.; Wang, L., Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nat Chem Biol* **2017**, *13* (8), 842-844.
101. Italia, J. S.; Peeler, J. C.; Hillenbrand, C. M.; Latour, C.; Weerapana, E.; Chatterjee, A., Genetically encoded protein sulfation in mammalian cells. *Nat Chem Biol* **2020**, *16* (4), 379-382.
102. Porter, J. J.; Jang, H. S.; Van Fossen, E. M.; Nguyen, D. P.; Willi, T. S.; Cooley, R. B.; Mehl, R. A., Genetically Encoded Protein Tyrosine Nitration in Mammalian Cells. *ACS Chem Biol* **2019**, *14* (6), 1328-1336.
103. Mondal, S.; Wang, S.; Zheng, Y.; Sen, S.; Chatterjee, A.; Thompson, P. R., Site-specific incorporation of citrulline into proteins in mammalian cells. *Nat Commun* **2021**, *12* (1), 45.
104. Tian, Y.; Lin, Q., Fitness Factors for Bioorthogonal Chemical Probes. *ACS Chem Biol* **2019**, *14* (12), 2489-2496.
105. Nguyen, S. S.; Prescher, J. A., Developing bioorthogonal probes to span a spectrum of reactivities. *Nature Reviews Chemistry* **2020**, *4* (9), 476-489.

106. Kolmel, D. K.; Kool, E. T., Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chem Rev* **2017**, *117* (15), 10358-10376.
107. Dirksen, A.; Hackeng, T. M.; Dawson, P. E., Nucleophilic catalysis of oxime ligation. *Angew Chem Int Ed Engl* **2006**, *45* (45), 7581-4.
108. Dirksen, A.; Dawson, P. E., Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjug Chem* **2008**, *19* (12), 2543-8.
109. Rayo, J.; Amara, N.; Krief, P.; Meijler, M. M., Live cell labeling of native intracellular bacterial receptors using aniline-catalyzed oxime ligation. *J Am Chem Soc* **2011**, *133* (19), 7469-75.
110. Liu, H.; Wang, L.; Brock, A.; Wong, C. H.; Schultz, P. G., A method for the generation of glycoprotein mimetics. *J Am Chem Soc* **2003**, *125* (7), 1702-3.
111. Stanley, M.; Virdee, S., Genetically Directed Production of Recombinant, Isosteric and Nonhydrolysable Ubiquitin Conjugates. *Chembiochem* **2016**, *17* (15), 1472-80.
112. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* **2002**, *41* (14), 2596-9.
113. Tornøe, C. W.; Christensen, C.; Meldal, M., Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* **2002**, *67* (9), 3057-64.
114. 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), 8720-1.
115. Eger, S.; Scheffner, M.; Marx, A.; Rubini, M., Synthesis of defined ubiquitin dimers. *J Am Chem Soc* **2010**, *132* (46), 16337-9.
116. Kofoed, C.; Riesenberger, S.; Smolikova, J.; Meldal, M.; Schoffelen, S., Semisynthesis of an Active Enzyme by Quantitative Click Ligation. *Bioconjug Chem* **2019**, *30* (4), 1169-1174.
117. Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V., Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org Lett* **2004**, *6* (17), 2853-5.
118. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* **2004**, *126* (46), 15046-7.
119. Gordon, C. G.; Mackey, J. L.; Jewett, J. C.; Sletten, E. M.; Houk, K. N.; Bertozzi, C. R., Reactivity of biarylazacyclooctynones in copper-free click chemistry. *J Am Chem Soc* **2012**, *134* (22), 9199-208.
120. Smeenk, M.; Agramunt, J.; Bongers, K. M., Recent developments in bioorthogonal chemistry and the orthogonality within. *Curr Opin Chem Biol* **2020**, *60*, 79-88.
121. Bernard, S.; Audisio, D.; Riomet, M.; Bregant, S.; Sallustrau, A.; Plougastel, L.; Decuypere, E.; Gabillet, S.; Kumar, R. A.; Elyian, J.; Trinh, M. N.; Koniev, O.; Wagner, A.; Kolodych, S.; Taran, F., Bioorthogonal Click and Release Reaction of Iminosydones with Cycloalkynes. *Angew Chem Int Ed Engl* **2017**, *56* (49), 15612-15616.
122. McKay, C. S.; Blake, J. A.; Cheng, J.; Danielson, D. C.; Pezacki, J. P., Strain-promoted cycloadditions of cyclic nitrones with cyclooctynes for labeling human cancer cells. *Chem Commun (Camb)* **2011**, *47* (36), 10040-2.
123. Blackman, M. L.; Royzen, M.; Fox, J. M., Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J Am Chem Soc* **2008**, *130* (41), 13518-9.
124. Darko, A.; Wallace, S.; Dmitrenko, O.; Machovina, M. M.; Mehl, R. A.; Chin, J. W.; Fox, J. M., Conformationally Strained trans-Cyclooctene with Improved Stability and Excellent Reactivity in Tetrazine Ligation. *Chem Sci* **2014**, *5* (10), 3770-3776.

125. Beliu, G.; Kurz, A. J.; Kuhlemann, A. C.; Behringer-Pliess, L.; Meub, M.; Wolf, N.; Seibel, J.; Shi, Z. D.; Schnermann, M.; Grimm, J. B.; Lavis, L. D.; Doose, S.; Sauer, M., Bioorthogonal labeling with tetrazine-dyes for super-resolution microscopy. *Commun Biol* **2019**, *2*, 261.
126. 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), 1861-1867.
127. 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. *Nat Chem* **2012**, *4* (4), 298-304.
128. Mayer, S. V.; Murnauer, A.; von Wrisberg, M. K.; Jokisch, M. L.; Lang, K., Photo-induced and Rapid Labeling of Tetrazine-Bearing Proteins via Cyclopropenone-Caged Bicyclononynes. *Angew Chem Int Ed Engl* **2019**, *58* (44), 15876-15882.
129. Eising, S.; Lelivelt, F.; Bongers, K. M., Vinylboronic Acids as Fast Reacting, Synthetically Accessible, and Stable Bioorthogonal Reactants in the Carboni-Lindsey Reaction. *Angew Chem Int Ed Engl* **2016**, *55* (40), 12243-7.
130. Kamber, D. N.; Liang, Y.; Blizzard, R. J.; Liu, F.; Mehl, R. A.; Houk, K. N.; Prescher, J. A., 1,2,4-Triazines Are Versatile Bioorthogonal Reagents. *J Am Chem Soc* **2015**, *137* (26), 8388-91.
131. Chen, Y.; Wu, K. L.; Tang, J.; Lored, A.; Clements, J.; Pei, J.; Peng, Z.; Gupta, R.; Fang, X.; Xiao, H., Addition of Isocyanide-Containing Amino Acids to the Genetic Code for Protein Labeling and Activation. *ACS Chem Biol* **2019**, *14* (12), 2793-2799.
132. Bandyopadhyay, A.; Cambay, S.; Gao, J., Fast Diazaborine Formation of Semicarbazide Enables Facile Labeling of Bacterial Pathogens. *J Am Chem Soc* **2017**, *139* (2), 871-878.
133. Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R., A Pictet-Spengler ligation for protein chemical modification. *Proc Natl Acad Sci U S A* **2013**, *110* (1), 46-51.
134. Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G., A genetically encoded boronate-containing amino acid. *Angew Chem Int Ed Engl* **2008**, *47* (43), 8220-3.
135. Tu, J.; Svatunek, D.; Parvez, S.; Liu, A. C.; Levandowski, B. J.; Eckvahl, H. J.; Peterson, R. T.; Houk, K. N.; Franzini, R. M., Stable, Reactive, and Orthogonal Tetrazines: Dispersion Forces Promote the Cycloaddition with Isonitriles. *Angew Chem Int Ed Engl* **2019**, *58* (27), 9043-9048.
136. Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J., Site-selective protein-modification chemistry for basic biology and drug development. *Nat Chem* **2016**, *8* (2), 103-13.
137. Klan, P.; Solomek, T.; Bochet, C. G.; Blanc, A.; Givens, R.; Rubina, M.; Popik, V.; Kostikov, A.; Wirz, J., Photoremovable protecting groups in chemistry and biology: reaction mechanisms and efficacy. *Chem Rev* **2013**, *113* (1), 119-91.
138. Luo, J.; Liu, Q.; Morihira, K.; Deiters, A., Small-molecule control of protein function through Staudinger reduction. *Nat Chem* **2016**, *8* (11), 1027-1034.
139. Li, J.; Jia, S.; Chen, P. R., Diels-Alder reaction-triggered bioorthogonal protein decaging in living cells. *Nat Chem Biol* **2014**, *10* (12), 1003-5.
140. Reille-Seroussi, M.; Mayer, S. V.; Dorner, W.; Lang, K.; Mootz, H. D., Expanding the genetic code with a lysine derivative bearing an enzymatically removable phenylacetyl group. *Chem Commun (Camb)* **2019**, *55* (33), 4793-4796.
141. Nguyen, D. P.; Mahesh, M.; Elsasser, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W., Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J Am Chem Soc* **2014**, *136* (6), 2240-3.

142. Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W., Genetically encoded photocontrol of protein localization in mammalian cells. *J Am Chem Soc* **2010**, *132* (12), 4086-8.
143. Luo, J.; Torres-Kolbus, J.; Liu, J.; Deiters, A., Genetic Encoding of Photocaged Tyrosines with Improved Light-Activation Properties for the Optical Control of Protease Function. *Chembiochem* **2017**, *18* (14), 1442-1447.
144. Wang, J.; Liu, Y.; Liu, Y.; Zheng, S.; Wang, X.; Zhao, J.; Yang, F.; Zhang, G.; Wang, C.; Chen, P. R., Time-resolved protein activation by proximal decaging in living systems. *Nature* **2019**, *569* (7757), 509-513.
145. Bose, M.; Groff, D.; Xie, J.; Brustad, E.; Schultz, P. G., The incorporation of a photoisomerizable amino acid into proteins in *E. coli*. *J Am Chem Soc* **2006**, *128* (2), 388-9.
146. Hoppmann, C.; Maslennikov, I.; Choe, S.; Wang, L., In Situ Formation of an Azo Bridge on Proteins Controllable by Visible Light. *J Am Chem Soc* **2015**, *137* (35), 11218-21.
147. Luo, J.; Samanta, S.; Convertino, M.; Dokholyan, N. V.; Deiters, A., Reversible and Tunable Photoswitching of Protein Function through Genetic Encoding of Azobenzene Amino Acids in Mammalian Cells. *Chembiochem* **2018**, *19* (20), 2178-2185.
148. Tsai, Y. H.; Essig, S.; James, J. R.; Lang, K.; Chin, J. W., Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nat Chem* **2015**, *7* (7), 554-561.
149. Pham, N. D.; Parker, R. B.; Kohler, J. J., Photocrosslinking approaches to interactome mapping. *Curr Opin Chem Biol* **2013**, *17* (1), 90-101.
150. Preston, G. W.; Wilson, A. J., Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem Soc Rev* **2013**, *42* (8), 3289-301.
151. Long, M. J.; Poganik, J. R.; Aye, Y., On-Demand Targeting: Investigating Biology with Proximity-Directed Chemistry. *J Am Chem Soc* **2016**, *138* (11), 3610-22.
152. Chmura, A. J.; Orton, M. S.; Meares, C. F., Antibodies with infinite affinity. *Proc Natl Acad Sci U S A* **2001**, *98* (15), 8480-4.
153. Stanton, B. Z.; Chory, E. J.; Crabtree, G. R., Chemically induced proximity in biology and medicine. *Science* **2018**, *359* (6380).
154. Kobayashi, T.; Hoppmann, C.; Yang, B.; Wang, L., Using Protein-Confined Proximity To Determine Chemical Reactivity. *J Am Chem Soc* **2016**, *138* (45), 14832-14835.
155. Xiang, Z.; Lacey, V. K.; Ren, H.; Xu, J.; Burban, D. J.; Jennings, P. A.; Wang, L., Proximity-enabled protein crosslinking through genetically encoding haloalkane unnatural amino acids. *Angew Chem Int Ed Engl* **2014**, *53* (8), 2190-3.
156. Furman, J. L.; Kang, M.; Choi, S.; Cao, Y.; Wold, E. D.; Sun, S. B.; Smider, V. V.; Schultz, P. G.; Kim, C. H., A genetically encoded aza-Michael acceptor for covalent cross-linking of protein-receptor complexes. *J Am Chem Soc* **2014**, *136* (23), 8411-7.
157. Xuan, W.; Li, J.; Luo, X.; Schultz, P. G., Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins. *Angew Chem Int Ed Engl* **2016**, *55* (34), 10065-8.
158. Xuan, W.; Shao, S.; Schultz, P. G., Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. *Angew Chem Int Ed Engl* **2017**, *56* (18), 5096-5100.
159. Wang, N.; Yang, B.; Fu, C.; Zhu, H.; Zheng, F.; Kobayashi, T.; Liu, J.; Li, S.; Ma, C.; Wang, P. G.; Wang, Q.; Wang, L., Genetically Encoding Fluorosulfate-l-tyrosine To React with Lysine, Histidine, and Tyrosine via SuFEx in Proteins in Vivo. *J Am Chem Soc* **2018**, *140* (15), 4995-4999.

160. 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), 15737-15741.
161. Nguyen, T. A.; Cigler, M.; Lang, K., Expanding the Genetic Code to Study Protein-Protein Interactions. *Angew Chem Int Ed Engl* **2018**, *57* (44), 14350-14361.
162. Geurink, P. P.; Prely, L. M.; van der Marel, G. A.; Bischoff, R.; Overkleeft, H. S., Photoaffinity labeling in activity-based protein profiling. *Top Curr Chem* **2012**, *324*, 85-113.
163. Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J. S., Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci* **2016**, *15*, 14.
164. Mackinnon, A. L.; Taunton, J., Target Identification by Diazirine Photo-Cross-linking and Click Chemistry. *Curr Protoc Chem Biol* **2009**, *1*, 55-73.
165. Sato, S.; Mimasu, S.; Sato, A.; Hino, N.; Sakamoto, K.; Umehara, T.; Yokoyama, S., Crystallographic study of a site-specifically cross-linked protein complex with a genetically incorporated photoreactive amino acid. *Biochemistry* **2011**, *50* (2), 250-7.
166. Young, T. S.; Young, D. D.; Ahmad, I.; Louis, J. M.; Benkovic, S. J.; Schultz, P. G., Evolution of cyclic peptide protease inhibitors. *Proc Natl Acad Sci U S A* **2011**, *108* (27), 11052-6.
167. Brunner, J.; Senn, H.; Richards, F. M., 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *J Biol Chem* **1980**, *255* (8), 3313-8.
168. Herner, A.; Marjanovic, J.; Lewandowski, T. M.; Marin, V.; Patterson, M.; Miesbauer, L.; Ready, D.; Williams, J.; Vasudevan, A.; Lin, Q., 2-Aryl-5-carboxytetrazole as a New Photoaffinity Label for Drug Target Identification. *Journal of the American Chemical Society* **2016**, *138* (44), 14609-14615.
169. Zhang, M.; Lin, S.; Song, X.; Liu, J.; Fu, Y.; Ge, X.; Fu, X.; Chang, Z.; Chen, P. R., A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nat Chem Biol* **2011**, *7* (10), 671-7.
170. Mori, H.; Ito, K., Different modes of SecY-SecA interactions revealed by site-directed in vivo photo-cross-linking. *Proc Natl Acad Sci U S A* **2006**, *103* (44), 16159-64.
171. Coin, I.; Perrin, M. H.; Vale, W. W.; Wang, L., Photo-cross-linkers incorporated into G-protein-coupled receptors in mammalian cells: a ligand comparison. *Angew Chem Int Ed Engl* **2011**, *50* (35), 8077-81.
172. Tian, Y.; Jacinto, M. P.; Zeng, Y.; Yu, Z.; Qu, J.; Liu, W. R.; Lin, Q., Genetically Encoded 2-Aryl-5-carboxytetrazoles for Site-Selective Protein Photo-Cross-Linking. *J Am Chem Soc* **2017**, *139* (17), 6078-6081.
173. Dugan, A.; Majmudar, C. Y.; Pricer, R.; Niessen, S.; Lancia, J. K.; Fung, H. Y.; Cravatt, B. F.; Mapp, A. K., Discovery of Enzymatic Targets of Transcriptional Activators via in Vivo Covalent Chemical Capture. *J Am Chem Soc* **2016**, *138* (38), 12629-35.
174. Coin, I.; Katritch, V.; Sun, T.; Xiang, Z.; Siu, F. Y.; Beyermann, M.; Stevens, R. 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), 1258-69.
175. Murray, C. I.; Westhoff, M.; Eldstrom, J.; Thompson, E.; Emes, R.; Fedida, D., Unnatural amino acid photo-crosslinking of the IKs channel complex demonstrates a KCNE1:KCNQ1 stoichiometry of up to 4:4. *Elife* **2016**, *5*.
176. Grunbeck, A.; Huber, T.; Abrol, R.; Trzaskowski, B.; Goddard, W. A., 3rd; Sakmar, T. P., Genetically encoded photo-cross-linkers map the binding site of an allosteric drug on a G protein-coupled receptor. *ACS Chem Biol* **2012**, *7* (6), 967-72.
177. Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H. R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson,

- J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J., Genomically recoded organisms expand biological functions. *Science* **2013**, *342* (6156), 357-60.
178. Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G., A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* **2013**, *52* (10), 1828-37.
179. Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Soll, D.; Liu, D. R., Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat Chem Biol* **2017**, *13* (12), 1253-1260.
180. Reinkemeier, C. D.; Girona, G. E.; Lemke, E. A., Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science* **2019**, *363* (6434).
181. Bartoschek, M. D.; Ugur, E.; Nguyen, T. A.; Rodschinka, G.; Wierer, M.; Lang, K.; Bultmann, S., Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells. *Nucleic Acids Res* **2021**.
182. 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), 276-284.
183. Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R., A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angew Chem Int Ed Engl* **2010**, *49* (18), 3211-4.
184. Willis, J. C. W.; Chin, J. W., Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat Chem* **2018**, *10* (8), 831-837.
185. Dunkelmann, D. L.; Willis, J. C. W.; Beattie, A. T.; Chin, J. W., Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat Chem* **2020**, *12* (6), 535-544.
186. Cervettini, D.; Tang, S.; Fried, S. D.; Willis, J. C. W.; Funke, L. F. H.; Colwell, L. J.; Chin, J. W., Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs. *Nat Biotechnol* **2020**, *38* (8), 989-999.
187. Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W., Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464* (7287), 441-4.
188. Fredens, J.; Wang, K.; de la Torre, D.; Funke, L. F. H.; Robertson, W. E.; Christova, Y.; Chia, T.; Schmied, W. H.; Dunkelmann, D. L.; Beranek, V.; Uttamapinant, C.; Llamazares, A. G.; Elliott, T. S.; Chin, J. W., Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **2019**, *569* (7757), 514-518.
189. Lopez-Otin, C.; Bond, J. S., Proteases: multifunctional enzymes in life and disease. *J Biol Chem* **2008**, *283* (45), 30433-7.
190. Overall, C. M.; Blobel, C. P., In search of partners: linking extracellular proteases to substrates. *Nat Rev Mol Cell Biol* **2007**, *8* (3), 245-57.
191. Sanman, L. E.; Bogyo, M., Activity-based profiling of proteases. *Annu Rev Biochem* **2014**, *83*, 249-73.
192. Chen, S.; Yim, J. J.; Bogyo, M., Synthetic and biological approaches to map substrate specificities of proteases. *Biol Chem* **2019**, *401* (1), 165-182.
193. Diamond, S. L., Methods for mapping protease specificity. *Curr Opin Chem Biol* **2007**, *11* (1), 46-51.
194. Agard, N. J.; Wells, J. A., Methods for the proteomic identification of protease substrates. *Curr Opin Chem Biol* **2009**, *13* (5-6), 503-9.
195. Vizovisek, M.; Vidmar, R.; Fonovic, M.; Turk, B., Current trends and challenges in proteomic identification of protease substrates. *Biochimie* **2016**, *122*, 77-87.

