Technische Universität München

TUM School of Natural Sciences

Towards Optical Control of Protein Function Using an Expanded Genetic Code

Increasing the Repertoire of Photoswitchable Azobenzene-Based Non-Canonical Amino Acids

Kristina Maria Marit Krauskopf

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:	Prof. Dr. Hubert A. Gasteiger
Prüfer*innen:	1. Prof. Dr. Kathrin Lang 2. Prof. Dr. Stephan A. Sieber

Die Disseration wurde am 07.08.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 04.09.2023 angenommen.

List of Publications

Parts of the work presented within this thesis, have been published in scientific journals:

K. Krauskopf, K. Lang, Increasing the chemical space of proteins in living cells via genetic code expansion. *Curr Opin Chem Bio* **58**, 112-120 (2020), doi: 10.1016/j.cbpa.2020.07.012.

Further publications, which are not part of this thesis are:

PRA Zanon, F. Yu, P. Musacchio, L. Lewald, M. Zollo, **K. Krauskopf**, et al. Profiling the proteome-wide selectivity of diverse electrophiles. *ChemRxiv. Cambridge: Cambridge Open Engage* (2021), doi: 10.26434/chemrxiv-2021-w7rss-v2. This content is a preprint and has not been peer-reviewed.

Parts of the work presented within this thesis, have been presented during the following scientific conferences:

K. Krauskopf, M. v. Wrisberg, K. Lang, Photoswitchable Unnatural Amino Acids – Towards Conformational Control of *Sa*ClpP. 26th Lecture Conference on Photochemistry (2018), Munich, Germany.

K. Krauskopf, K. Lang, Photoswitchable Unnatural Amino Acids – Towards Conformational Control of *Sa*ClpP. SFB1035 Conference on Conformational Transitions in Proteins (2018), Venice, Italy.

Danksagung

Die letzten zwei Jahre meines PhDs waren wohl mit die langatmigsten überhaupt. Und dennoch sind an meinem letzten Tag im KLangLab Tränen geflossen. Und das ist vor allem den unglaublich tollen Menschen zu verdanken, mit denen ich diese intensiven sechs Jahre verbringen durfte.

Zunächst möchte ich Kathrin für die Gelegenheit danken, meinen PhD in ihrem Labor gemacht haben zu dürfen. Mein Hauptprojekt begleitet uns beide nun bereits seit einigen Jahren, verbunden mit manchen Tiefschlägen, Diskussionen und wieder aufflammender Hoffnung. Auch wenn der Ausgang kaum unserer beider Erwartungen entspricht, will ich dir für deine Unterstützung in persönlichen Krisen und konstantes Vertrauen in mich danken. Manch biochemisches Experiment wäre anfangs vermutlich schneller verlaufen, hättest du mich nicht selbst die Biochemie erkunden lassen. Ich bin bei dir fachlich und persönlich enorm gewachsen und würde es mir nicht anders wünschen.

Und der Rest des KLangLabs, was soll ich sagen; wenn ich euch und die Zeit beschreiben müsste, würden ein paar Worte nicht ausreichen. Es war ein Erlebnis. Es war einzigartig, chaotisch, Halt gebend, verrückt, lustig, schräg, lehrreich und emotional. Egal ob mit Umarmungen, Tränen, Schreikreisen oder Freudentänzen. Ihr habt manch schwere, lange Tage erträglich gemacht. Da durfte auch das ein oder andere Bierchen nicht fehlen.

Meine Chemistry Crew, Toni, Pascal und Simon. Nicht nur seid ihr mir mit Rat und Tat bei chemischen Fragestellungen zur Seite gestanden, ihr habt auch fleißig die chemischen Teile meiner Dissertation gelesen. Danke euch dafür.

Toni, es tut mir wirklich, wirklich leid. Es war nicht meine Absicht, vor dir mit meinem PhD fertig zu werden… oh wait, war es doch! Sorry, Toni oder Moni, aber bald hast du es dann auch geschafft. Seit der Ablöse von Susanne und Marko waren wir die nächste Generation Chemiker und ich muss sagen, es hat extrem viel Spaß gemacht mit dir. Sogar der Chemikalien-Umzug nach Zürich, als auch das wiederholt gemeinsame Rätseln und Verzweifeln vor der LC-MS. Danke auch für das Teilhaben lassen an Kreuzworträtseln und das gemeinsame Schmettern von Disney und Anime Songs; und nicht zu vergessen, stop it now!

Pascal, Pascolo, mein Music- und Dance-buddy, der mit mir die Achterbahnfahrten des Lab-Alltags durchlebt hat, inklusive mancher failure. Danke, dass du mich auf meine alten Tage noch zur richtigen Chemikerin gemacht und zur Kristallisation ermuntert (herausgefordert?) hast. Mit dir zusammen zu Space-Taxi, Ghostbusters und What does the fox say? zu tanzen, war immer die richtige Maßnahme um unsere emotional damage zu verarbeiten oder einfach nur Spaß zu haben.

Simon, danke, dass ich immer zu dir kommen durfte um dir ein paar Snacks abzuluchsen, obwohl du die auch dringend für deine Gains benötigt hättest. Ich wünsch dir für deinen Weg alles Gute und danke dir für deine stete Hilfsbereitschaft und eine tolle Zeit im Lab.

Und nun, liebe Biochemiker, ich muss gestehen, Bio ist für mich nicht länger Abfall. Und das ist zum großen Teil euch zu verdanken. Danke für eure Hilfsbereitschaft, Geduld, Unterstützung und das Korrekturlesen der biochemischen Teile meiner Dissertation.

Allen vorweg, Dr Lord Max Fottner, oder Lord Dr Max Fottner?! Du bist seit sechs Jahren ein Teil vom KLangLab, ein Urgestein quasi und seit ich angefangen habe hast du mich in allen biochemischen Fragestellungen unterstützt, egal wie knapp deine Zeit bemessen war. Danke dafür und natürlich die gemeinsamen Bierchen, vor allem in meiner AugusTina Ära.

Und Vera, die Beschte, wie du zu Recht von Marko genannt wirst. Ich durfte so viel von dir lernen, vor allem im Bereich SPPS und Proteinaufreinigung. Es war mir wahrhaftig eine Freude und Ehre, mit dir so eng zusammen gearbeitet haben zu dürfen. Egal ob an Projekten oder zur Betreuung von Studenten. Egal ob erfolgreich, oder weniger erfolgreich. Diesbezüglich haben wir ja gelernt „It can happen to anyone of us“. Danke auch für einen Haufen Umarmungen, vor allem in den letzten Monaten. Forever, your dark sunshine.

Hannes, mein Synchron-buddy (sowas gibt es nur einmal im Leben), bei dir hat es sich einfach seit deinem Start im KLangLab angefühlt, als wärst du schon ewig bei uns (oder ging es nur mir so?). Allein dafür danke ich dir schon (ich weiß, das war einfach). Deine Kloniertips waren aber auch ok, seitdem gab es bei mir nur noch Ligation à la Hannes. Spaß beiseite, danke, dass ich bei dir Zellkultur-Praktikum machen durfte und für deine moralische Unterstützung im Endgame.

Tarun, thank you for your kindness and willingness to help, whenever I dropped by to ask for your experience in regards to amber suppression. Thank you also for your constant good mood and fun facial expressions that cheered me up.

Maria, mein Teaching-buddy, auch wenn dein PhD noch nicht so lange geht, so bist du doch schon viel länger Teil vom KLangLab. Ich konnte immer bei dir vorbeikommen für einen kurzen Austausch oder um deine Expertise bezüglich Expressionskonstrukten einzuholen. Danke dir auch für deine stets gelassene und fröhliche Art über die Jahre, die für eine tolle Atmosphäre im Lab und während den Übungen gesorgt hat.

Und zu guter Letzt Paul, danke dir für deine fröhliche Art und deine Hilfsbereitschaft bei den neuen Blottern (die versüßen einem das Western Blot Leben wahrhaftig). Es hat mich gefreut, dass wir zumindest noch eine kurze Zeit zusammen im KLangLab verbringen konnten.

Danke auch an Leyla und Barbara für eure stets freundliche Unterstützung. Ohne euch würde das KLangLab nur halb so rund laufen (oder gar nicht?).

Und nun zu den KLangLab Alumnis.

Wie seltsam es doch ist, dich, Marie, Lennox, Lenny oder Lenovo, zu den Alumnis zu zählen. Du hast zeitgleich mit mir den PhD gestartet und quasi auch beendet. Danke, für all deine Hilfe und moralische Unterstützung über die Jahre. Egal ob bezüglich Zellkultur, Up-Take Assays oder Korrekturlesen meiner Dissertation. Danke auch, dass du mich unabhängig von meinem Fitnesslevel zu sportlicher Bewegung mit dir ermuntert hast. Von Joggen, über Fahrradfahren bis Wandern, war das ein wertvoller Ausgleich zum Lab-Alltag. Und obwohl wir nie direkte Lab-buddys waren, haben wir es jetzt doch noch zusammen in eine WG geschafft.

Marko, danke dir für die einzigartigen Jahre back then in munich. Chemie, Drehschieberpumpen, Marvel, Rennbahn, Margaritas… die Erlebnisse sind zu zahlreich um ihnen gerecht zu werden. Danke für die geile Zeit und danke auch, dass du trotz deines Ausscheidens aus dem KLangLab bereit warst, einige Teile meiner Dissertation Korrektur zu lesen.

Anh, du hast mich als erste Person das Klonieren gelehrt und auch sonst konnte man selbst und die eigenen Projekte immer von deinen clever gestellten Fragen zehren. Danke dafür und für eine tolle Zeit in München mit unvergleichlicher Musik. You are Gold.

Danke auch Shubhendu für den wissenschaftlichen Input und die gute Zusammenarbeit, sowie Susanne und Marie für eure Hilfe und Unterstützung vor allem am Anfang von meinem PhD.

Ein besonderes Dankeschön geht an unsere zwei ehemaligen Masteranden Philipp aka Thor und Ilias. Philipp, es war richtig schön und lustig dich wieder im Lab in Zürich zu haben und mit dir zusammenzuarbeiten während deiner Master Thesis. Es hat mich besonders gefreut, dass dir Chemie am Ende sogar ein wenig Spaß gemacht hat. Und Ilias, du begleitest mich ja schon ein paar Jahre, wenn auch on und off seit deinem Synthi-Versuch bei mir in 2019. Damals noch für Azobenzene, bei der Bachelorarbeit und während dem HiWi drehte sich dann alles um elektrophile warheads. Danke für dein wiederholtes Vertrauen in mich als Betreuerin. Es war mir wirklich eine Freude mit dir arbeiten und wachsen zu dürfen und ich bin richtig stolz und gespannt, wo dein Weg dich noch hinführen wird.

Danke auch an Domi, Flo, Carlos, Maxl und Paul B. für eine tolle Atmosphäre und Zeit im Lab.

Eine Gruppe an Leuten, denen ich besonders danken möchte, sind meine ehemaligen Azubis und Studenten. Nicht nur hat es mir Spaß gemacht zu sehen, wie ihr lernt und euch weiterentwickelt, ich durfte selber auch unglaublich viel durch die Tätigkeit als Betreuerin wachsen. Danke dafür und eure konstante Mitarbeit an meinen Projekten. Im Speziellen, danke an Dennis, Max aka AzoM, Sara, Stigler aka Sebastian, Yassmine, Andrei und zu guter Letzt Tobi für eure Mitarbeit am Azobenzen-Projekt. Danke an Anna, Ilias (wiederholt) und Emilia für eure synthetischen Mühen rund um elektrophile warheads. Danke auch an unsere Azubis Tanja, Michi und Steffi für eure großartige Arbeit über die Jahre in München.

Ein großes Dankeschön an meine Kollaborationspartner der Sieber Gruppe, im speziellen an Thommy G und Laura, für eure Geduld und Hilfe bei Peptidase oder Protease Assays. Ohne euch hätten diese Experimente vermutlich wesentlich länger gedauert oder mehrfach wiederholt werden müssen. Danke an Prof. Dr. Stephan Sieber für die wissenschaftlichen Ratschläge und dass ich die letzten Tage meines PhDs im Sieber Lab verbringen durfte.

Tuan, du als einziger Buchner-Vertreter, danke für deinen Urea-Puffer und deine Tips vor ein paar Jahren (ja, vermutlich erinnerst du dich nicht mehr daran), als ich noch frisch und unerfahren in der Biochemie war. Danke auch für das Korrekturlesen, deine stets gute Laune und das viele Essen.

Ein abschließendes Dankeschön an alle weiteren TUM und ETH Angehörigen, die diese Jahre zu einem einzigartigen Erlebnis gemacht haben.

Zu guter Letzt möchte ich mich bei den Menschen bedanken, die mich vielleicht nicht professionell durch diese Zeit begleitet haben, mich aber privat enorm unterstützt haben. Ohne euch wäre ich heute nicht da wo ich bin.

Domi und Nikki, danke für euer stets offenes Ohr über die Jahre und/oder in letzter Zeit.

Jörg, ich weiß gar nicht wo ich anfangen soll, deswegen halte ich es simpel. Danke, dass ich bei dir einfach sein kann.

Torsten, mein bester Freund und Bruder. Du hast mich durch diese Zeit getragen wie kaum jemand anderes. Danke für dein konstantes Interesse an meiner Arbeit, obwohl du wusstest, dass es oft schwer zu verstehen sein würde. Mama und Papa, danke dass ich euch immer anrufen konnte und meine Frustration bei euch abladen durfte. Danke euch dreien, für euer konstantes Vertrauen in mich, vor allem dann, wenn es mir selber schwer fiel weiter an mich zu glauben.

Gilles, du hast hauptsächlich die schwersten Jahre meines PhDs abbekommen. Und nicht einmal hast du mir das Gefühl gegeben, dass es zu viel für dich ist. Im Gegenteil, du bist mehrfach nach Zürich gefahren, hast unzählige späte Telefonate mit mir geführt und mich wissen lassen, dass ich nicht alleine bin. Danke für alles.

Table of Contents

Abstract ..... **ix**
Zusammenfassung ..... **xi**
Part A – Towards Optical Control of Protein Function Using an Expanded Genetic Code **1**
1 Introduction **2**
1.1 And Then There Was Light – Optogenetics 2
1.2 Photosensitive Small Molecules 4
1.2.1 Photocaging 4
1.2.2 Photoswitches 6
1.2.3 Azobenzene 8
1.3 The Advance of Azobenzenes in Biology 12
1.3.1 Optical Modulation of Nucleic Acid Morphology by Azobenzenes 12
1.3.2 Photoswitchable Lipids and Ligands in Chemical Biology 13
1.3.3 Azobenzenes to Study Peptide and Protein Conformation 21
1.4 An Expanded Genetic Code 24
1.4.1 Ribosomal Protein Translation 24
1.4.2 Amber Codon Suppression 25
1.4.3 Orthogonal Synthetase/Aminoacyl-tRNA Pairs 26
1.4.4 Directed Evolution of Novel Synthetases 28
1.4.5 Recent Efforts Within the Field 30
1.5 Genetically Encoded Non-Canonical Amino Acids 34
1.5.1 Optical Control of Protein Function Using Photocaged ncAAs 35
1.5.2 Reversible Optical Control of Protein Function Using Azo-ncAAs 36
2 Project Aim **38**
2.1 Optical Modulation of Biological Processes 38
2.2 Envisioned Azo-ncAAs for Proximity-Induced Crosslinking 39
3 Novel Azo-ncAAs for the Optical Control of Protein Function **41**
3.1 Development of Novel Azo-ncAAs Bearing Cysteine-Reactive Moieties 41
3.1.1 General Synthesis of Azobenezenes 41
3.1.2 Prior Work 42
3.1.3 Design of Azo-ncAAs From This Project 43
3.1.4 Development of Azo-ncAAs for S_N2-Based Crosslinking 46
3.1.5 Development of Azo-ncAAs for Crosslinking via S_N2 and Michael Addition 51
3.1.6 Development of Azo-ncAAs for Crosslinking via S_NAr Reaction 55
3.1.7 Development of Azo-ncAAs Bearing tetra-*ortho* Halogenations 59
3.1.8 Synthesis of Model Trp-based Azo-ncAA 60
3.1.9 A Tripeptide Bearing Azo-ncAAs for Amber Suppression 61
3.2 Photophysical Characterization 63
3.3 Reversible Optical Control of Protein Function 65
3.3.1 Reversible Optical Control of PPIs 65
3.3.2 Reversible Optical Control of Enzyme Activity 77
3.3.3 Reversible Optical Control over Protein Oligomerization 86
4 Summary of Results and Outlook **91**

Part B – Electrophilic Warheads for the Generation of Activity-Based Probes Using Sortylation 96
5 Introduction 97
5.1 The Ubiquitin System 97
5.2 Activity-Based Probes for Ub Deconjugating Enzymes 98
5.3 Sortylation – Sortase-Mediated Ligation 99
6 Project Aim 100
6.1 Generation of ABPs Using Sortylation 100
6.2 Envisioned Design of Warheads 100
7 Electrophilic Warheads for the Generation of ABPs via Sortylation 102
7.1 Prior Work 102
7.2 Design of Warheads and Probes From This Project 102
7.3 Synthesis Towards Moderately Electrophilic Warheads 103
7.4 Synthesis Towards Masked Electrophilic Warheads 107
7.5 Synthesis of Fluorogenic Probes 111
8 Outlook 112
Part C – Supplementary Information 114
9 Experimental Procedures 115
9.1 Chemistry – General Remarks 115
9.2 UV-Vis Spectrophotometry for Photophysical Characterization 117
9.3 Synthetic Standard Procedures 118
9.4 Synthetic Procedures of azo-ncAAs 121
9.4.1 Synthesis of Precursor Boc-Phe(4-NH_2)-OH (12) 121
9.4.2 Synthesis of pnX Series for S_N2 122
9.4.3 Synthesis of mnX Series for S_N2 128
9.4.4 Synthesis of pQX Series for Activated S_N2 135
9.4.5 Synthesis of mQX/mQR Series for S_N2 and Michael Addition 140
9.4.6 Synthesis of pXN Derivatives for S_NAr 144
9.4.7 Synthesis of tetra-*ortho* Halogenated Azo-ncAAs 146
9.4.8 Synthesis of Trp-Based Azo-ncAA 148
9.4.9 Synthesis of Tripeptide KGpFN 150
9.4.10 Synthesis of Model Azo-ncAA pSEtN 153
9.4.11 Synthesis of Non-Crosslinking Azo-ncAA m1OH 155
9.5 Synthetic Procedures of Electrophilic Warheads and Fluorogenic Probes 156
9.5.1 Synthesis of Moderately Electrophilic Warhead G-VME 156
9.5.2 Synthesis of Moderately Electrophilic Warhead G-VCN 158
9.5.3 Synthesis Towards Moderately Electrophilic Warhead G-VtBuE 159
9.5.4 Synthesis of Moderately Electrophilic Warhead G-V(D)MA 161
9.5.5 Synthesis of Lysine-Based ncAAs (G-)VAK 164
9.5.6 Synthesis of Masked Electrophilic Warhead G-MVK 167
9.5.7 Synthesis Towards Masked Electrophilic Warhead G-ClMK 172
9.5.8 Synthesis of Fluorogenic Probes G-AMC and GG-AMC 176
9.6 Solid Phase Peptide Synthesis 178

9.7 Biology – General Remarks 179
9.7.1 Assay, Protein Purification and Storage Buffers 179
9.7.2 Buffers and Media 180
9.7.3 *E. coli* strains 181
9.7.4 Oligonucleotides 182
9.7.5 Plasmids 182
9.7.6 Stock Solutions 185
9.8 Microbiology Procedures 186
9.8.1 Preparation of Chemically Competent *E. coli* Cells 186
9.8.2 Heat-Shock Transformation of Chemically Competent *E. coli* Cells 186
9.9 Molecular Biology Procedures 186
9.9.1 Polymerase Chain Reaction (PCR) 186
9.9.2 Molecular Cloning 187
9.9.3 DNA Isolation and Purification 187
9.9.4 Agarose Gel Electrophoresis 188
9.9.5 DNA Sequencing 188
9.10 Protein Biochemistry Procedures 188
9.10.1 Standard Protein Expression and Purification of sfGFP-N149TAG-His_6 188
9.10.2 Expression and Purification of Calmodulin-His_6 Variants 189
9.10.3 Expression and Purification of Affibody-His_6 Variants 189
9.10.4 Expression and Purification of GST-Z-Domain 190
9.10.5 Expression and Purification of His_6-Throm-Rap80-TEV-StrepII Variants 190
9.10.6 Expression and Purification of anti-GFP Nanobody-His_6 Variants 191
9.10.7 Expression of AltTPase-StrepII Variants 191
9.10.8 Expression and Purification of PDZ3-His_6 Variants and His_6-TEV-PDZ3 191
9.10.9 Expression and Purification of MID1sc10-His_6, MID1sc10-TEV-His_6, His_6-(MBP)-TEV-MID1sc10 and His_6-MBP-TEV-MID1sc10 Variants 192
9.10.10 Expression and Purification of MID1sc10-CPD-His_6 Variants 192
9.10.11 Expression of OP-His_6 Variants 193
9.10.12 Expression and Purification of ClpP-StrepII Variants 193
9.11 Protein Characterization Methods 194
9.11.1 SDS-PAGE 194
9.11.2 Western Blot Analysis 194
9.11.3 Protein-MS 195
9.11.4 Circular Dichroism Spectroscopy 195
9.12 Protein Interaction and Activity Assays 195
9.12.1 Pull-Down Assay 195
9.12.2 Fluorogenic Peptidase Assay 196
9.12.3 Intrinsic Tryptophan Fluorescence Quenching Assay 196
I Literature 197
II List of Abbreviations 214
III Supplementary Figures 217
IV Supplementary Tables 222
V Eidesstattliche Erklärung 226

Abstract

Part A – Towards Optical Control of Protein Function Using an Expanded Genetic Code

The development of tools to probe and manipulate biological processes in living cells with high specificity is an important endeavour in chemical biology. Orthogonality, i.e. studying an individual biological process without affecting the corresponding endogenous processes, is a key requirement and often challenging to implement. Light, however, can be an ideal orthogonal trigger, as it is non-invasive in many settings from cellular to *in vivo* biology and thus represents an elegant approach to gain control over physiological processes with high spatial and temporal resolution. Consequently, a range of photo-activatable small molecules have been developed for the optical modulation of biological events. Whereas light-cleavable protecting groups, so-called "photocages", are irreversibly triggered to unmask biological functions, bistable photoswitches have the necessary properties to study complex biological processes in a reversible manner.

Azobenzenes are the most commonly used reversible photoswitches, since they meet many requirements for use in modulating biological processes *in vivo*, such as a well-studied photoisomerization process at wavelengths above 350 nm and an exceptionally high photostability over many cycles. Over the last decade, azobenzene derivatives have been widely employed as photoswitchable ligands to study biological processes. In the field of photopharmacology, they are typically employed as freely diffusible or tethered ligands, allowing for the regulation of membrane receptor dynamics in live neurons. Complementary to this, the function of proteins can also be modulated by the installation of photoswitchable crosslinkers or amino acids into secondary structural elements as a conformational switch. The introduction of reversible photoswitches directly into proteins in such a way, that site-specifically and homogenously modified azobenzene-protein conjugates are obtained, remains challenging.

The incorporation of azobenzene-based non-canonical amino acids (azo-ncAAs) into proteins via genetic code expansion represents a sophisticated way to tackle these challenges. Genetic code expansion uses orthogonal aminoacyl-tRNA synthetase/tRNA pairs to direct the site-specific incorporation of a non-canonical amino acid, typically in response to an amber stop codon (amber codon suppression).

The goal of this work was to develop novel azo-ncAAs as tools for a genetic code expansion approach, to reversibly modulate α-helix and thus protein conformation with light. These azo-ncAAs were designed to undergo a proximity-induced reaction with a proximally spaced cysteine, furnishing an azobenzene-based crosslink that would drive photoisomerization-based conformational switching of α-helix motifs. In this work, we synthesized a range of azo-ncAAs for different ligation reaction types, including S_N2, S_NAr and Michael addition. A screening of aminoacyl-tRNA synthetase variants identified several mutants which readily incorporated three novel azo-ncAAs into proteins in *E. coli*. One of these novel azo-ncAAs was further shown to give a full crosslink with a proximal cysteine (*i*, *i*+7) when incorporated into the α-helix of different target proteins.

This photoswitchable tool was tested in photophysical studies as well as proof-of-principle experiments to evaluate its potential for switching protein-protein interactions and controlling enzyme activity *in vitro*. This work primarily highlights the potential and limitations of azo-ncAAs as photoswitchable tools in genetic code expansion, while discussing alternative strategies to gain optical control over protein conformation and thus function.

Part B – Electrophilic Warheads for the Generation of Activity-Based Probes Using Sortylation

The reversible, post-translational modification of proteins with the highly conserved 76 amino acid protein ubiquitin (Ub), also termed ubiquitylation, serves as key regulator in almost all eukaryotic cellular processes such as proteasomal degradation or endocytosis. Ub-dependent signaling relies on three key classes of enzymes (writers, readers and erasers) that are essential in maintaining the highly dynamic process of installing, interpreting and erasing the different Ub topologies. Deconjugation of Ub is performed by highly specific proteases, of which the vast majority are classified as cysteine-proteases, so-called deubiquitylases (DUBs). A dysregulation of DUBs is typically associated with tumor growth, autoimmune diseases and neurodegenerative disorders, making the elucidiation of their function and substrate specificity highly important.

Here, the application of activity-based probes (ABPs) has proven particularly effective. By mimicking the natural substrate of enzymes without being processed by them, the application of ABPs allows to covalently trap the target enzyme. To this end, electrophilic warheads are attached to the recognition motif, which typically is a Ub monomer in the case of ABPs targeting DUBs. A main drawback of DUB-activity profiling using ABPs is the limitation to *in vitro* applications due to permeability issues of Ub-ABPs. An elegant solution to overcome these issues, would be the generation of Ub-ABPs inside living cells using sortase-mediated ligation (sortylation) and thereby facilitate the elucidation of DUB activity in mammalian cells.

The goal of this work was the design and synthesis of electrophilic warheads with varying reactivity to trap the active site cysteine of DUBs either via Michael addition or nucleophilic displacement. Conjugation of the respective warheads to a glycine-moiety would further allow for subsequent transpeptidation to the C-terminus of Ub using sortase.

In this work, we synthesized electrophilic warheads known from literature, as well as novel, less reactive Michael acceptors based on α,β-unsaturated amides. Two lysine-based α,β-unsaturated amides may further serve as valuable tools in probing DUB activity via a genetic code expansion approach. Further, we successfully established a selective synthesis route towards highly reactive α,β-unsaturated ketones masked by a photolabile protecting group. Upon irradiation with UV-light the reactivity towards nucleophiles would be restored, thus enabling trapping of DUBs after successful sortylation.

Zusammenfassung

Teil A – Entwicklungen hin zur Optischen Kontrolle von Proteinfunktionen mit Hilfe eines erweitereten genetischen Codes

Die Entwicklung von Werkzeugen um biologische Prozesse in lebenden Zellen mit hoher Spezifität untersuchen und beeinflussen zu können, ist ein wichtiges Unterfangen in der chemischen Biologie. Dabei ist Orthogonalität, d.h. die Untersuchung eines einzelnen biologischen Prozesses, ohne die jeweiligen endogenen Prozesse zu beeinträchtigen, ein wichtiges und oft schwer zu realisierendes Ziel. Die Verwendung von Licht als externer Auslöser ist jedoch in vielen Bereichen, von der Zellbiologie bis hin zur *in-vivo*-Biologie, orthogonal und nicht invasiv. Licht bietet somit einen eleganten Ansatz zur Kontrolle physiologischer Prozesse mit hoher räumlicher und zeitlicher Auflösung. Infolgedessen wurde eine Reihe lichtempfindlicher kleiner Moleküle, die bei Bestrahlung eine Reaktion zeigen, für die optische Regulierung biologischer Vorgänge entwickelt. Während durch Licht abspaltbare Schutzgruppen, so genannte "Photocages", irreversibel ausgelöst werden, um biologische Funktionen einmalig freizulegen, verfügen bistabile Photoschalter über die notwendigen Eigenschaften, um komplexe biologische Prozesse reversibel zu untersuchen.

Azobenzole sind die am häufigsten verwendeten reversiblen Photoschalter, da sie viele Anforderungen für die *in-vivo*-Modulation biologischer Prozesse erfüllen, wie z.B. einen gut untersuchten Photoisomerisierungsprozess bei Wellenlängen über 350 nm sowie eine außergewöhnlich hohe Photostabilität über mehrere Zyklen. In den letzten zehn Jahren wurden Azobenzolderivate in großem Maße als photoschaltbare Liganden zur Untersuchung biologischer Prozesse eingesetzt. Im Bereich der Photopharmakologie werden sie in der Regel als frei diffundierende oder gebundene Liganden eingesetzt, welche die Regulierung von Membranrezeptordynamiken in lebenden Neuronen ermöglichen. Ergänzend dazu kann die Funktion von Proteinen auch durch den Einbau von photoschaltbaren „crosslinkern“ oder Aminosäuren in sekundäre Strukturelemente als Konformationsschalter moduliert werden. Die Einführung von reversiblen Photoschaltern direkt in Proteine, so dass ortsspezifisch und homogen modifizierte Azobenzol-Protein-Konjugate entstehen, bleibt eine Herausforderung.

Der Einbau von nicht-kanonischen Aminosäuren auf Azobenzol-Basis (azo-ncAAs) in Proteine mittels genetischer Code-Erweiterung stellt eine ausgeklügelte Möglichkeit dar, diese Herausforderungen zu bewältigen. Bei der Erweiterung des genetischen Codes werden orthogonale Aminoacyl-tRNA-Synthetase/tRNA-Paare verwendet, um den positionsspezifischen Einbau einer nicht-kanonischen Aminosäure zu steuern, in der Regel als Reaktion auf ein Amber Stopp-Codon (amber codon suppression).

Ziel dieser Arbeit war es, neuartige azo-ncAAs zu entwickeln, die es ermöglichen, α-Helix-Strukturelemente in gefalteten Proteinen und damit Proteinkonformation reversibel mit Licht zu modulieren. Diese azo-ncAAs wurden so konzipiert, dass sie eine Proximitäts-induzierte Reaktion mit einem in räumlicher Nähe befindlichen Cystein eingehen. Hierdurch entsteht ein Azobenzol-basierter „crosslink“, welcher einen auf Photoisomerisierung basierenden Konformationswechsel von α-Helix-Motiven auslösen würde.

In dieser Arbeit haben wir eine Reihe von Azo-ncAAs für verschiedene Ligationsreaktionstypen synthetisiert, darunter S_N2, S_NAr und die Michael-Addition. Bei einem Screening von Aminoacyl-tRNA-Synthetase-Varianten wurden mehrere Mutanten identifiziert, die drei neue azo-ncAAs problemlos in *E. coli*-Proteine einbauen konnten. Bei einer dieser neuartigen azo-ncAAs wurde außerdem gezeigt, dass es beim Einbau in die α-Helix verschiedener Zielproteine zu einem vollständigen „crosslink“ mit einem proximalen Cystein (*i*, *i*+7) kommt.

Daraufhin folgten photophysikalische Studien sowie *in*-vitro-Experimente, in denen die Auswirkungen des erfolgreich entwickelten, photoschaltbaren Werkzeugs auf Protein-Protein-Interaktionen und Enzymaktivität untersucht wurden. Diese Arbeit zeigt vorallem die Möglichkeiten und Grenzen von azo-ncAAs als photoschaltbare Werkzeuge für die Erweiterung des genetischen Codes auf und erörtert gleichzeitig alternative Strategien zur optischen Kontrolle der Konformation von Proteinen und damit deren Funktion.

Teil B – Elektrophile „warheads" für die Herstellung von aktivitätsbasierten Sonden durch Sortase-vermittelte Ligation

Die reversible, posttranslationale Modifikation von Proteinen mit dem hochkonservierten 76-Aminosäuren-Protein Ubiquitin (Ub), auch Ubiquitinierung genannt, spielt eine entscheidende Rolle als Regulator bei fast allen eukaryotischen Zellprozessen wie dem proteasomalen Abbau von Proteinen oder der Endozytose. Die Ub-abhängige Signalübertragung beruht auf drei wesentlichen Enzymen („writers", „readers" und „erasers"), die für die Aufrechterhaltung des hochdynamischen Prozesses der Installation, Interpretation und Entfernung verschiedener Ub-Topologien unerlässlich sind. Die Dekonjugation von Ub erfolgt durch hochspezifische Proteasen, so genannte Deubiquitinasen (DUBs), von denen die überwiegende Mehrheit als Cystein-Proteasen klassifiziert werden. Eine Dysregulation von DUBs wird typischerweise mit Tumorwachstum, Autoimmunkrankheiten und neurodegenerativen Störungen in Verbindung gebracht, weshalb die Aufklärung ihrer Funktion und Substratspezifität von großer Bedeutung ist.

Hier hat sich der Einsatz von aktivitätsbasierten Sonden (ABPs) als besonders effektiv erwiesen. Durch die Nachahmung des natürlichen Substrats von Enzymen, ohne von diesen jedoch verarbeitet zu werden, ermöglicht die Anwendung von ABPs das kovalente Einfangen des Zielenzyms. Zu diesem Zweck werden elektrophile „warheads" an das Erkennungsmotiv angehängt, das im Falle von ABPs, die auf DUBs abzielen, in der Regel der C-Terminus eines Ub-Monomers ist. Ein Hauptnachteil der derzeitigen Erstellung von DUB-Aktivitätsprofilen mit Hilfe von ABPs ist die Beschränkung auf *in-vitro*-Anwendungen aufgrund von Permeabilitätsproblemen der Ub-ABPs. Eine elegante Lösung zur Überwindung dieser Probleme wäre die Erzeugung von Ub-ABPs in lebenden Zellen durch Sortase-vermittelte Ligation (Sortylierung) und damit die Erforschung der DUB-Aktivität in intakten Säugetierzellen.

Ziel dieser Arbeit war die Entwicklung und Synthese elektrophiler „warheads" mit unterschiedlicher Reaktivität, um das Cystein im aktiven Zentrum von DUBs entweder durch Michael-Addition oder nukleophile Substitution einzufangen. Die Konjugation der jeweiligen „warheads" an eine Glycineinheit soll eine anschließende Transpeptidierung an den C-Terminus von Ub mittels Sortase ermöglichen.

In dieser Arbeit haben wir sowohl literaturbekannte elektrophile „warheads" als auch neuartige, weniger reaktive Michael-Akzeptoren auf Basis von α,β-ungesättigten Amiden synthetisiert. Zwei auf Lysin basierende α,β-ungesättigte Amide könnten sich darüber hinaus als wertvolle Werkzeuge zur Untersuchung der DUB-Aktivität mittels einer Erweiterung des genetischen Codes erweisen. Des Weiteren ist es uns gelungen, einen selektiven Syntheseweg für hochreaktive α,β-ungesättigte Ketone zu etablieren, die durch eine photolabile Schutzgruppe maskiert sind. Bei Bestrahlung mit UV-Licht wird die Reaktivität gegenüber Nukleophilen wiederhergestellt, so dass DUBs nach erfolgreicher Sortylierung eingefangen werden können.

Part A – Towards Optical Control of Protein Function Using an Expanded Genetic Code

1 Introduction

1.1 And Then There Was Light – Optogenetics

Light is the ideal tool when it comes to probing biological processes in living cells and organisms with high spatial and temporal control. Applied correctly (e.g. visible wavelengths, low intensity) it is non-invasive and orthogonal to most elements in a living system. The last decade has given rise to light-controlled technologies that have revolutionized many research fields from cell biology to neuroscience and even medical approaches with potential therapeutic applications such as photodynamic therapy are currently under investigation.[1, 2]

Nature has long succeeded in using the unparalleled properties that light has to offer. Interestingly enough, only a few classes of chromophores are responsible for rendering certain biological functions responsive to light, the most prominent ones being retinal and flavins (Figure 1.1).

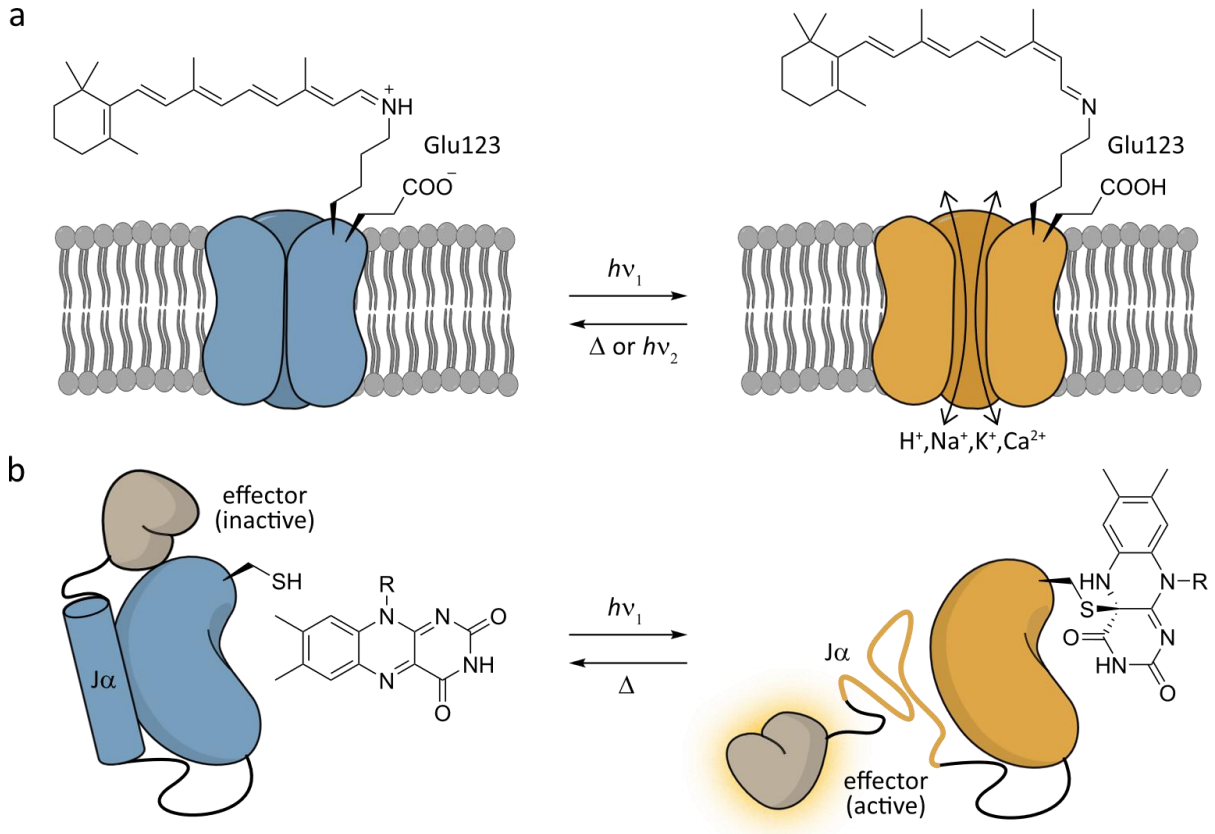


Figure 1.1 | **Light-sensitivity found in the genetically encoded proteins rhodopsin and LOV.** (**a**) Microbial opsins (type I) conduct ions across the membrane with a retinal molecule remaining associated to the opsin protein (contrary to type II opsins). Depicted here the algal channelrhodopsin (ChR), covalently bound through a lysine residue to the all-*trans* isomer of retinal (blue). Upon irradiation (λ = 470 nm), retinal isomerizes to the 13-*cis* isomer, triggering the transfer of its Schiff base proton to Glu123 of ChR. As a consequence, non-selective cation flux along the electrochemical gradient is initiated for ChR (orange). Isomerization back to the *trans*-configuration proceeds either through thermal relaxation Δ or irradiation with green light (λ = 520 nm).[3] (**b**) LOV domains are non-covalently associated with the oxidized flavin cofactor in the dark state (blue). Upon blue light-irradiation (λ = 450 nm), LOV domains form a covalent adduct with the flavin cofactor through their active site cysteine, resulting in large conformational changes. Shown here is the displacement and unfolding of the Jα helix in the LOV2 domain from *Avena sativa* (orange). When coupled to an effector (grey), the conformational change can switch the effector from its inactive form to its active form. The covalent adduct decays as soon as illumination ceases, returning the LOV2 domain back to its dark state conformation (blue).[4]

As key regulator involved in animal vision, the natural photoswitch retinal (vitamin A aldehyde) is typically found in the retina as part of photosensitive membrane proteins, termed rhodopsins. Interconversion between the *cis*- and *trans*-isomer of retinal mediates phototransduction to the brain, resulting in a visual signal.[5] The necessary light-targetable complex rhodopsin is created through Schiff base formation of retinal with a highly conserved lysine amino acid residue of microbial opsins (type I) or G protein-coupled receptor (GPCR) opsins (type II). While type I opsins modulate the ion flow across the plasma membrane in response to stimulation with visible light, type II opsins mainly initiate protein-protein interactions and play key roles in vision and the circadian rhythm.[3]

Flavins can be encountered in various redox processes, where they act as cofactors of enzymes in the form of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Amongst many others, such enzymes include so-called flavoprotein photosensors. Herein, the light-oxygen-voltage (LOV) protein family is best understood and used in optogenetics. Upon irradiation with blue light, these domains undergo large conformational changes through formation of a covalent thioether linkage between a conserved cysteine amino acid residue and FMN or FAD (Figure 1.1b). Together with the cryptochrome (Cry) protein family, these flavin-binding photoreceptors play important roles as blue-light sensors in circadian clocks (Crys) of both plants and animals, as well as cellular signal-transduction networks (LOV domains) of plants and microorganisms.[2, 6, 7]

These examples depict nature's capability of developing molecular devices that, upon an external trigger such as light, carry out fundamental tasks with exceptional precision and efficiency, even when embedded within complex environments. Mankind's search for a comparably precise tool that can regulate singular events with high spatiotemporal resolution within organisms, tissue or even on a subcellular level, however, has been an unmet need amongst scientists across various fields until the recent decade. This included Francis Crick who turned to neurobiology after his ground-breaking work in the field of molecular biology. Seeking to understand the elaborate neural activities around consciousness and visual perception, in his article "Thinking about the Brain" from 1979 he stated the need for a new technology that would allow "all neurons of just one type to be inactivated, leaving the others more or less unaltered".[8]

Although various light-gated proteins such as bacteriorhodopsin[9] or halorhodopsin[10] were already described in the early 1970s, it took another 35 years to address the problem repeatedly stated by Francis Crick and other neuroscientists. Following the characterization of channelrhodopsin in 2002[11], Deisseroth and co-workers succeeded in the targeted delivery of the algae protein channelrhodopsin-2 into neurons only three years later.[12] The introduction of light-sensitivity through microbial opsins allowed for targeted optical control of neuron excitability. This breakthrough paved the way for a quickly evolving area of scientific activity termed optogenetics the following year.[2, 13]

Similar to the scientific area of chemical biology, the interdisciplinary field of optogenetics strives to find and develop molecular tools to address major biological questions. Originally, the field of optogenetics arose with the aim of investigating the activity of neurons and other electrically excitable cells in a precise and non-invasive manner. It does so by expressing genetically encoded, light-targetable ion channels, pumps and other proteins, thus allowing for the spatio-temporal control of various biochemical signaling pathways.[14]

The application of optogenetics has quickly expanded since it was first introduced in neuroscience in 2005. Molecular engineering and directed evolution approaches coupled to screening pipelines[15] continue to increase the repertoire of photobiosensors, including a variety of ion-channels (esp. flavin/LOV-domains), G protein-coupled receptors (GPCRs, esp. retinal/opsins) and other light-sensitive proteins (e.g. photoactivatable fluorescent proteins such as Dronpa or the photoactive yellow protein). By now applications encompass the photocontrol of cellular processes such as motility[16], protein degradation[17], nuclear-cytosolic translocation[18], enzyme activity[19] and gene expression[20], as well as the preparation of photoswitchable nanobodies, so-called OptoNBs[21].

Still, rendering physiological processes light-responsive can only be done to a certain extent on a genetic level. Aspects like the need for successful gene delivery and/or expression, as well as the size of the described genetic modifications itself may pose substantial challenges depending on the cell type or biological question. Undesired side-effects due to overexpression altering cellular dynamics further limit the use of optogenetics.

An expansion of the field towards photoresponsive small molecules offers a way for optical control with minimal perturbation. The approach of endowing biomolecules with chemically synthetic tools, creates hybrid small molecule:protein devices, that are being reprogrammed to act upon an external stimulus – light.[22] Heavily used nowadays in chemical biology, initial efforts by chemists in the late 1970s have proven that the activity of biomolecules can be controlled through installation of light-sensitive synthetic molecules, such as “caged” cAMP or ATP derivatives.[23, 24] Since then, a broad array of small molecule inhibitors and metabolites, as well as nucleic acids, peptides and proteins have been rendered light-responsive through the installation of photosensitive small molecules, applicable not only *in vitro* or *in cellulo*, but also in complex cellular networks such as animals.

1.2 Photosensitive Small Molecules

Two of the most widely used chemical approaches to install light-sensitivity into biomolecules include the use of (i) light-cleavable protecting groups, so-called “photocages”, as irreversible triggers for unmasking or (ii) bistable photoswitches for reversibly switching protein function between an “on” and an “off” state.[25]

1.2.1 Photocaging

Photolabile protecting groups (PPG) suppress the bioactivity of molecules, until irradiation with appropriate wavelengths releases the original molecule *in situ* and thereby restores bioactivity. PPGs have been used extensively for optical regulation of oligonucleotides[26-28], leading to the recent development of a very fast CRISPR system (vfCRISPR) by installation of three caged nucleosides into the protospacer adjacent motif (PAM) region of the guide RNA[29]. Furthermore, PPGs have been applied to mask various metabolites[30], enabling time-resolved NMR studies of biochemical reactions[31]. Last but not least, photocages have been exploited for the light-regulation of a broad range of proteins, either by directly installing caged amino acid residues through genetic code expansion[32] (Figure 1.2a, Chapter 1.5.1), or through the use of masked inhibitors and effectors (Figure 1.2b)[33-35].

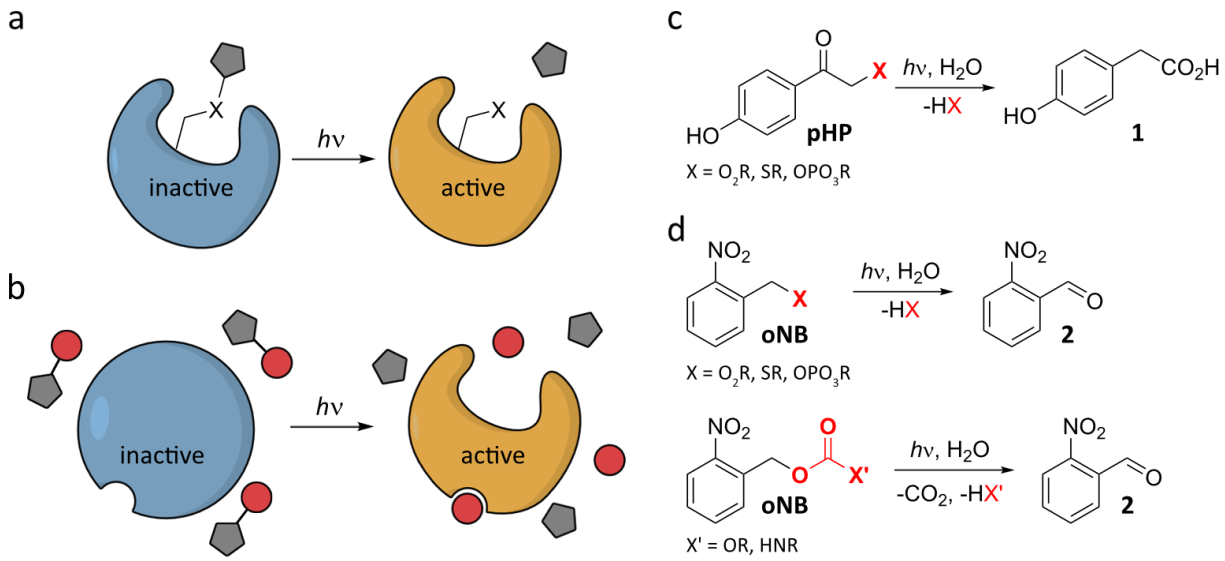


Figure 1.2 | Photolysis of caged molecules and their application to modulating protein function. (**a**) A photocaged amino acid analog installed onto the protein renders it inactive (blue). When irritated with UV-light, the photocage is released and the active site amino acid residue (X = SH or OH) unleashed. (**b**) The use of UV-light releases an effector (red) from the photocage (grey). This effector interacts with the target protein (blue), triggering a conformational change to its active form (orange). (**c, d**) Photolysis of **pHP** and **oNB** releases the cargo X (red) to yield 4-phenylhydroxyacetic acid (**1**) and 2-nitrosobenzaldehyde (**2**) as byproducts, respectively. Typical leaving groups are carboxylates, thiolates and phosphates. (**d**) Alcohols and amines (X') are implemented as carbonates and carbamates instead.

The most prominent PPGs include phenacyl, e.g. *p*-hydroxyphenylacyl (**pHP**) or nitrobenzyl compounds, such as *o*-nitrobenzyl (**oNB**) (Figure 1.2c, d). Mechanistically, the **pHP** removal proceeds via a skeletal rearrangement, the so-called "Favorskii intermediate", accompanied by a fast, water-assisted release of the substrate.[36] In contrast to **pHP**, liberation of the substrate from a hemiacetal intermediate of **oNB** is rather slow. This rate-determining step is especially poor for leaving groups such as methanol.[37] Nonetheless, **oNB** derivatives have been attached to various leaving groups, typically masking carboxylic acids, thiols or phosphates (Figure 1.2d, upper panel). Alcohols and amines are installed as the synthetically more readily accessible carbonic acid derivatives instead, thereby also turning them into better leaving groups via release of CO_2 (Figure 1.2d, lower panel).

"Decaging" of the described PPGs usually requires high-intensity UV-light and/or generates strongly-absorbing byproducts, e.g. nitrosoaldehydes. Such byproducts do not only act as quencher but they may also be toxic to the surrounding environment.[38] Therefore, scientific research has brought forth a repertoire of alternative photocages. These light-responsive molecules can undergo rapid photolysis due to large extinction coefficients, such as coumarin **3** and **4** or BODIPY chromophores **5** and **6** (Figure 1.3). Considerable efforts have focused on improving these visible light-activatable coumarine and BODIPY scaffolds further towards longer wavelengths[39, 40] and ultimately near-infrared light (NIR).[41] Together with the introduction of nitrodibenzofurane **7** as predictable two-photon caging group[42], the toolbox of PPGs vastly expands towards *in vivo* and potential therapeutic applications (Figure 1.3).[43, 44]

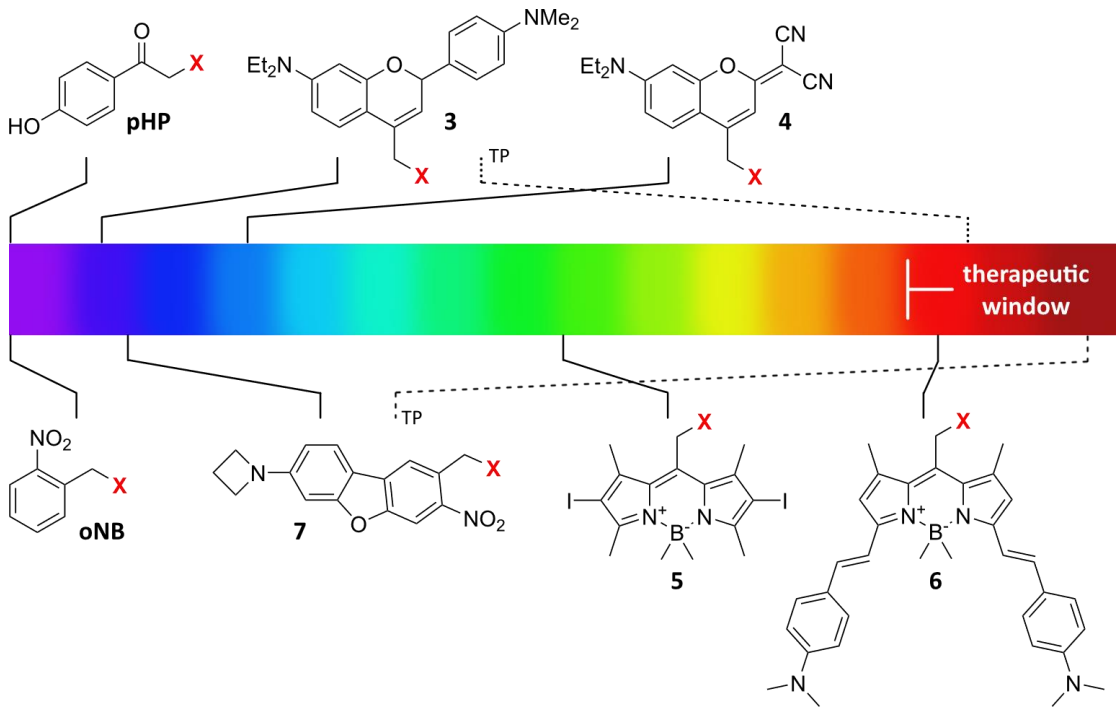


Figure 1.3 | Overview of different caging groups. The photocages are sorted according to the wavelengths which are needed for liberation of cargo X (red). Solid lines represent photolysis upon one photon absorption, while dashed lines represent photolysis upon two photon (TP) absorption. The UV-light activatable cages **pHP** and **oNB** are depicted as a reference. The donor-π-acceptor push-pull system in coumarin **4** allows for decaging using red-shifted wavelengths (λ_{Ex} = 480 nm).[45] A donor-π-donor coumarin backbone results in TP-activatable compound **3** (single photon, λ_{Ex} = 400 nm or TP, λ_{Ex} = 750 nm).[41] Boron-alkylation of the BODIPY scaffold yields photocage **5** with exceptional quantum yields *φ(λ)* of 95% (λ_{Ex} = 540 nm)[46], while derivative **6,** which has been further extended with styryl moieties, features absorptions with the highest known wavelengths using single photons (λ_{Ex} = 690 nm).[40] Azetidinyl-substituted nitrodibenzofurane **7** results in efficient cargo release using blue (λ_{Ex} = 420 nm), green (λ_{Ex} = 530 nm) or IR light (TP, λ_{Ex} = 850 nm).[42]

Despite these recent advances, PPGs only allow for the implementation of an irreversible "on" switch. Photoswitchable compounds on the other hand provide two key advantages over photocages. First, they provide a fast and reversible switching process, ideally with high fatigue resistance over several cycles (i.e. no photodegradation), and secondly, no toxic byproducts are formed upon low-intensity irradiation with light. Reversibility of the process is important in elucidating the function of proteins, since it serves as a strong experimental control and furthermore minimizes potential irreversible downstream effects, which are more frequently encountered when protein functions are being altered over a long period of time.[47]

1.2.2 Photoswitches

The process of photoregulation using molecular switches is, similar to photoreceptors such as retinal/opsin (Figure 1.1a), based on the light-induced interconversion between two states of the chromophore, hence the term "photoswitch". Over the last decades photoswitches have found applications ranging from material[48] to life sciences[49].

In general, photoswitches can be divided into two groups based on the photochemical process. Diarylethenes and spiropyranes undergo electrocyclizations between "closed" and "open" forms upon irradiation with light of a certain wavelength (Figure 1.4a) and will not be discussed in detail here. More interesting for the work proposed within this thesis are photoswitches undergoing geometric, E/Z isomerization of a double-bond. They are relatively small and can be easily incorporated into larger molecules. These include indigoid scaffolds, stilbenes and azobenzenes (Figure 1.4b).

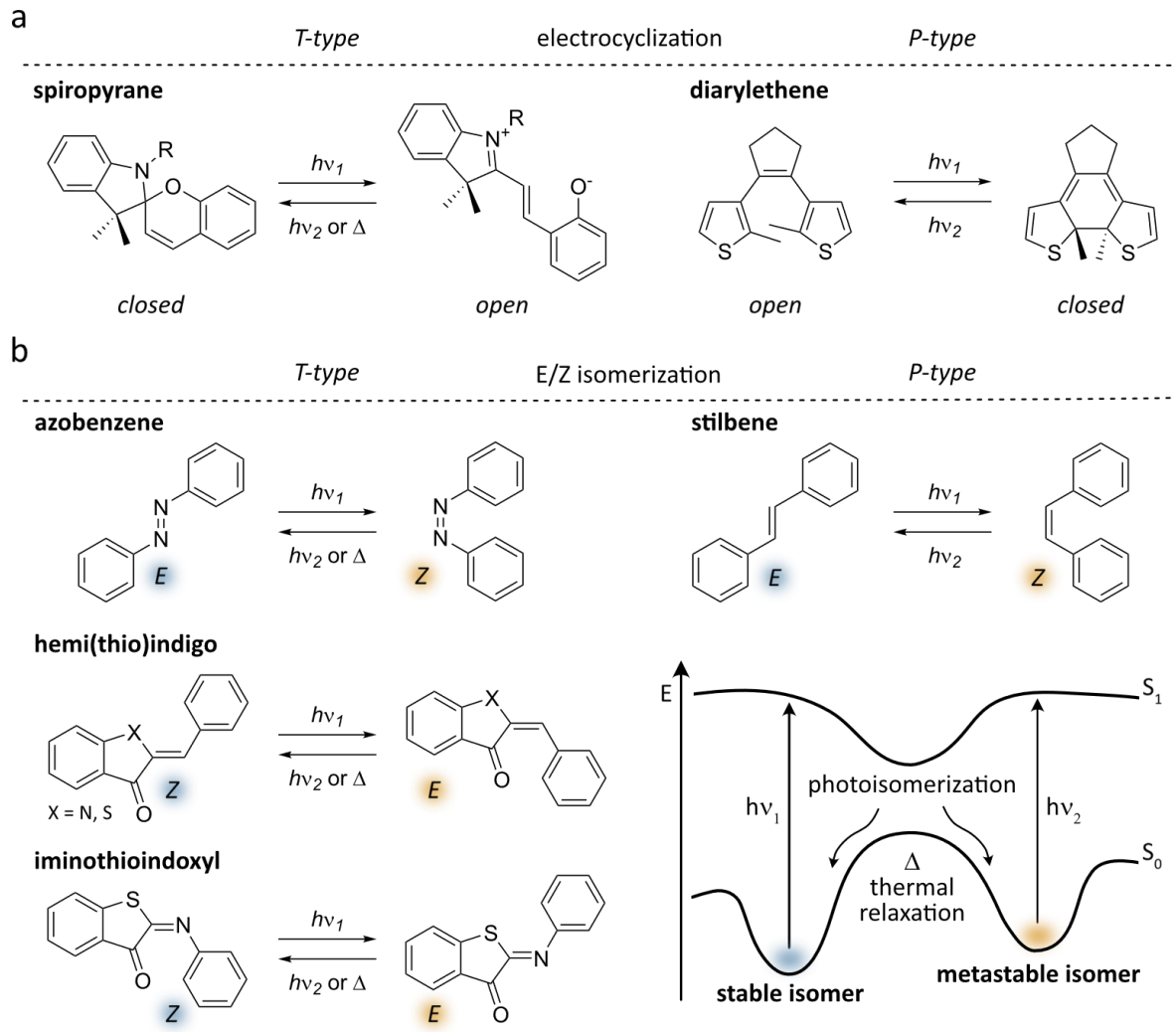


Figure 1.4 | Overview of different photoswitches and their isomerization process. (**a**) Diarylethenes switch between the closed and open form via a 6π electrocyclization, while spiropyranes undergo an additional *trans-cis* isomerization after the cyclization.[50] (**b**) Azobenzenes, stilbenes, hemi(thio)indigos and iminothioindoxyls only undergo the isomerization around the double-bond. The stable state is shown in blue, the metastable state is shown in orange. A general graph shows the potential energy surfaces (S_0 = ground state, S_1 = excited state) of a photoswitch during isomerization.[49]

The photoisomerization process usually involves a one-photon mechanism. Upon excitation (S_1) of the parent isomer, the chromophore relaxes back to the ground state (S_0) through isomerization around a double-bond. The new isomeric form is higher in energy than the original one and thereby often metastable, leading to spontaneous isomerization back to the thermodynamically stable form (Figure 1.4b). Depending on whether the thermal activation barrier between metastable and stable isomer is low or high, these photochromic molecules can be either classified as T-type, i.e. a significant amount of back isomerization is observed, or P-type, namely bi-stable, photoswitches (Figure 1.4). The thermal relaxation process of T-type switches is typically described by first order kinetics with a corresponding thermal relaxation time τ and half-life $t_{1/2}$.[49] The extent of photoisomerization is commonly characterized by the photostationary state (PSS). The PSS describes the dynamic equilibrium mixture of the two isomeric states, generated under irradiation with light of a specific wavelength.

Photoswitches are usually positively photochromic, which means the absorption maximum of the generated, metastable species is bathochromically shifted in respect to the stable isomer. Although photochromism is often described as a reversible, non-destructive process, isomerization performance can actually be lost over several cycles. This observation also referred to as "fatigue", has its origins in various side-reactions, e.g. photodegradation through oxidation, irreversible cyclization or dimerization.[51] For probing biological processes, these photoswitching molecules should not only show little to no fatigue, but they also need to be fast-switching. Therefore, they require high photoisomerization quantum yields $\varphi(\lambda)$ and large extinction coefficients $\varepsilon(\lambda)$ at the wavelengths of excitation, which ideally are innocuous for living cells.

The use of stilbene and stiff-stilbene, a 5-membered fused ring analog of stilbene, is quite challenging due to the prerequisite of harsh UV-light for isomerization (Figure 1.4b).[52] Instead, considerable effort has been spent on the development of indigoid and azobenzene photoswitches. Among the indigoid scaffold, hemithioindigos (HTI) might very well hold the biggest potential for applications in biology. These unsymmetrical, rigid chromophores consist of a thioindigo moiety and a stilbene moiety, connected via a central double-bond (Figure 1.4b). These extremely bi-stable photoswitches show little fatigue over many switching cycles and are typically isomerized with visible light wavelengths, accumulating the respective isomer in high yields. However, continuous irradiation is necessary to achieve near-quantitative photoisomerization, since the quantum yield $\varphi(\lambda)$ of hemithioindigos is comparatively low, ranging from 5-15%.[53] Still considered an attractive choice of photoswitch, early efforts have already proven that HTIs can drive structural change in peptides.[54] Iminothioindoxyls (ITI) combine photochromic fragments of thioindigos and azobenzenes (Figure 1.4b). These visible-light, fast-switching molecules show an unparalleled band separation of over 100 nm, thereby further expanding the toolbox of reversible photoswitches holding great potential in chemical biology.[55, 56]

While indigoid switches have only recently been more broadly explored for biological applications, the azobenzene is clearly the most-often-used molecular photoswitch. It comprises many advantageous key features, such as a well-studied and understood photoisomerization process, exceptional photostability over many cycles, facile syntheses and simple tuning towards the desired properties. Consequently, azobenzenes have dominated the field, when it comes to studying complex biological processes in a reversible manner.

1.2.3 Azobenzene

Over the last two centuries, the azobenzene molecule has evolved from a simple dye, starting out as aniline yellow in 1861, to one of the most broadly used molecular switches by far. Not only has it become ubiquitous in fundamental and applied research but also in many areas of our everyday lives, including textiles, cosmetics, food and medicine.[57] Azobenzene photoswitches have furthermore emerged as versatile tools in photopharmacology[58-60], robotics[61], solar energy storage[62] as well as self-immolative[63] and self-healing polymers[64].

In 1937, Hartley reported on the light-driven isomerization process of azobenzene for the first time since its introduction to the dye industry. He observed the isomerization from the *E*-form (*trans*-isomer) of azobenzene to its metastable *Z*-form (*cis*-isomer) and vice versa (Figure 1.5a).[65]

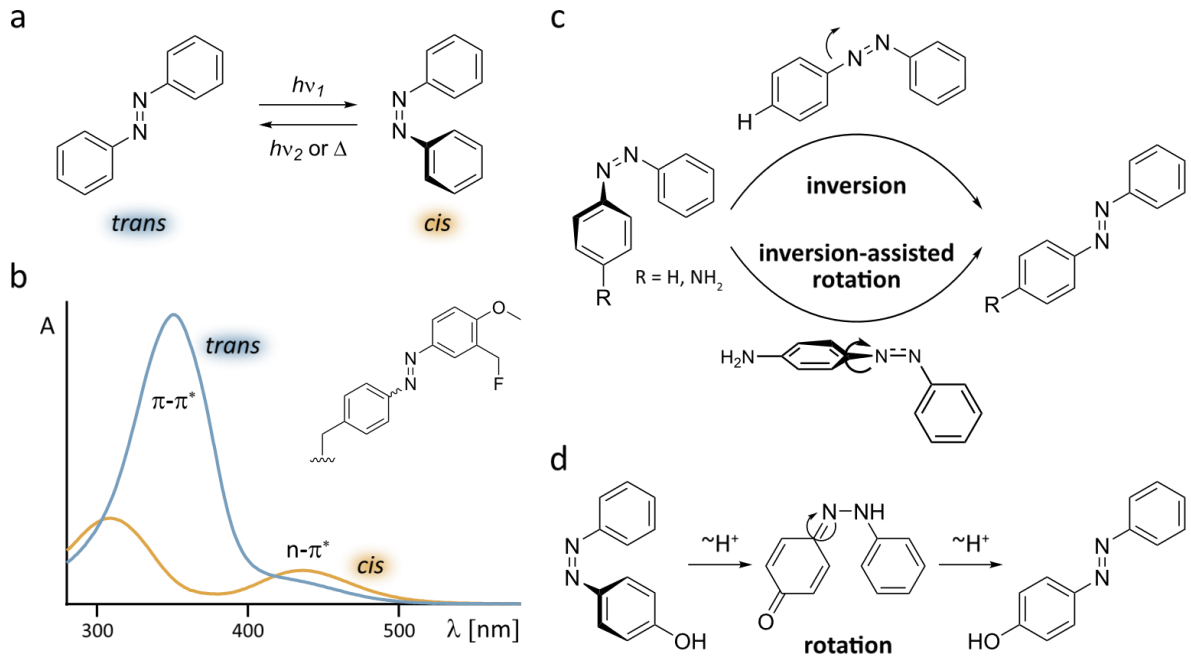


Figure 1.5 | Photoisomerization and thermal relaxation of azobenzenes. (**a**) General isomerization scheme for azobenzene. (**b**) Typical absorption spectrum for azobenzenes featuring an alkoxy moiety in *para* to the diazene. The absorption spectrum for the *trans*-isomer is depicted in blue, featuring a strong π-π* absorption band. For the *cis*-isomer a pronounced n-π* absorption band can be observed, shown in orange. (**c**) Thermal relaxation scheme for standard and *para*-amino substituted azobenzenes. (**d**) Thermal relaxation scheme for *para*-hydroxy substituted azobenzene via hydrazone-like intermediates.[49]

Photoisomerization of azobenzene occurs through excitation of the *trans*-isomer from its ground state (S_0) to its first (S_1) or predominantly to its second singlet excited state (S_2). The various reaction pathways and electronic excited states involved in the isomerization process, have been repeatedly discussed over the decades. With new insights still being revealed while also refuting earlier proposals, the mechanism will not be discussed in detail within this work.[66]

There are two main electronic transitions which are relevant to the photoisomerization process of azobenzene, namely the π-π* and *n*-π* transitions. The symmetry-allowed π-π* transition (S_2) gives rise to a strongly absorbing band in the UV region for the *trans*-isomer, while a weakly absorbing band in the visible region arises from the symmetry-forbidden *n*-π* transition (S_1). Contrary to the parent azobenzene, the *n*-π* transition is more pronounced for the *cis*-isomer, while the π-π* transition is only half as intense and shifted towards the UV region (Figure 1.5b). Generally, the decay to the S_0 state is accompanied by a rotation around the ruptured N-N single bond for π-π* transitions, while n-π* transitions can form the respective isomer via one of several controversial pathways.[67] Switching between the planar, thermodynamic ground-state *trans*-isomer and the metastable, twisted *cis*-isomer is extremely fast due to large extinction coefficients ($\varepsilon(\lambda) \approx 20\text{-}30\times10^3$ M^{-1} cm^{-1} for *trans*, $\varepsilon(\lambda) \approx 15\times10^2$ M^{-1} cm^{-1} for *cis*) and sufficient photoisomerization quantum yields ($\varphi(\lambda) \approx 0.15\text{-}0.4$ for *trans* and *cis* respectively).[67] Isomerization therefore can be triggered by low-intensity irradiation at wavelengths compatible with *in cellulo* settings (>350 nm). Photoisomerization is further accompanied by a substantial change in geometry, polarity and end-to-end distance (approx. 3.5 Å). The latter is often reasoned to be the driving force for conformational changes in peptides or proteins when using azobenzene crosslinkers.

Thermal relaxation of the *cis*-azobenzene back to the *trans*-isomer, usually through inversion, is associated with a relaxation half-life $t_{1/2}$ that is strongly depending on temperature, solvent, pH and polarity of the environment. Typical relaxation times range from hours to days at room temperature, while *para*- or *ortho*-amino substituted azobenzenes exhibit a half-life $t_{1/2}$ ranging from seconds to minutes. The relaxation of such derivatives proceeds via an inversion-assisted rotation mechanism instead (Figure 1.5c). Even faster thermal relaxation with a half-life $t_{1/2}$ ranging from nano- to milliseconds can be observed for *para*- or *ortho*-hydroxy substituted derivatives. This is most likely due to intramolecular interactions of hydroxyl groups with solvent molecules, favoring the formation of hydrazone-like intermediates which facilitates relaxation through rotation (Figure 1.5d).[68]

Oftentimes, a distinct overlap in the absorption spectra of both isomers can be observed (Figure 1.5b). For applications in biology this is quite problematic, since it renders complete *trans*→*cis* conversion (and vice versa) photochemically impossible. *trans*→*cis* photoisomerization typically leads to a PSS composed of approximately 80% *cis*-isomer at best, while for wavelengths favoring the *cis*→*trans* isomerization, the PSS is typically composed of 90% *trans*-isomer. Thermal relaxation on the other hand yields back the *trans*-isomer in quantitative yields.[67]

A large section of scientific research is therefore engaging in the development of azobenzene derivatives, which feature a more pronounced band separation. This can either be improved through the introduction of tetra-*ortho*-methoxy[69], -chloro[70] and -fluoro[71] substituents on the phenyl rings or by using constrained, (hetero-)bridged azobenzenes, so-called diazocines[72-75] (Figure 1.6).

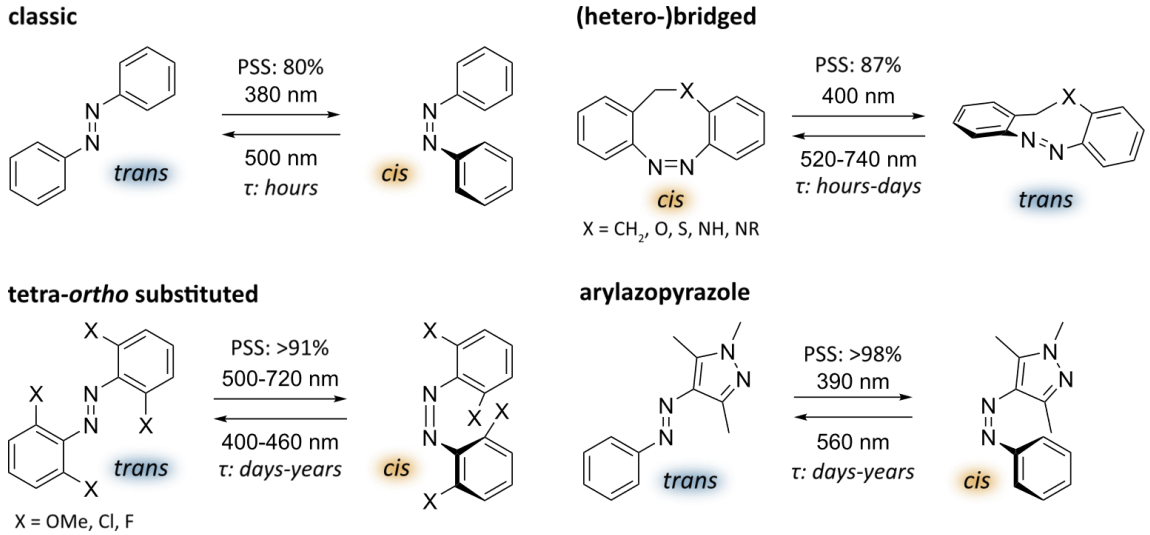


Figure 1.6 | Tailoring of azobenzenes for the optical control of protein function. General isomerization schemes for classic azobenzene, tetra-ortho substituted azobenzene, (hetero-)bridged and arylazopyrazole. Also depicted are the approximate wavelengths needed for switching, relaxation times τ and PSS for isomerization to the metastable isomer.

Pioneered by the groups of Woolley and Hecht, tetra-*ortho*-substituents on the chromophore separate the *n*-π* bands of the *trans*- and *cis*-isomer by 42 nm at best. Contrary to that, diazocine derivatives, whose thermodynamically stable isomer is the *cis*-form, are showing much larger band separations around 100 nm (Figure 1.6). Not only does the use of such molecules facilitate near-quantitative photoconversion, it is also concomitant with visible light being applied to drive the isomerization process.

Although UV-light seems to be tolerated to a large extent in cell cultures, a significant number of publications on red-shifted azobenzenes can be found in the literature. This is also due to the azobenzene's advance towards potential therapeutic applications, where the penetration-depth of light is extremely dependent on the wavelength, creating a demand for red/NIR light switchable azobenzenes.[47] Initially, modifications to shift absorption towards longer wavelengths, also lowered the thermal barrier for isomerization, i.e. leading to faster thermal relaxation. Classic examples are push-pull systems, in which the *para*-positions relative to the diazene are linked to electron donating groups on one side and electron withdrawing groups on the other. While beneficial regarding switching wavelength, such systems have been reported to have low thermal stability and also low photoisomerization yields, limiting their advance in biology.[69]

There are certain scenarios in which a fast, thermal relaxation process is however considered beneficial. This is predominantly the case for therapeutic approaches using antibiotics or for cancer treatment. Herein, an inactive, photoresponsive drug is administered and switched to the active form only at the target site, ideally by illumination with (near-)infrared light. Once outside the irradiated target area, it spontaneously isomerizes back to its thermodynamically stable, but inactive form, allowing for highest spatio-temporal control over drug activity.[50] Azobenzenes fulfilling such criteria are still very scarce. It has been shown that azonium ions derived from protonation of tetra-*ortho*-methoxy substituted azobenzenes exhibit a very short thermal half-life $t_{1/2}$ (microsecond) while responding to wavelengths in the region of NIR light.[76, 77]

For reversibly controlling protein structures, a bi-stable photoswitch with a large thermal half-life $t_{1/2}$ is essential (Figure 1.6). Here, arylazopyrazoles could be extremely useful due to their extraordinary isomerization efficiency and high thermal stability of the *cis*-isomer (Figure 1.6).[78] Furthermore, the aforementioned tetra-*ortho*-substituted or bridged azobenzene derivatives present a good choice. Tetra-*ortho*-methoxy substituted azobenzenes exhibit long-wavelength photoswitching exceeding 600 nm, while maintaining long half-lifes $t_{1/2}$ around two days. Application of this particular photoswitch however, is limited since physiological glutathione concentrations (10 mM) have led to its complete reduction within a few hours.[69] While tetra-*ortho*-fluorinated azobenzenes do not show the same degree of *n*-π* red-shifting, they are stable towards glutathione with a half-life $t_{1/2}$ of up to 2 years and have been used in cellular settings.[79]

Ultimately, the ideal azobenzene photoswitch for *in vivo* applications should be i) stable towards glutathione and reductases, ii) water-soluble, iii) fatigue-resistant over many irradiation cycles, iv) deeply red/NIR-shifted, v) bi-stable as well as vi) yielding complete photoconversion upon irradiation (in both directions). To that end, researchers are further working on the optimization of substitution patterns for the promising, tetra-*ortho*-substituted scaffolds through computational and experimental studies.[80-82]

While there have also been tremendous efforts put into implementing facile syntheses and late-stage functionalization of azobenzenes that are deserving of recognition, they will not be discussed in detail within the scope of this work. A short and basic overview of standard syntheses to access azobenzenes is given in Chapter 3.1.1.

1.3 The Advance of Azobenzenes in Biology

Over the last decade, the photoisomerization of azobenzene has been widely applied to biological systems, modulating the function of much larger biomolecules with unparalleled precision. Early aspirations aimed at probing and manipulating fundamental processes such as replication, transcription and translation. With photoswitches as potential control elements to study these processes with high spatio-temporal resolution, the introduction of azobenzenes into nucleic acids seemed particularly appealing and has led to numerous publications.

1.3.1 Optical Modulation of Nucleic Acid Morphology by Azobenzenes

The development of the nowadays widely used photoswitchable nucleoside **tAzo**, i.e. a threoninol moiety linked to an azobenzene allowed for facile incorporation of photoswitches as nucleoside surrogates (Figure 1.7).[83]

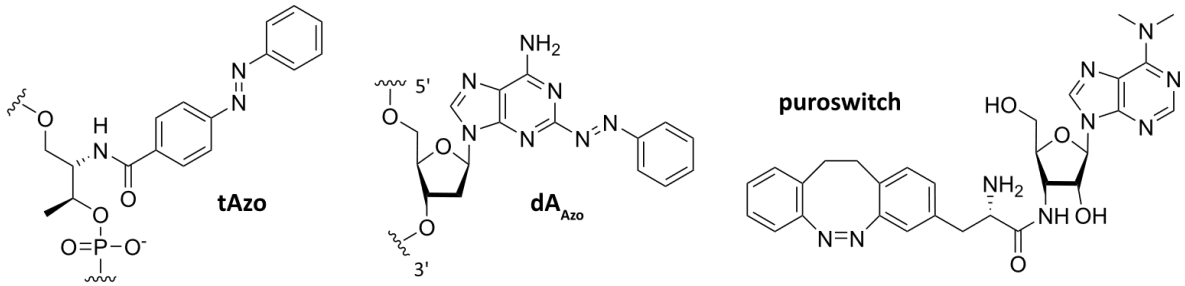


Figure 1.7 | Overview of azobenzene nucleoside surrogates to study nucleic acid duplex stabilization/destabilization and optically control transcription or translation. Directionality is indicated via 5' and 3'.

Usually, the *trans*-isomer stabilizes the DNA duplex through stacking interactions, which allowed for the regulation of DNA duplex stability[84] and hybridization[85] with light. Purine-based nucleoside surrogates, e.g. the photoswitchable nucleobase **dA$_{Azo}$**, improved the overall hybridization affinity further (Figure 1.7).

The unique properties of azobenzene photoisomerization can also be exhorted to regulate oligonucleotide hybridization through non-covalent interactions. Photoswitchable molecular DNA glues are typically comprised of a small molecule azobenzene decorated with cationic tails[86] or mismatch-binding ligands,[87] allowing for intercalation into complementary oligonucleotides and thus stabilization of the duplex DNA in the *trans*-form. Under UV-light irradiation, the molecular glue isomerizes to the non-planar *cis*-form, thereby disrupting π-stacking and reducing the stability of the DNA double-strand.

Recent advances have started to go beyond DNA hybridization studies and towards the optical control of gene expression. Amongst others, this was demonstrated by Asanuma and co-workers, who installed a photoresponsive T7 promotor using **tAzo**-analogs to regulate gene expression in a cell-free translation system by light.[88] Isomerization to the *cis*-isomer destabilizes the unwinding region and facilitates the binding of T7 RNA polymerase, thus initiating transcription. Similarly, photocontrollable caps (PC-cap), meaning mRNAs bearing azobenzene-based 5'-caps, were applied to control gene expression in zebrafish embryos in a reversible manner by photoisomerization between the *trans*- (i.e. no translation) and the *cis*-form (i.e. translation).[89]

Ko et al. recently developed a photoswitchable analog of the nucleoside antibiotic puromycin, termed **puroswitch** (Figure 1.7). Puromycin features a non-hydrolyzable amide linkage between the 3'-end of a nucleoside moiety and the carboxyl group of an amino acid (aa), effectively mimicking aa-tRNA conjugates. As a result, puromycin gets recognized by the endogenous translational machinery and covalently traps the nascent peptide chain, stalling the ribosomal translation. Installation of a diazocine amino acid afforded **puroswitch**, capable of reversibly controlling protein translation in near-quantitative fashion due to the superb photophysical properties of the diazocine scaffold. Proof-of-principle experiments in HEK 293T cells revealed cytotoxic effects of **puroswitch** solely for the thermodynamically less stable *trans*-isomer. Subsequent incorporation into *de novo* synthesized proteins in mouse primary neurons further demonstrated the applicability of **puroswitch** as reliable tool for nascent proteome profiling.[90]

However, photoswitchable nucleic acid materials have found much broader application in the development of biosensing and optogenetic tools as well as drug delivery systems.[91]. Prusty et al. constructed a light-sensitive, spherical nanocarrier by lipid-mediated self-assembly of aptamers and photoswitchable DNA hairpins, thus enabling for the light-triggered release of cargo loads with high spatio-temporal precision. Whereas introduction of a **tAzo** derivative into GC-rich DNA hairpins allowed for controlled release of the chemotherapeutic doxorubicin, the aptamers provided the necessary specificity, targeting the hepatocyte growth factor receptor (cMet) of H1838 tumor cells.[92]

1.3.2 Photoswitchable Lipids and Ligands in Chemical Biology

Beyond light-responsive nucleic acid nanostructures, photoswitchable surfactants have also found use in cargo delivery systems by construction of liposomes from azobenzene-based lipids. The design of photoswitchable lipids is typically based on the incorporation of an azobenzene moiety into the hydrophobic lipid tail (Figure 1.8).

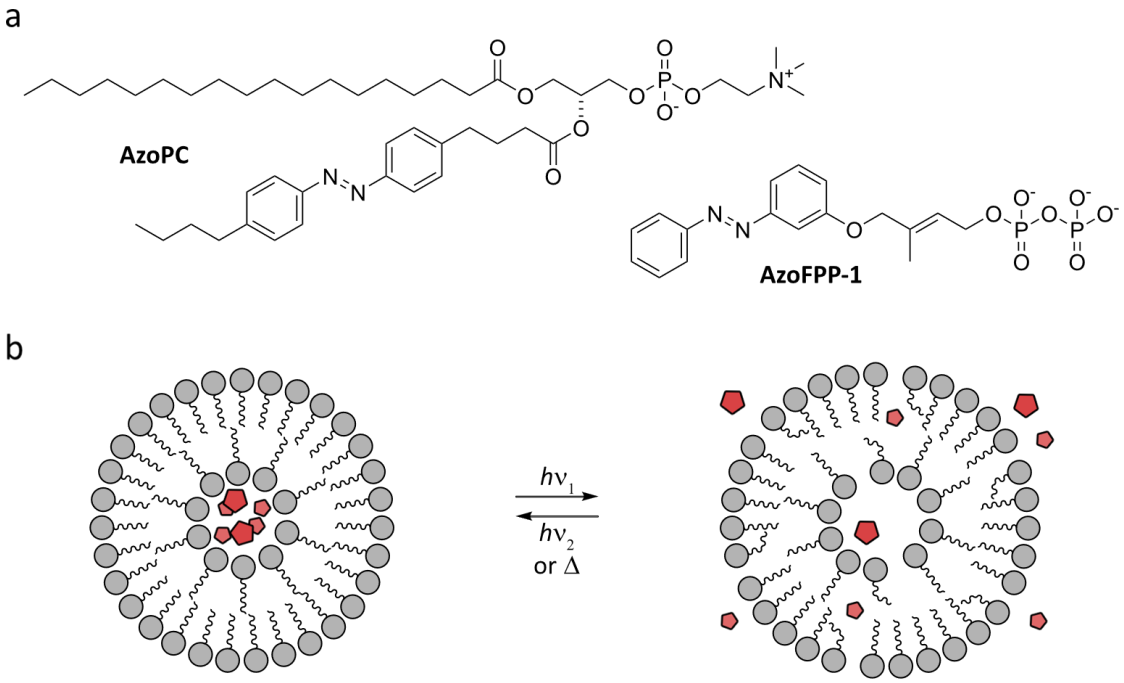


Figure 1.8 | Photoswitchable lipids applied in chemical biology. (**a**) Structural formulas of the photoswitchable phosphatidylcholine analog **AzoPC** and farnesyl diphosphate analog **AzoFPP-1**. (**b**) Schematic depiction of the photo-triggered release of cargo molecules (red) from azosome.

Initially introduced to study the effects of photoisomerization on lipid assembly, the azobenzene-based phosphatidylcholine **AzoPC** (Figure 1.8a) has become the major tool in studying membrane organization and liposome formation.[93, 94] Recent efforts have furnished **AzoPC** containing lipid nanoparticles (LNPs) for the light-induced release of the antimitotic drug doxorubicin *in vitro* and *in vivo* using red light, making them promising tools for therapeutic approaches.[95] Similarly, Xiong et al. developed a photosensitive liposome formulation comprising **AzoPC** for neuromodulation, termed azosome. Photoisomerization of **AzoPC** to the *cis*-isomer drastically changes the lipid bilayer, leading to an increase in permeability and thus release of neuromodulators (Figure 1.8b). Delivery and release of a dopamine D1-receptor agonist was successfully demonstrated *in vivo*, increasing the excitability of primary striatal neurons.[96]

Azobenzene moieties have further been introduced into lipids and fatty acid precursors to modulate biological functions such as membrane fluidity for protein secretion[97], cell signaling pathways for protein translocation[98] and cell growth[99] or the biosynthesis of lipids.[100] While the design of these photoswitchable lipids is mostly based on glycerol- and sphingolipids, Morstein et al. recently installed an azobenzene core into an isoprenoid tail for the first time. The thus obtained photoswitchable analog of farnesyl diphosphate (**AzoFPP-1**) was shown to control substrate prenylation by farnesyltransferase *in vitro* (Figure 1.8a), but had little effect on the subsequent processing steps i.e. proteolysis and carboxyl methylation.[101]

Beyond that, photoswitchable lipids have found broad application in the vibrant field of photopharmacology, where they are exhorted to reversibly modulate membrane receptor dynamics. The research area of photopharmacology, termed and heavily pioneered by Trauner and Feringa, can be defined as the optical regulation of essential biological processes such as cell signaling or proliferation via photoswitchable small molecules.[59, 60] Over the last decade, the pharmacodynamic and pharmacokinetic properties of azobenzenes have become broadly established, facilitating the application of synthetic switches based on lipids and ligands to a broad range of biological targets. Amongst these are various transmembrane proteins such as ion channels, ion pumps or GPCRs, but also soluble proteins including kinases and proteases.[102]

Ion channels, such as ligand-gated ion channels, voltage-gated ion channels or ionotropic glutamate receptors (iGluRs) are prime targets in photopharmacology due to well-matched kinetics of ion channel dynamics and azobenzene photoswitching. Early studies into voltage-gated potassium channels (Kvs), typically blocked through quaternary ammonium ions, led to the design of a variety of neuromodulators. These light-sensitive ligands bind at the tetraethyl ammonium binding site, typically blocking the ion channel in the *trans*-form (**BENAQ**, **DENAQ**, Figure 1.9a, b). Isomerization to the thermodynamically less stable *cis*-isomer leads to dissociation of the azobenzene blocker, affording optical control over neuronal firing *in cellulo*.[103] Follow-up studies have further demonstrated the successful application of **BENAQ** and **DENAQ** in hippocampal brain slices and as tool in vision restoration.[104, 105] Recent efforts have furnished diazocine-based potassium channel blockers that convert the typically dark-active ligands into dark-inactive ligands (**LAB-QA**, **CAL**, Figure 1.9b).[106, 107] Designing photochromic ligands (PCLs) in such a way that the thermodynamically stable isomer is the biologically inactive one, enables high spatio-temporal control over the induction of a biological effect.