196. Gosalia, D. N.; Salisbury, C. M.; Maly, D. J.; Ellman, J. A.; Diamond, S. L., Profiling serine protease substrate specificity with solution phase fluorogenic peptide microarrays. *Proteomics* **2005**, *5* (5), 1292-8.
197. Impens, F.; Colaert, N.; Helsens, K.; Plasman, K.; Van Damme, P.; Vandekerckhove, J.; Gevaert, K., MS-driven protease substrate degradomics. *Proteomics* **2010**, *10* (6), 1284-96.
198. Gillet, L. C.; Namoto, K.; Ruchti, A.; Hoving, S.; Boesch, D.; Inverardi, B.; Mueller, D.; Coulot, M.; Schindler, P.; Schweigler, P.; Bernardi, A.; Gil-Parrado, S., In-cell selectivity profiling of serine protease inhibitors by activity-based proteomics. *Mol Cell Proteomics* **2008**, *7* (7), 1241-53.
199. Sieber, S. A.; Niessen, S.; Hoover, H. S.; Cravatt, B. F., Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat Chem Biol* **2006**, *2* (5), 274-81.
200. Yanagisawa, T.; Hino, N.; Iraha, F.; Mukai, T.; Sakamoto, K.; Yokoyama, S., Wide-range protein photo-crosslinking achieved by a genetically encoded N(epsilon)-(benzyloxycarbonyl)lysine derivative with a diazirinyl moiety. *Mol Biosyst* **2012**, *8* (4), 1131-5.
201. Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S., Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2005**, *2* (3), 201-6.
202. Kleiner, R. E.; Hang, L. E.; Molloy, K. R.; Chait, B. T.; Kapoor, T. M., A Chemical Proteomics Approach to Reveal Direct Protein-Protein Interactions in Living Cells. *Cell Chem Biol* **2018**, *25* (1), 110-120 e3.
203. Yang, T.; Li, X. M.; Bao, X.; Fung, Y. M.; Li, X. D., Photo-lysine captures proteins that bind lysine post-translational modifications. *Nat Chem Biol* **2016**, *12* (2), 70-2.
204. He, D.; Xie, X.; Yang, F.; Zhang, H.; Su, H.; Ge, Y.; Song, H.; Chen, P. R., Quantitative and Comparative Profiling of Protease Substrates through a Genetically Encoded Multifunctional Photocrosslinker. *Angew Chem Int Ed Engl* **2017**, *56* (46), 14521-14525.
205. Yang, Y.; Song, H.; He, D.; Zhang, S.; Dai, S.; Lin, S.; Meng, R.; Wang, C.; Chen, P. R., Genetically encoded protein photocrosslinker with a transferable mass spectrometry-identifiable label. *Nat Commun* **2016**, *7*, 12299.
206. Deshwal, S.; Fiedler, K. U.; Langer, T., Mitochondrial Proteases: Multifaceted Regulators of Mitochondrial Plasticity. *Annu Rev Biochem* **2020**, *89*, 501-528.
207. Schapira, A. H., Mitochondrial disease. *Lancet* **2006**, *368* (9529), 70-82.
208. Nissanka, N.; Moraes, C. T., Mitochondrial DNA damage and reactive oxygen species in neurodegenerative disease. *FEBS Lett* **2018**, *592* (5), 728-742.
209. Moehle, E. A.; Shen, K.; Dillin, A., Mitochondrial proteostasis in the context of cellular and organismal health and aging. *J Biol Chem* **2019**, *294* (14), 5396-5407.
210. Bhandari, V.; Wong, K. S.; Zhou, J. L.; Mabanglo, M. F.; Batey, R. A.; Houry, W. A., The Role of ClpP Protease in Bacterial Pathogenesis and Human Diseases. *ACS Chem Biol* **2018**, *13* (6), 1413-1425.
211. Nouri, K.; Feng, Y.; Schimmer, A. D., Mitochondrial ClpP serine protease-biological function and emerging target for cancer therapy. *Cell Death Dis* **2020**, *11* (10), 841.
212. Szczepanowska, K.; Maiti, P.; Kukat, A.; Hofsetz, E.; Nolte, H.; Senft, K.; Becker, C.; Ruzzenente, B.; Hornig-Do, H. T.; Wibom, R.; Wiesner, R. J.; Kruger, M.; Trifunovic, A., CLPP coordinates mitoribosomal assembly through the regulation of ERAL1 levels. *EMBO J* **2016**, *35* (23), 2566-2583.
213. Haynes, C. M.; Petrova, K.; Benedetti, C.; Yang, Y.; Ron, D., ClpP mediates activation of a mitochondrial unfolded protein response in *C. elegans*. *Dev Cell* **2007**, *13* (4), 467-80.

214. Seiferling, D.; Szczepanowska, K.; Becker, C.; Senft, K.; Hermans, S.; Maiti, P.; Konig, T.; Kukat, A.; Trifunovic, A., Loss of CLPP alleviates mitochondrial cardiomyopathy without affecting the mammalian UPRmt. *EMBO Rep* **2016**, *17* (7), 953-64.
215. Gispert, S.; Parganlija, D.; Klinkenberg, M.; Drose, S.; Wittig, I.; Mittelbronn, M.; Grzmil, P.; Koob, S.; Hamann, A.; Walter, M.; Buchel, F.; Adler, T.; Hrabe de Angelis, M.; Busch, D. H.; Zell, A.; Reichert, A. S.; Brandt, U.; Osiewacz, H. D.; Jendrach, M.; Auburger, G., Loss of mitochondrial peptidase Clpp leads to infertility, hearing loss plus growth retardation via accumulation of CLPX, mtDNA and inflammatory factors. *Hum Mol Genet* **2013**, *22* (24), 4871-87.
216. Cole, A.; Wang, Z.; Coyaud, E.; Voisin, V.; Gronda, M.; Jitkova, Y.; Mattson, R.; Hurren, R.; Babovic, S.; Maclean, N.; Restall, I.; Wang, X.; Jeyaraju, D. V.; Sukhai, M. A.; Prabha, S.; Bashir, S.; Ramakrishnan, A.; Leung, E.; Qia, Y. H.; Zhang, N.; Combes, K. R.; Ketela, T.; Lin, F.; Houry, W. A.; Aman, A.; Al-Awar, R.; Zheng, W.; Wienholds, E.; Xu, C. J.; Dick, J.; Wang, J. C.; Moffat, J.; Minden, M. D.; Eaves, C. J.; Bader, G. D.; Hao, Z.; Kornblau, S. M.; Raught, B.; Schimmer, A. D., Inhibition of the Mitochondrial Protease ClpP as a Therapeutic Strategy for Human Acute Myeloid Leukemia. *Cancer Cell* **2015**, *27* (6), 864-76.
217. Seo, J. H.; Rivadeneira, D. B.; Caino, M. C.; Chae, Y. C.; Speicher, D. W.; Tang, H. Y.; Vaira, V.; Bosari, S.; Palleschi, A.; Rampini, P.; Kossenkov, A. V.; Languino, L. R.; Altieri, D. C., The Mitochondrial Unfoldase-Peptidase Complex ClpXP Controls Bioenergetics Stress and Metastasis. *PLoS Biol* **2016**, *14* (7), e1002507.
218. Ishizawa, J.; Zarabi, S. F.; Davis, R. E.; Halgas, O.; Nii, T.; Jitkova, Y.; Zhao, R.; St-Germain, J.; Heese, L. E.; Egan, G.; Ruvolo, V. R.; Barghout, S. H.; Nishida, Y.; Hurren, R.; Ma, W.; Gronda, M.; Link, T.; Wong, K.; Mabanglo, M.; Kojima, K.; Borthakur, G.; MacLean, N.; Ma, M. C. J.; Leber, A. B.; Minden, M. D.; Houry, W.; Kantarjian, H.; Stogniew, M.; Raught, B.; Pai, E. F.; Schimmer, A. D.; Andreeff, M., Mitochondrial ClpP-Mediated Proteolysis Induces Selective Cancer Cell Lethality. *Cancer Cell* **2019**, *35* (5), 721-737 e9.
219. Hackl, M. W.; Lakemeyer, M.; Dahmen, M.; Glaser, M.; Pahl, A.; Lorenz-Baath, K.; Menzel, T.; Sievers, S.; Bottcher, T.; Antes, I.; Waldmann, H.; Sieber, S. A., Phenyl Esters Are Potent Inhibitors of Caseinolytic Protease P and Reveal a Stereogenic Switch for Deoligomerization. *J Am Chem Soc* **2015**, *137* (26), 8475-83.
220. Gronauer, T. F.; Mandl, M. M.; Lakemeyer, M.; Hackl, M. W.; Messner, M.; Korotkov, V. S.; Pachmayr, J.; Sieber, S. A., Design and synthesis of tailored human caseinolytic protease P inhibitors. *Chem Commun (Camb)* **2018**, *54* (70), 9833-9836.
221. Wong, K. S.; Mabanglo, M. F.; Seraphim, T. V.; Mollica, A.; Mao, Y. Q.; Rizzolo, K.; Leung, E.; Moutaoufik, M. T.; Hoell, L.; Phanse, S.; Goodreid, J.; Barbosa, L. R. S.; Ramos, C. H. I.; Babu, M.; Mennella, V.; Batey, R. A.; Schimmer, A. D.; Houry, W. A., Acyldepsipeptide Analogs Dysregulate Human Mitochondrial ClpP Protease Activity and Cause Apoptotic Cell Death. *Cell Chem Biol* **2018**, *25* (8), 1017-1030 e9.
222. Graves, P. R.; Aponte-Collazo, L. J.; Fennell, E. M. J.; Graves, A. C.; Hale, A. E.; Dicheva, N.; Herring, L. E.; Gilbert, T. S. K.; East, M. P.; McDonald, I. M.; Lockett, M. R.; Ashamalla, H.; Moorman, N. J.; Karanewsky, D. S.; Iwanowicz, E. J.; Holmuhamedov, E.; Graves, L. M., Mitochondrial Protease ClpP is a Target for the Anticancer Compounds ONC201 and Related Analogues. *ACS Chem Biol* **2019**, *14* (5), 1020-1029.
223. Fischer, F.; Langer, J. D.; Osiewacz, H. D., Identification of potential mitochondrial CLPXP protease interactors and substrates suggests its central role in energy metabolism. *Sci Rep* **2015**, *5*, 18375.

224. Hofsetz, E.; Demir, F.; Szczepanowska, K.; Kukat, A.; Kizhakkedathu, J. N.; Trifunovic, A.; Huesgen, P. F., The Mouse Heart Mitochondria N Terminome Provides Insights into ClpXP-Mediated Proteolysis. *Mol Cell Proteomics* **2020**, *19* (8), 1330-1345.
225. Fux, A.; Korotkov, V. S.; Schneider, M.; Antes, I.; Sieber, S. A., Chemical Cross-Linking Enables Drafting ClpXP Proximity Maps and Taking Snapshots of In Situ Interaction Networks. *Cell Chem Biol* **2019**, *26* (1), 48-59 e7.
226. Kang, S. G.; Dimitrova, M. N.; Ortega, J.; Ginsburg, A.; Maurizi, M. R., Human mitochondrial ClpP is a stable heptamer that assembles into a tetradecamer in the presence of ClpX. *J Biol Chem* **2005**, *280* (42), 35424-32.
227. Kang, S. G.; Ortega, J.; Singh, S. K.; Wang, N.; Huang, N. N.; Steven, A. C.; Maurizi, M. R., Functional proteolytic complexes of the human mitochondrial ATP-dependent protease, hClpXP. *J Biol Chem* **2002**, *277* (23), 21095-102.
228. Kang, S. G.; Maurizi, M. R.; Thompson, M.; Mueser, T.; Ahvazi, B., Crystallography and mutagenesis point to an essential role for the N-terminus of human mitochondrial ClpP. *J Struct Biol* **2004**, *148* (3), 338-52.
229. Baker, T. A.; Sauer, R. T., ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta* **2012**, *1823* (1), 15-28.
230. Martin, A.; Baker, T. A.; Sauer, R. T., Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease. *Mol Cell* **2007**, *27* (1), 41-52.
231. Yu, A. Y.; Houry, W. A., ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Lett* **2007**, *581* (19), 3749-57.
232. Gersch, M.; Stahl, M.; Poreba, M.; Dahmen, M.; Dziedzic, A.; Drag, M.; Sieber, S. A., Barrel-shaped ClpP Proteases Display Attenuated Cleavage Specificities. *ACS Chem Biol* **2016**, *11* (2), 389-99.
233. Chou, C.; Uprety, R.; Davis, L.; Chin, J. W.; Deiters, A., Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chemical Science* **2011**, *2* (3), 480-483.
234. Ai, H. W.; Shen, W.; Sagi, A.; Chen, P. R.; Schultz, P. G., Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *Chembiochem* **2011**, *12* (12), 1854-7.
235. Stahl, M.; Korotkov, V. S.; Balogh, D.; Kick, L. M.; Gersch, M.; Pahl, A.; Kielkowski, P.; Richter, K.; Schneider, S.; Sieber, S. A., Selective Activation of Human Caseinolytic Protease P (ClpP). *Angew Chem Int Ed Engl* **2018**, *57* (44), 14602-14607.
236. 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. *J Am Chem Soc* **2014**, *136* (44), 15577-83.
237. de Los Milagros Bassani Molinas, M.; Beer, C.; Hesse, F.; Wirth, M.; Wagner, R., Optimizing the transient transfection process of HEK-293 suspension cells for protein production by nucleotide ratio monitoring. *Cytotechnology* **2014**, *66* (3), 493-514.
238. Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; Jensen, L. J.; Mering, C. V., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **2019**, *47* (D1), D607-D613.
239. Jourdain, A. A.; Koppen, M.; Wydro, M.; Rodley, C. D.; Lightowlers, R. N.; Chrzanowska-Lightowlers, Z. M.; Martinou, J. C., GRSF1 regulates RNA processing in mitochondrial RNA granules. *Cell Metab* **2013**, *17* (3), 399-410.
240. Antonicka, H.; Sasarman, F.; Nishimura, T.; Paupe, V.; Shoubridge, E. A., The mitochondrial RNA-binding protein GRSF1 localizes to RNA granules and is required for posttranscriptional mitochondrial gene expression. *Cell Metab* **2013**, *17* (3), 386-98.

241. Sasarman, F.; Brunel-Guitton, C.; Antonicka, H.; Wai, T.; Shoubridge, E. A.; Consortium, L., LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates posttranscriptional gene expression in mitochondria. *Mol Biol Cell* **2010**, *21* (8), 1315-23.
242. Chaudhury, A.; Chander, P.; Howe, P. H., Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. *RNA* **2010**, *16* (8), 1449-62.
243. Geuens, T.; Bouhy, D.; Timmerman, V., The hnRNP family: insights into their role in health and disease. *Hum Genet* **2016**, *135* (8), 851-67.
244. Jacques, S.; van der Sloot, A. M.; C, C. H.; Coulombe-Huntington, J.; Tsao, S.; Tollis, S.; Bertomeu, T.; Culp, E. J.; Pallant, D.; Cook, M. A.; Bonneil, E.; Thibault, P.; Wright, G. D.; Tyers, M., Imipridone Anticancer Compounds Ectopically Activate the ClpP Protease and Represent a New Scaffold for Antibiotic Development. *Genetics* **2020**, *214* (4), 1103-1120.
245. Picotti, P.; Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **2012**, *9* (6), 555-66.
246. Baker, M. J.; Tatsuta, T.; Langer, T., Quality control of mitochondrial proteostasis. *Cold Spring Harb Perspect Biol* **2011**, *3* (7).
247. Kirkinezos, I. G.; Moraes, C. T., Reactive oxygen species and mitochondrial diseases. *Semin Cell Dev Biol* **2001**, *12* (6), 449-57.
248. Pryde, K. R.; Taanman, J. W.; Schapira, A. H., A LON-ClpP Proteolytic Axis Degrades Complex I to Extinguish ROS Production in Depolarized Mitochondria. *Cell Rep* **2016**, *17* (10), 2522-2531.
249. Forkink, M.; Basit, F.; Teixeira, J.; Swarts, H. G.; Koopman, W. J. H.; Willems, P., Complex I and complex III inhibition specifically increase cytosolic hydrogen peroxide levels without inducing oxidative stress in HEK293 cells. *Redox Biol* **2015**, *6*, 607-616.
250. Castello, A.; Hentze, M. W.; Preiss, T., Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrinol Metab* **2015**, *26* (12), 746-757.
251. Carlson, E. A.; Marquez, R. T.; Du, F.; Wang, Y.; Xu, L.; Yan, S. S., Overexpression of 17beta-hydroxysteroid dehydrogenase type 10 increases pheochromocytoma cell growth and resistance to cell death. *BMC Cancer* **2015**, *15*, 166.
252. Reinhard, L.; Sridhara, S.; Hallberg, B. M., The MRPP1/MRPP2 complex is a tRNA-maturation platform in human mitochondria. *Nucleic Acids Res* **2017**, *45* (21), 12469-12480.
253. Gonzalez-Mariscal, I.; Martin-Montalvo, A.; Vazquez-Fonseca, L.; Pomares-Viciano, T.; Sanchez-Cuesta, A.; Fernandez-Ayala, D. J.; Navas, P.; Santos-Ocana, C., The mitochondrial phosphatase PPTC7 orchestrates mitochondrial metabolism regulating coenzyme Q10 biosynthesis. *Biochim Biophys Acta Bioenerg* **2018**, *1859* (11), 1235-1248.
254. Gourdoupis, S.; Nasta, V.; Calderone, V.; Ciofi-Baffoni, S.; Banci, L., IBA57 Recruits ISCA2 to Form a [2Fe-2S] Cluster-Mediated Complex. *J Am Chem Soc* **2018**, *140* (43), 14401-14412.
255. Ye, Z.; Wang, S.; Zhang, C.; Zhao, Y., Coordinated Modulation of Energy Metabolism and Inflammation by Branched-Chain Amino Acids and Fatty Acids. *Front Endocrinol (Lausanne)* **2020**, *11*, 617.
256. Szczepanowska, K.; Senft, K.; Heidler, J.; Herholz, M.; Kukat, A.; Hohne, M. N.; Hofsetz, E.; Becker, C.; Kaspar, S.; Giese, H.; Zwicker, K.; Guerrero-Castillo, S.; Baumann, L.; Kauppila, J.; Rumyantseva, A.; Muller, S.; Frese, C. K.; Brandt, U.; Riemer, J.; Wittig, I.; Trifunovic, A., A salvage pathway maintains highly functional respiratory complex I. *Nat Commun* **2020**, *11* (1), 1643.