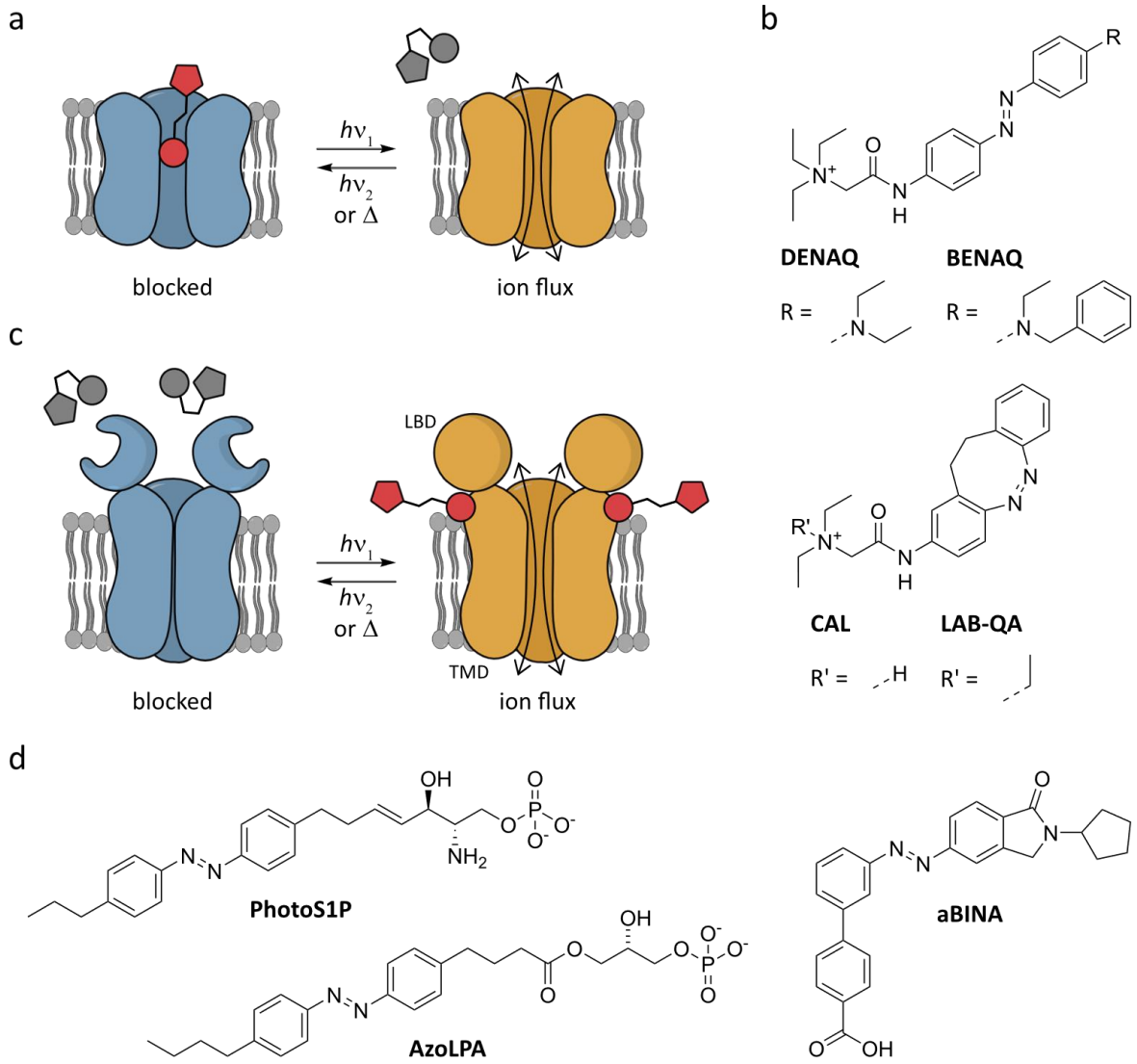


Figure 1.9 | PCLs used in the field of photopharmacology. (**a**) Schematic depiction of a typical blockage of ion channels using the active *trans*-isomer (red). Photoswitching of the neuromodulator to its *cis*-form reestablishes ion flow across the membrane. (**b**) Structural formulas of diverse azobenzene-based PCLs such as **DENAQ** and **BENAQ**, including the diazocine-based neuromodulators **CAL** and **LAB-QA** for reversible modulation of potassium ion channels. (**c**) Schematic overview showing the allosteric activation of GPCRs with PCLs in the *trans*-form (red). LBD = ligand binding domain, TBD = transmembrane domain (**d**) Structural formulas of lipid-based PCLs **PhotoS1P** and **AzoLPA** applied in the elucidation of lipid signaling pathways. Structural formula of **aBINA**, a PCL allosterically regulating the activity of GPCRs.

Beyond the various classes of ion channels, photopharmacology studies have predominantly focused on GPCRs, the most prominent class of membrane proteins in the human genome. GPCRs play key roles in cellular signaling processes and are widely implicated in a range of neurological and metabolic disorders, cancer or cardiac disease. Hence, they have emerged as prime drug targets.[108]

Some photoswitchable small molecules that have been used to modulate and study the biological function of GPCRs are in essence based on freely diffusible lipids or fatty acids such as sphingosine-1-phosphate (S1P) and lysophosphatic acid (LPA). Morstein et al. demonstrated the optical control of GPCRs $S1P_{1-3}$ using a photoswitchable, *trans*-active analog of S1P, termed **PhotoS1P** (Figure 1.9d). The high metabolic stability of **PhotoS1P** enabled to probe sphingolipid biology *in cellulo* as well as *in vivo* by spatio-temporally manipulating $S1P_3$-dependet heat hypersensitivity in mice.[109]

Similarly, the photoswitchable analog **AzoLPA** (Figure 1.9d) was shown to activate the GPCRs LPA_{1-5} in its *cis*-form, resulting in an increased intracellular concentration of Ca^{2+} in cells which were stably transfected with the respective LPA_{1-5} receptor. Morstein et al. further confirmed these results in the neuronal cell line NG108.15, demonstrating the light-dependent regulation of neurite branch outgrowth with **AzoLPA**.[110]

Beyond lipid-based GPCR modulators, Donthamsetti et al. developed an allosteric PCL (**aBINA**, Figure 1.9c, d) for the specific optical control of the metabotropic glutamate receptor 2 (mGluR2), which is an attractive drug target for treating Alzheimer's disease or schizophrenia. In general, the challenge with applying PCLs as GPCR-related therapeutics lies in their nonselective mode of action. Most GPCR agonists and antagonists bind highly conserved sites in the extracellular glutamate-binding ligand binding domain (LBD). Recent efforts have furnished a range of mGluR2-selective allosteric ligands that bind to the transmembrane domain (TMD) instead, such as benzyloxydimethylbenzene (BINA). The application of a photoswitchable analog of BINA, termed **aBINA** (Figure 1.9d), allowed for selective activation of mGluR2 in primary cortical neurons solely when in the *trans*-form.[111] Recent work in GPCR-related photopharmacology has further focused on the development of PCLs targeting adenosine receptors for pain relief[112] or the cannabinoid receptor 2, elucidating its therapeutic potential in neurodegenerative diseases.[113]

In general, azobenzene moieties can either be simply added to an existing molecular system, called "azoextension", or azobenzene isosters such as stilbenes or *N*-aryl benzamides can be exploited to replace the azobenzene, termed "azologization". Such approaches work for a broad variety of known ligands. However, the development of PCLs exhibiting a large enough difference in activity between the two isomers, therefore switching enzymes between fully active and fully inactive, remains challenging. In addition, PCLs are often unable to regulate the activity of a target protein specifically if it has several closely related family members. Similar to conventional therapeutics, PCLs also suffer from a lack in cell-type specificity.[60] These challenges can sometimes be tackled through the development of allosteric ligands as discussed before in the case of GPCRs, or by employing tethering approaches.[114]

In order to achieve cell-type or receptor-type specificity, the complementary photochromic tethered ligand (PTL)[115] and photochromic bioorthogonal ligand tethering (photoBOLT) approach[116] were developed (Figure 1.10). Ideally, the ligand cannot reach the binding site while in one isomeric form, but effectively binds when photoisomerized to the other configuration. For both approaches, a ligand binding the protein of interest (POI) is linked via a photoswitch to a reactive moiety. The latter can form a covalent bond with a bioconjugation motif, which was introduced into the target protein. In general, PTLs react with canonical amino acids, such as cysteines, while the attachment site for photoBOLT consists of a bioorthogonal handle that is site-specifically installed via genetic code expansion (Chapter 1.4). Application of photoBOLT afforded optical control over the kinase MEK1 by tethering of a photoswitchable inhibitor to the POI via bioorthogonal tetrazine inverse electron demand Diels–Alder reaction.[116] So far, photoBOLT has not been applied to ion channels or membrane receptors, mainly due to technical challenges arising from an amber suppression approach.

PTLs on the other hand, allow for a more general and practical approach towards tethered photopharmacology, ultimately facilitating *in vivo* studies.

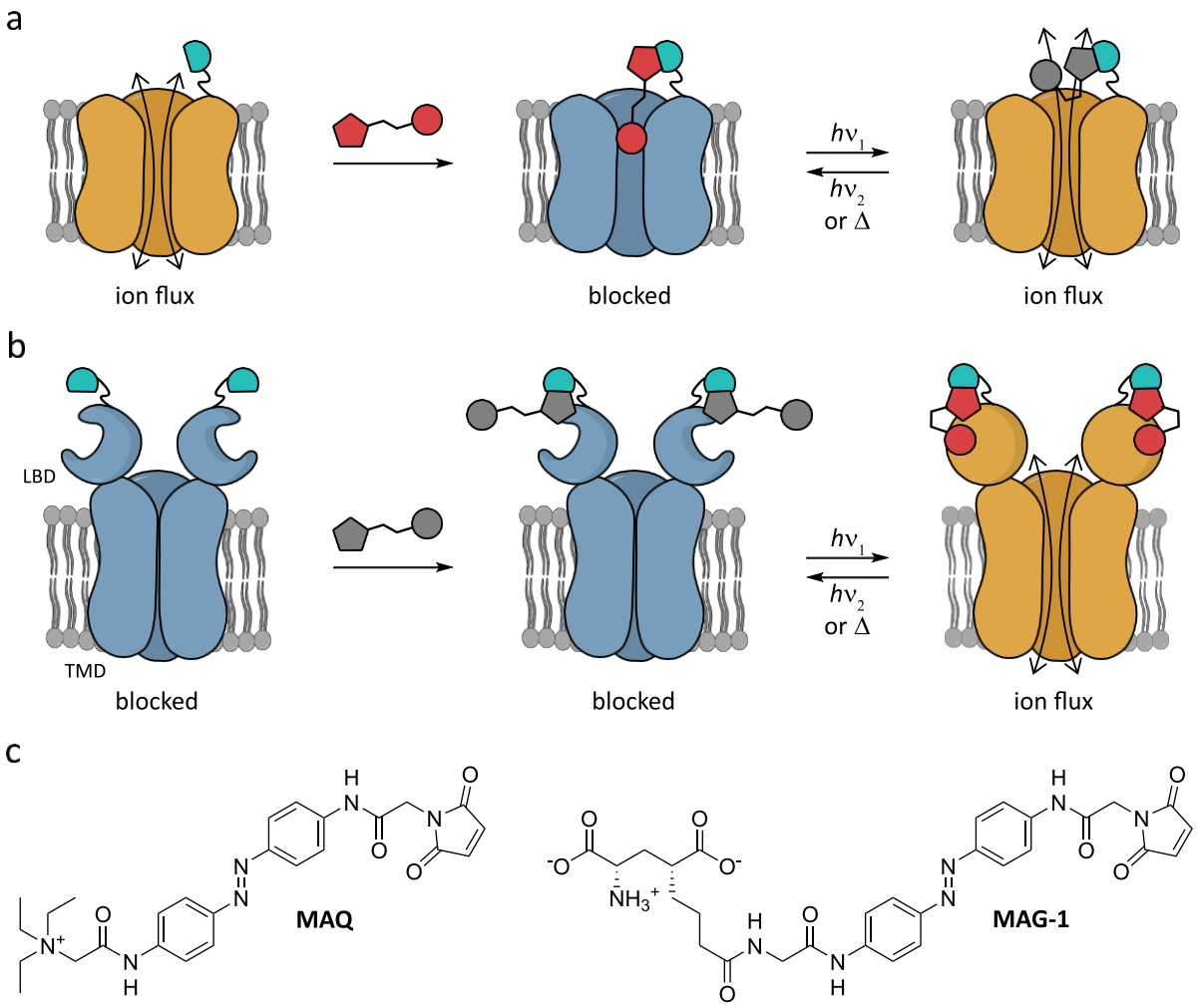


Figure 1.10 | Schematic depiction of the PTL approach. The photoswitchable ligand reacts with a bioconjugation handle (cyan) installed into the POI, which has either been introduced via genetic code expansion for photoBOLT or via site-directed mutagenesis for PTLs. (**a**) The tethered Kv channel blocker **MAQ** toggles between the active *trans*-form (red) and its inactive *cis*-form (grey), thus modulating ion flux across the membrane. (**b**) Tethering of the *trans*-inactive glutamate ligand **MAG-1** (grey) to the LBD of GPCRs allows for the activation of the receptor upon photoisomerization to the *cis*-isomer (red). LBD = ligand binding domain, TBD = transmembrane domain (**c**) Structural formulas of the Kv channel blocker **MAQ** and the glutamate PTL **MAG-1**.

Already prior to the development of PCLs based on quaternary ammonium ions targeting Kvs, a maleimide-functionalized PTL analog of **DENAQ** (Figure 1.9c), termed **MAQ** was developed. The PTL **MAQ** reacted with a cysteine residue in a multi-mutated Shaker-type Kv, yielding the light- and voltage-sensitive potassium channel SPARK (Figure 1.10a, c). **MAQ** enabled optical control over neuronal inhibition with high specificity for Shaker-type Kv that could not be surpassed by freely diffusible PCLs such as **DENAQ**, which affected a range of voltage-gated ion channels besides the Shaker-type Kv.[102, 117]

Similar to Kvs, iGluRs were amongst one of the first targets in photopharmacology due to their key role in fast synaptic transmission throughout the central nervous system. Tethering of the glutamate agonist **MAG-1** to the clamshell-type LBD of the receptor GluK2 gave the light-responsive, ligand-gated ion channel LiGluR (Figure 1.10b, c). Binding of the glutamate ligand in the *cis*-form of **MAG-1** led to full closure of the LBD, resulting in the allosteric activation of the TMD that forms the pore.

While proof-of principle studies focused on the electrophysiological evaluation of **MAG-1**,[118, 119] follow-up studies introduced LiGluR into neurons, providing robust control over action potentials in zebrafish larvae.[120] Glutamate is the dominant neurotransmitter found within the retinal network, thereby making the application of **MAG-1** particularly interesting for vision restoration. Expression of LiGluRs in retinal ganglion cells of the *rd1* mouse model restored both the pupillary reflex and natural light-avoidance behavior.[121] Clamshell-like LBDs are widespread across periplasmic binding proteins and consequently the strategy reliant on **MAGs** was also successfully applied to other receptor types such as the NMDA receptor, yielding LiGluNs.[122]

Whereas photoswitching of closely tethered ligands primarily increases the local concentration of the ligand, photoswitchable orthogonal remotely tethered ligands (PORTLs) mostly affect the efficacy of the ligand.[123] The bioconjugation motif of PORTLs is based on orthogonal, genetically encoded protein tags such as Halo, SNAP or CLIP (Figure 1.11a).

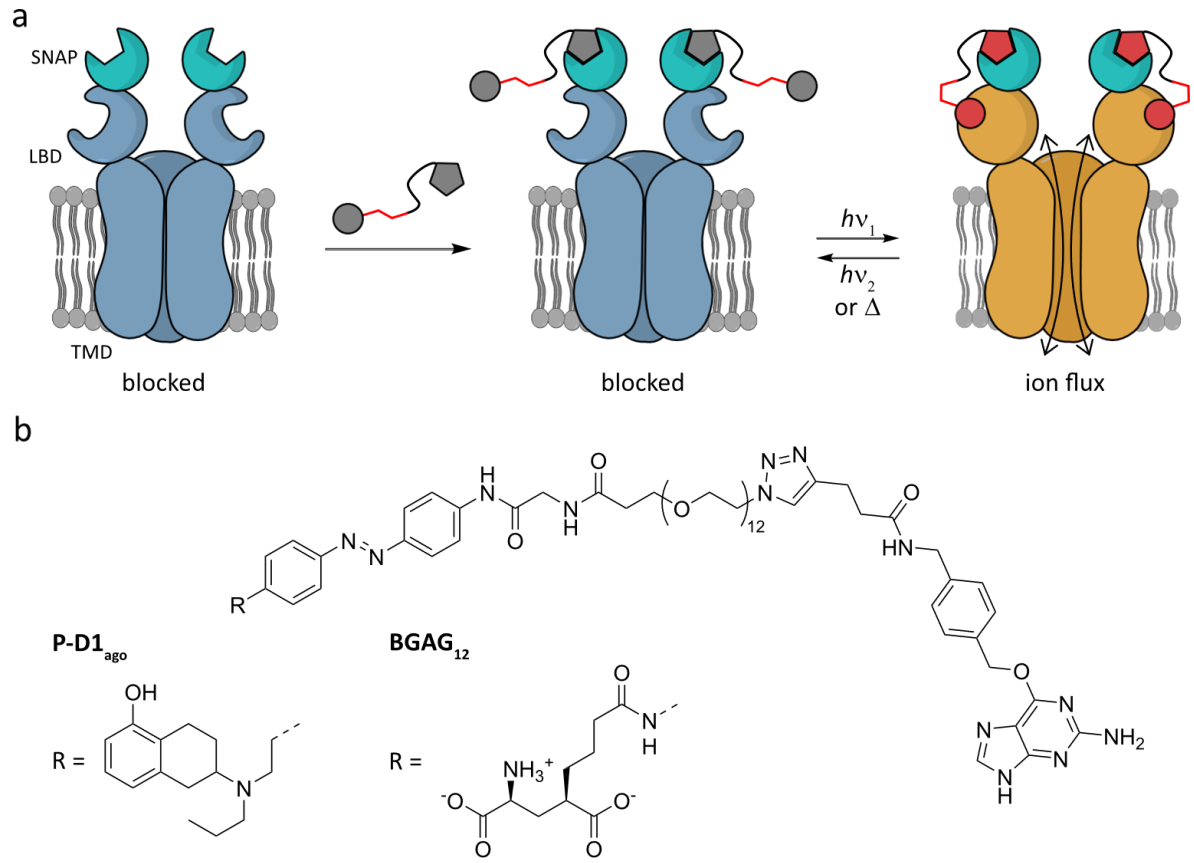


Figure 1.11 | Schematic depiction of the PORTL approach. (**a**) The photoswitchable ligands are tethered to the SNAP protein tag (cyan) via a benzylguanine moiety. The conjugated PORTL toggles between the inactive *trans*-form (grey) and its active *cis*-form (red), thus controlling the activity of GPCRs by light. LBD = ligand binding domain, TBD = transmembrane domain (**b**) Structural formulas of the glutamate PORTL **BGAG$_{12}$** and the PORTL **P-D1$_{ago}$** targeting dopamine receptors.

These self-labeling protein tags are tolerated by many receptors as fusion constructs and thus enable facile bioconjugation to a photoswitchable ligand. Introduction of the tethering site far away from the ligand binding site results in linkers that can be several nanometers long. Therefore, the azobenzene moiety has to be placed inside the pharmacophore to gain the desired effect over receptor activity. In contrast to PTLs, photoisomerization of PORTLs has an immediate effect on the affinity of the ligand.[114] The tethered photopharmacology approach PORTL has proven particularly successful for controlling GPCR signaling with molecular cell-type specificity. Broichhagen et al. first established PORTL on the retinal GPCR mGluR2 by *N*-terminally fusing it to the catalytic protein-tag SNAP.

Conjugation of the photoswitchable glutamate ligand **BGAG$_{12}$** to the SNAP domain was achieved via an orthogonal benzylguanine (BG) moiety, affording the light-responsive receptor SNAG-mGluR2 (Figure 1.11b).[124] Application of SNAG-mGluR2 further endowed a rapid light response onto retinal ganglion cells of blind *rd1* mice, thereby restoring patterned vision that is typically encountered with genetic disorders such as retinitis pigmentosa.[125] Recent improvements of this approach are based on branched PORTLs that contain between two to four photoswitchable glutamate ligands to further increase photoswitching efficiency. This also resulted in markedly enhanced optical control of mGluR3 that could not sufficiently be targeted with single-branched PORTLs.[126] Holt et al. tethered a four-branched photoswitchable glutamate ligand to a SNAP-mGluR2 fusion expressed in retinal ganglion cells via intravitreal adeno-associated virus (AAV) gene delivery, restoring pattern recognition in *rd1* mice with a similar acuity as reported for wt mouse vision.[127]

Beyond the optical modulation of mGluR receptors for vision restoration, PORTL has been successfully applied to serotonin GPCRs[128] and dopamine receptors (DARs).[129] The manipulation of dopamine signaling pathways is particularly challenging due to a diversity of neurons that are being regulated by dopamine as well as the complex localization of DARs. Moreover, dopamine can be co-released with the neurotransmitters glutamate and GABA, further complicating our understanding of how dopamine controls certain behavior such as movement, aversion or motivation. The recent development of a cell-specific, membrane-anchored agonist for the DAR subtype D1R, allowed Isacoff and co-workers to investigate the relationship between dopamine release and specific DAR activation. To this end, they expressed the SNAP-bearing membrane anchor for the photoswitchable ligand **P-D1$_{ago}$** in dopamine neurons of the dorsal striatum using AAV gene delivery in D1-Cre mice (Figure 1.11b). Optical control over locomotion was effectively demonstrated by photoisomerization of the thus locally and cell-type specifically restricted agonist **P-D1$_{ago}$**, indicating a direct link between D1R activation and enhanced movement.[129]

While the described examples are providing insight into how photopharmacology can be applied to probe cell signaling pathways *in vivo*, they are by far not all-encompassing and the reader is kindly referred to some excellent reviews giving a more detailed overview.[102, 114, 130]

By now, the repertoire of PCLs has vastly expanded beyond neuromodulators, ranging from photoswitchable hormones (**AzoGW**) to reversibly modulate transcription[131] over photoswitchable antibiotics (**8**) for minimized resistance development in bacteria[132, 133] to photoswitches targeting the cytoskeleton (**PST-1**, Figure 1.12).[134] Molecules like the colchicine domain microtubule inhibitor **PST-1** do not only exhibit great potential for targeted cancer therapy treatments,[135] but they prove valuable tools for probing the highly dynamic nature of the cytoskeleton.[136]

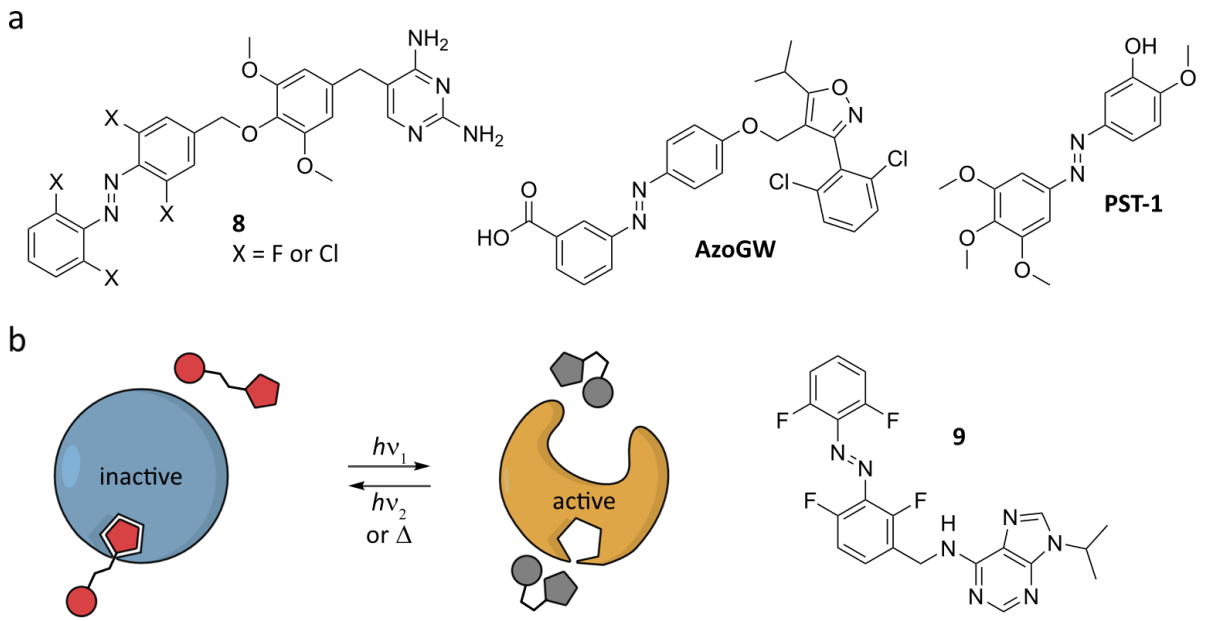


Figure 1.12 | Overview of small molecule inhibitors and modulators used in chemical biology. (**a**) Structural formulas of the photoswitchable antibiotic **8**, the azobenzene-based hormone **AzoGW** and the light-responsive combretastatin analog **PST-1**. (**b**) The azobenzene-based CKI inhibitor **9** toggles between the non-binding *cis*-isomer (grey) and the *trans*-isomer (red). Upon binding to the protein, CKI is rendered inactive (blue) only if **9** is present in the *trans*-form.

Feringa and co-workers recently developed a visible light-responsive inhibitor of casein kinase I (CKI) to control the period and phase of circadian rhythms in live cells and tissue, coining the term of chronophotopharmacology. The circadian rhythm is driven by the transcription of so-called *Period* (*Per*) and *Cryptochrome* (*Cry*) genes. Phosphorylation of the respective PER protein by CKI proteins typically leads to proteasomal degradation, keeping the circadian period at 24 h. Consequently, inhibition of CKI proteins leads to an elongation of the circadian cycle. The design of a drastically bistable photoswitchable inhibitor was crucial to allow for long-term modulation of the circadian period. To this end, Kolarski et al. designed the potent CKI inhibitor **9** based on *ortho*-fluorinated azobenzenes (Figure 1.12), facilitating period and phase regulation of circadian rhythms in U2OS cells.[137]

A novel technique which has only emerged during the last years are so-called PhotoPROTACs[138] or PHOTACs (PHOtochemically TArgeting Chimeras, Figure 1.13)[139], i.e. photoswitchable PROTACs (PROteiolysis TArgeting Chimeras). Through incorporation of an azobenzene moiety into the linker region of PROTACSs, a trifunctional molecule is generated where an azobenzene moiety bridges a ligand for an E3 ligase with a ligand for the target protein (Figure 1.13a). In general, PHOTACs show little to no proteolytic activity when in the dark-adapted *trans*-form. Photoisomerization to the *cis*-isomer facilitates binding to the E3 ligase and POI, thus promoting proteolysis of the POI in a light-dependent manner. The concept of activating proteasomal degradation with high temporal and spatial precision was successfully demonstrated using PHOTACs targeting bromodomains of BET proteins in acute lymphoblastic leukemia cells (Figure 1.13b). A new class of photoswitchable PROTACs makes use of the extremely bistable arylazopyrazole photoswitch.[140] Although these results are so far a mere proof of principle, future developments of PHOTACs might hold great potential as tools in cell biology and ultimately therapeutic approaches using (near-)infrared light.

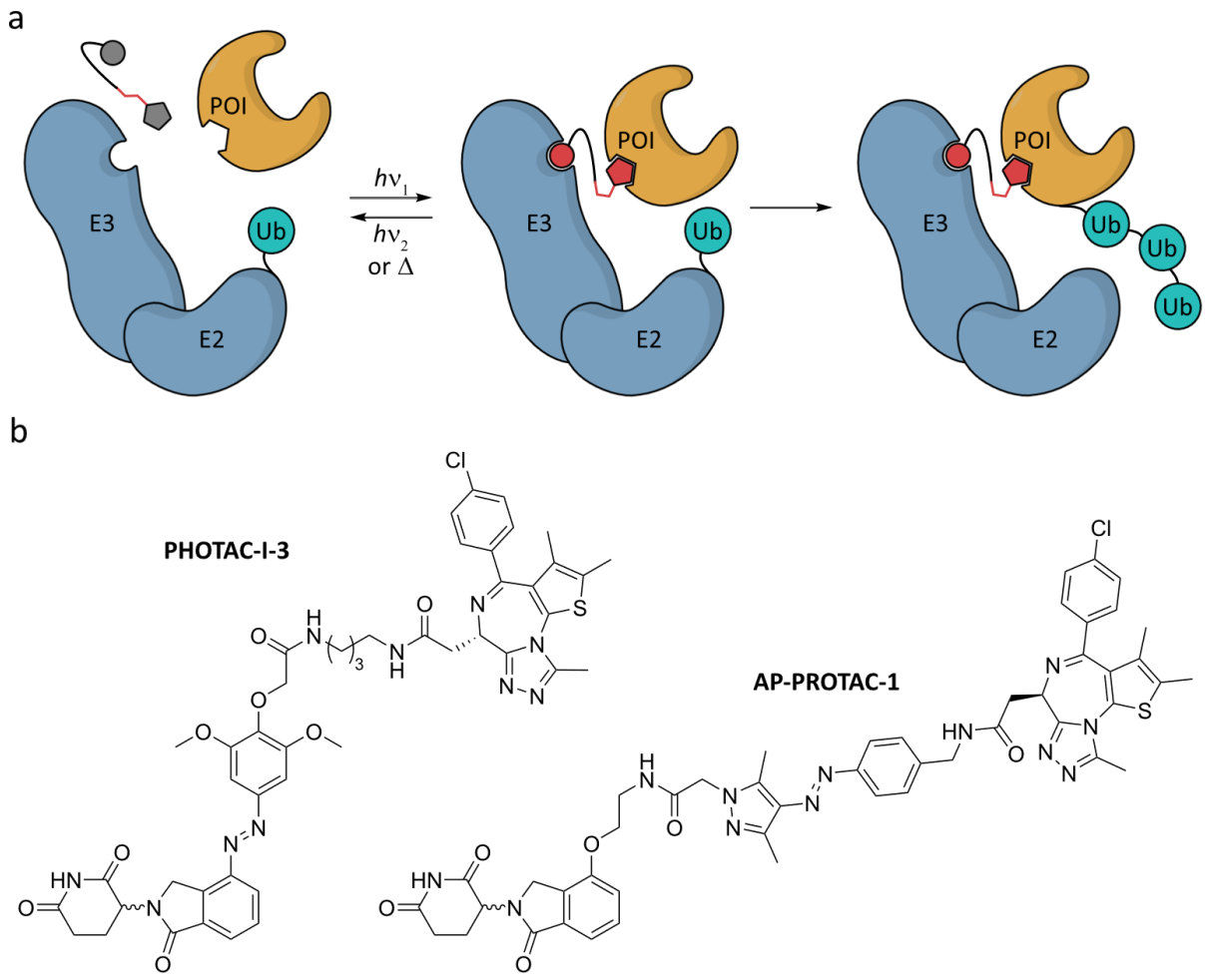


Figure 1.13 | Schematic concept of PHOTACs. (**a**) The molecule can be switched from its inactive *trans*-form (grey) to the active (red) *cis*-isomer, thus binding to the E3 ligase in complex with the E2 ligase (blue). Recruitment of the POI (orange) allows for targeted ubiquitylation (cyan) and subsequent proteasomal degradation of the POI. (**b**) Structural formulas of PHOTACs targeting bromodomains.

1.3.3 Azobenzenes to Study Peptide and Protein Conformation

Beyond the application of freely diffusible PCLs or tethered photoswitchable ligands (PTL and PORTL), protein function can be modulated with spatio-temporal precision by directly installing the azobenzene into secondary structural elements as a conformational switch.

The introduction of azobenzene moieties into the main chain of peptides (**11**, Figure 1.14a) strongly mimics ß-turns when isomerized to the *cis*-form. This allowed for light-regulated studies of ß-hairpin formation within tryptophan zipper motifs.[141] Through insertion of azobenzene **11** into polypeptide backbones as a ß-turn inducing segment, this approach could even be extended to the larger helical avian pancreatic polypeptide (aPP). Here, Hilvert and co-workers successfully demonstrated the power of azobenzenes as conformational switches by toggling between two different tertiary structures of aPP.[142] Similarly, Broichhagen et al. replaced the amino acid residues G22-Q23 between two α-helices of the incretin mimetic liraglutide with azobenzene **11**, furnishing the glucagon-like peptide-1 receptor (GLP-1R) agonist LirAzo to manipulate GLP-1R signaling and thus insulin secretion in pancreatic beta cells.[143]

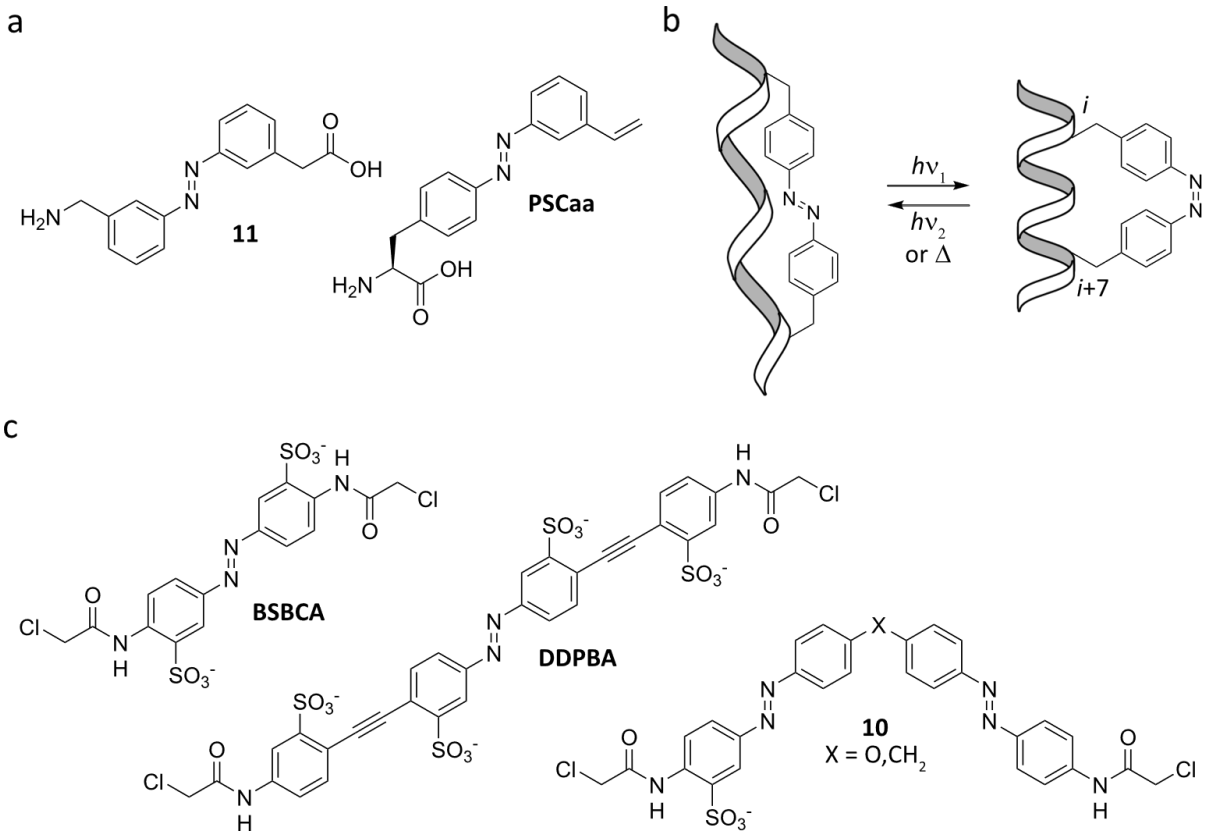


Figure 1.14 | Azobenzene-based ncAAs and crosslinkers for optical control of the secondary structure elements α-helix and β-turn. (**a**) Examples of azobenzenes used as mimics of β-turns (**11**) or as bridges in α-helices (**PSCaa**) via SPPS. (**b**) Schematic representation of a crosslinked azobenzene (e.g. **BSBCA**) and reversible switching of the formed azo-bridge to modulate α-helical content in a short peptide. (**c**) Overview of some water-soluble azo-crosslinkers, including the broadly used **BSBCA**.

Studies on conformational folding and unfolding processes of α-helices are the most prominent, however and typically rely on the introduction of azobenzene-based ncAAs (azo-ncAAs) via SPPS or on azobenzene-based crosslinkers (azo-crosslinkers). The photoswitchable side chain amino acid **PSCaa** (Figure 1.14a) was introduced into model peptides using SPPS and subsequently reacted with a proximal cysteine residue via thiol-ene click reaction, thereby installing a photocontrollable bridge.[144] Recent work has predominantly focused on the modulation of model coiled-coil motifs using single side-chain isomerization.[145, 146]

The alternative, cysteine-reactive azo-crosslinkers, heavily pioneered by Woolley and co-workers, have found much broader application when it comes to reversibly controlling α-helical content in peptides or ideally proteins. To this end, the azo-crosslinker is installed on proteins bearing two surface-exposed cysteine amino acid residues. Synthetic crosslinks in general tend to greatly stabilize α-helical secondary structures or even induce helicity in peptides, which are normally disordered in water, a phenomenon which is described through the term “stapled peptides”. An azo-crosslink on the other hand, destabilizes the helix relative to the non-crosslinked peptide, ideally when in the *trans*-form (Figure 1.14b). One example thereof is the broadly used, water-soluble azo-crosslinker **BSBCA** that is typically applied to cysteine residues spaced by *i,i*+7 (Figure 1.14c).[147] Upon photoisomerization to the *cis*-isomer, the destabilization is reversed since the amino acid residues at *i,i*+3 and *i,i*+4 are much more easily accommodated, finding enough space between the α-helix and azo-crosslinker (Figure 1.14b).[148]

The effect of photoswitching that can be observed for intramolecularly installed azobenzene-based bridges in α-helices is very much dependent on the change in end-to-end distance of the designated crosslinker and on the spacing of the two cysteine residues, i.e. *i*,*i*+4, *i*,*i*+7 or *i*,*i*+11.[149] Considerable efforts have therefore gone into increasing the end-to-end distance of azo-crosslinkers in order to maximize conformational change in the target peptide. Woolley and co-workers enlarged a water-soluble switch with rigid phenylethynyl units (**DDPBA**, Figure 1.14c), while at the same time avoiding the introduction of additional degrees of rotational freedom. Though the end-to-end distance was successfully increased from 4 Å (**BSBCA**) to 13 Å (**DDPBA**), this compound was suffering from a comparably low PSS, thus limiting its use.[150] An even larger system featuring a high PSS in both directions was recently introduced by Szymanski and co-workers. This crosslinker is based on a water-soluble bisazobenzene scaffold (**10**, Figure 1.14c) and exhibits an overall change in end-to-end distance of 23 Å.[151]

Photoswitchable crosslinkers have been broadly used to control enzymatic activities[152, 153] and helical coiled-coil motifs found in transcription factors[154]. Zhang et al. were the first ones to install an azo-crosslink into the dominant negative peptide Fos of the transcription factor AP-1, rendering the *trans*-state disordered. The helicity of Fos increased upon isomerization to the *cis*-form so that a coiled-coil structure was formed between Fos and its interacting peptide Jun. Ultimately, this led to inhibition of DNA binding, which was demonstrated in live HEK cells by expression of green fluorescent protein (GFP) under the control of an AP-1 dependent promotor.[155] Photoregulation using azobenzene crosslinkers has furthermore been employed to gain spatio-temporal control over the affibody-interactor Z-domain of protein A[156] and the allosteric protein domain PDZ[157, 158].

Still, in contrast to peptides and small proteins, conformational changes are much more difficult to predict and interpret within large and more complex proteins. In addition to rational design of photosensitive proteins, computational selection methods such as molecular dynamics (MD) and Rosetta-based simulations or structure-activity relationship (SAR) should be considered.[153, 159]

Despite the described advances, the installation of azo-ncAAs directly into proteins remains challenging. Approaches using SPPS methods for the *de novo* preparation of terminal protein fragments have been successfully applied for more complex proteins.[160] However, such strategies are tedious and might not be applicable to every POI. Similar considerations apply to azo-crosslinkers. Though they have revolutionized the field of optically controlling α-helical content in peptides and small proteins, crosslinking procedures are time-consuming and generally require reducing agents and elevated temperatures. Depending on the accessibility of the cysteine residues, often a heterogenous mixture of modified proteins is obtained, including intermolecular or mono-crosslinks.

Therefore, a technique which allows for the efficient and specific generation of homogeneous azobenzene-protein conjugates would be of great value. Genetic code expansion tackles these challenges in an elegant way by enabling the installation of small light-sensitive molecules already during protein biosynthesis into almost any region of the target protein.

1.4 An Expanded Genetic Code

1.4.1 Ribosomal Protein Translation

Protein biosynthesis is the precise translation of genetic information transcribed as messenger RNA (mRNA) into functional proteins with an astonishing accuracy of more than 99%.[161] The execution of this process is realized by a complex translational machinery and its key component, the ribosome, made from both ribosomal RNA (rRNA) and over 50 proteins (Figure 1.15).

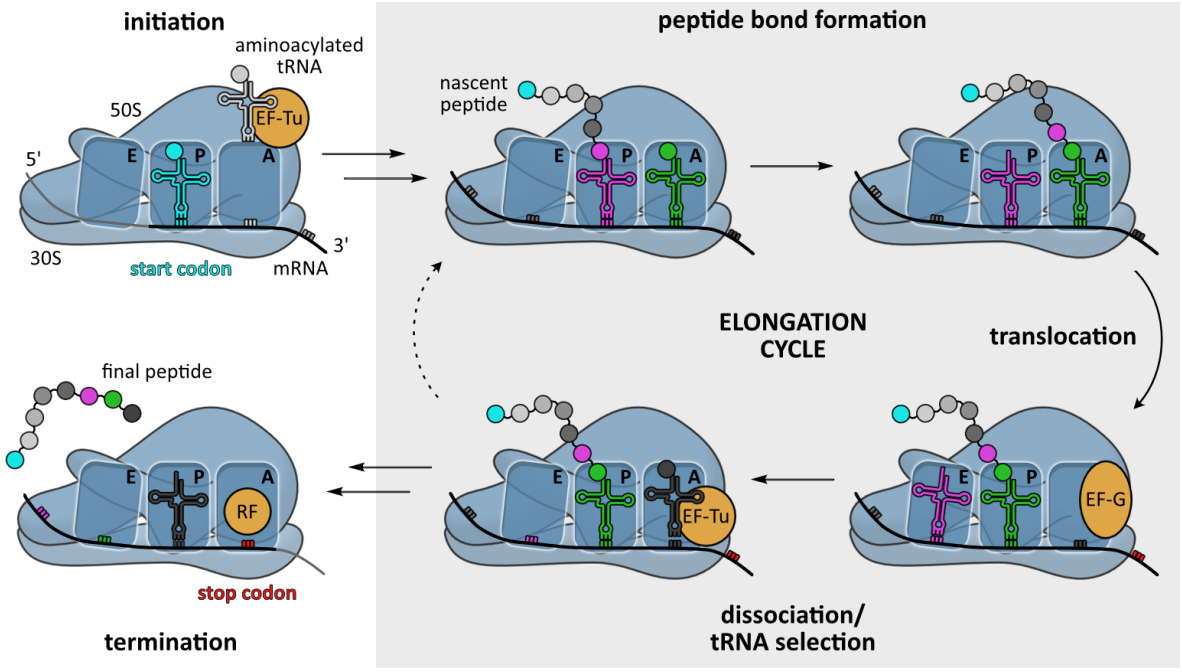


Figure 1.15 | Schematic overview of prokaryotic ribosomal translation. Initiation begins with the recruitment of the special initiator fMet-tRNA^{fMet} (cyan) to the P-site in the ribosomal 30S unit. Recruitment of the second aminoacylated tRNA via EF-Tu starts the elongation cycle (grey box). During elongation, a matching aminoacylated tRNA (green) is placed in the A-site next to the tRNA (purple) bearing the nascent peptide chain in the P-site. Subsequent peptide bond formation in the ribosomal 50S unit transfers the peptide chain to the A-site. Translocation of the ribosome is performed through binding of EF-G, thereby placing the now uncharged tRNA into the E-site and the peptide-bearing tRNA into the P-site. Lastly, the uncharged tRNA dissociates from the ribosomal complex and a new aminoacylated tRNA is guided to the A-site by EF-Tu, completing the elongation cycle (grey). As soon as a stop codon (red) is encountered, the recruitment of RF will lead to termination. The peptide chain is transferred to a water molecule and released from the ribosome (70S).[162]

The ribosome comprises two subunits, a large ribosomal subunit (LSU) needed for peptide bond formation and a small ribosomal subunit (SSU) which interprets the mRNA codons. The mRNA serves as the template for protein synthesis and is composed of triplet codons. These triplet codons have to be non-overlapping, in order for a combination of the four nucleobases (A, C, G, U) to be efficiently decoded. The decoding center (DCC) within the SSU then interprets each triplet codon, which corresponds to a specific amino acid. Prior to translation by the ribosome, the amino acids have to be covalently attached to the 3' end of so-called aminoacyl-transfer RNAs (tRNAs). This process, also called "charging" of aminoacyl-tRNAs, is catalyzed by a cognate and highly specific aminoacyl-tRNA synthetase (aaRS). Now, the matching substrate can get recruited to the peptidyl transferase center (PTC) within the LSU, meaning the aminoacylated tRNA bearing the complementary sequence (anticodon) in its stem loop. There are 64 possible triplet codons, out of which 61 encode the 20 proteinogenic amino acids. Thus, multiple codon combinations can code for a specific amino acid, rendering the genetic code redundant or degenerate. The residual three codons terminate translation and are thus termed stop codons. [163]

Bacterial protein translation is initiated by base-pairing of the rRNA in the 16S component of the SSU (30S) to a complementary sequence, called the Shine-Dalgarno sequence, which is found upstream of the start codon AUG (Figure 1.15). Then, the special initiator fMet-tRNA^{fMet}, a tRNA bearing the amino acid *N*-formylmethionine (fMet), binds to the AUG codon. This is accompanied by binding of the LSU (50S) and thus the mature ribosome (70S) is formed. With the fMet-tRNA^{fMet} having been positioned in the peptidyl site (P-site) of the ribosome, the SSU decodes the first triplet codon.

The correct aminoacyl-tRNA is thus recruited and guided to the aminoacyl site (A-site) of the PTC by the elongation factor (EF) Tu. Subsequent reaction between the amine group of the aminoacyl-tRNA and the activated carbonyl group of the fMet-tRNA^{fMet} within the P-site results in the formation of an amide bond and the transfer of the nascent peptide chain to the A-site tRNA. After EF-G catalyzed translocation of the ribosome, the now uncharged tRNA^{fMet} molecule is positioned in the exit site (E-site) and the peptidyl-bearing tRNA in the P-site. As a result, the next triplet codon on the mRNA becomes available for decoding. Binding of a new aminoacyl-tRNA to the A-site is concomitant with the dissociation of the uncharged tRNA^{fMet} from the ribosomal complex. This elongation cycle of decoding, peptide bond formation and translocation gets repeated until a so-called stop codon is encountered in the A-site, leading to the recruitment of release factor (RF) 1 or 2, and hence dissociation of the final peptide chain from the ribosome.[164]

1.4.2 Amber Codon Suppression

The three stop codons, termed “amber” (UAG), “opal” (UGA) and “ochre” (UAA), have been repurposed to push the boundaries imposed by the limited diversity of functionalities within the 20 canonical amino acids. Contrary to auxotrophic *E. coli* strains used for the residue-specific incorporation of non-canonical amino acids (ncAA), the recoding of stop codons aims for a single, site-specific ncAA incorporation. While residue-specific incorporation replaces e.g. all methionine residues for a surrogate such as homoallylglycine throughout the proteome[165], reassignment of the amber codon was much more promising, since UAG is (i) a blank codon and (ii) with 7.4% the rarest of all three stop codons in *E. coli*.[166]

Pioneered by Schultz and co-workers[167], the following exogenous parts are needed for “amber (codon) suppression” (Figure 1.16): i) a non-toxic ncAA of choice, ii) a gene of interest bearing the blank codon UAG at a user-defined position and iii) an orthogonal aaRS/tRNA$_{CUA}$ pair. The latter is of great importance, since the aaRS must not accept any of the canonical amino acids or endogenous tRNAs and vice versa. First, the ncAA supplied to the growth medium gets charged onto a complementary tRNA$_{CUA}$ by an orthogonal aaRS, evolved to accept the desired ncAA but none of the 20 canonical amino acids. Once the ribosome encounters an amber codon during translation, the ncAA gets incorporated into the nascent peptide chain, as long as cell-permeability and charging efficiencies are high and competition with RF1 is negligible. Usually, the full-length protein can then be easily purified via a C-terminal tag.[168]

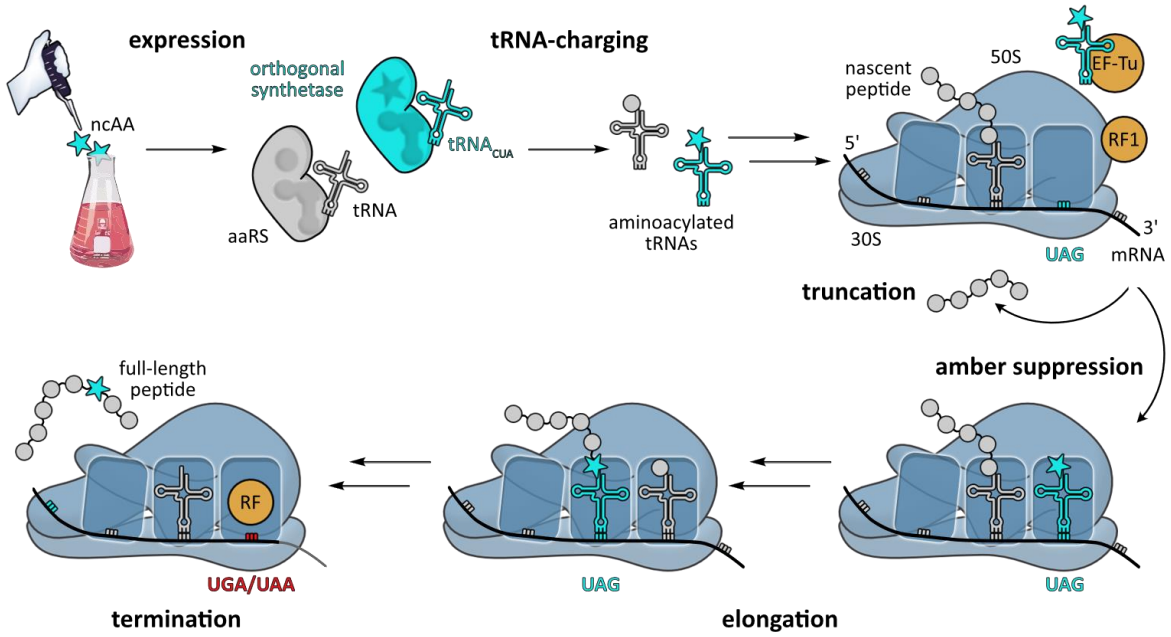


Figure 1.16 | Schematic overview of the amber suppression approach. After addition of the ncAA (cyan star) to the bacterial culture, expression of the orthogonal aaRS/ $tRNA_{CUA}$ pair (cyan) and the POI is induced. Once the tRNA has been successfully charged by the aaRS, EF-Tu guides the aminoacylated $tRNA_{CUA}$ to the ribosome in response to an amber stop codon (UAG). During this step, competition with RF1 will also yield truncated protein (grey). If amber suppression was successful, the ncAA is site-specifically incorporated into the nascent peptide chain until a UGA or UAA stop codon (red) trigger termination by RF (1 or 2).

1.4.3 Orthogonal Synthetase/Aminoacyl-tRNA Pairs

An orthogonal aaRS/$tRNA_{CUA}$ pair is crucial for the successful incorporation of structurally diverse ncAAs using an expanded genetic code in bacteria and eukaryotes. A common strategy for achieving a broad substrate scope while maintaining orthogonality is the engineering of naturally occurring, orthogonal aaRS/tRNA pairs from evolutionary distant organisms. What made amber suppression in particular so broadly applicable is the fact that some archaea had already reassigned the UAG codon for the incorporation of amino acids. The most common amongst them are the archaebacterium *Methanococcus jannaschii (Mj)* incorporating tyrosine[169] and the methanogenic archaeons *Metharnosarcina barkeri (Mb)*[170] and *Methanosarcina mazei (Mm)*,[171] both incorporating the 22nd proteinogenic amino acid pyrrolysine (Pyl) in response to the UAG codon. Unlike the *Mj*TyrRS/*Mj*tRNA^{Tyr} pair, the endogenous activity of PylRS towards a natural amino acid did not have to be deleted for the *Mb*PylRS/*Mb*tRNA^{Pyl} and *Mm*PylRS/*Mm*tRNA^{Pyl} pairs. Further, the PylRS/tRNA^{Pyl} pairs from *Mb* and *Mm* turned out to be orthogonal with respect to the endogenous synthetases and tRNAs in both bacteria and mammalian cells, while the *Mj*TyrRS/*Mj*tRNA^{Tyr} pair is orthogonal solely in bacteria (Figure 1.17a). As a result, the *Mb*PylRS/*Mb*tRNA^{Pyl} and *Mm*PylRS/*Mm*tRNA^{Pyl} pairs have quickly become the most widely used aaRS/tRNA systems in the field of genetic code expansion.

Over the years, several other aaRS/tRNA systems from different organisms have been discovered, such as the SepRS/phosphoseryl-tRNA pair from *Methanococcus maripaludis* (*Mmp*). Evolved *Mmp*SepRS/*Mj*tRNA^{Sep} pairs facilitated the incorporation of phosphorylated serine[172] and threonine[173], as well as their non-hydrolysable analogs[174] into proteins, first in *E. coli* and then also in mammalian cells[175].

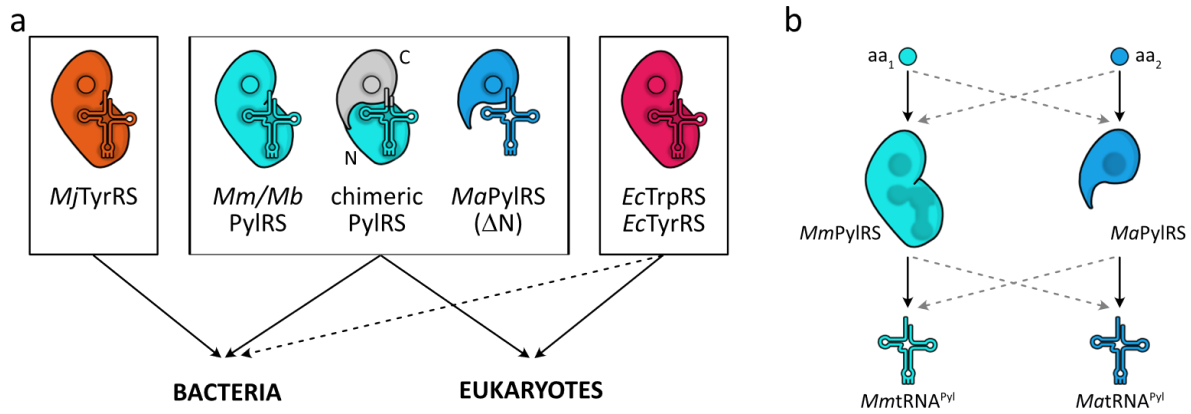


Figure 1.17 | Overview of main orthogonal aaRS/tRNA pairs used to expand the genetic code. (**a**) Most common aaRS/tRNA pairs and their application to eukaryotes and/or bacteria. Solid arrows indicate orthogonality within the respective organism, while a dashed arrow indicates orthogonality only under certain circumstances (e.g. only in specific *E. coli* strains). (**b**) Mutual orthogonality scheme for the *Mm* and *Ma* PylRS/tRNA^{Pyl} system. Solid arrows are needed for successful genetic code expansion. If the dashed arrows would apply, this would render the depicted system no longer mutually orthogonal within the same host.

Although initially used less frequently in favor of the versatile PylRS/tRNA^{Pyl} pairs, the *E. coli* (*Ec*) derived pairs for tyrosine (*Ec*TyrRS/*Ec*tRNA^{Tyr}) and tryptophane (*Ec*TrpRS/*Ec*tRNA^{Trp}) have become quite popular (Figure 1.17a)[176], especially after Chatterjee and co-workers further extended the application of the latter from yeast and mammalian cells to *E. coli.* Replacement of the endogenous TrpRS/tryptophanyl-tRNA pair with a counterpart *Sc*TrpRS/*Sc*tRNA^{Trp} from *Saccharomyces cerevisiae (Sc)* liberated the *Ec*TrpRS/*Ec*tRNA^{Trp} pair. This allowed for the straight-forward, directed evolution of *Ec*TrpRS in *E. coli* which can then be used for genetic code expansion in mammalian cells.[177] Similarly, the *Ec*TyrRS/*Ec*tRNA^{Tyr} pair was also liberated for evolution approaches in *E. coli* only a year later.[178]

Typical PylRS are comprised of a catalytically active C-terminal domain and an N-terminal domain, the latter being responsible for binding the T-arm and variable loop of its cognate tRNA.[179] The discovery of single-domain PylRS (ΔNPylRS) analogs lacking the N-terminal domain, e.g. from the human gut archaeon *Methanomethylophilus alvus (Ma)* (Figure 1.17a), provided a basis for the generation of mutually orthogonal PylRS/tRNA pairs in *E. coli* and mammalian cells (Figure 1.17b). Whereas the deletion of the N-terminal domain from *Mm*PylRS completely abolishes enzymatic activity, the ΔNPylRS analogs are exceptionally active *in vivo*.[180, 181] Recently, Dunkelmann et al. reported on the incorporation of three distinct amino acids using two mutually orthogonal ΔNPylRS/tRNA pairs and a new orthogonal *Mm*PylRS/*Spe*tRNA^{Pyl} pair in mammalian cells. Through extensive engineering of several new ΔNPylRS/tRNA pairs and the identification of *Spe*tRNA^{Pyl} from *Methanosarcina spelaei* functioning with the *Mm*PylRS with similar activity to the native *Mm*PylRS/*Mm*tRNA^{Pyl} pair, a triply orthogonal set could be created.[182]

Over the last decade, new orthogonal aaRS/tRNA systems have been continuously discovered and evolved to expand on the functionalities that can be incorporated into proteins. Beyond that, there are also approaches that focus on controlling gene expression to further decrease the level of background misincorporation encountered with tRNA mischarging. This is particularly crucial for amber suppression used in biocontainment strategies or evolutionary approaches based on the expression of toxic genes.

So-called split aaRSs, where C-terminal and N-terminal fragments of a split PylRS variant are fused to interacting polypeptides, can be applied to mitigate these limitations. Herein, the desired enzymatic activity of the aaRSs is solely restored by induction through protein-protein interactions or the exogenous addition of small molecules. This allows for the tight control of gene expression at the translational level, resulting in faster response times in comparison to methods employing transcriptional control.[183]

Still, other aaRS/tRNA pairs featuring high efficiency and orthogonality in *E. coli*, yeast and mammalian cells, would be of great value. One recent approach exploits the broad orthogonality of the PylRS/$tRNA^{Pyl}$ system and combines it with unique structural features of canonical aaRS/tRNA pairs, thereby generating chimeric aaRS/chimeric tRNA pairs. The chimeric aaRSs consist of the PylRS N-terminal domain fused to the C-terminal domain of endogenous aaRSs, thus exhibiting higher aminoacylation activities as well as improved solubility compared to the wild-type PylRS. Subsequent directed evolution furnished orthogonal chimeric histidine, phenylalanine and tryptophan pairs, which accepted a broad spectrum of aromatic amino acid derivatives.[184] By using such an evolved chimeric Pyl/Phe translational system[184], Zhao et al. did not only incorporate ncAAs with wild-type like efficiency, they also succeeded in developing another ncAA-dependent synthetic auxotroph, namely an organism that only grows when exogenously supplied with the synthetic amino acid.[185] Such strategies are highly sought after in order to (i) contain genetically modified organisms (GMOs)[186, 187] and (ii) expand on live attenuated virus vaccines[188], which would trigger an immune response without simultaneous infection of the host.

1.4.4 Directed Evolution of Novel Synthetases

The standardized directed evolution approach to obtain orthogonal aaRS/$tRNA^{CUA}$ pairs accepting novel ncAAs as substrates, is based on a two-step selection protocol in *E. coli*. First, a library containing between 10^7 to maximum of 10^{12} aaRS mutants is created. Typically, this is done via PCR, using so-called NNK primers (N = A,C,G,T and K = G,T) for randomization of the binding pocket. The created library containing the aaRS mutants is then co-transformed with a positive selection marker, such as the *chloramphenicol acetyltransferase* resistance gene bearing an amber codon at a non-essential position. Cell growth in the presence of chloramphenicol singles out mutants which either incorporated the desired ncAA or a proteinogenic amino acid. Early termination of translation on the other hand, will lead to cell death (Figure 1.18a). In a second negative selection step, obtained false positives are eliminated. Here, aaRS mutants from the first selection round are screened for their ability to facilitate the incorporation of any proteinogenic amino acid into a toxic gene. Typically, a gene encoding for the ribonuclease barnase bearing two amber codons is used as a negative selection marker. Contrary to the positive selection step, cells are now grown in absence of ncAA, so that only mutants incorporating any of the canonical amino acids will express the toxic gene.

Hence, aaRS variants, that exclusively incorporate the desired ncAA in response to the amber stop codon, are being selected (Figure 1.18a).[189] This two-step selection protocol is usually repeated several times to ensure that a aaRS/$tRNA^{CUA}$ pair with high specificity towards the desired ncAA has been evolved. Unfortunately, such standard evolutionary approaches are tedious, extremely time-consuming in typically furnish aaRS mutants with a greatly reduced activity. Thus, it comes as no surprise that complementary technologies have been developed to speed up the process of evolving highly efficient and orthogonal aminoacyl-tRNA/aaRS pairs.

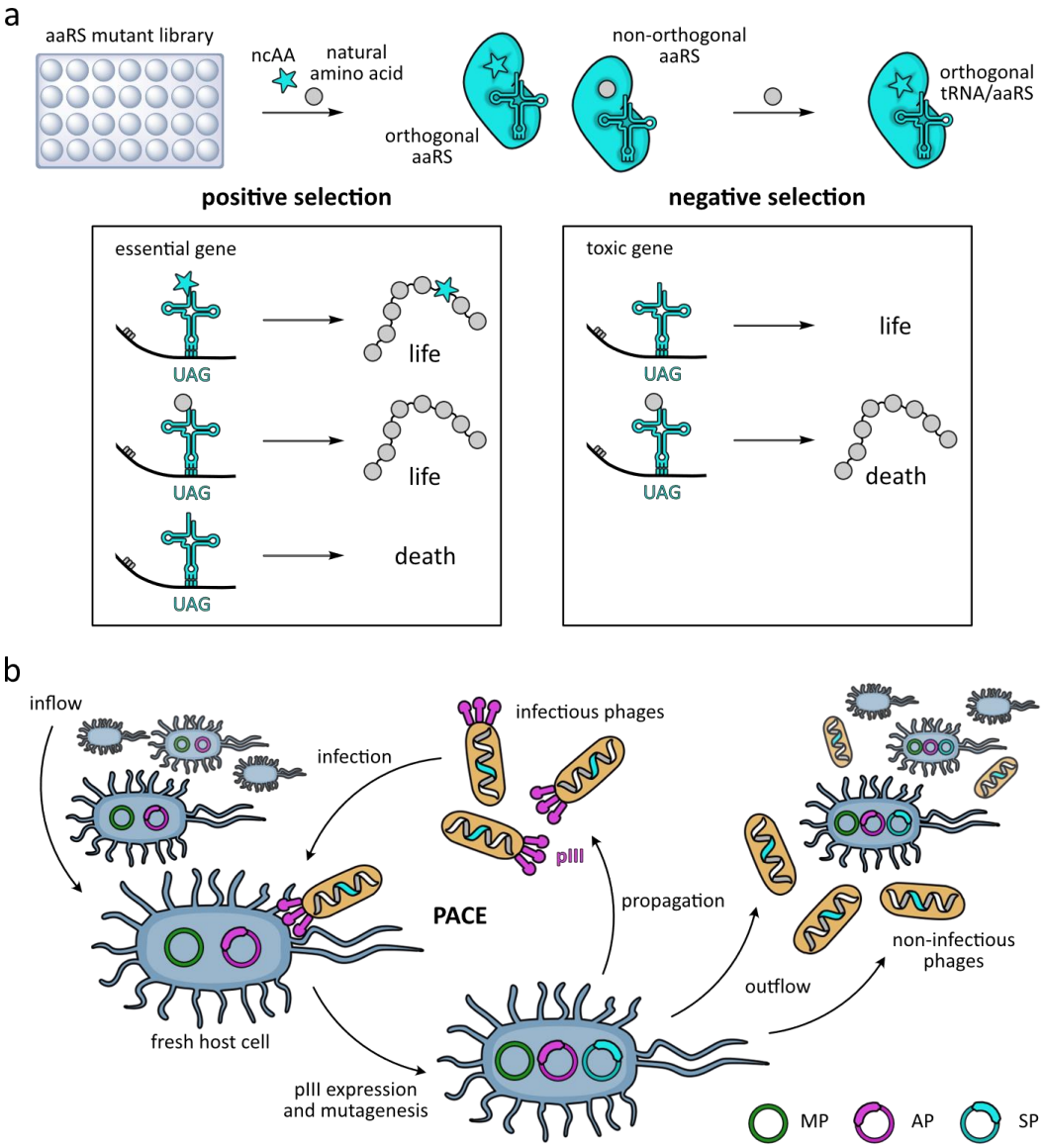


Figure 1.18 | Schematic depiction of directed evolution approaches for aaRS mutants. (**a**) Schematic depiction of the traditional directed evolution approach. In a first, positive selection step, aaRS mutants which incorporate either the ncAA (cyan star) or any proteinogenic amino acid (grey ball) in response to an amber stop codon introduced into an essential gene, survive. Subsequent negative selection by incorporation into a toxic gene bearing amber codons results in orthogonal aaRS mutants, which only accept the desired ncAA. (**b**) Phage-assisted directed evolution (PACE) for aaRS mutants by direct ncAA incorporation into pIII. The mutagenesis plasmid (MP, green) increases the mutagenesis rate during DNA replication while the accessory plasmid (AP, purple) encodes the aminoacyl-tRNA as well as gene III bearing an amber stop codon. Infection of fresh host cells bearing both plasmids leads to the expression of aaRS mutants from the selection phage (SP, cyan). A continuous in- and outflow regulates optimal phage propagation and minimal host replication during PACE.

One strategy tackling the lack of efficiency in traditional evolution of aaRS variants, is the development of phage-assisted continuous evolution (PACE)[190] and its application to aminoacylation activity, introduced by Liu and co-workers in 2017.[191] This technology increases the number of selection rounds from only a couple to hundreds, generating highly evolved aaRS variants within a short period of time.

In order to make phage infectivity dependent on tRNA aminoacylation, two different strategies based on the expression of protein pIII (encoded by gene *gIII*) were pursued. Either, the amber codon was directly placed into *gIII* (Figure 1.18b) or it was positioned in a *T7 RNA polymerase* (T7 RNAP) gene (encoded on an accessory plasmid), which allowed for the translation of *gene III* through an upstream T7 promotor. Placing the amber codon into the *T7 RNAP* gene is a less stringent selection scheme, since successful amber suppression will lead to the expression of several T7 RNAP copies. Still, both strategies rely on the successful expression of the protein pIII through the selection phage (SP) which bears the evolving aaRS mutant gene instead of *gIII*. This way, only the phages, which have infected the host cell with the gene encoding for a highly active and efficient aaRS mutant, are able to propagate. Mutagenesis of potential aaRS gene candidates is driven by an additional mutagenesis plasmid (MP), increasing the mutation rate on a genetic level upon arabinose induction during PACE.

Significant improvements, especially regarding time and effort, have been achieved by further coupling fluorescence-activated cell sorting (FACS) to the selection protocol. Early versions of this approach have relied on the incorporation of bioorthogonal handles into the *E. coli* outer membrane protein OmpC. Subsequent labeling with a fluorescent tag allowed to screen for suitable aaRS/tRNA pairs via FACS. Still, this approach was very much limited to ncAAs bearing a reactive moiety for specific labeling with fluorophores.[192]

Building on these early results, Lin and co-workers instead co-transformed a library containing aaRS mutants with a fluorescent reporter gene bearing an amber codon, e.g. the GFP.[193] This allowed for identification of functional aaRS/tRNA pairs featuring high suppression efficiencies within short times via FACS sorting. In addition, a similar approach led to the identification of a polyspecific PylRS variant (HpRS)[194] capable of incorporating 31 structurally diverse ncAAs.

Those recent advances have managed to significantly speed up and improve the evolutionary process for finding novel synthetases, which catalyze the aminoacylation of a broad range of ncAAs. However, efforts to improve the site-specific installation of new functionalities into proteins are by far not limited to the evolution of aaRSs.

1.4.5 Recent Efforts Within the Field

Low yields of ncAA incorporation encountered by employing amber suppression approaches can be attributed to multiple factors such as i) inefficient aaRSs (Chapter 1.4.4), ii) low intracellular concentrations of the suppressor-tRNAs or iii) the competition of *in vivo* stop codon suppression with release factors.

Particularly the incorporation of multiple ncAAs in response to multiple amber codons has been fairly limited, also due to the fact that yields are very much dependent on the suppression site within the target protein. Tools such as iPASS (identification of Permissive Amber Sites for Suppression) can contribute to the improvement of incorporation efficiencies by predicting suitable amber suppression positions within a target protein *in silico*. Bartoschek et al. computed data obtained from a proteomics-based approach that gave insight into what sequence context up- and downstream of UAG affects ncAA incorporation efficiencies in mammalian cells. The resulting iPASS predictions were then verified using a dual-fluorescence reporter in flow-cytometry by amber suppression of histones H2A and H3 in mammalian cells.[195]

Complementary to computationally aided strategies, the following components and factors can be targeted for improving the genetic code expansion approach towards the biosynthesis of proteins featuring one or multiple new functions: i) the aminoacyl-tRNA, (ii) the competition with release factors, iii) the ribosome itself or iv) using flexizymes (Fxs) for *in vitro* translation, thereby circumventing the need for orthogonal aaRS/tRNA pairs altogether.

Recent advances on improving overall suppression efficiencies by targeting the tRNA component, have focused on the development of engineered tRNAs or the exogenous application of chemically acylated tRNAs.[196, 197] Serfling et al. screened a small set of rationally designed tRNAs based on the most versatile and widespread archaeal $tRNA^{Pyl}$. Substitutions of single nucleotides and the installation of a canonical hinge between the D- and T-loop similar to human tRNAs, aimed at improving the recognition by the endogenous mammalian system. One such tRNA ($tRNA^{M15}$) featured higher intercellular concentrations and worked well together with *Mb*PylRS variants, thereby generating a very efficient PylRS system to simultaneously incorporate two distinct ncAAs into a GPCR for the first time.[196]

Tackling the competition of *in vivo* stop codon suppression with release factors (RFs) on the other hand, poses a great challenge within the field of genetic code expansion. In order to increase suppression efficiencies, the translational machinery has to clearly favor ncAA incorporation over termination of the protein through RFs. In *E. coli* this particularly concerns RF1, which recognizes the least frequently used amber (UAG) codon and ochre (UAA) codon, while RF2 recognizes the opal (UGA) and ochre stop codon. The development of an engineered ribosome, termed ribo-X, which no longer recognizes RF1, would ideally result in minimized RF1-mediated chain termination during amber suppression while not affecting the decoding of chromosomal amber codons. Mutations in the 3' end of the rRNA within the 30S subunit enabled guidance of an orthogonal ribosome (O-ribosome) to a complementary mRNA sequence (O-mRNA). Such O-ribosome/O-mRNA pairs can function in parallel to the endogenous translational machinery, rendering them completely orthogonal. Enhancement of amber suppression efficiency by ribo-X was successfully demonstrated using one and two amber codons.[198] Since then, multiple-site ncAA incorporation has also been facilitated by entirely deleting RF1 from *E. coli*, thereby reassigning the UAG codon as a sense codon.[199, 200]

The additional replacement of UAG with UAA or UGA codons throughout the *E. coli* genome yielded stable and fast-growing strains with improved protein yields.[201, 202] Similar strategies would unfortunately not suffice in eukaryotic cells, since a single release factor, namely eRF1, recognizes all three stop codons. This is partially why a significant drop in efficiencies can usually be observed for genetic code expansion in mammalian cells. eRF1 could still be engineered in such a way, that it led to enhanced multisite ncAA incorporation efficiencies without increasing readthrough of endogenous stop codons in mammalian cells.[203]

The use of blank codons beyond stop codons would not only circumvent the problematics encountered with RFs, but it would also facilitate the encoding of multiple diverse ncAAs. A fact that also occurred to Chin and co-workers, who pioneered the use of quadruplet codons and reassigned existing triplet sense codons. Non-canonical, quadruplet decoding methods using frame-shift tRNAs have been explored early on in genetic code expansion, due to the low abundance of endogenous tRNAs bearing extended anticodons.[204, 205] Initially, these approaches have mainly relied on the exogenous addition of pre-aminoacylated tRNAs bearing an extended anticodon loop. It was not until the evolution of new aaRSs which could efficiently aminoacylate quadruplet anticodon bearing tRNAs, that the expansion of the genetic code seemed no longer to be limited by 64 triplet codons.[205] Nonetheless, the mechanism of quadruplet decoding is to this day not fully understood, which has limited the advance of quadruplet codons in genetic code expansion.[206]

First efforts by Neumann et al. have addressed the endogenous ribosome as the main constraint contributing to inefficient suppression of quadruplet codons. Similar to the deletion of RF1 recognition, orthogonal ribosomes can be altered without adversely affecting the proteome within a cell, contrary to the natural ribosome. Ultimately, an evolved orthogonal ribosome, named ribo-Q, was able to accommodate tRNAs bearing an extended anticodon loop.[207] Since then, ribo-Q has been combined with engineered, mutually orthogonal aaRS/tRNA pairs, promoting quadruplet and enhanced amber-decoding to incorporate multiple distinct ncAAs (Figure 1.19).[208] Unfortunately, depending on the open reading frame (ORF), these suppression systems suffered from low protein yields.

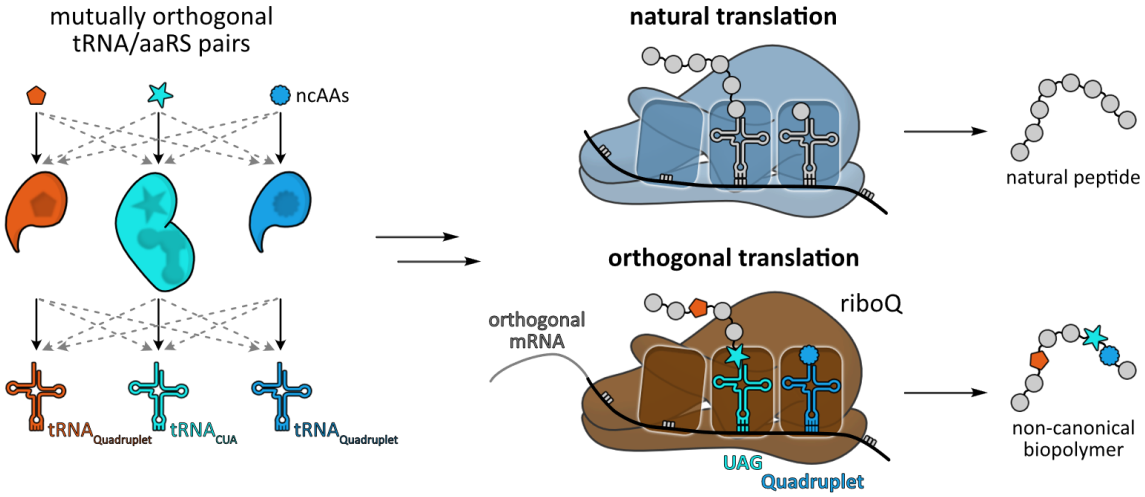


Figure 1.19 | Expansion of the genetic code using orthogonal ribosomes. Mutually orthogonal tRNA/aaRS pairs allow for the facile incorporation of three (or more) distinct ncAAs in response to the amber stop codon and quadruplet codons. If placed on an orthogonal mRNA, only the evolved orthogonal ribosome (**riboQ**, brown) can efficiently translate these codons. Non-canonical biopolymer synthesis can thus be performed simultaneously to natural peptide synthesis within the same cell.[209]

To address this, Dunkelmann et al. established an improved discovery system for orthogonal mRNA sequences, to which orthogonal ribosomes would bind more efficiently. Together with an operon-based expression system for mutually orthogonal aaRS/tRNA pairs, a genetic code containing 68 codons was successfully established. This approach gave maximized yields for proteins bearing four different ncAAs.[210]

A second strategy towards the incorporation of multiple ncAAs takes advantage of endogenous sense codons. The degeneracy of the genetic code allows for the liberation of some sense codons without impeding natural protein synthesis. Subsequent reassignment of these new blank codons allows for incorporation of ncAAs instead of natural amino acids. Thus, a semi-synthetic organism possessing a reduced genetic code of only 61 codons was created. Termed Syn61 by Chin and co-workers, three blank codons have been made available for incorporation of structurally diverse ncAAs with unparalleled efficiency.[211] Recent advances have demonstrated the outstanding potential of this *E. coli* strain containing a synthetic genome even further. Using three distinct codon reassignment schemes (TCG, TCA and TAG), the synthesis of diverse non-natural, macrocyclic peptides has been programmed into the genome of Syn61Δ3 cells.[212]

Beyond traditional *in vivo* incorporation of ncAAs using GCE, so-called flexizymes (Fxs) have been successfully applied *in vitro* to establish a second genetic code, which circumvents the need for suitable aaRS/tRNA pairs altogether. Heavily pioneered by Suga and co-workers, Fxs catalyze the aminoacylation of the 3’ end of tRNAs with activated esters as aminoacyl donors.

Therefore, these ribozymes are also seen as a prime example of how the ancient genetic code evolved within an RNA dominated world.[213] Self-aminoacylation of tRNAs might have taken place through aaRS-like ribozymes being attached to the 5' end of the tRNA, similar to 5' leader sequences of precursor tRNAs. After self-aminoacylation, these ribozymes would then get cleaved to form the mature, aminoacylated tRNA. This theory provided the basis for the evolution history of the first flexizyme **eFx** (Figure 1.20).[214] Further evolution of **eFx** has produced two more catalytically active Fx variants, **dFx** and **aFx**. All three Fx variants are capable of installing a broad range of structurally diverse ncAAs[215] through binding different aminoacyl donors (Figure 1.20).[216, 217]

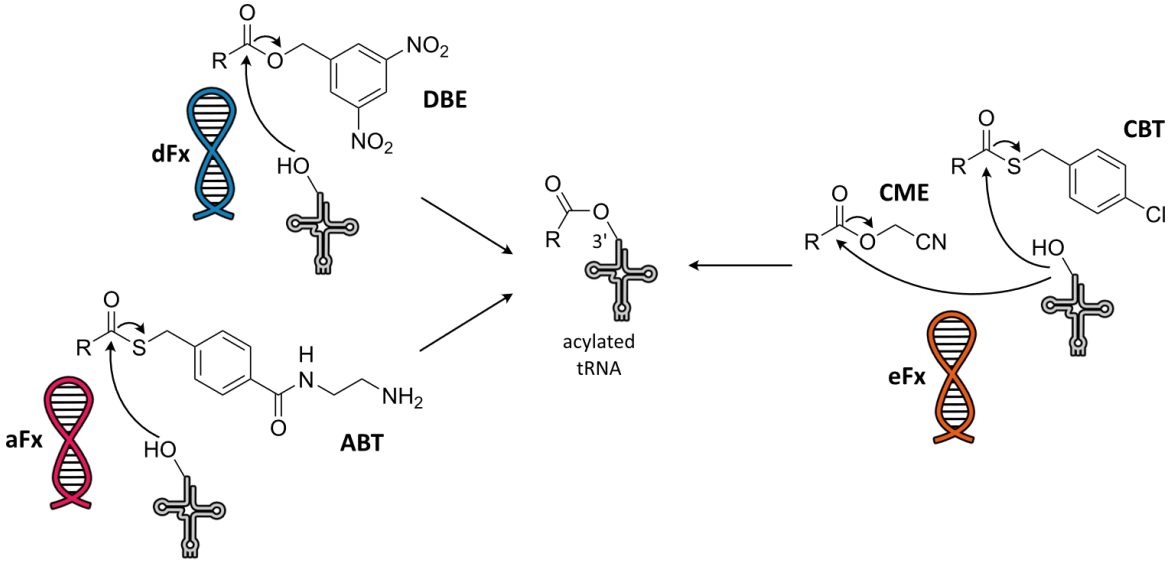


Figure 1.20 | Schematic depiction of tRNA acylation using flexizymes (Fx). Different activating groups are needed depending on the Fx. The first evolved variant **eFx** (orange) recognizes cyanomethyl esters (**CME**) and chlorobenzyl thioesters (**CBT**) as aminoacyl donors for aromatic phenylalanine analogs and non-aromatic side chains respectively. **dFX** (blue) catalyzes the acylation reaction of substrates bearing dinitrobenzyl esters (**DBE**) as leaving group. Lastly, **aFX** (pink) charges non-polar substrates activated as amino-functionalized benzyl thioester (**ABT**) onto the 3' end of tRNAs.[217]

This highly versatile technique, termed flexible *in vitro* translation (FIT), has not only allowed for the installation of ncAAs into proteins.[218] It has further facilitated the ribosomal preparation of aromatic foldamers[219] and complex macrocyclic peptides containing *N*-methylated, D-, cyclic γ-[220] or β-amino acids[221] as well as thiazole and oxazole moieties[222]. The development of novel macrocyclic peptide pharmacophores holds great potential for future therapeutic applications. Especially since insertion of such macrocyclic peptides into Fc regions of human immunoglobulin has demonstrated an enhanced transport across the blood-brain barrier *in vivo*.[223]

The within this chapter described efforts have propelled genetic code expansion and the preparation of designer proteins into new dimensions, including the facile, cell-based synthesis of complex new biopolymers which may also hold great potential for therapeutic applications. Still, there are even more ways to further push genetic code expansion beyond its long-believed limits, either through novel forms of engineered ribosomes[224, 225] or by increasing the coding possibilities of DNA through unnatural base pairs[226, 227]. The repertoire of functionalities that can be introduced into proteins is thereby also continuously growing, encompassing tools for visualization of proteins and mapping of protein-protein interactions (PPIs), as well as optical control over protein functions.

1.5 Genetically Encoded Non-Canonical Amino Acids

These recent efforts regarding time-saving evolutionary approaches for novel aaRS/tRNA pairs and the discovery of the N-terminal domain lacking PylRS from *Ma*, have further expanded the toolbox of ncAAs for *in vivo* amber suppression. Pioneered by Schultz and co-workers, the first ncAA was incorporated into proteins in *E. coli* more than two decades ago.[167]

By now, the chemical functionalities that proteins can be endowed with, include a broad variety of post-translational modifications (PTMs, Figure 1.21a)[228] as well as bioorthogonal handles[229], e.g. strained alkynes (**BCNK**) or tetrazine-based ncAAs[230] (**Tet4**, Figure 1.21a).

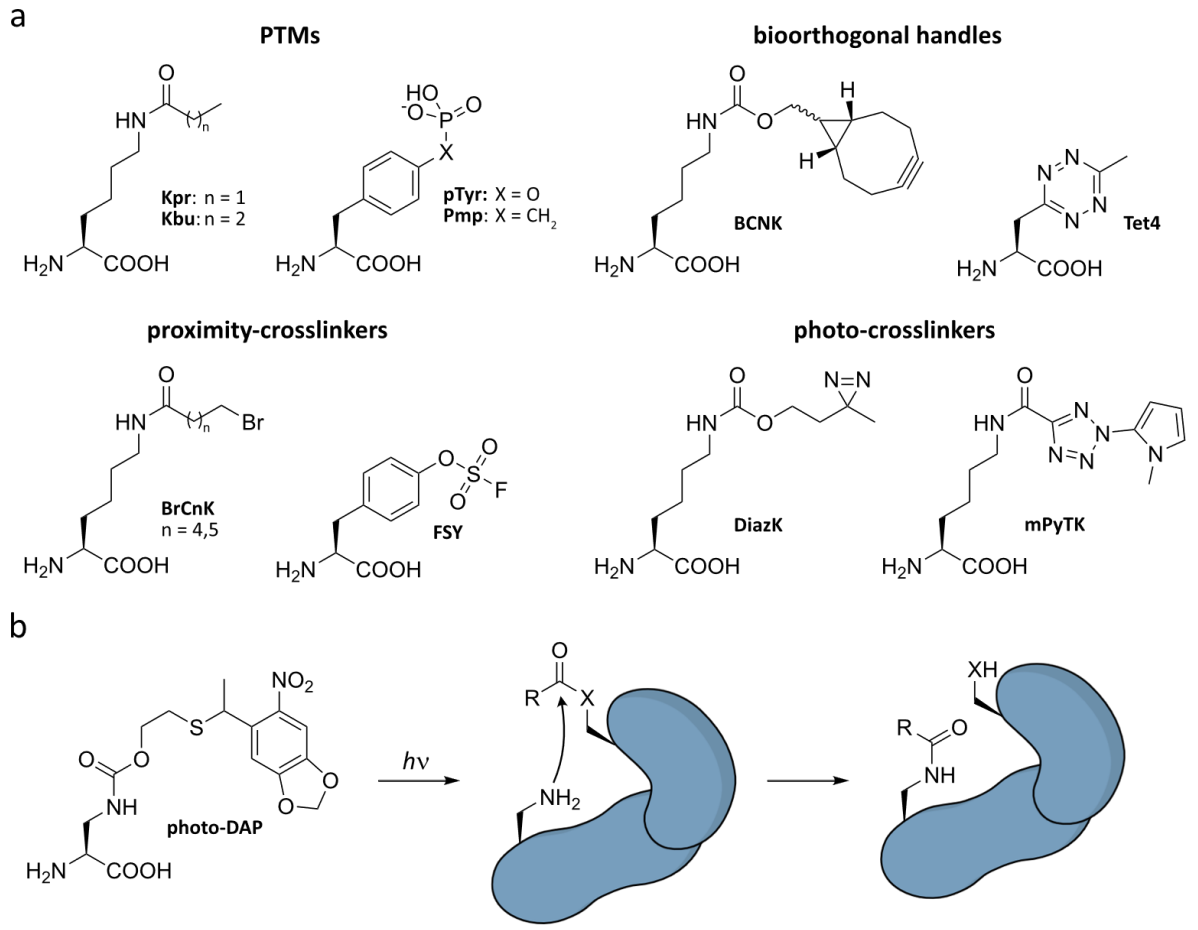


Figure 1.21 | Overview of ncAAs used in genetic code expansion. (**a**) Structural formulas of exemplary ncAAs[228] used as PTMs[231], bioorthogonal handles[229, 232] and (photo)crosslinkers[233, 234]. (**b**) Schematic depiction of a possible **photo-DAP** application. Upon photodecaging and subsequent spontaneous fragmentation, **DAP** traps the transient acyl-enzyme intermediate (X = O or S) through formation of a stable amide bond.[235]

The foundations for *in cellulo* bioorthogonal chemistry have been heavily pioneered by C. Bertozzi, M. Meldal and B. Sharpless. In 2022, their discoveries have been awarded with the Nobel Prize in Chemistry, since they enabled the nowadays broad application and development of well-known bioorthogonal reactions such as the traceless Staudinger ligation or inverse electron-demand Diels-Alder cycloaddition (iEDDAC) in live cells.[236] Installation of such moieties as bioorthogonally reactive ncAAs has allowed for the rapid and selective decoration of proteins with fluorophores, affinity tags and other small molecules within their physiological environment.

Beyond labeling approaches, ncAAs have proven valuable tools for mapping PPIs and stabilizing transient enzyme-substrate complexes. Photocrosslinking ncAAs based on e.g., diazirine (**DiazK**) or tetrazole scaffolds (**mPyTK**, Figure 1.21a) trap protein interaction partners upon UV-irradiation. Coupling this approach to subsequent MS/MS analyses renders photocrosslinking ncAAs a powerful tool for the identification of protein interaction partners and enzymatic substrate scopes. Because of their high reactivity, they are however not suited for the stabilization of low-affinity protein complexes. Here, proximity-driven crosslinker ncAAs bearing weak electrophilic moieties have been successfully applied in live cells instead (**BrCnK, FSY**, Figure 1.21a).[228, 237, 238]

Similarly, short-lived acyl-enzyme intermediates could be trapped using photocaged 2,3-diaminopropionic acid (**photo-DAP**, Figure 1.21b) as active-site Cys- or Ser-surrogate, resulting in a stable amide-bond.[235] Here, Huguenin-Dezot et al. employed the photocage mainly to facilitate incorporation of the desired ncAA. Beyond that however, photocaged ncAAs have found much broader application in the optical regulation of protein function.

1.5.1 Optical Control of Protein Function Using Photocaged ncAAs

In general, photocaged ncAAs are installed at functionally critical sites within a target protein, such as the active site or interface for protein interaction. To that end, a range of photocaged lysine, cysteine and tyrosine derivatives have been designed, which temporally block protein activity (**HCK**, **PCC**, **ONBY**, Figure 1.22a). Thus, it was not only possible to optically control the enzymatic activity of proteases[239], kinases[240] and phosphatases[241] but it also enabled photo-triggered activation of targeted protein degradation in live cells[242]. Further, a range of light-responsive anti-[243] and nanobodies[244, 245] were generated by incorporation of photocaged ncAAs.

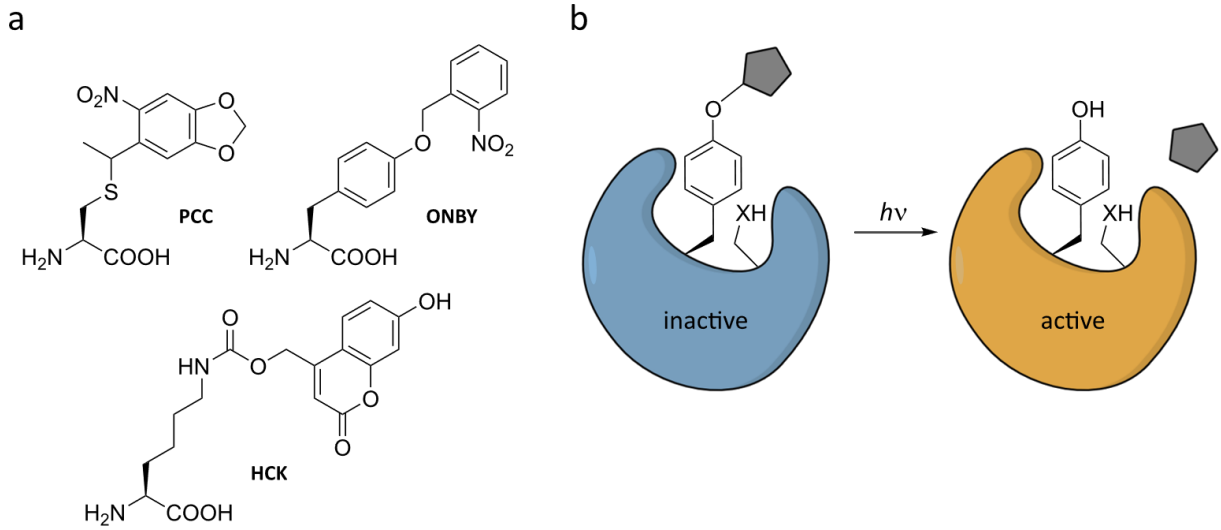


Figure 1.22 | Overview of photocaged ncAAs used in genetic code expansion. (**a**) Structural formulas of a subset of photocaged ncAAs.[228] (**b**) Schematic depiction of CAGE-prox. Photodecaging of a masked tyrosine residue in proximity to the active site residue (X = S or O) restores enzymatic activity (orange).[246]

Combination of the most prominent photocaged ncAA **ONBY** into a generalized concept for optical protein activation was realized by Chen and co-workers. CAGE-prox, meaning computationally aided and genetically encoded proximal decaging, uses an algorithm to predict suitable **ONBY**-incorporation sites in proximity to the functional site of a protein (Figure 1.22b).[246] The best-scoring sites can then be validated experimentally. Not only does this approach save experimental time by reducing the number of tested mutants, but it is also applicable to enzymes with yet undefined active sites.

Still, photocaged ncAAs only allow for a single, targeted activation of protein function, whereas azobenzene-based ncAAs have allowed for the reversible control over enzymatic activity.

1.5.2 Reversible Optical Control of Protein Function Using Azo-ncAAs

In general, there are two possibilities in applying azo-ncAAs for the reversible modulation of protein function. They can be incorporated at functionally critical sites within the POI either as non-crosslinking version (Figure 1.23a) or to subsequently crosslink with a proximal cysteine residue, furnishing an azo-bridge similar to azo-crosslinkers (Figure 1.23b).

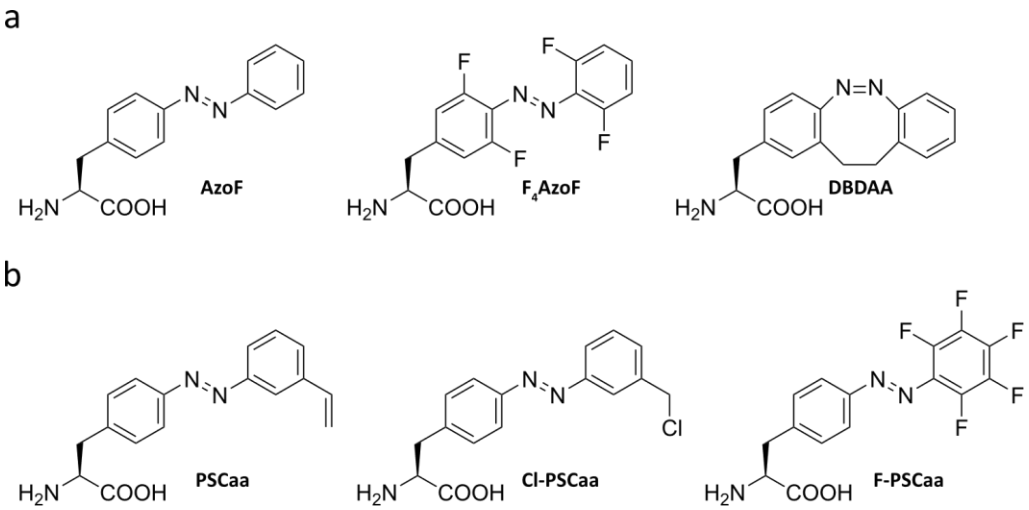


Figure 1.23 | Overview of azobenzene-based ncAAs used in genetic code expansion. (**a**) Structural formulas of non-crosslinking azo-ncAAs for side-chain isomerization. (**b**) Structural formulas of crosslinking azo-ncAAs for optical control over secondary motifs in proteins.

Although the non-crosslinking azobenzene-based ncAA (azo-ncAA) **AzoF** (Figure 1.23a) has already been incorporated into proteins in *E. coli* early on[204, 247], its successful application has been rather limited. An early example is the site-specific incorporation of **AzoF** into horse-radish peroxidase by a frame-shift suppressor tRNA to switch enzymatic activity between an "on"- and "off"-state.[204] However, the overall peroxidase activity was suffering greatly, merely retaining 3-10% of the wild-type activity in its "on"-state. Optical modulation of the restriction enzyme *Bam*HI through incorporation of **AzoF** into the key position K132 of the dimer interface followed. The *trans*-form of **AzoF** fully disrupted the H-bond between H133 and E167, which is crucial for the structure of the dimer interface and thus resulted in abolishment of enzymatic activity. Although photoisomerization to the *cis*-isomer successfully restored enzymatic activity of *Bam*HI, subsequent photoisomerization from the *cis*- to the *trans*-isomer by irradiation at 436 nm did not yield any substantial change in activity.[248]

Almost a decade later, a set of photoswitchable click amino acids (**PSCaas**, Figure 1.23b, Chapter 1.3.3) were incorporated into an α-helical motif within calmodulin.[249] Following incorporation, **Cl-PSCaa** reacted with a proximal cysteine, establishing a photoswitchable bridge with a spacing of *i*,*i*+7 that was capable of changing the conformation of calmodulin.

A year later, Hoppmann et al. demonstrated spatio-temporal control over the binding of calmodulin to its interacting peptide of the nitric oxide synthase 1 (NOS1) by incorporation of a pentafluoro-substituted PSCaa (**F-PSCaa,** Figure 1.23b).[250] Due to its *ortho*-fluoro substituents, conformational control over calmodulin could be exerted using visible light. Similarly, Deiters and co-workers developed a non-crosslinking, tetra-*ortho*-fluorinated azo-ncAA (**F_4AzoF**, Figure 1.23a), thereby gaining reversible photocontrol over firefly luciferase in live mammalian cells.[251]

The azo-ncAAs **PSCaa** and **AzoF** have since been used as non-crosslinking tools to optically modulate the activity of the glutamate receptor NMDA[252] as well as the imidazole glycerol phosphate synthase ImGPS,[253, 254] respectively. Choosing appropriate sites for the azo-ncAA installation can be tedious if done experimentally only, which is why computational strategies have oftentimes been utilized.[251, 253]

Although there have been examples of successful regulation of protein function with strategically placed azo-ncAAs, the repertoire of photoswitching ncAAs is still small. Further, little has been done to improve on the photophysical properties of said azo-ncAAs, even if **F_4AzoF** and a recently developed bridged azo-ncAA (**DBDAA**, Figure 1.23a)[255] hold potential for the future. We therefore set out to create synthetically easily accessible azo-ncAAs featuring optimized photocharacteristics, i.e. high PSS, long relaxation time τ and ideally visible-light wavelength photoswitching. As a result, we would contribute to the toolbox of azo-ncAAs available so far, enabling a modular approach for the spatio-temporal control of diverse protein functions.

2 Project Aim

Within this project we aimed at developing a range of novel, photoswitchable ncAAs as tools for spatio-temporal modulation of protein conformation and function via genetic code expansion. To this end, we designed and synthesized azobenzene-based ncAAs (azo-ncAAs) bearing cysteine-reactive groups that were to be incorporated into structurally relevant α-helix motifs by amber codon suppression. Proximity-enhanced reaction *in situ* with a site-selectively mutated cysteine along the same α-helix (*i,i+4*; *i,i+7* or *i,i+11*) would lead to the formation of an azobenzene-based crosslink. Similarly to crosslinks used for peptide stapling, we anticipated the *trans*-isomer of azo-ncAAs to be stabilizing the α-helix using a spacing of *i,i+7* or maybe even *i,i+11* (Figure 2.1). Subsequent irradiation using UV-light to drive photoisomerization to the *cis*-isomer should therefore destabilize α -helix conformation. In contrast, installation of the azobenzene by a spacing of *i,i+4* would lead to a decrease in α-helical propensity. Photoisomerization to the *cis*-isomer would then ideally restore α-helix conformation. Since here the *cis*-isomer might be necessary for quantitative crosslinking, we decided to focus on the installation of azo-ncAAs using the spacing *i,i+7*, similar to what can be found in literature.[250]

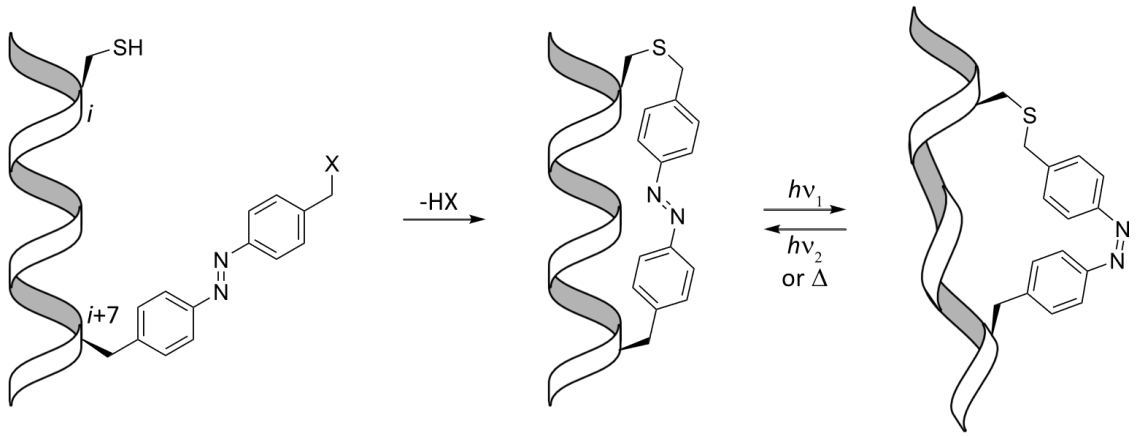


Figure 2.1 | Schematic depiction of spatio-temporally controlling α-helix propensity with azo-ncAAs. Proximity-induced reaction of an incorporated azo-ncAA with a cysteine amino acid residue at *i+7*, and reversible photoswitching of the formed azo-bridge. The *trans*-isomer stabilizes the α-helix, while photoisomerization to the *cis*-isomer results in disordered α-helical structures.

Consequently, reversible photoswitching of such azobenzene bridges would drive conformational switching of the α-helix and thus turn off protein function upon photoisomerization to the *cis*-isomer. The thermodynamically stable *trans*-state of the azo-ncAA equates to the protein being in its native conformation with operative function.

2.1 Optical Modulation of Biological Processes

Installation of azo-ncAAs into α-helices of proteins would render them light-sensitive. We thus envisioned modulation of different biological processes orthogonally by light.

We were particularly interested in investigating PPIs by application of azo-ncAAs due to their crucial importance for the functioning of most biological processes in all living organism. Ideally, stabilization of the α-helix in *trans* (for *i,i+7* and *i,i+11*) would allow the interacting partner to bind to the light-sensitive protein. Photoisomerization to the *cis*-isomer would then lead to dissociation of the respective binding partner (Figure 2.2a). Binding studies based on pulldown assays or fluorescence anisotropy would be applied to measure and evaluate the thus manipulated PPIs.

Beyond that, we were curious if our concept could also be applied to the spatio-temporal regulation of enzyme function. We proposed that by incorporation of azo-ncAAs into functionally critical α-helix motifs, we could reversibly switch between a stable and disordered α-helix conformation and thus specifically turn enzymatic activity off and on (Figure 2.2b). The generation of light-sensitive hydrolases or peptidases would allow us to simply and quickly assess the efficiency of azo-ncAA tools using fluorogenic substrates.

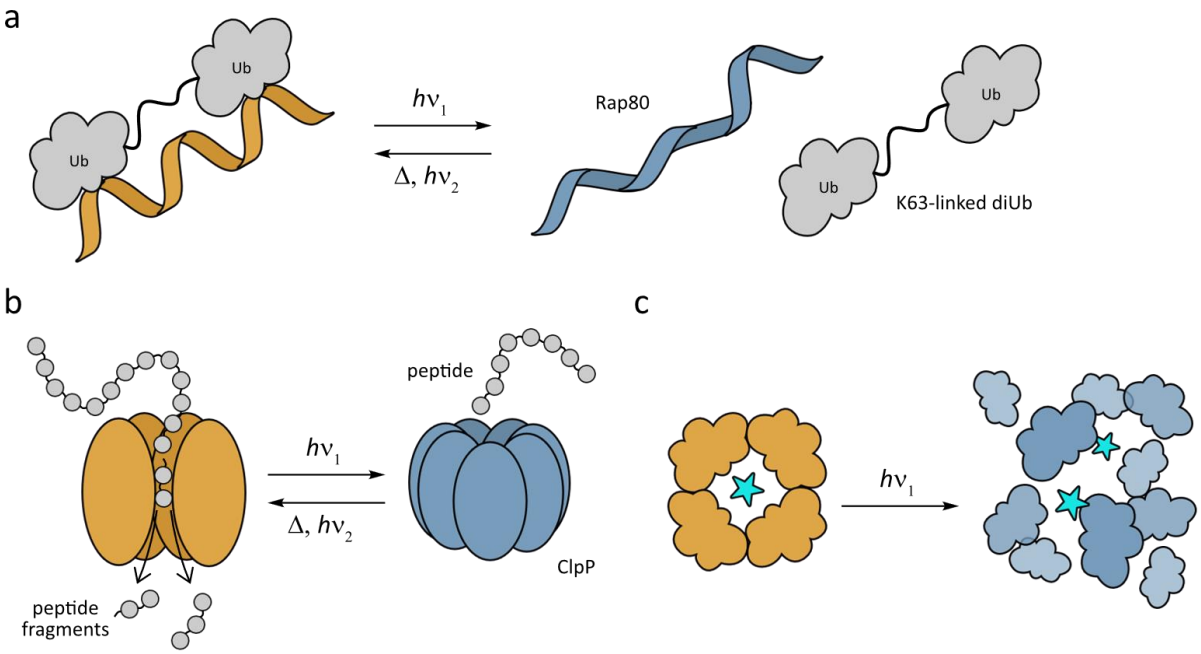


Figure 2.2 | Schematic overview over the optical modulation of various biological processes. (**a**) Schematic depiction of optically controlled PPIs, exemplarily shown for Rap80 and K63-linked diUb (Chapter 3.3.1). (**b**) Schematic depiction of enzyme activity regulated by photoirradiation, exemplarily shown for the serine protease ClpP (Chapter 3.3.2). (**c**) Schematic depiction of light-triggered cargo release from protein capsids, exemplarily shown for the protein cage OP (Chapter 3.3.3).

Finally, we envisioned to use our most promising azo-ncAA candidates to generate a light-sensitive protein assembly as a tool for targeted and controlled delivery of cargo molecules. Such tools are highly sought after and could largely benefit genetic engineering approaches. One possibility to address this would be to install azo-ncAAs into α-helix motifs crucial for protein cage assembly. We proposed, that photoisomerization to the *cis*-isomer would not only destabilize α-helix conformation but also protein cage assembly, leading to a light-triggered release of cargo molecules (Figure 2.2c).

2.2 Envisioned Azo-ncAAs for Proximity-Induced Crosslinking

The azo-ncAAs need to fulfill certain criteria to be considered successful tools for the optical modulation of protein function using a genetic code expansion approach. They need to (i) be readily synthesized on a large gram scale, (ii) be stable under physiological conditions, (iii) be efficiently incorporated into target proteins, (iv) undergo proximity-induced crosslinking with a cysteine in quantitative yields and (v) feature optimal photophysical properties.

To this end, we envisioned azo-ncAAs for several ligation reaction types, including S_N2, S_NAr and Michael addition (Figure 2.3). Varying the halide substituent and/or length of the alkyl-halide linkers would allow us to further fine-tune the reactivity of azo-ncAAs towards a proximal cysteine residue. Finally, installation of alkoxy groups in *para*-position and/or halide moieties in *ortho*-position to the diazene bond would improve the photophysical properties of the desired azo-ncAA tools, in particular the PSS of the *cis* → *trans* isomerization process.

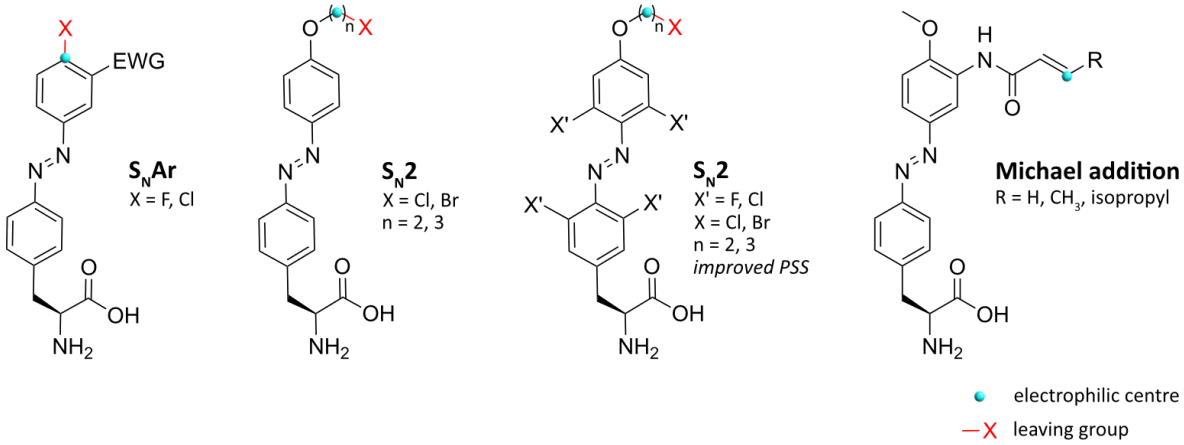


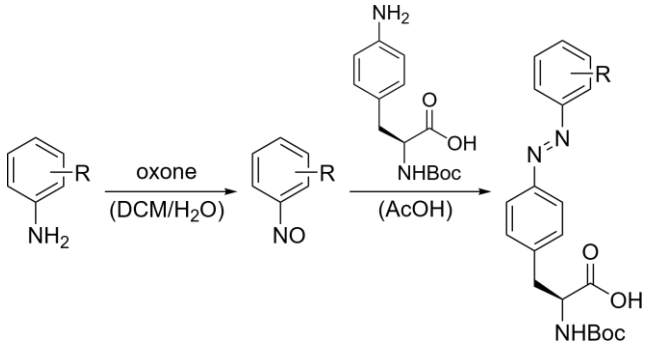
Figure 2.3 | Envisioned design of azo-ncAAs for the optical modulation of protein conformation. Structural formulas of azo-ncAAs for S_NAr, S_N2 and Michael addition reactions with a proximal cysteine residue. Decoration of azo-ncAAs with *para*-alkoxy moieties or tetra-*ortho*-halide substituents would improve the PSS of the *cis* → *trans* isomerization.

3 Novel Azo-ncAAs for the Optical Control of Protein Function

3.1 Development of Novel Azo-ncAAs Bearing Cysteine-Reactive Moieties

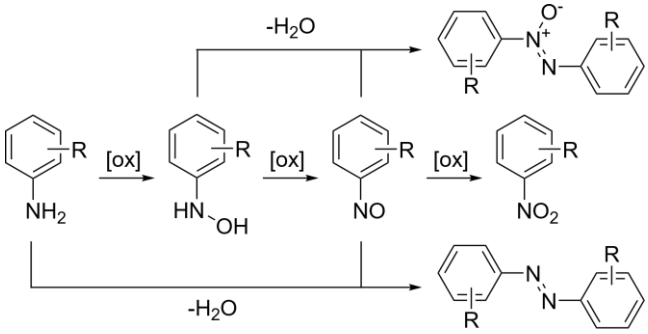
3.1.1 General Synthesis of Azobenezenes

A broad range of synthetic routes to access azobenzenes has been established since their emergence almost two centuries ago, especially regarding symmetrically substituted derivatives. The design of azo-ncAAs however needs syntheses that allow for the preparation of non-symmetric derivatives, which remains challenging to this day. Additionally, amino acid functionalities require adequate protecting group strategies, further limiting possible synthesis routes. The Baeyer-Mills reaction presents an attractive choice for the synthesis of azo-ncAAs (Scheme 3.1).



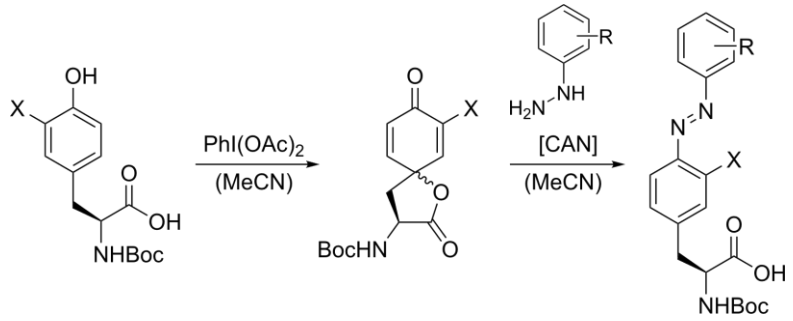
Scheme 3.1 | Synthesis of N_α-Boc protected azo-ncAAs via the Baeyer-Mills reaction.

It uses mild oxidative conditions for the preparation of nitrosoarenes that are subsequently condensed with N_α-protected 4-amino-phenylalanine derivatives in glacial AcOH, thus making it compatible with a variety of protecting groups. Due to these features, the Baeyer-Mills reaction has found broad application in the synthesis of azo-ncAAs.[249-251, 256] However, the yields of the Baeyer-Mills reaction often suffer from over-oxidation to the corresponding nitroarene or undesired condensation side-reactions with the starting aniline or the *N*-arylhydroxylamine intermediate (Scheme 3.2).



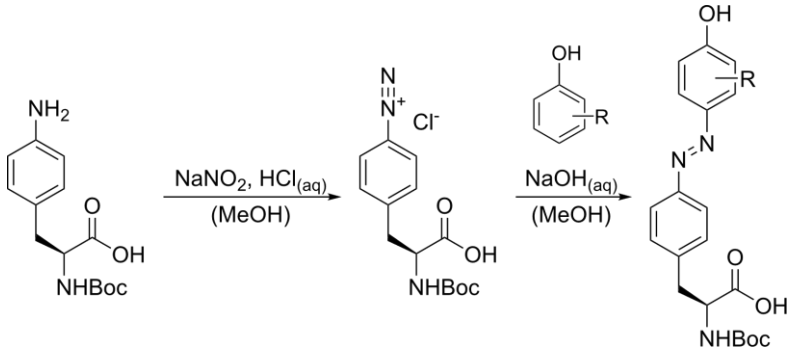
Scheme 3.2 | Side products observed during nitrosoarene formation of the Baeyer-Mills reaction.

While the use of biphasic systems like dichloromethane (DCM) and aqueous oxone® ($2KHSO_5 \cdot KHSO_4 \cdot K_2SO_4$) have proven to shield the hydrophobic nitrosoarene from most of the mentioned side-reactions, typical yields of the Baeyer-Mills reaction do not exceed 40%.[257] Therefore, additional approaches have been developed for the preparation of non-symmetric azo-ncAAs. The synthesis via quinonoidal spirolactones resulted in a series of N_α-Boc-protected and red-shifted azo-ncAAs, obtained in good to excellent yields (Scheme 3.3).



Scheme 3.3 | Synthesis of N_α-Boc protected azo-ncAAs via quinonoidal spirolactones (X = H,F,Cl).

Starting from N_α-Boc-tyrosine, a slight excess of the oxidant phenyliodine(III) diacetate leads to the formation of the quinone intermediate. Reaction with a phenylhydrazine derivative catalyzed by the one-electron oxidant ceric ammonium nitrate (CAN) gives the desired azo-ncAAs.[258] A facile synthesis route towards azobenzenes, which has so far been unexplored for the preparation of azo-ncAAs, employs azo-coupling reactions (Scheme 3.4).



Scheme 3.4 | Synthesis of N_α-Boc protected azo-ncAAs via azo-coupling to phenols.

Here, azobenzene derivatives featuring strongly electron-donating groups in *para* position to the diazene bond are obtained. First, *in situ* generation of nitrous acid leads to the diazotization of *N*-Boc protected 4-amino-phenylalanine under acidic conditions. Addition of the cooled diazonium solution to an electron-rich arene, usually a phenol or aniline, gives azo-ncAAs in good yields via electrophilic aromatic substitution (S_EAr) under mildly basic conditions (pH~9).

Since the azo-coupling route is highly pH-dependent in both the diazotization and the subsequent coupling step, a protocol compatible with the convenient Boc-protecting group was developed during this work. Typically, this required 10 eq of 0.7 M HCl in aqueous methanol for efficient diazotization without deprotection of the amino functionality.

3.1.2 Prior Work

Prior work on this project has centered on two different series of azo-ncAAs.[259] The standard series encompasses **TriF** for S_NAr reaction, **s2Br** for S_N2 reaction and **s2ene** for thiol-ene click reaction (Figure 3.1a), all synthetically accessible via the well-established Baeyer-Mills reaction. A second set of azo-ncAAs was designed to feature improved photophysical properties, comprising **m2Br** for S_N2 reaction and **m2ene** for thiol-ene click reaction (Figure 3.1b). To this end, both azo-ncAAs were substituted with a *para*-methoxy-group, allowing for the facile preparation employing the azo-coupling route. As a consequence, the cysteine-reactive linker was designed to be in *meta*-position to the diazene, contrary to the standard azo-ncAA series.

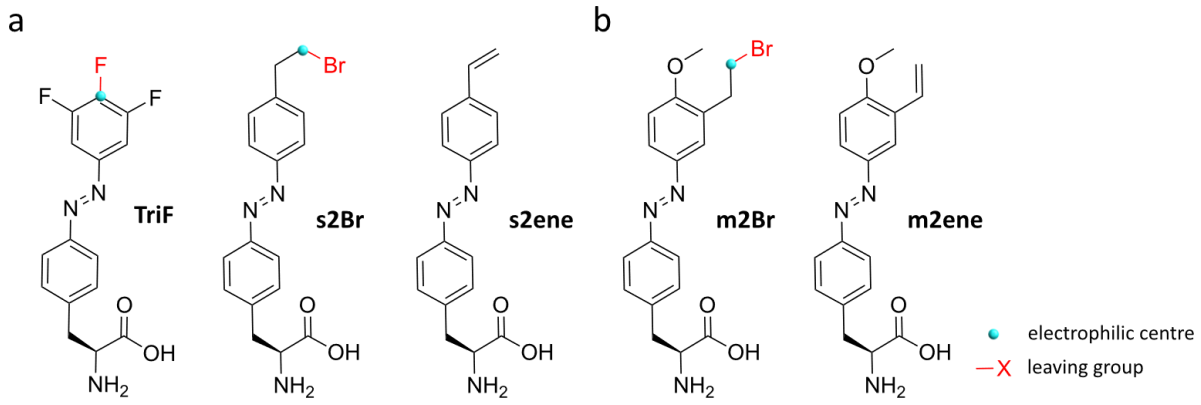


Figure 3.1 | Overview of azo-ncAAs developed in prior work. (**a**) Structural formulas of standard azo-ncAAs **TriF**, **s2Br** and **s2ene**. (**b**) Structural formulas of azo-ncAAs with improved photophysical properties **m2Br** and **m2ene**.

The development of **m2Br** and **m2ene** were considered a success, since the $PSS_{365\,nm}$ increased from 63% for standard azo-ncAAs to 85% and 78%, respectively. Unfortunately, successful incorporation into sfGFP-N149TAG-His_6 could only be observed for **TriF** using AzoB (A302T, L309A, N346A, C348G) and the two thiol-ene click reactive analogs **s2ene** and **m2ene** using the *Mm*PylRS AzoC (A302T, L309A, I322T, N346A, C348G)[249]. Since screening for suitable conditions of thiol-ene click reaction proved unsuccessful on small molecule level, we focused on the application of the successfully developed azo-ncAA **TriF** for S_NAr reactions.

Beyond that, prodrug-like strategies were explored, aiming at improved solubility and cellular bioavailability of the obtained azo-ncAAs. To this end, a lysine-azo-ncAA dipeptide that would get cleaved in *E. coli* by endogenous unspecific aminopeptidases has been prepared. While the dipeptide could indeed be successfully dissolved in water up to 100 mM, it did not prevent precipitation when added to the cell media.

3.1.3 Design of Azo-ncAAs From This Project

Within this project, we desired to design novel azo-ncAAs for several reaction ligation types, including S_NAr, S_N2 and Michael addition. For each ligation type we intended to design and synthesize a range of azo-ncAAs of varying reactivity. For proximity-induced S_NAr-reaction, the 4-halo-3-nitro azobenzene analogs **pClN** and **pFN** were designed to increase S_NAr reactivity in comparison to **TriF** (Figure 3.2a).

For classic S_N2 reaction three different avenues were pursued, predominantly based on *para*-alkoxylated scaffolds since prior work has proven their superiority in comparison to standard azo-ncAAs. Firstly, we designed a set of azo-ncAAs bearing typical S_N2-reactive alkyl chloride or the more reactive alkyl bromide linker in *para*-position to the diazene (Figure 3.2b). Unfortunately, finding a suitable PylRS accepting azo-ncAAs bearing the linker in *para*-position has been unsuccessful so far, as demonstrated in prior work. Therefore, we designed a second set of azo-ncAAs based on the highly promising scaffold of **m2Br** and **m2ene**. We included the less reactive alkyl chloride analog **m2Cl** as well as two benzyl halide derivatives (Figure 3.2c). Since benzyl halides are already more activated for S_N2-reactions, we focused on the benzyl chloride and even less reactive benzyl fluoride azo-ncAA. For the generation of a third set of activated and S_N2-reactive azo-ncAAs we desired to install α-halo acetamide linker in *para*- or *meta*-position to the diazene bond (Figure 3.2d). The last ligation type chosen was the highly reactive thiol Michael addition. Similar to the design of α-halo acetamide azo-ncAAs, we envisioned installation of α,β-unsaturated amide linker in *para*- or *meta*-position to the diazene bond (Figure 3.2e).

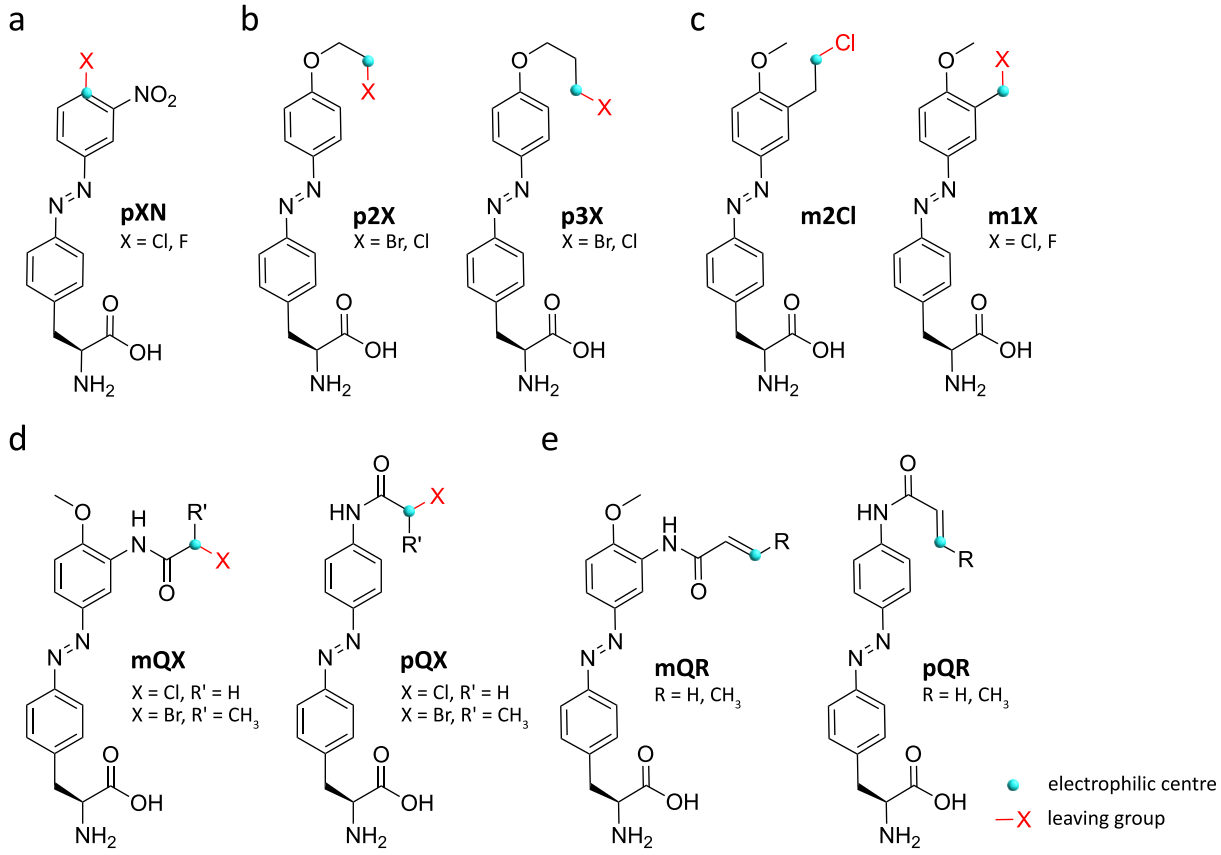


Figure 3.2 | Design of azo-ncAAs for different ligation reaction types. (**a**) Structural formula of azo-ncAAs **pXN** for S_NAr reactions. (**b**) Structural formula of azo-ncAAs **p2X** and **p3X** for S_N2 reactions. (**c**) Structural formula of azo-ncAAs **m2Cl** and **m1X** for S_N2 reactions. (**d**) Structural formula of azo-ncAAs **mQX** and **pQX** for activated S_N2 reactions. (**e**) Structural formula of azo-ncAAs **mQR** and **pQR** for thiol Michael addition reactions.

Once we would have identified those azo-ncAAs that get efficiently incorporated into target proteins and feature adequate reactivity for full crosslinking with a proximal cysteine, we envisioned further optimization of their photophysical properties. To this end, we intended to explore late stage tetra-*ortho* chlorination of azo-ncAAs as well as the synthesis of tetra-*ortho* fluorinated azo-ncAAs (Figure 3.3a).

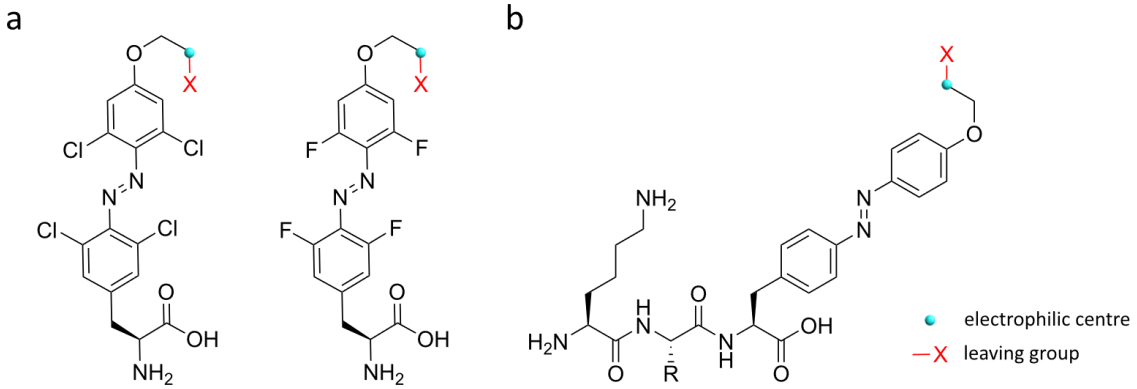


Figure 3.3 | Design of azo-ncAAs featuring improved photophysical properties and enhanced solubility in aqueous environments. (**a**) Structural formulas of exemplarily azo-ncAA **p2X** bearing either tetra-*ortho*-chloro or tetra-*ortho*-fluoro substituents to improve the photophysical properties. (**b**) Structural formula of azobenzene tripeptide scaffold bearing N-terminal lysine to enhance aqueous solubility and thus bioavailability of the exemplarily azo-ncAA **p2X**. R is preferably H or CH_3 to test the effect of small unreactive amino acids such as Gly or Ala, respectively.

This way, highly bistable azobenzenes with an improved PSS can be generated. Beyond that, it would also enable photoisomerization using wavelengths in the visible-light region, uniting many of the features that are typically desirable in a photoswitch for the spatio-temporal control of protein functions.

An additional feature of azo-ncAAs or any organic molecule in an aqueous environment to consider, is its solubility in water. While several strategies for the generation of water-soluble organic molecules exist, not many of them are applicable to azo-ncAAs as tools in genetic code expansion. We suspected that typical solubilizing groups such as charged phosphates, phosphonates or sulfonates would abolish incorporation by aaRS/tRNA pairs. Due to similar reasons, the addition of non-ionic solubilizing groups also applied in medicinal chemistry was not considered for our meticulously designed azo-ncAAs. While the addition of water-miscible co-solvents such as DMSO or various surfactants to the medium has been explored in prior work, it brought only minor improvements. Instead, we desired to expand the concept of lysine-bearing peptides for the solubilization of azo-ncAAs. From prior work we knew, that coupling of the hydrophobic azo-ncAAs to lysine did indeed improve solubility in water. To avoid precipitation after addition to the medium however, we anticipated improvements by using Lys-X-azo-ncAA tripeptide scaffolds (Figure 3.3b).

A challenge unmet so far in the design of azo-ncAAs is based on the fact that many biological assays rely on measuring absorbance, fluorescence or luminescence. Only few solutions have been proposed for the generation of orthogonal photoswitchable tools, oftentimes based on entirely different scaffolds than azobenzenes.[260] While the use of such assays in conjunction with azobenzenes can be tricky and easily lead to false results, careful adjustments of the used assays within this project (e.g. orthogonal fluorophores) make the application of the designed azo-ncAAs feasible.

The following subchapters give insight into the development of various sets of azo-ncAAs, including their syntheses, stability as well as screening for incorporation in *E. coli* by a suitable PylRS. Such test-expressions were typically monitored by SDS-PAGE and western blotting. The deprotected azo-ncAA salts were dissolved in DMSO (100 mM final stock concentration containing 200 mM TFA), added together with 2% (v/v) Tween20 as co-solvent and tested for incorporation into C-terminally His_6-tagged superfolder GFP (sfGFP) with the amber stop codon at position N149. Purification of full-length protein via Ni^{2+}-affinity chromatography and its subsequent analysis by MS were employed to rule out misincorporation and confirm the incorporation of intact azo-ncAA. Screening of in-house available or rationally designed PylRS (*Mb* and *Mm*) and *Mj*TyrRS variants led to the identification of several *Mm*PylRS mutants as most promising candidates for the incorporation of azo-ncAAs (Table 3.1).

Table 3.1 | Mutations of *Mm*PylRS used in this work for the incorporation of azo-ncAAs. AzoA, AzoC, KK4 and KK11 were known to incorporate azo-ncAAs from literature, while AzoB and KK10 were initially only cloned as intermediate plasmids.

*Mm*PylRS	A302	L305	Y306	L309	I322	N346	C348	Y384
AzoA[249]	T			S		V	G	
AzoB	T			S		A	G	
AzoC[249]	T			A	T	A	G	
KK4[251]			M	A		A	A	F
KK10			M	A		G	G	F
KK11[251]		F		M		G	G	F
KK13[249]	T			S	L	V	G	

Interestingly, attempts at amber suppression of sfGFP-N149TAG using the corresponding *Mb*PylRS analogs resulted in nearly no azo-ncAA-bearing sfGFP variants, but mainly in protein truncation. Therefore, the following work focuses on the identified *Mm*PylRS variants (Table 3.1), some of which were known to incorporate azo-ncAAs from literature.[249, 251]

3.1.4 Development of Azo-ncAAs for S_N2-Based Crosslinking

First, two sets of *para*-alkoxylated azo-ncAAs were synthesized, bearing the S_N2-reactive moiety either in *para* (pnX series) or in *meta* (mnX series) to the diazene.

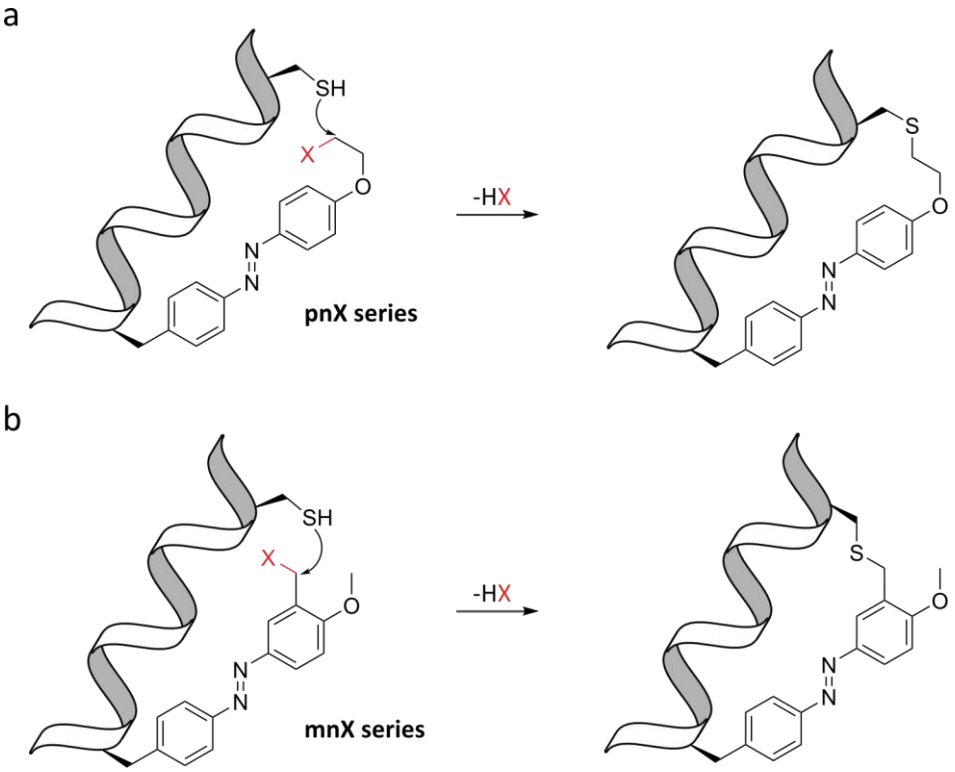
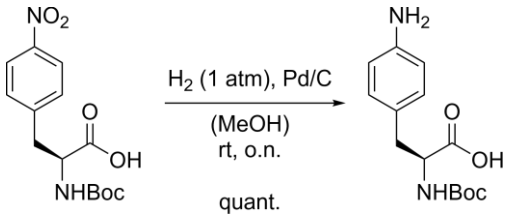


Figure 3.4 | Schematic depiction of the crosslinking reaction between an azo-ncAA and a proximal cysteine residue via S_N2. The proximity-induced reaction of an incorporated azo-ncAA with a cysteine residue is shown for the pnX series (**a**) and the mnX series (**b**), respectively.

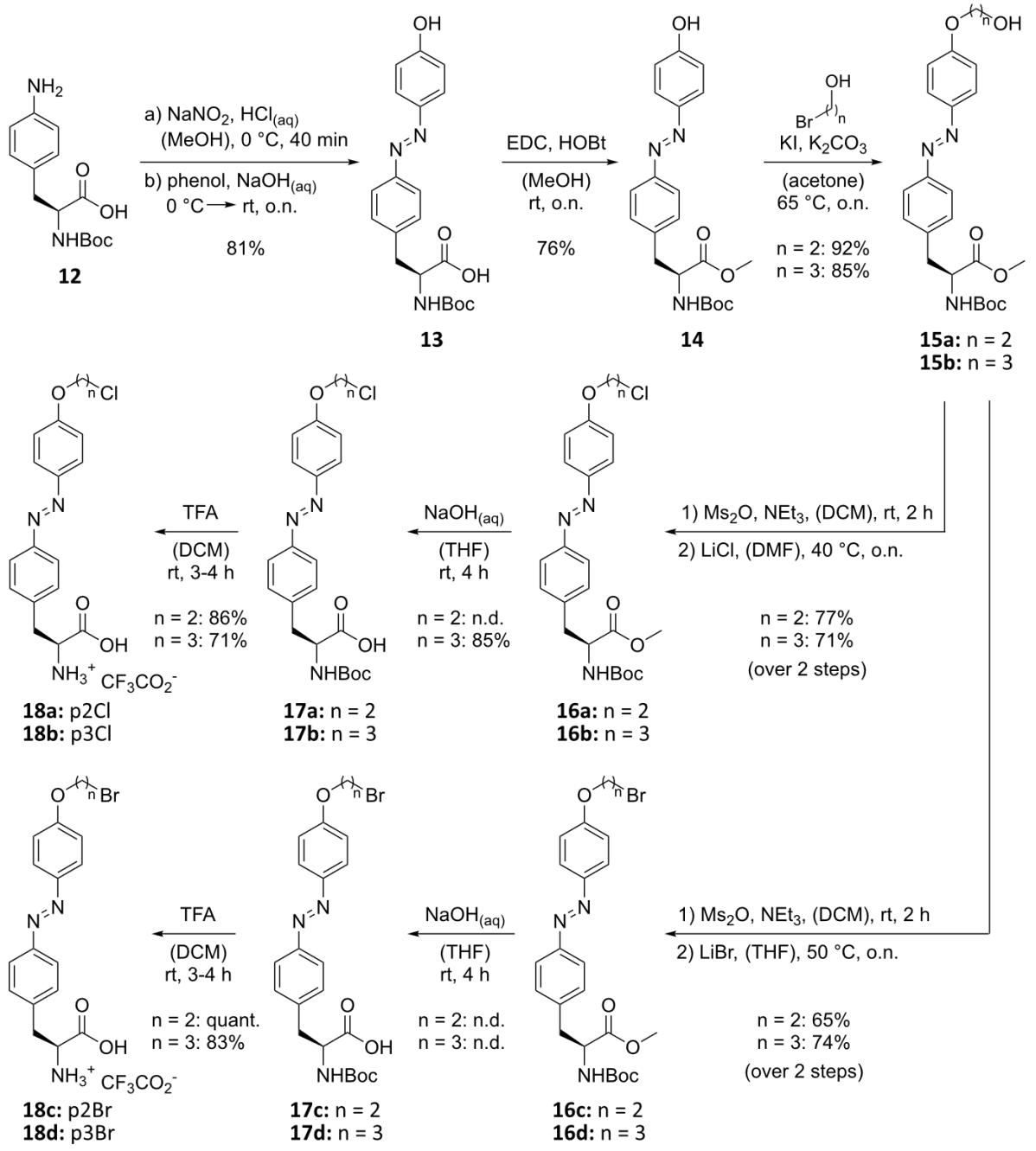
Both azo-ncAA sets were accessible via the high-yielding azo-coupling reaction, starting with reduction of commercially available Boc-Phe(4-NO_2)-OH **11** over Pd/C (Scheme 3.5).



11 **12**

Scheme 3.5 | Synthesis of the common precursor Boc-Phe(4-NH_2)-OH.

Subsequent diazotization of **12** and coupling to phenol gave the intermediate azobenzene **13** in very good yield (81%) (Scheme 3.6). Protection of the carboxyl group by EDC/HOBt in MeOH allowed for the selective alkylation of the phenolic alcohol in good yields (76%). Reaction of **14** with 2-bromo-1-ethanol or 3-bromo-1-propanol using catalytic amounts of KI, yielded **15a** and **15b**, respectively.

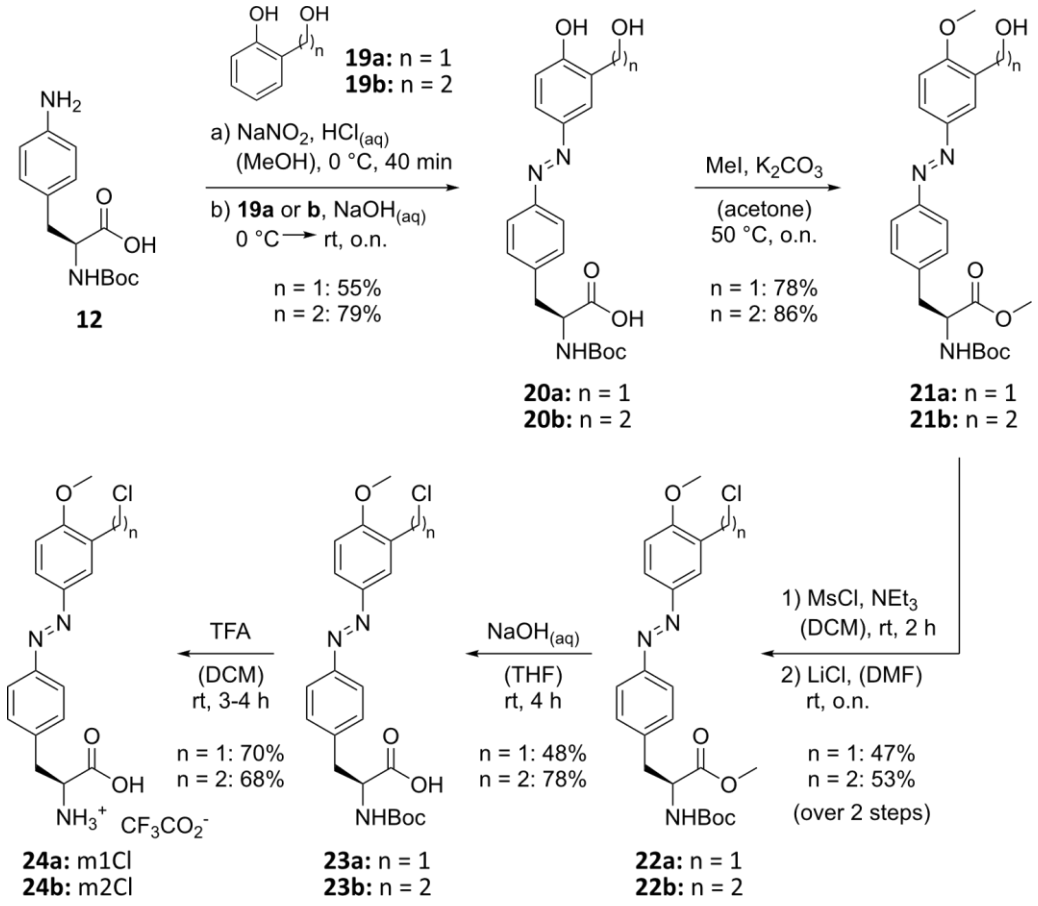


Scheme 3.6 | Synthesis route of *para*-alkoxylated azo-ncAAs p2Cl, p3Cl, p2Br and p3Br bearing S_N2-reactive linkers.

The aliphatic alcohol was mesylated and then treated with LiCl or LiBr to give the halides **16a-d** in good yields of 65-77%. The final azo-ncAAs **p2Cl**, **p3Cl**, **p2Br** and **p3Br** (**18a-d**, respectively) were obtained through hydrolysis of the methyl ester and subsequent Boc-deprotection, yielding the corresponding TFA-salts.

While initial results for the pnX series indicated successful incorporation using one published *Mm*PylRS KK4, analysis via LR-MS revealed possible misincorporation of assumedly phenylalanine (Supplementary Figure III.1). Unfortunately, no other synthetase could be found for any of the pnX analogs bearing the S_N2-reactive linker in *para*. Since these observations are in coherence with prior work from this project[259], we instead focused on the *meta*-substituted azo-ncAAs mnX.

Early UV-Vis analyses of the *para*-methoxylated mnX series had already confirmed our assumptions that *para*-alkoxylation of azo-ncAAs would give derivatives with improved photophysical features.[259] During previous studies, an exceptional PSS_{365} of ~85% had been observed for **m2Br**, around 7-22% higher in comparison to non-methoxy-bearing azobenzenes. During follow-up investigations of this particular azo-ncAA however, only misincorporation was observed with the *Mm*PylRS AzoA (Supplementary Figure III.2). Therefore, additional mnX analogs were designed and synthesized, such as the azo-ncAA **m2Cl** (**24b**) featuring a less reactive linker in comparison to its bromide analog **m2Br**, and the benzyl chloride derivative **m1Cl** (**24a**). Identical to the pnX series, azo-coupling of **12** with the respective phenol **19a** or **b** yielded the azo-ncAAs **20a** and **b** (Scheme 3.7).



Scheme 3.7 | Synthesis route of azo-ncAAs m1Cl and m2Cl bearing *para*-methoxy moieties for improved PSS.

Bismethylation of phenol and carboxyl group with methyl iodide furnished azobenzenes **21a** and **b** in good to very good yields (78% and 86%, respectively). The aliphatic alcohol was then mesylated and reacted with LiCl by S_N2 reaction, resulting in the fully protected chlorides **22a** and **b** in decent yields (47% and 53%, respectively). Subsequent hydrolysis of the methyl ester and Boc-deprotection with TFA yielded the target azo-ncAAs **m1Cl** (**24a**) and **m2Cl** (**24b**) as TFA-salts.

While initial screenings for incorporation into sfGFP-N149TAG in *E. coli* looked promising for both azo-ncAAs using the PylRS AzoC (Figure 3.5), **m2Cl** was later on discarded since no successful crosslink could be observed in target proteins (Chapter 3.3.2, ClpP).

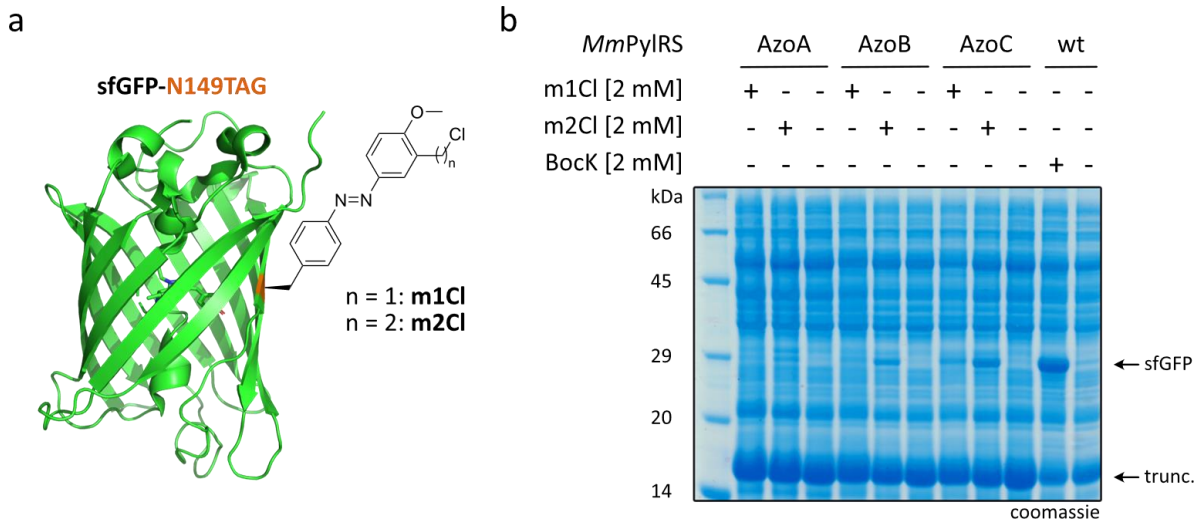


Figure 3.5 | Site-specific incorporation test for azo-ncAAs m1Cl and m2Cl into sfGFP-N149TAG-His$_6$ in *E. coli*. (**a**) Crystal structure of sfGFP (pdb: 2b3p) with the TAG-position (orange) bearing the azo-ncAA mnCl. (**b**) Coomassie-stained SDS-PAGE analysis of the azo-ncAAs **m1Cl** and **m2Cl**, tested for amber suppression of sfGFP-N149TAG using the PylRS variants AzoA-C. Boc-lysin (BocK), incorporated by the wild-type PylRS, was included as a positive control.

Unfortunately, the LR-MS results for **m1Cl** corresponded to a hydrolyzed version of the desired azo-ncAA (Figure 3.6a).

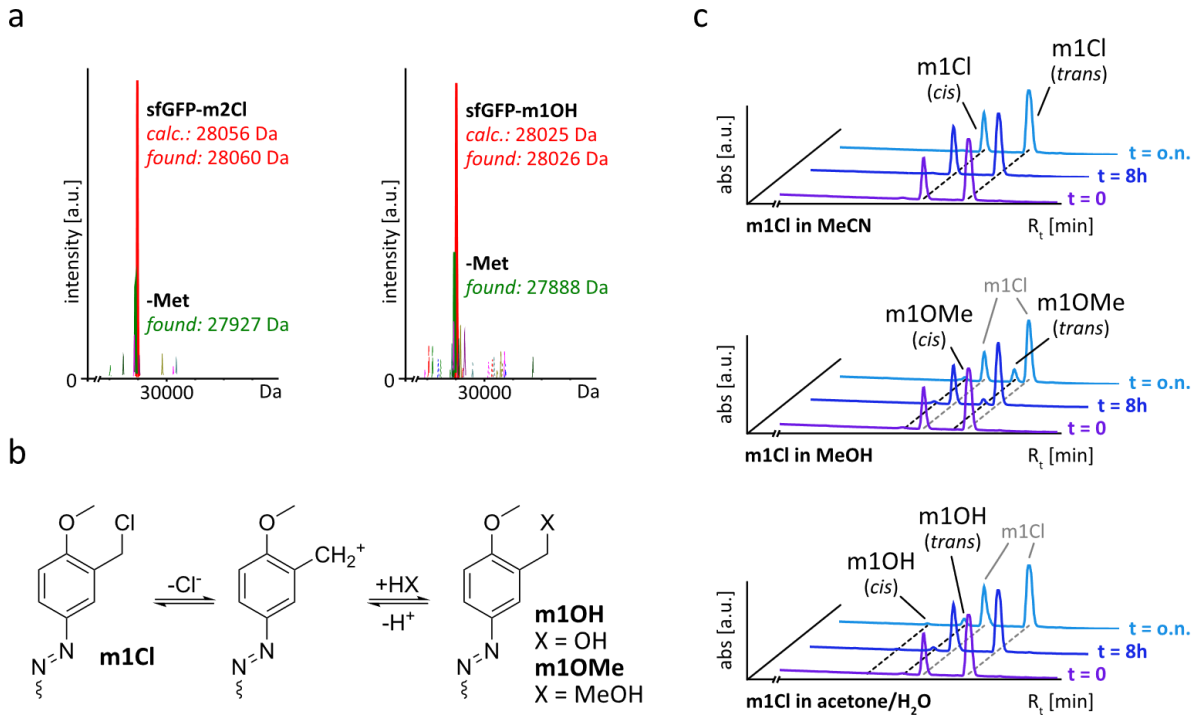
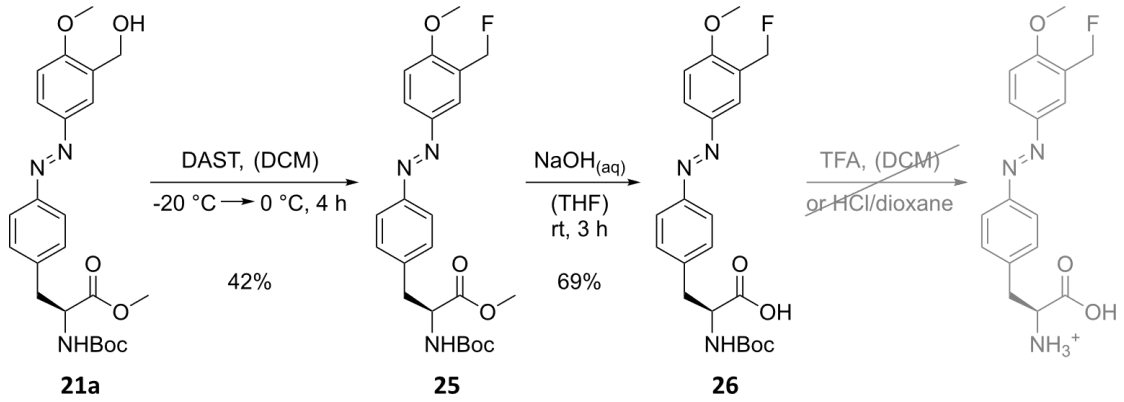


Figure 3.6 | Stability of incorporated azo-ncAAs m1Cl and m2Cl. (**a**) LR-MS results of sfGFP-N149TAG-His$_6$ amber suppressed with **m2Cl** and **m1Cl** using the PylRS AzoC. Full hydrolysis of the halide linker was observed for sfGFP-N149TAG-His$_6$ amber suppressed with **m1Cl**. (**b**) Proposed mechanism for the decomposition of the benzyl chloride when in a polar and protic environment. Following an S_N1-mechanism the resonance-stabilized benzylic cation is formed. Subsequent reaction with nucleophiles such as water or MeOH yields the benzylic alcohol and methyl ether, respectively. (**c**) LC-MS analysis of fully protected azo-ncAA Boc-m1Cl-OH in MeCN, MeOH or acetone/H_2O. Both isomeric forms, *cis* and *trans* could be observed in the obtained absorption spectra (λ = 193 nm). Incubation in the designated solvent at 37°C for 8 h and o.n. led to decomposition in both MeOH and acetone/H_2O.

In dielectric solvents the benzylic chloride may favor the resonance-stabilized benzylic cation, thus yielding the benzylic alcohol via S_N1-mechanism (Figure 3.6b). Additional stability assays of the fully protected azo-ncAA Boc-m1Cl-OMe in MeOH, MeCN or aqueous acetone confirmed these findings. Decomposition of the benzyl chloride could already be observed in both MeOH and aqueous acetone after eight hours (Figure 3.6c).

In order to generate a stable yet S_N2-reactive m1X analog, the bond strength between the benzylic carbon atom and the halide had to be increased. In other words, the leaving group character had to be reduced in comparison to the chloride. Thus, the azo-ncAA **m1F** (**29**) bearing a benzylic fluoride was designed (Scheme 3.8).

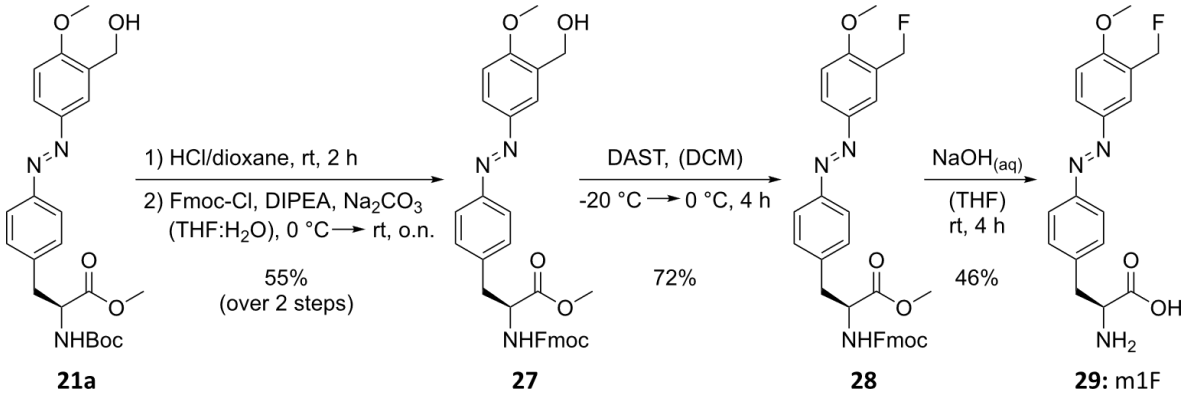


Scheme 3.8 | Failed synthesis route of azo-ncAA m1F. While the use of TFA yielded a mixture of trifluoroacetic acid ester and hydrolyzed linker, deprotection with 4 M HCl in dioxane led to a partial halide-exchange at the benzylic position. Due to nearly identical polarities, the hydrolysis-sensitive side-product **m1Cl** could not be separated from the desired azo-ncAA **m1F** (**29**).

Initial attempts at a successful synthesis of the target compound started from the intermediate azo-ncAA **21a**. Fluorination of the benzylic alcohol was performed with commercially available fluorinating reagent diethylaminosulfur trifluoride (DAST), giving the fully protected azo-ncAA **25** in fair yields of 42%.

Subsequent hydrolysis of the methyl ester yielded the azo-ncAA **26** without hydrolysis of the benzylic fluoride. Deprotection of the N_α-Boc protecting group using either TFA in DCM or 4 M HCl in dioxane both failed. Therefore, a different synthesis route towards **m1F** was explored.

Herein, the N_α-protecting group was exchanged for Fmoc prior to fluorination, since partial deprotection of the N_α-Fmoc group was observed during reduction of Fmoc-Phe-(4-NO_2)-OH with H_2 over Pd/C. Different reduction methods were explored but discarded in favor of the already established synthesis route towards the intermediate azobenzene **21a** (Scheme 3.7). First, the N_α-Boc protecting group of **21a** was removed using 4 M HCl in dioxane (Scheme 3.9).



Scheme 3.9 | Final synthesis of the azo-ncAA m1F bearing an S_N2-reactive linker. Removal of the N_α-Fmoc protecting group with aqueous NaOH in THF gave rise to a mixture of the desired azo-ncAA **29** and a cyclized azo-ncAA derivative, in which the deprotected amino functionality reacted with the benzylic fluoride. Purification via RP-HPLC yielded the azo-ncAA **m1F** (**29**).

The crude intermediate was then reacted with Fmoc-Cl to give the azo-ncAA **27** in decent yields of 55% over two steps (32% during scale-up). The fully protected azo-ncAA **28** was obtained in 72% yield (33% during scale-up) via standard fluorination with DAST. Parallel hydrolysis of the methyl ester and N_α-Fmoc protecting group with aqueous NaOH in THF gave the target azo-ncAA **m1F** (**29**) in decent yields of 46% after HPLC purification.

Successful incorporation of **m1F** (**29**) was observed using the *Mm*PylRS AzoB and AzoC. The HR-MS results confirmed the designed stability of m1F (Figure 3.7).

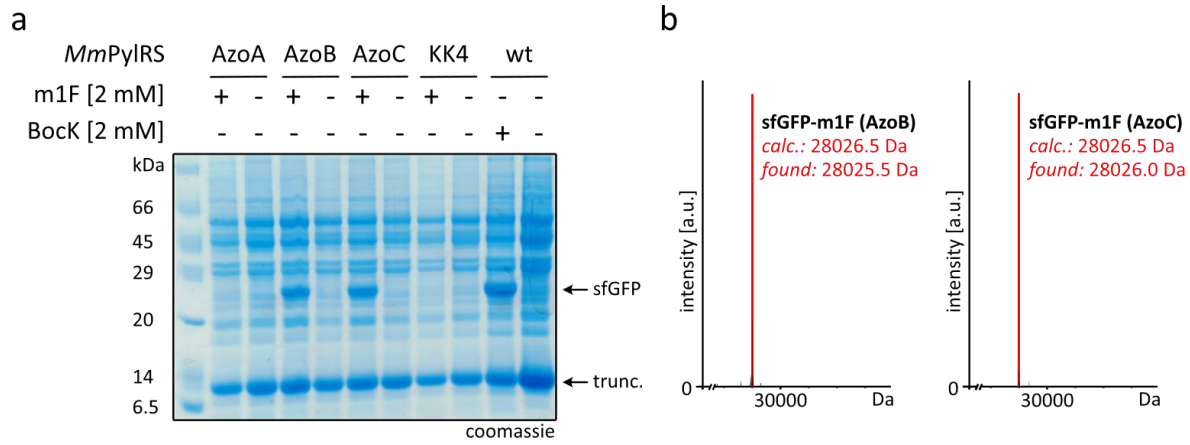


Figure 3.7 | | Amber suppression of sfGFP-N149TAG-His$_6$ with azo-ncAA m1F in *E. coli*. (**a**) Coomassie-stained SDS-PAGE analysis of **m1F** incorporation screening the PylRS variants AzoA-C and KK4. BocK, incorporated by the wt PylRS was used as positive control. (**b**) HR-MS analysis of **m1F**-bearing sfGFP variants amber suppressed with the PylRS AzoB (left panel) or AzoC (right panel).

Still, application of this particular azo-ncAA remained challenging throughout the project (Chapter 3.3). While solubility issues of the azo-ncAA in aqueous medium led to reduced amber suppression yields, stacking interactions of residual free azo-ncAA onto proteins, made protein purification particularly tedious and even led to precipitation of amber suppressed protein.

3.1.5 Development of Azo-ncAAs for Crosslinking via S_N2 and Michael Addition

The ideal azo-ncAA should be reactive enough to fully crosslink with a proximal cysteine yet stable enough to not get consumed by the abundance of nucleophiles within a cell. In order to find such azo-ncAAs, we wanted to explore a broad spectrum of linker reactivities, including more activated forms of S_N2 and Michael additions. For this purpose, two additional sets of azo-ncAAs were developed in parallel to the described pnX and mnX series. Again, the sets were either bearing the reactive moieties in *para*-position (pQX, Figure 3.8) or *meta*-position (mQR/mQX, Figure 3.9) to the diazene, linked via an aromatic amide bond.

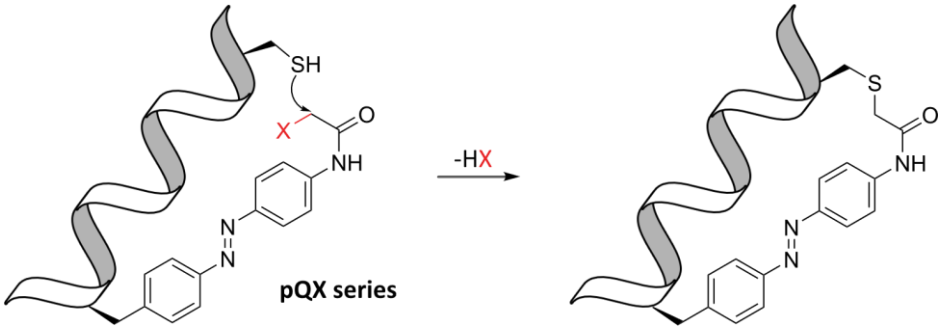
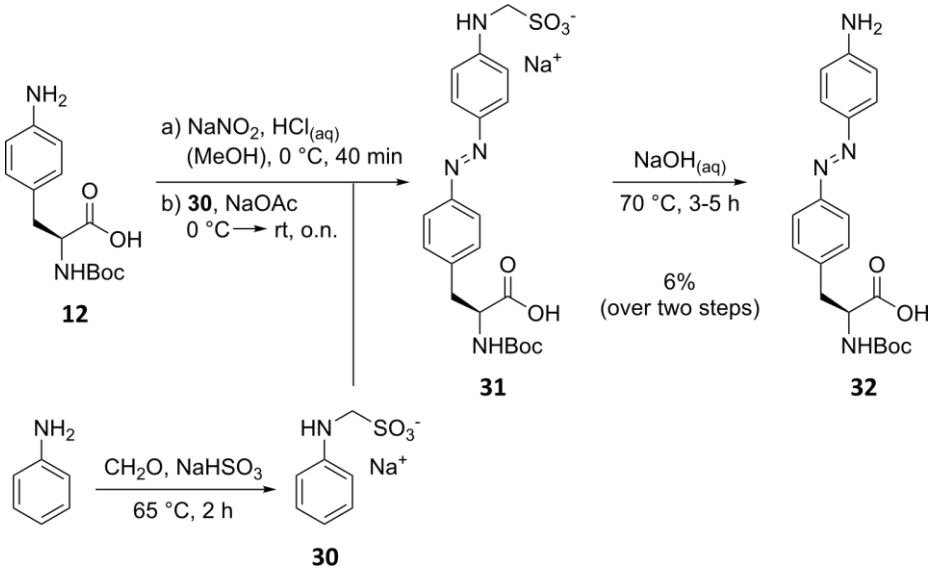


Figure 3.8 | Schematic depiction of the crosslinking reaction between an azo-ncAA of the pQX series and a proximal cysteine residue via activated S_N2.

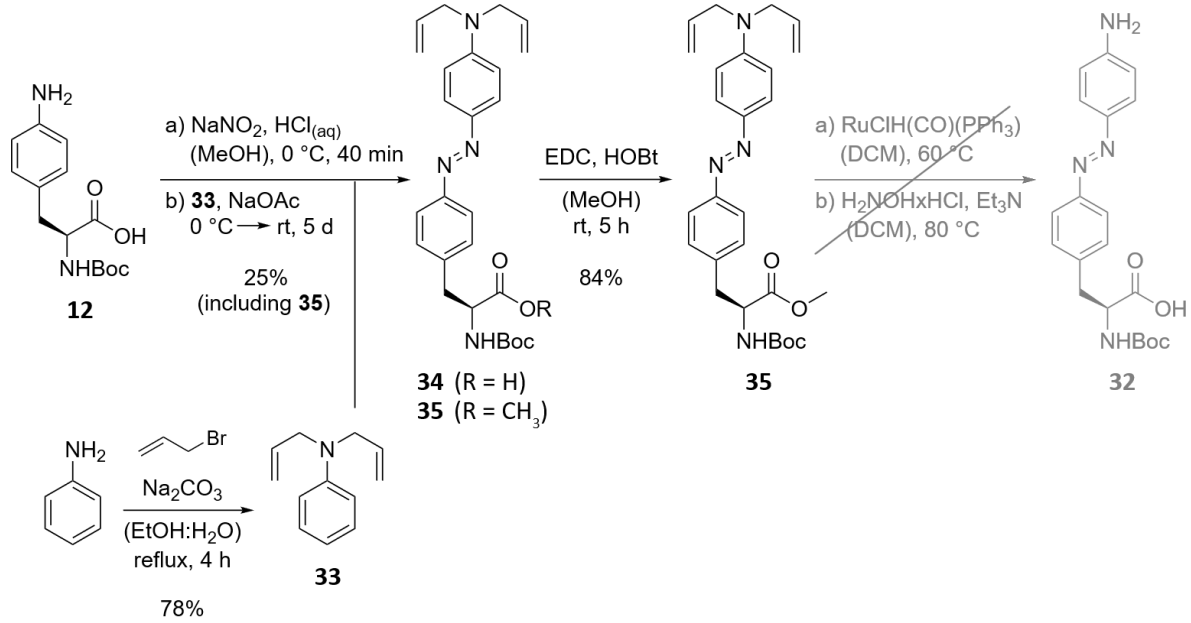
Since direct azo-coupling of aniline typically generates the triazene side-product, two different aniline protecting group strategies used for the synthesis of *para*-amino substituted azobenzenes were explored. One strategy uses anilinomethanesulfonate derivative **30** that was generated by reacting aniline with sodium bisulfite and formaldehyde (Scheme 3.10).[261]



Scheme 3.10 | Attempted synthesis route towards *para*-amino substituted azo-ncAA 32 via (phenylamino)methane-sulfonate.

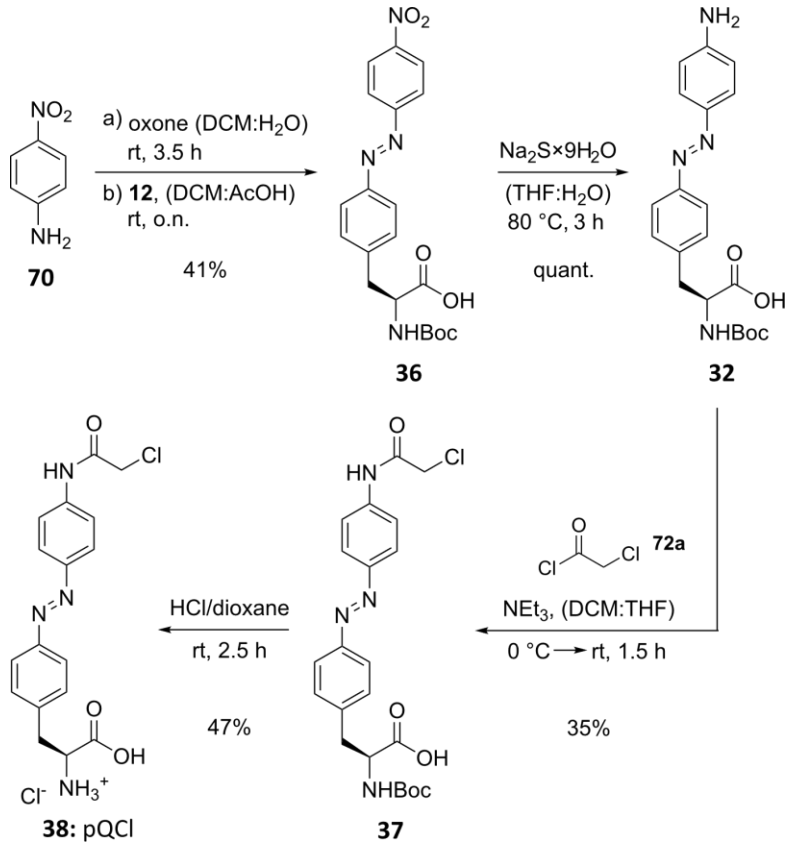
Although quantitative protection could be observed via LC-MS, *in situ* coupling with the diazonium ion of **12** was very low-yielding (6 %). It was suspected, that this observation was a result of the wrong pH during the coupling step. Azo-coupling of diazonium ions to anilines are in general much more pH-sensitive in comparison to phenols. Too high in pH and the aniline will participate in addition reactions, leading to triazene formation. If the pH is too low on the other hand, the aniline is not activated enough for S_EAr reactions with the diazonium. All attempts at improving the reaction conditions resulted in similarly poor yields, also due to decomposition of the diazonium ion and N_α-Boc deprotection of the target azo-ncAA.

A second strategy relying on the installation of an activating yet *N*-protected substituent to avoid triazene formation, employs the *N*,*N*-diallyl protected aniline **33** (Scheme 3.11).[262]



Scheme 3.11 | Failed synthesis route towards *para*-amino substituted azo-ncAA 32 via *N*,*N*-diallyl aniline. During the azo-coupling of **12** and **33** the respective methyl ester by-product formed (**35**), which was combined later on.

Although protection of the aniline gave rise to the desired product **33** in high yields, subsequent coupling to the diazonium ion was again very slow (>5 days) and low-yielding (25%). Protection of the carboxyl moiety of **34** employing EDC and HOBt in MeOH yielded the intermediate azobenzene **35**. To our disappointment, deallylation of the protected aromatic amine via ruthenium(II)-catalyzed isomerization to the divinyl amine and its subsequent hydrolysis failed repeatedly. Therefore, a different strategy was developed that would circumvent the low-yielding azo-coupling step to an aniline derivative entirely (Scheme 3.12).



Scheme 3.12 | Final synthesis route towards *para*-amino substituted azo-ncAA 32 and the target azo-ncAA pQCl bearing an S_N2-reactive linker.

Masking of the essential amino group as a nitro group finally enabled the facile synthesis of the azo-ncAA **32** via the Baeyer-Mills reaction and subsequent reduction using Na_2S in H_2O, leaving the diazene bond intact.[263] Reaction of **32** with 2-chloroacetyl chloride led to the formation of azo-ncAA **37** in fair yields (35%). Deprotection of the N_α-Boc group yielded the desired azo-ncAA **pQCl** (**38**) as an HCl-salt, bearing the S_N2-reactive linker in *para*-position to the diazene bond.

Again, no suitable PylRS variant was found during incorporation screening, similarly as for other *para*-linker bearing azo-ncAAs (Scheme 3.6). Thus, no additional linkers were investigated for this particular azo-ncAA scaffold and the following efforts focused instead on azo-ncAAs bearing the linker in the *meta*-position (Figure 3.9).

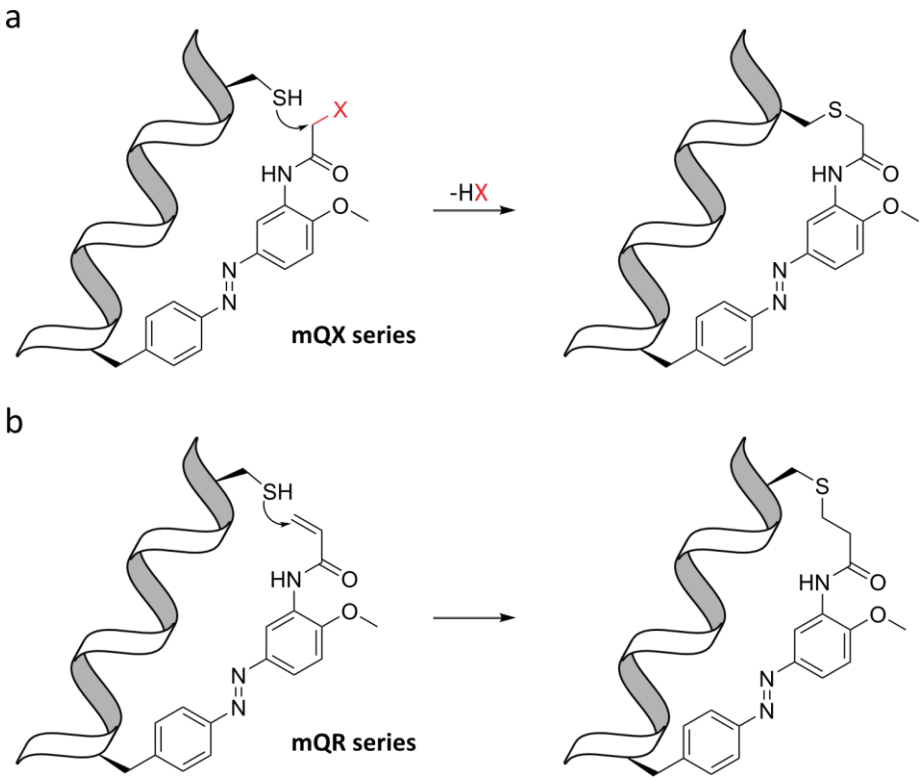
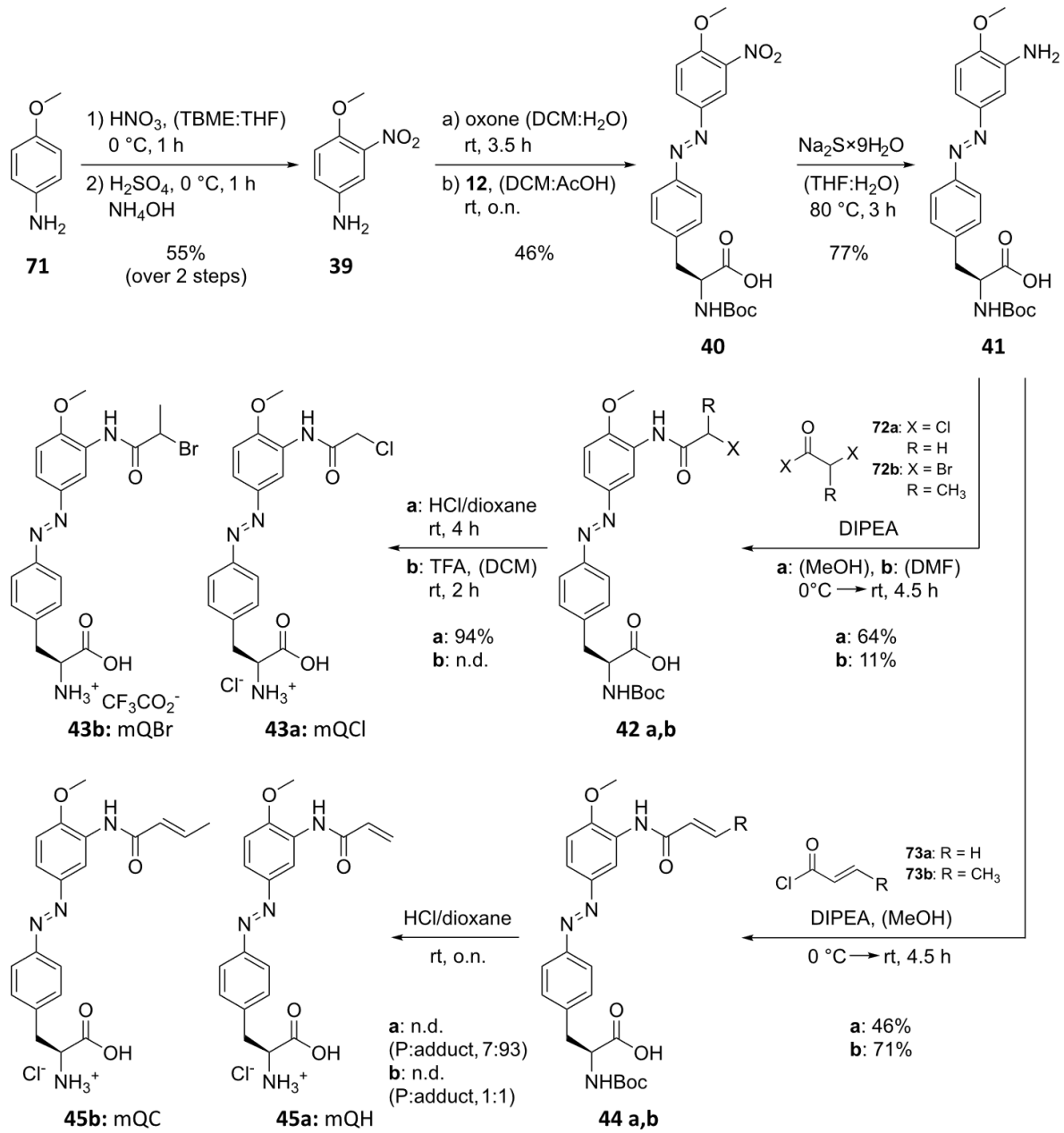


Figure 3.9 | Schematic depiction of the crosslinking reaction between an azo-ncAA and a proximal cysteine residue via activated S_N2 (a) or Michael addition (b). The proximity-induced reaction of an incorporated azo-ncAA with a cysteine residue is shown for the mQX series (**a**) and the mQR series (**b**), respectively.

Since masking of the aromatic amine as a nitro group worked well for the preparation of *para*-amino substituted azo-ncAA **32**, the same approach was pursued for a series of *meta*-substituted azo-ncAAs mQR and mQX. To this end, 4-methoxy aniline (**71**) was first regioselectively nitrated in *para*-position to the methoxy substituent (Scheme 3.13). Precipitation of the aniline as a nitrate salt prior to nitration did not only eliminate the *ortho*-directing qualities of the amine substituent during S_EAr,[264] but it also prevented over-nitration through *in situ* generation of equimolar amounts of nitronium ions, giving the desired aniline **39** in 55% yield. The azo-ncAA **40** was obtained via the Baeyer-Mills reaction in decent yields (46%). The aromatic nitro moiety was selectively reduced with Na_2S and the intermediate azo-ncAA **41** reacted with various acetyl halide linkers, yielding the various target compounds **42a** and **b** (mQX), and **44a** and **b** (mQR). Attempts at installing a butynoyl linker failed due to instability of the introduced linker. Acidic N_α-Boc deprotection yielded the desired azo-ncAAs **mQCl** (**43a**), **mQBr** (**43b**), **mQH** (**45a**) and **mQC** (**45b**). For the Michael acceptor bearing azo-ncAAs **45a** and **b** (mQR) a defined mixture of product and chloro-adduct (7:93 and 1:1, respectively) could be observed after o.n. deprotection with 4 M HCl in dioxane. Shorter reaction times would most likely improve the ratio between intact Michael acceptor and chloro-adduct, if not prevent formation of the chloro-adduct completely.



Scheme 3.13 | Synthesis of azo-ncAAs mQCl and mQBr for S_N2, and mQH and mQC for Michael addition. The side-product observed during deprotection of **44a** and **b** was identified as the respective chloro-adduct of the Michael acceptor. Ratios are given as product:chloro-adduct (P:adduct).

Prior to potential improvements of the synthesis route, the obtained azo-ncAAs were tested for incorporation into sfGFP-N149TAG in *E. coli*. Although *Mm*PylRS KK11 yielded at least detectable amounts of amber suppressed protein (Figure 3.13a, Chapter 3.1.6), subsequent analysis via LR-MS revealed cleavage of the aromatic amide bond by unspecific peptidases and phenylalanine misincorporation (Supplementary Figure III.3).

Not only were the mQX and mQR derivatives bad substrates but they also seemed to promote the incorporation of a natural amino acid with *Mm*PylRS KK11. This came as a surprise, since this particular *Mm*PylRS variant was highly specific when using different substrates (Chapter 3.1.7).

3.1.6 Development of Azo-ncAAs for Crosslinking via S_NAr Reaction

In addition to reactive linker-bearing azo-ncAAs, a set of azo-ncAAs for S_NAr was developed (Figure 3.10).