257. Strom, A.; Tong, C. L.; Wagner, C. R., Histidine triad nucleotide-binding proteins HINT1 and HINT2 share similar substrate specificities and little affinity for the signaling dinucleotide Ap4A. *FEBS Lett* **2020**, *594* (10), 1497-1505.
258. Rajasekaran, R.; Felser, A.; Nuoffer, J. M.; Dufour, J. F.; St-Pierre, M. V., The histidine triad nucleotide-binding protein 2 (HINT-2) positively regulates hepatocellular energy metabolism. *FASEB J* **2018**, *32* (9), 5143-5161.
259. Li, Z.; Wang, D.; Li, L.; Pan, S.; Na, Z.; Tan, C. Y.; Yao, S. Q., "Minimalist" cyclopropene-containing photo-cross-linkers suitable for live-cell imaging and affinity-based protein labeling. *J Am Chem Soc* **2014**, *136* (28), 9990-8.
260. Dunham, W. H.; Mullin, M.; Gingras, A. C., Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* **2012**, *12* (10), 1576-90.
261. Berggard, T.; Linse, S.; James, P., Methods for the detection and analysis of protein-protein interactions. *Proteomics* **2007**, *7* (16), 2833-42.
262. Hoppmann, C.; Lacey, V. K.; Louie, G. V.; Wei, J.; Noel, J. P.; Wang, L., Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells. *Angew Chem Int Ed Engl* **2014**, *53* (15), 3932-6.
263. Xie, X.; Li, X. M.; Qin, F.; Lin, J.; Zhang, G.; Zhao, J.; Bao, X.; Zhu, R.; Song, H.; Li, X. D.; Chen, P. R., Genetically Encoded Photoaffinity Histone Marks. *J Am Chem Soc* **2017**, *139* (19), 6522-6525.
264. Yamaguchi, A.; Matsuda, T.; Ohtake, K.; Yanagisawa, T.; Yokoyama, S.; Fujiwara, Y.; Watanabe, T.; Hoshika, T.; Sakamoto, K., Incorporation of a Doubly Functionalized Synthetic Amino Acid into Proteins for Creating Chemical and Light-Induced Conjugates. *Bioconjug Chem* **2016**, *27* (1), 198-206.
265. Lin, S.; He, D.; Long, T.; Zhang, S.; Meng, R.; Chen, P. R., Genetically encoded cleavable protein photo-cross-linker. *J Am Chem Soc* **2014**, *136* (34), 11860-3.
266. Joiner, C. M.; Breen, M. E.; Clayton, J.; Mapp, A. K., A Bifunctional Amino Acid Enables Both Covalent Chemical Capture and Isolation of in Vivo Protein-Protein Interactions. *Chembiochem* **2017**, *18* (2), 181-184.
267. Hoffmann, J. E.; Dziuba, D.; Stein, F.; Schultz, C., A Bifunctional Noncanonical Amino Acid: Synthesis, Expression, and Residue-Specific Proteome-wide Incorporation. *Biochemistry* **2018**, *57* (31), 4747-4752.
268. Dziuba, D.; Hoffmann, J. E.; Hentze, M. W.; Schultz, C., A Genetically Encoded Diazirine Analogue for RNA-Protein Photo-crosslinking. *Chembiochem* **2020**, *21* (1-2), 88-93.
269. Li, Z.; Hao, P.; Li, L.; Tan, C. Y.; Cheng, X.; Chen, G. Y.; Sze, S. K.; Shen, H. M.; Yao, S. Q., Design and synthesis of minimalist terminal alkyne-containing diazirine photo-crosslinkers and their incorporation into kinase inhibitors for cell- and tissue-based proteome profiling. *Angew Chem Int Ed Engl* **2013**, *52* (33), 8551-6.
270. Zacharias, D. A.; Violin, J. D.; Newton, A. C.; Tsien, R. Y., Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **2002**, *296* (5569), 913-6.
271. Eaton, D. L.; Bammler, T. K., Concise review of the glutathione S-transferases and their significance to toxicology. *Toxicol Sci* **1999**, *49* (2), 156-64.
272. Du, J.; Wrisberg, M. V.; Gulen, B.; Stahl, M.; Pett, C.; Hedberg, C.; Lang, K.; Schneider, S.; Itzen, A., Rab1-AMPylation by Legionella DrrA is allosterically activated by Rab1. *Nat Commun* **2021**, *12* (1), 460.
273. Pattabiraman, V. R.; Bode, J. W., Rethinking amide bond synthesis. *Nature* **2011**, *480* (7378), 471-9.
274. de Figueiredo, R. M.; Suppo, J. S.; Campagne, J. M., Nonclassical Routes for Amide Bond Formation. *Chem Rev* **2016**, *116* (19), 12029-12122.

275. Li, G.; Ma, S.; Szostak, M., Amide Bond Activation: The Power of Resonance. *Trends in Chemistry* **2020**, *2* (10), 914-928.
276. Massolo, E.; Pirola, M.; Benaglia, M., Amide Bond Formation Strategies: Latest Advances on a Dateless Transformation. *European Journal of Organic Chemistry* **2020**, *2020* (30), 4641-4651.
277. Wang, Z. A.; Cole, P. A., The Chemical Biology of Reversible Lysine Post-translational Modifications. *Cell Chem Biol* **2020**, *27* (8), 953-969.
278. Sabari, B. R.; Zhang, D.; Allis, C. D.; Zhao, Y., Metabolic regulation of gene expression through histone acylations. *Nat Rev Mol Cell Biol* **2017**, *18* (2), 90-101.
279. Xie, Z.; Zhang, D.; Chung, D.; Tang, Z.; Huang, H.; Dai, L.; Qi, S.; Li, J.; Colak, G.; Chen, Y.; Xia, C.; Peng, C.; Ruan, H.; Kirkey, M.; Wang, D.; Jensen, L. M.; Kwon, O. K.; Lee, S.; Pletcher, S. D.; Tan, M.; Lombard, D. B.; White, K. P.; Zhao, H.; Li, J.; Roeder, R. G.; Yang, X.; Zhao, Y., Metabolic Regulation of Gene Expression by Histone Lysine beta-Hydroxybutyrylation. *Mol Cell* **2016**, *62* (2), 194-206.
280. Peng, C.; Lu, Z.; Xie, Z.; Cheng, Z.; Chen, Y.; Tan, M.; Luo, H.; Zhang, Y.; He, W.; Yang, K.; Zwaans, B. M.; Tishkoff, D.; Ho, L.; Lombard, D.; He, T. C.; Dai, J.; Verdin, E.; Ye, Y.; Zhao, Y., The first identification of lysine malonylation substrates and its regulatory enzyme. *Mol Cell Proteomics* **2011**, *10* (12), M111 012658.
281. Hirschev, M. D.; Zhao, Y., Metabolic Regulation by Lysine Malonylation, Succinylation, and Glutarylation. *Mol Cell Proteomics* **2015**, *14* (9), 2308-15.
282. Zhang, Z.; Tan, M.; Xie, Z.; Dai, L.; Chen, Y.; Zhao, Y., Identification of lysine succinylation as a new post-translational modification. *Nat Chem Biol* **2011**, *7* (1), 58-63.
283. Jin, J.; He, B.; Zhang, X.; Lin, H.; Wang, Y., SIRT2 Reverses 4-Oxononanoyl Lysine Modification on Histones. *J Am Chem Soc* **2016**, *138* (38), 12304-7.
284. Rowland, E. A.; Snowden, C. K.; Cristea, I. M., Protein lipoylation: an evolutionarily conserved metabolic regulator of health and disease. *Curr Opin Chem Biol* **2018**, *42*, 76-85.
285. Mattioli, F.; Sixma, T. K., Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nat Struct Mol Biol* **2014**, *21* (4), 308-16.
286. Huang, H.; Zhang, D.; Wang, Y.; Perez-Neut, M.; Han, Z.; Zheng, Y. G.; Hao, Q.; Zhao, Y., Lysine benzylation is a histone mark regulated by SIRT2. *Nat Commun* **2018**, *9* (1), 3374.
287. Cao, J.; Sun, L.; Aramsangtienchai, P.; Spiegelman, N. A.; Zhang, X.; Huang, W.; Seto, E.; Lin, H., HDAC11 regulates type I interferon signaling through defatty-acylation of SHMT2. *Proc Natl Acad Sci U S A* **2019**, *116* (12), 5487-5492.
288. Jiang, H.; Khan, S.; Wang, Y.; Charron, G.; He, B.; Sebastian, C.; Du, J.; Kim, R.; Ge, E.; Mostoslavsky, R.; Hang, H. C.; Hao, Q.; Lin, H., SIRT6 regulates TNF-alpha secretion through hydrolysis of long-chain fatty acyl lysine. *Nature* **2013**, *496* (7443), 110-3.
289. Pattabiraman, V. R.; Ogunkoya, A. O.; Bode, J. W., Amide-Forming Ligation Reactions. In *Organic Reactions*, pp 231-592.
290. Saxon, E.; Armstrong, J. I.; Bertozzi, C. R., A "traceless" Staudinger ligation for the chemoselective synthesis of amide bonds. *Org Lett* **2000**, *2* (14), 2141-3.
291. Wang, Z. A.; Kurra, Y.; Wang, X.; Zeng, Y.; Lee, Y. J.; Sharma, V.; Lin, H.; Dai, S. Y.; Liu, W. R., A Versatile Approach for Site-Specific Lysine Acylation in Proteins. *Angew Chem Int Ed Engl* **2017**, *56* (6), 1643-1647.
292. Bode, J. W.; Fox, R. M.; Baucom, K. D., Chemoselective amide ligations by decarboxylative condensations of N-alkylhydroxylamines and alpha-ketoacids. *Angew Chem Int Ed Engl* **2006**, *45* (8), 1248-52.
293. Pusterla, I.; Bode, J. W., The mechanism of the alpha-ketoacid-hydroxylamine amide-forming ligation. *Angew Chem Int Ed Engl* **2012**, *51* (2), 513-6.

294. Wu, J.; Ruiz-Rodríguez, J.; Comstock, J. M.; Dong, J. Z.; Bode, J. W., Synthesis of human GLP-1 (7–36) by chemoselective α -ketoacid–hydroxylamine peptide ligation of unprotected fragments. *Chemical Science* **2011**, *2* (10), 1976-1979.
295. Ogunkoya, A. O.; Pattabiraman, V. R.; Bode, J. W., Sequential alpha-ketoacid-hydroxylamine (KAHA) ligations: synthesis of C-terminal variants of the modifier protein UFM1. *Angew Chem Int Ed Engl* **2012**, *51* (38), 9693-7.
296. Thuaud, F.; Rohrbacher, F.; Zwicky, A.; Bode, J. W., Photoprotected Peptide α -Ketoacids and Hydroxylamines for Iterative and One-Pot KAHA Ligations: Synthesis of NEDD8. *Helvetica Chimica Acta* **2016**, *99* (11), 868-894.
297. Wucherpennig, T. G.; Pattabiraman, V. R.; Limberg, F. R.; Ruiz-Rodriguez, J.; Bode, J. W., Traceless preparation of C-terminal alpha-ketoacids for chemical protein synthesis by alpha-ketoacid-hydroxylamine ligation: synthesis of SUMO2/3. *Angew Chem Int Ed Engl* **2014**, *53* (45), 12248-52.
298. 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), 5683-6.
299. Noda, H.; Eros, G.; Bode, J. W., Rapid ligations with equimolar reactants in water with the potassium acyltrifluoroborate (KAT) amide formation. *J Am Chem Soc* **2014**, *136* (15), 5611-4.
300. White, C. J.; Bode, J. W., PEGylation and Dimerization of Expressed Proteins under Near Equimolar Conditions with Potassium 2-Pyridyl Acyltrifluoroborates. *ACS Cent Sci* **2018**, *4* (2), 197-206.
301. Chiotellis, A.; Ahmed, H.; Betzel, T.; Tanriver, M.; White, C. J.; Song, H.; Da Ros, S.; Schibli, R.; Bode, J. W.; Ametamey, S. M., Chemoselective (18)F-incorporation into pyridyl acyltrifluoroborates for rapid radiolabelling of peptides and proteins at room temperature. *Chem Commun (Camb)* **2020**, *56* (5), 723-726.
302. Boross, G. N.; Schauenburg, D.; Bode, J. W., Chemoselective Derivatization of Folded Synthetic Insulin Variants with Potassium Acyltrifluoroborates (KATs). *Helvetica Chimica Acta* **2019**, *102* (2), e1800214.
303. Noda, H.; Bode, J. W., Synthesis and chemoselective ligations of MIDA acylboronates with O-Me hydroxylamines. *Chemical Science* **2014**, *5* (11), 4328-4332.
304. Arora, J. S.; Kaur, N.; Phanstiel, O. t., Chemoselective N-acylation via condensations of N-(benzoyloxy)amines and alpha-ketophosphonic acids under aqueous conditions. *J Org Chem* **2008**, *73* (16), 6182-6.
305. Deng, X.; Zhou, G.; Tian, J.; Srinivasan, R., Chemoselective Amide-Forming Ligation Between Acylsilanes and Hydroxylamines Under Aqueous Conditions. *Angew Chem Int Ed Engl* **2020**.
306. Phanstiel, O. I.; Wang, Q. X.; Powell, D. H.; Ospina, M. P.; Leeson, B. A., Synthesis of Secondary Amines via N-(Benzoyloxy)amines and Organoboranes. *J Org Chem* **1999**, *64* (3), 803-806.
307. 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. *Proc Natl Acad Sci U S A* **2007**, *104* (27), 11268-73.
308. Sheradsky, T.; Nov, E., Studies on the preparation of N-alkyl-O-phenylhydroxylamines. *Journal of the Chemical Society, Perkin Transactions 1* **1980**, (0), 2781-2786.
309. Kumar, S.; Sharma, R.; Garcia, M.; Kamel, J.; McCarthy, C.; Muth, A.; Phanstiel, O. t., Chemoselective amide formation using O-(4-nitrophenyl)hydroxylamines and pyruvic acid derivatives. *J Org Chem* **2012**, *77* (23), 10835-45.

310. Nevill, C. R.; Angell, P. T., A novel three-step hydroxy-deamination sequence: Conversion of lysine to 6-hydroxynorleucine derivatives. *Tetrahedron Letters* **1998**, *39* (32), 5671-5674.
311. Han, Y.; Corey, E. J., Method for the Direct Enantioselective Synthesis of Chiral Primary alpha-Amino Ketones by Catalytic alpha-Amination. *Org Lett* **2019**, *21* (1), 283-286.
312. Lin, H.; Su, X.; He, B., Protein lysine acylation and cysteine succination by intermediates of energy metabolism. *ACS Chem Biol* **2012**, *7* (6), 947-60.
313. Choudhary, C.; Weinert, B. T.; Nishida, Y.; Verdin, E.; Mann, M., The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol* **2014**, *15* (8), 536-50.
314. Simithy, J.; Sidoli, S.; Yuan, Z. F.; Coradin, M.; Bhanu, N. V.; Marchione, D. M.; Klein, B. J.; Bazilevsky, G. A.; McCullough, C. E.; Magin, R. S.; Kutateladze, T. G.; Snyder, N. W.; Marmorstein, R.; Garcia, B. A., Characterization of histone acylations links chromatin modifications with metabolism. *Nat Commun* **2017**, *8* (1), 1141.
315. Zorro Shahidian, L.; Haas, M.; Le Gras, S.; Nitsch, S.; Mourão, A.; Geerlof, A.; Margueron, R.; Michaelis, J.; Daujat, S.; Schneider, R., Succinylation of H3K122 destabilizes nucleosomes and enhances transcription. *EMBO reports* **2021**, *22* (3), e51009.
316. Molander, G. A.; Raushel, J.; Ellis, N. M., Synthesis of an acyltrifluoroborate and its fusion with azides to form amides. *J Org Chem* **2010**, *75* (12), 4304-6.
317. Ding, R.; Katebzadeh, K.; Roman, L.; Bergquist, K. E.; Lindstrom, U. M., Expanding the scope of Lewis acid catalysis in water: remarkable ligand acceleration of aqueous ytterbium triflate catalyzed Michael addition reactions. *J Org Chem* **2006**, *71* (1), 352-5.
318. Liu, S. M.; Wu, D.; Bode, J. W., One-Step Synthesis of Aliphatic Potassium Acyltrifluoroborates (KATs) from Organocuprates. *Org Lett* **2018**, *20* (8), 2378-2381.
319. Taguchi, J.; Ikeda, T.; Takahashi, R.; Sasaki, I.; Ogasawara, Y.; Dairi, T.; Kato, N.; Yamamoto, Y.; Bode, J. W.; Ito, H., Synthesis of Acylborons by Ozonolysis of Alkenylboronates: Preparation of an Enantioenriched Amino Acid Acylboronate. *Angew Chem Int Ed Engl* **2017**, *56* (44), 13847-13851.
320. Taguchi, J.; Takeuchi, T.; Takahashi, R.; Masero, F.; Ito, H., Concise Synthesis of Potassium Acyltrifluoroborates from Aldehydes through Copper(I)-Catalyzed Borylation/Oxidation. *Angew Chem Int Ed Engl* **2019**, *58* (22), 7299-7303.
321. Wu, D.; Fohn, N. A.; Bode, J. W., Catalytic Synthesis of Potassium Acyltrifluoroborates (KATs) through Chemoselective Cross-Coupling with a Bifunctional Reagent. *Angew Chem Int Ed Engl* **2019**, *58* (32), 11058-11062.
322. Wu, D.; Taguchi, J.; Tanriver, M.; Bode, J. W., Synthesis of Acylboron Compounds. *Angew Chem Int Ed Engl* **2020**, *59* (39), 16847-16858.
323. Schuhmacher, A.; Ryan, S. J.; Bode, J. W., Catalytic Synthesis of Potassium Acyltrifluoroborates (KATs) from Boronic Acids and the Thioimidate KAT Transfer Reagent. *Angewandte Chemie International Edition* **2021**, *60* (8), 3918-3922.
324. Lepage, M. L.; Lai, S.; Peressin, N.; Hadjerici, R.; Patrick, B. O.; Perrin, D. M., Direct Access to MIDA Acylboronates through Mild Oxidation of MIDA Vinylboronates. *Angew Chem Int Ed Engl* **2017**, *56* (48), 15257-15261.
325. Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I., NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **2010**, *29* (9), 2176-2179.

326. 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.
327. 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), 2601-7.
328. Schmidt, M. J.; Summerer, D., Directed Evolution of Orthogonal Pyrrolysyl-tRNA Synthetases in Escherichia coli for the Genetic Encoding of Noncanonical Amino Acids. In *Noncanonical Amino Acids: Methods and Protocols*, Lemke, E. A., Ed. Springer New York: New York, NY, 2018; pp 97-111.
329. Geiger, S. R.; Bottcher, T.; Sieber, S. A.; Cramer, P., A conformational switch underlies ClpP protease function. *Angew Chem Int Ed Engl* **2011**, *50* (25), 5749-52.
330. Rappsilber, J.; Ishihama, Y.; Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **2003**, *75* (3), 663-70.
331. Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K., A pyrrolysine analogue for protein click chemistry. *Angew Chem Int Ed Engl* **2009**, *48* (9), 1633-5.
332. Hu, H.; Faraldos, J. A.; Coates, R. M., Scope and mechanism of intramolecular aziridination of cyclopent-3-enyl-methylamines to 1-azatricyclo[2.2.1.0(2,6)]heptanes with lead tetraacetate. *J Am Chem Soc* **2009**, *131* (33), 11998-2006.
333. Yoshifuji, S.; Tanaka, K.-I.; Nitta, Y., Chemical Conversion of L- α , ω -Diamino Acids to L- ω -Carbamoyl- α -amino Acids by Ruthenium Tetroxide Oxidation. *CHEMICAL & PHARMACEUTICAL BULLETIN* **1987**, *35* (7), 2994-3001.
334. Yokokawa, F.; Sugiyama, H.; Shioiri, T.; Katagiri, N.; Oda, O.; Ogawa, H., An expeditious synthesis of pentosidine, an advanced glycation end product. *Tetrahedron* **2001**, *57* (22), 4759-4766.
335. 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), 7246-9.