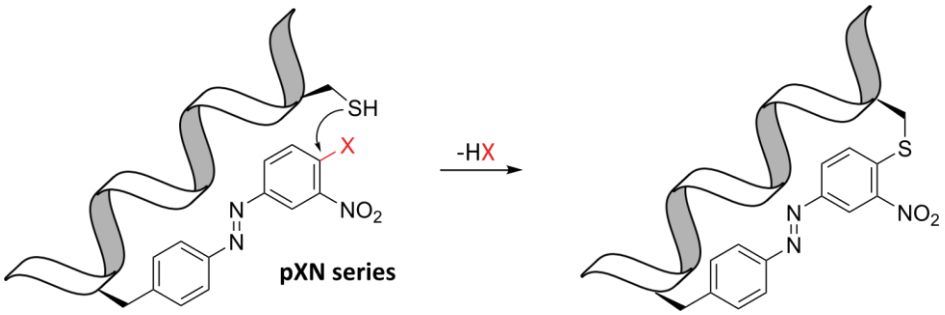


Figure 3.10 | Schematic depiction of the crosslinking reaction between an azo-ncAA of the pXN series and a proximal cysteine residue via S_NAr.

Prior work resulted in the azo-ncAA **TriF** (**46**), which was decently incorporated into sfGFP-N149TAG using the *Mm*PylRS AzoB (Figure 3.11).

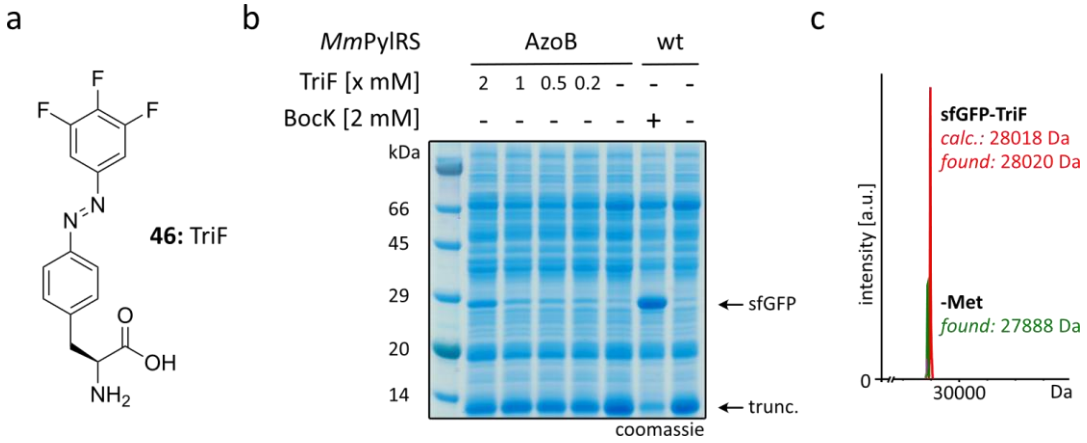
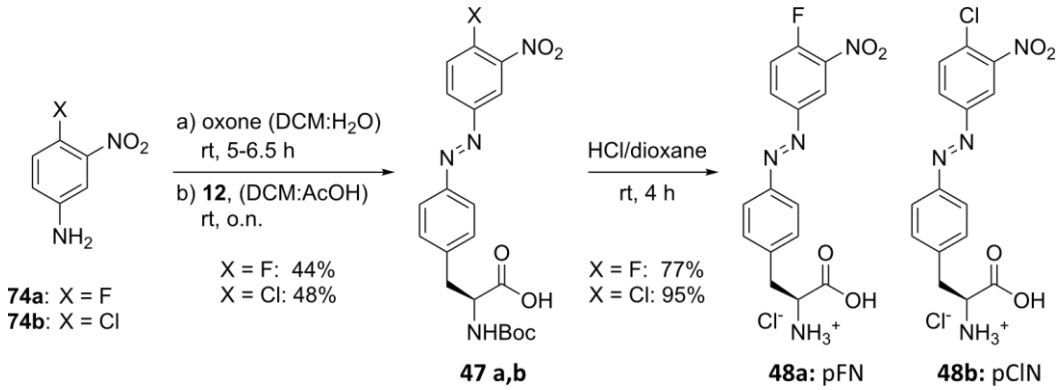


Figure 3.11 | Amber suppression of sfGFP-N149TAG-His_6 with azo-ncAA TriF in *E. coli*. (**a**) Structural formula of the azo-ncAA **TriF** (**46**) for S_NAr. (**b**) Coomassie-stained SDS-PAGE analysis of **TriF** incorporation by the PylRS AzoB using different concentrations of azo-ncAA (0.2 – 2 mM). BocK, incorporated by the wt PylRS was used as positive control. (**c**) LR-MS analysis of TriF-bearing sfGFP produced using the PylRS AzoB.

In order to generate a set of azo-ncAAs for S_NAr featuring a variable range of reactivity towards proximal cysteines, two pXN derivatives were designed (Scheme 3.14).



Scheme 3.14 | Synthesis of azo-ncAAs pFN and pClN bearing S_NAr-reactive linkers.

Coupling of commercially available 4-halide 3-nitroaniline derivatives and **12** via the Baeyer-Mills reaction furnished azobenzenes **47a** and **b** in fair yields of 44% and 48%, respectively. Subsequent N_α-Boc deprotection with 4 M HCl in dioxane afforded the desired azo-ncAAs **pFN** (**48a**) and **pClN** (**48b**) as HCl-salts.

A synthetase screening in *E. coli* identified two *Mm*PylRS, AzoB and AzoC that successfully incorporated both **pClN** and **pFN** (Figure 3.12a).

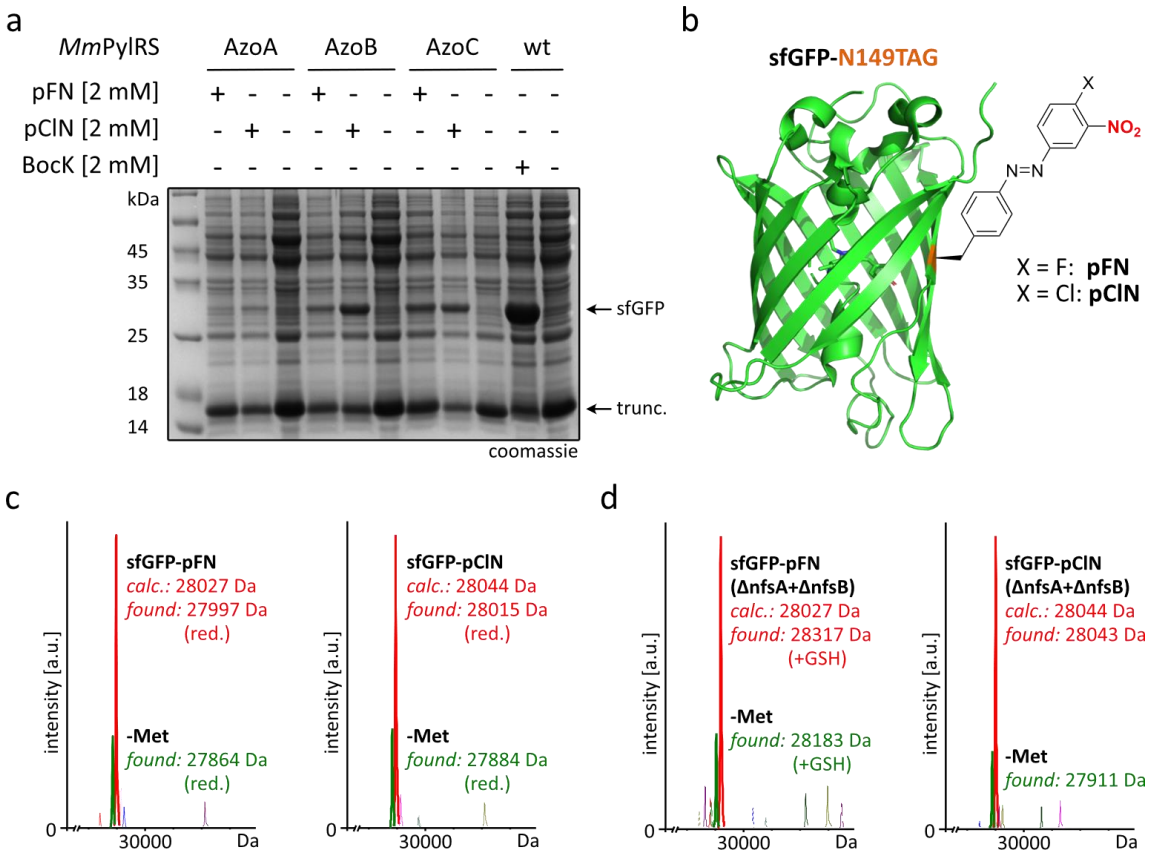


Figure 3.12 | Amber suppression of sfGFP-N149TAG-His$_6$ with azo-ncAAs pFN and pClN in *E. coli*. (**a**) Coomassie-stained SDS-PAGE analysis of pXN incorporation. BocK was used as positive control. (**b**) Crystal structure of sfGFP (pdb: 2b3p) with the TAG-position (orange) bearing the azo-ncAA derivatives pXN. The reduction-sensitive nitro-moiety is shown in red. (**c**) LR-MS analysis of pXN-bearing sfGFP produced using the PylRS AzoC. (**d**) LR-MS analysis of sfGFP amber suppressed with pXN in the *E. coli* strain K12 (ΔnfsA + ΔnfsB).

Unfortunately, a reduction of the nitro moiety could be observed via LR-MS (Figure 3.12c), rendering the pXN derivatives unreactive for S_NAr. Since primary amine groups are not easily accepted by the PylRS, it was suspected that endogenous nitro reductases of *E. coli* were post-translationally reducing the aromatic nitro groups. To this end, a nitro-reductase knockout strain, reported to show stable incorporation of nitro-substituted aromatic ncAAs, was investigated more closely. The *E. coli* strain referred to as B-95.ΔA comprises knockouts of six genes in total, namely *nfsA, nfsB, azoR, ydjA, nemA* and *rutE*, all based on the collection of carefully designed Keio single gene knockouts.[265, 266]

Some of the described Keio single gene knockouts were already available within the lab as a double (ΔnfsA, ΔnfsB) and triple knockout (ΔnfsA, ΔnfsB, ΔazoR). In the hope that the expression of one of those genes was responsible for the *in vivo* reduction of the aromatic nitro moieties, the in-house *E. coli* K12 strain (ΔnfsA, ΔnfsB) was tested first. To our satisfaction, the LR-MS data obtained from expression of amber suppressed sfGFP-N149TAG in the designated *E. coli* knockout strain did indeed correspond to intact nitro groups (Figure 3.12d). These findings further confirmed our suspicions that the pXN derivatives (**48a** and **b**) were getting reduced after ribosomal translation, since no increase in truncation levels was observed. Later experiments utilizing Keio single gene knockouts ΔnfsA or ΔnfsB did show both nitro reductases to be responsible for aromatic nitro reduction (Supplementary Figure III.4).

Notably, the more reactive azo-ncAA **pFN** had fully reacted with the cellular redox-regulator GSH, forming the respective adduct via S_NAr (Figure 3.12d). These results however did not provide information on the time frame in which the azo-ncAA got consumed by GSH.

Appropriate placement of **pFN** in proximity to a cysteine amino acid residue might very well promote full crosslinking before reaction with GSH (Chapter 3.3.1).

With the developed azo-ncAAs in hand, we set out to further improve their photophysical properties for the optical control of protein function. Key features, such as a high PSS and photoswitching utilizing visible light could both potentially be addressed through tetra-*ortho* halogenation of the developed azo-ncAAs. Further, we suspected that a published *Mb*PylRS mutant (L270F, L274M, N311G, C313G and Y349F) known to effectively incorporate the tetra-*ortho* fluorinated azo-ncAA F_4AzoF[251], would also accept other *ortho*-halogenated azo-ncAA scaffolds. Since incorporation of the standard azo-ncAA AzoF has been demonstrated using the same *Mb*PylRS variant, a range of our developed azo-ncAAs were tested for amber suppression of sfGFP-N149TAG with the respective *Mm*PylRS analog KK11 (Table 3.1).The *Mm*PylRS KK11 showed exceptional incorporation efficiencies for several azo-ncAAs including **m1F**, **TriF**, **pFN** and **pClN** (Figure 3.13a).

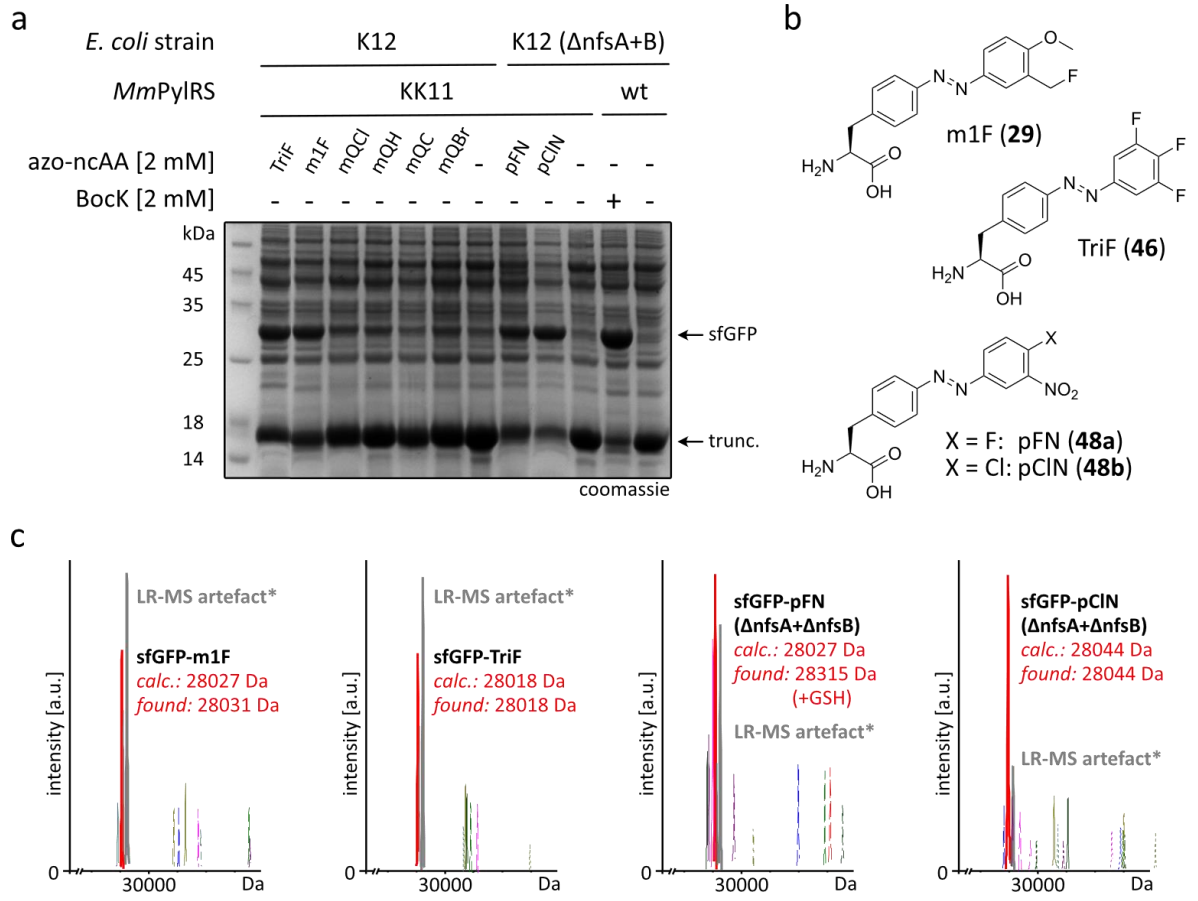


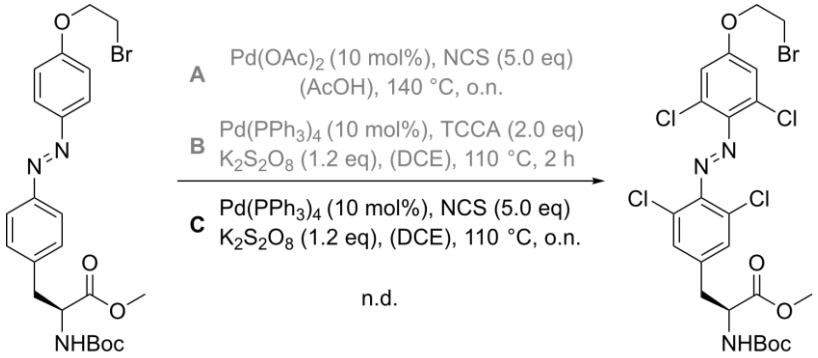
Figure 3.13 | Amber suppression of sfGFP-N149TAG-His_6 with various azo-ncAAs in *E. coli*. (**a**) Coomassie-stained SDS-PAGE analysis of azo-ncAA incorporation by *Mm*PylRS KK11 in the *E. coli* strains K12 or K12 (ΔnfsA + ΔnfsB). BocK, incorporated by the wt PylRS was used as positive control. (**b**) Structural formula of the azo-ncAAs **m1F**, **TriF**, **pFN** and **pClN**. (**c**) LR-MS analysis of amber suppressed sfGFP with the azo-ncAAs **m1F**, **TriF**, **pFN** and **pClN**. LR-MS artefacts* (+ 200 Da) were observed due to LC-MS instrument errors and are shown in grey.

To our satisfaction, LR-MS analysis of the respective amber suppressed sfGFP proteins confirmed the desired incorporation of all four azo-ncAAs (Figure 3.13b, c).

3.1.7 Development of Azo-ncAAs Bearing tetra-*ortho* Halogenations

Encouraged by these results using the *Mm*PylRS mutant KK11, we set out to synthesize tetra-*ortho* chlorinated and tetra-*ortho* fluorinated crosslinking azo-ncAAs. Although a range of synthetic methods are available for the preparation of tetra-*ortho* substituted symmetrical azobenzenes[267], synthesis routes towards non-symmetrical analogs are still scarce. Two promising publications providing high-yielding access to tetra-*ortho* chlorinated azo-benzenes via late-stage functionalization were investigated for the synthesis of respective azo-ncAAs.[268, 269] Both approaches employ Pd(0)-catalysts that coordinate one nitrogen of the diazene bond as oxidized Pd(II) species and thus selectively form C-Cl bonds in the *ortho*-positions of the azobenzene. Typical sources of chlorine radicals employed in such reactions are NCS and TCCA. The application of either synthetic strategy to the late-stage functionalization of N_α-Boc protected azo-ncAAs however, yielded intractable mixtures of di-, tri-, tetra- and even penta-halogenated derivatives. Various conditions had to be screened (Supplementary Table IV.1) to establish a suitable procedure for the synthesis of tetra-*ortho* chlorinated crosslinking azo-ncAAs.

Selective tetra-*ortho* chlorination of the azo-ncAA **16c** was finally observed using 10 mol-% $Pd(PPh_3)_4$ as catalyst and 5 eq NCS in dry DCE (Scheme 3.15).



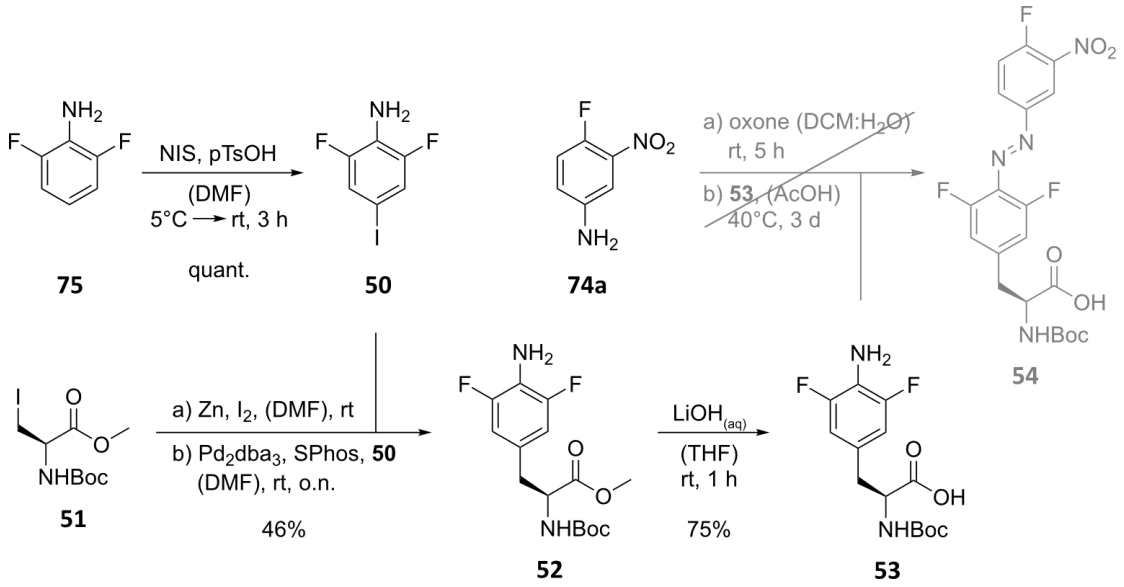
16c **49**

Scheme 3.15 | Failed (A and B) and final conditions (C) for the preparation of tetra-*ortho* chlorinated azo-ncAA 49.

The oxidant $K_2S_2O_8$ was added for *in situ* generation of chlorine radicals from NCS. To our disappointment, no pure product could be obtained after purification via flash column chromatography (FCC), presumably due to triphenylphosphine oxide. Complexation of this by-product prior to FCC would significantly improve the purity of tetra-*ortho* chlorinated products obtained by the described procedure.[270] Beyond that, scale up of the reaction with the established conditions led to reaction mixtures of tri-, tetra- and penta-chlorinated product. Additional amounts of NCS and $K_2S_2O_8$ and/or prolonged reaction times gave similar results with increasing amounts of N_α-Boc deprotected azo-ncAA. The development of tetra-*ortho* chlorinated azo-ncAAs was therefore set aside in favor of the fluorinated analogs.

In contrast to *ortho*-chlorination no strategies are available for late-stage *ortho*-fluorinations of azobenzenes. The appropriate building blocks have to be prepared prior to azobenzene formation, limiting the scope of azo-ncAA derivatives. Further, the use of *ortho*-fluorinated anilines immensely slows down the Baeyer Mills reaction that is typically used for the preparation of such azobenzenes. In general, elevated temperatures and prolonged reaction times are necessary to obtain the *ortho*-fluorinated azobenzenes in acceptable yields.[82, 251]

To obtain the fluorinated building blocks, commercially available 2,6-difluoro aniline was first reacted with NIS, thus yielding the desired iodo-arene **50** (Scheme 3.16).[71]



Scheme 3.16 | Synthesis route towards tetra-*ortho* fluorinated azo-ncAA 54. Coupling of **53** to 4-fluoro-3-nitroaniline via the Baeyer-Mills reaction failed so far and is still under investigation.

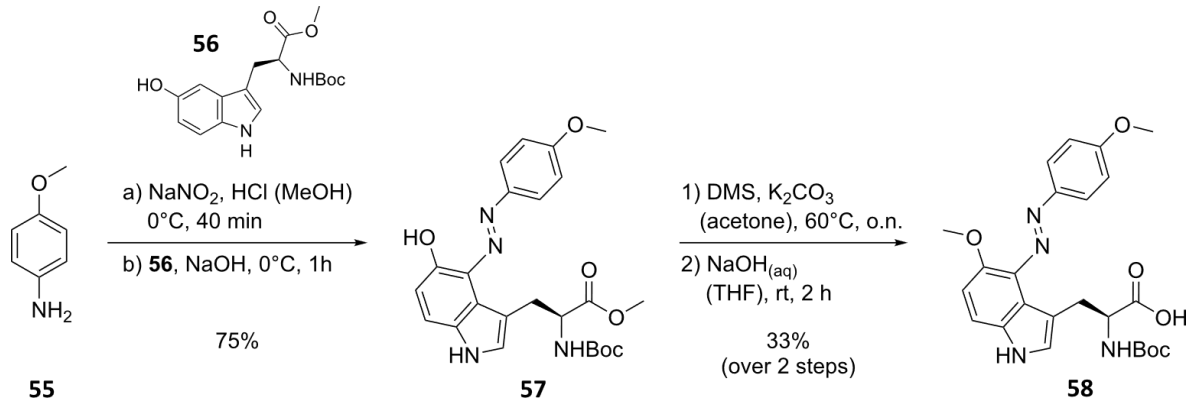
Subsequent Negishi cross-coupling of **50** with N_α-Boc-Ala(3-I)-OMe (**51**) employing a combination of $Pd_2(dba)_3$ and SPhos (1:2 molar ratio), afforded the fully protected phenylalanine derivative **52** in decent yields of 46%.[271] The methyl ester was hydrolyzed and the resulting aniline derivative **53** used in a Baeyer-Mills reaction. Attempts at coupling **53** to 4-fluoro-3-nitroaniline in glacial AcOH over prolonged reaction times and at elevated temperatures failed so far. Changing the solvent system to a reported mixture of toluene:AcOH:TFA (40:40:6) was also unsuccessful in yielding the desired azo-ncAA. Thus, future activities will be focusing on improving the Baeyer-Mills reaction for the preparation of tetra-*ortho* fluorinated azo-ncAAs.

3.1.8 Synthesis of Model Trp-based Azo-ncAA

Azoheteroarenes present an attractive choice of photoswitch if high thermal stability of the *cis*-isomer is desired. Additionally, such azobenzenes typically feature well separated π-π* and n-π* bands, leading to exceptional PSSs in both directions. Although azoheteroarenes based on 5-membered heterocycles such as pyrazole are the most well-studied[272], arylazoindoles[273] and arylazo-1,2,3-triazoles[274] have received some attention as of late.

The arylazoindole scaffold was of particular interest to us, since we envisioned potential incorporation by the *Ec*TrpRS/tRNA pair in an engineered *E. coli* strain. Chatterjee and co-workers successfully demonstrated the site-specific incorporation of Trp analogs such as 5-methoxy-Trp (5MTP) by polyspecific *Ec*TrpRS variants.[177, 275] Encouraged by these results we designed a potentially bistable azo-ncAA based on the 5MTP scaffold, that would ideally exhibit high thermal stability while the *ortho*-methoxy substitution would shift the absorption of the chromophore towards wavelengths in the visible region.

The Trp-based azo-ncAA **57** was obtained in good yields of 75% by azo-coupling of the diazonium-ion of **55** to the N_α-Boc protected methyl ester of 5MTP-OMe (**56**) (Scheme 3.17). Selective methylation of the phenol over the indole nitrogen required screening of various methylation reaction conditions. Methylation by MeI with different bases such as K_2CO_3, $CsCO_3$ and $AgCO_3$ was either too sluggish or giving intractable mixtures of the desired product and *N*-methylated indole species. Finally, DMS and K_2CO_3 in acetone yielded the mono-methylated Trp-based azo-ncAA. Hydrolysis of the methyl ester with aqueous NaOH gave the target compound **58** in fair yields of 33% over two steps.



Scheme 3.17 | Synthesis of Trp-based azo-ncAA 58 as a model compound for azoheteroarenes.

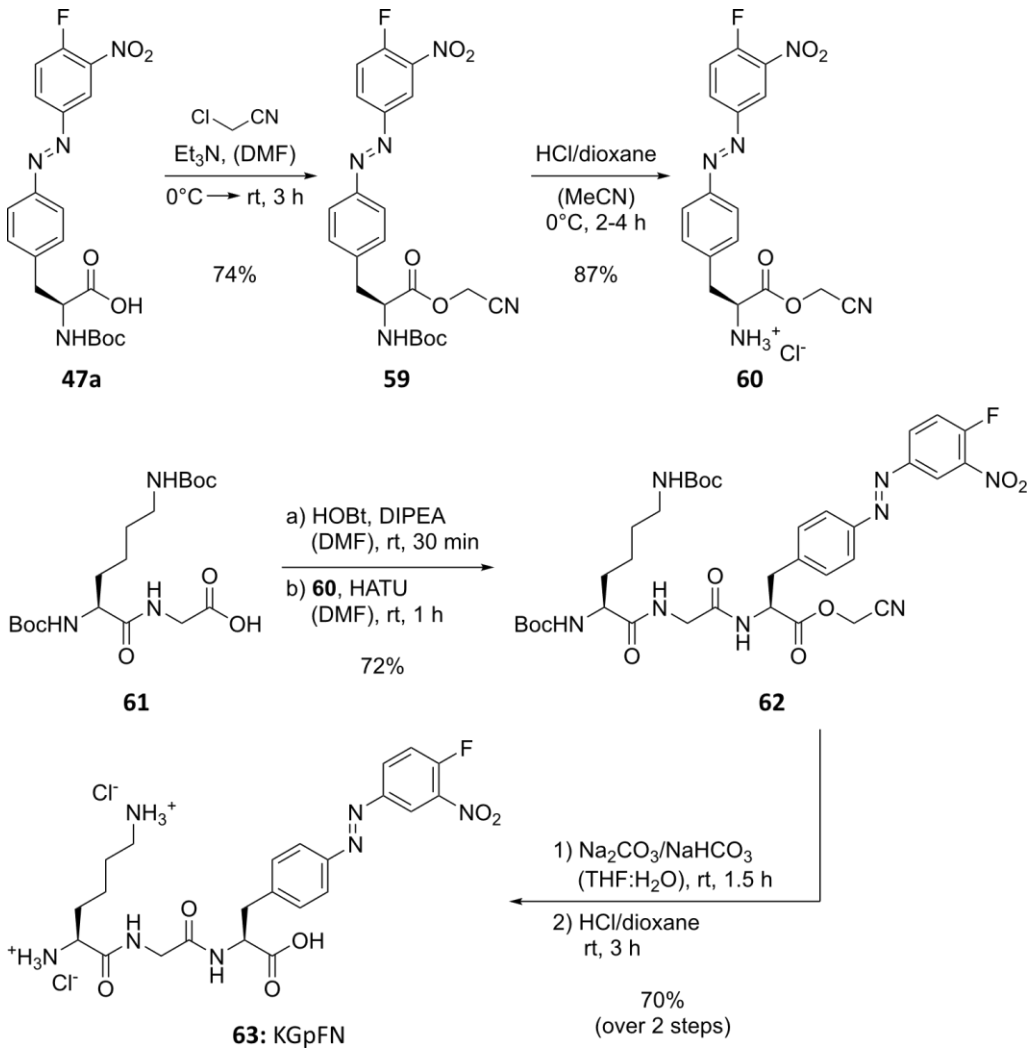
To our disappointment, this compound exhibited rather unusual photophysical properties (Chapter 3.2, Supplementary Figure III.6). As a consequence, this azo-ncAA scaffold was not pursued any further for the development of crosslinking azo-ncAAs.

3.1.9 A Tripeptide Bearing Azo-ncAAs for Amber Suppression

In order to increase the yields of amber suppressed target protein with the found azo-ncAAs **m1F**, **pFN** and **pClN**, we first turned our attention to the highly active *Ma*PylRS analog. The key advantage of *Ma*PylRS over full-length *Mm*PylRS is its greater solubility. This is attributed to the lack of an *N*-terminal domain in *Ma*PylRS, resulting in higher amber suppression yields in comparison to the *Mm*PylRS. Unfortunately, attempts at amber-suppressing sfGFP-N149TAG by using the analog *Ma*PylRS variants (Supplementary Table IV.2) resulted in protein truncation.

We instead focused on directly improving the aminoacylation activity of the identified *Mm*PylRS variants. Recent reports yielded more robust *Mm*PylRS variants by introduction of a single point mutation (P188G).[276] This way the *N*-terminal domain is no longer cleaved from the full-length protein, thus greatly enhancing incorporation of various ncAAs. We therefore created *Mm*PylRS analogs of the wt PylRS and KK11, bearing the additional point mutation P188G. While similar incorporation levels could be observed for the positive control BocK, introduction of the point mutation P188G into KK11 seemed to completely abolish its aminoacylation activity (Supplementary Figure III.5).

Instead, a different strategy to enhance amber suppression yields was explored. We anticipated, that by improving the solubility of our poorly water-soluble azo-ncAAs we would also increase their bioavailability. Based on results obtained from prior work on this project, we expected that installation of our azo-ncAAs into lysine-bearing tripeptide scaffolds such as KG-ncAA would greatly improve their solubility in water. To this end we synthesized a tripeptide bearing the promising azo-ncAA **pFN** (Scheme 3.18). Protection of the carboxyl moiety in N_α-Boc-protected **pFN** (**47a**) as cyanomethyl ester allowed for selective N_α-Boc deprotection, giving the azo-ncAA **60** in decent yields (64% over two steps). Standard Fmoc-SPPS afforded the dipeptide Boc-Lys(Boc)-Gly-OH (**61**), which was coupled to the intermediate azo-ncAA **60** utilizing HOBt and HATU, thus yielding the fully protected tripeptide **KGpFN** (**62**). The target tripeptide **KGpFN** (**63**) was obtained by hydrolysis of the cyanomethyl ester and subsequent N_α-Boc deprotection in good yields (70% over two steps).



Scheme 3.18 | Synthesis of the tripeptide KGpFN (63) bearing the azo-ncAA pFN.

To estimate whether the use of tripeptide **KGpFN** would result in enhanced amber suppression yields compared to the single azo-ncAA **pFN**, both substrates were tested for incorporation into sfGFP-N149TAG using PylRS KK11 (Figure 3.14).

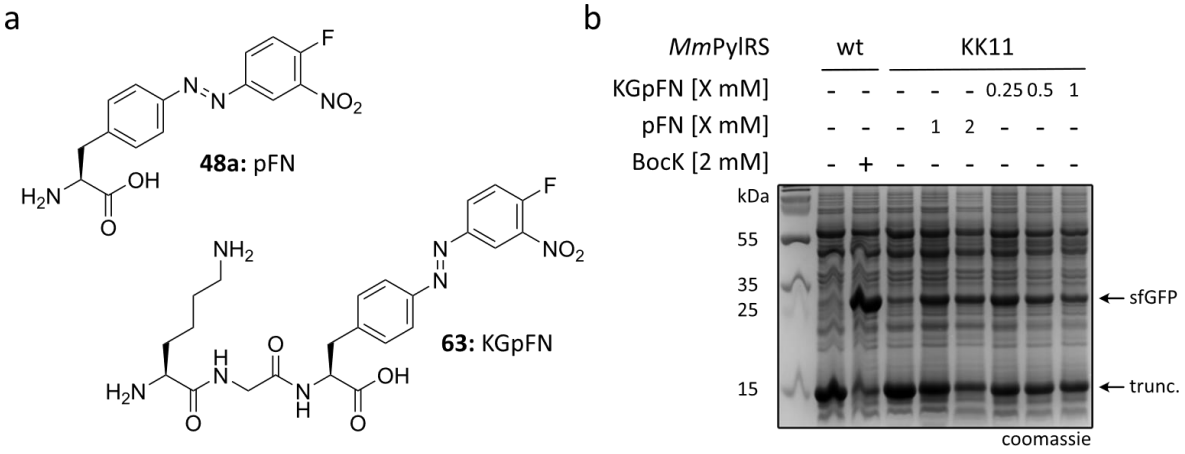


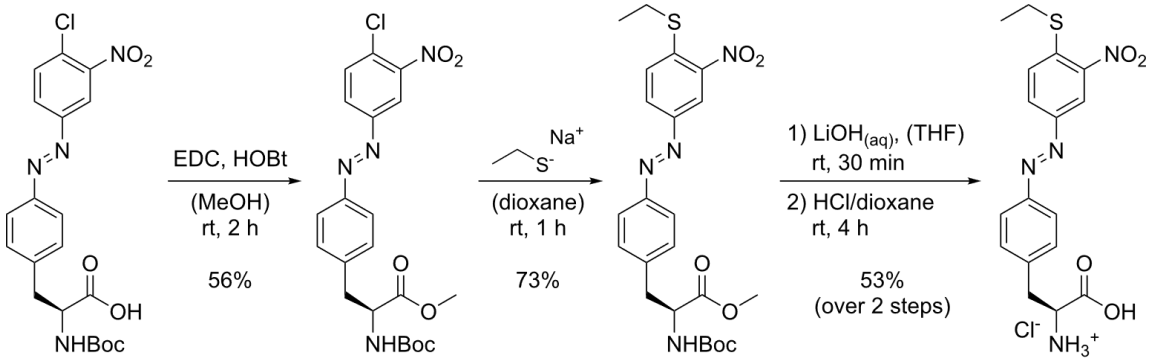
Figure 3.14 | Amber suppression of sfGFP-N149TAG-His$_6$ with the azo-ncAA pFN (48a) and tripeptide KGpFN (63) in *E. coli*. (**a**) Structural formula of the azo-ncAAs **pFN** and the tripeptide **KGpFN**. (**b**) Coomassie-stained SDS-PAGE analysis of pFN incorporation by the PylRS KK11 using different concentrations of **pFN** (1 or 2 mM) and **KGpFN** (0.25 – 1 mM). BocK, incorporated by the wt PylRS was used as positive control.

First, a 100 mM stock of **KGpFN** in H_2O was prepared. The tripeptide remained completely dissolved even after addition to the medium, while its ncAA counterpart **pFN** heavily precipitated. Overall, incorporation screenings in *E. coli* revealed similar expression levels of amber suppressed sfGFP-N149TAG for both **KGpFN** and **pFN**. In comparison to the single azo-ncAA however, substantially less substrate was needed (0.25-0.5 mM) when using the tripeptide **KGpFN**. Additionally, the use of the water-soluble tripeptide simplified protein purification, since no aggregates between typically precipitated azo-ncAA and amber suppressed protein were formed. The tripeptide-based approach thus provides a promising strategy to improve amber suppression yields while requiring less substrate. Unfortunately, the multi-step synthesis of **KGpFN** is rather low-yielding in comparison to **pFN**, thereby negating this advantageous effect. Consequently, we focused on the single azo-ncAA **pFN** within the scope of this work.

3.2 Photophysical Characterization

To analyze photoswitching and thermal relaxation in bulk samples, the most promising azo-ncAA scaffolds mnX, pXN and the model arylazoindole **58** were characterized by UV-Vis spectrophotometry as specified in chapter I.9.2.

UV-Vis analysis of the azoheteroarene model compound **58** showed a very unusual absorption spectrum (Supplementary Figure III.6). Although a strong visible band absorbing around 480 nm could observed, no typical π-π* and n-π* bands could be identified. Further, the absorption spectrum of the azoheteroarene **58** did not change upon irradiation with different wavelengths (λ = 365, 450 and 505 nm). This may be a result of the diazene being linked to the indole moiety in position 4, even if no comparable examples confirming our suspicions could be found in the literature. We instead focused on the more promising mnX and pnX derivatives. While the azo-ncAA **m1F** was characterized as exemplary azo-ncAA of the mnX series, reaction of the pNX derivatives with a proximal cysteine via S_NAr is directly influencing the diazene chromophore. Therefore, a model compound **pSEtN** bearing a thioether group in *para*-position to the diazene was synthesized for photocharacterization (Scheme 3.19).



47b **64** **65** **66:** pSEtN

Scheme 3.19 | Synthesis of the model azo-ncAA pSEtN (61) for photophysical characterization.

The respective absorption spectra of **m1F** and **pSEtN** are shown in Figure 3.15.

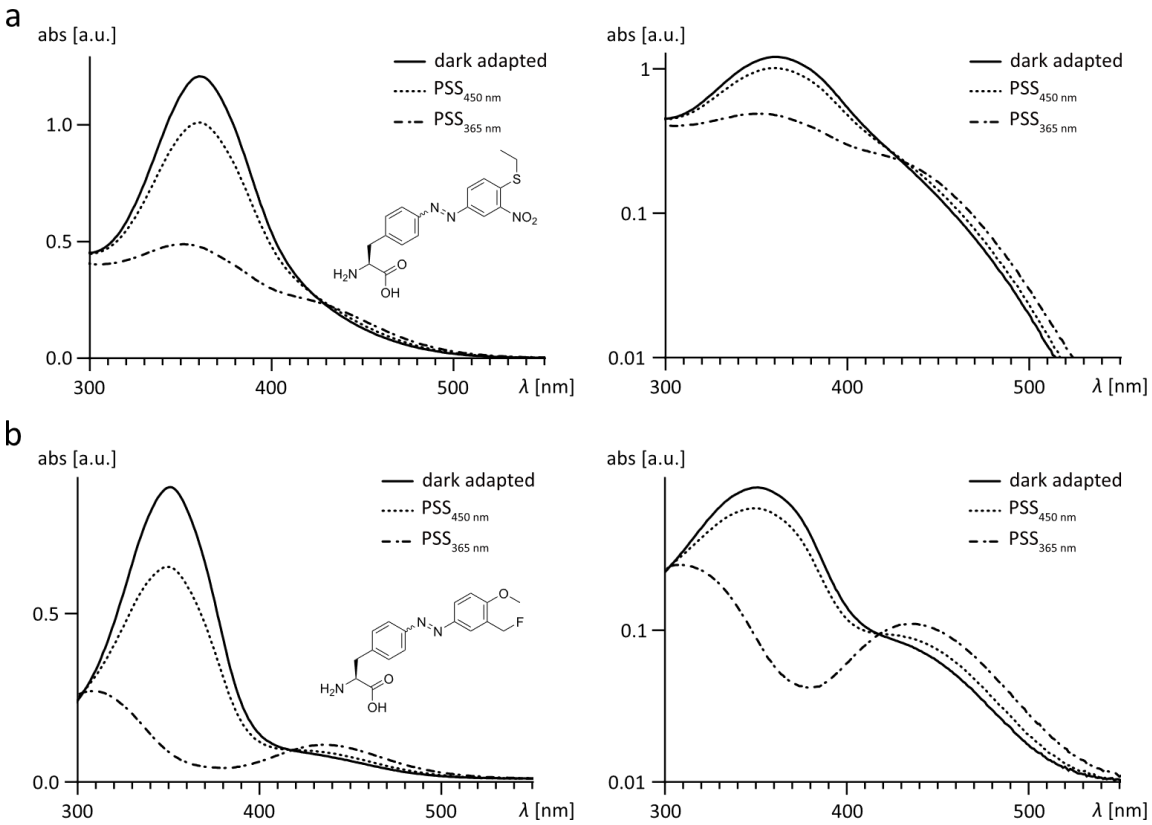


Figure 3.15 | Absorption spectra of the azo-ncAAs pSEtN and m1F in linear and log scale. (**a**) Absorption spectra of the model compound **pSEtN** in linear (left panel) and log scale (right panel). (**b**) Absorption spectra of the azo-ncAA **m1F** in linear (left panel) and log scale (right panel).

The absorption spectra of **pSEtN** feature a strong π-π* UV-band around 360 nm for the *trans*-isomer (Figure 3.15a). Surprisingly, the π-π* UV-band of the *cis*-isomer is (i) not shifted in comparison to the *trans*-isomer and (ii) barely separated from its weaker n-π* band around 425 nm in the visible region. Further, a large overlap in absorption spectra of the respective *cis*- and *trans*-isomer is observed, which is in accordance with literature on *ortho*-thiol substituted azobenzenes[277]. This renders complete conversion to one isomer photochemically impossible for the azo-ncAA **pSEtN**. Still, illumination of **pSEtN** at 365 nm resulted in greater absorption by the *trans*- than the *cis*-isomer, thereby generating a PSS containing a majority of the *cis*-isomer (Table 3.2). Similarly, illumination at 450 nm (or at 470 nm) generated a PSS containing a majority of the *trans*-isomer.

Table 3.2 | Photocharacterization of the azo-ncAAs pSEtN and m1F.

azo-ncAA	λ_{cis} [nm]	λ_{trans} [nm]	isosbestic point [nm]	A_{max} [nm]	$PSS_{365\ nm}$	$t_{½}$ [h]
pSEtN	365	450-470	430	360	0.60	2.9
m1F	365	450-470	418	351	0.90	26

Although the UV-Vis measurements show similar features for the azo-ncAA **m1F**, the band separation of the π-π* and weaker n-π* band is much more pronounced for the *cis*-isomer (Figure 3.15b). As expected, the methoxy-group in *para*-position of the diazene bond greatly improved the $PSS_{365\ nm}$ for **m1F** in comparison to **pSEtN** (Table 3.2).

Kinetic measurements showed that **m1F** is slow-relaxing, exhibiting a half-life $t_{½}$ that was too long to measure conveniently, but with a lower bound of at least 26 h (Figure 3.16).

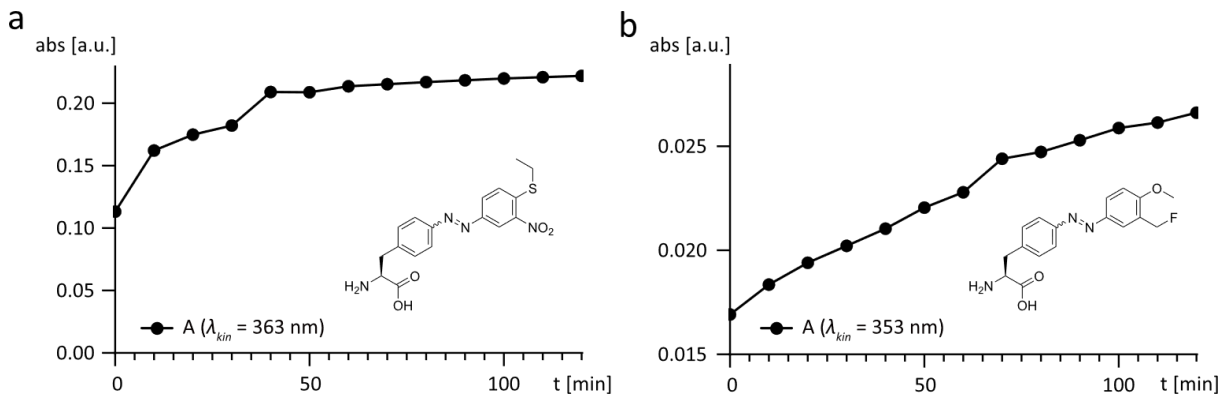


Figure 3.16 | Kinetic measurements of the *cis→trans* thermal relaxation of azo-ncAAs pSEtN and m1F. Details of the calculations for the relaxation half-life $t_{½}$ of both azo-ncAAs are shown in the supplementary information (Supplementary Table IV.3 and Supplementary Table IV.4). (**a**) Kinetic measurement of **pSEtN** was performed at λ_{kin} = 363 nm. (**b**) Kinetic measurement of **m1F** was performed at λ_{kin} = 353 nm.

The model azo-ncAA **pSEtN** exhibits a half-life $t_{½}$ of only 3 h, which is in accordance with literature-known thiolated azobenzenes[278] and should be considered when applying this azo-ncAA to the optical control of protein function.

3.3 Reversible Optical Control of Protein Function

With the development of the novel azo-ncAAs **m1F**, **TriF**, **pFN** and **pClN** we were able to demonstrate successful incorporation into sfGFP-N149TAG in *E. coli* using the *Mm*PylRS AzoC and KK11. With these photoswitchable tools in hand we set out to apply them to optically control diverse protein functions by introducing TAG-Cys mutations at user-defined positions within α-helical motifs. The following subchapters will give an overview of the different target proteins and our attempts to gain optical control over PPIs (Chapter 3.3.1) and enzyme activity (Chapter 3.3.2). Beyond that, efforts also went into the development of a photo-triggered cargo release from protein capsids (Chapter 3.3.3).

3.3.1 Reversible Optical Control of PPIs

Calmodulin

As a proof-of-principle target we chose the calcium-dependent protein calmodulin (Cam), since successful crosslinking using the photoswitchable click-amino acid **F-PSCaa** has already been demonstrated in Cam-M76TAG-E83C (Chapter 1.5.2).[250] Initial experiments therefore focused on reproducing the reported results. To this end we tested the different azo-ncAAs for incorporation into C-terminally His_6-tagged Cam-M76TAG-E83C in *E. coli* (Figure 3.17a, b, c). In order to estimate the crosslinking capabilities of our azo-ncAAs, we analyzed purified Cam variants bearing the desired azo-ncAAs via LR-MS (Figure 3.17d). While both **pXN** derivatives yielded full crosslinks with the proximal cysteine amino acid residue (*i,i+7*), no or only partial crosslinking could be observed for **m1F** and **TriF**, respectively. Consequently, we focused on the expression and purifications of Cam amber suppressed with one of the **pXN** derivatives, **pFN**.

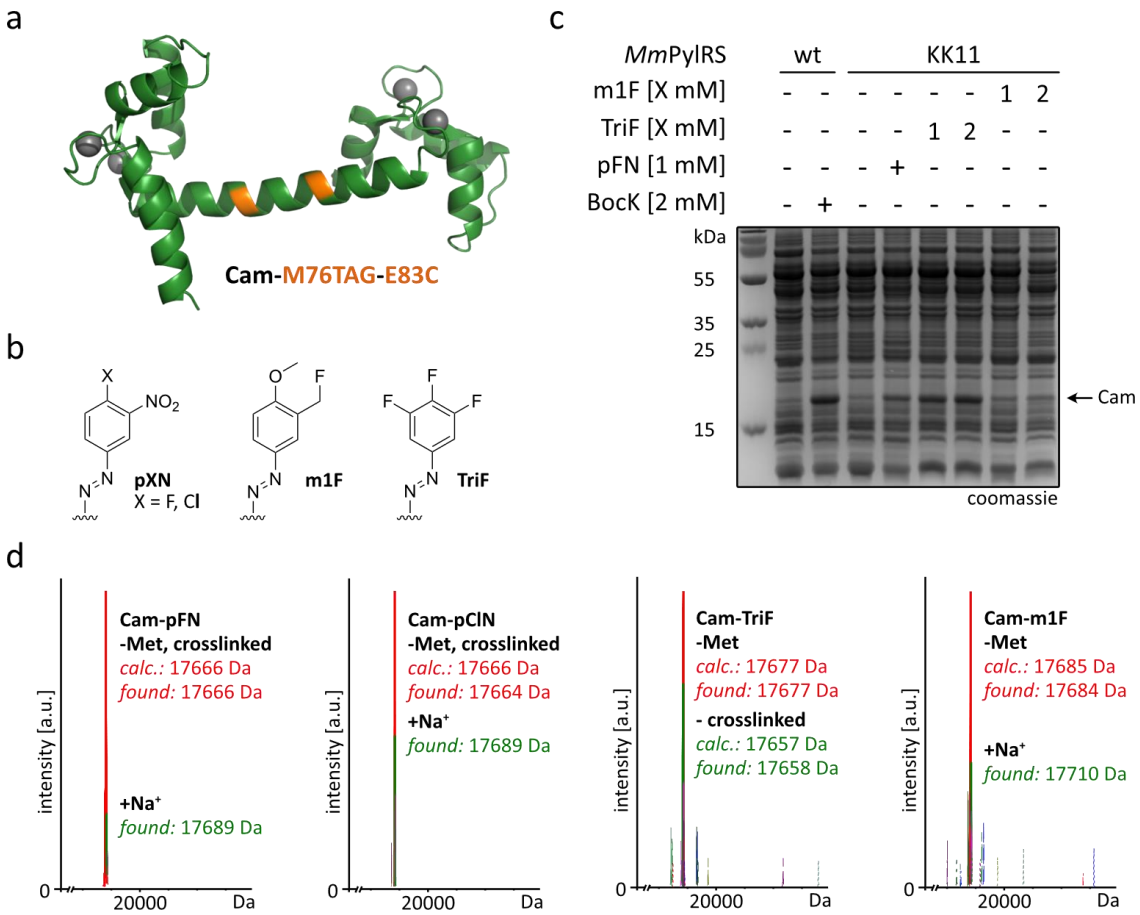


Figure 3.17 | Amber suppression of Cam-M76TAG-E83C-His$_6$ with the azo-ncAAs pXN (X = Cl, F), TriF and m1F in *E. coli*. (**a**) Crystal structure of Cam (pdb: 1cll) with the TAG- and Cys-position depicted in orange. Ca^{2+}-ions are shown as grey spheres. (**b**) Abbreviated structural formulas of the azo-ncAAs **pXN** (X = Cl, F), **TriF** and **m1F**. (**c**) Coomassie-stained SDS-PAGE analysis of azo-ncAA incorporation by *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). BocK, incorporated by the wt PylRS was used as positive control. (**d**) LR-MS analysis of amber suppressed Cam variants produced using the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB).

Contrary to reported purification procedures of amber suppressed Cam[250], supplementation of purification buffers with 5 mM $CaCl_2$ was crucial for successful affinity-chromatography. The obtained protein was subsequently purified via SEC and dialyzed into Ca^{2+}-free buffer for CD measurements (Supplementary Figure III.7).

The recorded CD-spectra of the different Cam samples all feature two distinct minima around 208 and 222 nm, typical for α-helical proteins (Figure 3.18a). Unfortunately, the expected difference in helicity upon *trans*→*cis* photoisomerization for amber suppressed Cam could not be observed. Instead, the biggest difference in helicity was observed for wt Cam samples in different buffer systems. Supplementation of the buffer with Ca^{2+} led to a noticeable decrease of predominantly the minimum at 222 nm, suggesting a conformational change of Cam. UV-Vis experiments of amber suppressed Cam-M76TAG-E83C revealed incomplete photoisomerization of the azobenzene (Figure 3.18b). Since the respective PSS was already reached after 15 s of irradiation, the observed incomplete photoisomerization is suspected to be a result of mainly two factors. First, *para*-thiol-substituted azobenzenes are known to undergo inefficient photoisomerization as could also be observed during small molecule experiments with our model azo-ncAA **pSEtN** (Chapter 3.2).[277] Second, the azo-ncAA **pFN** was designed to feature minimal degrees of rotational freedom when conjugated to α-helices with a spacing of *i,i+7* in its *trans*-form. It has already been shown with azo-crosslinkers that the PSS typically suffers from such rigidity when covalently attached to a target protein (Chapter 1.3.3).[150, 279]

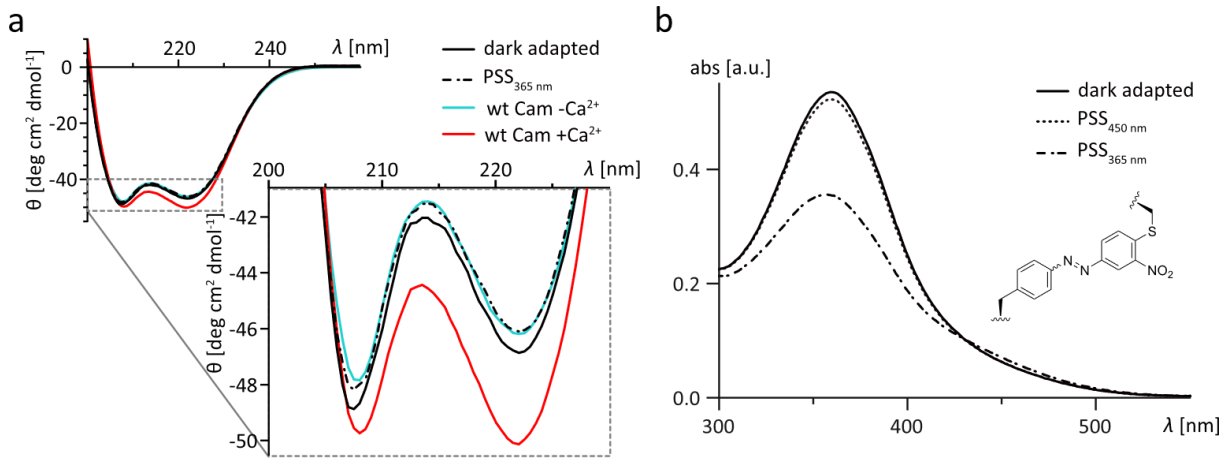


Figure 3.18 | CD spectroscopy and UV-Vis spectrophotometry of Cam-M76TAG-E83C amber suppressed with pFN. (**a**) CD spectra of **pFN**-bearing Cam without Ca^{2+} (black) and wt Cam with (red) and without Ca^{2+} (cyan). (**b**) Absorption spectra of F-bearing Cam with a $PSS_{365\ nm}$ of 35%. Equilibrium spectra were obtained by irradiation with the specified wavelengths for 15 s.

Taken aback by these initial results, we looked for suitable target proteins to estimate the maximum effect our azo-ncAA **pFN** could potentially have on protein function. We therefore investigated several small α-helical proteins such as affibody, Rap80 and the allosteric element PDZ3, binding to their respective interacting partner. This would allow us to use quantitative *in vitro* binding assays such as protein pull-down assays or fluorescence anisotropy.

Affibody

The small affinity protein affibody (6.5 kDa) was of particular interest to us, since reversible binding of its interacting partner Z-domain has already been demonstrated using a photoswitchable crosslinker molecule.[156] The affibody scaffold is based on a three-helical bundle derived from the B domain of staphylococcal protein A (Figure 3.19a), devoid of any natural cysteine amino acid residues. These key features and its stability towards mutations made affibody an ideal target for the site-specific installation of azo-ncAAs within the α-helical motifs.[280]

To this end, we tested a range of TAG-Cys affibody mutants for incorporation of the desired azo-ncAAs in *E. coli* (Supplementary Table IV.5). Only one out of 12 TAG-Cys mutants (Q26C-S33TAG) showed promising results when amber suppressed with **pXN** derivatives (Figure 3.19c). Western blot analysis of affibody variants amber suppressed with **m1F** or **TriF** resulted in no detectable amounts of protein. Consequently, subsequent experiments were performed exclusively with the azo-ncAAs **pFN** and **pClN**.

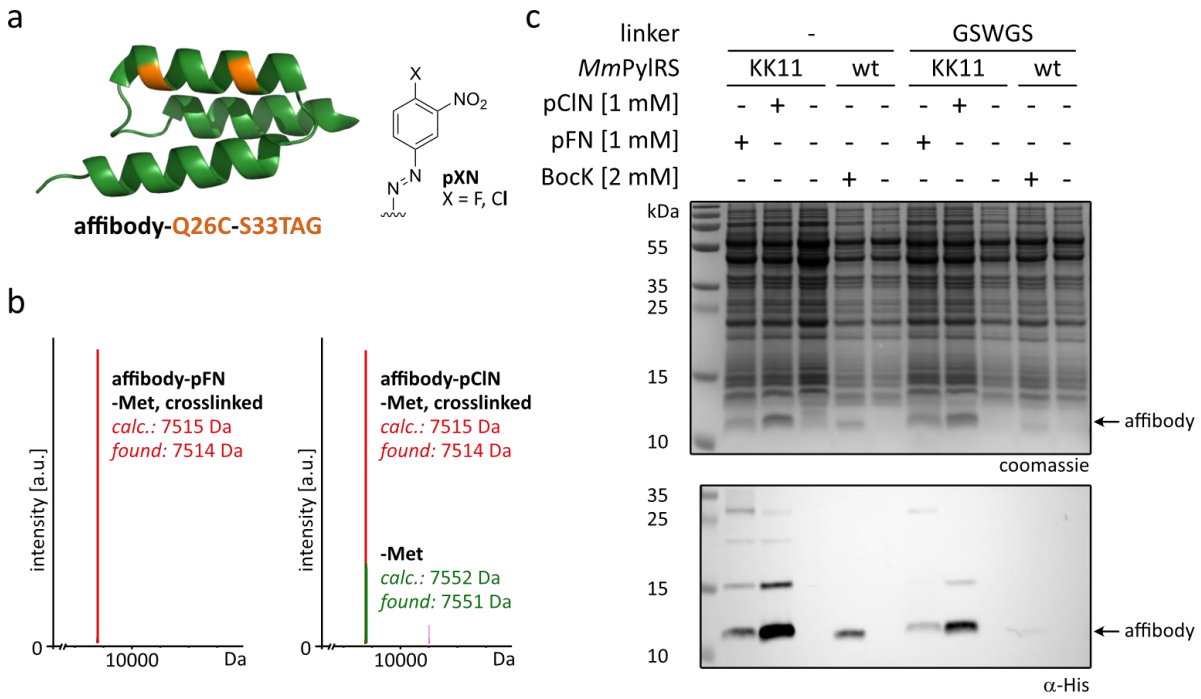


Figure 3.19 | Amber suppression of affibody-Q26C-S33TAG-(GSWGS)-His$_6$ with the azo-ncAAs pFN and pClN in *E. coli*. (**a**) Crystal structure of affibody (pdb: 1lp1) with the TAG- and Cys-position depicted in orange. The abbreviated structural formula of the **pXN** (X = Cl, F) derivatives is shown. (**b**) Coomassie-stained SDS-PAGE analysis of azo-ncAA incorporation by *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). BocK, incorporated by the wt PylRS was used as positive control. Two different constructs were used during protein expression, Cam-Q26C-S33TAG-His$_6$ (no linker) and Cam-GSWGS-His$_6$ (linker: GSWGS). (**d**) LR-MS analysis of amber suppressed Cam-Q26C-S33TAG-His$_6$ variants produced using the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB).

Purification via affinity chromatography followed by SEC of full-length affibody also yielded higher molecular weight proteins that correspond to a range of chaperones.[281] This was especially problematic using the azo-ncAA **pClN** (Figure 3.20a).

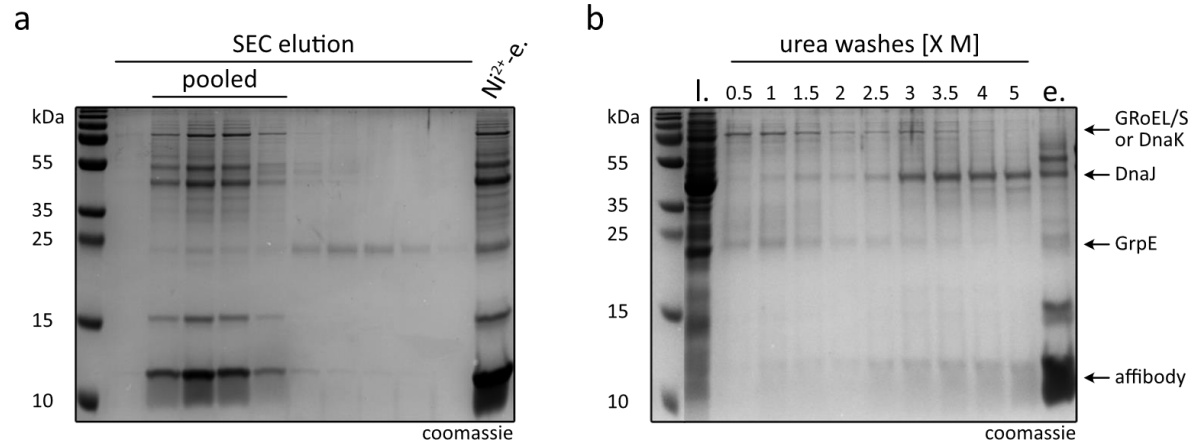


Figure 3.20 | Purification troubleshooting of affibody- Q26C-S33TAG-His$_6$ amber suppressed with the azo-ncAAs pClN. (**a**) Coomassie-stained SDS-PAGE analysis of SEC purification. Ni^{2+}-e. shows the amber suppressed affibody protein after affinity chromatography and prior to SEC purification. (**b**) Coomassie-stained SDS-PAGE analysis of affinity chromatography purification using washes containing increasing amounts of urea (0.5-5.0 M). l. shows the lysate after sonication, e. the eluted protein. Observed protein bands corresponding to suspected chaperones are annotated accordingly.

Cation exchange purification, high-salt (600 mM NaCl), ATP (10 mM) and even urea washes (5 M) were tested to remove the observed impurities. Unfortunately, little to no improvement could be observed via SDS-PAGE analysis (Figure 3.20b).

Additionally, a full crosslink could only be observed in LR-MS data of affibody bearing the azo-ncAA **pFN** but not with the less reactive **pClN** analog (Figure 3.19b). This might also explain, why proteins of higher molecular weight were predominantly observed for **pClN**-suppressed affibody. Partial intramolecular crosslinking would lead to more intermolecular crosslinks with proximal proteins such as chaperones. Efforts to enhance the amount of intramolecular crosslink obtained with **pClN** failed. Neither post-translational incubation at a more basic pH and/or at higher temperatures (Supplementary Figure III.8a), nor expression of amber suppressed affibody at different pH (Supplementary Figure III.8b) led to the desired result.

Based on reports of azobenzene-based coiled coil photoswitches[146] we have further designed a range of affibody variants that do not bear the TAG-Cys mutations within the typical α-helical motif (*i,i+7*). Anchoring of the azo-ncAAs into a more flexible region across the diagonal of two adjacent α-helices would promote optimal packing of hydrophobic residues within the coiled-coil motif in the *trans*-form. Photoisomerization to the *cis*-isomer was expected to disrupt the interhelical packing. Amber suppression of such additional mutants yielded barely detectable amounts of protein and only using the problematic azo-ncAA **pClN** (Supplementary Table IV.5). Once more, the obtained LR-MS data revealed no to little crosslinking of **pClN** in the respective affibody mutants.

We therefore focused our attention on amber suppression of C-terminally His_6-tagged affibody-Q26C-S33TAG (**TP1**) with the fully crosslinking azo-ncAA **pFN**. Typically, the obtained protein was purified only via affinity chromatography, since low amber suppression yields (0.3 mg/L) and insufficient protein detection methods prevented additional purification via SEC. Although introduction of a GSWGS-linker before the C-terminal His_6-tag (**TP2**) diminished amber suppression yields with **pFN** even further (0.13 mg/L) (Figure 3.19c), it finally allowed for subsequent purification via SEC.

In vitro pull-down assays were performed using affibody interactor Z-domain as bait protein. Expression as fusion-construct bearing N-terminal GST allowed for immobilization on GSH-sepharose beads (Figure 3.21a). Control experiments using wt affibody-GSWGS-His_6 (**TP3**) showed clear physical interaction between affibody and the Z-domain (Figure 3.21b). Unfortunately, no binding was observed for either amber suppressed affibody construct (**TP1** and **TP2**) (Figure 3.21b, c). One likely explanation for these negative results is based on chaperones sticking to amber suppressed affibody variants and thereby hindering interaction with the Z-domain (Figure 3.20). Although observed to a lesser extent when using the azo-ncAA **pFN,** problematic bands corresponding to proteins with a higher molecular weight could still be found after purification via SEC (Figure 3.21c). Since incorporation of azo-ncAAs into the affibody scaffold were rather low yielding and seemed to result in enhanced misfolding, we explored protein scaffolds beyond α-helical coiled coil motifs.

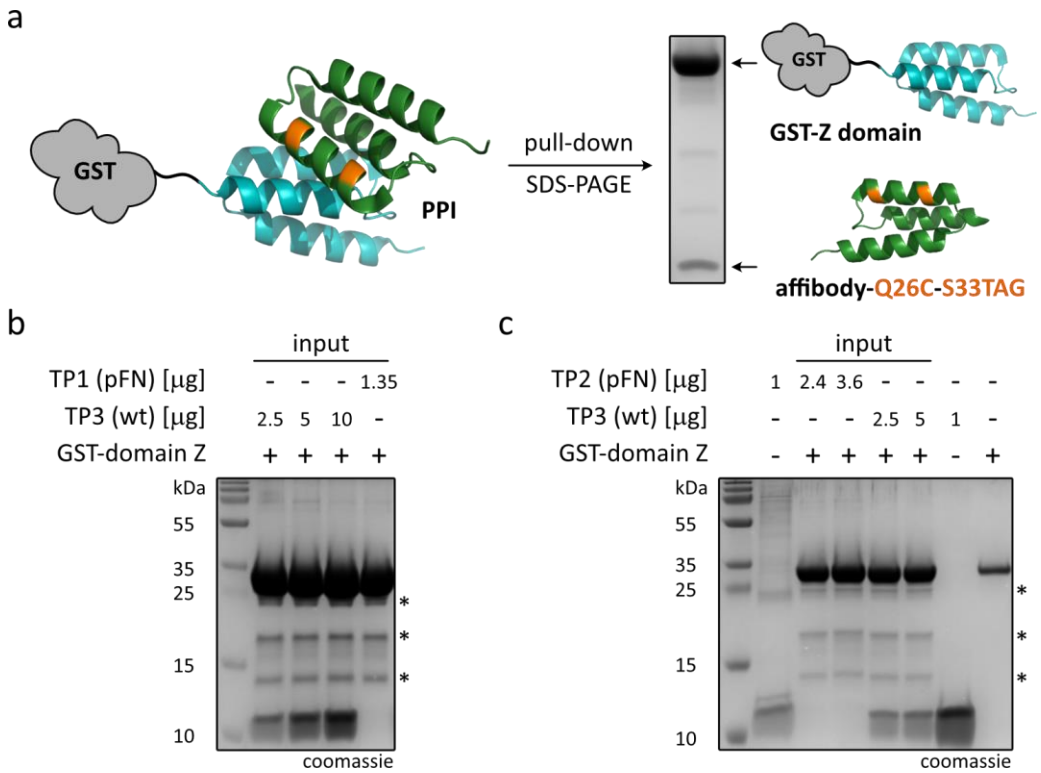


Figure 3.21 | *In vitro* pull-down assays using GST-Z-domain and affibody-Q26C-S33TAG amber suppressed with pFN. (**a**) Schematic depiction of pull-down assays using N-terminally GST-tagged Z-domain (cyan, pdb: 1lp1) and amber suppressed affibody-Q26C-S33TAG (green). (**b**) Coomassie-stained SDS-PAGE analysis of pull-down assays (20 µL loaded) using dark-adapted affibody-Q26C-S33TAG-His_6 (**TP1**) amber suppressed with **pFN** (purified via affinity chromatography prior to use). Installation of a GSWGS-linker did not impede pull-down assays as demonstrated for wt affibody-GSWGS-His_6 (**TP3**). Artefact bands observed for purified GST-protein Z-domain are annotated as *. Note that the concentration of **pFN**-bearing **TP1** was estimated using a NanoPhotometer® N60 (*Implen*) and might not be accurate, since the obtained protein was not pure after affinity chromatography. (**c**) Coomassie-stained SDS-PAGE analysis of pull-down assays (15 µL loaded) using dark-adapted affibody-Q26C-S33TAG-GSWGS-His_6 (**TP2**) amber suppressed with **pFN** (purified via SEC prior to use). Wt affibody-GSWGS-His_6 (**TP3**) was used as a positive control. Artefact bands observed for purified GST-protein Z-domain are annotated as *. Note that the concentration of **pFN**-bearing **TP2** was estimated using a NanoPhotometer® N60 (*Implen*) and might not be accurate, since the obtained protein was not pure even after purification via SEC.

Rap80

We further investigated human receptor-associated protein 80 (Rap80), which plays a critical role in signaling pathways by recruiting BRCA1-A for DNA damage repair. Rap80 is a small α-helical peptide containing so-called tandem ubiquitin interacting motifs (tUIMs) (Figure 3.22a). These tUIMs specifically recognize K63-linked di-ubiquitin (diUb) while only moderately binding mono-Ub.[282]

Since a binding assay for the investigation of the PPI between Rap80 and K63-linked diUb was already established in our lab[283], we envisioned optical control over di-Ub binding through incorporation of the designed azo-ncAAs into the α-helix of Rap80(79-126). The most promising expression construct of Rap80 was kindly provided by co-worker M. von Wrisberg. It comprises the tUIM of Rap80(79-126) flanked by an N-terminal His_6-tag that can be cleaved by Thrombin and a C-terminal StrepII-tag that can be removed using TEV protease (His_6-Throm-Rap80(79-126)-TEV-StrepII). We tested azo-ncAA incorporation into either native Rap80(79-126) or Rap80(79-126) featuring a mutated core-region (96-102). In this Rap80(79-126) variant, the central part of the α-helix comprises seven alanine amino acid residues (Rap80 7A) that promote α-helical formation.

Early incorporation assays into Rap80 7A-R97TAG were exerted with the azo-ncAA **m2Cl** (Chapter 3.1.4) but no full-length protein could be detected via western blot analysis. Also, purification of the respective azo-ncAA-bearing Rap80 variants via affinity chromatography gave no detectable amounts of protein. Amber suppression of Rap80-R97TAG-E104C and Rap80 7A-R97TAG-E104C with the most promising pXN derivatives yielded at least detectable amounts of protein (Figure 3.22c, d).

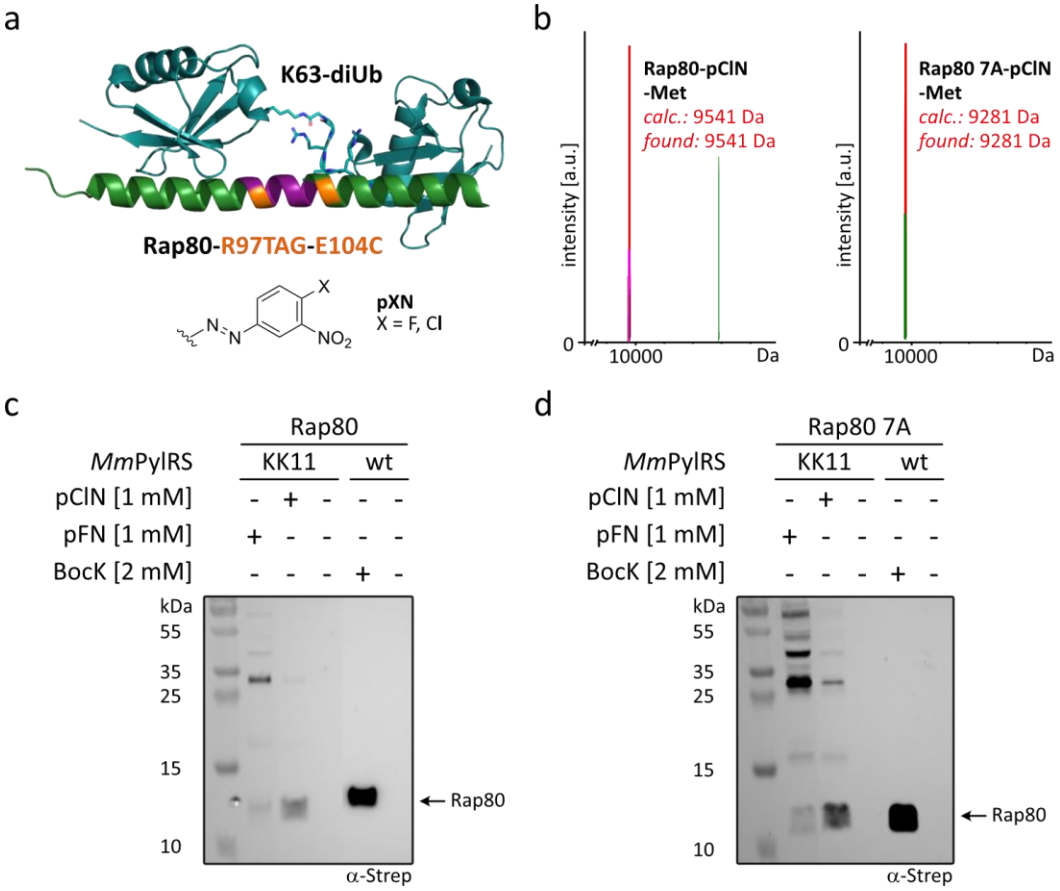


Figure 3.22 | Amber suppression of His_6-Throm-Rap80 (7A)-R97TAG-E104C-TEV-StrepII with the azo-ncAAs pFN and pClN in *E. coli*. (**a**) Crystal structure of Rap80 tUIM with the TAG- and Cys-position depicted in orange, the interacting K63-linked diUb in cyan (pdb: 3a1q). The amino acid residues mutated to alanine residues in Rap80 7A are shown in purple. The abbreviated structural formula of the pXN (X = Cl, F) derivatives is shown. (**b**) LR-MS analysis of Rap80 (7A)-R97TAG-E104C amber suppressed with pClN using the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). (**c,d**) Western blot analysis of pXN incorporation into Rap80 (c) and Rap80 7A (d) by *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). BocK, incorporated by the wt PylRS was used as positive control.

Similar to affibody though, bands corresponding to proteins of higher molecular weight could be observed via western blot analysis. To our dismay, incorporation of the typically fully crosslinking azo-ncAA **pFN** resulted almost exclusively in protein bands of a higher molecular weight. Once more, incorporation of **pClN** yielded a mixture of crosslinked and non-crosslinked Rap80 variants (Figure 3.22b).

In conclusion, incorporation of bulky and hydrophobic azo-ncAAs into essential α-helices of small and compact proteins is problematic. It was suspected, that the α-helices in proteins such as affibody and Rap80 could no longer properly fold, thereby attracting various chaperones (Figure 3.20). Consequently, protein bands with a higher molecular weight could be observed in western blot analysis. To circumvent these issues, we would need a protein that would allow for incorporation of azo-ncAAs into more flexible regions, thereby preventing misfolding, but still having the desired effect on its function.

PDZ3

To this end, we investigated the smallest allosteric element found in a broad variety of proteins, the PDZ-domains. Protein domains such as PDZ3 found in postsynaptic density protein 95 (PSD-95), function as mediators of PPIs by recognizing a short, C-terminal amino acid sequence (< 10 aa) within the target protein. The characteristic structure of the PDZ3 domain includes a C-terminal extension that forms an additional third α-helix (α3-helix) and functions as an allosteric element (Figure 3.23a).[284]

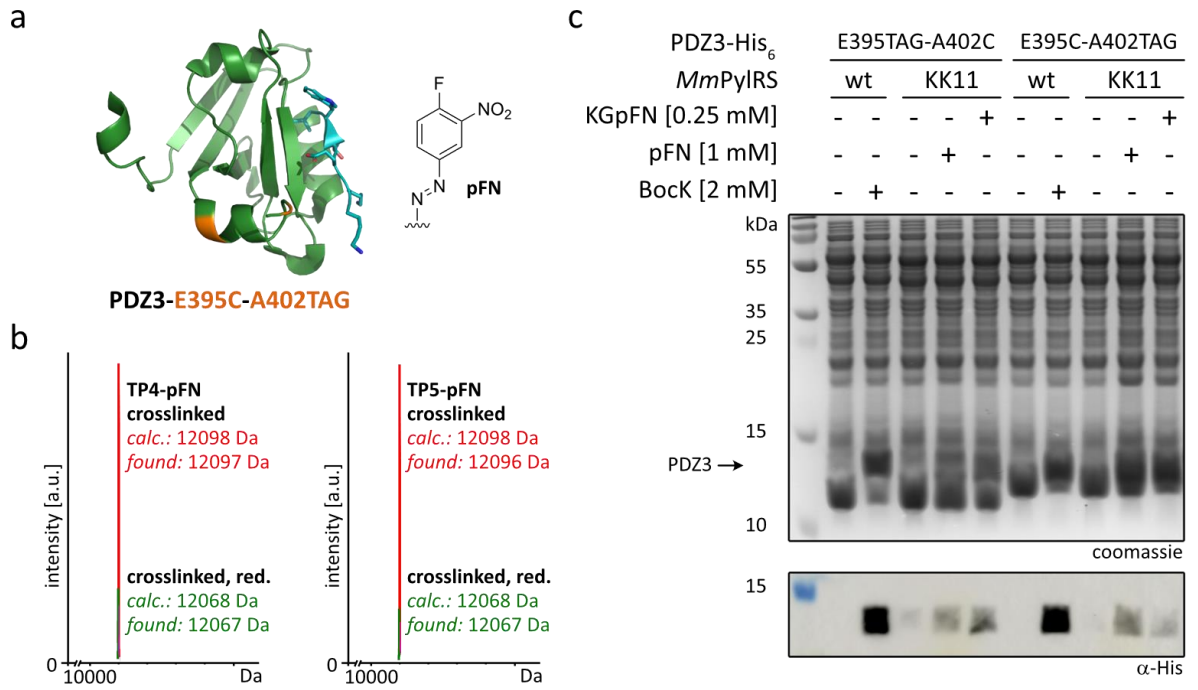
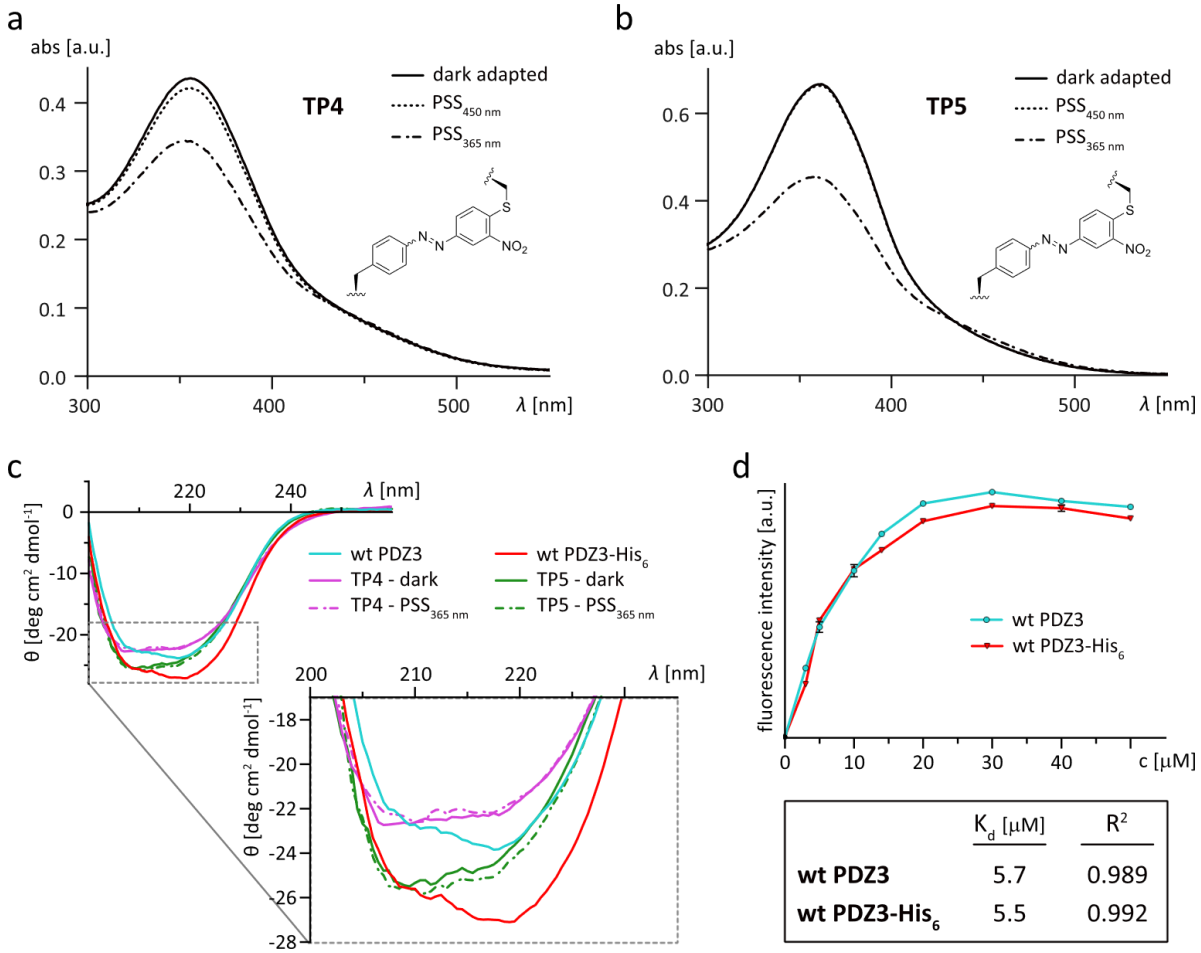


Figure 3.23 | Amber suppression of PDZ3-E395TAG-A402C-His$_6$ (TP4) and PDZ3-E395TAG-A402C-His$_6$ (TP5) with the azo-ncAA pFN in *E. coli*. (**a**) Crystal structure of PDZ3 from PSD-95 (pdb: 1tp5) with the TAG- and Cys-position depicted in orange, the interacting peptide KETWV in cyan. The abbreviated structural formula of **pFN** is shown. (**b**) LR-MS analysis of both PDZ3 variants amber suppressed with **pFN** using the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). (**c**) Coomassie-stained SDS-PAGE and western blot analysis of **pFN** incorporation into both PDZ3 variants by *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). pFN was either added as single ncAA or as tripeptide **KGpFN** (Chapter 3.1.9). BocK, incorporated by the wt PylRS was used as positive control.

By applying azo-ncAAs to this rather flexible α3-helix, we hoped to prevent misfolding of the protein. Further, recent work by Bozovic et al. demonstrated photoregulation of ligand binding to PDZ3 with an azo-crosslinker bound to the allosteric element α3-helix.[158] Dependent on the temperature, isomerization of the employed azo-crosslinker to the *cis*-form increased the binding affinity of a peptide ligand to the allosteric element by a factor of up to 120. Simultaneously, binding of the peptide ligand accelerated thermal relaxation from the *cis*- to the *trans*-isomer.

Encouraged by these reports, we tested our azo-ncAA pFN for incorporation into C-terminally His$_6$-tagged PDZ3-E395TAG-A402C (**TP4**) and PDZ3-E395C-A402TAG (**TP5**). While successful amber suppression was difficult to estimate by SDS-PAGE and western blot analysis (Figure 3.23c), purification via affinity-chromatography and SEC furnished the pure pFN-bearing PDZ3 variants (**TP4** and **TP5**) in yields of 0.6-1.4 mg/L, respectively. LR-MS data suggested full crosslinking of pFN in both PDZ3 mutants (Figure 3.23b). Interestingly, partial reduction of the nitro-moiety was observed, too. Since the nitro-moiety is only necessary for successful crosslinking, these results did not impede the following experiments.

We next analyzed the amber suppressed proteins via CD-measurements and UV-Vis spectrophotometry. C-terminally His_6-tagged wt PDZ3 and untagged wt PDZ3 were used as controls during CD-measurements. All spectra featured a broad band between 210-225 nm, which is in accordance with values found in literature (Figure 3.24c).[158]



	K_d [µM]	R^2
wt PDZ3	5.7	0.989
wt PDZ3-His$_6$	5.5	0.992

Figure 3.24 | CD spectroscopy and UV-Vis spectrophotometry of Cam-M76TAG-E83C amber suppressed with pFN. (**a, b**) Absorption spectra of **pFN**-bearing PDZ3-E395TAG-A402C (**TP4**) and PDZ3-E395C-A402TAG (**TP5**) with a PSS_{365nm} of 21% and 32%, respectively. Equilibrium spectra were obtained by irradiation with the specified wavelengths for 15 s. (**c**) CD spectra of **pFN**-bearing PDZ3-E395TAG-A402C (**TP4**, purple lines) and PDZ3-E395C-A402TAG (**TP5**, green lines). CD-spectra of untagged wt PDZ3 (cyan line) and wt PDZ3-His_6 (red line) were used as controls. (**d**) Intrinsic tryptophan fluorescence measurements using untagged wt PDZ3 (cyan line) and wt PDZ3-His_6 (red line). Fluorescence intensity was determined as a function of protein concentration with the initial ligand concentration (KETWV) kept fixed at 15 µM. K_d values were obtained by fitting the data to the function "Specific binding with Hill slope" using GraphPad prism.

Interestingly, the C-terminal His_6-tag in wt PDZ3 seems to slightly promote α-helical formation in comparison to untagged wt PDZ3. Similar to Cam however, the observed difference in CD-spectra obtained upon photoisomerization to the *cis*-isomer was marginal for both amber suppressed PDZ3 variants. Considering, that the α3-helix is small and only makes up around 5% of the overall structure, these findings are not entirely surprising. Moreover, the CD-spectra of both amber suppressed PDZ3 variants (**TP4** and **TP5**) featured a less intense minimum at 220 nm in comparison to wt PDZ3. This could be interpreted as a slight decrease in α-helical formation, possibly of the α3-helix.

Subsequent UV-Vis experiments showed, that the minimal change observed in the CD-spectra might be a result of insufficient photoisomerization. Similar to Cam and the model small molecule **pSEtN** (Chapter 3.2), a large overlap of the UV-Vis spectra for each isomer could be observed (Figure 3.24a, b). Although 15-30s of irradiation were sufficient to reach an equilibrium, the PPS_{365} of **pFN**-bearing PDZ3 variants was especially low for PDZ3-E395TAG-A402C (**TP4**, 21%).

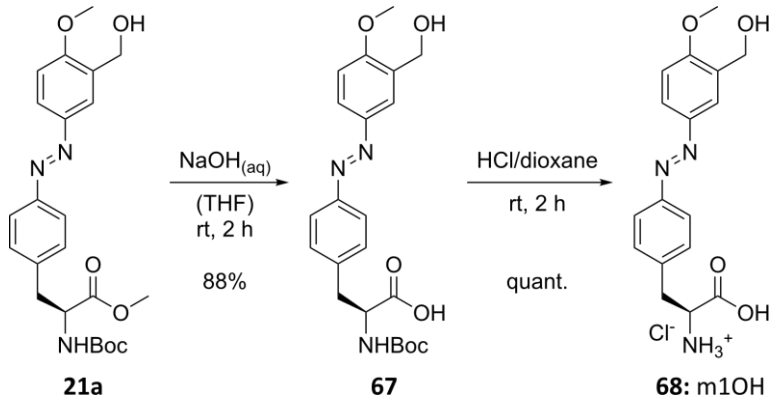
Binding affinity assays were therefore performed only with amber suppressed PDZ3-E395C-A402TAG (**TP5**) and both wt constructs as control. To measure binding affinities by intrinsic tryptophan fluorescence, the interacting pentapeptide KETWV was synthesized via standard Fmoc-SPPS. Control experiments using the wt constructs were performed at 21 °C as specified in Chapter 9.12.3 to estimate whether the C-terminal His_6-tag would affect binding of the pentapeptide (Figure 3.24d). To our satisfaction, the calculated K_d values were almost identical, so that we could proceed with C-terminally His_6-tagged PDZ3 constructs.

In contrast to the azo-crosslinker used by Bozovic et al., Trp-fluorescence emission at 325 nm would get absorbed by our azo-ncAA **pFN**, thereby leading to fluorescence quenching. Thus, future work will be focusing on the preparation of a fluorophore-labelled hexapeptide KKETWV that would instead allow us to measure binding affinities via fluorescence anisotropy. It is also worth noting, that conjugation of the azobenzene using a spacing of *i,i+7* corresponds to the *cis*-isomer of the reported azo-crosslinker, while the same spacing instead fits the *trans*-isomer of the azo-ncAA **pFN**. Testing a shorter spacing of *i,i+4* might proof valuable to render potential results more comparable.

antiGFP nanobodies

Nanobodies (nbs) are derived from single variable domains of heavy-chain-only antibodies and can therefore easily be selected to bind to a broad range of target epitopes. As such, they typically do so with high selectivity and affinity, spanning from picomolar to low nanomolar range.[285] A couple of commonly used nbs have been selected for binding to fluorescent protein variants such as GFP. Recent years have further developed such strategies towards spatio-temporal control over nb binding affinities, either by exogenously added small molecules[286] or genetically introduced modifications[21, 244]. Mootz and co-workers installed a single photocaged tyrosine derivative via genetic code expansion into anti GFP nb, thereby successfully impairing binding to its target protein GFP.[244] Native binding affinity was restored upon photodecaging. The concept of such photobodies has further been expanded to the high-affinity ALFA-Tag nb, enabling photo-triggered binding to any biomolecule bearing the short peptide ALFA-tag.[245]

These concepts however only allow for targeted activation of protein binding. Reversible switching of protein binding was so far only realized by insertion of rather large LOV domains into nb loops, termed opto-nbs.[21] In order to generate reversibly switchable nbs with minimal perturbation, we envisioned the installation of non-crosslinking azo-ncAAs in a similar fashion as demonstrated by Mootz and co-workers.[244] To this end we synthesized a non-crosslinking analog of the azo-ncAA **m1F** (**29**), since its exceptional photophysical properties (Chapter 3.2) would allow for near-quantitative optical control over antiGFP nb binding. The target azo-ncAA **m1OH** (**68**) was obtained through hydrolysis of the intermediate methyl ester **21a** and subsequent *N*-Boc-deprotection (Scheme 3.20).



21a **67** **68:** m1OH

Scheme 3.20 | Synthesis of the non-crosslinking azo-ncAA m1OH (68) for side-chain photoisomerization.

Two different antiGFP nb variants were to be investigated for the generation of photoswitchable nbs, a maximizer (nb(max)) and minimizer (nb(min)). While nb(max) enhances GFP fluorescence upon binding, a decrease in fluorescence is observed with the nb(min) when binding to GFP.[287] We chose the same installation sites for our azo-ncAAs as were reported for the generation of photobodies using photocaged tyrosine derivatives, namely nb(max)-Y37TAG and nb(min)-Y116TAG (Figure 3.25b, c).[244]

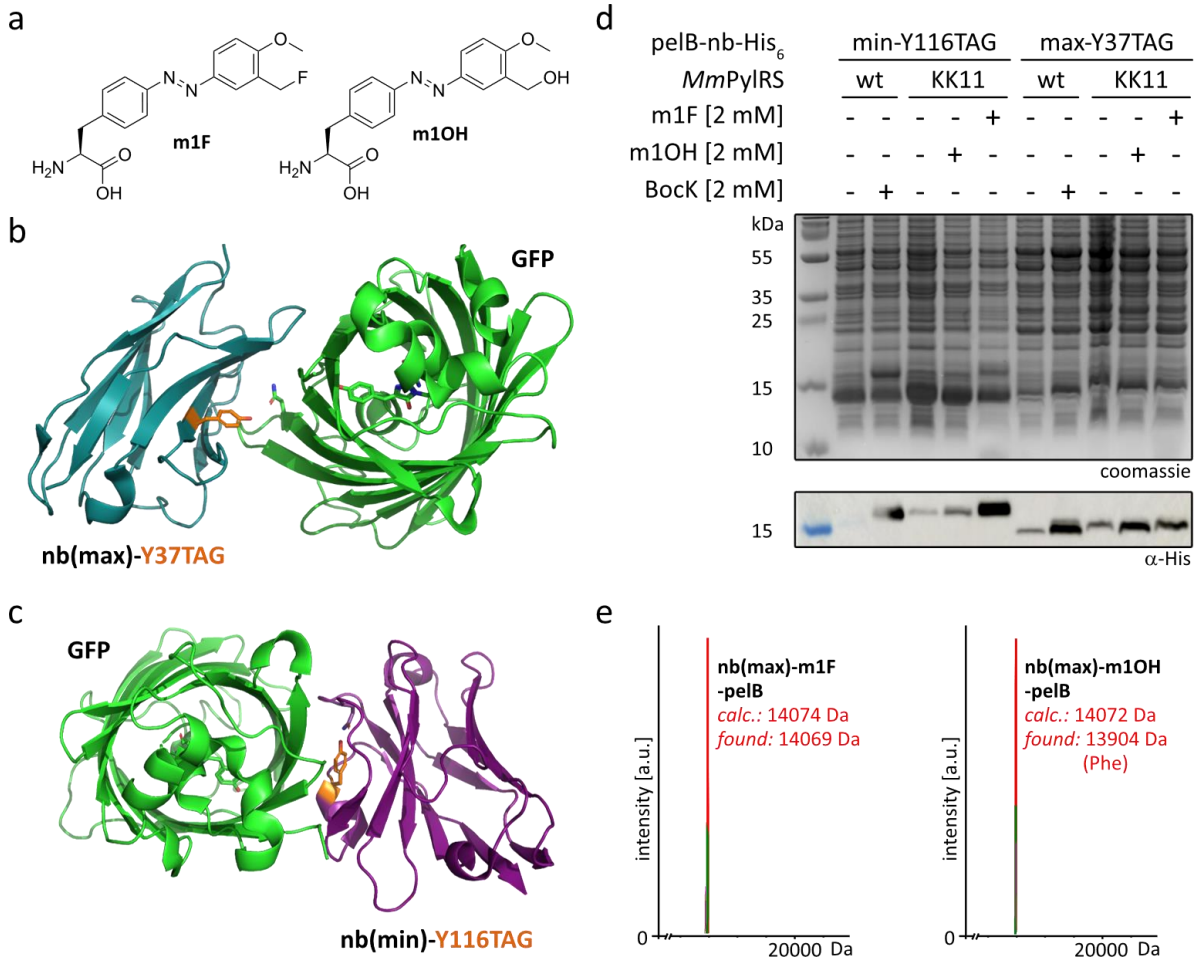


Figure 3.25 | Amber suppression of pelB-nb(min)-Y116TAG-His$_6$ (TP6) and pelB-nb(max)-Y37TAG-His$_6$ (TP7) with the azo-ncAAs m1F and m1OH in *E. coli*. (**a**) Structural formula of the azo-ncAAs **m1F** and **m1OH**. (**b,c**) Crystal structures of GFP nb(max) and nb(min) are depicted in cyan (pdb: 3k1k) and purple (pdb: 3g9a), respectively. The TAG- and Cys-positions are shown in orange, the interacting GFP in green. (**d**) Coomassie-stained SDS-PAGE and western blot analysis of azo-ncAA incorporation into both nb variants by *Mm*PylRS KK11 in the *E. coli* strain K12. BocK, incorporated by the wt PylRS was used as positive control. (**e**) LR-MS analysis of nb(max)-Y37TAG amber suppressed with **m1F** or **m1OH** using the *Mm*PylRS KK11. While the correct LR-mass was observed using **m1F**, the addition of **m1OH** only resulted in misincorporation of phenylalanine.

We reasoned that one isomeric form of the azo-ncAA, presumably the *trans*-isomer would abolish binding of the nb to its target GFP, similar to the reported photocaged tyrosine. Photoisomerization to the *cis*-isomer would ideally restore binding of the amber suppressed nb to GFP, giving not only a photoactivatable, but reversibly switchable anti GFP nb variant.

With the desired azo-ncAA **m1OH** in hand, we tested both nb(max) and nb(min) for amber suppression using PylRS KK11. To this end, we made use of a well-established nb expression construct that was already available within our lab.[288] Herein, the C-terminally His_6-tagged nb variants were N-terminally fused to a pelB signal sequence, thereby targeting the expressed nb variants to the periplasm. In addition to the azo-ncAA **m1OH**, both pelB-nb(min)-Y116TAG (**TP6**) and pelB-nb(max)-Y37TAG (**TP7**) were tested for amber suppression with **m1F** (Figure 3.25a). Unfortunately, decent amber suppression yields could only be observed with **m1F** and not with the designed non-crosslinking analog **m1OH** (Figure 3.25d). In addition, subsequent purification via affinity chromatography of nb(min) variants was unsuccessful, presumably due to the formation of inclusion bodies. In contrast to that, purification of nb(max) variants was straight-forward. While corresponding LR-MS data of nb(max) amber suppressed with **m1F** fit the calculated molecular weight minus the pelB signal sequence, the obtained LR-MS spectrum of nb(max) amber suppressed with **m1OH** corresponded to misincorporation of phenylalanine (Figure 3.25e).

Consequently, we decided to proceed our investigation of antiGFP nanobodies using the azo-ncAA **m1F**. Although originally designed for crosslinking with a proximal cysteine, the obtained data of this particular azo-ncAA suggests that the benzyl fluoride is barely reactive towards nucleophiles. We therefore expected little to no interference when using **m1F** for the generation of photoswitchable nbs. We then performed *in vitro* pull-down assays using the purified fusion protein GST-EGFP, kindly provided by co-worker M. Fottner (Figure 3.26).

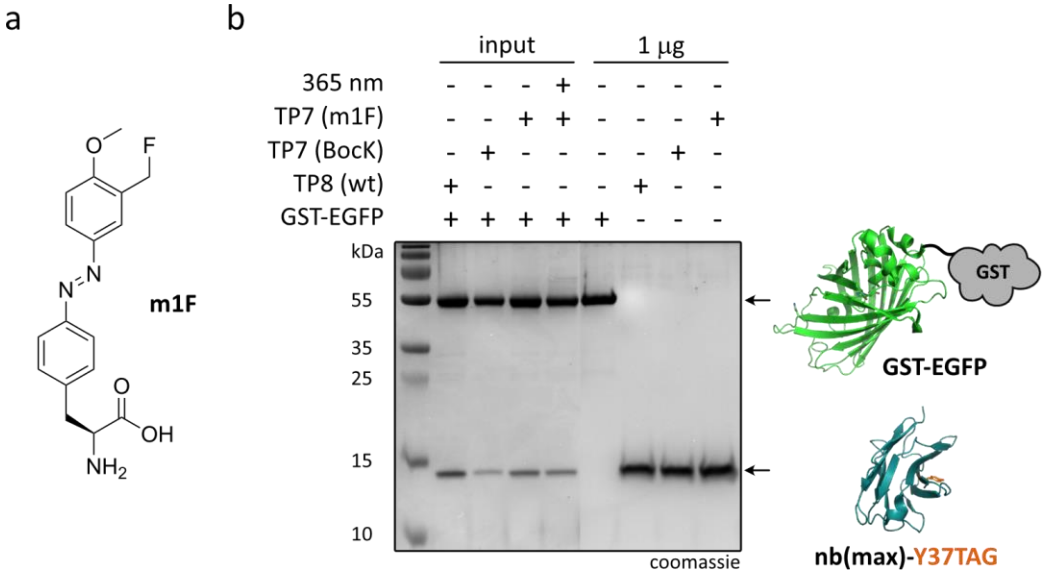


Figure 3.26 | *In vitro* pull-down assays using GST-EGFP and nb(max)-Y37TAG amber suppressed with m1F. (**a**) Structural formula of the azo-ncAA **m1F**. (**b**) Coomassie-stained SDS-PAGE analysis of pull-down assays (10 µL loaded) using nb(max)-Y37TAG-His_6 (**TP7**) amber suppressed with **m1F** (purified via affinity chromatography prior to use). 5.5 µg GST-EGFP were incubated with 3.5 µg nb(max) variant. Wt nb(max)-His_6 (**TP8**) and nb(max)-Y37TAG-His_6 (**TP7**) amber suppressed with BocK were used as controls.

GST-EGFP was immobilized on GSH-sepharose beads and then either incubated with dark-adapted **m1F**-bearing nb(max) or an equilibrium mixture obtained by irradiation with 365 nm (for 3 min, every 10 min). Control experiments were performed using C-terminally His_6-tagged wt nb(max) (**TP8**) as well as C-terminally His_6-tagged nb(max)-Y37TAG (**TP7**) amber-suppressed with BocK (Figure 3.26). Although binding could be observed for all nb(max) variants via coomassie-stained SDS-PAGE analysis, no difference in binding could be detected for dark-adapted and irradiated ($PSS_{365\ nm}$) nb(max)-Y37TAG bearing **m1F**.

Since wt nb(max) is reported to bind GFP with a K_d of 3.0 nM, we concluded that SDS-PAGE analysis is not sensitive enough to accurately determine whether a marginal difference in binding can be realized with photoswitchable nb(max). Further efforts will therefore focus on establishing FACS-assisted binding assays on the surface of *E. coli* cells using the AIDA autodisplay system as reported by Mootz and co-workers.[244]

3.3.2 Reversible Optical Control of Enzyme Activity

Parallel to investigating PPIs in small α-helical proteins such as affibody and Rap80, we also attempted spatio-temporal control over enzymatic activity by incorporation of our azo-ncAAs into the synthetic α-helical bundles AltTPase[289] and MID1sc10[290].

AltTPase

In contrast to its natural counterpart, the *de novo* designed alternative ATPase reported by Hecht and co-workers, is structurally much smaller and consists of four α-helices only (Figure 3.27a).[289] Together with a straight-forward readout based on a typical hydrolase assay, these features made the α-helical bundle the ideal target for optical control over enzymatic activity.

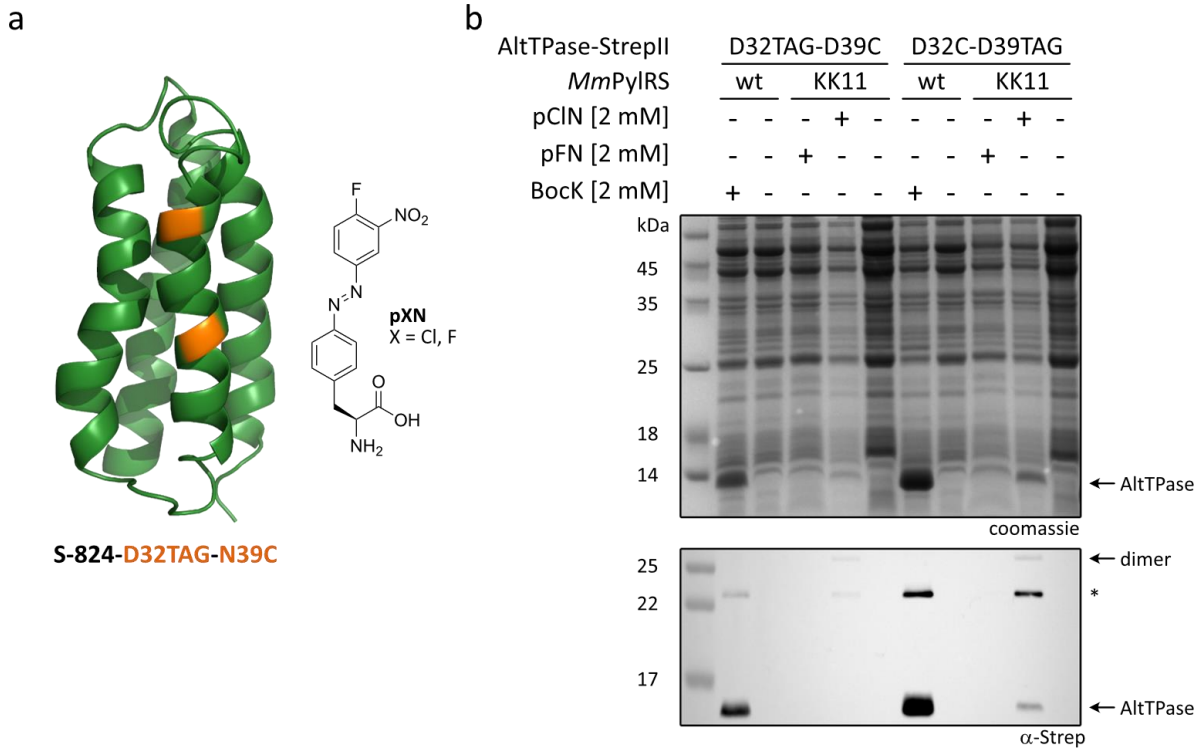


Figure 3.27 | Amber suppression of exemplary AltTPase-D32TAG-D39C-StrepII and AltTPase-D32C-D39TAG-StrepII with the azo-ncAAs pFN and pClN in *E. coli*. (**a**) Crystal structure of the structurally related four-helix bundle S-824[291] with the TAG- and Cys-position depicted in orange (pdb: 1p68). The abbreviated structural formula of the pXN (X = Cl, F) derivatives is shown. (**b**) Coomassie-stained SDS-PAGE and western blot analysis of pXN incorporation into AltTPase-D32TAG-D39C-StrepII and AltTPase-D32C-D39TAG-StrepII by *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). BocK, incorporated by the wt PylRS was used as positive control. * denotes unidentified artefact bands.

Initial expression constructs based on C-terminally His_6-tagged AltTPase were not compatible with the amber suppression approach. Since the four-helix bundle features multiple histidine amino acid residues, a separation of full-length protein and truncated protein using affinity chromatography was impossible. We therefore tested several TAG-Cys mutants of C-terminally StrepII-tagged AltTPase for amber suppression using the most promising azo-ncAAs **pFN** and **pClN** (Supplementary Table IV.6). Unfortunately, almost no full-length protein was observed via western blot analysis for any of the tested five mutants with our azo-ncAAs **pFN** and **pClN** (Figure 3.27b). Incorporation of BocK using the wt PylRS however resulted in easily detectable amounts of full-length protein for four mutants. Consequently, we put our efforts into finding other target proteins that would feature similarly appealing aspects as AltTPase and could be amber suppressed using our azo-ncAAs.

MID1sc10

In cooperation with the Hilvert group from ETH Zurich we instead focused on the computationally designed zinc-binding peptide MID1, comprised of a homodimeric α-helix.[292] Fusion of two such subunits afforded the single-chain MID1 variant, termed MID1sc. Progressive evolution and selection of more efficient variants brought forth the metalloenzyme MID1sc10. This four-helix protein shows an exceptional esterase activity towards a single enantiomer of fluorogenic esters.[290] These key features drove us to test our azo-ncAA pFN on MID1sc10 for the optical control of its esterase activity.

To this end, we designed various expression constructs that would ideally allow us to separate full-length from truncated protein via affinity chromatography (Table 3.3).

Table 3.3 | Description of MID1sc10 expression constructs. "yield (Ni^{2+})" shows the protein yield after affinity chromatography, while "final yield" shows the protein yield after processing (tag-cleavage) and final purification via SEC. * **TP9** was not entirely pure and cleaved MBP could still be detected via SDS-PAGE analysis; # **TP9** and **TP10** partially precipitated during purification via reverse affinity chromatography using buffers supplemented with 2 mM $ZnSO_4$.

construct	fusion	His_6-tag	cleavage	yield (Ni^{2+}) [mg/L]	final yield [mg/L]
TP9	MBP-MID1sc10	N-terminal	TEV	5.8	0.1*#
TP10	-	N-terminal	TEV	5.4	1.0#
TP11	-	C-terminal	-	2.3	0.7
TP12	-	C-terminal	TEV	1.4	0.4
TP13	MID1sc10-CPD	C-terminal	$InsP_6$	13.4	1.4

In addition, we desired to cleave off any tags prior to *in vitro* esterase assays, thereby avoiding any interference in zinc-binding or substrate-binding and -hydrolysis. Decent protein expression levels could be observed for almost all constructs, yielding sufficient amounts of the 11 kDa wt MID1sc10 after protein purification (Supplementary Figure III.9a). Further, we observed the correct LR-mass for all wt MID1sc10 constructs, prior and post cleavage (Supplementary Table IV.7). Interestingly, switching from N-terminal to C-terminal His_6-tag led to a significant decrease in protein expression apart from one fusion construct (Table 3.3). Herein, MID1sc10 is C-terminally fused to a cysteine protease domain (CPD). CPDs are typically found in multifunctional, auto processing repeats-in-toxin (MARTX) toxins.[293]

Not only does CPD seem to boost protein expression in fusion constructs but it is also described to undergo autoproteolysis when activated by exogenously adding inositol hexakisphosphate ($InsP_6$). This way, untagged MID1sc10 was obtained in highest yields (Table 3.3, Supplementary Figure III.9b).

Consequently, first efforts towards amber suppression of the metalloesterase MID1sc10 were exerted using the expression construct MID1sc10-CPD-His_6 (**TP13**). Contrary to previous efforts around α-helical bundles such as affibody or AltTPase, we deliberately placed the azobenzene within a hydrophobic fold between two α-helices. We thereby hoped to avoid misfolding of the amber suppressed protein, since the azobenzene was no longer exposed to an aqueous environment. Initial screens for incorporation of **pFN** into various TAG-Cys mutants yielded detectable amounts of full-length protein for all six tested mutants (Figure 3.28a, c and Supplementary Table IV.6).

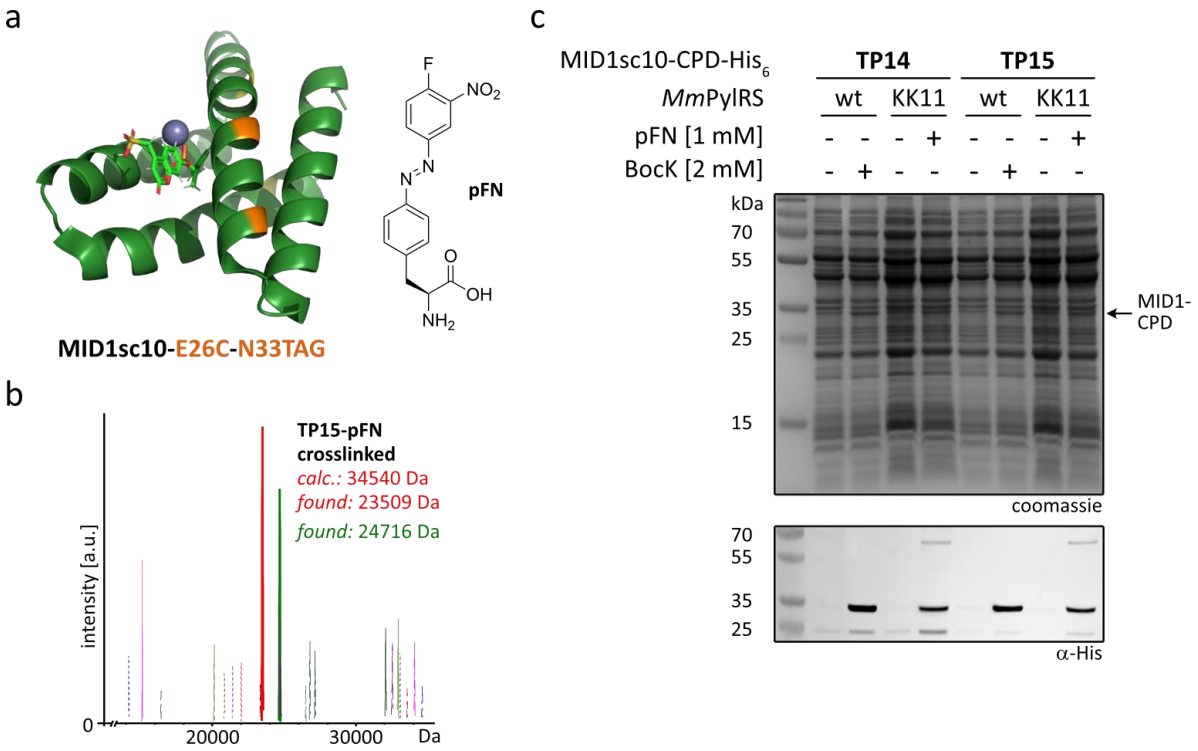


Figure 3.28 | Amber suppression of exemplary MID1sc10-E26TAG-N33C-CPD-His_6 (TP14) and MID1sc10-E26C-N33TAG-CPD-His_6 (TP15) with the azo-ncAA pFN in *E. coli*. (**a**) Crystal structure of MID1sc10 (pdb: 5od1) binding the fluorogenic coumarin ester of 2-phenylpropionate as substrate. The TAG- and Cys-positions are depicted in orange, the Zn^{2+}-ion is shown as a grey sphere. The abbreviated structural formula of **pFN** is shown on the right side. (**b**) LR-MS analysis of MID1sc10-E26C-N33TAG-CPD-His_6 (**TP15**) amber suppressed with **pFN** using the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). (**c**) Coomassie-stained SDS-PAGE and western blot analysis of **pFN** incorporation into both MID1sc10 variants by *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). BocK, incorporated by the wt PylRS was used as positive control.

To our dismay, only dirty, if barely any protein at all could be obtained from purification efforts via affinity chromatography (Supplementary Figure III.9c). The LR-MS data of exemplary, amber suppressed MID1sc10-E26C-N33TAG-CPD-His_6 (**TP15**) was inconclusive as well and did not correspond to the right mass at all (Figure 3.28b). We feared that these issues were a result of the catalytic cysteine amino acid residue within CPD, potentially crosslinking with our azo-ncAA **pFN**.

We therefore designed an additional set of TAG-Cys mutants using the original expression construct His_6-MBP-TEV-MID1sc10 (Supplementary Table IV.6). In those mutants, the azobenzene was once more placed onto more solvent-exposed areas of α-helices. Again, no decent amber suppression levels could be detected using **pFN** incorporated by the *Mm*PylRS KK11. Promising amber suppression results were only obtained for a single mutant designed to promote inter-crosslinking between two different α-helices (Figure 3.29a, c).

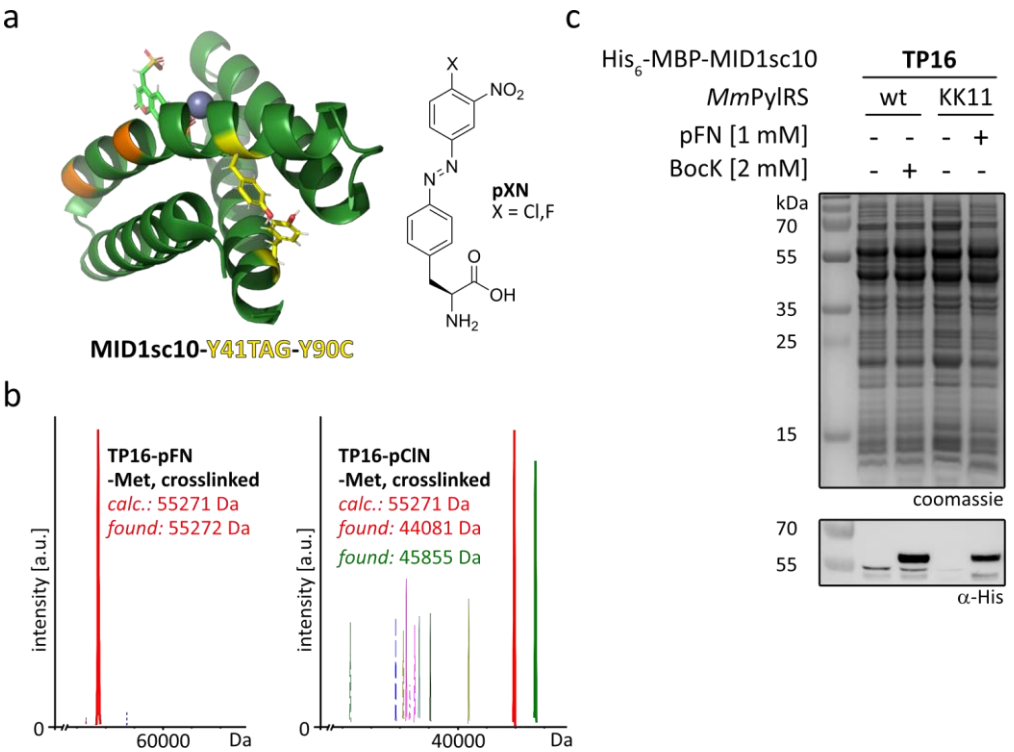


Figure 3.29 | Amber suppression of His_6-MBP-TEV-MID1sc10-Y41TAG-Y90C (TP16) with the azo-ncAA pFN in *E. coli*. (**a**) Crystal structure of MID1sc10 (pdb: 5od1) binding the fluorogenic ester substrate. The tyrosine amino acid residues are shown in yellow. The TAG- and Cys-positions of previously explored MID1sc10-E26TAG-N33C are depicted in orange, the Zn^{2+}-ion is shown as a grey sphere. The abbreviated structural formula of pXN (X = Cl,F) is shown on the right side. (**b**) LR-MS analysis of His_6-MBP-TEV-MID1sc10-Y41TAG-Y90C (**TP16**) amber suppressed with pXN (X = Cl,F)using the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). (**c**) Coomassie-stained SDS-PAGE and western blot analysis of pFN incorporation into MID1sc10-Y41TAG-Y90C by *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). BocK, incorporated by the wt PylRS was used as positive control.

By replacing two aromatic tyrosine amino acid residues we expected facile amber suppression due to the azobenzene's resemblance in size and polarity. We further envisioned that photoisomerization of such a crosslinked azobenzene would lead to a distortion in α-helix alignment and thus substrate binding. LR-MS data obtained of purified His_6-MBP-TEV-MID1sc10-Y41TAG-Y90C (**TP16**) indicated full formation of a crosslink using the azo-ncAA pFN (Figure 3.29b). Amber suppression using **pClN** however gave C-terminally degraded protein fragments that did not correspond to truncated protein (Figure 3.29b).

Since protein purification using the N-terminally His_6-tagged MBP-fusion construct is very low-yielding (see Table 3.3), we tested amber suppression of MID1sc10-Y41TAG-Y90C as C-terminally His_6-tagged and CPD-fusion construct (**TP17**) instead. While no protein was observed using MID1sc10-Y41TAG-Y90C-His_6, two main protein bands were obtained from purifying amber suppressed MID1sc10-Y41TAG-Y90C-CPD-His_6 (**TP17**) via affinity chromatography (Figure 3.30a).

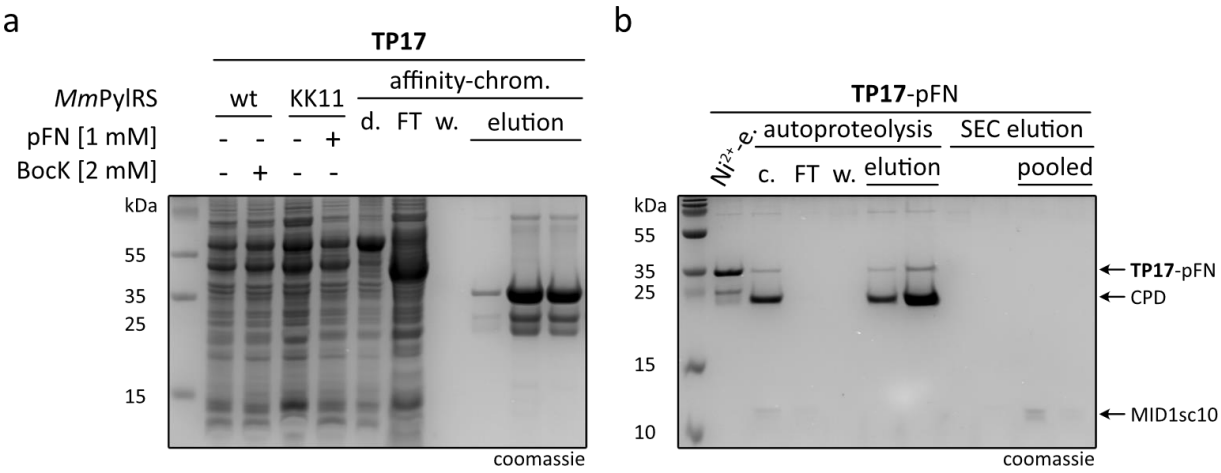


Figure 3.30 | Purification and autoproteolysis of MID1sc10-Y41TAG-Y90C-CPD-His$_6$ (TP17) amber suppressed with the azo-ncAA pFN. (**a**) Coomassie-stained SDS-PAGE analysis of MID1sc10-Y41TAG-Y90C (**TP17**) amber suppression and purification via affinity-chromatography. d. = debris after sonication, FT = flow-through, w. = wash. (**b**) Coomassie-stained SDS-PAGE analysis of the autoproteolysis and subsequent purification via SEC. Ni^{2+}-e. shows the amber suppressed MID1sc10 variant prior to autoproteolysis. c. = autoproteolysis after 1.5 h, FT = flow-through, w. = wash.

Further processing of the obtained protein mix via autoproteolysis, subsequent reverse affinity chromatography and final purification via SEC yielded barely detectable amounts of protein (Figure 3.30b). Concentrating the eluted protein fractions further led to precipitation so that the desired MID1sc10 variant bearing the azo-ncAA **pFN** could not be isolated. Incorporation of azo-ncAAs into MID1sc10 presumably destabilizes the protein, which would be in correspondence with protein precipitation and degraded protein fragments being observed during protein purifications (Figure 3.29b).

In alignment with results obtained for affibody, Rap80 and AltTPase, amber suppression and protein purification yields of MID1sc10 were continuously bad. We assumed that incorporation of azo-ncAAs into small α-helical bundles predominantly resulted in the expression of misfolded target proteins (Chapter 3.3.1). This in turn is associated with low expression yields, purification issues, protein precipitation and degradation. We reasoned that by using larger and more flexible protein folds in comparison to structurally restricted α-helical coiled-coils, misfolding upon azo-ncAA incorporation could be decreased or even avoided.

While this assumption widens the scope of potential protein targets for amber suppression with azo-ncAAs, suitable targets would still need to be easy to express and purify. In general, this leads to a limit in size, which is also beneficial during analysis of amber suppression efficiency using standard biochemical methods as well as LR-MS. Homo-oligomers made up of several, sequentially identical proteins, can unite many of these advantageous key features.

ClpP

The oligomeric caseinolytic protease proteolytic subunit (ClpP) was investigated in cooperation with the Sieber group from TU Munich since it unites many of the desirable key features. The tetradecameric serine protease is a highly conserved virulence regulator in bacteria such as *Staphylococcus aureus* (*Sa*). Its monomers assemble into a double-ring cylinder in which proteolytic degradation takes place (Figure 3.31).

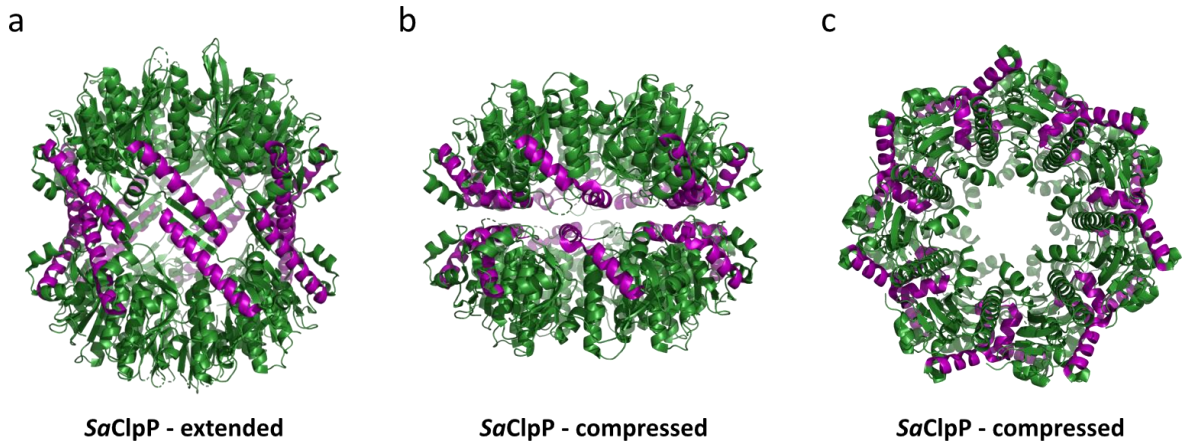


Figure 3.31 | Tetradecameric *Sa*ClpP in its active, extended conformation and in its inactive, compressed conformation. The α-helical motif helix E (133-158) is shown in purple. (**a**) Crystal structure of *Sa*ClpP in its extended state (pdb: 3sta) with helix E adopting a straight conformation. (**b**) Crystal structure of *Sa*ClpP in its compressed state (pdb: 3st9) with helix E adopting a kinked conformation around Lys145. (**c**) Top-view of *Sa*ClpP in its compressed state (pdb: 3st9).

ClpP can do so by itself, showing peptidase activity against fully unfolded peptides, or in association with ATP-driven hexameric chaperone ClpX or ClpA from the AAA+ family. Within the proteolytic machinery ClpXP, ClpX first binds and unfolds the target protein. Subsequent translocation to ClpP then leads to degradation of the unfolded protein. The two heptameric rings of ClpP interact via dynamic handle domains that comprise an α-helical structural motif, termed helix E (133-158). It has been shown, that helix E can undergo dramatic conformational changes that regulate enzymatic activity. While helix E motifs of *Sa*ClpP are extended if in the proteolytically active state (Figure 3.31a), they are interrupted by a kink around Lys145 when in the inactive state (Figure 3.31b, c). Thereby two short helices are formed, resulting in a compressed conformation in which the catalytic triads are misaligned. The dependency of proteolytic activity on this α-helical structural element provided the rationale for incorporation of our azo-ncAAs into *Sa*ClpP.[294, 295]

Early screenings for incorporation of the ncAA BocK using wt *Mm*PylRS into C-terminally His_6-tagged *Sa*ClpP revealed decent amber suppression levels for a range of ClpP-N141TAG and ClpP-R152TAG variants (Figure 3.32a, b). We then investigated incorporation of the azo-ncAAs **TriF** and **m2Cl** into ClpP-R152TAG and ClpP-N141TAG variants, respectively (Figure 3.32c-e). Placement of these two azo-ncAAs was chosen according to computational simulations performed by the Kaila group from Stockholm University (former TU Munich). Herein it was suggested that either a spacing of *i,i+11* using **TriF** or *i,i+4* and *i,i+7* using **m2Cl** would have the biggest effect on α-helix E conformation upon photoisomerization. Analysis via HR-MS of respective purified full-length ClpP variants bearing **TriF** yielded degraded protein fragments that could not be identified (Figure 3.32d). Although HR-MS data obtained for ClpP variants bearing **m2Cl** did show successful incorporation of the target azo-ncAA, no crosslink with the proximal cysteine could be observed (Figure 3.32e).

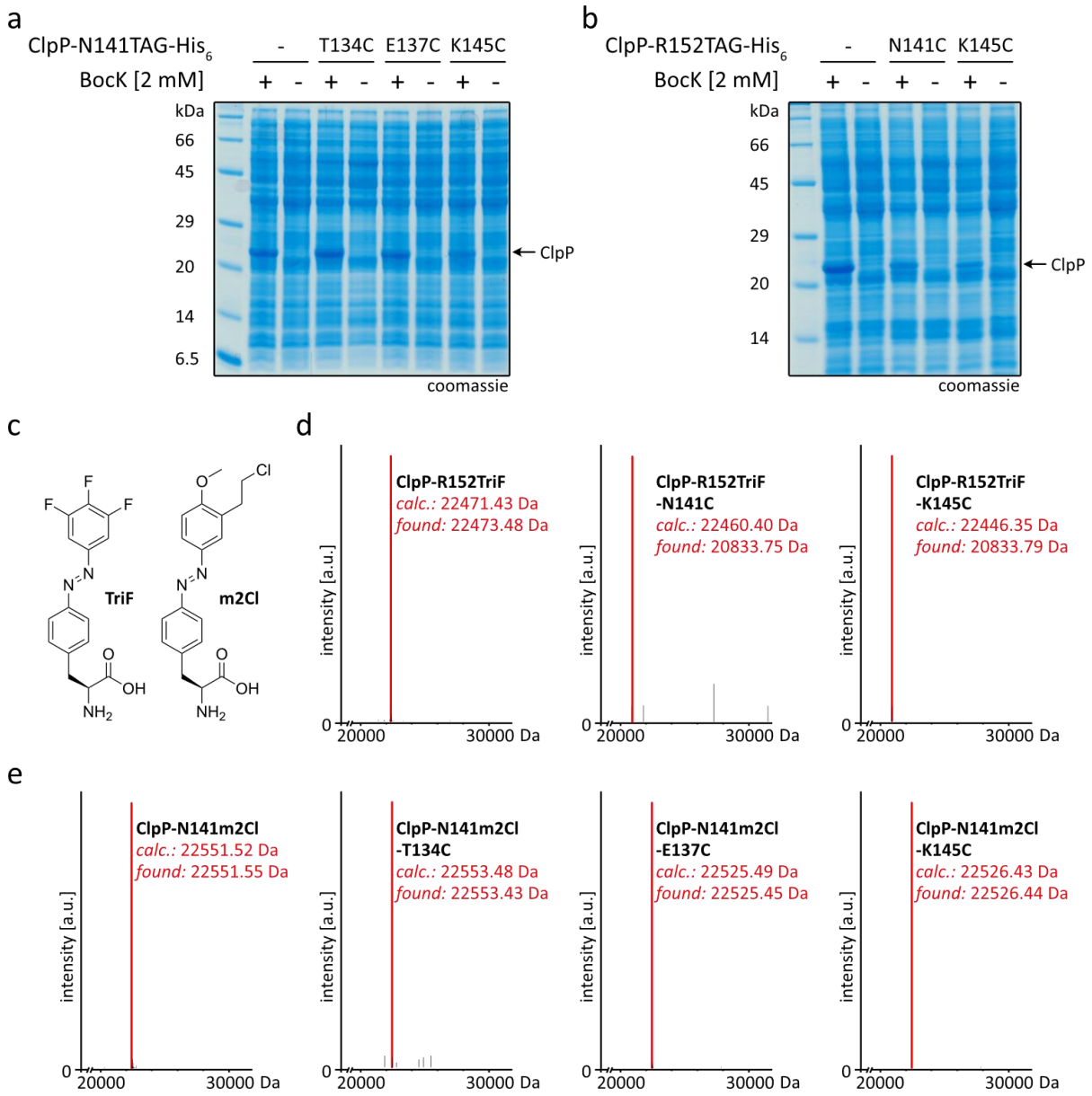


Figure 3.32 | Amber suppression of *Sa*ClpP-His$_6$ variants with the azo-ncAAs TriF and m2Cl in *E. coli*. (**a, b**) Coomassie-stained SDS-PAGE analysis of BocK incorporation into C-terminally His$_6$-tagged *Sa*ClpP variants by wt *Mm*PylRS. (**c**) Structural formulas of the azo-ncAAs **TriF** and **m2Cl**. (**d**) HR-MS analysis of *Sa*ClpP-R152TAG-His$_6$ variants amber suppressed with **TriF** using the *Mm*PylRS AzoC. (**e**) HR-MS analysis of *Sa*ClpP-N141TAG-His$_6$ variants amber suppressed with **m2Cl** using the *Mm*PylRS AzoC.

Since amber suppression of ClpP or crosslinking using the azo-ncAA **TriF** and **m2Cl** was unsuccessful, we decided to explore the possibility of azo-crosslinkers to test the best positions for optical control over helix E conformation (Supplementary Figure III.10). To this end, we expressed a range of C-terminally StrepII-tagged *Sa*ClpP variants bearing cysteine mutations according to *i,i+4*, *i,i+7* or *i,i+11*. Although decent expression levels could be observed for all cysteine-mutants (Supplementary Figure III.10b), SEC elution profiles and SDS-PAGE analysis of purified protein variants hinted towards the formation of protein aggregates in comparison to wt *Sa*ClpP-StrepII (Supplementary Figure III.10c, d). In addition, attempts at bioconjugation of a symmetric chloro-acetamide bearing azo-crosslinker (**CAC**) as reported by Yasuike et al.,[296] failed (Supplementary Figure III.10a). Bioconjugation of cysteine-reactive crosslinkers typically yields heterogenous mixtures of crosslinked proteins. Unsurprisingly, similar observations were made for *Sa*ClpP, a tetradecameric protein with the target α-helix motif of monomers in close proximity to each other (Figure 3.31).

Consequently, a considerable amount of effort would have to be spent on (i) improving the bioconjugation procedure for *Sa*ClpP and (ii) separating intramolecularly crosslinked monomers from intermolecularly crosslinked dimers. Instead we focused on the development of more reactive azo-ncAAs such as **m1F** (Chapter 3.1.4) and **pFN** (Chapter 3.1.6) that were expected to crosslink more efficiently as compared to **m2Cl** (Figure 3.32e).

In parallel, we performed *in vitro* peptidase activity assays as reported by Lakemeyer et al. with the purified *Sa*ClpP variants bearing two cysteine mutations (Figure 3.33).[297] We thereby wanted to estimate which positions within helix E would tolerate our azo-ncAAs best.

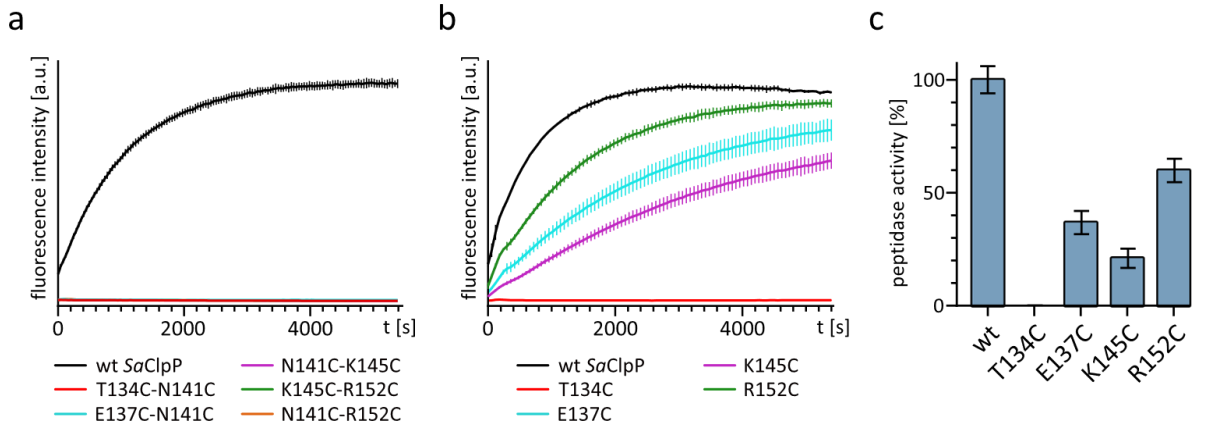


Figure 3.33 | *Sa*ClpP *in vitro* peptidase assays in ClpP assay buffer containing 1% (v/v) DMSO. Fluorescence measurements (λ_{Ex} = 380 nm, λ_{Em} = 430 nm) were performed in triplicates at 32 °C. (**a**) Peptidolysis of the published substrate *N*-Ac-Ala-hArg-2-Aoc-ACC[297] (200 µM) by *Sa*ClpP variants (1 µM) bearing two cysteine mutations in helix E. (**b**) Peptidolysis of the published substrate *N*-Ac-Ala-hArg-2-Aoc-ACC[297] (200 µM) by *Sa*ClpP variants (1 µM) bearing a single cysteine mutation in helix E. (**c**) The initial slope (0-540 s) of the fluorescence over time signal of **b** was calculated by linear regression fitting using GraphPad Prism. The wt *Sa*ClpP-StrepII control sample was normalized to 100% activity and the residual activity of variants bearing a single cysteine mutation was determined.

Unfortunately, none of the tested mutants were enzymatically active (Figure 3.33a). Since a double-cysteine mutant might not necessarily resemble *Sa*ClpP variants bearing an azo-ncAA and a single cysteine mutation, we also investigated the peptidase activity of several *Sa*ClpP variants bearing only one cysteine mutation (Figure 3.33b). Three out of four tested mutants retained between 22-60 % of their enzymatic activity in comparison to wt *Sa*ClpP (Figure 3.33c). Although encouraging, these results might not necessarily apply to *Sa*ClpP variants bearing crosslinked azo-ncAAs.

Still, we tested amber suppression of C-terminally StrepII-tagged ClpP-N141TAG-R152C (**TP18**) and ClpP-K145TAG-R152C (**TP19**) with the azo-ncAA **m1F** using the *Mm*PylRS AzoC (Figure 3.34). As already mentioned before (Chapter 3.1.4), application of this particular azo-ncAA was tedious since precipitation of the free azo-ncAA also led to heavy precipitation of the hydrophobically interacting protein. This resulted in diminished protein yields ranging from 0.3 to 1.0 mg/L (estimated via densitometry) and also complicated analysis of full-length protein via LR-MS. While no crosslink was observed for the ClpP variant featuring a spacing of *i,i+11* (**TP18**), ClpP-K145TAG-R152C-StrepII (**TP19**) amber suppressed with **m1F** yielded a partial crosslink (Figure 3.34d, e). Due to these results and difficulties during protein purification when employing this particular azo-ncAA, we instead focused on amber suppression of *Sa*ClpP using **pFN**.

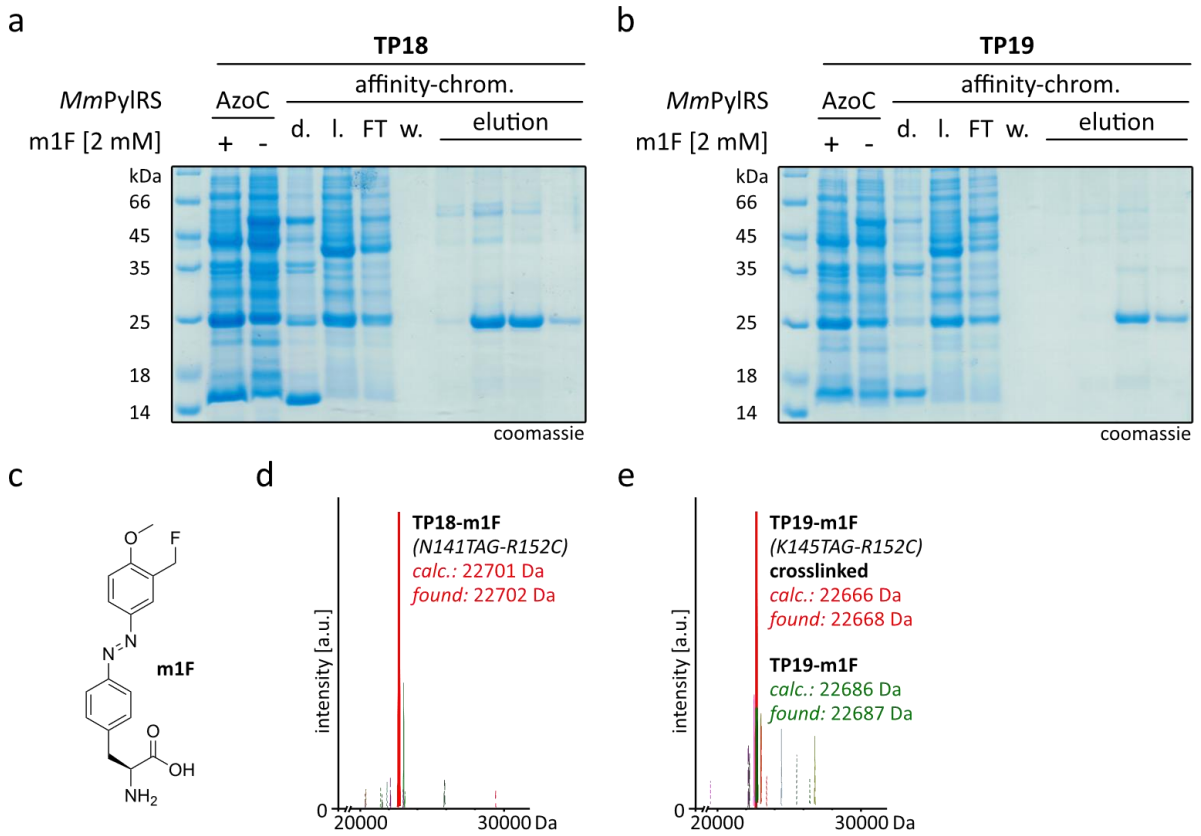


Figure 3.34 | Amber suppression of ClpP-N141TAG-R152C-StrepII (TP18) and ClpP-K145TAG-R152C-StrepII (TP19) with the azo-ncAA m1F in *E. coli*. (**a, b**) Coomassie-stained SDS-PAGE analysis of *Sa*ClpP variant (**TP18** and **TP19**) amber suppression and purification via affinity-chromatography. d. shows the debris after sonication. l. = lysate, FT = flow-through, w. = wash. (**c**) Structural formula of the azo-ncAA **m1F**. (**d**) LR-MS analysis of ClpP-N141TAG-R152C-StrepII (**TP18**) amber suppressed with m1F using the *Mm*PylRS AzoC. (**e**) LR-MS analysis of ClpP-K145TAG-R152C-StrepII (**TP19**) amber suppressed with **m1F** using the *Mm*PylRS AzoC.

We first investigated four C-terminally StrepII-tagged ClpP variants bearing the crosslinked azo-ncAA **pFN** at R152 (Figure 3.35), since mutations at this position were tolerated fairly well (Figure 3.33c). Again, no crosslink was observed for ClpP variants featuring a spacing of *i,i+11* at N141-R152 (Figure 3.35c). Both mutants (**TP20** and **TP18**) were of particular interest to us, because early computational simulations by the Kaila group from Stockholm University (former TU Munich) predicted large conformational changes within helix E upon photoisomerization of the similar azo-ncAA scaffold **TriF**. In contrast, full crosslinking was detected for both ClpP-K145TAG-R152C (**TP19**) and ClpP-K145C-R152TAG (**TP21**) featuring a spacing of *i,i+7* (Figure 3.35d). Unfortunately, according to computational simulations, photoisomerization of the crosslinked azo-ncAA **pFN** spaced by *i,i+7* would have no to little effect on the conformation of helix E.

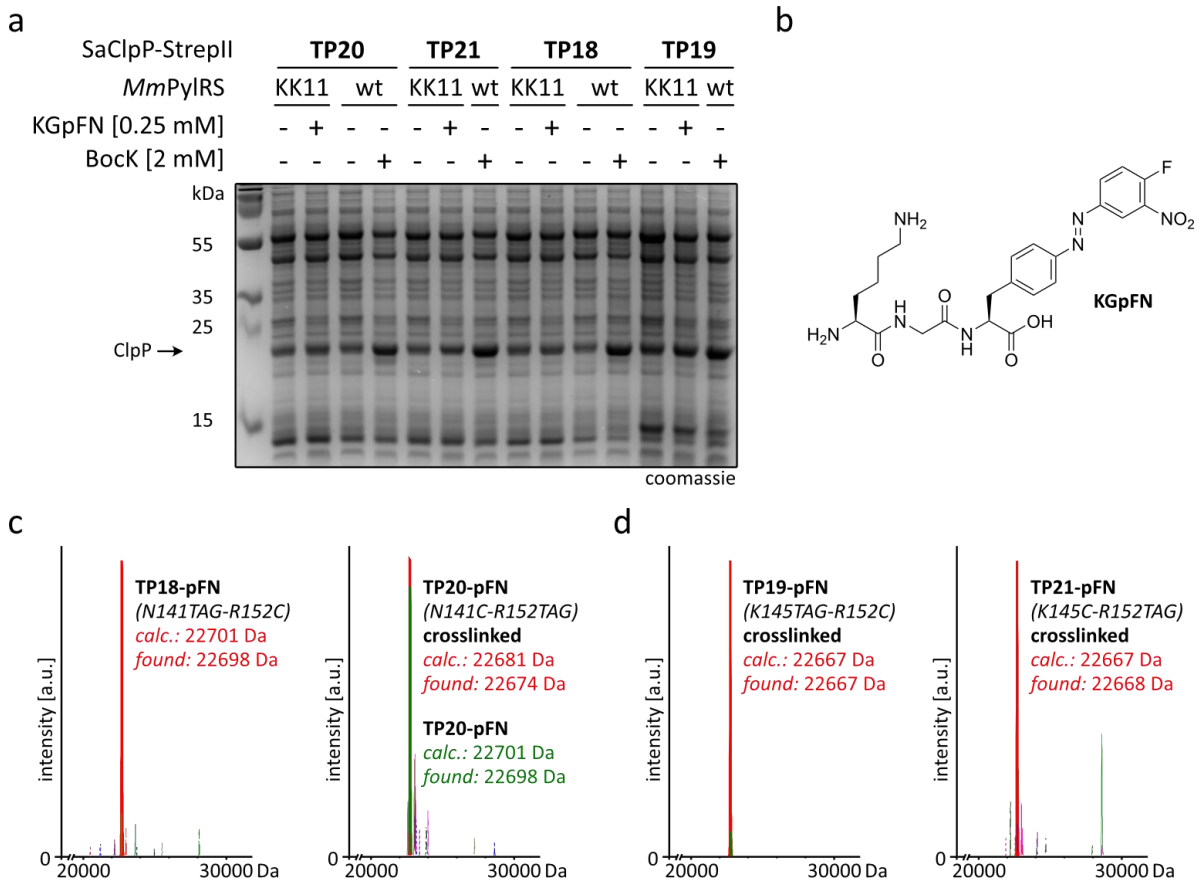


Figure 3.35 | Amber suppression of *Sa*ClpP variants (TP18-21) with the azo-ncAA pFN in *E. coli*. (**a**) Coomassie-stained SDS-PAGE analysis of *Sa*ClpP variants (**TP18-21**) amber suppressed with **pFN** incorporated by the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). BocK, incorporated by the wt PylRS was used as positive control. (**b**) Structural formula of the tripeptide **KGpFN** bearing the azo-ncAA **pFN**. (**c**) LR-MS analysis of ClpP-N141TAG-R152C-StrepII (**TP18**) **and** ClpP-N141C-R152TAG-StrepII (**TP20**) amber suppressed with **pFN** using the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB). (**d**) LR-MS analysis of ClpP-K145TAG-R152C-StrepII (**TP19**) **and** ClpP-K145C-R152TAG-StrepII (**TP21**) amber suppressed with **pFN** using the *Mm*PylRS KK11 in the *E. coli* strain K12 (ΔnfsA + ΔnfsB).

Further efforts will be focusing on experimentally validating the enzymatic activity of ClpP variants **TP19** and **TP21** bearing the crosslinked azo-ncAA **pFN**. To this end it might be beneficial to switch to *in vitro* protease activity assays using the *S. aureus* chaperone ClpX. Similar assays have already been published, investigating the degradation of SsrA-tagged eGFP.[298] This way, fluorescence measurements could be exerted at wavelengths compatible with wavelengths used during photoisomerization of model azo-ncAA **pSEtN** (Chapter 3.2). It would also prove valuable to incorporate **pFN** into ClpP variants featuring a spacing of *i,i+4* instead, such as E137-N141, N141-K145, E148-R152 or R152-E156. All of the mentioned positions could already be amber suppressed using BocK incorporated by the wt *Mm*PylRS (Figure 3.32a). If successfully crosslinked, photoisomerization to the *cis*-form would allow for the accommodation of amino acid side chains, while hopefully disrupting α-helix formation in the *trans*-form.[148]

3.3.3 Reversible Optical Control over Protein Oligomerization

Beyond optically controlling enzymatic activity of oligomeric proteins, we were curios if reversible switching of protein function could be coupled to the oligomerization state itself. The artificial protein cage OP developed by the Hilvert group from ETH Zurich presented an interesting target to explore this idea.

Inspired by viral genome packaging particles, the engineered protein capsid OP was initially developed to pack and deliver oligonucleotides such as short interfering RNA (siRNA) to mammalian cells.[299] Subsequent cargo release by diffusion led to efficient gene knockdown. The porous protein capsule comprises 24 identical monomers and features a highly positively charged interior (Figure 3.36a).

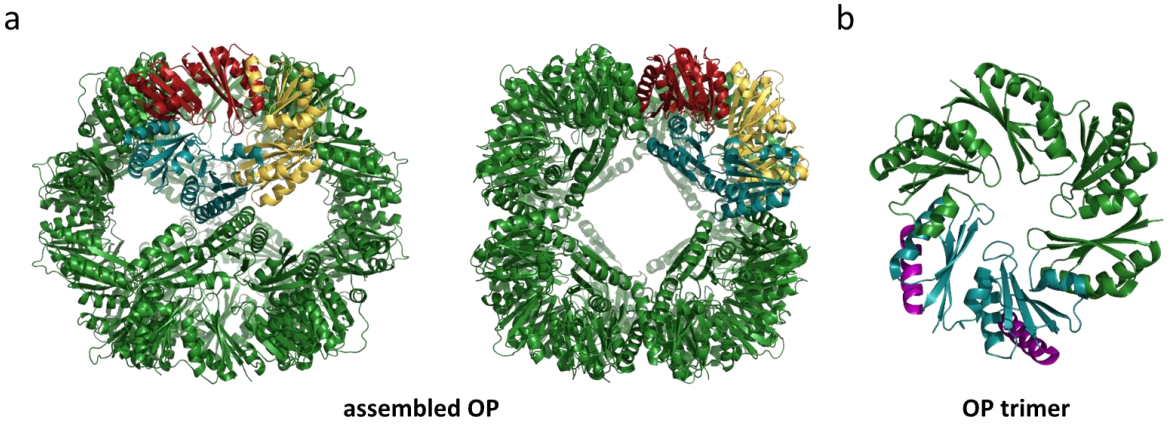


Figure 3.36 | Oligomeric protein cage OP. (**a**) Crystal structure of fully assembled OP cage with one trimer shown as red, yellow and cyan monomers (pdb: 6fdb). (**b**) Crystal structure of OP trimer with one single monomer shown in cyan. The α-helices targeted for amber suppression with our azo-ncAAs are shown in purple.

Further development of this self-assembling protein cage enabled delivery of cargo molecules beyond negatively charged oligonucleotides. Using a two-tier encapsulation concept, the OP cavity is first encapsulated with anionic surfactants, thereby generating a hydrophobic core. This allows for the packing of nonpolar small molecules in a second encapsulation step.[300]

By installing azo-ncAAs into OP monomers we desired to gain spatio-temporal control over protein cage disassembly and thus cargo delivery. A similar strategy has already been realized by Witzigmann and co-workers using photoswitchable phosphatidylcholine analogs to control drug release from lipid nanoparticles.[95] Beyond that, an engineered viral vector[301] and protein chaperonin[302] have been used to drive the development towards optically controlled cargo delivery systems. Using the artificial protein cage OP instead would remove the need for stable lipid nanoparticle pre-formation and broaden the scope of potential cargo molecules beyond proteins and nucleic acids.

We therefore anticipated that the lipoprotein-resembling OP would highly benefit from the installation of our azo-ncAAs. In cooperation with the Hilvert group from ETH Zurich, we decided to focus on two different α-helices for amber suppression which were known to tolerate single point mutations without impeding protein cage assembly (Figure 3.36b). Initial test expressions of C-terminally His_6-tagged wt OP hinted at diminished protein expression levels in the *E. coli* strain K12 which was typically used for amber suppression within this work (Figure 3.37a). In comparison, protein expression of OP cloned downstream of a T7 promoter, an expression construct kindly provided by T. Edwardson (Hilvert group, ETH Zurich), yielded decent amounts of wt OP in the *E. coli* strain BL21(DE3) (Figure 3.37a).

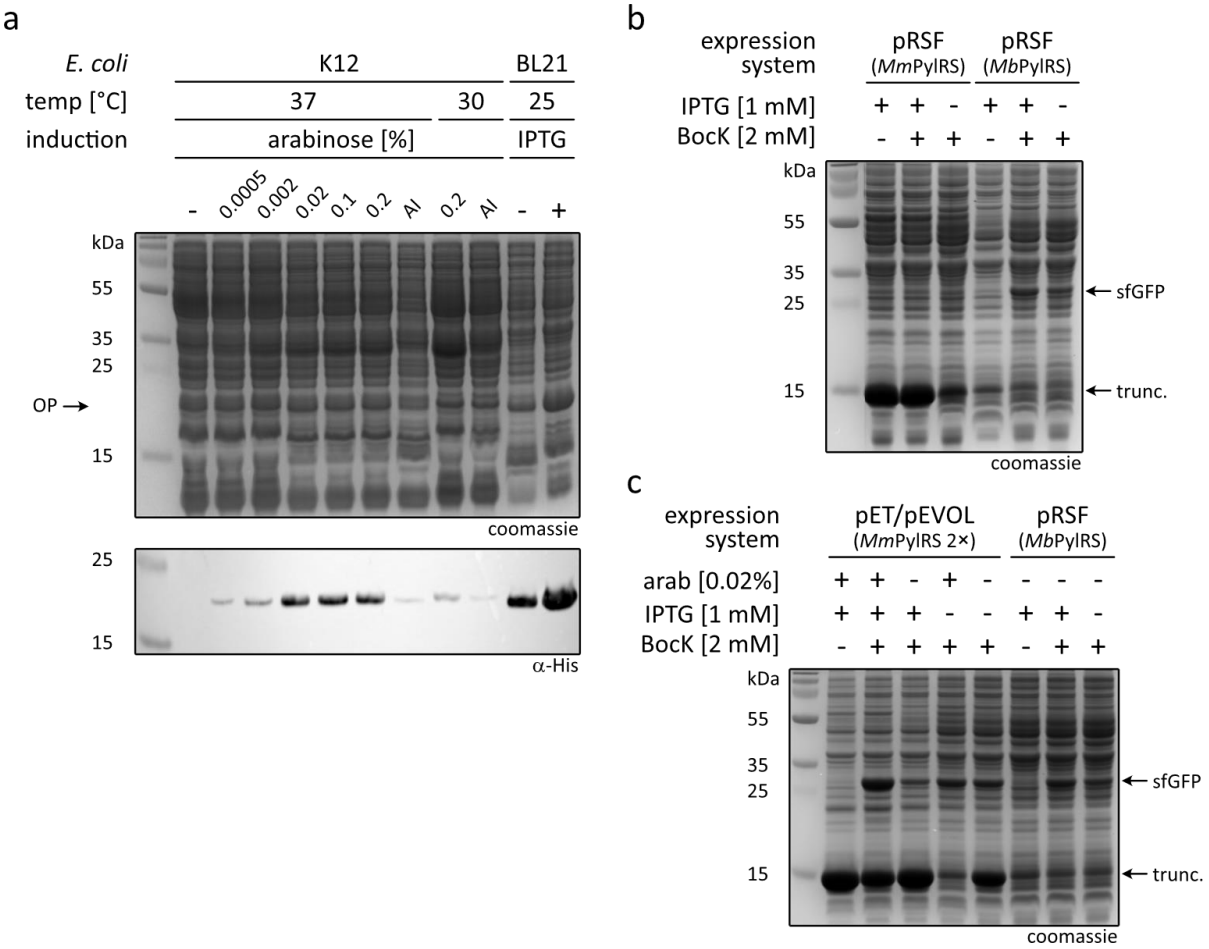


Figure 3.37 | Test expression of wt OP-His$_6$ and establishment of suitable amber suppression conditions in the *E. coli* strain BL21(DE3). (**a**) Coomassie-stained SDS-PAGE and western blot analysis of wt OP-His$_6$ expression in the *E. coli* strain K12. Various conditions were screened, including arabinose content and expression temperature. AI denotes auto-inducing medium with 0.05% arabinose. Expression of wt OP-His$_6$ in the *E. coli* strain BL21(DE3) was induced with 0.1 mM IPTG and used as control. (**b**) Coomassie-stained SDS-PAGE analysis of sfGFP-N149TAG-His$_6$ amber suppressed with BocK using the wt *Mm*PylRS (pRSFDuet) in BL21(DE3). Amber suppression using the pRSFDuet vector containing the *Mb*PylRS, which was already established within our lab, was used as positive control. (**c**) Coomassie-stained SDS-PAGE of sfGFP-N149TAG-His$_6$ amber suppressed with BocK using the wt *Mm*PylRS (pEVOL) expressed as two copies in BL21(DE3). Amber suppression using the pRSFDuet vector containing the *Mb*PylRS, which was already established within our lab, was used as a comparison.

Consequently, a different protein expression system for the amber suppression of OP with our azo-ncAAs in *E. Coli* BL21(DE3) had to be established. Two different expression systems were tested for amber suppression of C-terminally His$_6$-tagged sfGFP-N149TAG with BocK using the wt *Mm*PylRS in BL21(DE3) (Figure 3.37b, c). The first system comprised two different plasmids, one encoding the *Mm*PylRS (two copies) under an araBAD promotor (pEVOL), the second plasmid bearing the gene of the target protein OP downstream of a T7 promoter (pET). The second system was based on a single plasmid (pRSFDuet) encoding two multiple cloning sites each of which is preceded by a T7 promotor.

Since only the expression vectors based on pET/pEVOL yielded decent amounts of sfGFP-N149TAG-His$_6$ amber suppressed with BocK, we proceeded using pEVOL plasmids bearing the desired *Mm*PylRS variants which were already available within the lab. Since the respective *E. coli* strain BL21 (DE3, ΔnfsA + ΔnfsB) for amber suppression of OP with our reduction-sensitive azo-ncAAs **pFN** and **pClN** did not exist yet, we first investigated amber suppression of OP with **TriF**. The azo-ncAA **TriF** has been shown to yield similar expression levels as **pFN** when using the *Mm*PylRS KK11 (Figure 3.13 and Figure 3.17), thereby allowing us to estimate whether OP can be amber suppressed with our desired azo-ncAA **pFN**.

To our disappointment, only few out of eight mutants showed detectable amounts of full-length protein with the azo-ncAA **TriF** (Figure 3.38c and Supplementary Table IV.8). Interestingly, most of them seemed to feature read-through of the amber stop codon when expressed in absence of ncAA using the *Mm*PylRS KK11 or even wt *Mm*PylRS (**TP25**, Figure 3.38c, right panel).

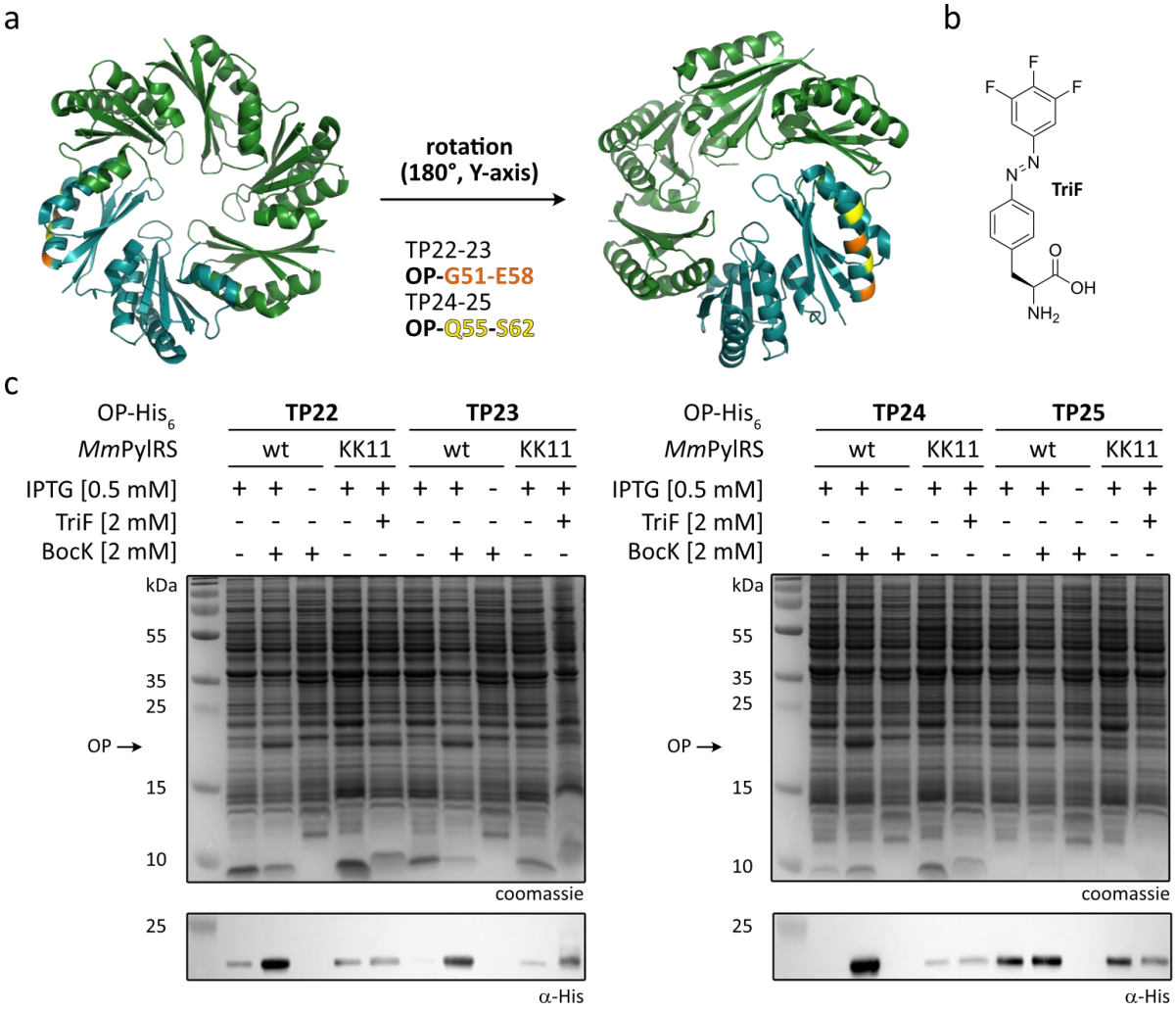


Figure 3.38 | Amber suppression of exemplary OP-His$_6$ variants (TP22-25) with the azo-ncAA TriF in the *E. coli* strain BL21(DE3) using pET/pEVOL expression vectors. (**a**) Crystal structure of OP trimer with one single monomer shown in cyan (pdb: 6fdb). The respective TAG and Cys-mutations are shown in orange and yellow. (**b**) Structural formula of the azo-ncAA **TriF**. (**c**) Coomassie-stained SDS-PAGE and western blot analysis of TriF incorporation into OP-G51TAG-E58C-His$_6$ (**TP22**), OP-G51C-E58TAG-His$_6$ (**TP23**), OP-Q55TAG-S62C-His$_6$ (**TP24**) and OP-Q55C-S62TAG-His$_6$ (**TP25**) by *Mm*PylRS KK11 in the *E. coli* strain BL21(DE3). BocK incorporated by wt PylRS was used as positive control.

In parallel, we investigated whether a nitro reductase single gene knockout of either *nfsA* or *nfsB* would be sufficient for stable incorporation of the reduction-sensitive analogs pXN in BL21(DE3). To this end, we performed amber suppression experiments of C-terminally His$_6$-tagged sfGFP-N149TAG with **pClN** in the Keio single gene knockouts K12 (ΔnfsA) and K12 (ΔnfsB).[266] The obtained LR-MS data of purified sfGFP-N149TAG-His$_6$ variants indicated partial reduction of the nitro-moiety by nfsA and complete reduction by nfsB (Supplementary Figure III.4). Since the K12 double knockout (ΔnfsA + ΔnfsB) strain has been successfully employed to incorporate the desired azo-ncAAs **pClN** and **pFN** (Chapter 3.1.6), we suspected that a respective double gene knockout of *nfsA* and *nfsB* in BL21(DE3) would also yield stable incorporation into OP.

Therefore, it might still prove valuable to prepare the *E. coli* double knockout BL21 (DE3, ΔnfsA + ΔnfsB) and test amber suppression of OP with our fully crosslinking azo-ncAA **pFN**. We decided however, to focus on more promising target proteins such as PDZ3 (Chapter 3.3.1) and *Sa*ClpP (Chapter 3.3.2), for which **pFN** is already easily incorporated in decent yields without reduction of the nitro moiety.

4 Summary of Results and Outlook

In summary, we designed and synthesized a broad range of novel azo-ncAAs featuring moderate to high reactivity towards proximal nucleophiles such as the cysteine side chain (Chapter 3.1.3). With five azo-ncAAs getting incorporated into proteins by variants of *Mm*PylRS in *E. coli*, we substantially contribute to the modest number of photoswitchable modulators used in GCE (Figure 4.1a).

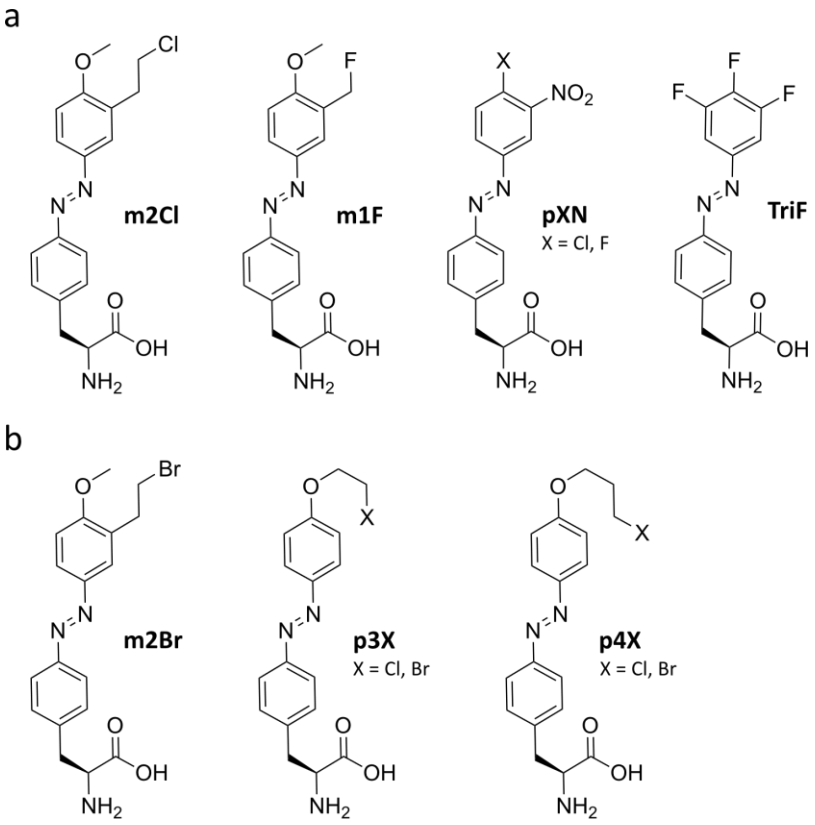


Figure 4.1 | Structural formulas of azo-ncAAs for amber suppression developed in this project. (**a**) Structural formulas of azo-ncAAs **m2Cl**, **m1F**, **pFN**, **pClN** and **TriF** that have successfully been incorporated into target proteins using the *Mm*PylRS AzoC and KK11 in *E. coli*. The azo-ncAAs **m1F** and **TriF** were further successfully incorporated into sfGFP-N149TAG-His_6 using the *Mm*PylRS KK11 in mammalian cells. (**b**) Structural formulas of azo-ncAAs **m2Br**, **p3Cl**, **p3Br**, **p4Cl** and **p4Br** that hold great potential for the optical modulation of protein function if suitable PylRS/tRNA pairs can be found.

Further, we also succeeded in incorporating two of these azo-ncAAs, **TriF** and **m1F**, into sfGFP-N149TAG-His_6 in HEK293T cells. The respective mammalian cell experiments were kindly performed by colleague J. Heimgärtner. Finally, five more derivatives, namely **m2Br**, **p3Cl**, **p3Br**, **p4Cl** and **p4Br** hold great potential of becoming valuable tools, if appropriate PylRS/tRNA pairs can be found or engineered (Figure 4.1b). All five derivatives feature very good photophysical properties and medium to high reactivity towards nucleophiles.

We demonstrated the synthesis of basic S_NAr-reactive azo-ncAAs **pFN** and **pClN** over two steps via the Baeyer-Mills reaction in decent yields up to 40% (Chapter 3.1.6). The facile synthesis route makes these two azo-ncAAs easily accessible in multigram scale. Incorporation of **TriF** and the **pXN** derivatives was achieved in standard *E. coli* K12 or in the nitro reductase knockout strain K12(ΔnfsA + ΔnfsB), respectively. Beyond that, we synthesized a variety of azo-ncAAs bearing alkyl- and acyl-halide linkers for S_N2 reaction or Michael-acceptors for thiol-Michael addition (Chapter 3.1.4 and Chapter 3.1.5). To this end, we adapted the azo-coupling reaction for the facile preparation of *N*-Boc protected azo-ncAAs, which has not been done so far (Chapter 3.1.4).

Although the azo-coupling reaction was high-yielding using phenol analogs, we were not able to establish a similar synthesis route using aniline derivatives. Instead, we found the preparation of nitro-substituted azo-ncAAs with subsequent reduction of the nitro-moiety under mild conditions to be an elegant alternative synthesis route towards amino-substituted azo-ncAAs (Chapter 3.1.5). Unfortunately, no PylRS was found for the incorporation of azo-ncAAs bearing the cysteine-reactive linker in *para*-position to the diazene bond. This was valid for both *para*-alkoxy (Chapter 3.1.4) and *para*-amide substitutions (Chapter 3.1.5). A range of azo-ncAAs bearing the cysteine-reactive linker in *meta*-position to the diazene instead did get incorporated by the published *Mm*PylRS AzoC (A302T, L309A, I322T, N346A and C348G)[249] and KK11 (L305F, L309M, N346G, C348G and Y384F)[251]. Whereas amber suppression with azo-ncAAs of the mQX and mQR series resulted in cleavage of the linker-bearing amide bond (Chapter 3.1.5), derivatives of the mnX series yielded the desired, amber suppressed sfGFP variants (Chapter 3.1.4).

The azo-ncAAs of the mnX series featured an exceptional $PSS_{365\ nm}$ (90%) due to substitution with a *para*-methoxy group, which was exemplified for **m1F** (Chapter 3.2). In contrast, S_NAr-reactive azo-ncAAs featured a $PSS_{365\ nm}$ of only 60%. This was shown for the model compound **pSEtN**, mimicking the pXN analogs after successful crosslinking (Chapter 3.2). Consequently, the optical control over protein conformation and function that can be accomplished by application of this particular azo-ncAA scaffold is significantly diminished. This could also be observed for protein variants bearing the fully crosslinked azo-ncAA **pFN**, e.g. Cam or PDZ3 (Chapter 3.3.1). To our dismay, the remaining four azo-ncAAs **pClN**, **TriF**, **m2Cl** and **m1F** yielded only partial or no crosslinks at all when incorporated into target proteins bearing the optimally spaced cysteine mutation.

As a result, we focused on the optical control of protein function using the azo-ncAA **pFN**, for which full crosslinking with a proximal cysteine side chain was observed. We encountered various issues when incorporating the azo-ncAA **pFN** and its less reactive analog **pClN** into small, α-helical proteins. Amber suppression and purification of affibody, AltTPase and MID1sc10 was low-yielding and mostly resulted in the detection of proteins corresponding to a higher molecular weight. We suspected that incorporation of our azo-ncAAs into rigid α-helical coiled-coil motifs led to the expression of misfolded proteins, thus attracting a variety of chaperons. Consequently, we encountered low expression yields, purification issues, protein precipitation and degradation. While similar results were also observed for Rap80 bearing pFN, it might be worth investigating different expression constructs of Rap80. A fusion construct of Rap80 bearing an N-terminal small ubiquitin-related modifier (SUMO) and C-terminal CPD would boost expression yields and could further benefit α-helix folding. Similarly, a Rap80 fusion construct bearing a C-terminal monoUb and N-terminal monoUb could be considered. With an adequate linker, binding of both monoUbs to Rap80 might promote folding and thus prevent chaperones from binding to Rap80. If successful, 7A-Rap80 could provide an ideal target protein for the characterization of previously developed and novel azo-ncAAs. It features a segment in the middle of the α-helix that has high intrinsic helical propensity due to the introduction of seven alanine mutations. It has been shown, that optical modulation of α-helical conformation is most effective if this central region is C-terminally followed by a region of lower intrinsic helicity.[303]

In general, incorporation of azo-ncAAs into small α-helical proteins has proven problematic. The hydrophobic and ionic interactions found between different helices of coiled-coil motifs promote helical packing and thus yield highly stable protein scaffolds. As a result, optical modulation will be difficult to achieve without inserting additional point mutations to disrupt helical packing and/or helix propensity.

Similar to our approach using GCE, A. Karlström and co-workers employed an azo-crosslinker to generate photoreceptive Z domain variants.[156] Successful binding to an IgG-sepharose column was observed for two different azo-conjugated Z domain variants. The first variant Z_{C3} was bearing the azo-crosslinker in one of the two IgG-binding helices, the second variant Z_{C2} was designed for crosslinking in the non-binding, stabilizing third helix. The latter, however, did not elute upon photoisomerization to the *cis*-isomer. Although the elution assay was more successful for Z_{C3}, photoirradiation did not result in full elution of the bound Z domain variant.[156] Such reports further confirmed our assumption that the highly stable protein conformation of coiled-coil motifs cannot be easily disrupted using small molecule azobenzenes. Different strategies exploring photoreceptor fusions with the photoactivatable yellow protein (PYP) instead, appear to be much more promising. By inserting PYP into a loop of the Z domain, A. Woolly and co-workers created a photoswitchable affibody scaffold. Proper folding and thus binding of the Z domain to its target IgG could be initiated upon irradiation with blue light, which was further demonstrated *in vivo* using yeast two hybrid assays. Successful transfer of this design to a different affibody scaffold proved the broad versatility of the described approach, enabling the generation of photoswitchable affibody scaffolds for a broad range of targets.[304]

In contrast to α-helical coiled-coil motifs, the conformationally less rigid target proteins Cam, PDZ3 and ClpP were readily expressed and purified bearing the fully crosslinked azo-ncAA **pFN**. Analysis of photoisomerization efficiency within Cam and PDZ3 using CD-spectroscopy and UV-Vis spectrophotometry confirmed the obtained results on small molecule level (Chapter 3.2). No conformational change was observed upon photoisomerization during CD spectroscopy experiments. From the obtained UV-Vis spectra we deduced that photoisomerization of **pFN** was further restricted once crosslinked within the target protein. This led to an even further diminished $PSS_{365\,nm}$ (35% for Cam, 21% and 32% for PDZ3) compared to the one observed on small molecule level (60%). As reported in literature for less flexible azo-crosslinker, the rigidity of our azo-ncAA scaffold led to lower photoisomerization quantum yields when attached to the protein.[150] Here, it might prove beneficial to incorporate the rigid azo-ncAA **pFN** with a different spacing of *i,i+4.* In contrast to a S_NAr-reactive azo-ncAA spaced by *i,i+7*, we anticipate the *cis*-isomer spaced by *i,i+4* to stabilize α-helical conformation similarly to stapled peptides, thus increasing the change in concentration of the active species upon photoisomerization from 1.5-fold for *i,i+7* to potentially 35-fold for *i,i+4*.[279]

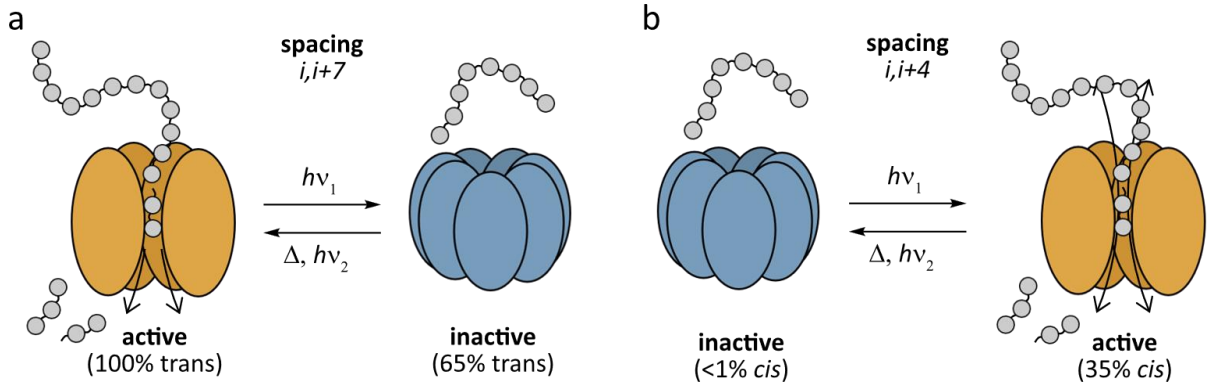


Figure 4.2 | Schematic depiction of spatio-temporal control over ClpP enzymatic activity with azo-ncAAs, dependent on the azo-ncAA and Cys spacing. (**a**) The *trans*-isomer stabilizes the α-helix when spaced according to *i,i+7*. Here, the dark-adapted form of ClpP is active and a 1.5-fold increase in concentration of the active species can be expected upon photoisomerization. (**b**) The *cis*-isomer stabilizes the α-helix for a spacing of *i,i+4*, rendering it inactive in the dark-adapted *trans*-form. Additionally, the change in concentration of the active species upon photoisomerization increases to potentially 35-fold.