7 List of abbreviations

The standard abbreviations for units, amino-acids (one letter and three letter code) and nucleobases were used.

aa	amino-acid
AAA+	ATPases associated with diverse activities
aaRS	aminoacyl-tRNA-synthetase
ABP	activity-based profiling
AcOH	acetic acid
AfBP	affinity-based profiling
AI	auto-induction
Akt1	RAC- α -serine/threonine protein kinase
Amp ^R	ampicillin resistance
BCA	bicinchoninic acid
BCAA	branched-chain amino acid
BCKD	branched chain ketoacid dehydrogenase
BCKDHA	2-oxoisovalerate dehydrogenase subunit
BCNK	bicyclo[6.1.0]nonyne lysine
BioID	proximity dependent biotin identification
BirA	biotin ligase
Boc	<i>tert</i> -Butoxycarbonyl
BocK	N ϵ -(<i>tert</i> -Butoxycarbonyl)-L-lysine
BONCAT	bioorthogonal non-canonical amino acid tagging
bp	base pair
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CaM	calmodulin
Cam ^R	chloramphenicol resistance
CaP	catabolite activator protein
CAT	chloramphenicol acetyltransferase
Cbz	carboxybenzyl
<i>C. pneumoniae</i>	<i>Chlamydophila pneumoniae</i>
<i>C. difficile</i>	<i>Clostridiodes difficile</i>
ClpP	caseinolytic protease P
ClpX	caseinolytic mitochondrial matrix peptidase chaperone subunit X
CRISPR/Cas9	clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9
CuAAC	copper catalyzed azide-alkyne cycloaddition
CV	column volume
DA	Diels-Alder
DBT	lipamide acyltransferase component of branched-chain α -keto acid dehydrogenase complex

DCM	dichlormethane
DegP	periplasmic serine endoprotease P
DIPEA	diisopropylethylamine
DMAP	4-(dimethylamino)pyridine
DMEM	Dulbecco's Modified Eagle Medium
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DrrA	effector protein of <i>Legionella pneumophila</i>
DTT	dithiothreitol
E	electrophile
ECHS1	Enoyl-CoA hydratase
EF1 α	elongation factor 1- α
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EPL	expressed protein ligation
eq.	equivalent
ERAL1	Era-like 12S mitochondrial rRNA chaperone 1
ESI-MS	electrospray ionization mass spectrometry
Et ₂ O	diethyl ether
Et ₃ N	triethylamine
EtOAc	ethyl acetate
EWG	electron withdrawing group
FA	formic acid
FBS	fetal bovine serum
FLAG	FLAG-tag, protein enrichment tag with the following sequence: DYKDDDDK
FUNCAT	fluorescent non-canonical amino acid tagging
Gal4-Gal80	galactose-responsive transcription factor 4 – transcription regulator GAL80
GCE	genetic code expansion
GFP	green fluorescent protein
GLP-1	glucagon-like peptide 1
GMMML	glycerol minimal media supplemented with leucine
GO	gene ontology
GRSF1	G-rich sequence factor 1
GTPase	hydrolases, which bind nucleotide guanosine triphosphate
HA	hemagglutinin, protein enrichment tag with the following sequence: YPYDVPDYA
HdeA	acid stress chaperone HdeA
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFIP	hexafluoroisopropanol

His ₆	hexahistidine
HSPE1	heat shock protein 10 kDa
Hsc70	heat shock cognate 71 kDa protein
HINT2	histidine triad nucleotide-binding protein 2
hnRNP	heterogeneous ribonucleoprotein particle
HRP	horseradish peroxidase
HSD17B10	= MRPP 2
Hz	Hertz
IBA57	putative transferase CAF17
iBOLT	inhibition bioorthogonal ligand tethering
iEDDA	inverse electron demand Diels-Alder
IL	interleukin
iPrOH	2-propanol
IRP1	human iron regulatory protein 1
ISG15	interferon-stimulated gene 15, ubiquitin-like
KAHA	α -ketoacid-hydroxylamine
Kan ^R	kanamycin resistance
KAT	potassium acyltrifluoroborates
KDAC	lysine deacylase
KO	knockout
LB	lysogeny broth
LC-MS	liquid chromatography mass spectrometry
LC-MS/MS	liquid chromatography tandem mass spectrometry
LONP1	Lon protease homolog
LR-MS	low resolution mass spectrometry
LRPPRC	leucine-rich PPR motif containing
<i>M. alvus</i>	<i>Methanomethylophilus alvus</i>
mAU	milli absorption unit
MeCN	acetonitrile
MEK1/2	MAPK/ERK kinase 1/2
MeOH	methanol
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MIDA	<i>N</i> -methyliminodiacetyl acylboronates
<i>Mj</i>	<i>Methanocaldococcus jannaschii</i>
<i>Mb/Mm</i>	<i>Methanosarcina barkeri/mazei</i>
MMP	matrix metallo protease
MRPP1/2	= TRMT10C/HSD17B10; complex of tRNA methyltransferase 10 homolog C/3-hydroxyacyl-CoA dehydrogenase type 2
mt	mitochondrial
MWCO	molecular weight cut-off
NAD ⁺	nicotinamide adenine dinucleotide, oxidised
NC	nitrocellulose

NCL	native chemical ligation
NDUFV2	NADH dehydrogenase [ubiquinone] flavoprotein 2
NEDD8	neural precursor cell expressed, developmentally downregulated 8
NiNTA	Ni ²⁺ -nitriloacetic acid agarose resin
NMP	<i>N</i> -methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
NSI	nanospray ionisation
Nu	nucleophile
<i>OaAEP1</i>	<i>Oldenlandia affinis</i> asparaginyl endopeptidase 1
OD	optical density
OmpC	outer membrane porin C
o.n.	overnight
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PDH	pyruvate dehydrogenase
PDHX	pyruvate dehydrogenase protein X
PEG	polyethylene glycol
<i>P. falciparum</i>	<i>Plasmodium falciparum</i>
<i>P. anserina</i>	<i>Podospora anserina</i>
POI	protein of interest
PPI	protein-protein interaction
PPTC7	protein phosphatase PTC7 (protein phosphatase 2C homolog 7) homolog
PSM	peptide spectrum matches
PTAD	phenyl-3H-1,2,4-triazole-3,5(4H)-diones
<i>p</i> TsOH	para-toluenesulfonic acid
PyIRS	pyrrolysyl tRNA synthetase
Rab1b	Ras-related protein Rab1b
RCC	respiratory chain complex
RF	release factor
RNA; m/t	ribonucleic acid; messenger/transfer
ROS	reactive oxygen species
RP-HPLC	reverse phase high-performance liquid chromatography
RPS17	40S ribosomal protein S17
r.t.	room temperature
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sfGFP	super-folder green fluorescent protein
SEC	size-exclusion chromatography
SLIM	site-directed ligation independent mutagenesis
SLIRP	mitochondrial SRA stem-loop-interacting RNA-binding protein
SOB	super optimal broth

SOC	super optimal broth with catabolite repression
SPPS	solid-phase peptide chemistry
SrtA	sortase A
Strep	Strep-tag, protein enrichment tag with the following sequence: WSHPNFEK
STRING	search tool for the retrieval of interacting genes/proteins
SUMO	small ubiquitin-like modifier
T _a /T _m	annealing temperature/melting temperature
TAE	buffer consisting of Tris base, acetic acid and EDTA
TAMRA	carboxytetramethylrhodamine
TBS-T	tris buffered saline with Tween 20
TCA	tricarboxylic acid cycle
Tet ^R	tetracycline resistance
TEV	tobacco etch virus
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIC	total ion count
TLC	thin-layer chromatography
TOMM	mitochondrial import receptor subunit
TPAC	thiophosphoro alkyne dichloridate
TRAP1	heat shock protein 75 kDa
Tris	tris(hydroxymethyl)aminomethane
TRMT10C	= MRPP1
UAA	unnatural amino acid
Ub	ubiquitin
Ubc9	SUMO-conjugating enzyme Ubc9
Ubl	ubiquitin-like
UFM1	ubiquitin-fold modifier 1
UPR ^{mt}	mitochondrial unfolded protein response
UQCRC1	ubiquinol cytochrome c reductase subunit 1
USP2	ubiquitin specific peptidase 2
UV	ultraviolet
WB	western blot
wt	wildtype
Ypt1	GTP-binding protein Ypt1
<i>Y. pestis</i>	<i>Yersinia pestis</i>

8 Appendix

8.1 Supporting information for chapter 2

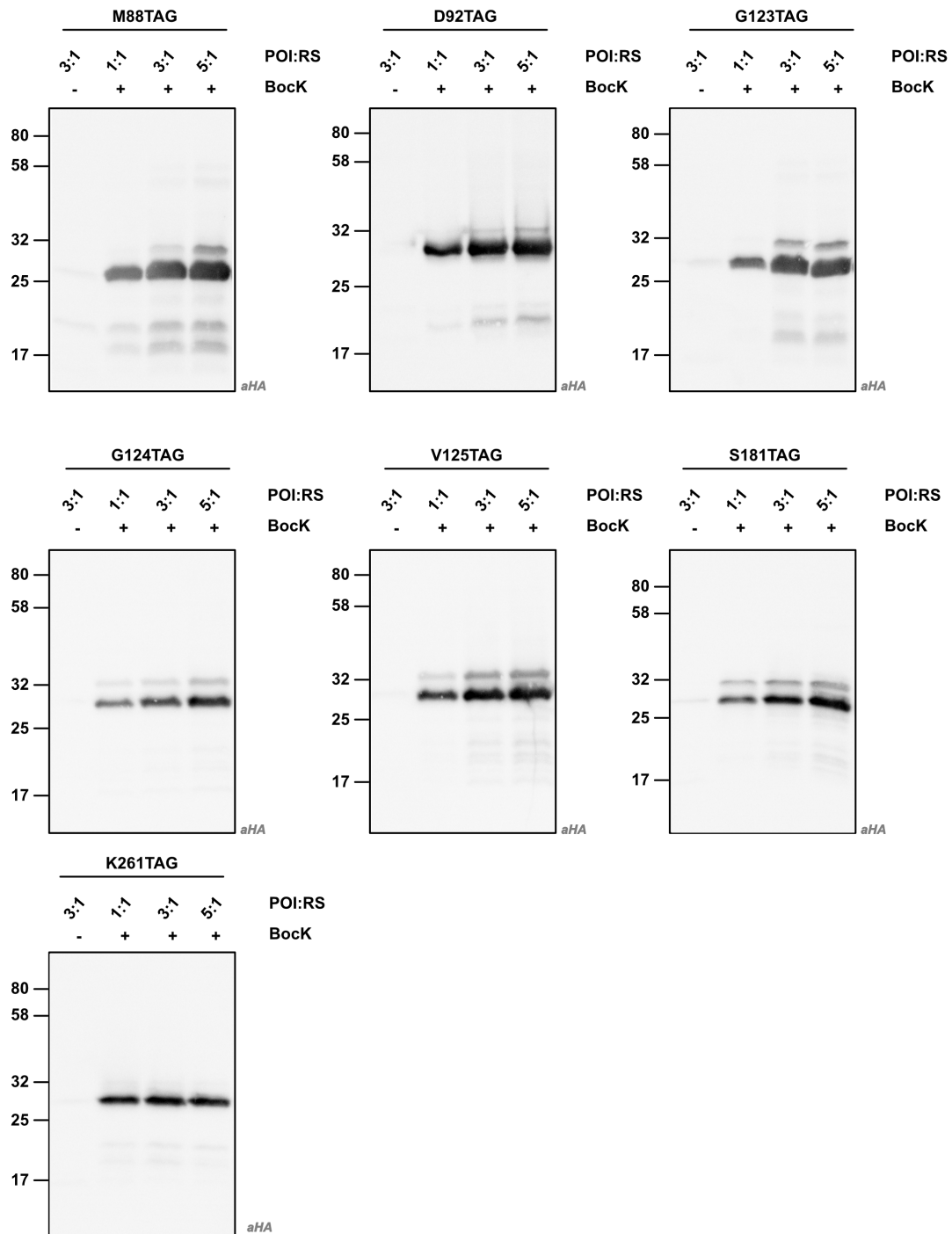


Fig. SI 2.1 I Transfection ratio screen for optimal amber suppression of hClpP-TAG. Transfections were performed with a constant amount DNA as listed in table 5.4. Protein amounts were adjusted according BCA assay and incorporation efficiencies detected by WB with an anti-HA-HRP antibody. A 3:1 POI:aaRS was found optimal in terms of signal/noise ratio.

Trapping experiments with hClpP amber suppressed mutants in HEK293T

The following tables list all enriched proteins from the crosslinking and proteomics studies. The tables were adapted from the PhD thesis (2021) of Thomas F. Gronauer (TU Munich) with permission.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
10 kDa heat shock protein, mitochondrial	HSPE1	2,528060913	2,357710492	-
Dihydrolypoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	2,465864182	2,80002771	+
39S ribosomal protein L13, mitochondrial	MRPL13	2,259292603	2,699404838	+
Serine beta-lactamase-like protein LACTB, mitochondrial	LACTB	2,176890055	2,572455941	-
Clathrin light chain B	CLTB	2,020890554	2,431020243	-
Clathrin light chain A	CLTA	1,966786702	2,557270996	-
Acetyltransferase component of pyruvate dehydrogenase complex	DLAT	1,89736557	3,709135937	+
Clathrin heavy chain; Clathrin heavy chain 1	CLTC	1,736051559	2,961035244	-
Mitochondrial import receptor subunit TOM22 homolog	TOMM22	1,710225423	1,440738034	-
Hsc70-interacting protein	ST13	1,659377416	1,952781386	-
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	1,573442459	2,736985754	+
Protein scribble homolog	SCRIB	1,551549276	3,539765357	-
Unconventional myosin-Id	MYO1D	1,484951655	1,64892459	-
Stress-induced-phosphoprotein 1	STIP1	1,29959933	2,367580932	+
Non-specific protein-tyrosine kinase	YES1	1,120776494	1,381649232	-
Malate dehydrogenase; Malate dehydrogenase, mitochondrial	MDH2	1,095053355	2,650466362	+
Enoyl-CoA hydratase, mitochondrial	ECHS1	1,090739568	1,345539571	+
Histone-arginine methyltransferase CARM1	CARM1	1,032866796	2,484322024	-

Table SI 2.1 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments in HEK293T with hClpP-M88dK. For annotation of hClpP substrates ref. ^{212, 216, 218, 223-224, 244, 256} were considered and proteins marked with '+', if they were previously mentioned in any of the references. Otherwise, protein were marked with '-', if no previous connection to ClpP was identified.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	6,518294017	3,110163196	-
Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	4,536731084	3,787498035	-
Lupus La protein	SSB	4,239100138	3,281673152	-
RNA-binding motif, single-stranded-interacting protein 1	RBMS1	4,080912272	3,334308368	-
Threonine synthase-like 1	THNSL1	4,00594902	2,76971573	+

Nuclease-sensitive element-binding protein 1	YBX1	3,788115819	2,647808991	-
5-3 exoribonuclease 2	XRN2	3,784579595	1,987684658	-
Polyadenylate-binding protein 2	PABPN1	3,684401194	5,265217984	-
Polyadenylate-binding protein 4	PABPC4	3,624721527	3,165265389	-
Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	3,583220164	2,000476081	-
Glycine cleavage system H protein, mitochondrial	GCSH	3,544822693	2,06076592	+
Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	3,502468745	4,691630576	-
Spermatid perinuclear RNA-binding protein	HEL162	3,452546438	4,487742728	-
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	3,413604736	4,007088155	-
Probable tRNA pseudouridine synthase 2	TRUB2	3,179822286	2,300139499	-
THUMP domain-containing protein 1	THUMPD1	3,174427032	2,603579596	-
Polyadenylate-binding protein	PABPC1	3,137008031	1,753759417	-
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	2,919965108	2,420918282	-
10 kDa heat shock protein, mitochondrial	HSPE1	2,761142095	3,09089524	-
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	2,755172729	4,298819047	-
28S ribosomal protein S28, mitochondrial	MRPS28	2,729255676	3,446636279	+
RNA-binding protein Musashi homolog 2	MSI2	2,702177684	2,971068803	-
Isoleucine--tRNA ligase, mitochondrial	IARS2	2,593344371	2,965459546	+
Insulin-like growth factor 2 mRNA-binding protein 2	IGF2BP2	2,534356435	2,700772355	-
ATP-dependent RNA helicase DHX36	DHX36	2,501374563	4,009703465	-
28S ribosomal protein S34, mitochondrial	MRPS34	2,4945666	3,644553291	-
28S ribosomal protein S6, mitochondrial	MRPS6	2,381146113	2,339273589	+
39S ribosomal protein L1, mitochondrial	MRPL1	2,373535792	1,615456931	+
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	2,367306391	3,13102073	-
Enoyl-CoA hydratase, mitochondrial	ECHS1	2,342359543	3,047345112	+
Probable E3 ubiquitin-protein ligase makorin-2	MKRN2	2,319894155	3,55549555	-
39S ribosomal protein L3, mitochondrial	MRPL3	2,30510203	2,612886642	-
28S ribosomal protein S23, mitochondrial	MRPS23	2,268669128	1,625540672	+
Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	2,225276311	2,443275937	-
G-rich sequence factor 1	GRSF1	2,191397985	2,477793642	+
Dihydrolipoyl dehydrogenase; Dihydrolipoyl dehydrogenase, mitochondrial	DLD	2,153752009	2,769800596	+
39S ribosomal protein L15, mitochondrial	MRPL15	2,151285172	2,905576994	+
28S ribosomal protein S18b, mitochondrial	MRPS18B	2,08224678	2,633008426	+
Insulin-like growth factor 2 mRNA-binding protein 1	IGF2BP1	2,064293543	2,941183194	-
La-related protein 1	LARP1	2,006263097	2,120226653	-
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	2,002729416	1,877540646	+
39S ribosomal protein L11, mitochondrial	MRPL11	2,002288183	2,247417398	-
28S ribosomal protein S9, mitochondrial	MRPS9	1,978295008	2,517465445	+
39S ribosomal protein L4, mitochondrial	MRPL4	1,969324112	5,952863687	-
Heterogeneous nuclear ribonucleoprotein L	HNRNPL	1,955095927	2,677170871	-
39S ribosomal protein L43, mitochondrial	MRPL43	1,823653539	3,138160528	+
2-oxoglutarate dehydrogenase, mitochondrial	OGDH	1,819985708	2,856320033	+