A similar approach could be attempted for gaining spatio-temporal control over the oligomeric protease ClpP (Figure 4.2), although initial peptidase activity assays of cysteine mutants hint towards reduced peptidase activity also for amber suppressed variants (Chapter 3.3.2). The installation of an azo-bridge in α-helical motifs may not be the only way to control protein function. Instead, strategically placed non-crosslinking azo-ncAAs featuring an exceptional $PSS_{365\ nm}$, such as **m1F** or **m2Cl** (90%), could also enable spatio-temporal control over protein function. To this end, we explored the concept of photoswitchable anti-GFP nbs using the GFP-orthogonal azo-ncAA **m1F**, which further holds great potential for incorporation into mammalian cells (Chapter 3.3.1). Similar to photocaged nbs,[244, 245] we would anticipate robust function directly when expressed in an intracellular environment, thereby generating optically programmable tools for subcellular protein localization and cell signaling manipulation.

Still, further improving the photophysical properties of the azo-ncAA **pFN** is indispensable. Synthetic efforts towards a di-*ortho*-fluoro substituted analog of **pFN** (**F_2-pFN**, Figure 4.3, Chapter 3.1.7) would hopefully separate the respective absorption spectra of *trans*- and *cis*-isomer,[79] while still being incorporated by the same *Mm*PylRS KK11.

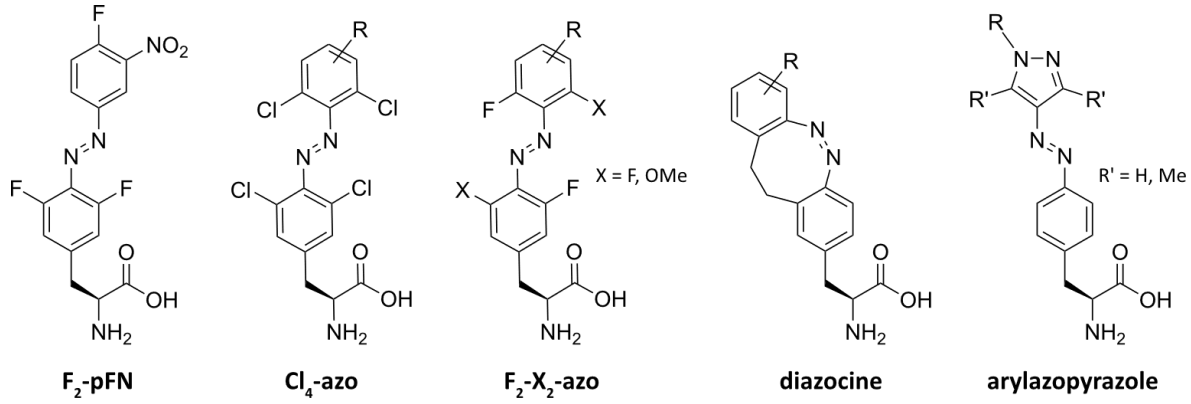


Figure 4.3 | Structural formulas of potential azo-ncAA scaffolds to spatio-temporal control protein function. R indicates potential decoration of the azo-ncAA core structures with Cys-reactive linkers.

Not only, would this increase the PSS but could also enable photoisomerization using wavelengths within the visible light spectrum. Since **pFN** was unfortunately the only azo-ncAA giving a full crosslink in target proteins, it could be advantageous to consider directed evolution approaches for azo-ncAA scaffolds bearing S_N2-reactive bromo-alkyl linkers in *para*- or *meta*-position (Figure 4.1b, Chapter 3.1.4). It has been shown that bromo-alkyl linkers can undergo facile proximity-induced reactions with cysteine side chains.[305]

Beyond that, the design of novel photoswitchable ncAA scaffolds could provide the necessary means to efficiently modulate protein conformation and function. While late-stage functionalization of azo-ncAAs towards tetra-*ortho*-chloro substituted derivatives has been unsuccessful so far (Chapter 3.1.7), it would give access to a range of azo-ncAAs with exceptional photophysical properties (**Cl_4-azo**, **F_2-Cl_2-azo**, Figure 4.3).[80, 81] Similarly, recent work has demonstrated the superior switching behavior of di-*ortho*-fluoro and di-*ortho*-methoxy substituted azobenzenes (**F_2-$(OMe)_2$-azo**, Figure 4.3).[306] Also visible-light sensitive diazocines and the exceptionally bistable arylazopyrazole scaffold could be investigated for site-specific incorporation into target proteins via GCE. Both, the diazocine and also the arylazopyrazole scaffold have recently been employed as tools in amber suppression.[255, 307]

Although suitable aaRS/tRNA$_{CUA}$ pairs would need to be found or engineered for most of these novel photoswitchable scaffolds, it has been shown that the *Ma*PylRS/tRNA$_{CUA}$ pair holds great potential for the incorporation of such structurally diverse ncAAs.[308] Preparation as tripeptide KG-ncAA would further prevent solubility issues of these hydrophobic molecules during evolutionary approaches (Chapter 3.1.9).

An advantageous key feature of azo-ncAAs over other small molecule azobenzenes is that they can be used to efficiently and specifically generate homogeneous azobenzene-protein conjugates. However, they do not present a broadly applicable solution to study protein function under spatio-temporal control. Determined by protein size, structure, localization and function, it might prove beneficial to employ an azo-crosslinker, photosensitive proteins such as LOV domains or regulate protein activity via photoswitchable ligands. This could improve protein expression and folding, while also allowing for the application of a photoswitch featuring the best possible photophysical properties.

Depending on the addressed biological question, it might be sufficient to use a photoswitch featuring a PSS of 60%, thereby making the application of azo-ncAA pFN attractive. In other cases, a PSS of 60% might not be sufficient to gain the desired effect on protein function, especially if the PSS is further decreased upon incorporation into the protein. Overall, the effect of photoisomerization on protein conformation and thus function is difficult to predict, especially for large and complex proteins. Therefore, computational simulations can prove highly beneficial for the application of azobenzene-based tools.

With the successful development and characterization of the azo-ncAAs presented within this work, we are able to contribute to the range of photoresponsive tools that are currently available within the complex fields of chemical biology, optogenetics and photopharmacology. The insightful consideration discussed throughout this work will prove valuable to the scientific community working towards the optic modulation of protein function. Outstanding issues may be reasonably addressed in future work to improve and tune azo-ncAAs as photoswitchable tools in GCE.

Part B – Electrophilic Warheads for the Generation of Activity-Based Probes Using Sortylation

5 Introduction

5.1 The Ubiquitin System

The 76 aa protein ubiquitin (Ub) is highly conserved across all eukaryotic life from yeast to humans. The reversible, post-translational modification of proteins with Ub, also termed ubiquitylation, serves as key regulator in almost every cellular process. To regulate functions such as proteasomal degradation, endocytosis or DNA repair, the C-terminal glycine of Ub is attached via an isopeptide bond to the ε-amino group of a lysine residue within the substrate. Further, ubiquitylation has also been reported on protein N-termini or on thiol and hydroxyl moieties in cysteine, serine or threonine residues.[309, 310] Ub-dependent signaling typically not only involves the installation of a single Ub monomer but different topologies of ubiquitylation, including the attachment of multiple Ub monomers. Beyond that, Ub can also be ubiquitylated on seven lysine residues or the N-terminus, thereby forming complex polymeric Ub chains.[311] These can either be of a homotypic nature, comprising the same lysine linkage type, or of a heterotypic nature. The latter encompasses a variety of linkages, thereby leading to the formation of mixed or even branched chains (Figure 5.1).[312]

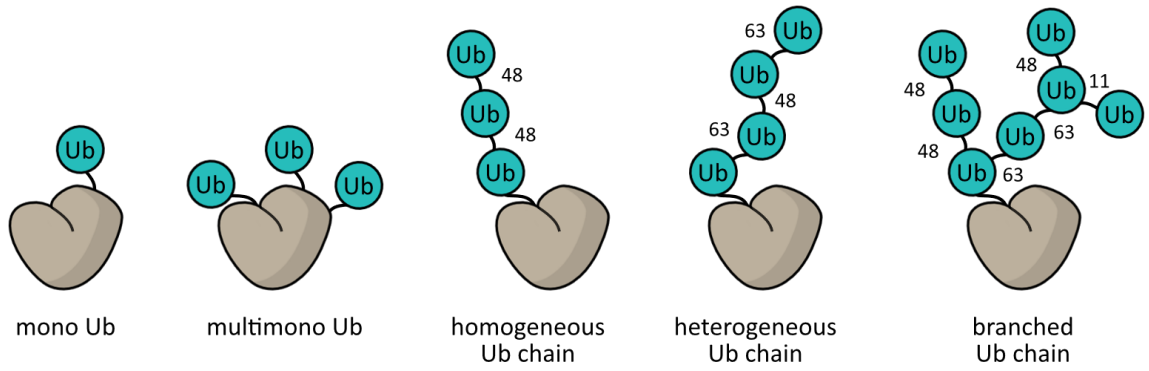


Figure 5.1 | Schematic depiction of various Ub topologies. Besides mono ubiquitylation of the substrate (grey), Ub (cyan) can also become a target for ubiquitylation itself, thereby forming Ub chains. Homogeneous chains consist of the same linkage throughout the chain, here shown for K48. Heterogeneous and branched Ub chains comprise a variety of linkages, here shown using the lysine residues K11, K48 and K63.

Three key classes of enzymes are essential in the highly dynamic process of installing, interpreting and erasing the various Ub topologies, termed writers, readers and erasers, respectively. The so-called writers of ubiquitylation are ultimately a multi-enzyme cascade, comprising E1, E2 and E3 enzymes. Ub is first activated by the ATP-driven E1 enzyme to then form a thioester acyl intermediate with the E2. An E3 ligase mediates the transfer of Ub from the E2 enzyme to the target substrate, conjugating Ub to a nucleophilic acceptor.[313]

Proteins related in structure to Ub, termed ubiquitin-like proteins (Ubls), get attached by a similar enzymatic cascade consisting of their own family of E1, E2 and E3 enzymes.[314] This results in additional diversification of Ub/Ubl topologies using Ubls such as the small ubiquitin-related modifier (SUMO). Both, Ubiquitylation and SUMOylation are highly dynamic processes due to the opposing activity of writer and eraser enzymes. Deconjugation of Ub and SUMO involves highly specific proteases, of which the vast majority are classified as cysteine-proteases. These include six Ub-specific protease families (USPs, UCHs, OTUs, MJDs, MINDYs and ZUP1) as well as all SUMO-specific protease families (SENPs, DESIs, USPL1). Nucleophilic attack of the active site cysteine of deubiquitylases (DUBs) or SUMO-specific proteases on the isopeptide bond leads to formation of a thioester acyl intermediate that is subsequently hydrolyzed, thus releasing the respective Ubl.[315, 316]

DUB and SENPs are important key players in regulating biological functions that are essential for cell viability and their dysregulation is heavily associated with tumor growth, autoimmune diseases and neurodegenerative disorders. Consequently, DUBs have become prime therapeutic targets, driving the development of probes and inhibitors for the elucidation of their function and substrate specificity.[317]

5.2 Activity-Based Probes for Ub Deconjugating Enzymes

The application of activity-based probes (ABPs) has proven particularly effective in the elucidation of DUB activity, due to the presence of a nucleophilic active site cysteine. In general, an ABP mimics the natural substrate of an enzyme without being processed by it. Instead, it is covalently attached to the enzyme, thereby trapping it. To this end, a recognition element is typically linked to an electrophilic warhead and reporter tag (Figure 5.2a).[318]

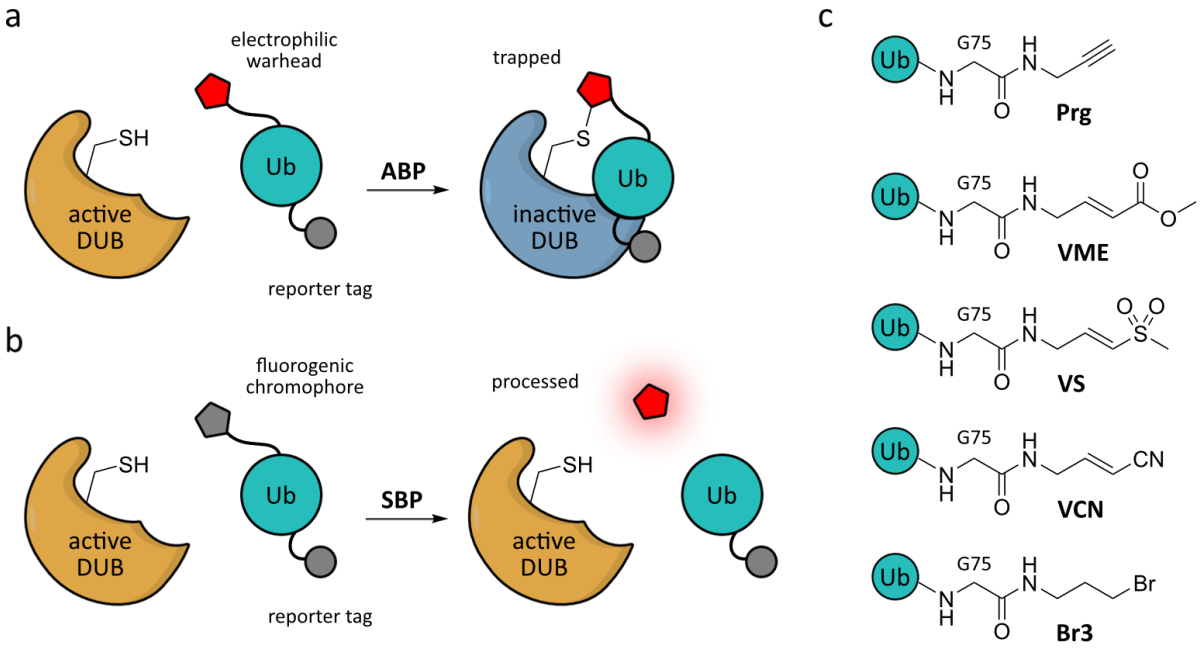


Figure 5.2 | Schematic depiction of activity-based probes (ABPs) and substrate-based probes (SBPs) targeting DUBs. (**a**) The warhead (red) connected to the recognition element Ub reacts with the active site cysteine of DUBs (orange), thus rendering the trapped enzyme inactive (blue). (**b**) Kinetics of DUBs (orange) are typically measured using SBPs that get processed, thereby releasing the fluorogenic chromophore (red). An additional reporter or affinity tag is therefore optional. (**c**) Structural formulas of the most prominent warheads used in ABPs targeting the conjugating and deconjugating enzymes of the Ub system. Here, Ub indicates Ub_{1-74} without the two C-terminal Gly residues.

The electrophilic warhead is the key component that differentiates an ABP from a substrate-based probe (SBP). A range of electrophilic warheads has been employed as ABPs to study Ub conjugation and deconjugation, particularly targeting active site cysteine residues. Amongst them, the most prominent ones are (i) propargyl warheads, reacting via direct (1,2) thiol-addition, (ii) vinyl sulfones and a vinyl methyl ester warhead for conjugate (1,4) addition reactions and (iii) alkyl halides, leading to nucleophilic displacement (Figure 5.2c).[318]

The recognition element most often encountered in ABPs targeting DUBs, is monomeric Ub, although a variety of diUb and triUb ABPs have been reported as well.[319] Together with an affinity or bioorthogonal handle as reporter tag, ABPs allow not only for the detection of DUB activity, but also for the discovery and identification of novel deconjugating enzymes within the ubiquitin system.

Substrate-based probes (SBPs) such as fluorogenic Ub-conjugates (Figure 5.2b) further complement ABPs in the elucidation of DUB activity and specificity.[316, 320]

A main drawback of DUB-activity profiling using ABPs is the limitation to *in vitro* applications. ABPs are predominantly generated using semi-synthetic or chemoenzymatic strategies. Subsequent delivery into cells is thus very much hindered by the size of the recognition element, which renders them cell-impermeable. Although strategies such as cell-penetrating peptides[321] have been developed to tackle these challenges, novel approaches which allow for the investigation of DUB activity dynamics under physiologically relevant conditions are highly sought after.[322]

5.3 Sortylation – Sortase-Mediated Ligation

An elegant solution to overcome permeability issues of ABPs would be their generation inside living cells using sortase-mediated ligation, also termed sortylation. This reaction exploits the natural enzymatic activity of the bacterial transpeptidase sortase A from *Staphylococcus aureus*. After recognition of a five amino acid motif LPXTG in the C-terminal region of its substrate, sortase A cleaves this motif downstream of the threonine residue by nucleophilic attack of its active site cysteine. The resulting thioester intermediate is subsequently attacked by the N-terminal amino group of a pentaglycine moiety found in peptidoglycans, resulting in an amide bond between acyl donor and acceptor (Figure 5.3).[323]

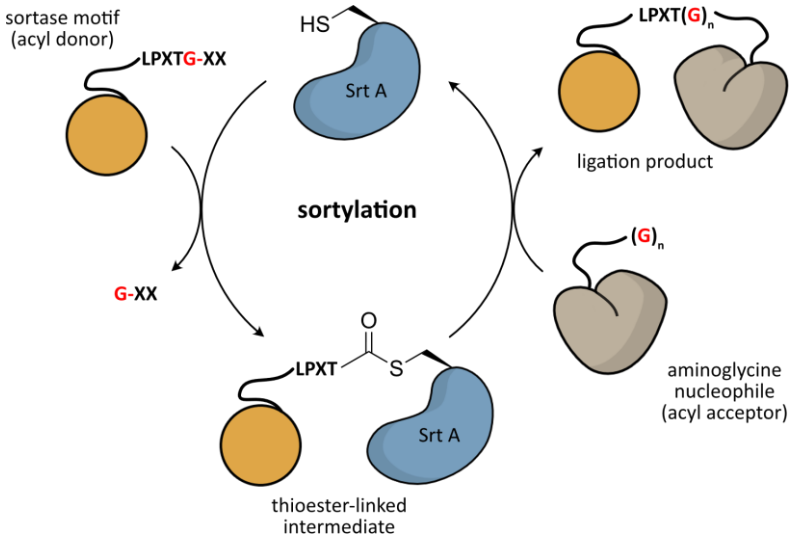


Figure 5.3 | Schematic depiction of sortase-mediated ligation. Sortase A (Srt A) attacks the amide-bond downstream of the threonine residue within the sortase motif, thereby forming an activated thioester intermediate and releasing G-XX. Nucleophilic attack by an aminoglycine nucleophile leads to a novel amide bond being formed between the acyl acceptor and acyl donor.

Sortase A has been successfully used in protein engineering, yielding variants with strongly enhanced activity such as sortase 2A (Srt 2A).[324] In addition, Srt 2A also targets a different recognition motif LAXTG, making it particularly interesting for sortylation approaches using C-terminally modified Ub- and SUMO-variants. Fottner et al. combined genetic code expansion with sortylation to site-specifically attach Ub or SUMO to target proteins in mammalian cells. Di-Ubs generated in this manner were further shown to be stable towards DUB cleavage.[325]

Similarly, sortylation by Srt 2A may be exploited for the generation of DUB-resistant ABPs bearing Ub recognition motifs, thereby potentially overcoming the current limitations of cell permeability.

6 Project Aim

6.1 Generation of ABPs Using Sortylation

Within this project we envisioned the generation of ABPs via sortylation for the elucidation of DUB and SENP activity in mammalian cells. Expression of Srt 2A and a respective Ubl bearing the C-terminal sortylation recognition motif together with cell-permeable electrophilic warheads would provide the necessary means for *in vivo* assembly of the ABP (Figure 6.1).

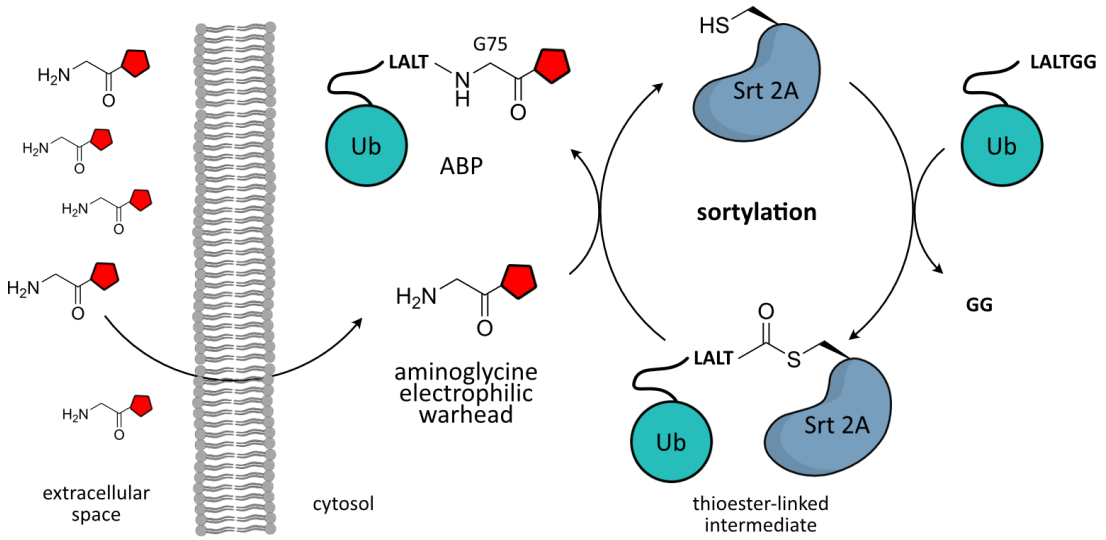


Figure 6.1 | Schematic depiction of generating ABPs in live cells using sortylation. Cell-permeable electrophilic warheads are taken up by the cell to then get conjugated to the C-terminus of Ub via sortylation. Sortase 2A (Srt 2A) attacks the amide-bond downstream of the threonine residue within the C-terminus of Ub bearing the mutated motif LALTG. Nucleophilic attack on the thioester intermediate by the aminoglycine moiety linked to the electrophilic warhead (red) leads to formation of the ABP in live cells.

6.2 Envisioned Design of Warheads

As part of this project we aimed at the design and synthesis of electrophilic warheads with varying reactivity to trap the active site cysteine of DUBs or SENPs either via Michael addition or nucleophilic displacement. Conjugation of the respective warheads to a Gly-moiety would allow for transpeptidation using sortase, thereby replacing Gly_{76} of Ub or SUMO with the electrophilic warhead (Figure 6.1). The evaluation and application of the obtained warheads was pursued by my colleague V. Wanka and will thus be discussed within the scope of her doctoral thesis.

For ligations via Michael addition, a range of α,β-unsaturated esters and amides was designed to reduce reactivity in comparison to the literature-known vinyl methyl ester warhead G-VME (Figure 6.2a, blue range).[318] Michael acceptors featuring a greater electrophilicity instead, such as α,β-unsaturated nitriles[318] or α,β-unsaturated ketones, were also included (Figure 6.2a, red range).

Similarly, we planned on synthesizing a series of α-halo ketone warheads for nucleophilic displacement, including an α-fluoroketone which was synthetically inaccessible during prior work (Figure 6.2a). In case enhanced reactions with nucleophiles such as GSH would be observed, a photolabile ketone protecting group could be installed as reported in literature for α-halo ketone moieties.[326, 327] Upon irradiation with UV-light the reactivity towards nucleophiles would be restored, thus enabling trapping of DUBs or SENPs after successful sortylation.

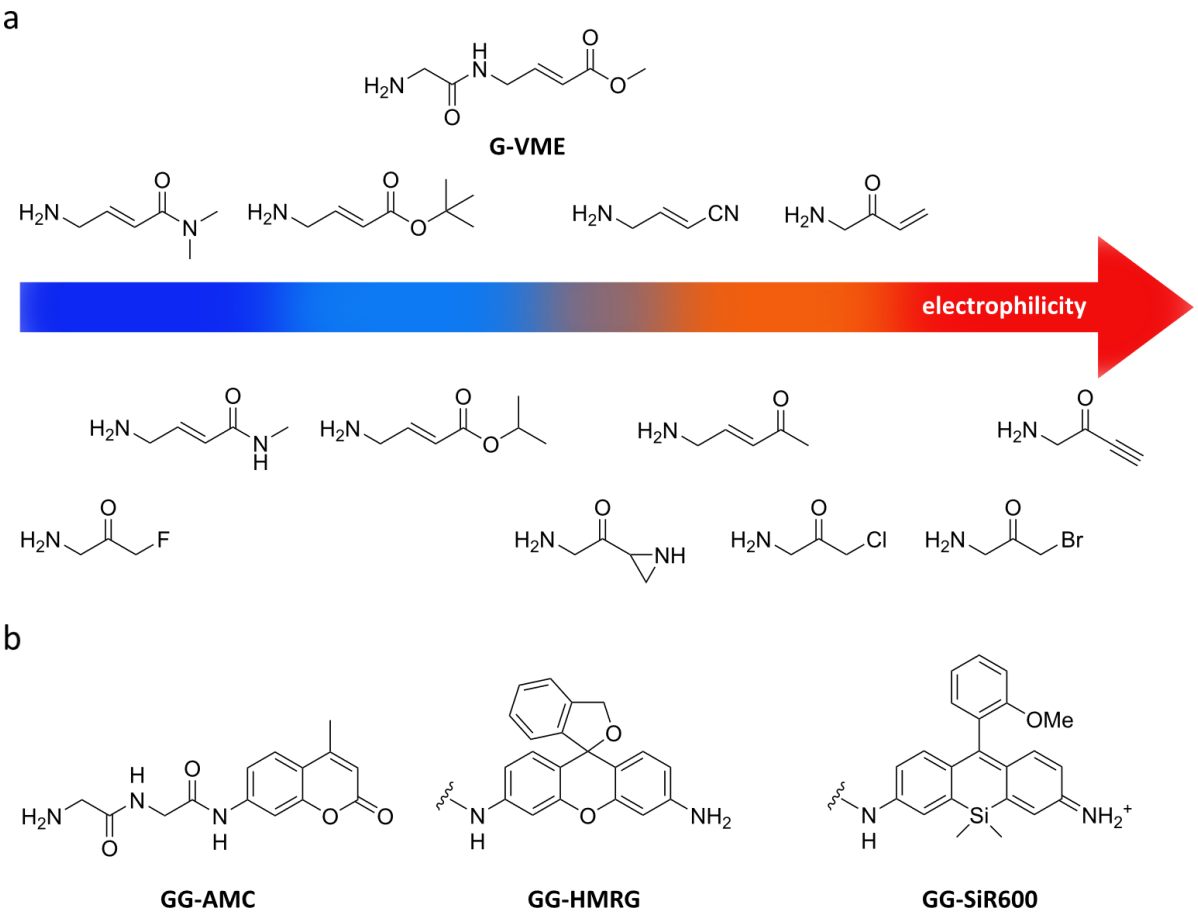


Figure 6.2 | Schematic overview of potential target electrophilic warheads and fluorogenic diglycine peptides in this project. (**a**) All depicted warheads will be conjugated to an additional Gly moiety, as shown for G-VME. The electrophilic small molecules are arranged according to their expected electrophilicity.[328, 329] Warheads which are expected to be more reactive than G-VME might need to be masked using a photolabile protecting group.[326] (**b**) Structural formulas of fluorogenic diglycine peptides to probe DUB and SENP activity.

Finally, we aimed to investigate DUB and SENP activity using fluorogenic SBPs. To this end, we planned to prepare one of the coumarin- or rhodamine-conjugated dipeptides GG-AMC, GG-HMRG and GG-SiR600[330], a red fluorescent scaffold exhibiting little to no background signal (Figure 6.2b). Whereas AMC has been successfully used to probe DUB[320] and SENP[316] activity *in vitro*, HMRG has been used to generate cancer-selective probes that allowed for the elucidation of γ-glutamyltranspeptidase activity *in vivo*.[331]

7 Electrophilic Warheads for the Generation of ABPs via Sortylation

7.1 Prior Work

Prior work on this project by M. Cigler has focused on the preparation and *in vitro* evaluation of electrophilic warheads known from literature (Figure 7.1).[318]

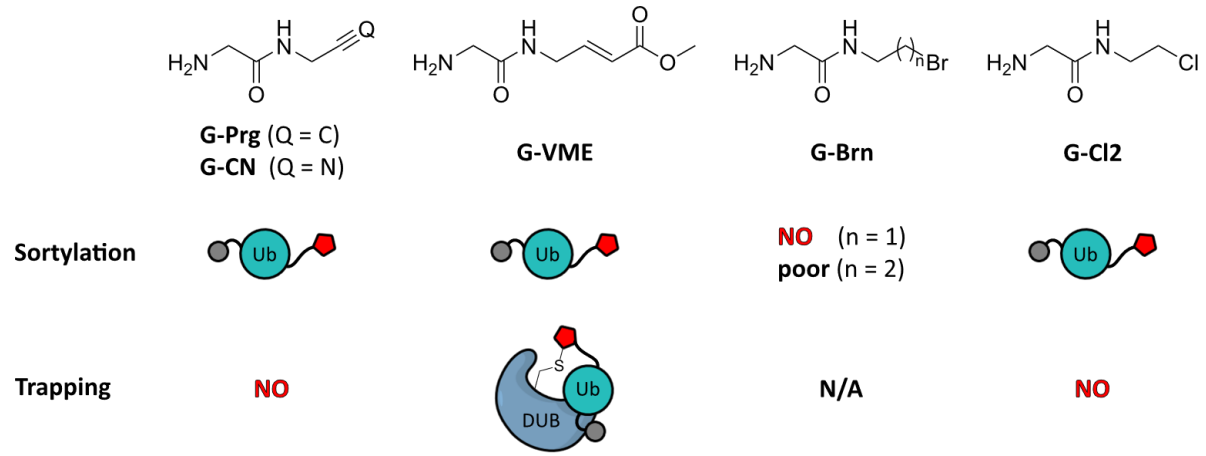


Figure 7.1 | Schematic overview of electrophilic warheads synthesized and evaluated in prior work. Structural formulas of warheads synthesized in prior work. While successful *in vitro* sortylation was observed for G-Prg, G-CN, G-VME and G-Cl2, *in vitro* trapping could only be observed using the vinyl methyl ester G-VME for a subset of DUBs.

Amongst them are the propargyl warhead G-Prg and nitrile warhead G-CN for (1,2) thiol-addition, the vinyl methyl ester warhead G-VME reacting via Michael addition and several alkyl halides (G-Br2, G-Br3, G-Cl2) for S_N2 reaction with the active site cysteine. While synthesis of the depicted electrophilic warheads was straight-forward, successful *in vitro* sortylation and *in vitro* DUB trapping could only be observed using the vinyl methyl ester G-VME (Figure 7.1).

Preliminary results have further shown the competitive reaction of Ub-LALTG-VME with abundant thiols such as GSH, making *in cellulo* applications of this particular warhead challenging. Therefore, we aimed at expanding the current scope of electrophiles capable of trapping the active site cysteine of DUBs and SENPs in live cells.

7.2 Design of Warheads and Probes From This Project

The target compounds synthesized in this project can be found in Figure 7.2. They include a series of moderately electrophilic (Figure 7.2a) and masked electrophilic warheads (Figure 7.2b), as well as coumarin-based fluorogenic probes (Figure 7.2c).

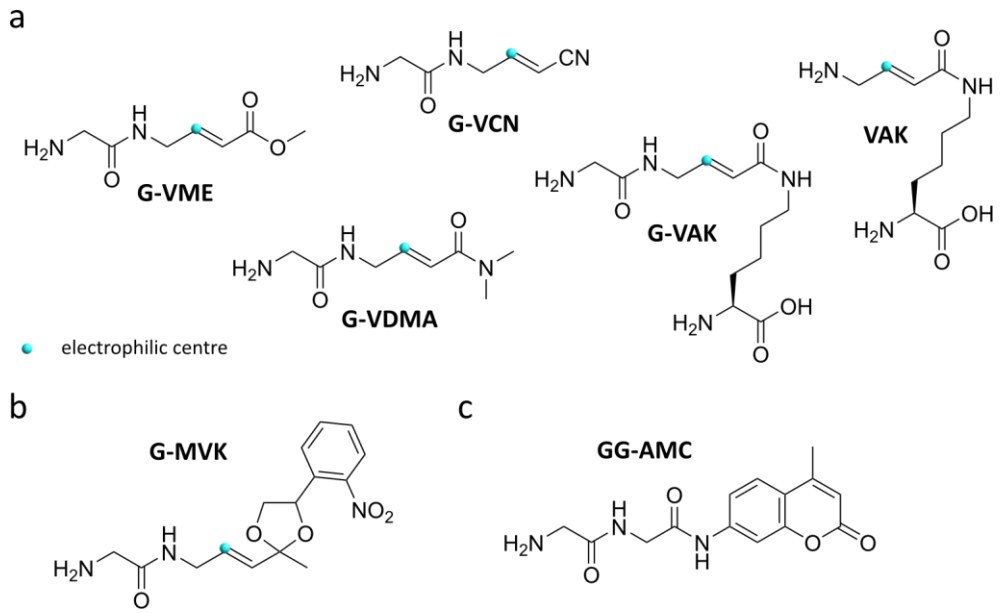
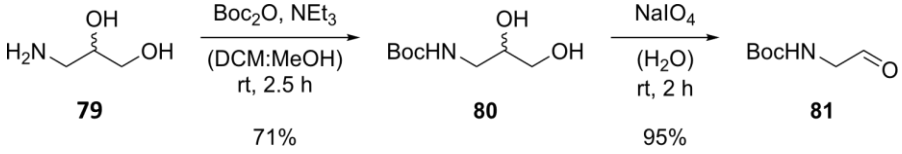


Figure 7.2 | Overview of compounds synthesized in this work. (**a**) Overview of moderately electrophilic warheads based on Michael acceptors with varying reactivity towards nucleophiles, including the least reactive derivative **G-VDMA**, lysine-based ncAAs **G-VAK** and **VAK**, the published reference compounds **G-VME** and **G-VCN**.[318] (**b**) Structural formulas of the photocaged electrophilic warhead analog **G-MVK**. (**c**) Structural formula of the published coumarin-based fluorogenic DUB-/SENP-activity probe **GG-AMC**.[320]

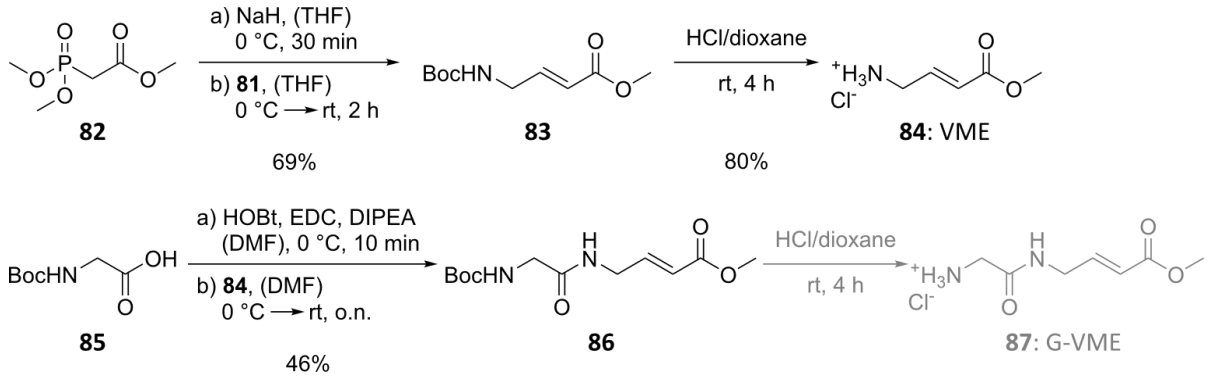
7.3 Synthesis Towards Moderately Electrophilic Warheads

In general, the *E*-isomer of all moderately electrophilic warheads was accessed by applying stereoselective olefination methods such as the Horner-Wadsworth-Emmons (HWE) or Wittig reaction. Standard *N*-Boc protection of commercially available 3-amino-1,2-propanediol (**79**) and subsequent oxidative diol cleavage employing $NaIO_4$ gave rise to the common precursor *N*-Boc glycinal **81** in fair yields (67% over two steps, Scheme 7.1).



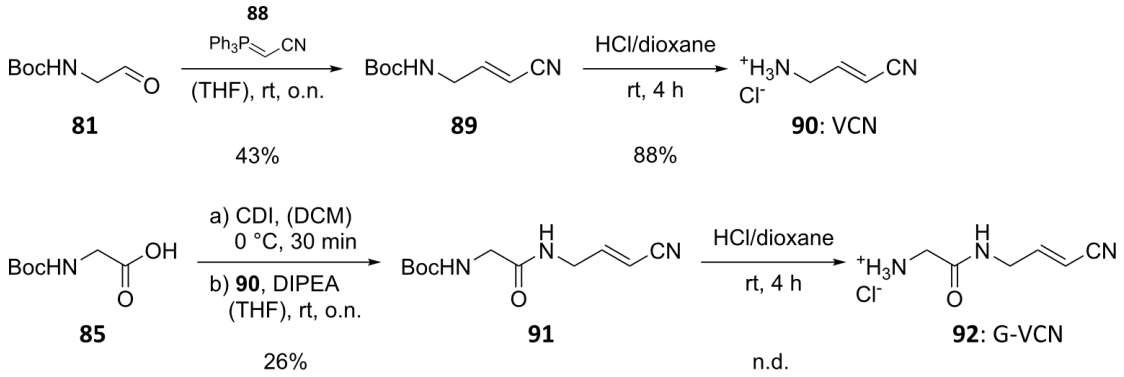
Scheme 7.1 | Synthesis of the common precursor *N*-Boc glycinal (81) used to synthetically access moderately electrophilic warheads.

The *N*-Boc protected vinyl methyl ester **83** was prepared from commercially available trimethyl phosphonoacetate (**82**) (Scheme 7.2).



Scheme 7.2 | Synthetic route towards the electrophilic warhead G-VME (87) known from literature.

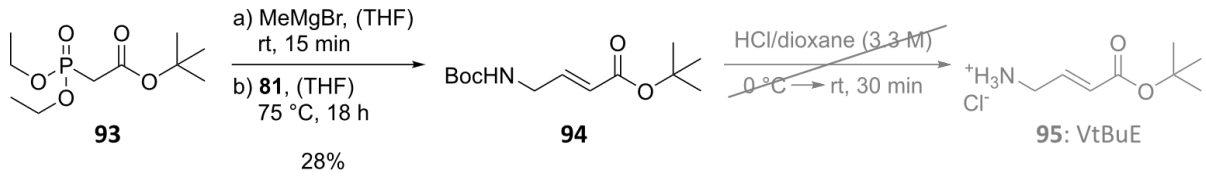
While the respective HWE reaction was successfully performed using KOtBu for *in situ* generation of the stabilized phosphonate anion, the initial yield of 60% was difficult to reproduce (29% and 40%). Although KOtBu is commercially available, it is highly moisture-sensitive and thus the amount of intact reagent can vary. Therefore, we investigated the formation of *N*-Boc protected vinyl methyl ester **83** using NaH and LiHMDS as base. Whereas NaH worked well for the preparation of **83**, LiHMDS was not effective in yielding the desired *E*-isomer. Instead, a *Z*:*E* ratio of approximately 1:1 was obtained, further diminishing the yield of the desired isomer with a maximum of 40%. In contrast, employment of NaH as base during the HWE reaction solely yielded the *E*-isomer. Finally, acidic deprotection gave the desired Michael acceptor **VME** (**84**) as HCl-salt in very good yields (80%). The *N*-Boc-protected target warhead **86** was obtained in decent yields (46%) by coupling *N*-Boc glycine (**85**) to **84**, utilizing the coupling reagents HOBt and EDC. Standard *N*-Boc deprotection with 4 M HCl in dioxane would then furnish the desired target compound **G-VME** (**87**) as HCl-salt (Scheme 7.2, depicted in grey). We then set out to prepare a more reactive Michael acceptor based on α,β-unsaturated nitriles (Scheme 7.3).



Scheme 7.3 | Synthetic route towards the highly reactive Michael acceptor G-VCN (92) known from literature.

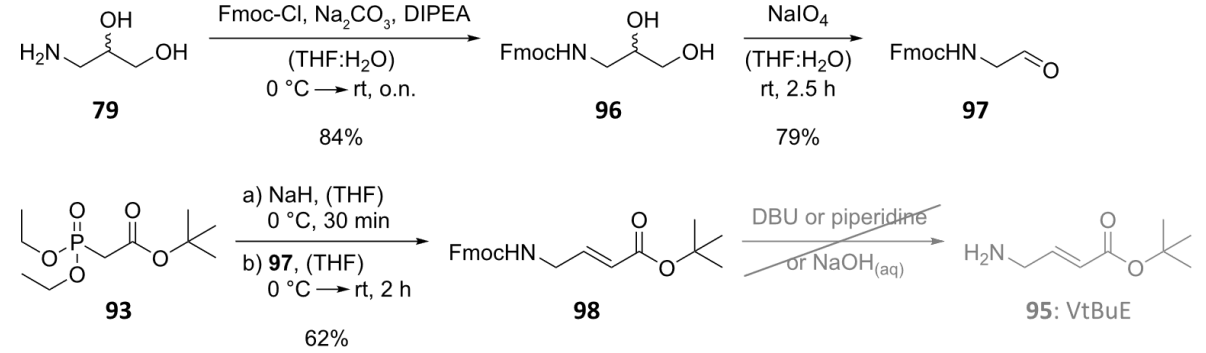
To this end, *N*-Boc glycinal **81** was reacted with the commercially available Wittig reagent (Triphenylphosphoranylidene)-acetonitrile (**88**), providing the desired product **89** in fair yields (43%). Due to the increased reactivity of α,β-unsaturated nitriles towards nucleophiles, acidic deprotection using 4 M HCl in dioxane resulted in a mixture of product **VCN** (**90**) and the respective chloro adduct (0.20 eq). While the *N*-Boc protected target compound **91** was obtained in acceptable yields (26%) applying the coupling reagent CDI, its subsequent deprotection employing 4 M HCl in dioxane did once more result in the formation of chloro adduct (0.55 eq). Although shorter reaction times might improve the ratio between intact **G-VCN** (**92**) and the respective chloro adduct, we concluded that application of α,β-unsaturated nitrile warheads for the generation of ABPs in live cells would be nearly impossible.

We instead focused on the synthesis of highly reactive but masked electrophilic warheads (Chapter 7.4) and moderately electrophilic warheads in comparison to **G-VME**. To this end, we investigated the synthesis of a vinyl-*tert*-butyl ester warhead (**95**) that should be less electrophilic than its vinyl methyl ester analog **84**.[329] Initial attempts at obtaining the target compound **94** by esterification of the corresponding α,β-unsaturated carboxylic acid **109** failed. Preparation via HWE reaction finally yielded the desired *E*-isomer **94** in high selectivity by employing MeMgBr as base (Scheme 7.4).[332]



Scheme 7.4 | Attempted synthetic route towards the vinyl-*tert*-butyl ester VtBuE (95) via *N*-Boc protection strategy.

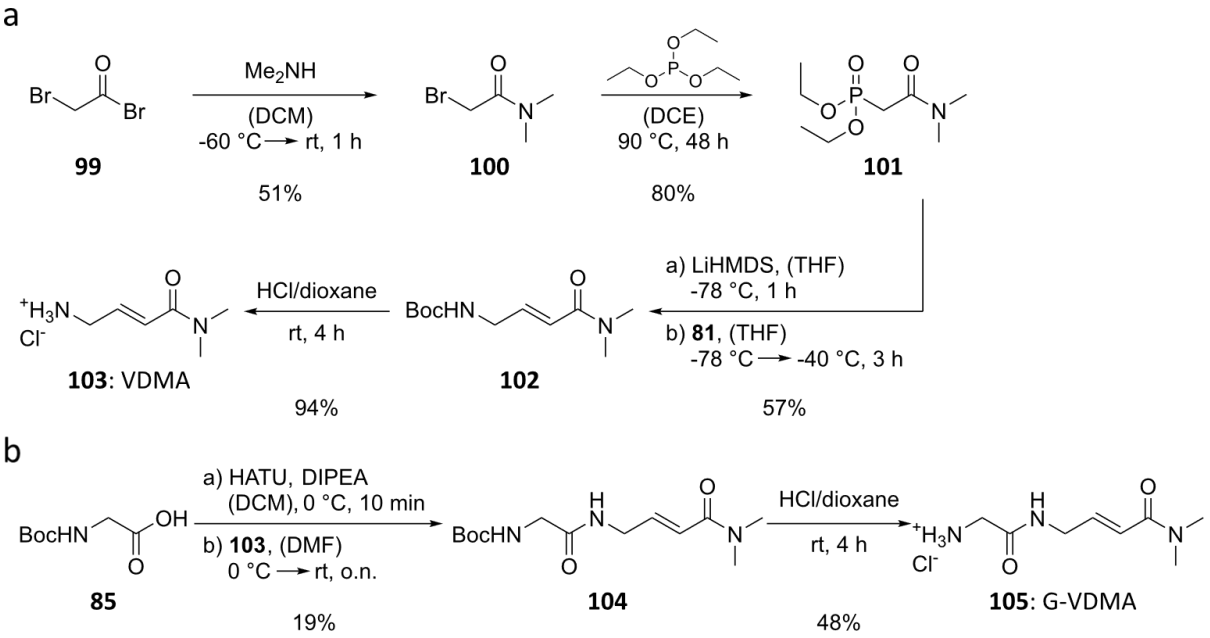
While literature suggested selective removal of the *N*-Boc protecting group in the presence of *tert*-butyl esters using dry HCl in EtOAc[333] or dioxane,[334] investigation of these reaction conditions resulted in unselective removal of the *N*-Boc protecting group and *tert*-butyl ester. Even running the reaction and subsequent evaporation at lower temperatures (0 °C and 25 °C, respectively) did result in hydrolysis of the *tert*-butyl ester in addition to removal of the *N*-Boc protecting group. Consequently, we instead synthesized *N*-Fmoc protected glycinal **97** to obtain the desired *N*-Fmoc protected vinyl-*tert*-butyl ester **98** via HWE reaction using NaH in decent yields (62%) (Scheme 7.5).



Scheme 7.5 | Attempted synthetic route towards the vinyl-*tert*-butyl ester VtBuE (95) via *N*-Fmoc protection strategy.

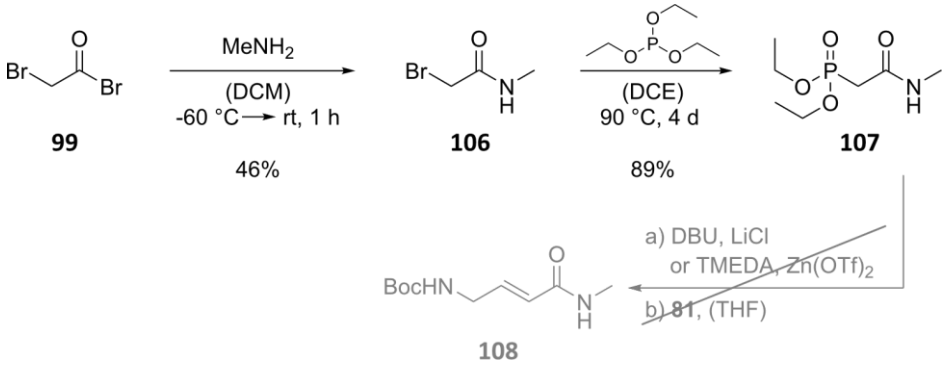
While MeMgBr and LiHMDS were also investigated for the preparation of **98**, both routes were discarded due to poor yields (17%) or high amounts of *Z*-isomer (0.27 eq) being formed, respectively. Subsequent deprotection using amine bases such as piperidine or DBU resulted in complex, inseparable product mixtures, presumably due to Michael addition of nucleophilic amine bases. An alternative deprotection route using aqueous NaOH in THF led to complete hydrolysis of the *tert*-butyl ester. Although it might be worth exploring additional *N*-Boc deprotecting strategies using oxalyl chloride[335] or install a different protecting group such as *N*-phthaloyl, we decided to focus on electrophilic warheads that would be synthetically easier to access.

Amongst those warheads are the α,β-unsaturated amides **G-VDMA** and **(G-)VAK**. While both are less reactive towards nucleophiles than their ester counterparts, the *N*,*N*-dimethyl vinyl amide **G-VDMA** is the least electrophilic warhead presented within this work.[328, 329] The latter was similarly to **G-VME** (**87**) accessed via the HWE reaction. The required olefination reagent **101** was synthesized via the Michaelis-Arbuzov reaction (Scheme 7.6a). Bromoacetyl bromide (**99**) was first treated with *N*,*N*-dimethylamine and the resulting alkyl halide **100** reacted with triethyl phosphite, giving the desired alkyl phosphonate **101** in fair yields (41% over two steps). The α,β-unsaturated amide **102** was obtained in decent yields (57%) via the HWE reaction, applying LiHMDS to generate the stabilized phosphorous ylide of **101**. Subsequent acidic deprotection yielded the Michael acceptor **VDMA** (**103**) as HCl-salt (Scheme 7.6a). *N*-Boc glycine (**85**) was then coupled to **103** using HATU, thus obtaining the *N*-Boc protected electrophilic warhead **104** in poor, yet acceptable yields (19%). Standard *N*-Boc deprotection using 4 M HCl in dioxane finally gave the desired target compound **G-VDMA** (**105**) as HCl-salt (Scheme 7.6b).



Scheme 7.6 | Synthetic route of the least reactive Michael acceptor G-VDMA (105).

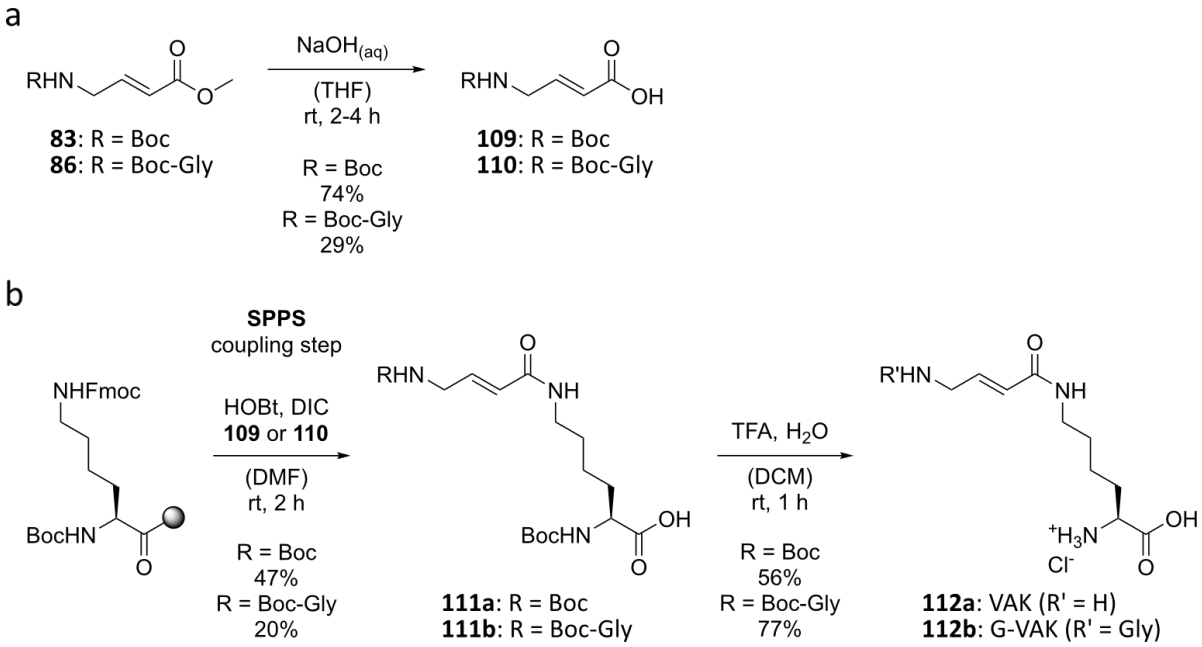
Preparation of a more reactive *N*-methyl vinyl amide warhead was attempted analog to the synthesis of **G-VDMA** (**105**). While the alkyl phosphonate **107** was straightforwardly acquired in decent yields over two steps (41%), subsequent HWE reaction with *N*-Boc glycinal **81** required screening of appropriate reaction conditions (Scheme 7.7).



Scheme 7.7 | Attempted synthetic route towards the *N*-methyl vinyl amide warhead 108 screening metal-promoted HWE reaction conditions.

Unfortunately, only little product (18%) was obtained using strong bases such as NaH or LiHMDS. In addition, only the *Z*-isomer of the desired product could be isolated and not the desired *E*-isomer (**108**). Therefore, we explored the use of mild tertiary amine bases in the presence of lithium halides[336] or Lewis acids[337] such as $Zn(OTf)_2$. Metal-promoted HWE reactions enhance the acidity of the methylene proton by coordination to the acyl and phosphonate group of the HWE reagent. Both methodologies are reported to be highly efficient and *E*-selective in the preparation of α,β-unsaturated amides. Unfortunately, neither the application of DBU in the presence of LiCl, nor TMEDA in the presence of $Zn(OTf)_2$ gave the desired result (Scheme 7.7).

We then synthesized the α,β-unsaturated carboxylic acid derivatives **109** and **110** by aqueous hydrolysis of **83** and **86**, respectively (Scheme 7.8a).



Scheme 7.8 | Synthesis of lysine-based ncAAs (G-)VAK (112a and b) bearing α,β-unsaturated amide handles via SPPS.

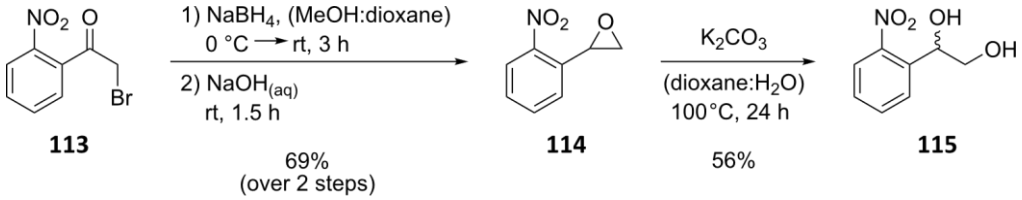
Subsequent attempts at coupling **109** to methylamine utilizing HOBt and EDC, failed. Due to these repeated issues in the synthesis towards an *N*-methyl vinyl amide warhead, we decided to focus on the preparation of lysine-based α,β-unsaturated amides **(G-)VAK**. Not only did we anticipate facile preparation of **(G-)VAK** via SPPS, but it would also allow to explore substrate- or linkage-specific DUB-trapping using both derivatives as tools in a genetic code expansion and sortylation approach. To this end, we coupled both α,β-unsaturated carboxylic acid derivatives **109** and **110** to lysine via standard Fmoc-SPPS, employing DIC and HOBt (Scheme 7.8b).

The corresponding *N*-Boc protected ncAAs **VAK** (**111a**) and **G-VAK** (**111b**) were obtained in acceptable yields of 47% and 20%, respectively. Subsequent acidic deprotection using TFA in DCM gave the target electrophilic ncAAs **VAK** (**112a**) and **G-VAK** (**112b**) as double TFA-salt (Scheme 7.8b). A scale-up of the developed synthesis route and application of **(G-)VAK** as tools in GCE was further pursued by P. Ruckgaber during his master's thesis.[338]

7.4 Synthesis Towards Masked Electrophilic Warheads

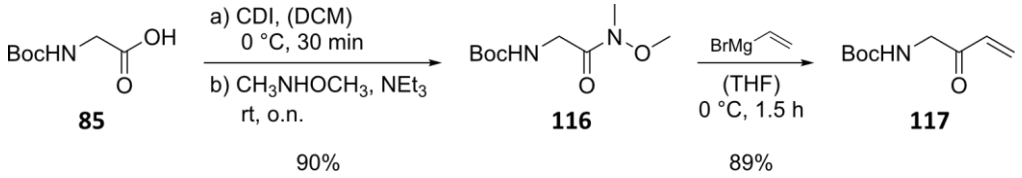
Besides the synthesis of moderately electrophilic warheads for the generation of ABPs via sortylation, we envisioned the development of highly reactive, but masked warheads. This way, the warhead is stable towards nucleophiles encountered in a physiological environment or during sortylation, but still quantitatively trapping the active-site cysteine of DUBs/SENPs upon "unmasking". An elegant way to accomplish such a scenario is the application of photocages, which allow for orthogonal uncaging by irradiation with UV-light. We intended to install a photocage at the carbonyl-moiety of α,β-unsaturated ketones or α-halo ketones, yielding the respective acetal which is inert towards nucleophiles.

An adequate photocage found in literature[326, 327, 339] was synthesized over three steps, starting from commercially available 2-bromo-2'-nitroacetophenone (Scheme 7.9). One-pot reduction with $NaBH_4$, followed by intramolecular S_N2 reaction under aqueous basic conditions gave the respective epoxide **114** in good yields (69%). The desired photocage **115** was thus obtained by refluxing **114** under aqueous basic conditions using K_2CO_3, leading to epoxide ring-opening.



Scheme 7.9 | Synthesis of the photolabile protecting group 115 which was used to mask highly electrophilic warheads.

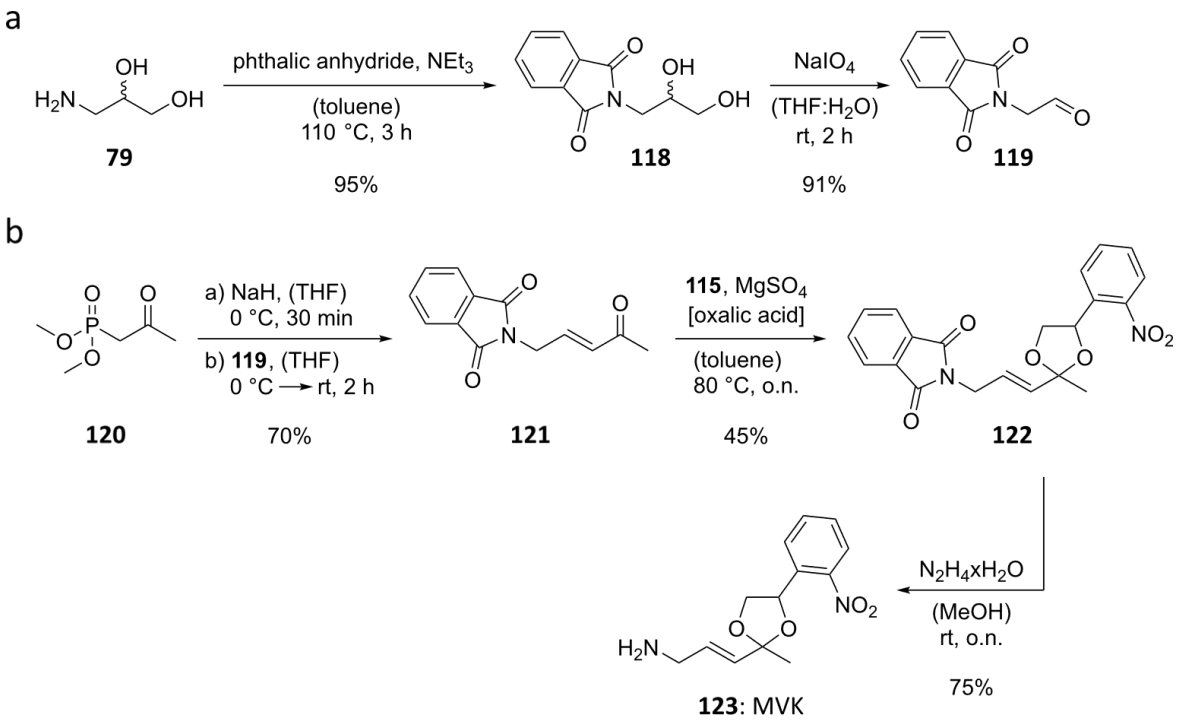
Initially, we aimed at preparing a highly reactive but masked, terminal, α,β-unsaturated ketone via a Grignard reaction (Scheme 7.10).



Scheme 7.10 | Synthesis of a highly reactive, but labile α,β-unsaturated ketone (117). The *N*-Boc protected warhead **117** was labile during acidic deprotection and at moderately elevated temperatures (>30 °C).

First, we reacted commercially available *N*-Boc glycine (**85**) with CDI, followed by *N*,*O*-dimethylhydroxylamine to obtain the Weinreb amide **116** in excellent yields (90%). Subsequent Grignard reactions with a commercially available solution of vinylmagnesium bromide however, yielded considerable amounts of side product. This side product could be identified as the Michael addition product of **117** with the *in situ* generated nucleophile *N*,*O*-dimethylhydroxylamine. Although reaction conditions could be optimized to generate and isolate the desired product **117** in very good yields (89%), the α,β-unsaturated ketone was not stable during acidic deprotection or at elevated temperatures needed for quantitative acetal formation (Scheme 7.10). Consequently, we explored a less electrophilic α,β-unsaturated ketone derivative and α-halo ketones for the generation of photocaged warheads.

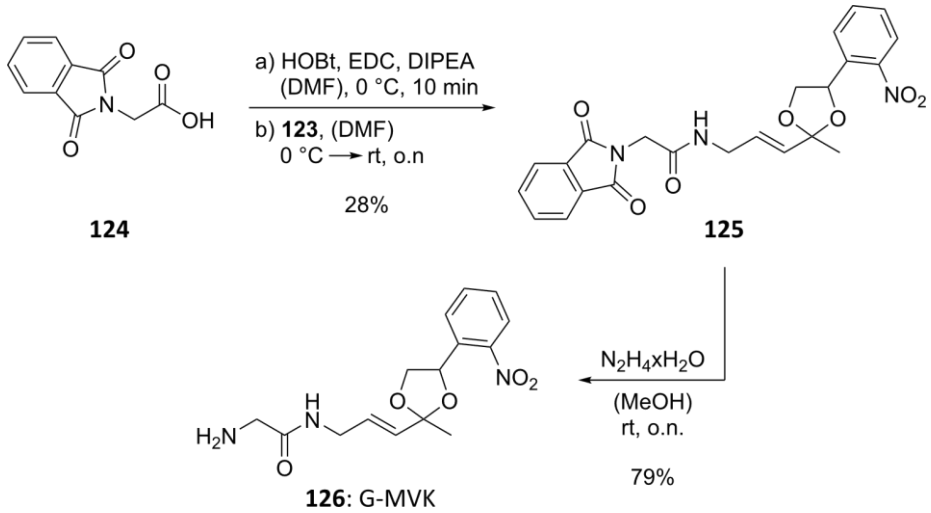
To obtain the less reactive target methyl vinyl ketone warhead **G-MVK** (**126**), we first synthesized *N*-phthaloyl protected glycinal **119** (Scheme 7.11a).



Scheme 7.11 | Synthesis route towards the highly electrophilic Michael acceptor MVK (123), masked by a photolabile protecting group.

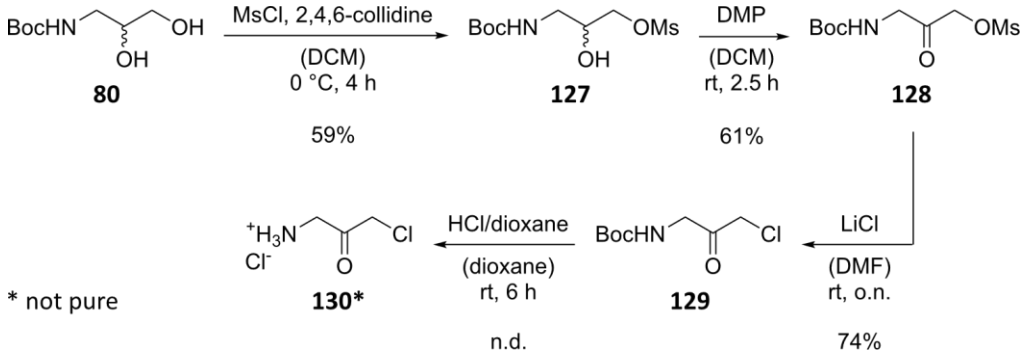
In contrast to previously used *N*-Boc protection, *N*-phthaloyl protection would allow for orthogonal installation of the photocage **115** as well as selective deprotection of the amine functionality in presence of an acetal moiety. Analogously to *N*-Boc and *N*-Fmoc glycinal (Chapter 7.3), **119** was obtained in very good yields over two steps (86%). HWE reaction of **119** with commercially available dimethyl-2-oxo-propylphosphonate (**120**) was promoted by the addition of NaH, thus giving the methyl vinyl ketone **121** in good yields (70%, Scheme 7.11b). Initial attempts at acetal formation with photocage **115** were performed as reported, applying catalytic amounts of PPTS (pKa 5.21) and $MgSO_4$ as dehydrating agent.[326, 327] Since no product formation could be observed, we tested catalysis of acetal formation using the stronger acid *p*-TsOH (pKa 1.00).[339, 340] Although the acetal product **122** was successfully isolated, the respective yields remained low (19%). In addition, we observed considerable amounts of homologation to the inactive β,γ-unsaturated acetal. Based on reports found in literature, we anticipated that the use of milder acids such as oxalic acid (pKa 1.23) or phthalic acid (pKa 2.89) might prevent homologation to the β,γ-unsaturated analog while still leading to acetal formation in good yields.[341] Whereas no product could be observed by employing phthalic acid, we were indeed able to generate the desired α,β-unsaturated acetal **122** in decent yields (45%) when catalyzed by oxalic acid (Scheme 7.11b). Subsequent deprotection using hydrazine hydrate gave the target warhead **MVK** (**123**) in good yields (75%, Scheme 7.11b).

The *N*-phthaloyl protected acetal **G-MVK** (**125**) was obtained in acceptable yields (28%) by coupling of *N*-phthaloyl glycine (**124**) to **123**, employing the standard coupling reagents HOBt and EDC. Deprotection with hydrazine hydrate then gave the desired target compound **G-MVK** (**126**) as photocaged, electrophilic warhead (Scheme 7.12).



Scheme 7.12 | Synthesis of the highly electrophilic target warhead G-MVK (126), masked by a photolabile protecting group.

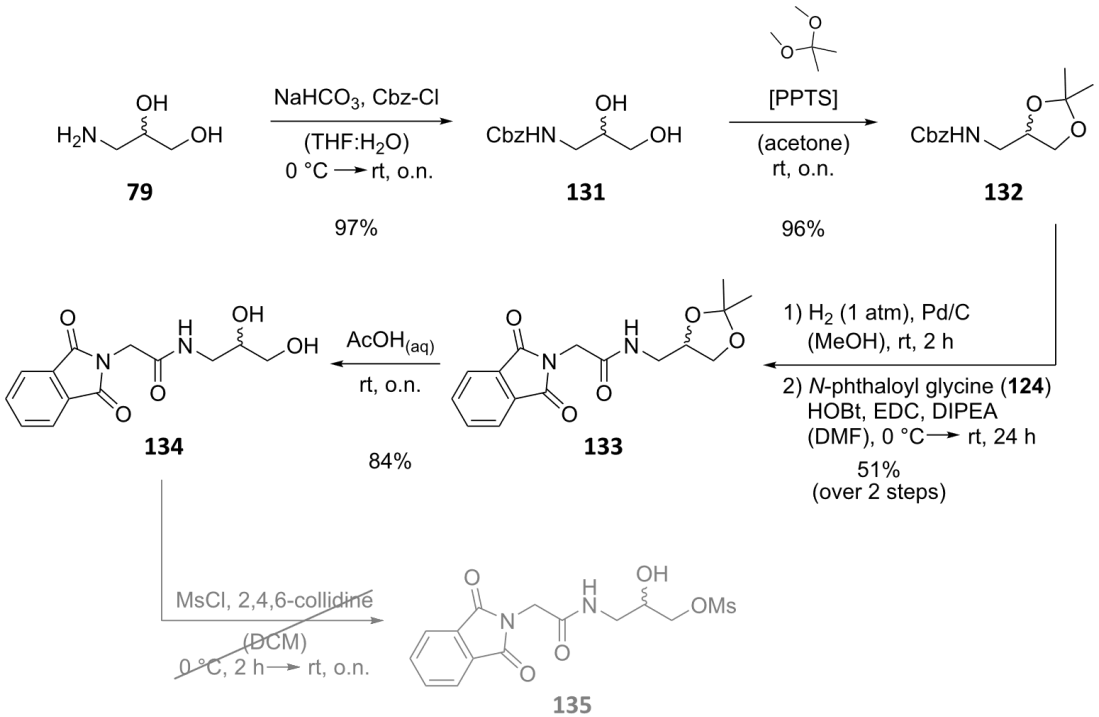
Early attempts towards a masked α-chloroketone warhead aimed at the preparation of α-chloroketone **130** (Scheme 7.13), followed by amide-coupling to *N*-phthaloyl glycine and acetal formation. To this end, we explored selective mesylation of the primary over the secondary alcohol in 3-amino-1,2-propandiol (Scheme 7.13). As reported in literature for vicinal diols, the desired product **127** was prepared in fair yields (59%) by reacting **80** with mesyl chloride and 2,4,6-collidine under kinetic control.[342] The secondary alcohol was then oxidized with DMP to yield the α-mesylated ketone **128** which was subsequently reacted with LiCl to give the *N*-Boc protected α-chloro ketone **129** in acceptable yields over two steps (45%). Acidic deprotection using 4 M HCl in dioxane yielded the HCl-salt **130** as an impure product mixture (Scheme 7.13).



Scheme 7.13 | | Failed synthesis route towards warheads based on an α-chloroketone. The *N*-Boc protected warhead **129** was labile during acidic deprotection and was thus obtained as an impure mixture (annotated as *).

Although the obtained impurities could neither be identified via LR-MS nor by NMR, an amide coupling to *N*-phthaloyl glycine was still attempted. To our dismay, an intractable reaction mixture was obtained from which the desired product could not be isolated. We concluded that a different synthesis route would have to be explored due to the α-chloro ketone's reactivity towards nucleophiles during *N*-Boc deprotection and subsequent amide coupling.

Similar to the successful synthesis of **G-MVK** (**126**), we aimed at installing the photocage prior to amide bond formation. Initial attempts at synthesizing the *N*-phthaloyl protected analog of **129** were however quickly discarded, since mesylation was substantially unselective and subsequent oxidation was low-yielding (9%). An alternative synthesis route was based on the preparation of *N*-Cbz protected 3-amino-1,2-propandiol (**131**) to then protect the vicinal diol as acetal in excellent yields (93% over two steps). Hydrogenation of **132** gave the volatile, crude amine which was used without further purification in an amide coupling reaction with *N*-phthaloyl glycine (**124**), yielding **133** (Scheme 7.14).



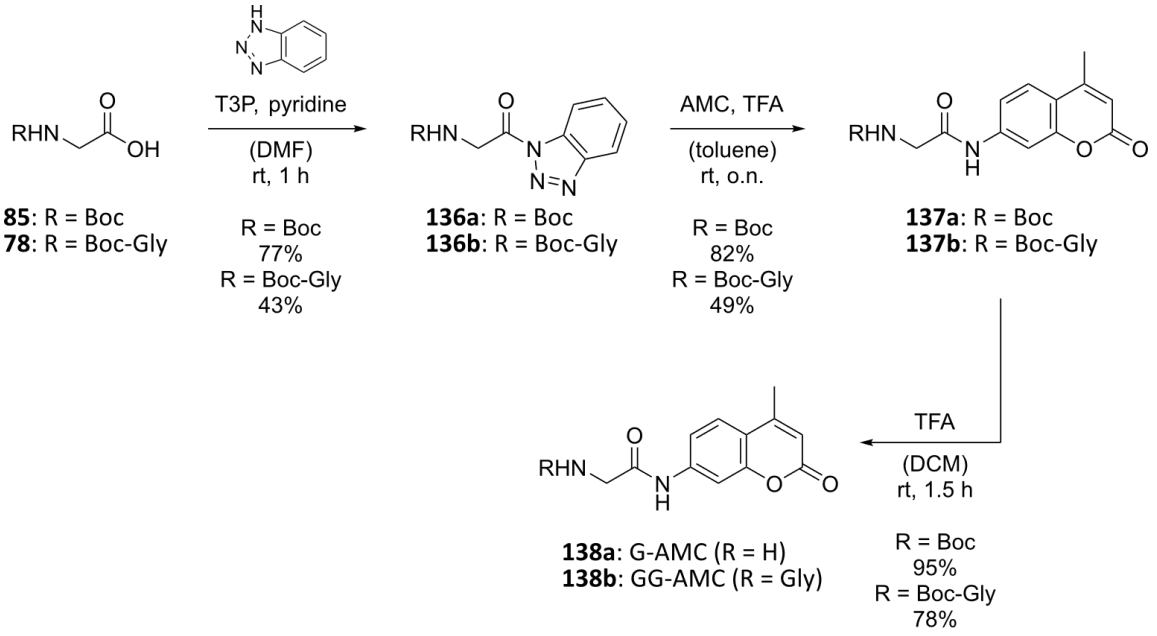
Scheme 7.14 | Failed synthesis route towards a warhead based on an α-chloro ketone, masked by a photolabile protecting group. *N*-phthaloyl protected vicinal diol **134** was unselectively mesylated.

Previous attempts at directly coupling 3-amino-1,2-propandiol (**79**) to *N*-phthaloyl glycine (**124**) failed repeatedly due to its insolubility in organic solvents such as DCM, THF or DMF. Contrary to these earlier attempts, compound **133** could be obtained in decent yields over two steps (51%) by masking the polarity of the vicinal diol as acetal (Scheme 7.14). Subsequent deprotection in aqueous AcOH gave the desired target molecule **134** in very good yields (84%). Unfortunately, selective mesylation of the primary alcohol failed again, this time due to solubility issues of **134** in the organic solvent DCM. Consequently, different solvents or alternative synthesis routes will have to be explored (Chapter 8).

7.5 Synthesis of Fluorogenic Probes

Beyond the development of novel electrophilic warheads for the generation of ABPs, we synthesized fluorogenic probes based on 7-amino-4-methylcoumarin (AMC) for the generation of substrate-based probes. AMC is a typical chromophore used for probing peptidase activity. Herein, we wanted to probe DUB and SENP activity by applying the published, fluorogenic target molecules **(G)G-AMC**.[320]

Whereas amide couplings presented within this work have successfully been performed utilizing standard coupling reagents such as HOBt/EDC or HATU, amide bond formation with the weakly nucleophilic, aromatic amine AMC required a different procedure. A reported reaction of *N*-acyl-benzotriazoles with aromatic amines such as AMC, promoted by equimolar amounts of TFA, seemed particularly promising for the preparation of **(G)G-AMC**.[343] To this end, we first prepared the *N*-acyl-benzotriazoles **136a** and **136b** in decent yields (77% and 43%, respectively, Scheme 7.15).



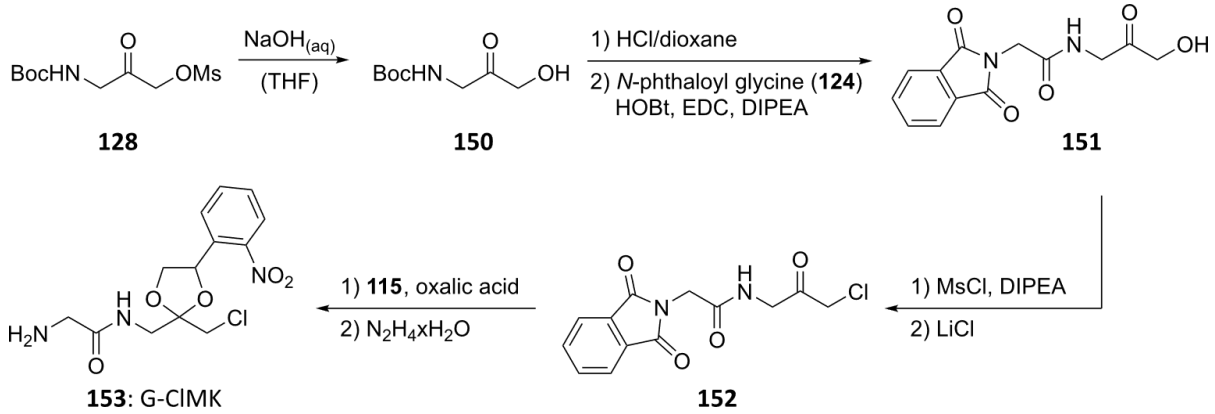
Scheme 7.15 | | Synthesis of the fluorogenic targets (G)G-AMC (138 a and b) to probe DUB and SENP activity.

Subsequent reaction with AMC under slightly acidic conditions gave the desired *N*-Boc protected compounds **137a** and **137b** in very good to fair yields (82% and 49%, respectively). The corresponding fluorogenic target compounds **G-AMC** (**138a**) and **GG-AMC** (**138b**) were obtained as TFA salts by acidic *N*-Boc deprotection.

8 Outlook

In summary, we synthesized a range of electrophilic warheads featuring moderate to high reactivity towards nucleophiles such as the active site cysteine of DUBs and SENPs (Chapter 7.2). Amongst them are the electrophilic warheads **G-VME** and **G-VCN** known from literature, as well as novel, less reactive Michael acceptors based on α,β-unsaturated amides (**G-VDMA** and **(G-)VAK**). The lysine-based α,β-unsaturated amides **(G-)VAK** may further serve as valuable tools in probing DUB and SENP activity via a genetic code expansion approach, as explored by P. Ruckgaber in his master's thesis. Beyond that, we established a reliable and selective synthesis route towards highly reactive α,β-unsaturated ketones masked by a photolabile protecting group, such as **G-MVK**.

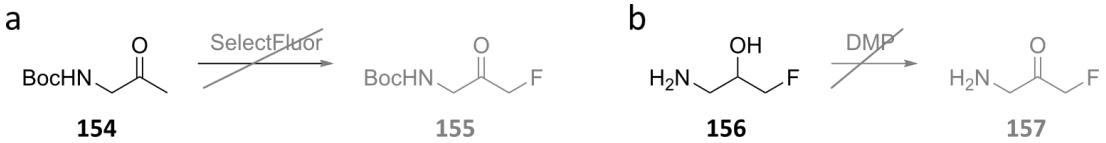
The adaption of this synthesis route for the preparation of masked α-halo ketone **G-ClMK** however, failed (Scheme 7.14). A potential alternative is the preparation of the *N*-Boc protected key intermediate α-hydroxyketone **150** by hydrolysis of the mesylated alcohol **128** (Scheme 8.1).



Scheme 8.1 | Proposed synthesis route to access the α-chloro ketone warhead G-ClMK (153) via intermediate 151.

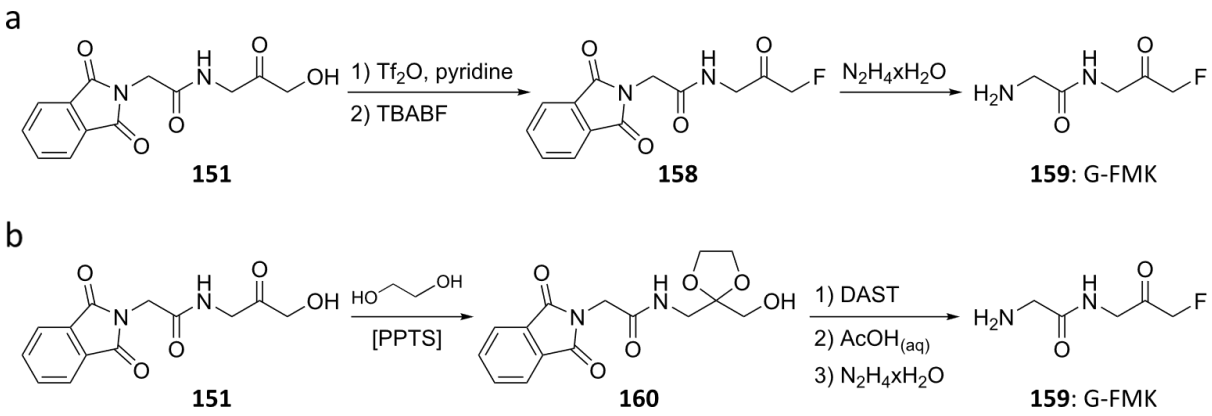
This way, selective mesylation and oxidation can be exerted as demonstrated for the high-yielding synthesis of *N*-Boc protected α-chloro ketone **129**. Subsequent deprotection and coupling to *N*-phthaloyl glycine (**124**) enables chlorination and protection of the ketone moiety as acetal with the photolabile protecting group **115**.

Intermediate **151** could further be used to prepare the moderately reactive α-fluoro ketone warhead **G-FMK**. Prior work focused on the preparation of G-FMK with the commercially available fluorinating reagent SelectFluor® on methyl ketone **154** (Scheme 8.2a) or oxidation of 1-amino-3-fluoropropan-2-ol (**156**) employing DMP (Scheme 8.2b).



Scheme 8.2 | Failed attempts at preparing a moderately electrophilic warhead based on an α-fluoro ketone moiety.

Since both attempts failed, it might be worth investigating the synthesis of G-FMK by first activating the α-hydroxy moiety of **80** to then react the activated alcohol with TBABF (Scheme 8.3a).[344] Beyond that, the fluorinating reagent DAST has proven efficient in transforming alcohols into fluoroalkanes (Chapter 3.1.4). To this end, the carbonyl moiety should be protected first to avoid transformation of the ketone into geminal difluoro alkanes (Scheme 8.3b).[345]



Scheme 8.3 | Proposed synthesis route to access the α-fluoroketone warhead G-FMK (159) via intermediate 151.

Although aziridines have been reported to preferentially react with carboxylates over thiols, investigation of this warhead as part of irreversible covalent inhibitors has resulted in selective labeling of a K-Ras cysteine mutant G12C over the respective aspartic acid mutant G12D. Its diminished electrophilicity in comparison to α,β-unsaturated ketones or α-chloro ketones, makes the aziridine **G-MAzK** a potentially interesting warhead for targeting the active site cysteine of DUBs and SENPs (Figure 8.1a).[346]

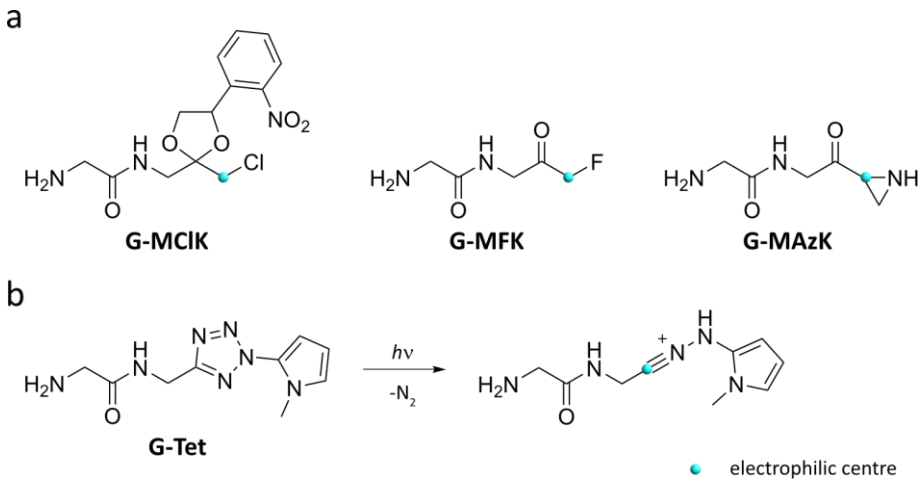


Figure 8.1 | Overview of proposed warheads for the generation of ABPs targeting DUBs and SENPs.

Lastly, a tetrazole-based substrate for light-triggered reaction with proximal nucleophiles might be explored as an alternative to photocaged, α,β-unsaturated ketones or α-chloro ketones. Not only should the warhead G-Tet feature enhanced aqueous solubility (Figure 8.1b), but it has further been successfully employed to probe DUB activity in live cells by conjugation to a cell-penetrating peptide.[321]

The recommendations given herein are by far not covering the whole range of potential warheads applicable to the generation of ABPs targeting DUBs and SENPs. However, they depict the most promising candidates, complementing the electrophilic warheads that have been synthesized so far within this project. Depending on the results obtained from evaluating and applying these warheads to DUB and SENP profiling, it is recommended to revisit the design of additional warheads, respectively.

Part C – Supplementary Information

9 Experimental Procedures

9.1 Chemistry – General Remarks

Automated Flash Column Chromatography

Preparative purification was performed on a Büchi *Pure C-850 Flash* chromatography instrument with UV detection at 254 and 280 nm. Commercially available Büchi *FlashPure* Silica cartridges with the indicated size were used as stationary phase.

Chemicals and Solvents

All commercially available chemicals were obtained from the companies abcr, Acros Organics, Alfa Aeasar, Carbolution, Fluorochem, Honeywell Fluka, Roth, Sigma-Aldrich, TCI Europe or Thermo Fisher Scientific and used without further purification. Dry solvents (water content < 50 ppm) were obtained from Acros Organics or Sigma-Aldrich and stored under argon atmosphere over molecular sieves. Other solvents were either obtained from the companies Honeywell Fluka, Roth, Sigma-Aldrich and Thermo Fisher Scientific or as technical grade solvents from TU Munich and ETH Zurich. Technical grade solvents were distilled once before employed in column chromatography.

Flash Column Chromatography

Preparative purification was performed via flash column chromatography under pressure with silica gel in high-purity grade (pore size 60 Å, 230 – 400 mesh and 40-63 µm particle size) provided by Supelco as stationary phase.

HPLC

Preparative HPCL was either performed on a *LC-20* system (Shimadzu) equipped with a MWD or on a *1290 Infinity II LC* system (Agilent) equipped with a DAD. The respective HPLC system was coupled to a *Luna* C18(2) (5 µm, 100 Å, 250×10 mm) column or a *Luna* C18(2) (10 µm, 100 Å, 250×21.2 mm), both provided by Phenomenex. Mobile phase A (0.1% formic acid or TFA in Milli-Q water) and mobile phase B (0.1% formic acid or TFA in MeCN) were applied for gradient elution. The eluent was typically removed by lyophilization on an *Alpha 2-4 LSCbasic* (Christ) equipped with a rotary vane vacuum pump (Pfeiffer).

HR-MS

High resolution mass spectra were recorded in the ESI mode with either an *LTQ-FT Ultra*^{TM} (Thermo Fisher Scientific) coupled to an *UltiMate 3000* HPLC system (Thermo Fisher Scientific) provided by the group of Prof. Sieber at the TU Munich or on a Bruker *Daltonics maXis*^{TM} ESI-QTOF spectrometer provided by Molecular and Biomolecular Analysis Service (MoBiAS) at ETH Zurich.

IUPAC Designations

IUPAC nomenclature of compounds are given according to the designations by *ChemDraw Professional 20.0.0.41.*

LC-MS

For low resolution mass spectra or reaction control purposes, samples were analyzed by a *6130 Quadrupole* mass spectrometer (Agilent Technologies) with a MWD detector, coupled to a *1260 Infinity* HPLC-system (Agilent Technologies) with EM detector. The HPLC system was either equipped with a *Luna* C18(2)-HST (2.5 µm, 100 Å, 100×2.0 mm) column or a *Luna Omega PS* C18 (3 µm, 100 Å, 100×2.1 mm) column, both provided by Phenomenex.

Mobile phase A (0.1% formic acid in *Optima*^{TM} water) and mobile phase B (0.1% formic acid in *Optima*^{TM} MeCN) were applied for gradient elution (15 to 95 % B in 3.5 min or 3.0 min, respectively) applying a flow rate of 0.36 mL/min or 0.55 mL/min, respectively. Samples were analyzed by UV absorbance at 193, 254 and 280 nm together with both the positive and negative ESI-mode.

NMR

^{13}C-NMR, ^{1}H-NMR and ^{19}F-NMR spectra as well as 2D spectra (COSY, HSQC) were measured at room temperature (300 K) on a Bruker *AVHD* 300 MHz, a Bruker *AVHD* 400 MHz and a Bruker *AVHD* 500 MHz provided by the TU Munich or on a Bruker *Avance III* 400 MHz provided by the ETH Zurich. ^{13}C- and ^{19}F-NMR experiments were performed with ^{1}H-decoupling. The following abbreviations were used for the multiplicity of signals: s (singlet), d (duplet), t (triplet), q (quartet), p (pentet), m (multiplet), br (broad) and their combinations. Chemical shifts of ^{1}H-NMR were calibrated to the residual solvent signal (δ = 2.50 ppm (p) for DMSO-d_6; δ = 3.31 ppm (p) for CD_3OD; δ = 7.26 ppm (s) for $CDCl_3$). Chemical shifts of ^{13}C-NMR were calibrated to the ^{13}C-D multiplets of residual solvent signal (δ = 39.52 ppm (m) for DMSO-d_6; δ = 49.00 ppm (m) for CD_3OD; δ = 77.16 ppm (t) for $CDCl_3$). Coupling constants in ^{1}H-NMR spectra refer to ^{1}H,^{1}H-couplings, if not specified otherwise.

The software *MestReNova 14.2.1* from Mestrelab Research S.L.K was used for the analysis of NMR spectra. The assigned number of carbon atoms per ^{13}C signal were marked with an asterisk (*), if signals could not be assigned clearly.

Rotary Evaporator

A rotary evaporator *R-300* with a heating bath *B-300* Base from Büchi was used for removing solvents in vacuo and for distilling technical grade solvents.

Thin-Layer Chromatography

TLC was performed using normal-phase silica TLC plates (silica gel 60G F_{254} on aluminium by Merck) for qualitatively controlling reaction progress and checking fractions.

Aromatic and fluorescent compounds were typically made visible by irradiation of the TLC plates with UV light (λ_1 = 254 nm, λ_2 = 366 nm).

Oxidizable compounds were visualized by dipping the TLC plate in a ceric ammonium molybdate (CAM) solution (12 g ammonium molybdate, 0.50 g ceric ammonium molybdate, 15 mL sulfuric acid, in 235 mL water) or potassium permanganate solution (1.5 g potassium permanganate, 10 g potassium carbonate, 1.25 mL 10 % (w/v) aqueous NaOH, in 200 mL water) followed by heating.

Amino acids were also visualized by dipping the TLC plate in a ninhydrin solution (1.5 g ninhydrin in 100 mL of EtOH and 3 mL AcOH) followed by heating.

No R_f-value was given for very polar compounds.

Working Methods

Unless otherwise specified, all reactions were carried out under standard laboratory conditions. All air-sensitive reactions were performed under *Schlenk* conditions under an inert argon atmosphere.

9.2 UV-Vis Spectrophotometry for Photophysical Characterization

General

Absorption spectra were acquired on an Agilent Cary 3500 UV-Vis. For photoisomerization measurements, Starna quartz cuvettes (type 16R/160/Z15, 10 mm pathway) taking 200 µL volume to top of the optical window were used. This way, the vertical pathlength of the isomerization light was less than 30 mm from the light source to the bottom of the cuvette. Photoisomerization in the cuvette and relaxation rate measurements were both performed at 25 °C. Illuminations at different wavelengths prior or during the UV-Vis scans were exerted using 1 W "Star" LEDs (H2A1 series spanning 365 – 590 nm by Roithner Lasertechnik).

Solvent of Choice

To better mimic physiological conditions at pH 7.4, the azo-ncAAs were per default dissolved in a PBS:MeCN mixture (3:1) containing 5% DMSO, giving a final concentration of 45 µM.

Kinetic Measurements and Relaxation Half-Life $t_{1/2}$

For kinetic measurements over 2 hours, a concentration of 10 µM was chosen to prevent precipitation effects. In addition, the amount of co-solvent was kept at 1% DMSO to avoid altered relaxation rates. One data point was taken only every ten minutes to avoid alteration of relaxation rates by too frequent absorption measurements.

The acquired data points A_{kin} (t_n) were transformed with respect to the measured absorption spectra, using A_{trans} (λ_{kin}) and A_{cis} (λ_{kin}) to obtain the fraction of relaxed azo-ncAA F_{relax} (t_n) each ten minutes (n = 0,1,…,12) (equation 1).

Equation 1

$$F_{relax}(t_n) = \frac{\text{change in relaxation}}{\max \text{relaxed azo} - \text{ncAA}} \times 100 = \frac{A_{kin}(t_n) - A_{kin}(t_0)}{A_{kin}(t_0) \times \frac{A_{trans}(\lambda_{kin})}{A_{cis}(\lambda_{kin})}} \times 100$$

The unimolecular *cis*→*trans* thermal relaxation process obeys first-order rate laws as depicted in equation 2.

Equation 2

$$A(t) = A_0 \times e^{-kt}$$

The associated reaction rate constant k can be calculated by taking the difference in the amount of relaxed azo-ncAA between $t_{n=0}$ = 0 h and $t_{n=12}$ = 2 h (equation 3).

Equation 3

$$k[h^{-1}] = \ln\frac{A(t)}{A_0} \times t^{-1} = \ln\left(\frac{(100 - F_{relax}(t_{12}))}{(100 - F_{relax}(t_0))}\right) \times \frac{1}{2}$$

The lower bound for the relaxation half-life $t_{1/2}$ was then calculated according to equation 4.

Equation 4

$$t_{1/2}[h] = \frac{ln2}{k} \text{ with } A\left(t_{1/2}\right) = \frac{1}{2} \times A_0$$

9.3 Synthetic Standard Procedures

Standard Procedure A: Mills Reaction

To the respective aniline derivative (1.0 eq) dissolved in DCM (typically 3 mL/mmol) was added an aqueous solution of Oxone® (2.0 eq, typically 3 mL/mmol). The biphasic system was then stirred vigorously for 4-6 h, checking the reaction progress by TLC. After full conversion, the green organic layer was separated, washed with 1 M aqueous HCl, water, brine and then dried over Na_2SO_4. After filtration, the crude mixture was concentrated under reduced pressure (not below 100 mbar) yielding a green oil which was typically used without any further purification.

The crude nitroso compound was dissolved in glacial AcOH or in DCM (approx. 3 mL/mmol) containing 10% (v/v) AcOH. The second aniline derivative (1.0 eq) was added to the solution and the reaction mixture was stirred o.n. at rt.

The solvent was removed under reduced pressure and the crude resuspended in an aqueous solution of 10% (w/v) KH_2PO_4 (pH 3.5–4). The aqueous mixture was extracted with EtOAc (3×) and the combined organic layers washed with 10% (w/v) KH_2PO_4 (1×), brine (1×), dried over Na_2SO_4 and filtered. The crude solution was concentrated, dry-loaded onto silica and purified by flash column chromatography.

Standard Procedure B: Diazonium Coupling

To the respective aniline derivative (1.0 eq) were added 0.7 M HCl in MeOH:H_2O (3:7, 11.0 eq) and, if necessary, a small amount of MeOH until dissolution, while cooling the mixture in an ice bath. Then, an aqueous 2 M solution of $NaNO_2$ (1.4 eq) was added dropwise and the mixture was left stirring for 40 min at 0 °C.

The phenol (1.1 eq) was dissolved in 8.0 eq of cold aqueous 1 M NaOH with the minimum volume of MeOH or dioxane needed for completely solubilizing the respective compound. The cooled diazonium solution was subsequently transferred dropwise into the well-stirred cold phenolate solution. Throughout the addition, the pH was maintained at 8<pH<10 and adjusted with 1 M NaOH if necessary. The mixture was left stirring in the cold until complete consumption of the starting material (typically 1-2 h).

Methanol was removed under reduced pressure and the crude aqueous mixture was acidified to pH 3-4 using aqueous 10% (w/v) KH_2PO_4 and aqueous 1 M HCl and finally extracted with EtOAc (3×). The combined organic layers were then washed with aqueous 10% (w/v) KH_2PO_4 (1×), brine (1×), dried over Na_2SO_4 and filtered. The crude was concentrated, dry-loaded onto silica and purified by flash column chromatography.

Standard Procedure C: Esterification of a Carboxyl Group

To the carboxylic acid (1.0 eq) were added MeOH (typically 10-15 mL/mmol until complete dissolution), EDC×HCl (1.2 eq) and HOBt (1.2 eq). The reaction mixture was then stirred o.n. at rt and the reaction progress was checked by TLC.

Methanol was removed under reduced pressure and the crude residues extracted with EtOAc (3×) against water. The combined organic layers were washed with water (1×), brine (1×), dried over Na_2SO_4, filtered and concentrated. The crude products were purified by flash column chromatography.

Standard Procedure D: Alkylation of a Phenolic Hydroxyl Group

To the phenol (1.0 eq) were added acetone (typically 10-15 mL/mmol until complete dissolution), solid K_2CO_3 (6.0 eq) and the alkylating agent (4.0-6.0 eq, plus 1.5 eq of KI if the alkylating agent was not an alkyl iodide).

The resulting heterogeneous mixture was left stirring in an oil bath o.n. at 50 °C. The reaction progress was monitored by TLC and if necessary, more alkylating agent was added until complete alkylation.

The organic solvent was then removed under reduced pressure and the crude resuspended in a mixture of water and EtOAc. The organic phase was separated and the aqueous layer extracted with EtOAc (3×). The combined organic layers were washed with water (1×), brine (1×), dried over Na_2SO_4, filtered and concentrated. The crude products were purified by flash column chromatography.

Standard Procedure E: Halogenation of an Alcohol

To the alcohol (1.0 eq) were added DCM (typically 10-15 mL/mmol until complete dissolution) and the solution was cooled to 0 °C. To this cooled solution NEt_3 (1.5-3.0 eq) and MsCl or MsO_2 (1.5-2.0 eq) were added and the reaction subsequently stirred at rt. The reaction was monitored by LC-MS. After complete mesylation (typically under 2 h), the mixture was diluted with water and extracted with DCM (2×). The combined organic layers were then washed with water (1×), brine (1×), dried over Na_2SO_4 and filtered.

The crude was concentrated and either dissolved in DMF (typically 5 mL/mmol) or THF (approx. 10-15 mL/mmol) and LiCl or LiBr (8.0 eq) were added for halogenation, respectively. The reaction was left stirring o.n. either at r.t. (for chlorination), or heating in an oil bath at 50 °C (bromination) until full disappearance of the mesylate (checked by LCMS). The crude mixture was then diluted with water, extracted with EtOAc (3×) and the combined organic layers were washed with water (1×), brine (1×), dried over Na_2SO_4, filtered and concentrated. The crude product was purified by flash column chromatography.

Standard Procedure F: Hydrolysis of a Methyl Ester

To the methyl ester (1.0 eq), THF (until complete dissolution) and aqueous 2 M NaOH or 1 M LiOH (2-4 eq) were added. The resulting suspension was stirred at rt and reaction progress was checked by LC-MS. After complete consumption of the starting material (typically under 3 h), the reaction mixture was acidified with aqueous 10% (w/v) KH_2PO_4 until pH 3-4, extracted with EtOAc (3×) and the combined organic layers were washed with water (1×), brine (1×), dried over Na_2SO_4, filtered and concentrated. The crude product was purified by flash column chromatography.

Standard Procedure G: Deprotection of the *N*-Boc-Group Using TFA in DCM

The protected carbamate (1.0 eq) was dissolved in DCM (typically 10-20 mL/mmol) and TFA (typically 5-10 mL/mmol) was added in excess. The reaction mixture was stirred at rt and the reaction progress was checked by TLC. After complete deprotection (typically under 4 h), TFA and DCM were removed under reduced pressure. The crude residue was then redissolved in minimal amounts of MeOH and precipitated in cold Et_2O (max. 2 mL MeOH in 100 mL Et_2O). Centrifugation (4000×g, 4 °C) for 10 min and disposal of the supernatant gave the crude product as TFA salt, which was typically used without further purification.

Standard Procedure H: Deprotection of the *N*-Boc-Group Using 4 M HCl in dioxane

The protected carbamate (1.0 eq) was dissolved in dry dioxane (typically 1.5 mL/mmol) and an equal volume of 4 M HCl in dioxane (6 eq.) was added. The reaction mixture was stirred at rt and the reaction progress was checked by TLC. After complete deprotection (typically under 4 h), HCl and dioxane were removed under reduced pressure. The crude residue was then redissolved in minimal amounts of MeOH and precipitated in cold Et_2O (max. 2 mL MeOH in 100 mL Et_2O). Centrifugation (4000×g, 4 °C) for 10 min and disposal of the supernatant gave the crude product as HCl salt, which was typically used without further purification.

Standard Procedure I: Acylation of Amines With Carboxylic Acid Halides

To the amine (1.0 eq) was added an aprotic, polar organic solvent until complete dissolution and the solution was cooled to 0 °C. To this cooled solution a tertiary amine base (typically NEt_3 or DIPEA, 3.0 eq) and the carboxylic acid halide (1.1-2.2 eq) were added and the reaction subsequently stirred at rt. The reaction was monitored by LCMS and TLC. After complete acylation (typically 2-5 h), the mixture was quenched with an aqueous solution of saturated $NaHCO_3$ (4 mL/mmol) and MeOH was removed under reduced pressure. The crude aqueous mixture was acidified to pH 3-4 using aqueous 10% (w/v) KH_2PO_4 and finally extracted with EtOAc (3×). The combined organic layers were then washed with aqueous 10% (w/v) KH_2PO_4 (1×), brine (1×), dried over Na_2SO_4 and filtered. The crude was concentrated, dry-loaded onto silica and purified by flash column chromatography.

Standard Procedure K: Horner-Wadsworth-Emmons Reaction Using NaH as Base

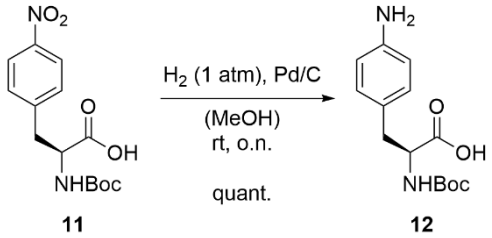
To a heat-dried *Schlenk*-flask under an inert argon atmosphere was added NaH (1.2 eq) as a 60% dispersion in mineral oil and suspended in dry THF (typically 4.5 mL/mmol) and the solution was cooled to 0 °C using an ice bath. To this cooled solution the olefination reagent (1.2 eq) was added dropwise and the reaction stirred at 0 °C for 30 min, resulting in a white slurry. Next, the aldehyde (1.0 eq) was dissolved in dry THF (typically 4.0 mL/mmol) and added dropwise to the cooled reaction slurry, resulting in a faintly yellow suspension which was stirred at 0°C for 30 min before warming to ambient temperature. The reaction mixture was stirred at rt until complete conversion was observed via TLC (typically 1-2 h), meanwhile turning into a light brown suspension. The mixture was diluted with water and THF evaporated under reduced pressure. The aqueous mixture was subsequently extracted with Et_2O (3×) and the combined organic layers washed with a saturated solution of aqueous $NaHCO_3$ (1×), aqueous 10% (w/v) $KHSO_4$ (1×) and brine (1×). The crude product was dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica and subsequently purified by flash column chromatography.

Standard Procedure L: Amide Coupling Promoted by EDC and HOBt

The *N*-Boc protected amino acid (1.0 eq) and HOBt (1.1 eq) were dissolved in dry DMF (typically 2-3 mL/mmol) and the solution was cooled to 0 °C using an ice bath. To this cooled solution, EDC×HCl (1.1 eq) and DIPEA (2.1 eq) were added and the reaction stirred at 0 °C for 10 min. After addition of the second amino acid (1.0 eq), typically protected as ester, dissolved in dry DMF (typically 1-2 mL/mmol), the cooled reaction was allowed to warm to ambient. The reaction mixture was stirred at rt until complete conversion was observed via TLC (typically 3-4 h, but oftentimes stirred o.n. without complications). The mixture was diluted with 10 % (w/v) LiCl and extracted with EtOAc (3×).The combined organic layers were then washed with a saturated solution of aqueous $NaHCO_3$ (1×), aqueous 10% (w/v) $KHSO_4$ (1×) and brine (1×). The crude product was dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica and subsequently purified by flash column chromatography.

9.4 Synthetic Procedures of azo-ncAAs

9.4.1 Synthesis of Precursor Boc-Phe(4-NH_2)-OH (12)



11 **12**

3-(4-aminophenyl)-2-((*tert*-butoxycarbonyl)amino)propanoic acid (12):

To Pd/C (loading 10 wt%, 1.13 g) in MeOH (500 mL) was added glacial AcOH (5 mL) and **11** (20.0 g, 64.5 mmol). The atmosphere was evacuated and then replaced with H_2 five times over the course of 3 h. The reaction mixture was filtered through a sand/cotton wool plug and the filtrate concentrated to give the product Boc-Phe(4-NH_2)-OH (**12**) as a beige foam (17.7 g, 63.2 mmol, quant.).

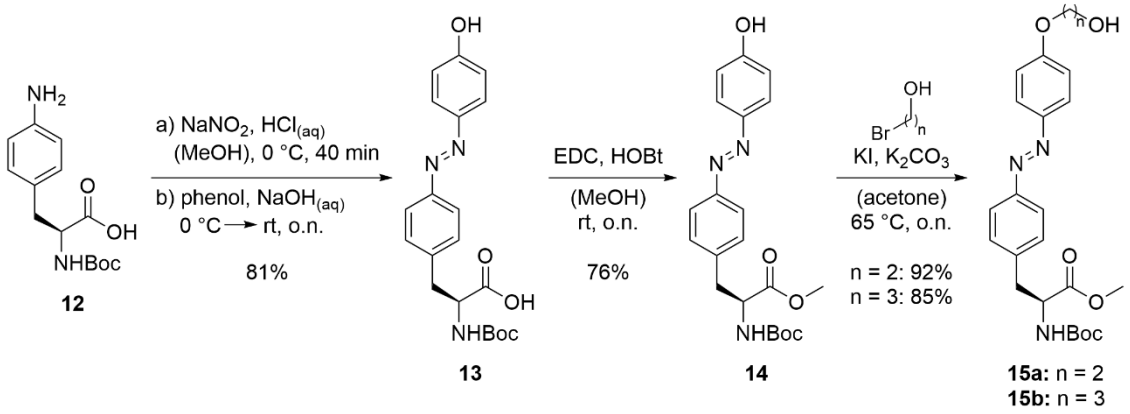
R_f = 0.15 (DCM/MeOH, 19:1 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (500 MHz, CD_3OD, 300 K): δ[ppm] = 7.02 – 6.96 (m, 2H), 6.72 – 6.67 (m, 2H), 4.26 (dd, *J* = 8.4, 5.2 Hz, 1H), 3.02 (dd, *J* = 13.9, 5.2 Hz, 1H), 2.80 (dd, *J* = 13.9, 8.4 Hz, 1H), 1.39 (s, 9H).

^{13}C NMR (101 MHz, CD_3OD, 300 K): δ[ppm] = 175.80 (s, 1C), 157.78 (s, 1C), 146.38 (s, 1C), 131.03 (s, 2C), 128.67 (s, 1C), 117.11 (s, 2C), 80.45 (s, 1C), 56.65 (s, 1C), 38.05 (s, 1C), 28.69 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{14}H_{21}N_2O_4^+$ ($[M+H]^+$): 281.1; found: 281.2.

9.4.2 Synthesis of pnX Series for S_N2



(*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-hydroxyphenyl)diazenyl)phenyl)propanoic acid (13): Diazonium coupling of **12** and phenol by standard procedure B.

After diazotization of **12** (2.72 g, 9.70 mmol, 1.0 eq) with 2 M $NaNO_2$ (6.80 mL, 13.6 mmol, 1.4 eq) in 0.7 M HCl in MeOH (153 mL, 107 mmol, 11 eq) to the corresponding diazonium ion, the solution was added to the phenol (0.910 g, 9.70 mmol, 1.0 eq) in 1 M NaOH (80 mL) to give the crude product, which was dry-loaded onto silica and purified via flash column chromatography (Pent/EtOAc gradient, 2.5:1 + 1 vol% AcOH → 1:1 + 1 vol% AcOH). The desired azobenzene **13** was obtained as an orange foam (3.03 g, 7.87 mmol, 81%).

R_f = 0.28 (Pent:EtOAc, 1:1 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] = 7.83 – 7.75 (m, 4H), 7.38 (d, *J* = 8.3 Hz, 2H), 6.94 – 6.88 (m, 2H), 4.41 (dd, *J* = 9.1, 5.0 Hz, 1H), 3.25 (dd, *J* = 13.9, 5.0 Hz, 1H), 2.99 (dd, *J* = 13.9, 9.1 Hz, 1H), 1.38 (s, 9H).

^{13}C NMR (75 MHz, CD_3OD, 300 K): δ[ppm] = 175.1 (s, 1C), 161.9 (s, 1C), 157.7 (s, 1C), 152.9 (s, 1C), 147.4 (s, 1C), 141.2 (s, 1C), 131.1 (s, 2C), 125.9 (s, 2C), 123.4 (s, 2C), 116.7 (s, 2C), 80.6 (s, 1C), 56.0 (s, 1C), 38.5 (s, 1C), 28.6 (s, 3C).

HRMS-ESI (m/z): calc. for $C_{20}H_{24}N_3O_5^+$ ($[M+H]^+$): 386.1710; found: 386.1699.

methyl (*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-hydroxyphenyl)diazenyl)phenyl)-propanoate (14): Methylation of **13** by standard procedure C.

To the phenolic carboxylic acid **13** (1.4 g, 3.7 mmol, 1.0 eq) were added MeOH until complete dissolution (37 mL), HOBt (0.68 g, 4.5 mmol, 1.2 eq) and EDC×HCl (0.65 g, 4.5 mmol, 1.2 eq). The reaction mixture was stirred o.n. at rt. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 2.5:1) furnished **14** as an orange solid (1.1 g, 2.8 mmol, 76%).

R_f = 0.50 (Pent:EtOAc, 2.5:1) [UV, CAM].

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] = 7.74 – 7.64 (m, 4H), 7.26 (d, *J* = 8.4 Hz, 2H), 6.84 – 6.78 (m, 2H), 4.33 (dd, *J* = 9.1, 5.6 Hz, 1H), 3.61 (s, 3H), 3.09 (dd, *J* = 13.8, 5.6 Hz, 1H), 2.89 (dd, *J* = 13.8, 9.1 Hz, 1H), 1.28 (s, 9H).

^{13}C NMR (75 MHz, $CDCl_3$, 300 K): δ[ppm] = 172.5 (s, 1C), 159.3 (s, 1C), 155.4 (s, 1C), 151.8 (s, 1C), 146.9 (s, 1C), 138.6 (s, 1C), 130.1 (s, 2C), 125.3 (s, 2C), 122.9 (s, 2C), 116.0 (s, 2C), 80.6 (s, 1C), 54.6 (s, 1C), 52.6 (s, 1C), 38.5 (s, 1C), 28.4 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{21}H_{26}N_3O_5^+$ ($[M+H]^+$): 400.2; found: 400.7.

methyl (*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(2-hydroxyethoxy)phenyl)diazenyl)-phenyl)propanoate (15a): Alkylation of the phenol **14** by standard procedure D.

To a solution of phenol **14** (1.0 g, 2.51 mmol, 1.0 eq) in acetone (40 mL) was added solid K_2CO_3 (2.1 g, 15.1 mmol, 6.0 eq), KI (0.83 g, 5.02 mmol, 2.0 eq) and bromoethanol (1.8 mL, 25.2 mmol, 10 eq). The reaction mixture was stirred o.n. at 65 °C. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:2.5) gave **15a** as an orange solid (1.0 g, 2.31 mmol, 92%).

R_f = 0.67 (Pent:EtOAc, 1:2.5) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.90 (d, *J* = 9.0 Hz, 2H), 7.81 (d, *J* = 8.3 Hz, 2H), 7.26 (d, *J* = 8.3 Hz, 2H), 7.02 (d, *J* = 9.0 Hz, 2H), 5.04 (d, *J* = 8.3 Hz, 1H), 4.63 (d, *J* = 7.4 Hz, 1H), 4.18 – 4.15 (m, 2H), 4.00 (td, *J* = 5.2, 3.9 Hz, 2H), 3.72 (s, 3H), 3.16 (m, 2H), 1.42 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{23}H_{30}N_3O_6^+$ ($[M+H]^+$): 444.2; found: 444.0.

methyl (*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(3-hydroxypropoxy)phenyl)di-azenyl)phenyl)propanoate (15b): Alkylation of the phenol **14** by standard procedure D.

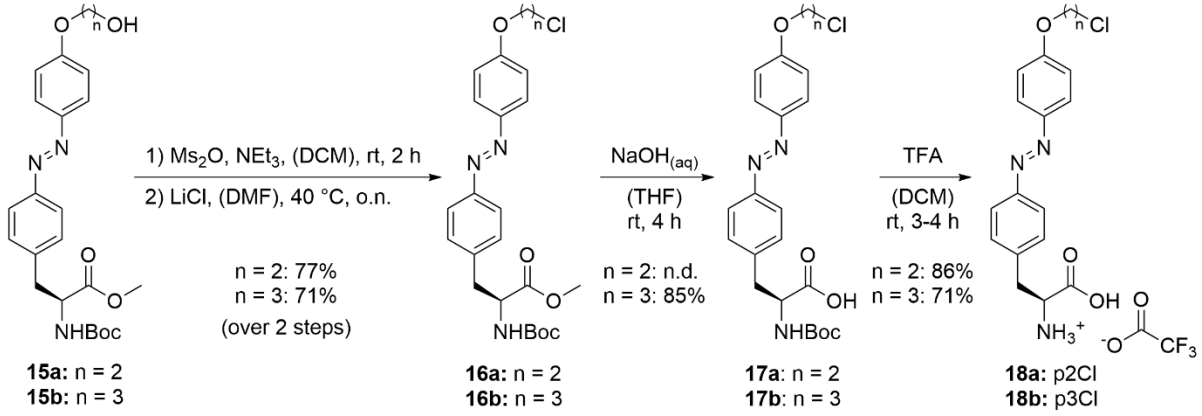
To a solution of phenol **14** (1.0 g, 2.51 mmol, 1.0 eq) in acetone (40 mL) was added solid K_2CO_3 (2.1 g, 15.1 mmol, 6.0 eq), KI (0.62 g, 3.76 mmol, 1.5 eq) and bromopropanol (0.91 mL, 10.0 mmol, 4.0 eq). The reaction mixture was stirred o.n. at 65 °C. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:2.5) yielded **15b** as an orange solid (1.0 g, 2.14 mmol, 85%).

R_f = 0.50 (Pent:EtOAc, 1:2.5) [UV, CAM].

^{1}H NMR (400 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.90 (d, *J* = 9.0 Hz, 2H), 7.81 (d, *J* = 8.3 Hz, 2H), 7.26 (d, *J* = 8.3 Hz, 2H), 7.01 (d, *J* = 9.0 Hz, 2H), 5.02 (d, *J* = 8.3 Hz, 1H), 4.63 (d, *J* = 7.5 Hz, 1H), 4.21 (t, *J* = 6.0 Hz, 2H), 3.89 (t, *J* = 6.0 Hz, 2H), 3.72 (s, 3H), 3.16 (m, 2H), 2.09 (p, *J* = 6.0 Hz, 2H), 1.43 (s, 9H).

^{13}C NMR (101 MHz, $CDCl_3$, 300 K): δ[ppm] = 172.3 (s, 1C), 161.5 (s, 1C), 155.2 (s, 1C), 152.0 (s, 1C), 147.2 (s, 1C), 138.8 (s, 1C), 130.2 (s, 2C), 124.9 (s, 2C), 122.9 (s, 2C), 114.9 (s, 2C), 80.2 (s, 1C), 65.9 (s, 1C), 60.2 (s, 1C), 54.5 (s, 1C), 52.4 (s, 1C), 38.4 (s, 1C), 32.1 (s, 1C), 28.4(s, 3C).

HRMS-ESI (m/z): calc. for $C_{24}H_{32}N_3O_6^+$ ($[M+H]^+$): 458.2286; found: 458.2279.



methyl (*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(2-chloroethoxy)phenyl)diazenyl)-phenyl)propanoate (16a): Chlorination of the aliphatic alcohol **15a** by standard procedure E.

To a cooled solution of alcohol **15a** (0.74 g, 1.68 mmol, 1.0 eq) in dry DCM (35 mL) were added NEt_3 (0.70 mL, 5.01 mmol, 3.0 eq) and Ms_2O (0.44 g, 2.52 mmol, 1.5 eq) and the reaction mixture stirred for 2 h at rt. The crude product was then split in half (approx. 0.84 mmol, 1.0 eq) and subsequently reacted with LiCl (0.28 g, 6.72 mmol, 8.0 eq) in DMF (15 mL), stirring o.n. at 40 °C. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 10:1 → 2.5:1) gave **16a** as an orange solid (0.30 g, 0.65 mmol, 77%).

R_f = 0.59 (Pent:EtOAc, 2.5:1) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.91 (d, *J* = 9.0 Hz, 2H), 7.82 (d, *J* = 8.3 Hz, 2H), 7.26 (d, *J* = 8.3 Hz, 2H), 7.03 (d, *J* = 9.0 Hz, 2H), 5.01 (d, *J* = 8.2 Hz, 1H), 4.63 (d, *J* = 7.4 Hz, 1H), 4.32 (t, *J* = 5.9 Hz, 2H), 3.86 (t, *J* = 5.9 Hz, 2H), 3.73 (s, 3H), 3.18 (m, 2H), 1.43 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{23}H_{29}ClN_3O_5^+$ ($[M+H]^+$): 462.2; found: 462.0.

methyl (*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(3-chloropropoxy)phenyl)diazenyl)-phenyl)propanoate (16b): Chlorination of the aliphatic alcohol **15b** by standard procedure E.

To a cooled solution of alcohol **15b** (0.49 g, 1.07 mmol, 1.0 eq) in dry DCM (20 mL) were added NEt_3 (0.22 mL, 1.61 mmol, 1.5 eq) and MsCl (0.12 mL, 1.61 mmol, 1.5 eq) and the reaction mixture stirred for 1 h at rt. The crude product was subsequently reacted with LiCl (0.27 g, 6.48 mmol, 6.0 eq) in DMF (10 mL), stirring o.n. at rt. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 10:1 → 2.5:1) gave **16b** as an orange solid (0.36 g, 0.76 mmol, 71%).

R_f = 0.36 (Pent:EtOAc, 5:1) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.93 (d, *J* = 9.0 Hz, 2H), 7.83 (d, *J* = 8.4 Hz, 2H), 7.26 (d, *J* = 8.4 Hz, 2H), 7.02 (d, *J* = 9.0 Hz, 2H), 5.01 (d, *J* = 8.3 Hz, 1H), 4.63 (d, *J* = 7.8 Hz, 1H), 4.22 (t, *J* = 6.1 Hz, 2H), 3.77 (t, *J* = 6.1 Hz, 2H), 3.73 (s, 3H), 3.24 – 3.08 (m, 2H), 2.36 – 2.22 (m, 2H), 1.43 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{24}H_{31}ClN_3O_5^+$ ($[M+H]^+$): 476.2; found: 476.0.

(*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(2-chloroethoxy)phenyl)diazenyl)phenyl)-propanoic acid (17a): Hydrolysis of the methyl ester **16a** by standard procedure F.

To **16a** (0.30 g, 0.65 mmol, 1.0 eq) were added THF (2.6 mL) and 1 M aqueous NaOH (2.6 mL, 2.6 mmol, 4.0 eq) and the reaction mixture was stirred for 2.5 h at rt. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:1 + 1 vol% AcOH) gave **17a** as an orange solid (yield n.d.).

R_f = 0.26 (Pent:EtOAc, 1:1+ 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] = 7.93 – 7.87 (m, 2H), 7.80 (d, *J* = 8.1 Hz, 2H), 7.40 (d, *J* = 8.1 Hz, 2H), 7.14 – 7.06 (m, 2H), 4.46 – 4.32 (m, 3H), 3.95 – 3.87 (m, 2H), 3.29 – 3.19 (m, 1H), 3.07 – 2.94 (m, 1H), 1.39 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{22}H_{27}ClN_3O_5^+$ ([M+H]$^+$): 448.2; found: 448.0.

(*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(3-chloropropoxy)phenyl)diazenyl)phenyl)-propanoic acid (17b): Hydrolysis of the methyl ester **16b** by standard procedure F.

To **16b** (0.36 g, 0.76 mmol, 1.0 eq) were added THF (3.0 mL) and 1 M aqueous NaOH (3.0 mL, 3.0 mmol, 4.0 eq) and the reaction mixture was stirred for 4 h at rt. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 2.5:1 + 1 vol% AcOH) gave **17b** as an orange solid (0.30 g, 0.65 mmol, 85%).

R_f = 0.13 (Pent:EtOAc, 2.5:1+ 1 vol% AcOH) [UV, CAM].

^{1}H NMR (400 MHz, CD_3OD, 300 K): δ[ppm] = 7.82 – 7.76 (m, 2H), 7.69 (d, *J* = 8.1 Hz, 2H), 7.30 (d, *J* = 8.1 Hz, 2H), 7.03 – 6.96 (m, 2H), 4.31 (dd, *J* = 9.0, 5.0 Hz, 1H), 4.13 (t, *J* = 5.9 Hz, 2H), 3.69 (t, *J* = 6.4 Hz, 2H), 3.19 – 3.10 (m, 1H), 2.90 (dd, *J* = 13.7, 9.0 Hz, 1H), 2.17 (p, *J* = 6.1 Hz, 2H), 1.29 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{23}H_{29}ClN_3O_5^+$ ([M+H]$^+$): 462.2; found: 462.0.

(*S*,*E*)-2-amino-3-(4-((4-(2-chloroethoxy)phenyl)diazenyl)phenyl)propanoic acid (18a, p2Cl): *N*-Boc deprotection of **17a** according to standard procedure G.

Deprotection of **17a** (0.29 g, 0.64 mmol) with excess TFA (5 mL) in DCM (10 mL) gave the TFA-salt **18a** as an orange solid (0.24 g, 0.52 mmol, 81%) which was used without further purification.

^{1}H NMR (300 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.31 (d, br, *J* = 5.3 Hz, 3H), 7.92 – 7.85 (m, 2H), 7.82 (d, *J* = 8.4 Hz, 2H), 7.46 (d, *J* = 8.4 Hz, 2H), 7.22 – 7.13 (m, 2H), 4.42 – 4.34 (m, 2H), 4.25 (dd, *J* = 11.7, 5.5 Hz, 1H), 4.03 – 3.95 (m, 2H), 3.27 – 3.11 (m, 2H).

LRMS-ESI (m/z): calc. for $C_{17}H_{19}ClN_3O_3^+$ ([M+H]$^+$): 348.1; found: 348.0.

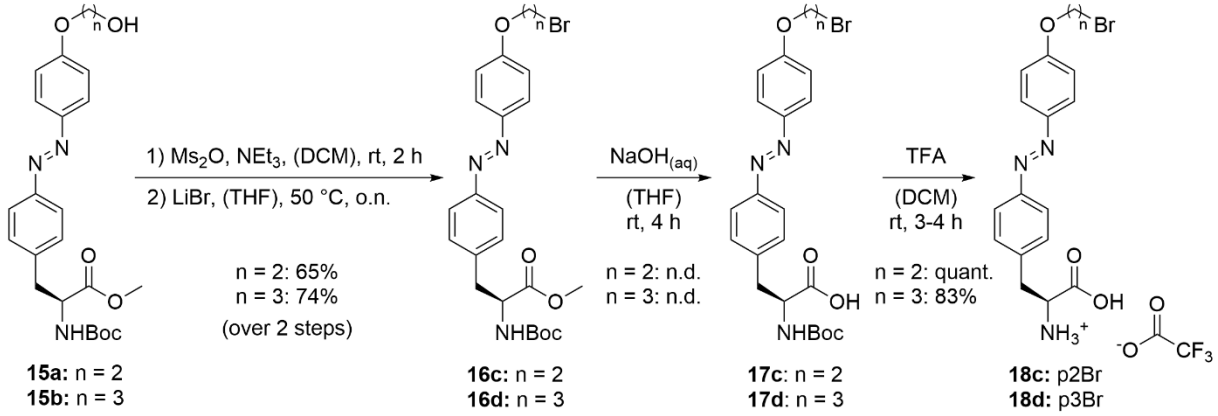
(*S*,*E*)-2-amino-3-(4-((4-(3-chloropropoxy)phenyl)diazenyl)phenyl)propanoic acid (18b, p3Cl): *N*-Boc deprotection of **17b** according to standard procedure G.

Deprotection of **17b** (0.30 g, 0.65 mmol) with excess TFA (5 mL) in DCM (20 mL) gave the TFA-salt **18b** as an orange solid (0.22 g, 0.46 mmol, 71%) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.31 (d, br, *J* = 5.4 Hz, 3H), 7.91 – 7.85 (m, 2H), 7.81 (d, *J* = 8.4 Hz, 2H), 7.46 (d, *J* = 8.4 Hz, 2H), 7.17 – 7.12 (m, 2H), 4.27 (q, *J* = 5.9 Hz, 1H), 4.20 (t, *J* = 6.1 Hz, 2H), 3.81 (t, *J* = 6.4 Hz, 2H), 3.25 – 3.13 (m, 2H), 2.21 (p, *J* = 6.2 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 170.5 (s, 1C), 161.3 (s, 1C), 151.5 (s, 1C), 146.4 (s, 1C), 138.0 (s, 1C), 130.7 (s, 2C), 124.7 (s, 2C), 122.6 (s, 2C), 115.2 (s, 2C), 65.0 (s, 1C), 53.2 (s, 1C), 41.9 (s, 1C), 35.8 (s, 1C), 31.7 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{18}H_{21}ClN_3O_3^+$ ([M+H]$^+$): 362.1266; found: 362.1262.



methyl (*S*,*E*)-3-(4-((4-(2-bromoethoxy)phenyl)diazenyl)phenyl)-2-((*tert*-butoxycarbonyl)-amino)propanoate (16c): Bromination of the aliphatic alcohol **15a** by standard procedure E.

To a cooled solution of alcohol **15a** (1.02 g, 2.31 mmol, 1.0 eq) in dry DCM (40 mL) were added NEt_3 (0.965 mL, 6.93 mmol, 3.0 eq) and Ms_2O (0.603 g, 3.47 mmol, 1.5 eq) and the reaction mixture stirred for 2 h at rt.

The crude product was then split in half (approx. 1.16 mmol, 1.0 eq) and subsequently reacted with LiBr (1.01 g, 11.6 mmol, 10 eq) in THF (15 mL), stirring o.n. at 50 °C. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 10:1 → 1:1) gave **16c** as an orange solid (0.38 g, 0.75 mmol, 65%).

R_f = 0.63 (Pent:EtOAc, 2.5:1) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.94 – 7.87 (m, 2H), 7.81 (d, *J* = 8.4 Hz, 2H), 7.26 (d, *J* = 8.4 Hz, 2H), 7.06 – 6.99 (m, 2H), 5.02 (d, *J* = 8.2 Hz, 1H), 4.63 (d, *J* = 6.4 Hz, 1H), 4.38 (t, *J* = 6.3 Hz, 2H), 3.72 (s, 3H), 3.68 (t, *J* = 6.3 Hz, 2H), 3.18 (m, 2H), 1.43 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{23}H_{29}BrN_3O_5^+$ ($[M+H]^+$): 506.1; found: 506.0.

methyl (*S*,*E*)-3-(4-((4-(3-bromopropoxy)phenyl)diazenyl)phenyl)-2-((*tert*-butoxycarbonyl)-amino)propanoate (16d): Bromination of the aliphatic alcohol **15b** by standard procedure E.

To a cooled solution of alcohol **15b** (0.34 g, 0.74 mmol, 1.0 eq) in dry DCM (15 mL) were added NEt_3 (0.30 mL, 2.16 mmol, 2.9 eq) and Ms_2O (0.20 mg, 1.15 mmol, 1.6 eq) and the reaction mixture stirred for 2 h at rt. The crude product was subsequently reacted with LiBr (0.39 g, 4.48 mmol, 6.0 eq) in THF (15 mL), stirring o.n. at 50 °C. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 10:1 → 2.5:1) gave **16d** as an orange solid (0.28 g, 0.55 mmol, 74%).

R_f = 0.78 (Pent:EtOAc, 2.5:1) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.93 – 7.88 (m, 2H), 7.82 (d, *J* = 8.3 Hz, 2H), 7.26 (d, *J* = 8.3 Hz, 2H), 7.05 – 6.99 (m, 2H), 5.00 (d, *J* = 8.2 Hz, 1H, 1H), 4.63 (m, 1H), 4.20 (t, *J* = 5.8 Hz, 2H), 3.73 (s, 3H), 3.63 (t, *J* = 6.4 Hz, 2H), 3.15 (m, 2H), 2.37 (p, *J* = 6.1 Hz, 2H), 1.42 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{24}H_{31}BrN_3O_5^+$ ($[M+H]^+$): 520.1; found: 520.1.

(*S*,*E*)-3-(4-((4-(2-bromoethoxy)phenyl)diazenyl)phenyl)-2-((*tert*-butoxycarbonyl)amino)-propanoic acid (17c): Hydrolysis of the methyl ester **16c** by standard procedure F.

To **16c** (0.26 g, 0.51 mmol, 1.0 eq) were added THF (2.1 mL) and 1 M aqueous NaOH (2.1 mL, 2.1 mmol, 4.1 eq) and the reaction mixture was stirred for 4 h at rt. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:1 + 1 vol% AcOH) gave **17c** as an orange solid (yield n.d.).

R_f = 0.33 (Pent:EtOAc, 1:1+ 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] = 7.94 – 7.87 (m, 2H), 7.80 (d, *J* = 8.2 Hz, 2H), 7.40 (d, *J* = 8.2 Hz, 2H), 7.13 – 7.07 (m, 2H), 4.45 – 4.37 (m, 3H), 3.75 (m, *J* = 6.3, 5.1 Hz, 2H), 3.27 – 3.20 (m, 1H), 3.00 (dd, *J* = 13.7, 9.1 Hz, 1H), 1.38 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{22}H_{27}BrN_3O_5^+$ ($[M+H]^+$): 492.1; found: 492.0.

(*S*,*E*)-3-(4-((4-(3-bromopropoxy)phenyl)diazenyl)phenyl)-2-((*tert*-butoxycarbonyl)amino)-propanoic acid (17d): Hydrolysis of the methyl ester **16d** by standard procedure F.

To **16d** (0.28 g, 0.55 mmol, 1.0 eq) were added THF (2.2 mL) and 1 M aqueous NaOH (2.2 mL, 2.2 mmol, 4.0 eq) and the reaction mixture was stirred for 4 h at rt. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:1 + 1 vol% AcOH) gave **17d** as an orange solid (yield n.d.).

R_f = 0.38 (Pent:EtOAc, 1:1+ 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] = 7.92 – 7.86 (m, 2H), 7.80 (d, *J* = 8.2 Hz, 2H), 7.40 (d, *J* = 8.2 Hz, 2H), 7.13 – 7.06 (m, 2H), 4.41 (dd, *J* = 9.1, 5.0 Hz, 1H), 4.23 (t, *J* = 5.9 Hz, 2H), 3.66 (t, *J* = 6.5 Hz, 2H), 3.29 – 3.19 (m, 1H), 3.06 – 2.93 (m, 1H), 2.35 (p, *J* = 6.2 Hz, 2H), 1.39 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{23}H_{29}BrN_3O_5^+$ ($[M+H]^+$): 506.1; found: 506.0.

(*S*,*E*)-2-amino-3-(4-((4-(2-bromoethoxy)phenyl)diazenyl)phenyl)propanoic acid (18c, p2Br): *N*-Boc deprotection of **17c** according to standard procedure G.

Deprotection of **17c** (0.26 g, 0.53 mmol) with excess TFA (4 mL) in DCM (20 mL) gave the TFA-salt **18c** as an orange solid (0.29 g, 0.57 mmol, quant.) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.31 (d, br, *J* = 5.3 Hz, 3H), 7.92 – 7.86 (m, 2H), 7.82 (d, *J* = 8.4 Hz, 2H), 7.46 (d, *J* = 8.4 Hz, 2H), 7.20 – 7.14 (m, 2H), 4.44 (dd, *J* = 6.2, 4.7 Hz, 2H), 4.29 (dt, *J* = 6.7, 3.9 Hz, 1H), 3.85 (dd, *J* = 5.8, 4.7 Hz, 2H), 3.26 – 3.13 (m, 2H).

HRMS-ESI (m/z): calc. for $C_{17}H_{19}BrN_3O_3^+$ ($[M+H]^+$): 392.0604; found: 392.0598.

(*S*,*E*)-2-amino-3-(4-((4-(3-bromopropoxy)phenyl)diazenyl)phenyl)propanoic acid (18d, p3Br): *N*-Boc deprotection of **17d** according to standard procedure G.

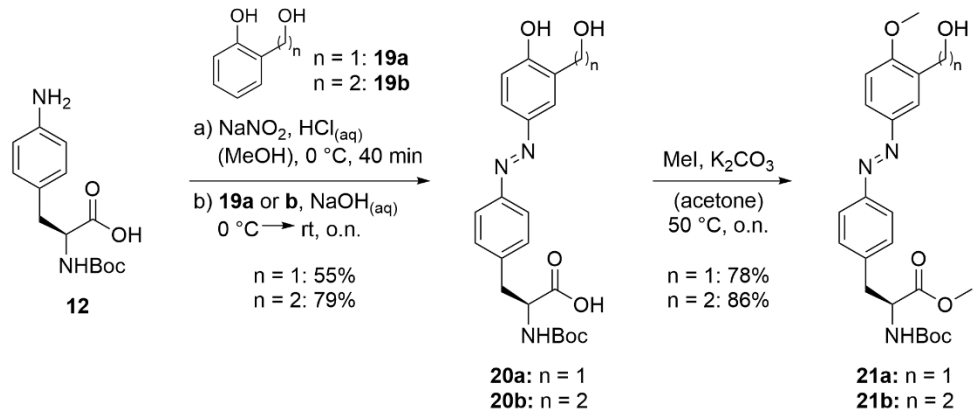
Deprotection of **17d** (0.27 g, 0.55 mmol) with excess TFA (5 mL) in DCM (20 mL) gave the TFA-salt **18d** as an orange solid (0.24 g, 0.45 mmol, 83%) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.32 (d, br, *J* = 5.3 Hz, 3H), 7.90 – 7.85 (m, 2H), 7.81 (d, *J* = 8.4 Hz, 2H), 7.46 (d, *J* = 8.4 Hz, 2H), 7.16 – 7.11 (m, 2H), 4.27 (q, *J* = 6.0 Hz, 1H), 4.19 (t, *J* = 6.0 Hz, 2H), 3.68 (t, *J* = 6.5 Hz, 2H), 3.26 – 3.13 (m, 2H), 2.29 (p, *J* = 6.3 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 170.6 (s, 1C), 161.3 (s, 1C), 151.5 (s, 1C), 146.5 (s, 1C), 138.0 (s, 1C), 130.7 (s, 2C), 124.7 (s, 2C), 122.6 (s, 2C), 115.2 (s, 2C), 66.0 (s, 1C), 53.2 (s, 1C), 35.8 (s, 1C), 31.9 (s, 1C), 31.1 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{18}H_{21}BrN_3O_3^+$ ($[M+H]^+$): 406.0761; found: 406.0749.

9.4.3 Synthesis of mnX Series for S_N2



(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-hydroxy-3-(hydroxymethyl)phenyl)-diazenyl)phenyl)propanoic acid (20a): Diazonium coupling of **12** and phenol **19a** by standard procedure B.

After diazotization of **12** (2.40 g, 8.66 mmol, 1.1 eq) with 2 M $NaNO_2$ (5.70 mL, 11.4 mmol, 1.4 eq) in 0.7 M HCl in MeOH (130 mL, 89.3 mmol, 11 eq) to the corresponding diazonium ion, the solution was added to cooled salicyl alcohol **19a** (1.00 g, 8.10 mmol, 1.0 eq) in 1 M NaOH (70 mL) to give the crude product, which was dry-loaded onto silica and purified via flash column chromatography (Pent/EtOAc gradient, 5:1 + 1 vol% AcOH → 1:2.5 + 1 vol% AcOH). The desired azobenzene **20a** was obtained as an orange foam (1.83 g, 4.42 mmol, 55%).

R_f = 0.27 (Pent:EtOAc, 1:2.5 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 10.27 (s, 1H), 7.97 – 7.89 (m, 1H), 7.74 (d, *J* = 8.4 Hz, 2H), 7.68 (dd, *J* = 8.5, 2.6 Hz, 1H), 7.42 (d, *J* = 8.4 Hz, 2H), 7.16 (d, *J* = 8.4 Hz, 1H), 6.95 (d, *J* = 8.5 Hz, 1H), 5.17 (t, *J* = 5.7 Hz, 1H), 4.55 (d, *J* = 4.1 Hz, 2H), 4.17 (ddd, *J* = 10.2, 8.4, 4.6 Hz, 1H), 3.11 (dd, *J* = 13.9, 4.6 Hz, 1H), 2.91 (dd, *J* = 13.9, 10.2 Hz, 1H), 1.32 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 173.4 (s, 1C), 157.5 (s, 1C), 155.5 (s, 1C), 150.9 (s, 1C), 145.2 (s, 1C), 140.8 (s, 1C), 130.0 (s, 2C), 129.9 (s, 1C), 124.1 (s, 1C), 121.9 (s, 2C), 120.7 (s, 1C), 114.9 (s, 1C), 78.1 (s, 1C), 57.9 (s, 1C), 54.9 (s, 1C), 36.3 (s, 1C), 28.1 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{21}H_{26}N_3O_6^+$ ($[M+H]^+$): 416.1; found: 416.1.

(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-hydroxy-3-(2-hydroxyethyl)phenyl)-diazenyl)phenyl)propanoic acid (20b): Diazonium coupling of **12** and phenol **19b** by standard procedure B.

After diazotization of **12** (2.55 g, 9.11 mmol, 1.0 eq) with 2 M $NaNO_2$ (6.40 mL, 12.8 mmol, 1.4 eq) in 0.7 M HCl in MeOH (150 mL, 100 mmol, 11 eq) to the corresponding diazonium ion, the solution was added to cooled 2-hydroxyphenethyl alcohol **19b** (1.26 g, 9.13 mmol, 1.0 eq) in 1 M NaOH (75 mL) to give the crude product, which was dry-loaded onto silica and purified via flash column chromatography (Pent/EtOAc gradient, 2.5:1 + 1 vol% AcOH → 1:2.5 + 1 vol% AcOH). The desired azobenzene **20b** was obtained as an orange foam (3.09 g, 7.21 mmol, 79%).

R_f = 0.14 (Pent:EtOAc, 1:2.5 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 12.36 (s, 1H), 10.21 (s, 1H), 7.77 – 7.67 (m, 3H), 7.64 (dd, *J* = 8.5, 2.5 Hz, 1H), 7.41 (d, *J* = 8.1 Hz, 2H), 7.16 (d, *J* = 8.4 Hz, 1H), 6.95 (d, *J* = 8.5 Hz, 1H), 4.70 (s, 1H), 4.16 (ddd, *J* = 10.4, 4.6 Hz, 1H), 3.64 (q, *J* = 6.4 Hz, 2H), 3.10 (dd, *J* = 13.9, 4.6 Hz, 1H), 2.91 (dd, *J* = 13.9, 10.4 Hz, 1H), 2.78 (t, *J* = 6.9 Hz, 2H), 1.32 (s, 9H).

^{13}C NMR (75 MHz, DMSO-d_6, 300 K): δ[ppm] = 173.4 (s, 1C), 158.9 (s, 1C), 155.5 (s, 1C), 150.8 (s, 1C), 145.1 (s, 1C), 140.7 (s, 1C), 130.0 (s, 2C), 126.6 (s, 1C), 125.0 (s, 1C), 123.0 (s, 1C), 121.9 (s, 2C), 115.3 (s, 1C), 78.1 (s, 1C), 60.4 (s, 1C), 54.9 (s, 1C), 36.3 (s, 1C), 33.6 (s, 1C), 28.1 (s, 3C).
HRMS-ESI (m/z): calc. for $C_{22}H_{28}N_3O_6^+$ ([M+H]$^+$): 430.1973; found: 430.1966.

methyl (*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(hydroxymethyl)-4-methoxyphenyl)-diazenyl)phenyl)propanoate (21a): Bismethylation of **20a** by standard procedure D.

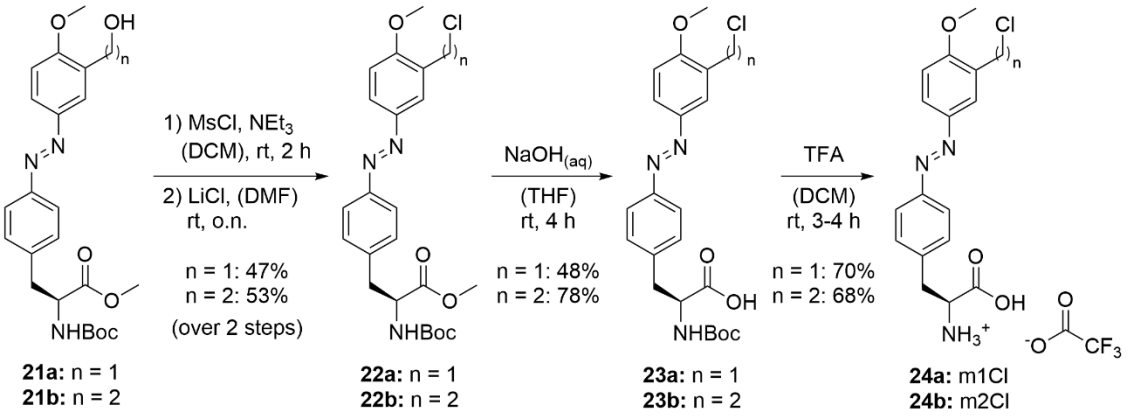
To the phenolic carboxylic acid **20a** (1.72 g, 4.14 mmol, 1.0 eq) were added acetone until complete dissolution (100 mL), K_2CO_3 (4.60 g, 33.2 mmol, 8.0 eq) and MeI (1.55 mL, 24.9 mmol, 6.0 eq) and the reaction mixture was stirred o.n. at 50 °C. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:1) furnished the bismethylated product **21a** as an orange solid (1.43 g, 3.23 mmol, 78%).

R_f = 0.43 (Pent:EtOAc, 1:1) [UV, CAM].
^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.97 (dd, *J* = 2.6, 1.2 Hz, 1H), 7.84 (dd, *J* = 8.7, 2.6 Hz, 1H), 7.78 (d, *J* = 8.4 Hz, 2H), 7.43 (d, *J* = 8.4 Hz, 2H), 7.38 (d, *J* = 8.2 Hz, 1H), 7.16 (d, *J* = 8.7 Hz, 1H), 5.24 (t, *J* = 5.7 Hz, 1H), 4.55 (d, *J* = 5.7 Hz, 2H), 4.25 (ddd, *J* = 10.0, 8.2, 5.2 Hz, 1H), 3.89 (s, 3H), 3.63 (s, 3H), 3.09 (dd, *J* = 13.8, 5.2 Hz, 1H), 2.95 (dd, *J* = 13.8, 10.0 Hz, 1H), 1.32 (s, 9H).
^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 172.4 (s, 1C), 158.6 (s, 1C), 155.4 (s, 1C), 150.8 (s, 1C), 145.9 (s, 1C), 140.6 (s, 1C), 131.7 (s, 1C), 130.1 (s, 2C), 125.1 (s, 1C), 122.1 (s, 2C), 119.1 (s, 1C), 110.4 (s, 1C), 78.3 (s, 1C), 57.7 (s, 1C), 55.8 (s, 1C), 54.9 (s, 1C), 51.8 (s, 1C), 36.2 (s, 1C), 28.1 (s, 3C).
HRMS-ESI (m/z): calc. for $C_{23}H_{30}N_3O_6^+$ ([M+H]$^+$): 444.2129; found: 444.2129.

methyl (*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(2-hydroxyethyl)-4-methoxyphenyl)-diazenyl)phenyl)propanoate (21b): Bismethylation of **20b** by standard procedure D.

To the phenolic carboxylic acid **20b** (3.09 g, 7.21 mmol, 1.0 eq) were added acetone until complete dissolution (160 mL), K_2CO_3 (8.00 g, 58.0 mmol, 8.0 eq) and MeI (2.70 mL, 43.3 mmol, 6.0 eq) and the reaction mixture was stirred o.n. at 50 °C. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:2.5) furnished the bismethylated product **21b** as an orange solid (2.83 g, 6.19 mmol, 86%).

R_f = 0.31 (Pent:EtOAc, 2.5:1) [UV, CAM].
^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.81 – 7.72 (m, 4H), 7.42 (d, *J* = 8.1 Hz, 2H), 7.38 (d, *J* = 8.2 Hz, 1H), 7.16 (d, *J* = 8.8 Hz, 1H), 4.65 (t, *J* = 5.3 Hz, 1H), 4.25 (ddd, *J* = 10.0, 8.2, 5.2 Hz, 1H), 3.89 (s, 3H), 3.67 – 3.57 (m, 5H), 3.09 (dd, *J* = 13.8, 5.2 Hz, 1H), 2.95 (dd, *J* = 13.8, 10.0 Hz, 1H), 2.80 (t, *J* = 6.9 Hz, 2H), 1.32 (s, 9H).
^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 172.4 (s, 1C), 160.1 (s, 1C), 155.4 (s, 1C), 150.8 (s, 1C), 145.7 (s, 1C), 140.6 (s, 1C), 130.1 (s, 2C), 128.3 (s, 1C), 123.9 (s, 1C), 123.6 (s, 1C), 122.1 (s, 2C), 110.9 (s, 1C), 78.3 (s, 1C), 60.3 (s, 1C), 55.9 (s, 1C), 54.9 (s, 1C), 51.8 (s, 1C), 36.2 (s, 1C), 33.5 (s, 1C), 28.1 (s, 3C).
LRMS-ESI (m/z): calc. for $C_{24}H_{32}N_3O_6^+$ ([M+H]$^+$): 458.2; found: 458.2.



methyl (*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(chloromethyl)-4-methoxyphenyl)-diazenyl)phenyl)propanoate (22a): Chlorination of the benzyl alcohol **21a** by standard procedure E.

To a cooled solution of alcohol **21a** (0.23 g, 0.53 mmol, 1.0 eq) in dry DCM (10 mL) were added NEt_3 (0.30 mL, 2.1 mmol, 4.0 eq) and MsCl (0.16 mL, 2.1 mmol, 4.0 eq) and the reaction mixture stirred o.n. at rt to yield a mixture of the mesylated alcohol and benzyl chloride.

The crude product was then reacted with LiCl (0.18 g, 4.2 mmol, 8.0 eq) in DMF (10 mL), stirring for 6 h at rt. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 10:1 → 2.5:1) gave **22a** as an orange solid (0.11 g, 0.25 mmol, 47%).

R_f = 0.74 (Pent:EtOAc, 1:1) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.98 (d, *J* = 2.4 Hz, 1H), 7.92 (dd, *J* = 8.8, 2.4 Hz, 1H), 7.81 (d, *J* = 8.4 Hz, 2H), 7.26 (d, *J* = 8.4 Hz, 2H), 7.01 (d, *J* = 8.8 Hz, 1H), 5.04 (d, *J* = 8.2 Hz, 1H), 4.71 (s, 2H), 4.66 – 4.58 (m, 1H), 3.96 (s, 3H), 3.72 (s, 3H), 3.16 (m, 2H), 1.42 (s, 9H).

^{13}C NMR (75 MHz, $CDCl_3$, 300 K): δ[ppm] = 172.2 (s, 1C), 159.8 (s, 1C), 155.0 (s, 1C), 151.8 (s, 1C), 146.7 (s, 1C), 139.0 (s, 1C), 130.2 (s, 2C), 126.7 (s, 1C), 126.4 (s, 1C), 124.3 (s, 1C), 122.9 (s, 2C), 110.9 (s, 1C), 80.1 (s, 1C), 56.1 (s, 1C), 54.4 (s, 1C), 52.4 (s, 1C), 41.4 (s, 1C), 38.1 (s, 1C), 28.4 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{23}H_{29}ClN_3O_5^+$ ($[M+H]^+$): 462.2; found: 462.1.

methyl (*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(2-chloroethyl)-4-methoxyphenyl)-diazenyl)phenyl)propanoate (22b): Chlorination of the aliphatic alcohol **21b** by standard procedure E.

To a cooled solution of alcohol **21b** (1.4 g, 3.1 mmol, 1.0 eq) in dry DCM (20 mL) were added NEt_3 (1.3 mL, 9.3 mmol, 3.0 eq) and MsCl (0.71 mL, 9.3 mmol, 3.0 eq) and the reaction mixture stirred for 3 h at rt. The crude product was then split 7:1 (approx. 2.70 mmol, 1.0 eq) and subsequently reacted with LiCl (1.0 g, 24 mmol, 7.8 eq) in DMF (20 mL), stirring o.n. at 40 °C. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:1) gave **22b** as an orange solid (0.78 g, 1.6 mmol, 53%).

R_f = 0.81 (Pent:EtOAc, 1:1) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.88 – 7.77 (m, 4H), 7.26 (d, *J* = 8.4 Hz, 2H), 6.97 (d, *J* = 8.7 Hz, 1H), 5.03 (d, *J* = 8.3 Hz, 1H), 4.63 (d, *J* = 7.6 Hz, 1H), 3.91 (s, 3H), 3.80 – 3.70 (m, 5H), 3.25 – 3.08 (m, 4H), 1.43 (s, 9H).

^{13}C NMR (75 MHz, $CDCl_3$, 300 K): δ[ppm] = 172.3 (s, 1C), 160.2 (s, 1C), 155.2 (s, 1C), 151.9 (s, 1C), 146.7 (s, 1C), 138.8 (s, 1C), 130.1 (s, 2C), 127.2 (s, 1C), 125.0 (s, 1C), 124.4 (s, 1C), 122.8 (s, 2C), 110.5 (s, 1C), 80.2 (s, 1C), 55.8 (s, 1C), 54.5 (s, 1C), 52.4 (s, 1C), 43.5 (s, 1C), 38.3 (s, 1C), 34.4 (s, 1C), 28.4 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{24}H_{31}ClN_3O_5^+$ ([M+H]$^+$): 476.2; found: 476.2.

(*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(chloromethyl)-4-methoxyphenyl)diazenyl)-phenyl)propanoic acid (23a): Hydrolysis of the methyl ester **22a** by standard procedure F.

To **22a** (0.21 g, 0.45 mmol, 1.0 eq) were added THF (1.8 mL) and 1 M aqueous NaOH (1.8 mL, 1.8 mmol, 4.0 eq) and the reaction mixture was heated to reflux using a heat-gun until complete hydrolysis of the methyl ester was observed via TLC. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 2.5:1 → 1:2.5 + 1 vol% AcOH) gave **23a** as an orange solid (0.10 g, 0.22 mmol, 48%).

No R_f-value was determined due to the susceptibility of the benzyl chloride towards nucleophiles.

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] = 7.97 – 7.89 (m, 2H), 7.80 (d, *J* = 8.3 Hz, 2H), 7.40 (d, *J* = 8.3 Hz, 2H), 7.15 (d, *J* = 8.7 Hz, 1H), 4.72 (s, 2H), 4.42 (dd, *J* = 9.1, 5.0 Hz, 1H), 3.96 (s, 3H), 3.25 (dd, *J* = 13.9, 5.0 Hz, 1H), 3.00 (dd, *J* = 13.9, 9.1 Hz, 1H), 1.38 (s, 9H).

^{13}C NMR (75 MHz, CD_3OD, 300 K): δ[ppm] = 175.1 (s, 1C), 161.2 (s, 1C), 157.8 (s, 1C), 152.9 (s, 1C), 147.8 (s, 1C), 141.9 (s, 1C), 131.2 (s, 2C), 128.2 (s, 1C), 127.1 (s, 1C), 125.2 (s, 1C), 123.6 (s, 2C), 112.2 (s, 1C), 80.6 (s, 1C), 56.6 (s, 1C), 56.1 (s, 1C), 41.8 (s, 1C), 38.6 (s, 1C), 28.7 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{22}H_{27}ClN_3O_5^+$ ([M+H]$^+$): 448.2; found: 448.0.

(*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(2-chloroethyl)-4-methoxyphenyl)diazenyl)-phenyl)propanoic acid (23b): Hydrolysis of the methyl ester **22b** by standard procedure F.

To **22b** (0.30 g, 0.63 mmol, 1.0 eq) were added THF (2.5 mL) and 1 M aqueous NaOH (2.5 mL, 2.5 mmol, 4.0 eq) and the reaction mixture was heated to reflux using a heat-gun until complete hydrolysis of the methyl ester was observed via TLC. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 2.5:1 → 1:5 + 1 vol% AcOH) gave **23b** as an orange solid (0.23 g, 0.49 mmol, 78%).

R_f = 0.27 (Pent:EtOAc, 1:1+ 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] = 7.82 – 7.70 (m, 4H), 7.37 (d, *J* = 8.4 Hz, 2H), 7.02 (d, *J* = 8.8 Hz, 1H), 4.43 (dd, *J* = 9.1, 5.0 Hz, 1H), 3.86 (s, 3H), 3.72 (t, *J* = 7.3 Hz, 2H), 3.24 (dd, *J* = 13.9, 5.0 Hz, 1H), 3.08 (t, *J* = 7.3 Hz, 2H), 2.99 (dd, *J* = 13.9, 9.1 Hz, 1H), 1.37 (s, 9H).

^{13}C NMR (75 MHz, CD_3OD, 300 K): δ[ppm] = 175.1 (s, 1C), 161.4 (s, 1C), 157.7 (s, 1C), 152.8 (s, 1C), 147.7 (s, 1C), 141.5 (s, 1C), 131.1 (s, 2C), 128.3 (s, 1C), 125.6 (s, 1C), 125.4 (s, 1C), 123.5 (s, 2C), 111.6 (s, 1C), 80.6 (s, 1C), 56.3 (s, 1C), 56.1 (s, 1C), 44.2 (s, 1C), 38.5 (s, 1C), 35.2 (s, 1C), 28.7 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{23}H_{29}ClN_3O_5^+$ ([M+H]$^+$): 462.2; found: 462.2.

(*S*,*E*)-2-amino-3-(4-((3-(chloromethyl)-4-methoxyphenyl)diazenyl)phenyl)propanoic acid (24a, m1Cl): *N*-Boc deprotection of **23a** according to standard procedure G.

Deprotection of **23a** (97 mg, 0.22 mmol) with excess TFA (4 mL) in DCM (6 mL) gave the TFA-salt **24a** as an orange solid (71 mg, 0.15 mmol, 70%). The benzyl chloride linker was labile towards various nucleophiles, including MeOH and H_2O and was therefore not used further.

^{1}H NMR (300 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.44 – 8.29 (m, br, 3H), 8.02 – 7.92 (m, 2H), 7.83 (d, *J* = 8.2 Hz, 2H), 7.47 (d, *J* = 8.2 Hz, 2H), 7.27 (d, *J* = 8.7 Hz, 1H), 4.81 (s, 2H), 4.27 (s, 1H), 3.96 (s, 3H), 3.20 (d, *J* = 6.1 Hz, 2H).

^{13}C NMR (75 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 170.4 (s, 1C), 159.9 (s, 1C), 151.3 (s, 1C), 145.7 (s, 1C), 138.2 (s, 1C), 130.6 (s, 2C), 126.6 (s, 1C), 126.5 (s, 1C), 123.8 (s, 1C), 122.6 (s, 2C), 111.9 (s, 1C), 56.3 (s, 1C), 53.1 (s, 1C), 41.3 (s, 1C), 35.7 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{17}H_{19}ClN_3O_3^+$ ([M+H]$^+$): 348.1; found: 348.1.

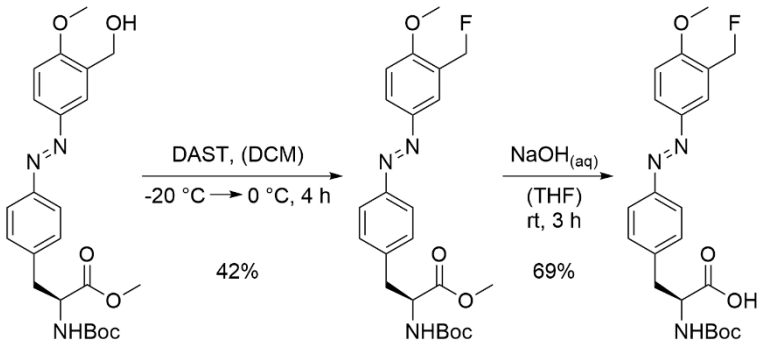
(*S,E*)-2-amino-3-(4-((3-(2-chloroethyl)-4-methoxyphenyl)diazenyl)phenyl)propanoic acid (24b, m2Cl): *N*-Boc deprotection of **23b** according to standard procedure G.

Deprotection of **23b** (0.58 g, 1.25 mmol) with excess TFA (15 mL) in DCM (35 mL) gave the TFA-salt **24b** as an orange solid (0.40 g, 0.85 mmol, 68%) which was used without further purification.

^{1}H NMR (300 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.31 (d, br, J = 6.0 Hz, 3H), 7.88 – 7.77 (m, 4H), 7.46 (d, J = 8.4 Hz, 2H), 7.18 (d, J = 8.9 Hz, 1H), 4.27 (m, 1H), 3.90 (s, 3H), 3.83 (t, J = 7.1 Hz, 2H), 3.28 – 3.14 (m, 2H), 3.10 (t, J = 7.1 Hz, 2H).

^{13}C NMR (75 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 170.6 (s, 1C), 160.4 (s, 1C), 151.6 (s, 1C), 146.0 (s, 1C), 138.0 (s, 1C), 130.7 (s, 2C), 127.1 (s, 1C), 124.7 (s, 1C), 124.3 (s, 1C), 122.6 (s, 2C), 111.4 (s, 1C), 56.2 (s, 1C), 53.3 (s, 1C), 43.8 (s, 1C), 35.9 (s, 1C), 33.5 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{18}H_{21}ClN_3O_3^+$ ([M+H]$^+$): 362.1266; found: 362.1262.



methyl (*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(fluoromethyl)-4-methoxyphenyl)-diazenyl)phenyl)propanoate (25):

DAST (92.0 mg, 0.57 mmol, 1.1 eq) dissolved in dry DCM (5 mL) was added to an evacuated, two-neck round-bottom flask and cooled to -20 °C (NaCl-ice bath, 1:3). To this cooled solution was added dropwise the benzyl alcohol **21a** (229 mg, 0.52 mmol, 1.0 eq) dissolved in dry DCM (5 mL) and the reaction mixture allowed to warm to 0 °C. After stirring at 0 °C for 4 h, the reaction was quenched by pouring into a cold, saturated aqueous $NaHCO_3$-solution. The crude mixture was then extracted with DCM (3 × 20 mL) and the combined organic layers were washed with water (1 × 30 mL), brine (1 × 30 mL), dried over Na_2SO_4, filtered and concentrated. Subsequent purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 10:1 → 1:1) yielded **25** as an orange solid (98.0 mg, 0.22 mmol, 42%).

R_f = 0.77 (Pent:EtOAc, 1:1) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.99 (td, J = 1.4, 0.8 Hz, 1H), 7.94 (ddd, J = 8.8, 2.5, 1.4 Hz, 1H), 7.82 (d, J = 8.4 Hz, 2H), 7.26 (d, J = 8.4 Hz, 2H), 7.00 (dd, J = 8.8, 1.2 Hz, 1H), 5.50 (d, $^2J_{H\text{-}F}$ = 47.6 Hz, 2H), 5.05 (d, J = 8.2 Hz, 1H), 4.63 (m, 1H), 3.92 (s, 3H), 3.72 s, 3H), 3.16 (m, 2H), 1.42 (s, 9H).

^{13}C NMR (75 MHz, $CDCl_3$, 300 K): δ[ppm] = 172.2 (s, 1C), 159.5 (d, $^3J_{C\text{-}F}$ = 4.2 Hz, 1C), 155.2 (s, 1C), 151.8 2 (s, 1C), 146.7 2 (s, 1C), 139.0 (s, 1C), 130.1 (s, 2C), 126.4 (d, $^5J_{C\text{-}F}$ = 2.5 Hz, 1C), 125.7 (d, $^2J_{C\text{-}F}$ = 17.1 Hz, 1C), 122.9 (s, 2C), 122.7 (d, $^3J_{C\text{-}F}$ = 8.0 Hz, 1C), 110.5 (s, 1C), 80.3 (d, $^1J_{C\text{-}F}$ = 165.9 Hz, 1C), 80.2 (s, 1C), 55.9 (s, 1C), 54.5 (s, 1C), 52.4 (s, 1C), 38.3 (s, 1C), 28.4 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{23}H_{29}FN_3O_5^+$ ([M+H]$^+$): 446.2; found: 446.2.

(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(fluoromethyl)-4-methoxyphenyl)diazenyl)-phenyl)propanoic acid (26): Hydrolysis of the methyl ester **25** by standard procedure F.

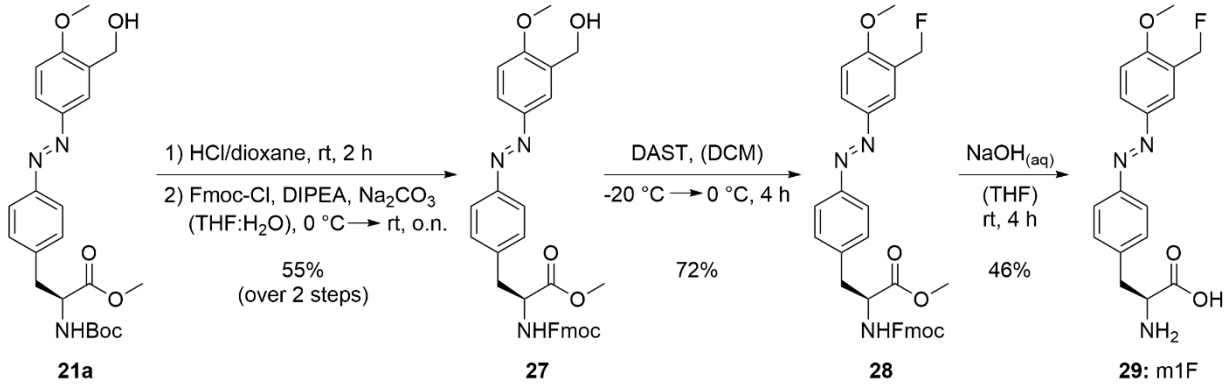
To **25** (136 mg, 0.310 mmol, 1.0 eq) were added THF (2.00 mL) and 1 M aqueous NaOH (1.25 mL, 1.25 mmol, 4.0 eq) and the reaction mixture was stirred at rt for 3 h until complete hydrolysis of the methyl ester was observed via TLC. The desired azobenzene **26** was obtained as an orange solid (92 mg, 0.21 mmol, 69%) after purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:2.5 + 1 vol% AcOH)

R_f = 0.25 (Pent:EtOAc, 1:1+ 1 vol% AcOH) [UV, CAM].

^{1}H NMR (400 MHz, CD_3OD, 300 K): δ[ppm] = 7.94 – 7.88 (m, 2H), 7.79 (d, J = 8.0 Hz, 2H), 7.38 (d, J = 8.0 Hz, 2H), 7.11 (d, J = 9.4 Hz, 1H), 5.45 (d, $^2J_{H\text{-}F}$ = 47.7 Hz, 2H), 4.42 (dd, J = 9.1, 5.0 Hz, 1H), 3.90 (s, 3H), 3.25 (dd, J = 13.9, 5.0 Hz, 1H), 2.99 (dd, J = 13.9, 9.1 Hz, 1H), 1.38 (s, 9H).

^{13}C NMR (101 MHz, CD_3OD, 300 K): δ[ppm] = 175.1 (s, 1C), 160.8 (d, $^3J_{C\text{-}F}$ = 4.1 Hz), 157.8 (s, 1C), 152.8 (s, 1C), 147.8 (s, 1C), 141.8 (s, 1C), 131.2 (s, 2C), 127.3 – 126.9 (m, 2C), 123.6 (s, 2C), 123.2 (d, $^3J_{C\text{-}F}$ = 8.2 Hz, 1C), 111.8 (s, 1C), 80.9 (d, $^1J_{C\text{-}F}$ = 165.1 Hz, 1C), 80.6 (s, 1C), 56.4 (s, 1C), 56.1 (s, 1C), 38.6 (s, 1C), 28.7 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{22}H_{27}FN_3O_5^+$ ([M+H]$^+$): 432.2; found: 432.2.



methyl (*S,E*)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-(4-((3-(hydroxymethyl)-4-methoxyphenyl)diazenyl)phenyl)propanoate (27): *N*-Boc deprotection of **21a** according to standard procedure H. Note that deprotection of **21a** according to standard procedure G using TFA resulted in considerable amounts of trifluoroacetic acid benzyl ester as by-product (> 50%).

Deprotection of **21a** (3.7 g, 8.0 mmol, 1.0 eq) employing 4 M HCl/dioxane (8.0 mL, 32 mmol, 4.0 eq) diluted with dry dioxane (8 mL) gave the HCl-salt which was used without further purification in the subsequent *N*-Fmoc protecting step.

To the crude azobenzene in THF (25 mL) were added DIPEA (1.7 mL, 9.7 mmol, 1.2 eq) and a solution of Na_2CO_3 (0.85 g, 8.0 mmol, 1.0 eq) in H_2O (25 mL). After cooling the solution to 0 °C using an ice bath, solid Fmoc-Cl (2.3 g, 8.8 mmol, 1.1 eq) was added portion wise. The reaction mixture was allowed to warm to ambient and stirred o.n. at rt. Then, THF was removed under reduced pressure, the crude aqueous mixture acidified with aqueous 1 M HCl and extracted with EtOAc (3 × 200 mL). The combined organic layers were washed with aqueous 1 M HCl (1 × 200 mL), water (1 × 200 mL), brine (1 × 200 mL), dried over Na_2SO_4, filtered and concentrated. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:2.5) furnished **27** as an orange solid (2.5 g, 4.4 mmol, 55% over two steps).

R_f = 0.48 (Pent:EtOAc, 1:1) [UV, CAM].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.98 (dd, *J* = 2.6, 1.2 Hz, 1H), 7.93 (d, *J* = 8.3 Hz, 1H), 7.87 (dt, *J* = 7.4, 1.0 Hz, 2H), 7.84 (dd, *J* = 8.8, 2.6 Hz, 1H), 7.77 (d, *J* = 8.4 Hz, 2H), 7.65 – 7.60 (m, 2H), 7.44 (d, *J* = 8.4 Hz, 2H), 7.39 (q, *J* = 7.4 Hz, 2H), 7.30 (dtd, *J* = 12.2, 7.4, 1.1 Hz, 2H), 7.16 (d, *J* = 8.8 Hz, 1H), 5.24 (t, *J* = 5.7 Hz, 1H), 4.56 (d, *J* = 5.7 Hz, 2H), 4.33 (ddd, *J* = 10.3, 8.3, 5.0 Hz, 1H), 4.25 (dd, *J* = 10.4, 6.9 Hz, 2H), 4.18 (t, *J* = 6.9 Hz, 1H), 3.89 (s, 3H), 3.65 (s, 3H), 3.16 (dd, *J* = 13.7, 5.0 Hz, 1H), 2.99 (dd, *J* = 13.7, 10.3 Hz, 1H).

^{13}C NMR (75 MHz, DMSO-d_6, 300 K): δ[ppm] = 172.6 (s, 1C), 159.1 (s, 1C), 156.4 (s, 1C), 151.4 (s, 1C), 146.4 (s, 1C), 144.2 (s, 1C), 141.2 (s, 2C), 141.0 (s, 1C), 132.1 (s, 2C), 130.6 (s, 2C), 128.1 (s, 2C), 127.5 (s, 2C), 125.6 (s, 2C), 125.5 (s, 1C), 122.6 (s, 2C), 120.6 (s, 2C), 119.6 (s, 1C), 110.9 (s, 1C), 66.1 (s, 1C), 58.2 (s, 1C), 56.3 (s, 1C), 55.7 (s, 1C), 52.5 (s, 1C), 47.0 (s, 1C), 36.7 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{33}H_{32}N_3O_6^+$ ($[M+H]^+$): 566.2; found: 566.2.

methyl (*S,E*)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-(4-((3-(fluoromethyl)-4-methoxyphenyl)diazenyl)phenyl)propanoate (28):

DAST (0.10 g, 0.62 mmol, 1.1 eq) dissolved in dry DCM (5 mL) was added to an evacuated, two-neck round-bottom flask and cooled to -20 °C (NaCl-ice bath, 1:3). To this cooled solution was added dropwise the benzyl alcohol **27** (0.25 g, 0.56 mmol, 1.0 eq) dissolved in dry DCM (5 mL) and the reaction mixture allowed to warm to 0 °C. After stirring at 0 °C for 4 h, the reaction was quenched by pouring into a cold, saturated aqueous $NaHCO_3$-solution. The crude mixture was then extracted with DCM (3 × 20 mL) and the combined organic layers were washed with water (1 × 30 mL), brine (1 × 30 mL), dried over Na_2SO_4, filtered and concentrated. Subsequent purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 10:1 → 2.5:1) yielded **28** as an orange solid (0.18 g, 0.40 mmol, 72%).

R_f = 0.31 (Pent:EtOAc, 2.5:1) [UV, CAM].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.98 (dt, *J* = 8.7, 1.9 Hz, 1H), 7.96 – 7.90 (m, 2H), 7.89 – 7.85 (m, 2H), 7.78 (d, *J* = 8.4 Hz, 2H), 7.62 (t, *J* = 7.0 Hz, 2H), 7.45 (d, *J* = 8.4 Hz, 2H), 7.39 (q, *J* = 7.7 Hz, 2H), 7.33 – 7.26 (m, 3H), 5.51 (d, $^2J_{H\text{-}F}$ = 47.6 Hz, 2H), 4.34 (ddd, *J* = 10.3, 8.3, 5.0 Hz, 1H), 4.29 – 4.20 (m, 2H), 4.17 (t, *J* = 7.0 Hz, 1H), 3.94 (s, 3H), 3.65 (s, 3H), 3.16 (dd, *J* = 13.8, 5.0 Hz, 1H), 3.00 (dd, *J* = 13.8, 10.3 Hz, 1H).

^{13}C NMR (75 MHz, $CDCl_3$, 300 K): δ[ppm] = 171.9 (s, 1C), 159.5 (d, $^3J_{C\text{-}F}$ = 4.2 Hz, 1C), 155.7 (s, 1C), 151.9 (s, 1C), 146.8 (s, 1C), 144.0 (s, 1C), 143.8 (s, 1C), 141.5 (s, 2C), 138.6 (s, 1C), 130.2 (s, 2C), 127.9 (s, 2C), 127.2 (s, 2C), 126.4 (d, $^5J_{C\text{-}F}$ = 2.6 Hz, 1C), 125.8 (d, $^2J_{C\text{-}F}$ = 17.1 Hz, 1C), 125.2 (s, 1C), 125.1 (s, 1C), 123.0 (s, 2C), 122.80 (d, $^3J_{C\text{-}F}$ = 8.0 Hz, 1C), 120.1 (s, 2C), 110.6 (s, 1C), 80.32 (d, $^1J_{C\text{-}F}$ = 165.8 Hz, 1C), 67.1 (s, 1C), 56.0 (s, 1C), 54.9 (s, 1C), 52.6 (s, 1C), 47.3 (s, 1C), 38.3 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{33}H_{31}FN_3O_5^+$ ($[M+H]^+$): 568.2; found: 568.2.

(*S,E*)-2-amino-3-(4-((3-(fluoromethyl)-4-methoxyphenyl)diazenyl)phenyl)propanoic acid (29, m1F):

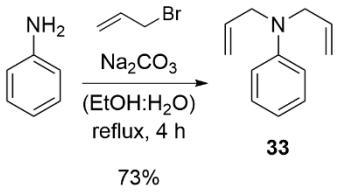
To **28** (0.81 g, 1.4 mmol, 1.0 eq) were added THF (23 mL) and aqueous 1 M NaOH (5.7 mL, 5.7 mmol, 4.0 eq) and the reaction mixture was stirred at rt for 4 h until complete consumption of starting material was observed via TLC. THF was then removed under reduced pressure and the crude aqueous mixture washed with Et_2O (2 × 20 mL). The lyophilized crude product was subsequently purified by HPLC (10 mL/min, 30% B → 55% B + 0.1 vol% FA in 18 min) injecting 1.5 mL of a 100 mM stock solution in DMSO onto a *Luna* C18(2) (10 µm, 100 Å, 250×21.2 mm) column. The desired azo-ncAA **29** was thus obtained as an orange powder (0.25 g, 0.66 mmol, 46%).

^{1}H NMR (400 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.31 (d, br, J = 5.0 Hz, 3H), 7.99 (dt, J = 8.9, 1.9 Hz, 1H), 7.94 (t, J = 1.9 Hz, 1H), 7.84 (d, J = 8.4 Hz, 2H), 7.47 (d, J = 8.4 Hz, 2H), 7.28 (dd, J = 8.9, 1.1 Hz, 1H), 5.51 (d, $^{2}J_{H\text{-}F}$ = 47.5 Hz, 2H), 4.28 (m, 1H), 3.94 (s, 3H), 3.25 – 3.14 (m, 2H).

^{13}C NMR (101 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 170.5 (s, 1C), 159.7 (s, 1C), 151.3 (s, 1C), 145.8 (s, 1C), 138.1 (s, 1C), 130.7 (s, 2C), 126.8 (d, $^{5}J_{C\text{-}F}$ = 2.7 Hz, 1C), 125.2 (d, $^{2}J_{C\text{-}F}$ = 16.7 Hz, 1C), 122.6 (s, 2C), 122.3 (d, $^{3}J_{C\text{-}F}$ = 7.2 Hz, 1C), 111.7 (s, 1C), 79.8 (d, $^{1}J_{C\text{-}F}$ = 162.9 Hz, 1C), 56.2 (s, 1C), 53.1 (s, 1C), 35.7 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{17}H_{19}FN_3O_3^+$ ([M+H]$^+$): 332.1405; found: 332.1401.

9.4.4 Synthesis of pQX Series for Activated S_N2



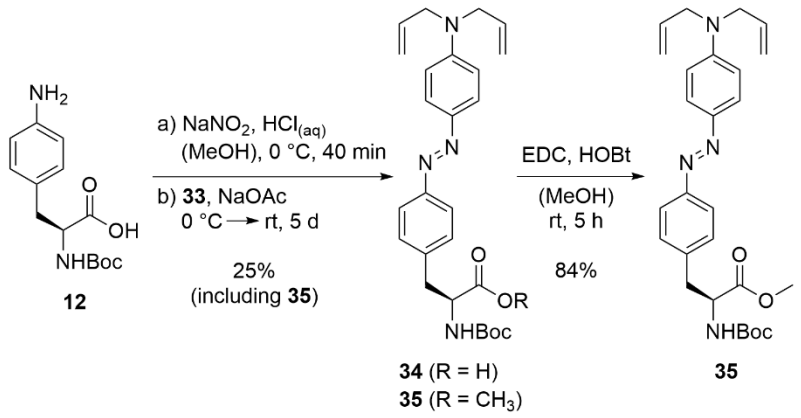
***N*,*N*-diallylaniline (33):**

The synthesis of **33** was carried out similar to known literature procedures.[262]

To aniline (2.0 g, 22 mmol, 1.0 eq) dissolved in EtOH (44 mL) were added a solution of Na_2CO_3 (2.8 g, 26 mmol, 1.2 eq) in H_2O (11 mL) and allyl bromide (4.6 mL, 53 mmol, 2.4 eq). The reaction mixture was heated to reflux at 90 °C for 4 h until complete consumption of aniline. After cooling the mixture to ambient, EtOH was then removed under reduced pressure and the crude aqueous mixture extracted with EtOAc (3 × 200 mL). The combined organic layers were washed with water (1 × 200 mL), brine (1 × 200 mL), dried over Na_2SO_4, filtered and concentrated to furnish **33** as a beige solid (2.8 g, 16 mmol, 73%) which was used without further purification.

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.16 – 7.08 (m, 2H), 6.66 (d, J = 8.0 Hz, 2H), 6.61 – 6.55 (m, 1H), 5.84 (ddt, J = 16.9, 10.2, 5.0 Hz, 2H), 5.15 (dq, J = 7.7, 1.7 Hz, 2H), 5.11 (t, J = 1.7 Hz, 2H), 3.91 (dt, J = 5.0, 1.7 Hz, 4H).

LRMS-ESI (m/z): calc. for $C_{12}H_{16}N^+$ ([M+H]$^+$): 174.1; found: 174.2.



***S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(diallylamino)phenyl)diazenyl)phenyl)-propanoic acid (34):**

The synthesis of **34** was carried out similar to known literature procedures.[262]

After diazotization of **12** (1.88 g, 6.7 mmol, 1.0 eq) with 2 M $NaNO_2$ (3.70 mL, 7.4 mmol, 1.1 eq) in 0.7 M HCl in MeOH (105 mL, 74 mmol, 11 eq) according to standard procedure B, the excess of $NaNO_2$ was quenched by the addition of sulfamic acid (130 mg, 1.3 mmol, 0.20 eq). *N*,*N*-diallylaniline **33** (1.16 g, 6.7 mmol, 1.0 eq) dissolved in MeOH (120 mL) and H_2O (60 mL) was then added dropwise to the cooled diazonium salt solution, thereby turning the reaction mixture pink. Subsequent addition of excess solid NaOAc (4.06 g) adjusted the pH to approx. 3-4 and the resulting mixture was stirred o.n. at 0 °C, warming up to ambient. The next day, the reaction mixture was cooled to 0 °C using a fresh ice bath and stirred at 0 °C for 4 days, also warming up to ambient. After removal of MeOH under reduced pressure, the crude mixture was extracted with EtOAc (3 × 150 mL) and the combined organic layers were washed with aqueous 10% (w/v) KH_2PO_4 (1 × 200 mL), brine (1 × 200 mL), dried over Na_2SO_4, filtered and concentrated. Subsequent purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:1 + 1 vol% AcOH) yielded **34** (0.51 g, 1.1 mmol, 16 %) and the methyl ester derivative **35** (0.27 g, 0.58 mmol, 9%) both as red solids.

Note that the yields might improve by stopping the reaction after 1.5 – 2 days, thereby avoiding *N*-Boc deprotection (approx. 20%, observed via LC-MS).

R_f = 0.24 (Pent:EtOAc, 2.5:1 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.73 (d, *J* = 9.2 Hz, 2H), 7.68 (d, *J* = 8.3 Hz, 2H), 7.38 (d, *J* = 8.3 Hz, 2H), 7.14 (d, *J* = 8.4 Hz, 1H), 6.80 (d, *J* = 9.2 Hz, 2H), 5.89 (ddt, *J* = 17.4, 9.6, 4.9 Hz, 2H), 5.22 – 5.12 (m, 4H), 4.14 (m, 1H), 4.05 (d, *J* = 4.9 Hz, 4H), 3.09 (dd, *J* = 13.7, 4.7 Hz, 1H), 2.89 (dd, *J* = 13.8, 10.2 Hz, 1H), 1.32 (s, 9H).