28S ribosomal protein S5, mitochondrial	MRPS5	1,779591242	2,439905779	+
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT	1,758091609	3,175318594	-
Heat shock protein 75 kDa, mitochondrial	TRAP1	1,741571426	3,627208492	+
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	1,727851868	3,730399994	+
28S ribosomal protein S16, mitochondrial	MRPS16	1,673002879	3,86417554	+
Constitutive coactivator of PPAR-gamma-like protein 1	FAM120A	1,664419174	2,432006898	-
Acetyltransferase component of pyruvate dehydrogenase complex	DLAT	1,646781286	2,619956461	+
28S ribosomal protein S31, mitochondrial	MRPS31	1,642515182	2,495011687	+
NF-kappa-B-repressing factor	NKRF	1,629261017	2,198007411	-
39S ribosomal protein L21, mitochondrial	MRPL21	1,614315033	2,751716284	+
39S ribosomal protein L13, mitochondrial	MRPL13	1,536867778	1,761671464	+
39S ribosomal protein L45, mitochondrial	MRPL45	1,51011912	2,818758442	+
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	1,493288676	4,410884257	+
Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	1,483260473	2,589449466	+
Protein deglycase DJ-1	PARK7	1,470893224	1,40334102	-
28S ribosomal protein S22, mitochondrial	MRPS22	1,408541361	3,16183956	+
Interleukin enhancer-binding factor 3	ILF3	1,39076678	3,376662231	-
Heterogeneous nuclear ribonucleoprotein K	HNRPK	1,387111028	4,044415826	-
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	1,363213857	3,179289191	+
Heterogeneous nuclear ribonucleoprotein A0	HNRNPA0	1,350752513	3,024946219	-
SRA stem-loop-interacting RNA-binding protein, mitochondrial	SLIRP	1,34758695	2,431424984	+
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	1,34037145	2,4306181	-
40S ribosomal protein S17	RPS17	1,305379232	3,539899275	-
39S ribosomal protein L51, mitochondrial	MRPL51	1,294469198	3,03400986	-
Williams-Beuren syndrome chromosomal region 16 protein	WBSCR16	1,292111715	1,659171632	+
39S ribosomal protein L39, mitochondrial	MRPL39	1,289264679	6,064794061	+
Pentatricopeptide repeat domain-containing protein 3, mitochondrial	PTCD3	1,275725047	2,906635339	+
28S ribosomal protein S26, mitochondrial	MRPS26	1,262624741	3,15316929	+
RNA-binding motif protein, X chromosome	RBMX	1,254060109	1,732152367	-
Protein ELYS	AHCTF1	1,225773493	1,577281647	-
Interleukin enhancer-binding factor 2	ILF2	1,18741099	1,811837932	-
28S ribosomal protein S11, mitochondrial	MRPS11	1,187133153	2,377316847	+
39S ribosomal protein L17, mitochondrial	MRPL17	1,182820002	3,031551281	+
Far upstream element-binding protein 3	FUBP3	1,172859828	1,804180042	-
28S ribosomal protein S29, mitochondrial	DAP3	1,166245143	1,970002487	+
Cysteine desulfurase, mitochondrial	NFS1	1,087621053	2,13564208	+
L-2-hydroxyglutarate dehydrogenase, mitochondrial	L2HGDH	1,063916524	1,589935736	-
39S ribosomal protein L44, mitochondrial	MRPL44	1,001045227	1,504273452	+

Table SI 2.2 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments in HEK293T with hClpP-D92dK. For annotation of hClpP substrates ref. ²¹², ²¹⁶, ²¹⁸, ²²³⁻²²⁴, ²⁴⁴, ²⁵⁶ were considered and proteins marked with '+', if they were previously mentioned in any of the references. Otherwise, protein were marked with '-', if no previous connection to ClpP was identified.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
26S proteasome non-ATPase regulatory subunit 10	PSMD10	2,599224091	3,099309091	-
Low molecular weight phosphotyrosine protein phosphatase	ACP1	2,394634883	2,20235606	-
Cytochrome c oxidase subunit 5A, mitochondrial	COX5A	1,666353861	2,481374747	+
L-2-hydroxyglutarate dehydrogenase, mitochondrial	L2HGDH	1,296777725	2,262133108	-
Acetyltransferase component of pyruvate dehydrogenase complex	DLAT	1,292052587	1,997436201	+
Dihydrolipoylysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	1,283274333	2,210274591	+
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	1,192612966	3,005726645	+
Vimentin	HEL113	1,184209188	2,364686983	+
SEC23-interacting protein	SEC23IP	1,117851893	1,597814887	-
Tubulin alpha-1A chain;Tubulin alpha-3C/D chain;Tubulin alpha-3E chain	TUBA1A	1,113746643	1,816653826	-
MAP7 domain-containing protein 1	MAP7D1	1,087779363	1,465578942	-
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT	1,006003698	3,013900794	-

Table SI 2.3 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments in HEK293T with hClpP-G123dK. For annotation of hClpP substrates ref. ^{212, 216, 218, 223-224, 244, 256} were considered and proteins marked with '+', if they were previously mentioned in any of the references. Otherwise, protein were marked with '-', if no previous connection to ClpP was identified.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
Glycine cleavage system H protein, mitochondrial	GCSH	3,83144188	3,39578492	+
Cytochrome c oxidase subunit 4 isoform 1, mitochondrial	COX4I1	2,07114728	3,26673057	+
Low molecular weight phosphotyrosine protein phosphatase	ACP1	1,80448977	1,30789463	-
10 kDa heat shock protein, mitochondrial	HSPE1	1,68412463	2,08915157	-
Inorganic pyrophosphatase 2, mitochondrial	PPA2	1,62349447	2,91607884	+
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT	1,56262271	2,99743635	-
Mitochondrial import inner membrane translocase subunit Tim8 B	TIMM8B	1,42010689	3,72810838	-
Stress-induced-phosphoprotein 1	STIP1	1,37785149	2,74930439	+
Polyadenylate-binding protein 2	PABPN1	1,31002935	1,34611824	-
Growth arrest and DNA damage-inducible proteins-interacting protein 1	GADD45GIP1	1,22106043	1,4733748	-
Mitochondrial import inner membrane translocase subunit Tim13	TIMM13	1,18352318	2,32393757	-
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	1,05083529	3,76064284	+
Heat shock protein 75 kDa, mitochondrial	TRAP1	1,01244863	3,36180087	+

Table SI 2.4 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments in HEK293T with hClpP-G124 dK. For annotation of hClpP substrates ref. ^{212, 216, 218, 223-224, 244, 256} were considered and proteins

marked with '+', if they were previously mentioned in any of the references. Otherwise, protein were marked with '-', if no previous connection to ClpP was identified.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
Protein deglycase DJ-1	PARK7	2,57212321	2,39455614	-
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT	2,52292633	3,0810378	-
L-2-hydroxyglutarate dehydrogenase, mitochondrial	L2HGDH	2,40566762	1,6841459	-
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	1,88935089	3,07300658	+
Proteasome activator complex subunit 3	PSME3	1,80785497	2,90815977	-
Acetyltransferase component of pyruvate dehydrogenase complex	DLAT	1,74389521	2,96192323	+
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	1,62713242	2,04765556	+
2-oxoglutarate dehydrogenase, mitochondrial	OGDH	1,41948636	4,23615689	+
Mitochondrial import receptor subunit TOM70	TOMM70A	1,4188385	2,55492142	-
Dihydrolipoyl dehydrogenase, mitochondrial	DLD	1,3963871	2,67590128	+
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	1,3935407	3,05525474	+
Pyruvate dehydrogenase E1 component subunit alpha	PDHA1	1,36758105	2,68051774	-
Coiled-coil domain-containing protein 80	CCDC80	1,16726494	1,46698247	-

Table SI 2.5 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments in HEK293T with hClpP-V125dK. For annotation of hClpP substrates ref. ^{212, 216, 218, 223-224, 244, 256} were considered and proteins marked with '+', if they were previously mentioned in any of the references. Otherwise, protein were marked with '-', if no previous connection to ClpP was identified.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
39S ribosomal protein L39, mitochondrial	MRPL39	3,69523875	2,87025648	+
Mitochondrial import receptor subunit TOM40 homolog	TOMM40	3,61561521	3,17518351	-
SRA stem-loop-interacting RNA-binding protein, mitochondrial	SLIRP	2,97568321	1,65137757	+
Cleavage and polyadenylation specificity factor subunit 4	CPSF4	2,9491717	2,7285779	-
Small nuclear ribonucleoprotein G	SNRPG	2,80916087	2,04616458	-
Tripeptidyl-peptidase 1	TPP1	2,78866768	1,32640606	-
Transducin beta-like protein 3	TBL3	2,78135109	1,43187895	-
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT	2,37619591	3,14234269	-
Septin-2	Sep 02	2,26404635	2,4847726	-
Mitochondrial import receptor subunit TOM6 homolog	TOMM6	2,14990807	1,42698883	+
Short/branched chain specific acyl-CoA dehydrogenase, mitochondrial	ACADSB	2,05969874	3,37409558	+
26S proteasome non-ATPase regulatory subunit 4	PSMD4	1,93259303	2,5893379	+

Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	1,77110227	4,17378336	+
Acetyltransferase component of pyruvate dehydrogenase complex	DLAT	1,65835698	3,3096671	+
Mitochondrial import receptor subunit TOM5 homolog	TOMM5	1,62214979	2,86460387	-
Nucleoporin Nup43	NUP43	1,62046878	1,42195773	-
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	1,518013	3,07525562	+
Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	1,43370883	1,71404677	-
Enoyl-CoA hydratase, mitochondrial	ECHS1	1,40692139	2,27185593	+
10 kDa heat shock protein, mitochondrial	HSPE1	1,394804	1,56747413	-
rRNA methyltransferase 2, mitochondrial	HEL97	1,27911504	2,2129185	-
Cyclin-dependent kinase 4	CDK4	1,20119413	3,4476943	-
Mitochondrial import receptor subunit TOM22 homolog	TOMM22	1,15445201	2,30656297	-
Heat shock protein 75 kDa, mitochondrial	TRAP1	1,08820279	2,82503219	+

Table SI 2.6 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments in HEK293T with hClpP-S181dK. For annotation of hClpP substrates ref. ^{212, 216, 218, 223-224, 244, 256} were considered and proteins marked with '+', if they were previously mentioned in any of the references. Otherwise, protein were marked with '-', if no previous connection to ClpP was identified.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
Enoyl-CoA hydratase, mitochondrial	ECHS1	7,08297412	6,14249209	+
10 kDa heat shock protein, mitochondrial	HSPE1	5,22580592	1,72892543	-
Isovaleryl-CoA dehydrogenase, mitochondrial	IVD	5,19725927	3,41344426	+
Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	HADH	4,82912254	4,47425768	+
Malate dehydrogenase; Malate dehydrogenase, mitochondrial	MDH2	4,56429545	5,94513238	+
Sepiapterin reductase	SPR	4,56307348	3,7642302	-
99,3% Presequence Protease, pirtilysin metalloproteinase	PITRM1	4,4543368	5,58843754	+
Aconitate hydratase, mitochondrial	ACO2	4,26304881	4,06072425	+
Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	4,13931402	3,85969321	+
Short/branched chain specific acyl-CoA dehydrogenase, mitochondrial	ACADSB	3,87617238	2,68326046	+
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	3,68018341	4,93766787	+
Citrate synthase; Citrate synthase, mitochondrial	CS	3,55343119	5,48344718	+
Acetyltransferase component of pyruvate dehydrogenase complex	DLAT	3,54185422	1,80263618	+
Polyribonucleotide nucleotidyltransferase 1, mitochondrial	PNPT1	3,53656324	5,08480506	+
Isocitrate dehydrogenase [NAD] subunit, mitochondrial; Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	IDH3A	3,50106875	3,3788779	+
Single-stranded DNA-binding protein, mitochondrial	SSBP1	3,35598755	3,86522097	+
Serine hydroxymethyltransferase, mitochondrial	SHMT2	3,34568469	5,2228862	+

Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	3,28835551	3,95726853	+
Tyrosine--tRNA ligase, mitochondrial	YARS2	3,21464602	2,86100621	+
Dihydrolipoyl dehydrogenase, mitochondrial	DLD	3,1332798	4,18157035	+
2-oxoglutarate dehydrogenase, mitochondrial	OGDH	3,05487124	3,66429743	+
Acyl-coenzyme A thioesterase 1; Acyl-coenzyme A thioesterase 2, mitochondrial	ACOT1	2,97574298	3,16847496	+
Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT	2,96562449	2,48880334	-
Electron transfer flavoprotein subunit beta	ETFB	2,90824827	2,26341865	+
Fumarate hydratase, mitochondrial	FH	2,86964989	2,44974423	-
Mitochondrial ribonuclease P protein 1	TRMT10C	2,78555107	1,87222244	+
Mitochondrial import receptor subunit TOM70	TOMM70A	2,76650874	1,39449238	-
Acyl-CoA dehydrogenase family member 9, mitochondrial	ACAD9	2,71775881	4,27364214	+
Protein deglycase DJ-1	PARK7	2,71552658	1,65359826	-
Heat shock protein 75 kDa, mitochondrial	TRAP1	2,70960172	3,19950177	+
Cob(I)yrinic acid a,c-diamide adenosyltransferase, mitochondrial	MMAB	2,69214185	2,97033291	-
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	2,63960838	3,00161336	+
Isoleucine-tRNA ligase, mitochondrial	IARS2	2,61310514	3,81692843	+
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	2,56129265	2,89057994	+
ATP synthase subunit d, mitochondrial	ATP5H	2,55917295	1,38680909	+
Glutaminase kidney isoform, mitochondrial	GLS	2,5211792	2,01331491	+
3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	2,46673584	4,53563713	+
ATP synthase F(0) complex subunit B1, mitochondrial	ATP5F1	2,35659091	1,82896105	-
Mitochondrial intermediate peptidase	MIPEP	2,28592555	2,47335817	+
Glycine--tRNA ligase	GARS	2,27018166	1,59514186	+
Aldehyde dehydrogenase, mitochondrial	ALDH2	2,2700901	3,4454459	-
Aspartate--tRNA ligase, mitochondrial	DARS2	2,26568476	2,72366543	-
NAD-dependent malic enzyme, mitochondrial	ME2	2,16055552	3,58115927	-
Monofunctional C1-tetrahydrofolate synthase, mitochondrial	MTHFD1L	2,1106542	3,40480894	+
Electron transfer flavoprotein subunit alpha, mitochondrial	ETFFA	1,99602763	4,13475842	+
ATP synthase subunit beta; ATP synthase subunit beta, mitochondrial	ATP5B	1,99509048	3,31934939	+
Hydroxymethylglutaryl-CoA lyase, mitochondrial	HMGCL	1,94898097	1,65773869	+
ATP synthase mitochondrial F1 complex assembly factor 2	ATPAF2	1,93302917	1,60720171	-
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6	NDUFA6	1,90193367	1,82497091	+
Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial	PDHA1	1,89166832	3,68221395	+
Ornithine aminotransferase, mitochondrial	OAT	1,83684095	2,7689186	+
Isocitrate dehydrogenase [NAD] subunit, mitochondrial	IDH3B	1,83460172	3,62711305	+

Succinyl-CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial	SUCLG1	1,83005333	4,50265362	+
Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADM	1,81834157	4,32905502	+
Complex III assembly factor LYRM7	LYRM7	1,74750392	1,82877052	+
Glycine dehydrogenase (decarboxylating), mitochondrial	GLDC	1,65087446	1,52630458	+
2,4-dienoyl-CoA reductase, mitochondrial	DECR1	1,62497393	1,49161777	+
Mitochondrial import inner membrane translocase subunit TIM44	TIMM44	1,61522675	2,26591336	+
G-rich sequence factor 1	GRSF1	1,60489273	2,3446218	+
ATP synthase subunit g, mitochondrial	ATP5L	1,57912763	1,53196088	-
Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	ECH1	1,57240359	2,11907915	+
ATP synthase subunit O, mitochondrial	ATP5O	1,54561996	2,93276679	-
39S ribosomal protein L44, mitochondrial	MRPL44	1,54289563	2,88366699	+
Cytoplasmic dynein 1 light intermediate chain 1	DYNC1LI1	1,47920227	1,35880247	-
3-hydroxyisobutyrate dehydrogenase, mitochondrial	HIBADH	1,45746994	2,16669445	-
Probable leucine--tRNA ligase, mitochondrial	LARS2	1,41318893	1,96446322	+
Glutamate dehydrogenase 1, mitochondrial	GLUD1	1,39283371	3,3153833	+
Aspartate aminotransferase, mitochondrial	GOT2	1,38335546	1,97231956	+
28S ribosomal protein S28, mitochondrial	MRPS28	1,33673096	2,01300666	+
28S ribosomal protein S27, mitochondrial	MRPS27	1,3307972	2,73975322	-
3-hydroxyisobutyryl-CoA hydrolase, mitochondrial	HIBCH	1,30554263	2,89547207	-
Ran GTPase-activating protein 1	RANGAP1	1,25149218	1,7788858	-
28S ribosomal protein S34, mitochondrial	MRPS34	1,22402382	1,5503205	-
39S ribosomal protein L1, mitochondrial	MRPL1	1,22068787	1,7740464	+
Cytosol aminopeptidase	LAP3	1,2072506	1,48275179	-
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	MCCC2	1,13636843	3,04760163	+
SRA stem-loop-interacting RNA-binding protein, mitochondrial	SLIRP	1,12037404	1,84222529	+
Ubiquitin carboxyl-terminal hydrolase; Probable ubiquitin carboxyl-terminal hydrolase FAF-X	USP9X	1,05227025	1,31671925	-
ATP synthase subunit alpha, mitochondrial	ATP5A1	1,00348409	3,88912011	+

Table SI 2.7 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments in HEK293T with hClpP-K261dK. For annotation of hClpP substrates ref. ^{212, 216, 218, 223-224, 244, 256} were considered and proteins marked with '+', if they were previously mentioned in any of the references. Otherwise, protein were marked with '-', if no previous connection to ClpP was identified.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
CDKN2A-interacting protein	CDKN2AIP	4.040013777	5.829336166	+
Glycine cleavage system H protein, mitochondrial	GCSH	3.547040595	5.622687658	-
THUMP domain-containing protein 1	THUMPD1	2.25853187	4.467810313	-
L-2-hydroxyglutarate dehydrogenase, mitochondrial	L2HGDH	2.627576481	3.54887708	-
RNA-binding protein Musashi homolog 1	MSI1	3.157466777	3.508930206	+
RNA-binding protein Musashi homolog 2	MSI2	2.923710133	3.060661316	-
Threonine synthase-like 1	THNSL1	2.611644794	2.965991338	-

Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2;	3.977559625	2.948937734	-
Polyadenylate-binding protein 2	PABPN1	2.594563442	2.659805934	-
Inorganic pyrophosphatase 2, mitochondrial	PPA2	1.517691785	2.637054443	-
Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	2.11972959	2.558508555	-
Protein phosphatase PTC7 homolog	PPTC7	2.367260172	2.429484685	+
Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	3.285908114	2.380776087	-
NAD-dependent malic enzyme, mitochondrial	ME2	2.705599794	2.338485718	-
RNA-binding motif, single-stranded-interacting protein 1	RBMS1	4.040859241	2.317210515	-
39S ribosomal protein L53, mitochondrial	MRPL53	1.644229884	2.278296153	+
39S ribosomal protein L52, mitochondrial	MRPL52	1.684215422	2.268756866	+
39S ribosomal protein L33, mitochondrial	MRPL33	3.921010963	2.196620941	+
Dihydrolipoyl dehydrogenase, mitochondrial	DLD	3.568975709	2.178907394	-
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	3.274047919	2.133876801	-
Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	2.604936259	1.989514669	-
Probable E3 ubiquitin-protein ligase makorin-2	MKRN2	3.091680264	1.952359517	-
Putative transferase CAF17, mitochondrial	IBA57	2.28678109	1.927953084	+
5-3 exoribonuclease 2	XRN2	2.899070362	1.894339879	-
RNA-binding motif, single-stranded-interacting protein 2	RBMS2	2.731841921	1.76350975	+
Scaffold attachment factor B2	SAFB2	2.6444863	1.705375671	+
Heterogeneous nuclear ribonucleoprotein D0	HNRPD	2.88015151	1.705236435	-
Enoyl-CoA hydratase, mitochondrial	ECHS1	2.209923076	1.640617371	-
3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	3.046260469	1.62707456	+
Heterogeneous nuclear ribonucleoprotein L-like	HNRNPDL	1.787033256	1.5285333	-
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	2.346542802	1.46873792	-
Acetyltransferase component of pyruvate dehydrogenase complex; Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	DLAT	2.174352815	1.389623642	-
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	3.256026935	1.369992574	-
39S ribosomal protein L11, mitochondrial	MRPL11	2.657172419	1.355416616	-
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	2.908116296	1.348293304	-
10 kDa heat shock protein, mitochondrial	HSPE1	2.380939534	1.346793493	-
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	2.900621023	1.346787135	-
Heat shock protein 75 kDa, mitochondrial	TRAP1	3.443833745	1.285238902	-
Hsc70-interacting protein; Putative protein FAM10A4; Putative protein FAM10A5	ST13	2.343873193	1.213008881	-
Isoleucine--tRNA ligase, mitochondrial	IARS2	2.172471039	1.210509618	-
ATP synthase mitochondrial F1 complex assembly factor 2	ATPAF2	2.36997155	1.207989375	-
2-oxoglutarate dehydrogenase, mitochondrial	OGDH	3.257591742	1.207115173	-
ATP-dependent RNA helicase DHX36	DHX36	2.20148615	1.187526067	-
Presequence protease, mitochondrial	PITRM1	2.587880206	1.16511027	-

Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	DBT	2.845733961	1.144769669	-
39S ribosomal protein L54, mitochondrial	MRPL54	1.658340646	1.111061732	+
39S ribosomal protein L2, mitochondrial	MRPL2	2.507914069	1.111005147	+
39S ribosomal protein L18, mitochondrial	MRPL18	2.994042816	1.10972023	+
2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial	BCKDHA	1.593489521	1.109059652	+
39S ribosomal protein L44, mitochondrial	MRPL44	3.076966717	1.096445719	-
Lupus La protein	SSB	2.815146607	1.095890681	-
Mitochondrial ribonuclease P protein 1	TRMT10C	1.630214209	1.090220133	+
THUMP domain-containing protein 3	THUMPD3	1.871894752	1.081604004	+
Polyadenylate-binding protein;Polyadenylate-binding protein 4	PABPC4	2.757012728	1.064692815	-
Zinc finger protein ubi-d4	DPF2	1.477201208	1.047100703	+
39S ribosomal protein L39, mitochondrial	MRPL39	2.360246654	1.037794749	-

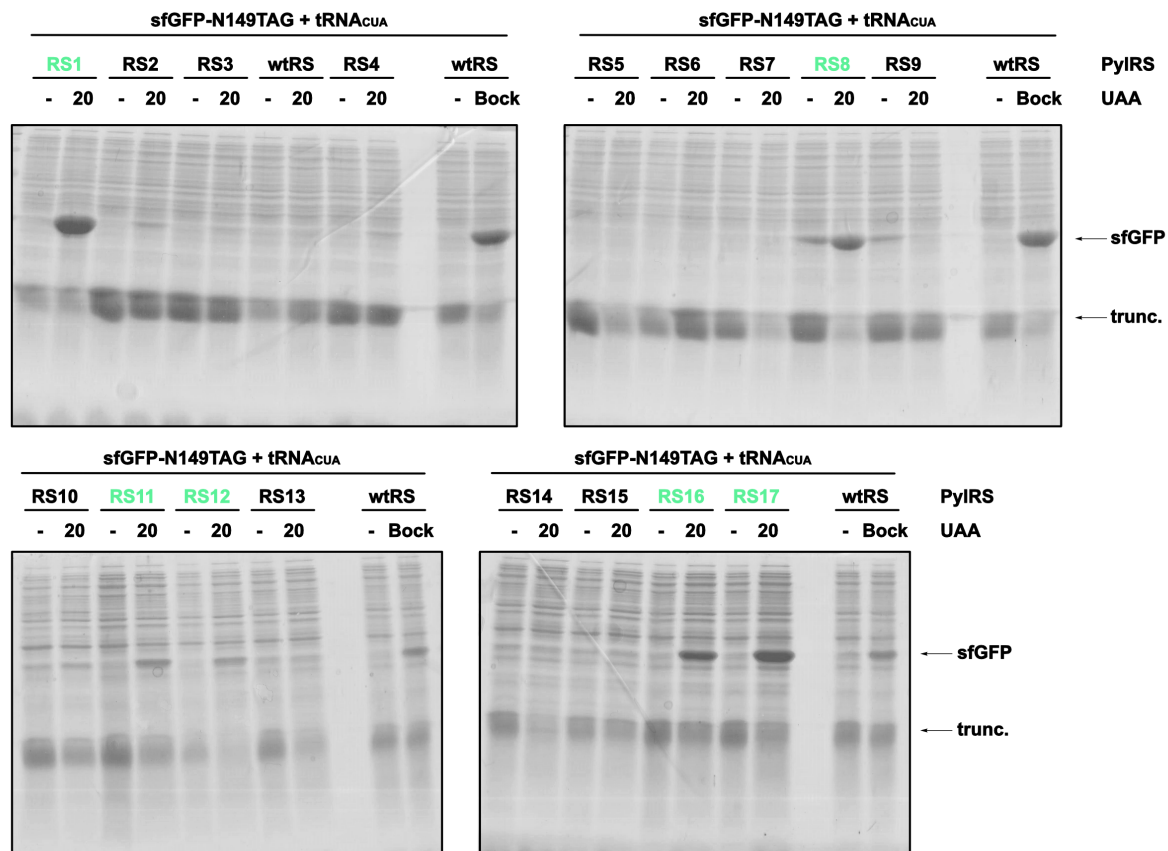
Table SI 2.8 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments of rotenone treated HEK293T with hClpP-D92dK. Proteins that were enriched only in rotenone treated samples and in no other mutants are marked with '+', if they are uniquely enriched and with '-', if they are not uniquely enriched.

Protein names	Gene name	enrichment factor (\log_2)	p-value t-test ($-\log_{10}$)	Annotated as substrate
Enoyl-CoA hydratase, mitochondrial	ECHS1	2.003698056	5.274079005	-
10 kDa heat shock protein, mitochondrial	HSPE1	2.416902285	3.559659958	-
Tyrosine--tRNA ligase, mitochondrial;Tyrosine--tRNA ligase	YARS2	2.676012595	3.065404256	-
Malate dehydrogenase;Malate dehydrogenase, mitochondrial	MDH2	2.538980576	2.600123723	-
Kappa-casein	CSN3	3.196187718	2.359725952	+
Single-stranded DNA-binding protein;Single-stranded DNA-binding protein, mitochondrial	SSBP1	3.129474044	2.081043879	-
Polyribonucleotide nucleotidyltransferase 1, mitochondrial	PNPT1	1.332356849	1.964795431	-
Isovaleryl-CoA dehydrogenase, mitochondrial	IVD	2.242487823	1.945773443	-
Cyclin-dependent kinase 4	CDK4	1.647243029	1.944344203	+
Presequence protease, mitochondrial	PITRM1	2.271304681	1.935722987	-
Inorganic pyrophosphatase 2, mitochondrial	PPA2	2.006109134	1.739587784	-
Tubulin beta chain	TUBB	1.715995809	1.616382599	-
Histidine triad nucleotide-binding protein 2, mitochondrial	HINT2	2.158128659	1.565341949	+
Citrate synthase;Citrate synthase, mitochondrial	CS	2.614555541	1.521754583	-
Putative transferase CAF17, mitochondrial	IBA57	1.419573126	1.488054911	+
Mitochondrial import receptor subunit TOM40 homolog	TOMM40	1.483859403	1.405598323	-
Acetyltransferase component of pyruvate dehydrogenase complex	DLAT	3.254700786	1.354489009	-
Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	HADH	2.839319267	1.331139247	-
Acyl-CoA dehydrogenase family member 9, mitochondrial	ACAD9	2.455665811	1.292857488	-
Mitochondrial ribonuclease P protein 1	TRMT10C	2.355294345	1.282247543	-
Aconitate hydratase, mitochondrial	ACO2	2.855714637	1.259790421	-
Heat shock protein 75 kDa, mitochondrial	TRAP1	3.172907231	1.242417653	-
Host cell factor 1	HCFC1	1.306021892	1.204651515	+

SRA stem-loop-interacting RNA-binding protein, mitochondrial	SLIRP	3.272002454	1.125771205	-
NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	NDUFV2	1.840864095	1.117319743	+
Glutamate dehydrogenase 1, mitochondrial	GLUD1	2.092659677	1.045743306	-
Poly [ADP-ribose] polymerase 1	PARP1	1.622107904	1.03410085	+
5-AMP-activated protein kinase catalytic subunit alpha-2;5-AMP-activated protein kinase catalytic subunit alpha-1	PRKAA1	2.443330979	1.022970835	+

Table SI 2.9 | Significantly enriched proteins (\log_2 ratio > 1, $-\log_{10}$ t-test p-value > 1.30) from trapping experiments of rotenone treated HEK293T with hClpP-K261dK. Proteins that were enriched only in rotenone treated samples and in no other mutants are marked with '+', if they are uniquely enriched and with '-', if they are not uniquely enriched.

8.2 Supporting information for chapter 3



PyIRS, Mb numbering

RS 1: Y271M, I274G, C313A
 RS 2: Y271V, L274M
 RS 3: M241F, A267S, Y271C, L274M
 RS 4: A267S, L274G, C313V
 RS 5: A267S, C313V, M315F
 RS 6: Y271A, L274M, C313A
 RS 7: Y271L, L274A, C313F
 RS 8: Y271G, C313V
 RS 9: L266M, L270I, Y271F, L274A, C313F

RS 10: C313V
 RS 11: Y271M, L274G, C313A, Y349W
 RS 12: Y271M, L274A, C313A
 RS 13: N311M, C313Q, V366G, W382N
 RS 14: N311Q, C313S, V366G, W382N
 RS 15: N311Q, C313A, V366G, W382N
 RS 16: Y271A, Y349F
 RS 17: Y271A, C313A, Y349F

Fig. SI 3.1 | aaRS screen for the selective incorporation of UAA 20 into sfGFP149TAG. UAA 20 was used in 2 mM concentration of screening. As judged by SDS-PAGE analyses, selective incorporations were achieved with aaRSs and their sequences as highlighted in green. Incorporation of BockK and wtRS was used as a reference for amber suppression of sfGFP. Transformation, expression, induction and SDS-PAGE analysis were performed as described in chapter 5.3.2.2.

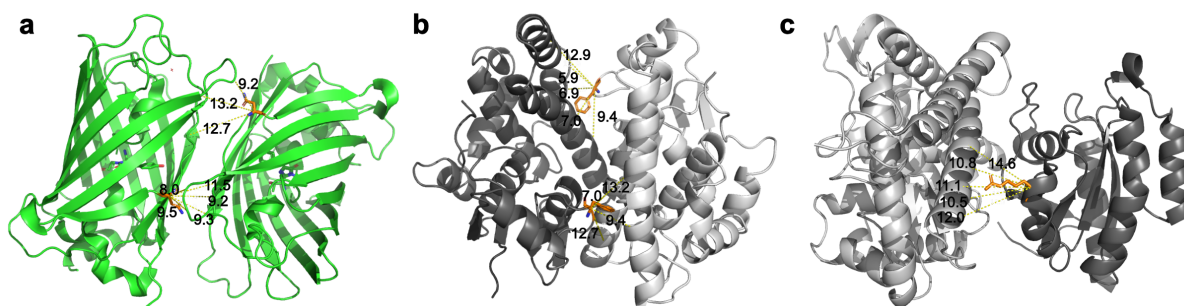


Fig. SI 3.2 I Interaction interface of PPIs investigated in this chapter. Depicted are the dimerization interfaces of **(a)** GFP (PDB: 5NHN) and **(b)** GST (PDB: 1GSD) as well as **(c)** the Rab/DrrA interaction (PDB: 6XY5). The amino acids, at which the different UAAs are incorporated at, are highlighted in orange. The various depicted distances (in Å) from the C α -atom of the orange residue to the opposing interactor sites were calculated with Pymol v.2.4.1.

8.3 Supporting information for chapter 4

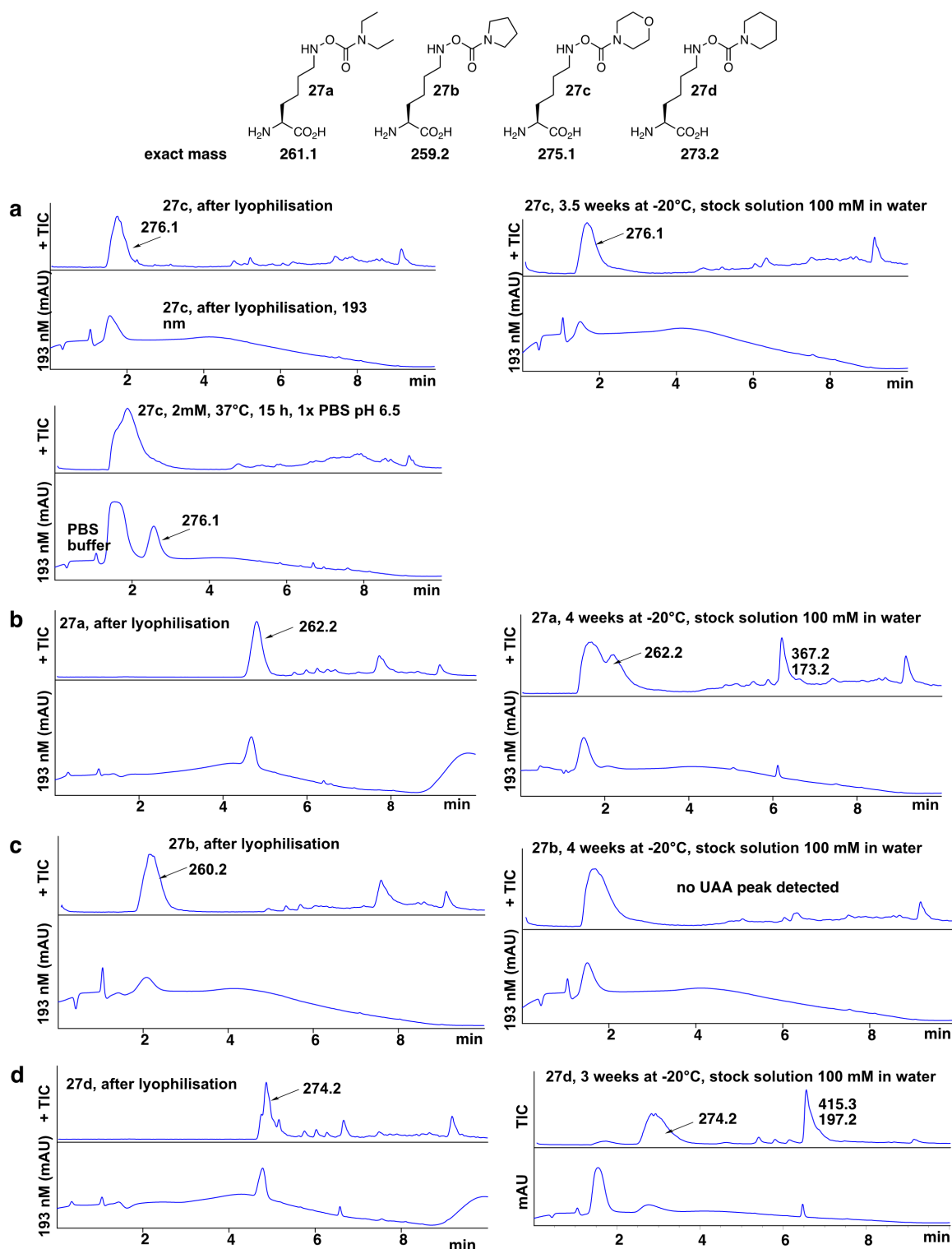


Fig. SI 4.1 | Stability measurements of UAA 27a-d. UAAs 27a-d were subjected to LC-MS analysis at different time-points (after lyophilisation and 3-4 weeks after storage in its stock solution at -20°C). Depicted are the positive TIC (+ TIC) and 193 nm traces. **(a)** UAA **27c** appeared stable in its stock solution over 3.5 weeks (stored at -20°) and at 37°C for 15 h. The remaining UAAs **(b-d)** showed additional peaks as indicated in the TIC traces indicating or no sign of UAA in case of 27b when stored for prolonged time at -20°C, indicating decomposition already during storage. The UAAs were thus not considered further.

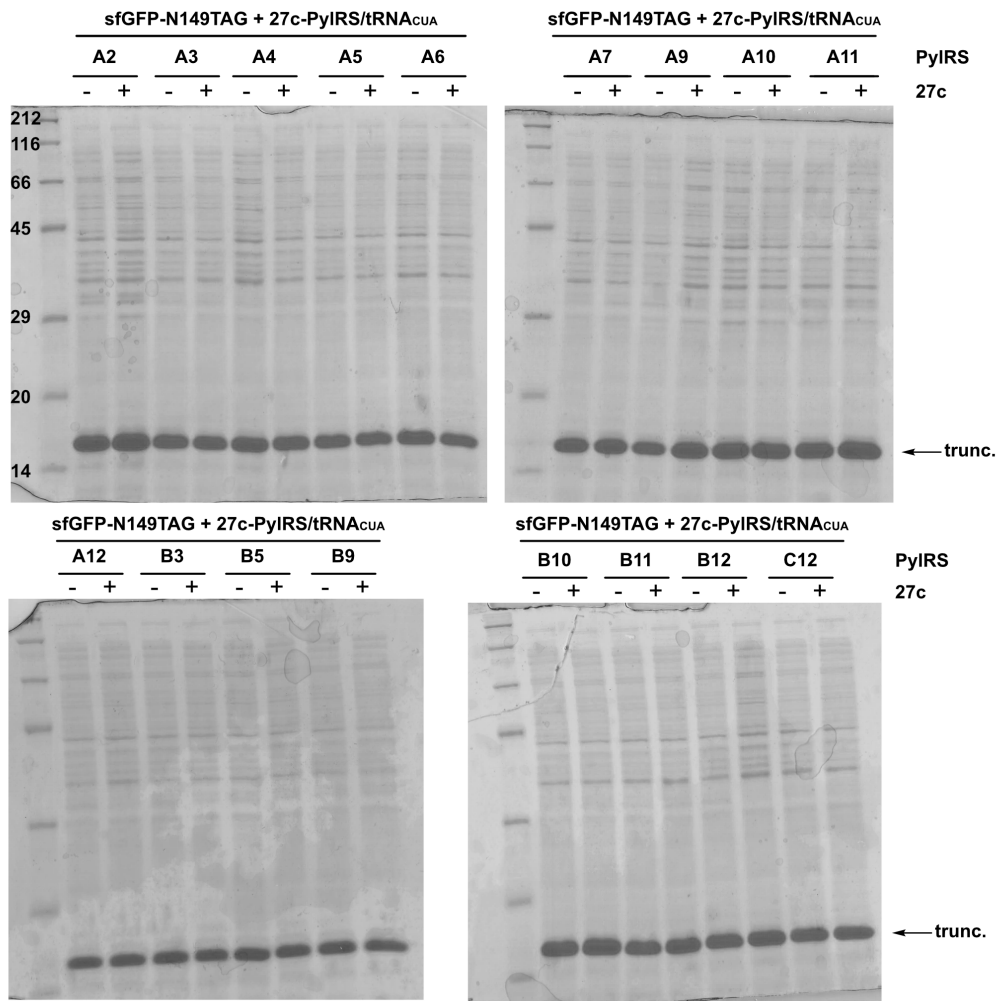
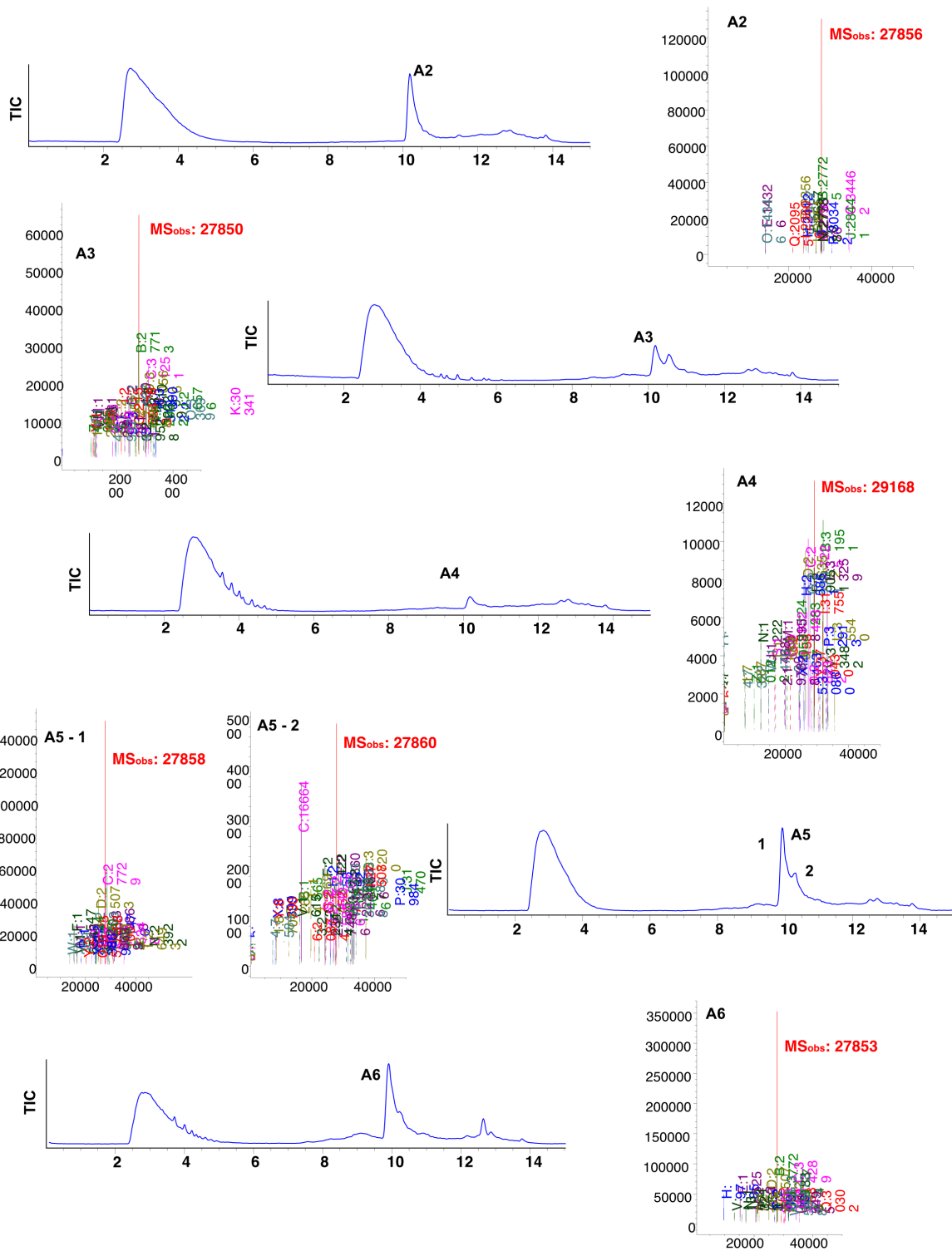
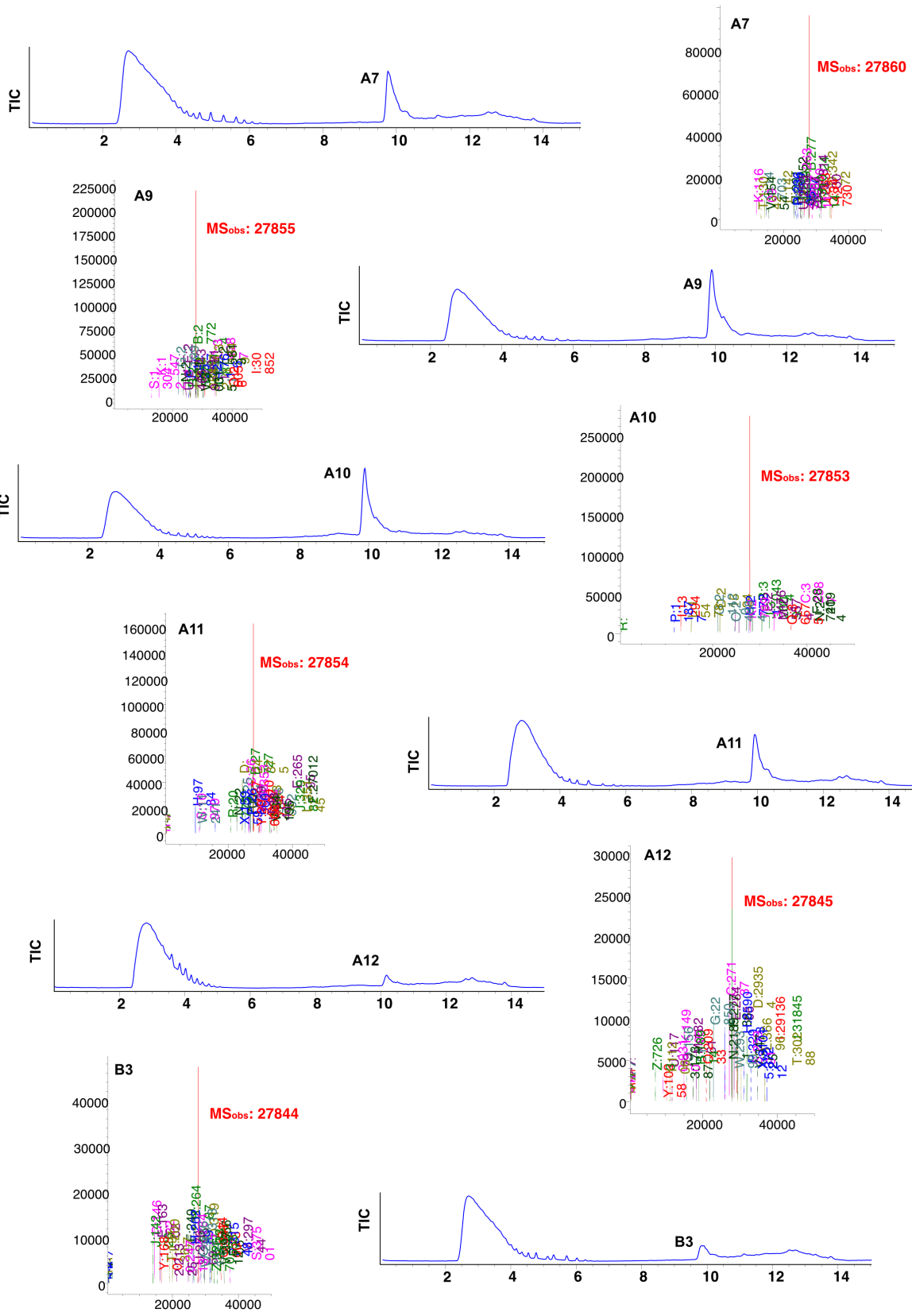


Fig. SI 4.2 | Incorporation test of evolved PyIRS after selection with UAA 27c and library AB1. Isolated PyIRSs were cotransformed with pPyIt sfGFP-149TAG and UAA 27c (1.5 mM) and expressed in AI media, o.n. at 37°C. Samples were taken and subjected to SDS-PAGE analysis. No sign of sfGFP expressions were observed in the SDS-PAGE.





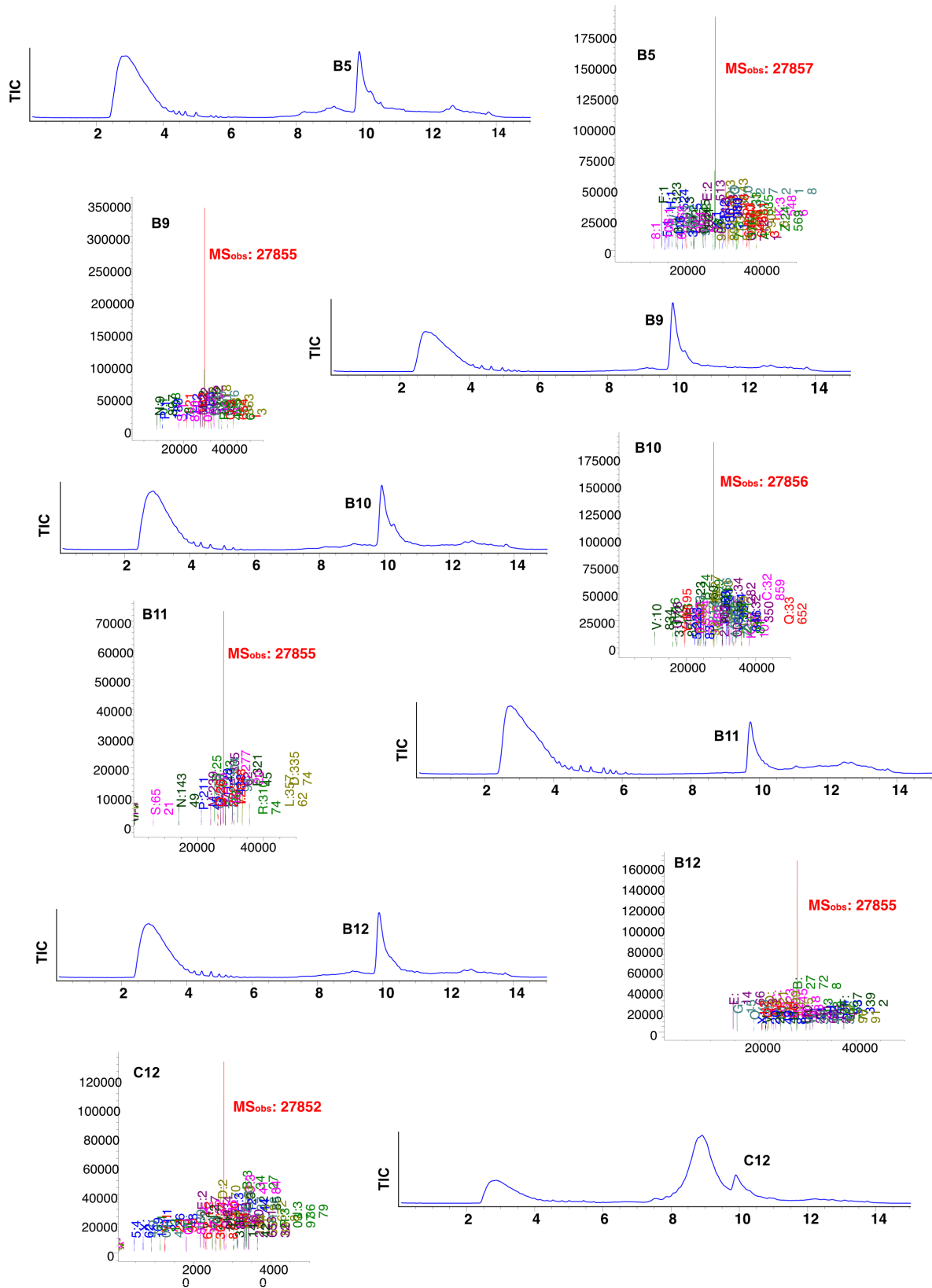


Fig. SI 4.3 | Expressed sfGFP masses with various PylRS mutants from PylRS selection with UAA 27c. TIC and deconvoluted MS of sfGFP masses listed in Table 4.1. The expected MS values of 27970 or 27839 (-Met) corresponding to sfGFP bearing UAA 27c instead of N149 were not observed in any cases. The observed MS values would correspond to Phe ($MS_{calc}: 27859$) or hydroxylysine incorporation ($MS_{calc}: 27857$).

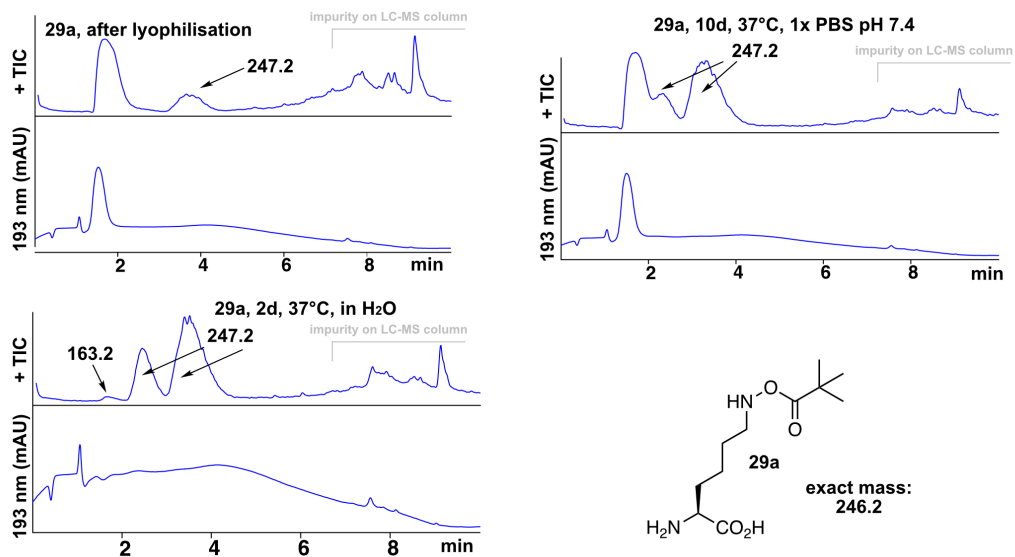


Fig. SI 4.4 | Stability profile of UAA 29a in PBS buffer and H₂O over indicated time points. UAA 29a appeared stable over prolonged time at 37°C, when dissolved in 1x PBS. Degradation of UAA 29a was observed after 2 d, when UAA was incubated in H₂O at 37°C. The observed MS value of 163.2 may correspond free hydroxylamine lysine. The Nu pathway may be a possible pathway in sole water, despite the putative E-pathway occurring in cellular environment.

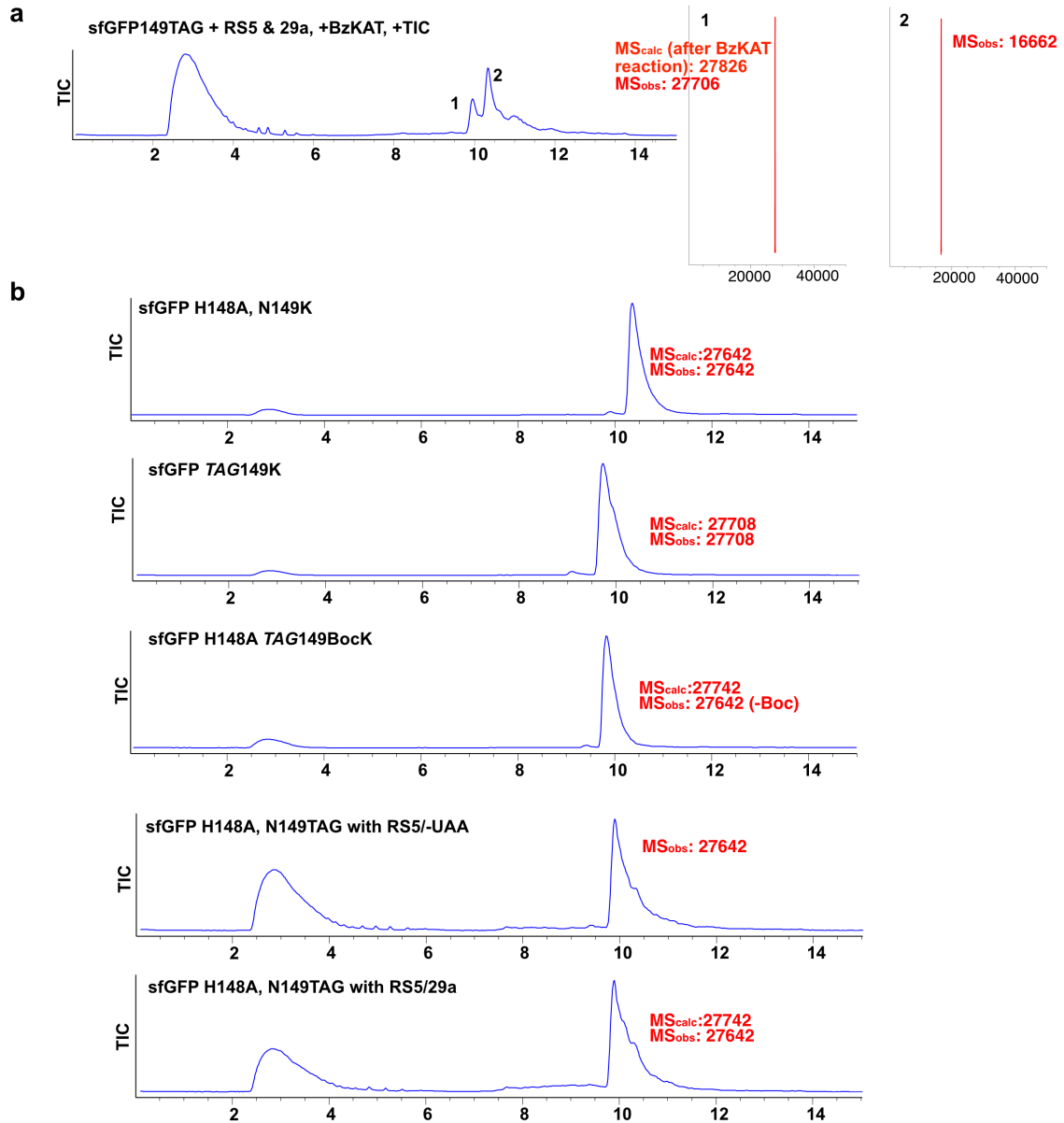


Fig. SI 4.5 | Investigation of sfGFP-N149TAG expressions with UAA 29a and PylRS5 isolated from the selection. (a) To investigate whether 29a was incorporated at position 149 of sfGFP, 2mM BzKAT was added to the lysate during NiNTA incubation. However, the respective MS value corresponding to BzKAT modified sfGFP was not detected, highlighting lysine misincorporation instead. **(b)** To investigate the effect of H148/K149 in the observed sfGFP (2-148) truncation, several mutants were created and expressed. Neither the preceding H148 nor the H148A mutation of N149(TAG) did not lead to autocatalytic cleavage of sfGFP. Position N149TAG in combination with H148A may be fully suppressed as indicated with the successful incorporation of BocK with wtRS, while sfGFP production with RS5 was independent of presence of UAA 29a. The observed mass fits to lysine incorporation at position 149, indicating Pyl-RS5 to be an UAA unspecific PylRS primed for Lys incorporation.

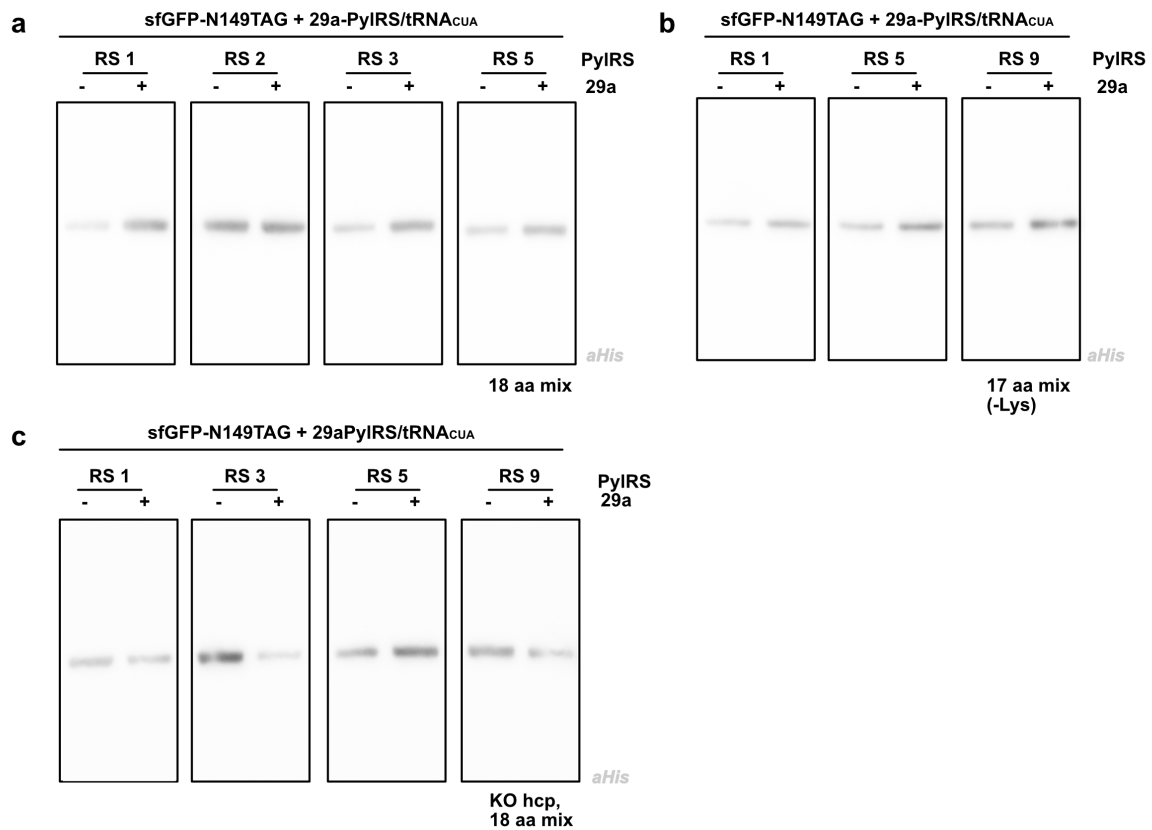
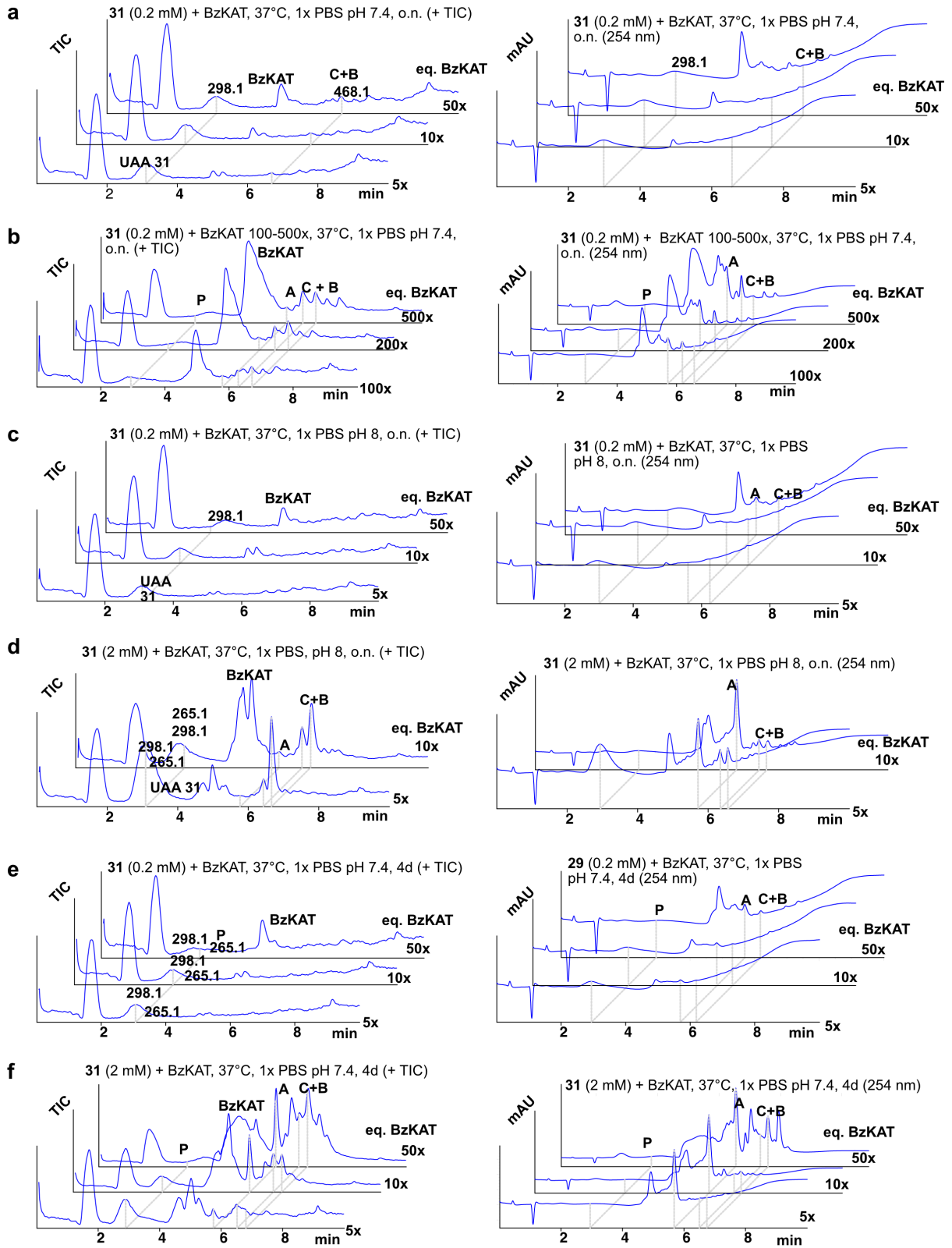


Fig. SI 4.6 I Reevaluation of sfGFP-N149TAG expression with UAA 29a and PyIRS5 from the selection procedure. **(a)** The selectivity of the evolved aaRS for 29a was reevaluated with a newly synthesised batch of UAA. Misincorporation of UAA in sfGFP-N149TAG in absence of UAA 29a was observed (in AI medium), as indicated by WB. **(b)** PyIRS selectivity could not be rescued when lysine was omitted from the AI media or **(c)** when sfGFP were expressed in a KO hcp (hydroxylamine reductase) strain. This overall indicates that the selected PyIRs misincorporates lysine instead of UAA 29a.



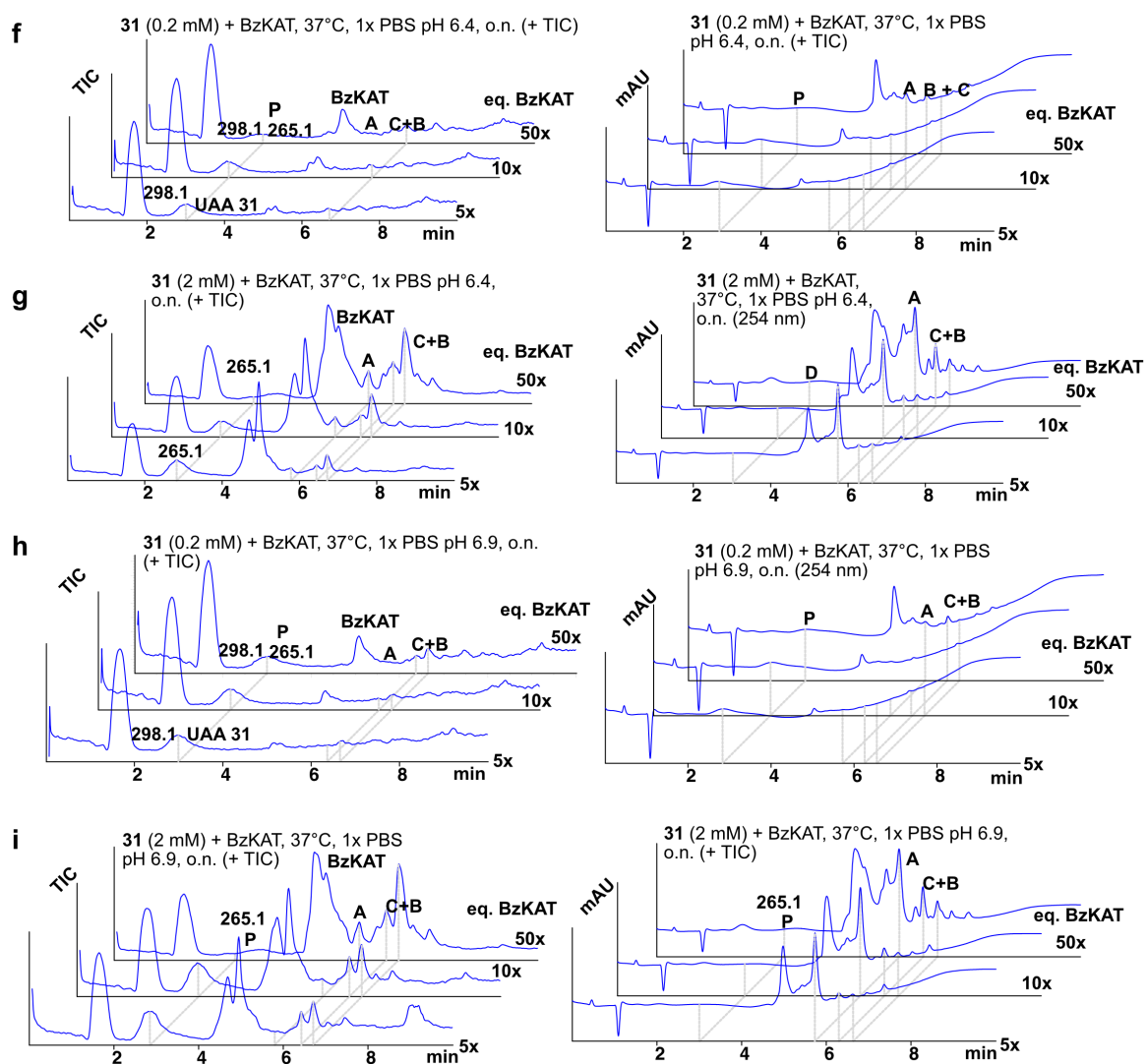


Fig. SI 4.7 | KAT reactivity profile of UAA 31. UAA 31 was incubated with BzKAT at various concentrations at 37°C, o.n. and conditions as specified in the chromatograms above (a-i). Reactions were monitored by LC-MS. TIC as well as 193 nm chromatogram of reach reaction are depicted. In any reaction conditions, various intermediates as described in the main section were detected. However full conversion of starting material 31 to product P was never observed with any of the reaction conditions.

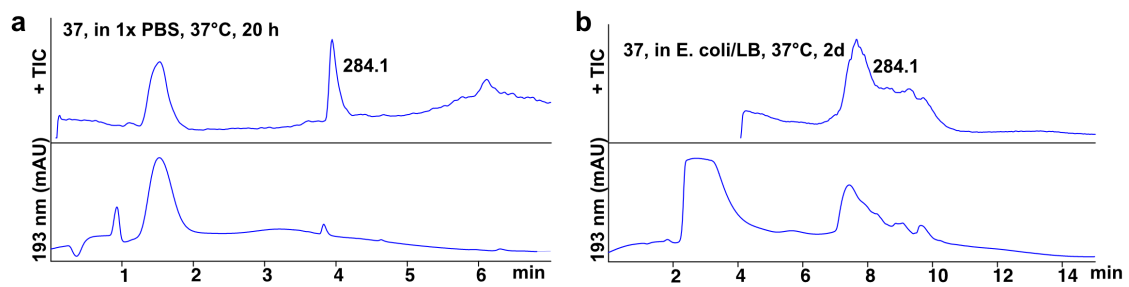


Fig. SI 4.8 | Stability profile of UAA 37. UAA 37 (2 mM) was incubated in (a) PBS and (b) E. coli in LB over indicated time-points and analysed by LC-MS for indication of degradation. UAA 37 was still detected and appeared stable during those conditions.

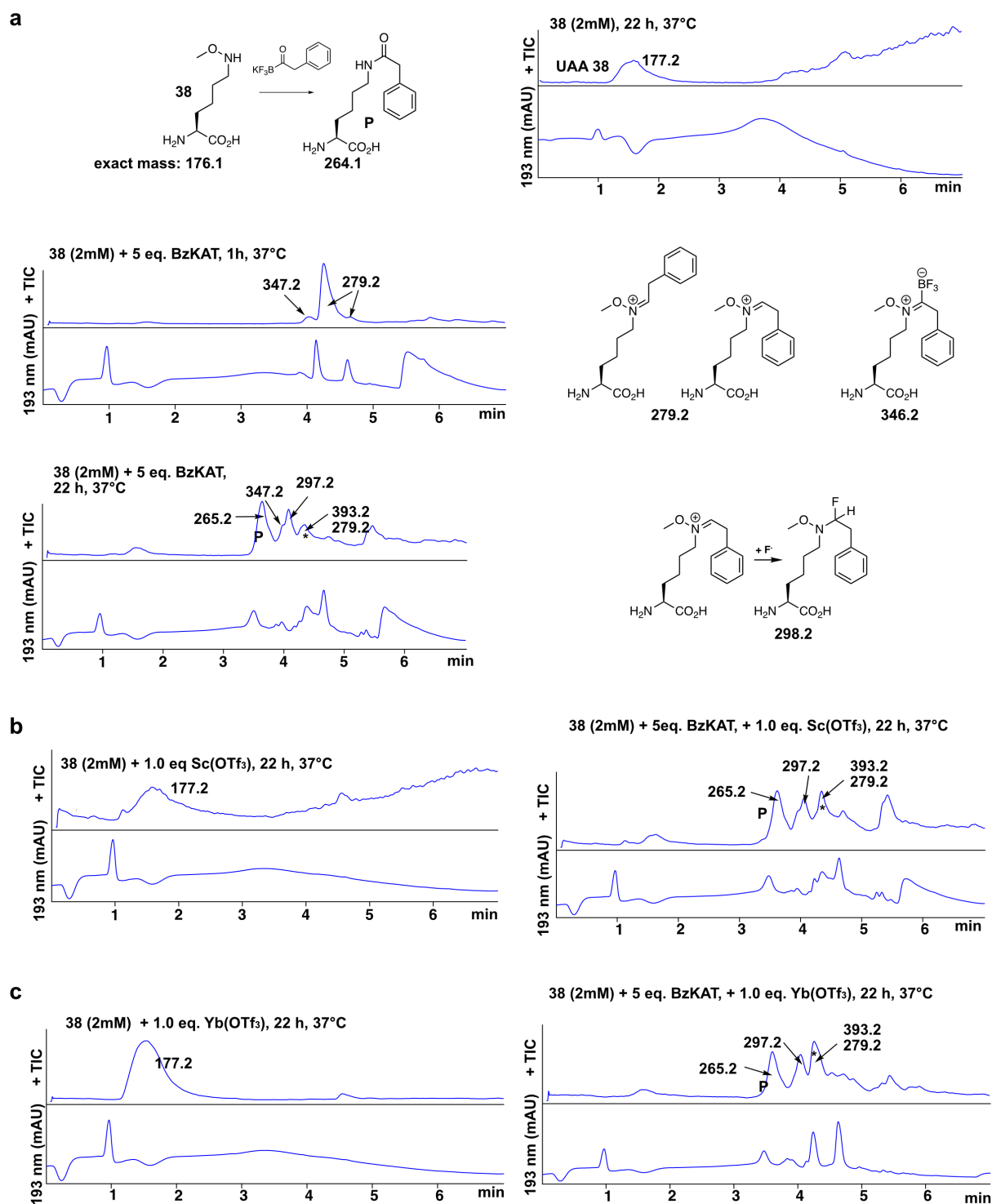


Fig. SI 4.9 | KAT reactivity profile of UAA 38 with Sc(OTf)₃ and Yb(OTf)₃. Depicted are the reaction profiles (positive TIC and 193 nm UV trace) of UAA 38 with or without Lewis acids as indicated in the figure caption. The complex peak pattern and analysis thereof was difficult and could only be partially attributed to the KAT intermediates. **(a)** The KAT reaction of UAA 38 with BzKAT optimally leads to conversion to the amide product P as described in the previous subchapters. UAA 38 appears stable in H₂O and 37°C over the monitored reaction time. Incubation with 5.0 eq. BzKAT already results in intermediate trapping after one hour as previously observed and proposed in ref.³⁰³ Further incubation of the reaction (in total 22 h) resulted into a more complex mixture. While product P (observed mass: 265.2) evolved, competitive pathways resulted into alternative products. The iminium intermediate, which already appeared after one hour (observed mass: 279.2) harbours a highly electrophilic carbon centre, which may be attacked by a fluorine (from BF₃) over time resulting into the monofluorinated hydroxylamine derivative (calc. mass: 298.2). The observed mass of 297.2 may result from the elimination of the highly acidic proton during ESI-MS. The other observed mass value of 393.3 (*) could not be identified. Fragmentation

pattern with an observed mass of 279.2 however suggests, that this species may be another modified product of the iminium intermediate. Overall, due to appearance of the side-products, KAT ligation with alkyl hydroxylamines may never reach quantitative values. Addition of $\text{Sc}(\text{OTf})_3$ and $\text{Yb}(\text{OTf})_3$ as lewis acids (**b** and **c**) did not result in improved reaction performance. In both cases, similar intermediates as observed in (**a**) were also detected.

9 Eidesstattliche Erklärung

Hiermit erkläre ich eidesstattlich, dass ich die vorliegende Arbeit selbständig 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

So, das Puzzle ist nun fertig und ich denke, es sieht am Ende doch ganz okay aus.