^{13}C NMR (75 MHz, DMSO-d_6, 300 K): δ[ppm] = 182.9 (s, 1C), 164.9 (s, 1C), 160.6 (s, 1C), 160.3 (s, 1C), 152.2 (s, 1C), 149.4 (s, 1C), 143.0 (s, 2C), 139.4 (s, 2C), 134.0 (s, 2C), 131.1 (s, 2C), 125.7 (s, 2C), 121.3 (s, 2C), 87.5 (s, 1C), 64.5 (s, 1C), 61.9 (s, 2C), 45.7 (s, 1C), 37.6 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{26}H_{33}N_4O_4^+$ ($[M+H]^+$): 465.2; found: 465.3.

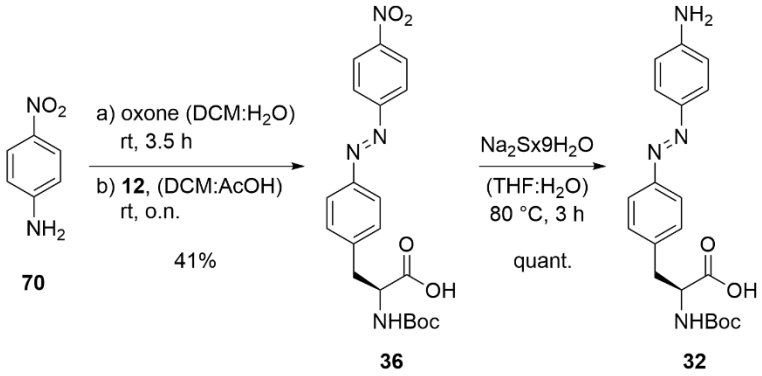
methyl (*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(diallylamino)phenyl)diazenyl)-phenyl)propanoate (35): Methylation of **34** by standard procedure C.

To the carboxylic acid **34** (1.9 g, 4.0 mmol, 1.0 eq) were added MeOH until complete dissolution (40 mL), HOBt (0.74 g, 4.8 mmol, 1.2 eq) and EDC×HCl (0.92 g, 4.8 mmol, 1.2 eq) and the reaction mixture was stirred at rt for 5 h. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 2.5:1) furnished **35** as a red solid (1.6 g, 3.3 mmol, 84%).

R_f = 0.71 (Pent:EtOAc, 2.5:1) [UV, CAM].

LRMS-ESI (m/z): calc. for $C_{27}H_{35}N_4O_4^+$ ([M+H]$^+$): 479.3; found: 479.3.

Further isomerization of the divinyl amine **35** employing the transition metal catalyst $RuClH(CO)(PPh_3)_3$ followed by hydrolysis with $H_2NOH \times HCl$ to liberate the primary aromatic amine group as reported in literature were unsuccessful.[262] Since the synthesis route was discarded, no high quality NMR was measured for **35**.



(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-nitrophenyl)diazenyl)phenyl)propanoic acid (36): Nitroso coupling of **12** and aniline **70** by standard procedure A.

After oxidation of 4-nitro aniline (**70**) (0.45 g, 3.26 mmol, 1.0 eq) with Oxone® (2.0 g, 6.52 mmol, 2.0 eq) to the corresponding nitroso derivative in a biphasic system of DCM (20 mL) and H_2O (20 mL), condensation with **12** (0.91 g, 3.26 mmol, 1.0 eq) in DCM (20 mL) and glacial AcOH (2 mL) gave the crude product which was dry-loaded onto silica for further purification via flash column chromatography (Pent/EtOAc gradient, 5:1 → 2.5:1, then EtOAc + 1 vol% AcOH). The desired azobenzene **36** was obtained (0.55 g, 1.34 mmol, 41%) as a red-orange foam.

R_f = 0.27 (Pent:EtOAc, 1:1 + 1 vol% AcOH) [UV, CAM].

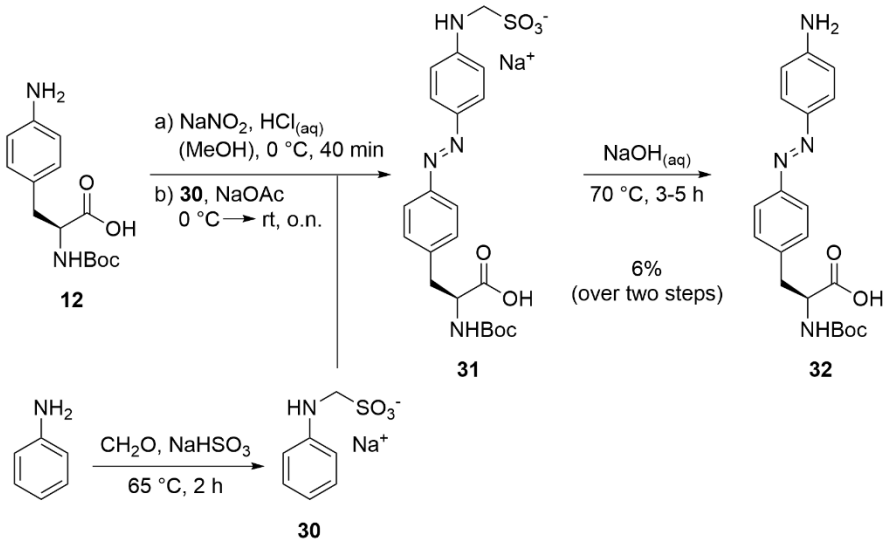
^{1}H NMR (500 MHz, CD_3OD, 300 K): δ[ppm] = 8.44 – 8.41 (m, 2H), 8.10 – 8.06 (m, 2H), 7.93 (d, *J* = 8.3 Hz, 2H), 7.47 (d, *J* = 8.3 Hz, 2H), 4.43 (dd, *J* = 9.3, 5.0 Hz, 1H), 3.32 - 3.28 (m, 1H), 3.03 (dd, *J* = 13.9, 9.3 Hz, 1H), 1.38 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{20}H_{23}N_4NaO_6^+$ ([M+Na]$^+$): 437.1; found: 437.2.

(*S,E*)-3-(4-((4-aminophenyl)diazenyl)phenyl)-2-((*tert*-butoxycarbonyl)amino)propanoic acid (32):

*Successful synthesis of **32** via reduction of **36**.*

To the azobenzene **36** (0.50 g, 1.2 mmol, 1.0 eq) in THF (15 mL) was added a solution of $Na_2S \times 9H_2O$ (0.87 g, 3.6 mmol, 3.0 eq) in H_2O (5 mL) and refluxed at 80 °C for 3 h. The reaction mixture was allowed to cool to ambient and THF was removed under reduced pressure. The crude aqueous mixture was acidified to pH 3-4 using aqueous 10% (w/v) KH_2PO_4 and aqueous 1 M HCl and finally extracted with EtOAc (3 × 40 mL). The combined organic layers were then washed with aqueous 10% (w/v) KH_2PO_4 (1 × 60 mL) and brine (1 × 60 mL), dried over Na_2SO_4 and filtered. The crude was concentrated to yield **32** (0.48 g, 1.2 mmol, quant.) as a red-orange foam which was used without further purification.



*Attempted synthesis of **32** via azo-coupling reaction using anilinomethanesulfonate derivative **30**.*

The attempted synthesis of **32** was carried out similar to known literature procedures.[261]

For the protection of aniline, $NaHSO_3$ (175 mg, 1.7 mmol, 1.2 eq) and an aqueous solution of 37 wt% formaldehyde (0.120 mL, 1.6 mmol, 1.1 eq) containing 15% MeOH as stabilizer were dissolved in water (500 µL) and the reaction mixture was stirred for 1 h at 65 °C. Then aniline (136 mg, 1.5 mmol, 1.0 eq) was added and the solution was stirred for 2 h at 65 °C.

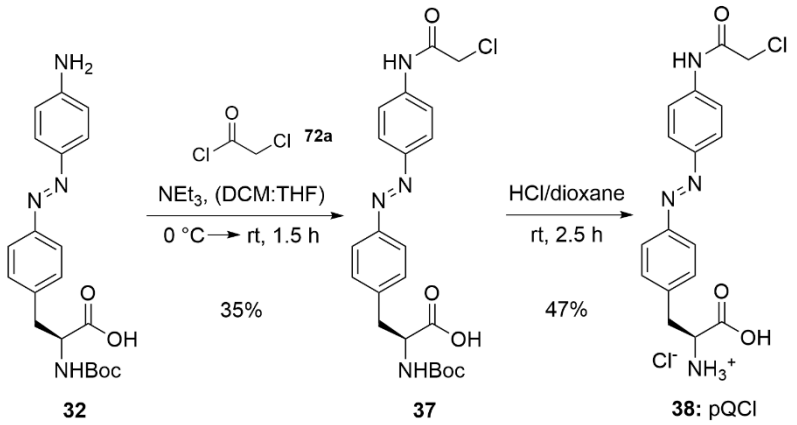
Parallel to diazotization of **12** (409 mg, 1.5 mmol, 1.0 eq) with 2 M $NaNO_2$ (0.880 mL, 1.8 mmol, 1.2 eq) in 0.7 M HCl in MeOH (23.0 mL, 16 mmol, 11 eq) according to standard procedure B, the protected aniline **30** was cooled to rt. The precipitated aniline derivative **30** was then dissolved in a minimum amount of dioxane and added dropwise to the corresponding diazonium salt solution at 0 °C. Next, solid NaOAc (15 g/mmol) was added to adjust the pH to approx. 5, thereby changing the colour of the reaction mixture from pink to orange. The mixture was allowed to warm to ambient and stirred o.n. at rt. Afterwards, MeOH was removed under reduced pressure and an aqueous solution of 10 wt% NaOH was added dropwise to adjust the pH to 9-10. The reaction mixture was stirred for 5 h at 70 °C. The pH of the aqueous mixture was adjusted to 3-4 using an aqueous solution of 1 M HCl and subsequently extracted with EtOAc (3 × 50 mL). The combined organic layers were washed with water (1 × 30 mL) and brine (1 × 30 mL), dried over Na_2SO_4, filtered and concentrated. The crude product was finally purified by flash column chromatography over silica gel (Pent/EtOAc gradient, 2.5:1 → 1:2.5 + 1 vol% AcOH), yielding the desired azobenzene **32** (32 mg, 0.083 mmol, 6%) as a red-orange solid.

R_f = 0.49 (Pent:EtOAc, 1:2.5 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] = 7.75 – 7.67 (m, 4H), 7.35 (d, *J* = 8.3 Hz, 2H), 6.77 – 6.69 (m, 2H), 4.40 (dd, *J* = 9.0, 5.1 Hz, 1H), 3.23 (dd, *J* = 13.9, 5.1 Hz, 1H), 2.98 (dd, *J* = 13.9, 9.0 Hz, 1H), 1.38 (s, 9H).

^{13}C NMR (75 MHz, CD_3OD, 300 K): δ[ppm] = 175.3 (s, 1C), 157.8 (s, 1C), 153.5 (s, 1C), 145.7 (s, 2C*), 140.5 (s, 1C), 131.0 (s, 2C), 126.1 (s, 2C), 123.1 (s, 2C), 115.2 (s, 2C), 80.6 (s, 1C), 56.2 (s, 1C), 38.5 (s, 1C), 28.7(s, 3C).

LRMS-ESI (m/z): calc. for $C_{20}H_{25}N_4O_4^+$ ($[M+H]^+$): 385.2; found: 385.2.



(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(2-chloroacetamido)phenyl)diazenyl)-phenyl)propanoic acid (37): Acylation of amine **32** according to standard procedure I.

A solution of azobenzene **32** (0.47 g, 1.2 mmol, 1.0 eq) and NEt_3 (0.22 mL, 1.6 mmol, 1.3 eq) in THF (6 mL) and DCM (4 mL) was reacted with 2-chloroacetyl chloride (**72a**) (0.10 mL, 1.6 mmol, 1.1 eq) by stirring at rt for 1.5 h. The crude was purified by flash column chromatography (Pent/EtOAc gradient, 1:1 → 1:2.5 + 1 vol% AcOH) to yield **37** (0.20 g, 0.43 mmol, 35%) as an orange foam.

R_f = 0.53 (Pent:EtOAc, 1:5 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 10.64 (s, 1H), 7.93 – 7.86 (m, 2H), 7.84 – 7.76 (m, 4H), 7.45 (d, *J* = 8.2 Hz, 2H), 7.19 (d, *J* = 8.5 Hz, 1H), 4.31 (s, 2H), 4.17 (ddd, *J* = 10.3, 8.5, 4.6 Hz, 1H), 3.12 (dd, *J* = 13.8, 4.6 Hz, 1H), 2.92 (dd, *J* = 13.8, 10.3 Hz, 1H), 1.31 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 173.4 (s, 1C), 165.1 (s, 1C), 155.5 (s, 1C), 150.7 (s, 1C), 148.0 (s, 1C), 141.7 (s, 1C), 141.3 (s, 1C), 130.2 (s, 2C), 123.6 (s, 2C), 122.2 (s, 2C), 119.6 (s, 2C), 78.1 (s, 1C), 54.9 (s, 1C), 43.6 (s, 1C), 36.3 (s, 1C), 28.1 (s, 3C).

HRMS-ESI (m/z): calc. for $C_{22}H_{26}ClN_4O_5^+$ ($[M+H]^+$): 461.1586; found: 461.1591.

(*S,E*)-2-amino-3-(4-((4-(2-chloroacetamido)phenyl)diazenyl)phenyl)propanoic acid (38, pQCl): *N*-Boc deprotection of **37** according to standard procedure H.

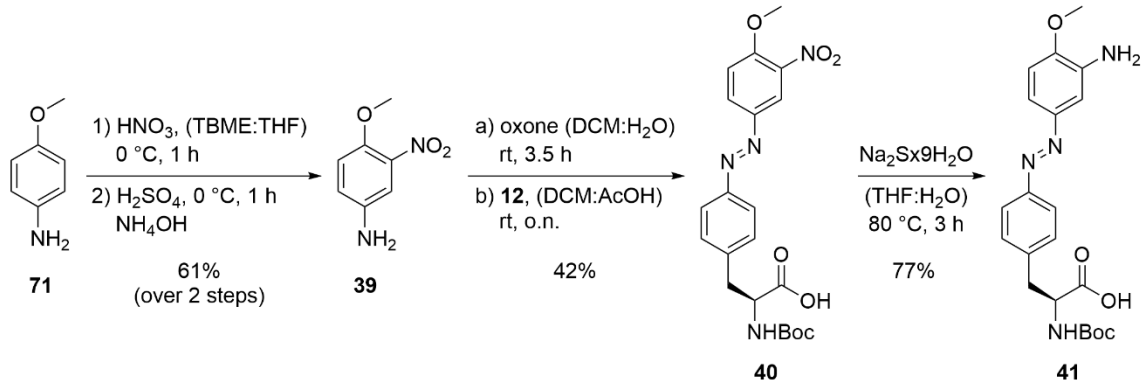
Deprotection of **37** (0.10 g, 0.22 mmol, 1.0 eq) by excess 4 M HCl/dioxane (0.66 mL, 2.6 mmol, 12.0 eq) diluted with dry dioxane (1 mL) gave the HCl-salt **38** as an orange solid (41 mg, 0.10 mmol, 47%) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 10.88 (s, 1H), 8.45 (s, br, 3H), 7.92 – 7.79 (m, 6H), 7.49 (d, *J* = 8.5 Hz, 2H), 4.35 (s, 2H), 4.24 (t, *J* = 6.5 Hz, 1H), 3.23 (d, *J* = 6.5 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 170.3 (s, 1C), 165.1 (s, 1C), 151.3 (s, 1C), 147.9 (s, 1C), 141.6 (s, 1C), 138.4 (s, 1C), 130.6 (s, 2C), 123.7 (s, 2C), 122.5 (s, 2C), 119.6 (s, 2C), 53.0 (s, 1C), 43.6 (s, 1C), 35.5 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{17}H_{18}ClN_4O_3^+$ ($[M+H]^+$): 361.1; found: 361.1.

9.4.5 Synthesis of mQX/mQR Series for S_N2 and Michael Addition



4-methoxy-3-nitroaniline (39):

The synthesis of **39** was carried out similar to known literature procedures.[264]

A solution of *para*-anisidine (**71**) (5.0 g, 41 mmol, 1.0 eq) dissolved in THF (30 mL) and TBME (30 mL) was cooled to 0 °C using an ice bath. To this cooled solution, 65 wt% nitric acid (3.1 mL, 45 mmol, 1.1 eq) was added dropwise and the reaction mixture stirred at 0 °C for 1 h. The precipitate was filtered, washed with cold Et_2O (3×) and dried to give the nitrate salt of **71** as a lilac solid which was used without further purification.

The nitrate salt of **71** was subsequently suspended in DCM (50 mL) and added portion wise to stirring, concentrated H_2SO_4 (13 mL) previously cooled to 0 °C using an ice bath. After stirring at 0 °C for 1 h the mixture was poured onto ice water and an aqueous solution of 25 wt% NH_4OH was added to adjust the pH to approx. 10. The crude aqueous mixture was then extracted with DCM (3 × 200 mL) and the combined organic layers were washed with water (1 × 200 mL), brine (1 × 200 mL), dried over Na_2SO_4, filtered and concentrated to furnish the desired aniline **33** (4.1 g, 22 mmol, 55%) which was used without further purification.

Note that a scale up employing 10 g of *para*-anisidine (**71**) resulted in tedious extractions after the nitration step.

^{1}H NMR (500 MHz, CD_3OD, 300 K): δ[ppm] = 7.13 (d, *J* = 2.8 Hz, 1H), 7.03 (d, *J* = 8.9 Hz, 1H), 6.94 (dd, *J* = 8.9, 2.8 Hz, 1H), 3.84 (s, 3H).

LRMS-ESI (m/z): calc. for $C_7H_9N_2O_3^+$ ($[M+H]^+$): 169.1; found: 169.1.

(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-methoxy-3-nitrophenyl)diazenyl)phenyl)-propanoic acid (40): Nitroso coupling of **12** and aniline **39** by standard procedure A.

After oxidation of 4-methoxy-3-nitroaniline (**39**) (3.3 g, 20 mmol, 1.1 eq) with Oxone® (12 g, 39 mmol, 2.2 eq) to the corresponding nitroso derivative in a biphasic system of DCM (125 mL) and H_2O (125 mL), condensation with **12** (5.1 g, 18 mmol, 1.0 eq) in DCM (150 mL) and glacial AcOH (15 mL) gave the crude product which was dry-loaded onto silica for further purification via flash column chromatography (DCM/EtOAc gradient, 19:1 → 4:1 + 1 vol% AcOH). The desired azobenzene **40** was obtained (3.8 g, 8.4 mmol, 46%) as a red-orange foam.

R_f = 0.39 (Pent:EtOAc, 1:2.5 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.34 (d, *J* = 2.5 Hz, 1H), 8.22 (dd, *J* = 9.0, 2.5 Hz, 1H), 7.82 (d, *J* = 8.2 Hz, 2H), 7.58 (d, *J* = 9.0 Hz, 1H), 7.47 (d, *J* = 8.2 Hz, 2H), 7.19 (d, *J* = 8.5 Hz, 1H), 4.18 (ddd, *J* = 10.4, 8.5, 4.6 Hz, 1H), 4.03 (s, 3H), 3.13 (dd, J = 13.9, 4.6 Hz, 1H), 2.94 (dd, *J* = 13.9, 10.4 Hz, 1H), 1.31 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{21}H_{25}N_4O_7^+$ ($[M+H]^+$): 445.2; found: 445.2.

(*S,E*)-3-(4-((3-amino-4-methoxyphenyl)diazenyl)phenyl)-2-((*tert*-butoxycarbonyl)amino)-propanoic acid (41):

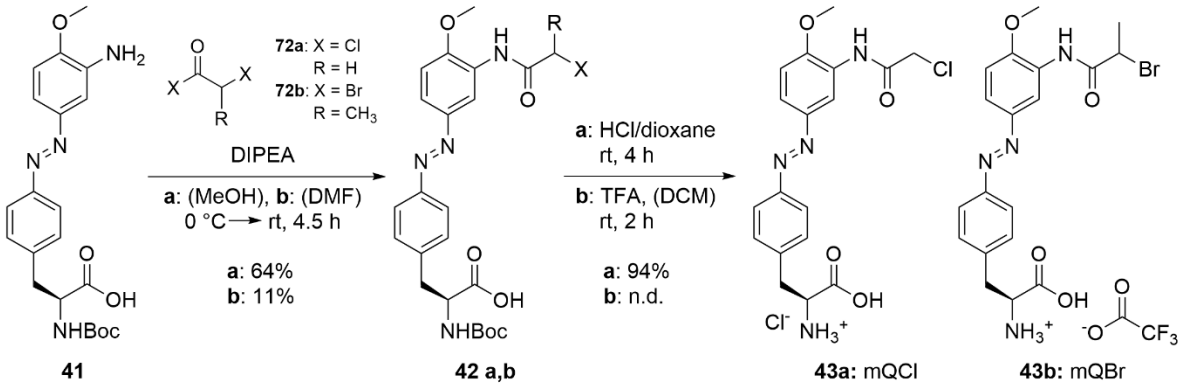
To the azobenzene **40** (3.8 g, 8.45 mmol, 1.0 eq) in THF (90 mL) was added a solution of $Na_2S \times 9H_2O$ (6.1 g, 25.4 mmol, 3.0 eq) in H_2O (30 mL) and refluxed at 80 °C for 3 h. The reaction mixture was allowed to cool to ambient and THF was removed under reduced pressure. The crude aqueous mixture was acidified to pH 3-4 using aqueous 10% (w/v) KH_2PO_4 and aqueous 1 M HCl and finally extracted with EtOAc (3 × 100 mL). The combined organic layers were then washed with aqueous 10% (w/v) KH_2PO_4 (1 × 100 mL) and brine (1 × 100 mL), dried over Na_2SO_4 and filtered. The crude was concentrated, dry-loaded onto silica and purified by flash column chromatography (DCM/EtOAc gradient, 4:1 → 1:1 + 1 vol% AcOH) to yield the intermediate azobenzene **41** (2.6 g, 6.28 mmol, 74%) as a red-orange foam.

R_f = 0.29 (DCM:EtOAc, 3:7 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (400 MHz, CD_3OD, 300 K): δ[ppm] = 7.75 (d, *J* = 8.1 Hz, 2H), 7.43 (dd, *J* = 8.5, 2.4 Hz, 1H), 7.40 (d, *J* = 2.4 Hz, 1H), 7.37 (d, *J* = 8.1 Hz, 2H), 7.01 (d, *J* = 8.5 Hz, 1H), 4.41 (dd, *J* = 9.1, 5.0 Hz, 1H), 3.93 (s, 3H), 3.24 (dd, *J* = 13.9, 5.0 Hz, 1H), 2.98 (dd, *J* = 13.9, 9.1 Hz, 1H), 1.38 (s, 9H).

^{13}C NMR (101 MHz, CD_3OD, 300 K): δ[ppm] = 175.2 (s, 1C), 157.8 (s, 1C), 152.9 (s, 1C), 152.5 (s, 1C), 148.5 (s, 1C), 141.5 (s, 1C), 136.7 (s, 1C), 131.1 (s, 2C), 123.4 (s, 2C), 120.0 (s, 1C), 111.1 (s, 1C), 108.0 (s, 1C), 80.6 (s, 1C), 56.4 (s, 1C), 56.1 (s, 1C), 38.5 (s, 1C), 28.7 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{21}H_{27}N_4O_5^+$ ($[M+H]^+$): 415.2; found: 415.2.



(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(2-chloroacetamido)-4-methoxyphenyl)-diazenyl)phenyl)propanoic acid (42a): Acylation of amine **41** according to standard procedure I.

A solution of azobenzene **41** (0.42 g, 1.0 mmol, 1.0 eq) and DIPEA (0.52 mL, 3.0 mmol, 3.0 eq) in dry MeOH (5 mL) was reacted with 2-chloroacetyl chloride (**72a**) (0.18 mL, 2.2 mmol, 2.2 eq) by stirring at rt for 1.5 h. The crude was purified by flash column chromatography (DCM/EtOAc gradient, 9:1 → 7:3 + 1 vol% AcOH) to yield **42a** (0.32 g, 0.64 mmol, 64%) as a yellow-orange foam.

R_f = 0.23 (Pent:EtOAc, 1:2.5 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 9.69 (s, 1H), 8.59 (d, *J* = 2.4 Hz, 1H), 7.81 – 7.73 (m, 3H), 7.44 (d, *J* = 8.4 Hz, 2H), 7.29 (d, *J* = 8.9 Hz, 1H), 7.17 (d, *J* = 8.5 Hz, 1H), 4.43 (s, 2H), 4.17 (ddd, *J* = 10.3, 8.5, 4.6 Hz, 1H), 3.98 (s, 3H), 3.11 (dd, *J* = 13.9, 4.6 Hz, 1H), 2.92 (dd, *J* = 13.9, 10.3 Hz, 1H), 1.32 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{23}H_{28}ClN_4O_6^+$ ($[M+H]^+$): 491.2; found: 491.2.

(2*S*)-3-(4-((*E*)-(3-(2-bromopropanamido)-4-methoxyphenyl)diazenyl)phenyl)-2-((*tert*-butoxycarbonyl)amino)propanoic acid (42b): Acylation of amine **41** according to standard procedure I.

A solution of azobenzene **41** (0.50 g, 1.2 mmol, 1.0 eq) and DIPEA (0.62 mL, 3.6 mmol, 3.0 eq) in dry DMF (5 mL) was reacted with 2-bromopropionyl bromide (**72b**) (0.28 mL, 2.7 mmol, 2.2 eq) by stirring at rt for 4.5 h. The crude was purified by flash column chromatography (DCM/EtOAc gradient, 25:1 → 6:1 + 1 vol% AcOH) to yield **42b** (70 mg, 0.13 mmol, 11%) as an orange foam.

R_f = 0.45 (DCM:EtOAc, 1:1) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 9.75 (s, 1H), 8.61 (d, *J* = 2.4 Hz, 1H), 7.83 – 7.71 (m, 3H), 7.44 (d, *J* = 8.2 Hz, 2H), 7.29 (dd, *J* = 8.9, 3.5 Hz, 1H), 7.19 (d, *J* = 8.5 Hz, 1H), 5.11 (q, *J* = 6.6 Hz, 1H), 4.17 (m, 1H), 3.97 (s, 3H), 3.11 (dd, *J* = 13.9, 4.7 Hz, 1H), 2.92 (dd, *J* = 13.9, 10.2 Hz, 1H), 1.76 (d, *J* = 6.6 Hz, 3H), 1.32 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{24}H_{30}BrN_4O_6^+$ ($[M+H]^+$): 549.1; found: 549.2.

(*S*,*E*)-2-amino-3-(4-((3-(2-chloroacetamido)-4-methoxyphenyl)diazenyl)phenyl)propanoic acid (43a, mQCl): *N*-Boc deprotection of **42a** according to standard procedure H.

Deprotection of **42a** (0.45 g, 0.91 mmol, 1.0 eq) employing 4 M HCl/dioxane (1.1 mL, 4.6 mmol, 5.0 eq) diluted with dry dioxane (1.5 mL) gave the HCl-salt **43a** as an orange solid (0.37 g, 0.86 mmol, 94%) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 9.74 (s, 1H), 8.61 (d, *J* = 2.4 Hz, 1H), 8.46 (s, br, 3H), 7.83 (d, *J* = 8.2 Hz, 2H), 7.77 (dd, *J* = 8.8, 2.4 Hz, 1H), 7.49 (d, *J* = 8.2 Hz, 2H), 7.30 (d, *J* = 8.8 Hz, 1H), 4.44 (s, 2H), 4.25 (m, 1H), 3.98 (s, 3H), 3.23 (d, *J* = 6.4 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 170.3 (s, 1C), 165.2 (s, 1C), 152.2 (s, 1C), 151.2 (s, 1C), 145.6 (s, 1C), 138.1 (s, 1C), 130.6 (s, 2C), 127.4 (s, 1C), 122.5 (s, 3C), 113.4 (s, 1C), 111.4 (s, 1C), 56.4 (s, 1C), 53.0 (s, 1C), 43.4 (s, 1C), 35.5 (s, 1C).

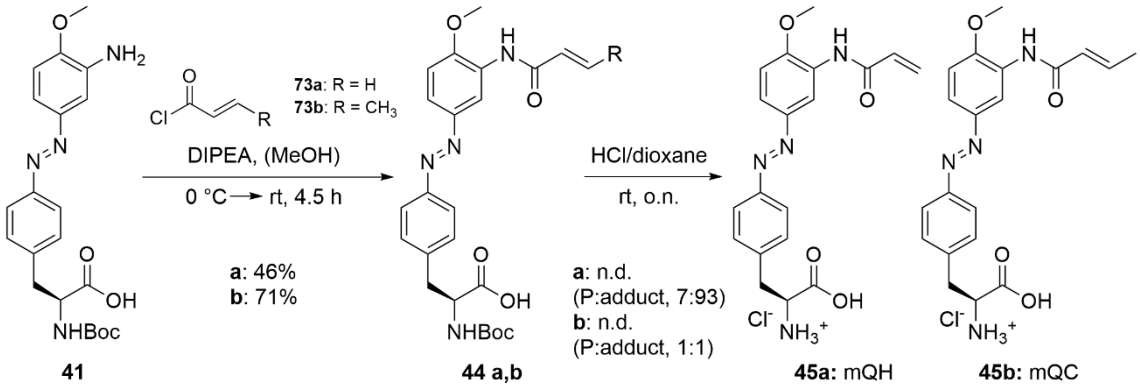
HRMS-ESI (m/z): calc. for $C_{18}H_{20}ClN_4O_4^+$ ($[M+H]^+$): 391.1168; found: 391.1161.

(2*S*)-2-amino-3-(4-((*E*)-(3-(2-bromopropanamido)-4-methoxyphenyl)diazenyl)phenyl)-propanoic acid (43b, mQBr): *N*-Boc deprotection of **42b** according to standard procedure G.

Deprotection of **42b** (60 mg, 0.11 mmol) with excess TFA (2 mL) in DCM (3 mL) gave the TFA-salt **43b** as a brown solid (yield n.d.) which was used without further purification.

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 9.61 (s, 1H), 8.64 (d, *J* = 2.5 Hz, 1H), 8.26 (s, br, 3H), 7.81 (d, *J* = 8.4 Hz, 2H), 7.71 (dd, *J* = 8.7, 2.5 Hz, 1H), 7.42 (d, *J* = 8.4 Hz, 2H), 7.17 (d, *J* = 8.7 Hz, 1H), 5.07-5.01 (m, 1H), 4.26-4.17 (m, 1H), 3.92 (s, 3H), 3.26-3.10 (m, 2H), 1.72 (d, *J* = 6.8 Hz, 3H).

HRMS-ESI (m/z): calc. for $C_{19}H_{22}BrN_4O_4^+$ ($[M+H]^+$): 449.0819; found: 449.0809.



(*S*,*E*)-3-(4-((3-acrylamido-4-methoxyphenyl)diazenyl)phenyl)-2-((*tert*-butoxycarbonyl)-amino)propanoic acid (44a): Acylation of amine **41** according to standard procedure I.

A solution of azobenzene **41** (0.50 g, 1.2 mmol, 1.0 eq) and DIPEA (0.62 mL, 3.6 mmol, 3.0 eq) in dry MeOH (5 mL) was reacted with acryloyl chloride (**73a**) (0.18 mL, 2.2 mmol, 2.2 eq) by stirring at rt for 5.5 h. The crude was purified by flash column chromatography (DCM/EtOAc gradient, 24:1 → 21:4 + 1 vol% AcOH) to yield **44a** (0.26 g, 0.55 mmol, 46%) as an orange foam.

R_f = 0.23 (DCM:EtOAc, 1:1 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 12.65 (s, br, 1H), 9.56 (s, 1H), 8.68 (d, *J* = 2.4 Hz, 1H), 7.78 (d, *J* = 8.3 Hz, 2H), 7.73 (dd, *J* = 8.7, 2.4 Hz, 1H), 7.44 (d, *J* = 8.3 Hz, 2H), 7.27 (d, *J* = 8.7 Hz, 1H), 7.17 (d, *J* = 8.5 Hz, 1H), 6.76 (dd, *J* = 17.0, 10.2 Hz, 1H), 6.28 (dd, *J* = 17.0, 2.0 Hz, 1H), 5.76 (dd, *J* = 10.2, 2.0 Hz, 1H), 4.17 (ddd, *J* = 10.3, 8.5, 4.6 Hz, 1H), 3.97 (s, 3H), 3.12 (dd, *J* = 13.9, 4.6 Hz, 1H), 2.92 (dd, *J* = 13.9, 10.3 Hz, 1H), 1.32 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{24}H_{29}N_4O_6^+$ ([M+H]$^+$): 469.2; found: 469.2.

((*S*)-3-(4-((*E*)-(3-((*E*)-but-2-enamido)-4-methoxyphenyl)diazenyl)phenyl)-2-((*tert*-butoxy-carbonyl)amino)propanoic acid (44b): Acylation of amine **41** according to standard procedure I.

A solution of azobenzene **41** (0.42 g, 1.0 mmol, 1.0 eq) and DIPEA (0.52 mL, 3.0 mmol, 3.0 eq) in dry MeOH (5 mL) was reacted with crotonyl chloride (**73b**) (0.21 mL, 2.2 mmol, 2.2 eq) by stirring at rt for 4.5 h. The crude was purified by flash column chromatography (DCM/EtOAc gradient, 24:1 → 21:4 + 1 vol% AcOH) to yield **44b** (0.34 g, 0.71 mmol, 71%) as an orange foam.

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 9.31 (s, 1H), 8.67 (d, *J* = 2.4 Hz, 1H), 7.78 (d, *J* = 8.3 Hz, 2H), 7.71 (dd, *J* = 8.7, 2.4 Hz, 1H), 7.45 (d, *J* = 8.3 Hz, 2H), 7.26 (d, *J* = 8.7 Hz, 1H), 7.17 (d, *J* = 8.5 Hz, 1H), 6.83 (dd, *J* = 15.2, 6.9 Hz, 1H), 6.45 (dd, *J* = 15.2, 1.6 Hz, 1H), 4.17 (m, 1H), 3.97 (s, 3H), 3.12 (dd, *J* = 13.9, 4.5 Hz, 1H), 3.00 – 2.87 (m, 1H), 1.88 (dd, *J* = 6.9, 1.6 Hz, 3H), 1.32 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{25}H_{31}N_4O_6^+$ ([M+H]$^+$): 483.2; found: 483.3.

(*S*,*E*)-3-(4-((3-acrylamido-4-methoxyphenyl)diazenyl)phenyl)-2-aminopropanoic acid (45a, mQH): *N*-Boc deprotection of **44a** according to standard procedure H.

Deprotection of **44a** (0.26 g, 0.55 mmol, 1.0 eq) employing 4 M HCl/dioxane (0.69 mL, 2.8 mmol, 5.0 eq) diluted with dry dioxane (1.0 mL) gave the HCl-salt **45a** as a brown solid (0.23 g, n.d.) which was used without further purification.

Since the reaction mixture was stirred o.n., a considerable amount of chloro-adduct was obtained as side product (approx. 6.5% of product **45a** were obtained, as judged by ^{1}H-NMR). This may potentially be avoided by shortening reaction times or switching to a different *N*-Boc deprotection method. Consequently, no yield was determined and no NMR is given.

HRMS-ESI (m/z): calc. for $C_{19}H_{21}N_4O_4^+$ ([M+H]$^+$): 369.1557; found: 369.1561.

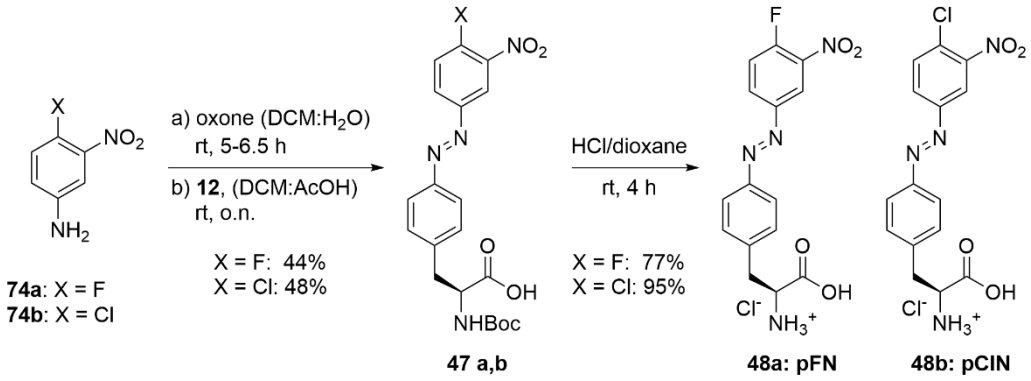
(*S*)-2-amino-3-(4-((*E*)-(3-((*E*)-but-2-enamido)-4-methoxyphenyl)diazenyl)phenyl)propanoic acid (45b, mQC): *N*-Boc deprotection of **44b** according to standard procedure H.

Deprotection of **44b** (0.34 g, 0.71 mmol, 1.0 eq) employing 4 M HCl/dioxane (0.90 mL, 3.6 mmol, 5.0 eq) diluted with dry dioxane (1.0 mL) gave the HCl-salt **45b** as a brown solid (0.28 g, n.d.) which was used without further purification.

Since the reaction mixture was stirred o.n., a considerable amount of chloro-adduct was obtained as side product (approx. 50% of product **45b** were obtained, as judged by ^{1}H-NMR). This may potentially be avoided by shortening reaction times or switching to a different *N*-Boc deprotection method. Consequently, no yield was determined and no NMR is given.

HRMS-ESI (m/z): calc. for $C_{20}H_{23}N_4O_4^+$ ([M+H]$^+$): 383.1714; found: 383.1711.

9.4.6 Synthesis of pXN Derivatives for S_NAr



(*S*,*E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-fluoro-3-nitrophenyl)diazenyl)phenyl)-propanoic acid (47a): Nitroso coupling of **12** and aniline **74a** by standard procedure A.

After oxidation of 4-fluoro-3-nitroaniline (**74a**) (1.7 g, 11 mmol, 1.1 eq) with Oxone® (6.7 g, 22 mmol, 2.2 eq) to the corresponding nitroso derivative in a biphasic system of DCM (30 mL) and H_2O (30 mL), condensation with **12** (2.8 g, 9.9 mmol, 1.0 eq) in DCM (100 mL) and glacial AcOH (10 mL) gave the crude product which was dry-loaded onto silica and purified via automated flash column chromatography (40 g silica, cyclohexane/EtOAc gradient, 19:1 → 17:3 → 6:4 + 1 vol% AcOH). The desired azobenzene **40** was obtained (1.9 g, 4.3 mmol, 44%) as an orange foam.

R_f = 0.37 (Pent:EtOAc, 1:1 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 12.68 (s, 1H), 8.52 (dd, *J* = 7.1, 2.5 Hz, 1H), 8.33 (ddd, *J* = 8.9, 4.2, 2.5 Hz, 1H), 7.88 (d, *J* = 8.4 Hz, 2H), 7.83 (dd, *J* = 10.9, 8.9 Hz, 1H), 7.50 (d, *J* = 8.4 Hz, 2H), 7.21 (d, *J* = 8.5 Hz, 1H), 4.19 (ddd, *J* = 10.4, 8.5, 4.5 Hz, 1H), 3.14 (dd, *J* = 13.8, 4.5 Hz, 1H), 2.95 (dd, *J* = 13.8, 10.4 Hz, 1H), 1.31 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 173.3 (s, 1C), 155.8 (d, $^1J_{C\text{-}F}$ = 266.3 Hz, 1C), 155.4 (s, 1C), 150.2 (s, 1C), 147.8 (d, $^4J_{C\text{-}F}$ = 3.4 Hz, 1C), 143.3 (s, 1C), 137.6 (d, $^2J_{C\text{-}F}$ = 8.8 Hz, 1C), 130.5 (d, $^3J_{C\text{-}F}$ = 9.7 Hz, 1C), 130.4 (s, 2C), 122.8 (s, 2C), 119.8 (d, $^2J_{C\text{-}F}$ = 22.3 Hz, 1C), 118.9 (s, 1C), 78.1 (s, 1C), 54.8 (s, 1C), 36.4 (s, 1C), 28.1 (s, 3C).

HRMS-ESI (m/z): calc. for $C_{20}H_{21}FN_4NaO_6^+$ ([M+Na]$^+$): 455.1337; found: 455.1343.

(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-chloro-3-nitrophenyl)diazenyl)phenyl)-propanoic acid (47b): Nitroso coupling of **12** and aniline **74b** by standard procedure A.

After oxidation of 4-chloro-3-nitroaniline (**74b**) (5.5 g, 32 mmol, 1.1 eq) with Oxone® (20 g, 64 mmol, 2.2 eq) to the corresponding nitroso derivative in a biphasic system of DCM (95 mL) and H_2O (95 mL), condensation with **12** (8.1 g, 29 mmol, 1.0 eq) in DCM (250 mL) and glacial AcOH (25 mL) gave the crude product which was dry-loaded onto silica for further purification via flash column chromatography (Pent/EtOAc gradient, 5:1 → 1:1 + 1 vol% AcOH). The desired azobenzene **47b** was obtained (6.3 g, 14 mmol, 48%) as an orange foam.

R_f = 0.36 (Pent:EtOAc, 1:1 + 1 vol% AcOH) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 12.69 (s, 1H), 8.47 (d, *J* = 2.3 Hz, 1H), 8.17 (dd, *J* = 8.6, 2.3 Hz, 1H), 8.00 (d, *J* = 8.6 Hz, 1H), 7.87 (d, *J* = 8.0 Hz, 2H), 7.50 (d, *J* = 8.0 Hz, 2H), 7.21 (d, *J* = 8.5 Hz, 1H), 4.19 (ddd, *J* = 10.4, 8.5, 4.5 Hz, 1H), 3.15 (dd, *J* = 13.9, 4.5 Hz, 1H), 2.95 (dd, *J* = 13.9, 10.4 Hz, 1H), 1.31 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 173.3 (s, 1C), 155.4 (s, 1C), 150.6 (s, 1C), 150.3 (s, 1C), 148.3 (s, 1C), 143.6 (s, 1C), 132.9 (s, 1C), 130.4 (s, 1C), 127.6 (s, 2C), 127.0 (s, 1C), 122.9 (s, 2C), 118.6 (s, 1C), 78.1 (s, 1C), 54.8 (s, 1C), 36.4 (s, 1C), 28.1 (s, 3C).

HRMS-ESI (m/z): calc. for $C_{15}H_{14}ClN_4O_4^+$ ([M-Boc+H]$^+$): 349.0698; found: 349.0698.

(*S,E*)-2-amino-3-(4-((4-fluoro-3-nitrophenyl)diazenyl)phenyl)propanoic acid (48a, pFN): *N*-Boc deprotection of **47a** according to standard procedure H.

Deprotection of **47a** (5.59 g, 13 mmol, 1.0 eq) employing 4 M HCl/dioxane (19.4 mL, 77 mmol, 6.0 eq) diluted with dry dioxane (12.9 mL) gave the HCl-salt **48a** as an orange solid (3.69 g, 1.0 mmol, 77%) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.63 (s, br, 3H), 8.52 (dd, *J* = 7.0, 2.5 Hz, 1H), 8.33 (ddd, *J* = 8.9, 4.2, 2.5 Hz, 1H), 7.91 (d, *J* = 8.4 Hz, 2H), 7.84 (dd, *J* = 10.8, 8.9 Hz, 1H), 7.56 (d, *J* = 8.4 Hz, 2H), 4.23 (t, *J* = 6.4 Hz, 1H), 3.33 – 3.24 (m, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 170.2 (s, 1C), 155.84 (d, $^1J_{C\text{-}F}$ = 266.6 Hz, 1C), 150.7 (s, 1C), 147.78 (d, $^4J_{C\text{-}F}$ = 3.6 Hz, 1C), 140.2 (s, 1C), 137.62 (d, $^2J_{C\text{-}F}$ = 9.0 Hz, 1C), 130.8 (s, 2C), 130.58 (d, $^3J_{C\text{-}F}$ = 9.7 Hz, 1C), 123.1 (s, 2C), 119.87 (d, $^2J_{C\text{-}F}$ = 22.3 Hz, 1C), 119.01 (d, $^3J_{C\text{-}F}$ = 1.9 Hz, 1C), 53.0 (s, 1C), 35.5 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{15}H_{14}FN_4O_4^+$ ([M+H]$^+$): 333.0994; found: 333.0993.

(*S,E*)-2-amino-3-(4-((4-chloro-3-nitrophenyl)diazenyl)phenyl)propanoic acid (48b, pClN): *N*-Boc deprotection of **47b** according to standard procedure H.

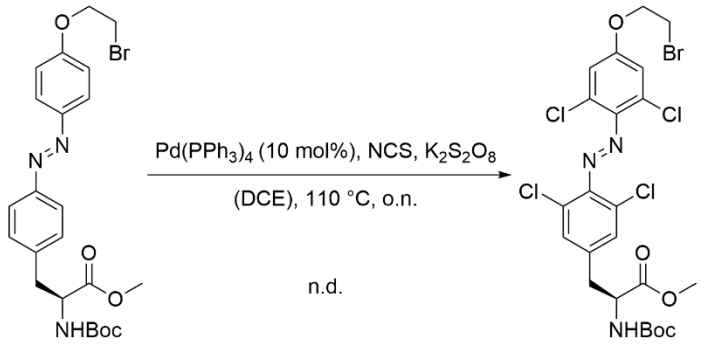
Deprotection of **47b** (6.3 g, 13.9 mmol, 1.0 eq) by excess 4 M HCl/dioxane (29 mL, 112 mmol, 8.0 eq) diluted with dry dioxane (14 mL) gave the HCl-salt **48b** as an orange solid (5.1 g, 13.2 mmol, 95%) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.56 (d, br, *J* = 5.2 Hz, 3H), 8.47 (d, *J* = 2.3 Hz, 1H), 8.18 (dd, *J* = 8.6, 2.3 Hz, 1H), 7.99 (d, *J* = 8.6 Hz, 1H), 7.90 (d, *J* = 8.2 Hz, 2H), 7.55 (d, *J* = 8.2 Hz, 2H), 4.25 (m, 1H), 3.27 (d, *J* = 6.5 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 170.3 (s, 1C), 150.9 (s, 1C), 150.6 (s, 1C), 148.4 (s, 1C), 140.4 (s, 1C), 133.0 (s, 1C), 131.0 (s, 2C), 127.7 (s, 1C), 127.3 (s, 1C), 123.3 (s, 2C), 118.7 (s, 1C), 53.0 (s, 1C), 35.7 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{15}H_{14}ClN_4O_4^+$ ([M+H]$^+$): 349.0698; found: 349.0698.

9.4.7 Synthesis of tetra-*ortho* Halogenated Azo-ncAAs



16c **49**

methyl (*S,E*)-3-(4-((4-(2-bromoethoxy)-2,6-dichlorophenyl)diazenyl)-3,5-dichlorophenyl)-2-((*tert*-butoxycarbonyl)amino)propanoate (49):

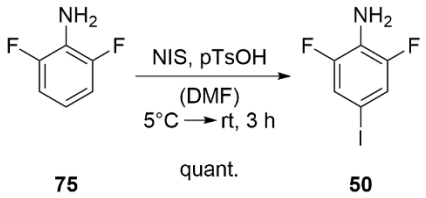
The synthesis of **49** was carried out similar to known literature procedures.[268, 269]

The azobenzene **16c** (88.0 mg, 0.17 mmol, 1.0 eq) was given into a pressure tube and dissolved in dry DCE (5 mL). To this, NCS (116 mg, 0.87 mmol, 5.1 eq), $Pd(PPh_3)_4$ (20.0 mg, 10 mol-%) and $K_2S_2O_8$ (56.0 mg, 0.21 mmol, 1.2 eq) were added and the reaction mixture refluxed o.n. at 110 °C. The reaction mixture was allowed to cool to ambient temperature, filtered and the organic solvent was removed under reduced pressure. The crude was subsequently purified by flash column chromatography (Pent/EtOAc gradient, 10:1 → 2.5:1) to yield the tetra-*ortho*-chlorinated azobenzene **49**. Repeated purification via flash column chromatography did not improve the purity of the obtained azo-ncAA. Therefore, no yield was determined and no NMR is given.

R_f = 0.77 (Pent:EtOAc, 2.5:1) [UV, CAM].

HRMS-ESI (m/z): calc. for $C_{23}H_{25}BrCl_4N_3NaO_5^+$ ([M+Na]$^+$): 663.9546; found: 663.9542.

Details of the various conditions and test compounds that had to be screened to establish the procedure for the synthesis of tetra-*ortho* chlorinated crosslinking azo-ncAAs on a small scale are disclosed in the Bachelor Thesis by Max Bottlinger.[347] Note that scale up of the reaction with the established conditions led to reaction mixtures of tri-, tetra- and penta-chlorinated product. Therefore, purification of these particular azo-ncAAs was not investigated any further.



75 50

2,6-difluoro-4-iodoaniline (50):

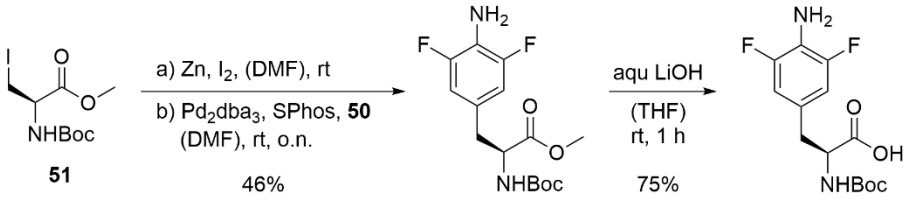
The synthesis of **50** was carried out similar to known literature procedures.[71]

A solution of 2,6-difluoroaniline (**75**) (2.3 mL, 20 mmol, 1.0 eq) and *p*-TsOH (0.76 g, 4.0 mmol, 0.20 eq) in dry DMF (35 mL) was cooled to 0 °C. To this cooled solution, NIS (5.0 g, 22 mmol, 1.1 eq) dissolved in dry DMF (25 mL) was added dropwise and the faint yellow solution allowed to warm to ambient temperature. The reaction mixture was then stirred at rt for 3 h until complete consumption of the starting material was observed via LC-MS. The crude was diluted with aqueous 10% (w/v) LiCl (500 mL), extracted with EtOAc (3 × 100 mL) and the combined organic layers were washed with water (3 × 400 mL), brine (1 × 200 mL), dried over Na_2SO_4, filtered and concentrated to furnish the desired aniline **50** (5.3 g, 21 mmol, quant.) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.26 (dd, J = 6.3, 1.9 Hz, 2H), 5.39 (s, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 151.1 (dd, $^1J_{C\text{-}F}$ = 243.7, $^3J_{C\text{-}F}$ = 9.5 Hz, 2C), 126.0 (t, $^2J_{C\text{-}F}$ = 16.4 Hz, 1C), 120.0 – 119.5 (m, 2C), 72.0 (t, $^3J_{C\text{-}F}$ = 9.7 Hz, 1C).

HRMS-ESI (m/z): calc. for $C_6H_5F_2IN^+$ ([M+H]$^+$): 254.9356; found: 254.9348.



51 46% 52 75% 53

methyl (*S*)-3-(4-amino-3,5-difluorophenyl)-2-((*tert*-butoxycarbonyl)amino)propanoate (52):

The synthesis of **52** was carried out similar to known literature procedures.[271]

To an evacuated *Schlenk* tube was added Zn-dust (195 mg, 3.0 mmol, 3.0 eq) and the system flame-dried and flushed with argon (2×). Then iodine (38.0 mg, 0.15 mmol, 0.15 eq)and dry DMF (1 mL) were added and the reaction mixture stirred vigorously (750 rpm) at rt. Immediately after changing its color from brown back to being colorless (typically 1 min), **51** (329 mg, 1.0 mmol, 1.0 eq) and additional iodine (38.0 mg, 0.15 mmol, 0.15 eq) were added to the reaction mixture, triggering the same color change as before. The colorless suspension was stirred at rt for 10 min before the addition of Pd_2dba_3 (23.0 mg, 2.5 mol-%), SPhos (21.0 mg, 5 mol-%) and **50** (332 mg, 1.3 mmol, 1.3 eq). A color change from dark-brown to slightly yellow was observed (typically within 30 min). The reaction mixture was subsequently stirred o.n. at rt and reaction progress was checked via TLC. The crude mixture was then diluted with aqueous 10% (w/v) LiCl (100 mL), extracted with EtOAc (3 × 30 mL) and the combined organic layers were washed with water (3 × 100 mL), brine (1 × 50 mL), dried over Na_2SO_4, filtered and concentrated. Purification by automated flash column chromatography (24 g silica, cyclohexane/EtOAc gradient, 19:1 → 4:1) furnished the desired phenylalanine derivative **51** (152 mg, 0.46 mmol, 46%) as a brown oil.

R_f = 0.32 (cyclohexane:EtOAc, 7:3) [UV, ninhydrin].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.25 (d, J = 8.2 Hz, 1H), 6.79 (dd, J = 7.7, 2.2 Hz, 2H), 5.02 (s, 2H), 4.11 (ddd, J = 10.3, 8.2, 5.0 Hz, 1H), 3.62 (s, 3H), 2.87 (dd, J = 13.8, 5.0 Hz, 1H), 2.69 (dd, J = 13.8, 10.3 Hz, 1H), 1.33 (s, 9H).

HRMS-ESI (m/z): calc. for $C_{15}H_{20}F_2N_2NaO_4^+$ ([M+H]$^+$): 353.1283; found: 353.1286.

(*S*)-3-(4-amino-3,5-difluorophenyl)-2-((*tert*-butoxycarbonyl)amino)propanoic acid (53): Hydrolysis of the methyl ester **52** by standard procedure F.

To **52** (0.15 g, 0.46 mmol, 1.0 eq) were added THF (1.0 mL) and 1 M aqueous LiOH (1.0 mL, 1.0 mmol, 2.0 eq) and the reaction mixture was stirred at rt for 1 h until complete hydrolysis of the methyl ester was observed via TLC. The desired product **53** was obtained (0.11 g, 0.34 mmol, 75%) after purification via automated flash column chromatography over silica gel (4 g silica, cyclohexane/EtOAc gradient, 4:1 → 1:4 + 1 vol% AcOH).

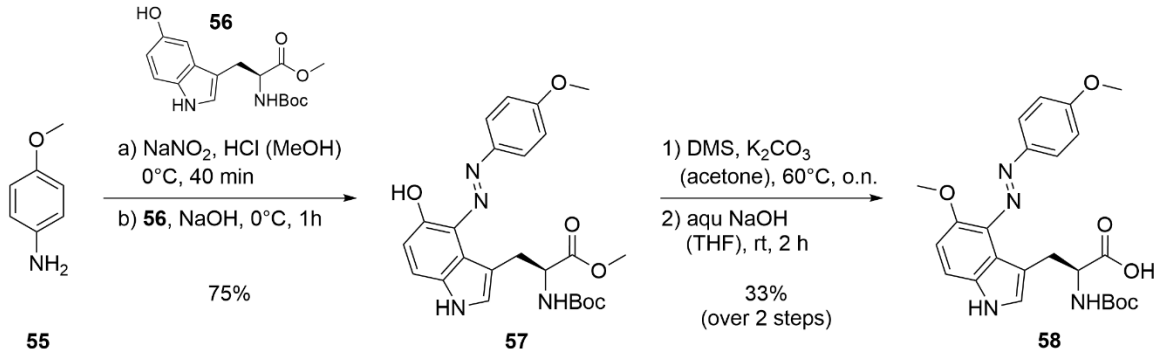
R_f = 0.11 (cyclohexane:EtOAc, 1:1+ 1 vol% AcOH) [UV, ninhydrin].

^{1}H NMR (400 MHz, $CDCl_3$, 300 K): δ[ppm] =6.75 – 6.60 (m, 2H), 6.06 (s, 2H), 5.05 (d, J = 8.0 Hz, 1H), 4.53 (m, 1H), 3.12 – 2.89 (m, 2H), 1.39 (s, 9H).

HRMS-ESI (m/z): calc. for $C_{14}H_{18}F_2N_2NaO_4^+$ ([M+H]$^+$): 339.1127; found: 339.1122.

Subsequent attempts at coupling **53** to 4-fluoro-3-nitroaniline (**74a**) according to standard procedure A in glacial AcOH over prolonged reaction times and at elevated temperatures failed so far and needs to be investigated in more detail.

9.4.8 Synthesis of Trp-Based Azo-ncAA



methyl (*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(5-hydroxy-4-((4-methoxyphenyl)diazenyl)-1H-indol-3-yl)propanoate (57): Diazonium coupling of **55** and phenol **56** by standard procedure B.

After diazotization of **55** (0.22 mg, 1.80 mmol, 1.1 eq) with 2 M $NaNO_2$ (1.1 mL, 2.28 mmol, 1.4 eq) in 0.7 M HCl in MeOH (27 mL, 19.1 mmol, 12 eq) to the corresponding diazonium ion, the trypthophane derivative **56** (0.54 mg, 1.63 mmol, 1.0 eq) and aqueous 1 M NaOH was added to give the crude product, which was dry-loaded onto silica and purified via flash column chromatography (Pent/EtOAc gradient, 5:1 → 1:2.5). The desired azobenzene **57** was obtained as an red-orange solid (0.57 mg, 1.22 mmol, 75%).

R_f = 0.52 (Pent:EtOAc, 1:2.5) [UV, CAM].

^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ[ppm] = 8.76 (s, br, 1H), 7.85 (d, J = 8.5 Hz, 2H), 7.28 (d, J = 8.8 Hz, 1H), 7.11 – 6.98 (m, 3H), 6.79 (d, J = 8.8 Hz, 1H), 5.55 (d, J = 7.4 Hz, 1H), 4.66 – 4.52 (m, 1H), 3.88 (s, 3H), 3.69 (s, 3H), 3.63 – 3.40 (m, 2H), 1.24 (s, 9H).

^{13}C NMR (75 MHz, $CDCl_3$, 300 K): δ[ppm] = 173.6 (s, 1C), 161.4 (s, 1C), 155.7 (s, 1C), 151.5 (s, 1C), 143.8 (s, 1C), 131.2 (s, 1C), 130.7 (s, 1C), 126.1 (s, 1C), 125.4 (s, 1C), 123.5 (s, 2C), 118.7 (s, 1C), 114.8 (s, 2C), 114.0 (s, 1C), 112.0 (s, 1C), 79.7 (s, 1C), 55.8 (s, 1C), 55.4 (s, 1C), 52.2 (s, 1C), 29.9 (s, 1C), 28.3 (s, 3C).

(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(5-methoxy-4-((4-methoxyphenyl)diazenyl)-1H-indol-3-yl)propanoic acid (58): Methylation of **57** and subsequent hydrolysis of the methyl ester by standard procedure F.

To the tryptophane-based azo-ncAA **57** (218 mg, 0.47 mmol, 1.0 eq) was added dry acetone until complete dissolution (30 mL), solid K_2CO_3 (519 mg, 3.8 mmol, 8.0 eq) and dimethyl sulfate (0.180 mL, 1.9 mmol, 4.0 eq). The reaction mixture was then stirred o.n. at 60 °C until complete consumption of the starting material was observed via LC-MS. The organic solvent was removed under reduced pressure and the crude resuspended in a mix of water (20 mL) and EtOAc (20 mL). The organic phase was separated and the aqueous layer extracted with EtOAc (3 × 20 mL).

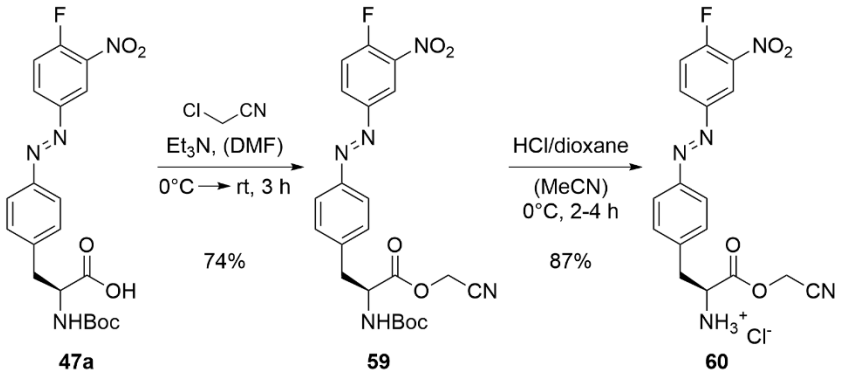
The combined organic layers were washed with water (1 × 30 mL), brine (1 × 30 mL), dried over Na_2SO_4, filtered and concentrated. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 2.5:1 → 1:2.5) furnished the alkylated product as an orange solid, which was subsequently dissolved in THF (2.00 mL). To this, 1 M aqueous NaOH (2.00 mL, 2.0 mmol, 4.3 eq) was added and the reaction mixture stirred o.n. at rt until complete hydrolysis of the methyl ester was observed via TLC. The desired azobenzene **58** was obtained as an orange solid (73 mg, 0.16 mmol, 33%) after purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 2.5:1 → 1:2.5 + 1 vol% AcOH).

^{1}H NMR (300 MHz, CD_3OD, 300 K): δ[ppm] =7.92 – 7.82 (m, 2H), 7.31 (d, *J* = 9.0 Hz, 1H), 7.09 – 6.98 (m, 3H), 6.72 (dd, *J* = 9.0, 3.8 Hz, 1H), 4.46 (dd, *J* = 9.5, 5.6 Hz, 1H), 3.85 (s,3H), 3.71 (s, 3H), 3.61 – 3.51 (m, 1H), 3.28 – 3.18 (m, 1H), 1.25 (s, 9H).

^{13}C NMR (75 MHz, CD_3OD, 300 K): δ[ppm] = 176.4 (s, 1C), 162.8 (s, 1C), 157.8 (s, 1C), 151.7 (s, 1C), 145.7 (s, 2C), 133.6 (s, 1C), 132.2 (s, 1C), 126.8 (s, 1C), 124.5 (s, 2C), 117.3 (s, 1C), 115.7 (s, 2C), 113.8 (s, 1C), 111.8 (s, 1C), 80.2 (s, 1C), 56.1 (s, 2C), 32.9 (s, 1C), 30.8 (s, 1C), 28.6 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{24}H_{29}N_4O_6^+$ ($[M+H]^+$): 469.2; found: 469.2.

9.4.9 Synthesis of Tripeptide KGpFN



cyanomethyl (*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-fluoro-3-nitrophenyl)diazenyl)-phenyl)propanoate (59):

The azobenzene **47a** (0.83 g, 1.9 mmol, 1.0 eq) and NEt_3 (0.53 mL, 3.8 mmol, 2.0 eq) were dissolved in dry DMF (10 mL) and cooled to 0 °C. The cooled solution was treated with excess chloroacetonitrile (0.97 g, 15 mmol, 8.0 eq) and allowed to warm to ambient temperature. The reaction mixture was then stirred at rt for 3 h until complete consumption of the starting material was observed via TLC. The crude mixture was diluted with aqueous 10% (w/v) LiCl (200 mL), extracted with EtOAc (3 × 50 mL) and the combined organic layers were washed with water (3 × 200 mL), brine (1 × 75 mL), dried over Na_2SO_4, filtered and concentrated. Purification by automated flash column chromatography (12 g silica, cyclohexane/EtOAc gradient, 17:3 → 7:13) to yield the fully protected azobenzene **41** (0.67 g, 1.4 mmol, 74%) as an orange solid.

R_f = 0.47 (cylcohexane:EtOAc, 1:1) [UV, CAM].

^{1}H NMR (400 MHz, $CDCl_3$, 300 K): δ[ppm] = 8.62 (dd, *J* = 7.0, 2.5 Hz, 1H), 8.22 (ddd, *J* = 8.9, 4.1, 2.5 Hz, 1H), 7.92 (d, *J* = 8.4 Hz, 2H), 7.46 (dd, *J* = 10.1, 8.9 Hz, 1H), 7.35 (d, *J* = 8.4 Hz, 2H), 4.97 (d, *J* = 8.3 Hz, 1H), 4.83 (d, *J* = 15.7 Hz, 1H), 4.71 (m, 2H), 3.21 (m, 2H), 1.43 (s, 9H).

^{13}C NMR (101 MHz, $CDCl_3$, 300 K): δ[ppm] = 170.6 (s, 1C), 156.6 (d, $^1J_{C\text{-}F}$ = 269.7 Hz, 1C), 151.4 (s, 1C), 148.6 (d, $^4J_{C\text{-}F}$ = 3.4 Hz, 1C), 140.0 (s, 1C), 130.4 (s, 2C), 130.3 (s, 1C), 129.98 (d, $^3J_{C\text{-}F}$ = 9.2 Hz, 1C), 123.9 (s, 2C), 121.0 (s, 1C) 120.2 (d, $^3J_{C\text{-}F}$ = 2.1 Hz, 1C), 119.2 (d, $^2J_{C\text{-}F}$ = 22.1 Hz, 1C), 113.8 (s, 1C), 80.9 (s, 1C), 54.3 (s, 1C), 49.1 (s, 1C), 38.2 (s, 1C), 28.4 (s, 3C).

HRMS-ESI (m/z): calc. for $C_{22}H_{23}FN_5O_6^+$ ($[M+H]^+$): 472.1627; found: 472.1627.

cyanomethyl (*S,E*)-2-amino-3-(4-((4-fluoro-3-nitrophenyl)diazenyl)phenyl)propanoate (60):

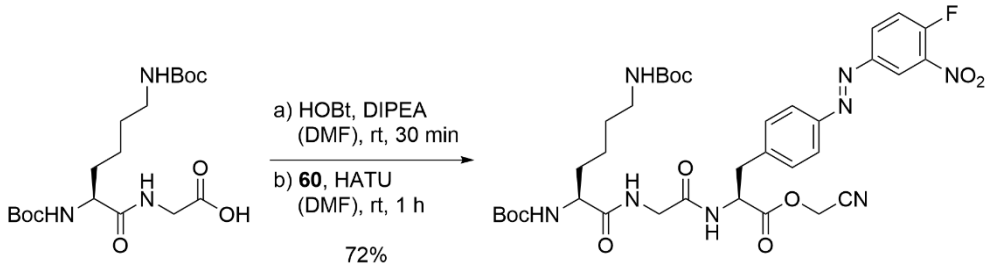
The synthesis of **60** was carried out similar to known literature procedures.[348]

The *N*-Boc protecting group of azobenzene **59** (266 mg, 0.56 mmol, 1.0 eq) was removed by dissolving **59** in dry MeCN (4 mL), cooling the mixture to 0 °C and subsequent addition of 4 M HCl/dioxane (0.56 mL, 2.2 mmol, 4.0 eq). The reaction mixture was then stirred at 15 °C for 4 h until complete consumption of the starting material was observed via TLC. MeCN was removed under reduced pressure (80 mbar) while heating to 30 °C The crude residue was taken up in minimal amounts of MeOH and precipitated in cold Et_2O (max. 2 mL MeOH in 100 mL Et_2O). Centrifugation (4000×g, 4 °C) for 10 min and disposal of the supernatant gave the crude product **60** as HCl salt (198 mg, 0.49 mmol, 87%) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.86 (s, br, 3H), 8.55 (dd, J = 7.0, 2.5 Hz, 1H), 8.35 (ddd, J = 8.9, 4.2, 2.5 Hz, 1H), 7.92 (d, J = 8.4 Hz, 2H), 7.85 (dd, J = 10.9, 8.9 Hz, 1H), 7.56 (d, J = 8.4 Hz, 2H), 5.12 (d, J = 0.6 Hz, 2H), 4.53 (m, 1H), 3.41 – 3.22 (m, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 168.1 (s, 1C), 155.9 (d, $^1J_{C\text{-}F}$ = 266.6 Hz, 1C), 150.8 (s, 1C), 147.8 (d, $^4J_{C\text{-}F}$ = 3.4 Hz, 1C), 139.3 (s, 1C), 137.7 (d, $^3J_{C\text{-}F}$ = 8.8 Hz, 1C), 130.8 (s, 2C), 130.6 (d, $^3J_{C\text{-}F}$ = 9.8 Hz, 1C), 123.2 (s, 2C), 119.9 (d, $^2J_{C\text{-}F}$ = 22.3 Hz, 1C), 119.0 (s, 1C), 115.2 (s, 1C), 52.7 (s, 1C), 50.2 (s, 1C), 35.5 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{17}H_{15}FN_5O_4^+$ ([M+H]$^+$): 372.1103; found: 372.1102.



61 62

***N*^{2},*N*^{6}-bis(*tert*-butoxycarbonyl)-*L*-lysylglycine (61):** Synthesis via SPPS according to Chapter 9.6.

Fmoc-Gly-OH (1.8 g, 6.0 mmol, 2.3 eq) was immobilized on CTC resin (2.0 g, 2.6 mmol, 1.0 eq) using DIPEA (1.8 mL, 10 mmol, 4.0 eq) in DCM (30 mL). After incubation at rt for 1.5 h and on-resin Fmoc-deprotection, the coupling reaction mixture comprising Boc-Lys(Boc)-OH (2.1 g, 3.9 mmol, 1.5 eq), DIC (0.56 mL, 3.6 mmol, 1.4 eq) and HOBt (0.70 g, 5.2 mmol, 2.0 eq) in DMF (30 mL) was applied to the resin. Incubation at rt for 1.5 h and subsequent cleavage from the CTC resin furnished the desired dipeptide **61** (0.58 g, 1.4 mmol, 56%) after purification via flash column chromatography (DCM/MeOH gradient, 49:1 → 23:2 + 1 vol% AcOH).

***R*$_f$** = 0.22 (DCM:MeOH, 23:2 + 1 vol% AcOH) [ninhydrin].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.03 (t, J = 5.8 Hz, 1H), 6.80 (d, J = 8.3 Hz, 1H), 6.74 (s, 1H), 3.94 – 3.86 (m, 1H), 3.73 (m, 2H), 2.92 – 2.81 (m, 2H), 1.65 – 1.54 (m, 1H), 1.54 – 1.43 (m, 1H), 1.43 – 1.16 (m, 22H).

cyanomethyl (10*S*,16*S*)-10-((*tert*-butoxycarbonyl)amino)-16-(4-((*E*)-(4-fluoro-3-nitrophenyl)diazenyl)benzyl)-2,2-dimethyl-4,11,14-trioxo-3-oxa-5,12,15-triazaheptadecan-17-oate (62):

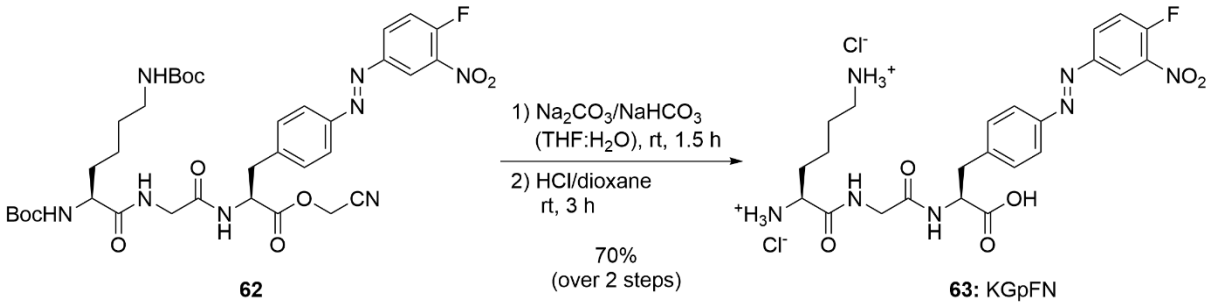
To the dipeptide **61** (0.56 g, 1.3 mmol, 1.05 eq) and HOBt (0.28 g, 1.8 mmol, 1.50 eq) in dry DMF (10 mL) was added DIPEA (0.63 mL, 3.7 mmol, 3.00 eq) and the reaction mixture was stirred at rt for 30 min. The reaction mixture was then added to a solution of **60** (0.49 g, 1.2 mmol, 1.00 eq) in dry DMF (8 mL) before reacting it with HATU (0.69g, 1.8 mmol, 1.50 eq). The reaction mixture was then stirred at rt for 1 h until complete consumption of the starting material was observed via TLC. The crude solution was diluted with aqueous 10% (w/v) LiCl (200 mL), extracted with EtOAc (3 × 50 mL) and the combined organic layers were washed with water (3 × 200 mL), brine (1 × 75 mL), dried over Na_2SO_4 and filtered. The crude tripeptide was concentrated, dry-loaded onto silica and purified by automated flash column chromatography (12 g silica, DCM/MeOH gradient, 99:1 → 9:1) to give the fully protected tripeptide **61** (0.66 g, 0.87 mmol, 72%) as an orange solid.

***R*$_f$** = 0.29 (DCM:MeOH, 19:1) [UV, ninhydrin].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.53 (dd, *J* = 7.0, 2.5 Hz, 1H), 8.45 (d, *J* = 7.6 Hz, 1H), 8.33 (ddd, *J* = 8.9, 4.1, 2.5 Hz, 1H), 8.03 (t, *J* = 5.6 Hz, 1H), 7.88 (d, *J* = 8.5 Hz, 1H), 7.83 (dd, *J* = 10.8, 8.9 Hz, 1H), 7.49 (d, *J* = 8.5 Hz, 2H), 6.86 (d, *J* = 7.9 Hz, 1H), 6.72 (t, *J* = 5.7 Hz, 1H), 5.00 (s, 2H), 4.67 (ddd, *J* = 9.0, 7.6, 5.7 Hz, 1H), 3.87 (td, *J* = 8.4, 4.7 Hz, 1H), 3.71 (d, *J* = 5.6 Hz, 2H), 3.24 – 3.17 (m, 1H), 3.08 (dd, *J* = 13.8, 9.0 Hz, 1H), 2.90 – 2.81 (m, 2H), 1.63 – 1.41 (m, 2H), 1.37 (s, 9H), 1.36 (s, 9H), 1.34 – 1.15 (m, 4H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 172.5 (s, 1C), 170.3 (s, 1C), 169.1 (s, 1C), 155.8 (d, $^1J_{C\text{-}F}$ = 266.1 Hz, 1C), 155.5 (s, 1C), 155.4 (s, 1C), 150.4 (s, 1C), 147.8 (d, $^4J_{C\text{-}F}$ = 3.5 Hz, 1C), 141.7 (s, 1C), 137.6 (d, $^2J_{C\text{-}F}$ = 9.1 Hz, 1C), 130.6 (s, 1C), 130.5 (s, 2C), 123.0 (s, 2C), 119.8 (d, $^2J_{C\text{-}F}$ = 22.4 Hz, 1C), 119.0 (s, 1C), 115.6 (s, 1C), 78.1 (s, 1C), 77.3 (s, 1C), 54.3 (s, 1C), 53.0 (s, 1C), 49.5 (s, 1C), 41.5 (s, 1C), 39.2 (s, 1C), 36.3 (s, 1C), 31.5 (s, 1C), 29.2 (s, 1C), 28.3 (s, 3C), 28.2 (s, 3C), 22.8 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{35}H_{46}FN_8O_{10}^+$ ([M+H]$^+$): 757.3315; found: 757.3312.



62 **63:** KGpFN

(*S*)-2-(2-((*S*)-2,6-diaminohexanamido)acetamido)-3-(4-((*E*)-(4-fluoro-3-nitrophenyl)diazenyl)phenyl)propanoic acid (63, KGpFN): Hydrolysis of the cyanomethyl ester **62** and subsequent *N*-Boc deprotection according to standard procedure H.

To the tripeptide **62** (0.66 g, 0.87 mmol, 1.0 eq) in THF (25 mL) was added a solution of Na_2CO_3 (0.28 g, 2.6 mmol, 3.0 eq) and $NaHCO_3$ (0.15 g, 1.7 mmol, 2.0 eq) in H_2O (25 mL) and the reaction mixture was stirred at rt for 1.5 h until complete hydrolysis of the cyanomethyl ester was observed via TLC. The reaction mixture was acidified to pH 3-4 using aqueous 10% (w/v) KH_2PO_4 and aqueous 1 M HCl and THF was removed under reduced pressure. The crude aqueous mixture was then extracted with EtOAc (3 × 100 mL) and the combined organic layers were washed with aqueous 10% (w/v) KH_2PO_4 (1 × 100 mL) and brine (1 × 100 mL), dried over Na_2SO_4 and filtered. The crude was concentrated, dry-loaded onto silica and purified by automated flash column chromatography (12 g silica, DCM/MeOH gradient, 49:1 → 23:2 + 1 vol% AcOH) to yield the hydrolyzed tripeptide (0.51 g, 0.71 mmol, 81%) as an orange foam.

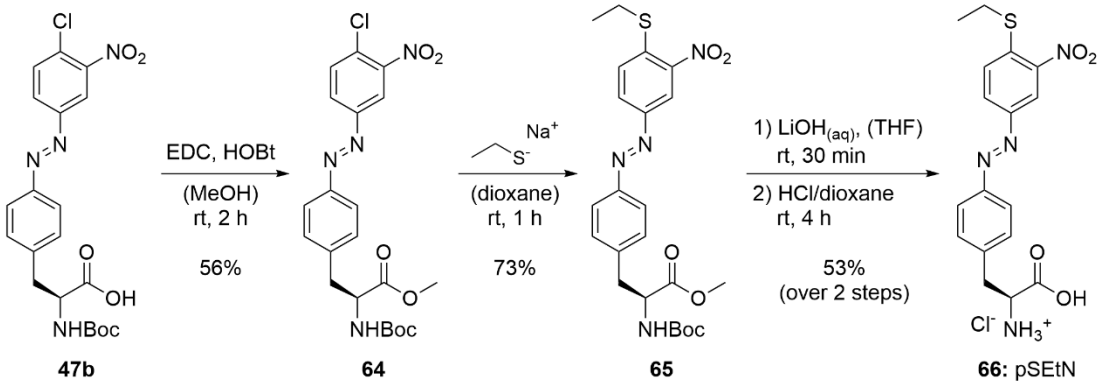
N-Boc deprotection of the hydrolyzed tripeptide by excess 4 M HCl/dioxane (2.1 mL, 8.5 mmol) diluted with dry dioxane (2.1 mL) gave the HCl-salt **63** as an orange solid (0.36 g, 0.61 mmol, 86% or 70% over two steps) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.83 (t, *J* = 5.8 Hz, 1H), 8.52 (dd, *J* = 7.0, 2.5 Hz, 1H), 8.48 (d, *J* = 8.1 Hz, 1H), 8.40 – 8.30 (m, br, 4H), 8.04 (s, br, 3H), 7.88 (d, *J* = 8.4 Hz, 2H), 7.85 – 7.81 (m, 1H), 7.52 (d, *J* = 8.4 Hz, 1H), 4.59 – 4.50 (m, 1H), 3.85 – 3.76 (m, 3H), 3.24 – 3.17 (m, 1H), 3.06 (dd, *J* = 13.7, 9.3 Hz, 1H), 2.72 (d, *J* = 7.1 Hz, 2H), 1.73 (q, *J* = 7.3 Hz, 2H), 1.56 (p, *J* = 7.5 Hz, 2H), 1.37 (t, *J* = 11.8 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 172.5 (s, 1C), 168.8 (s, 1C), 168.2 (s, 1C), 155.8 (d, $^{1}J_{C\text{-}F}$ = 266.4 Hz, 1C), 150.7 (s, 1C), 147.8 (d, $^{4}J_{C\text{-}F}$ = 3.4 Hz, 1C), 142.8 (s, 1C), 137.6 (d, $^{2}J_{C\text{-}F}$ = 9.0 Hz, 1C), 130.6 (s, 1C), 130.5 (s, 2C), 122.9 (s, 2C), 119.9 (d, $^{2}J_{C\text{-}F}$ = 22.3 Hz, 1C), 118.9 (s, 1C), 53.3 (s, 1C), 51.8 (s, 1C), 41.6 (s, 1C), 38.2 (s, 1C), 36.6 (s, 1C), 30.2 (s, 1C), 26.2 (s, 1C), 21.0 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{23}H_{29}FN_7O_6^+$ ([M+H]$^+$): 518.2158; found: 518.2159.

9.4.10 Synthesis of Model Azo-ncAA pSEtN



methyl (*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-chloro-3-nitrophenyl)diazenyl)-phenyl)propanoate (64): Methylation of **47b** by standard procedure C.

To the phenolic carboxylic acid **47b** (260 mg, 0.58 mmol, 1.0 eq) were added dry MeOH until complete dissolution (15 mL), HOBt (94.0 mg, 0.69 mmol, 1.2 eq) and EDC×HCl (133 mg, 0.69 mmol, 1.2 eq).and the reaction mixture was stirred o.n. at rt. Purification via flash column chromatography over silica gel (cyclohexane/EtOAc gradient, 5:1 → 2.5:1) furnished **47b** as an orange solid (150 mg, 0.32 mmol, 56%).

R_f = 0.52 (cyclohexane:EtOAc, 2:1) [UV, CAM].

^{1}H NMR (400 MHz, $CDCl_3$, 300 K): δ[ppm] = 8.40 (d, *J* = 2.3 Hz, 1H), 8.08 (dd, *J* = 8.6, 2.3 Hz, 1H), 7.89 (d, *J* = 8.4 Hz, 1H), 7.70 (d, *J* = 8.6 Hz, 1H), 7.32 (d, *J* = 8.4 Hz, 2H), 5.04 (d, *J* = 8.2 Hz, 1H), 4.65 (d, *J* = 7.3 Hz, 1H), 3.74 (s, 3H), 3.29 – 3.07 (m, 2H), 1.43 (s, 9H).

HRMS-ESI (m/z): calc. for $C_{21}H_{24}ClN_4O_6^+$ ([M+H]$^+$): 463.1379; found: 463.1378.

methyl (*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((4-(ethylthio)-3-nitrophenyl)diazenyl)-phenyl)propanoate (65):

To the azobenzene **64** (150 mg, 0.32 mmol, 1.0 eq) in dry 1,4-dioxane (10 mL) was added sodium ethanethiolate (61.0 mg, 0.65 mmol, 2.0 eq) and the reaction mixture stirred at rt for 1 h until complete consumption of the starting material was observed via TLC. The crude mixture was diluted with water (5 mL) and excess sodium ethanethiolate was quenched with aqueous 5% (v/v) H_2O_2 (2 mL). The organic solvent was then removed under reduced pressure and the aqueous mixture extracted with DCM (3 × 20 mL). The combined organic layers were washed with water (1 × 20 mL) and brine (1 × 20 mL), dried over Na_2SO_4, filtered and concentrated. Purification by flash column chromatography on silica (cyclohexane/EtOAc gradient, 5:1 → 2.5:1) yielded the desired thiolated azobenzene **65** (114 mg, 0.23 mmol, 73%) as an orange solid.

R_f = 0.41 (cyclohexane:EtOAc, 2.5:1) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.60 (d, *J* = 2.2 Hz, 1H), 8.20 (dd, *J* = 8.8, 2.2 Hz, 1H), 7.88 (d, *J* = 8.4 Hz, 2H), 7.85 (d, *J* = 8.8 Hz, 1H), 7.48 (d, *J* = 8.4 Hz, 3H), 7.39 (d, *J* = 8.2 Hz, 1H), 4.27 (m, 1H), 3.63 (s, 3H), 3.18 (q, *J* = 7.4 Hz, 2H), 3.12 (dd, *J* = 13.7, 5.2 Hz, 1H), 2.97 (dd, *J* = 13.7, 10.1 Hz, 1H), 1.37 – 1.30 (m, 12H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 172.4 (s, 1C), 155.4 (s, 1C), 150.4 (s, 1C), 148.2 (s, 1C), 145.8 (s, 1C), 142.5 (s, 1C), 140.1 (s, 1C), 130.3 (s, 2C), 128.1 (s, 1C), 127.3 (s, 1C), 122.8 (s, 2C), 119.3 (s, 1C), 78.3 (s, 1C), 54.8 (s, 1C), 51.9 (s, 1C), 36.3 (s, 1C), 28.1 (s, 3C), 25.7 (s, 1C), 12.8 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{23}H_{29}N_4O_6S^+$ ([M+H]$^+$): 489.1802; found: 489.1800.

(*S,E*)-2-amino-3-(4-((4-(ethylthio)-3-nitrophenyl)diazenyl)phenyl)propanoic acid (66, pSEtN): Hydrolysis of the methyl ester **65** by standard procedure F and subsequent *N*-Boc deprotection according to standard procedure H.

To **65** (49 mg, 0.1 mmol, 1.0 eq) were added THF (0.2 mL) and 1 M aqueous LiOH (0.2 mL, 0.2 mmol, 2.0 eq) and the reaction mixture was stirred at rt for 30 min until complete hydrolysis of the methyl ester was observed via TLC. The hydrolyzed methyl ester was obtained (41 mg, 0.09 mmol, 86%) after purification via automated flash column chromatography (4 g silica, cyclohexane/EtOAc gradient, 4:1 → 1:1).

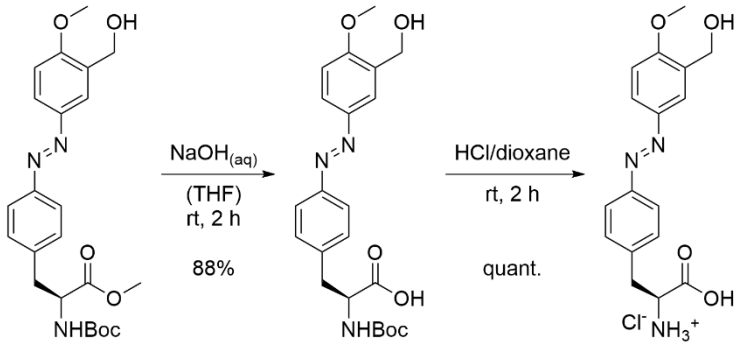
Subsequent deprotection employing 4 M HCl/dioxane (0.13 mL, 0.52 mmol, 8.0 eq) diluted with dry dioxane (0.13 mL) gave the HCl-salt **66** as an orange solid (22 mg, 0.05 mmol, 62% or 53% over two steps) which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 8.61 (d, *J* = 2.2 Hz, 1H), 8.38 (d, br, *J* = 5.3 Hz, 3H), 8.20 (dd, *J* = 8.8, 2.2 Hz, 1H), 7.92 (d, *J* = 8.4 Hz, 2H), 7.85 (d, *J* = 8.8 Hz, 1H), 7.52 (d, *J* = 8.4 Hz, 2H), 4.34 – 4.23 (m, 1H), 3.28 – 3.13 (m, 4H), 1.34 (t, *J* = 7.3 Hz, 3H).

^{13}C NMR (101 MHz, DMSO-d_6 + TFA, 300 K): δ[ppm] = 170.4 (s, 1C), 151.0 (s, 1C), 148.2 (s, 1C), 145.9 (s, 1C), 140.4 (s, 1C), 139.6 (s, 1C), 130.8 (s, 2C), 128.2 (s, 1C), 127.4 (s, 1C), 123.1 (s, 2C), 119.4 (s, 1C), 53.0 (s, 1C), 35.7 (s, 1C), 25.8 (s, 1C), 12.8 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{17}H_{19}N_4O_4S^+$ ([M+H]$^+$): 375.1122; found: 375.1127.

9.4.11 Synthesis of Non-Crosslinking Azo-ncAA m1OH



(*S,E*)-2-((*tert*-butoxycarbonyl)amino)-3-(4-((3-(hydroxymethyl)-4-methoxyphenyl)-diazenyl)phenyl)propanoic acid (67): Hydrolysis of the methyl ester **21a** by standard procedure F.

To **21a** (0.93 g, 2.1 mmol, 1.0 eq) were added THF (9.0 mL) and 1 M aqueous NaOH (8.5 mL, 8.5 mmol, 4.1 eq) and the reaction mixture was stirred at rt for 2 h until complete consumption of the methyl ester was observed via TLC. The desired azobenzene **67** was obtained as an orange foam (0.79 g, 1.84 mmol, 88%) after purification via automated flash column chromatography (12 g silica, cyclohexane/EtOAc gradient, 7:3 → 2:3 + 1 vol% AcOH)

R_f = 0.22 (cyclohexane:EtOAc, 1:5+ 1 vol% AcOH) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] =7.97 (dt, *J* = 2.4, 1.1 Hz, 1H), 7.84 (dd, *J* = 8.6, 2.4 Hz, 1H), 7.78 (d, *J* = 8.4 Hz, 2H), 7.44 (d, *J* = 8.4 Hz, 2H), 7.20 – 7.14 (m, 2H), 5.24 (t, *J* = 5.7 Hz, 1H), 4.55 (d, *J* = 3.9 Hz, 2H), 4.16 (ddd, *J* = 10.3, 8.4, 4.6 Hz, 1H), 3.89 (s, 3H), 3.18 – 3.08 (m, 1H), 2.92 (dd, *J* = 13.8, 10.3 Hz, 1H), 1.32 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 173.4 (s, 1C), 158.6 (s, 1C), 155.5 (s, 1C), 150.8 (s, 1C), 145.9 (s, 1C), 141.2 (s, 1C), 131.7 (s, 1C), 130.1 (s, 2C), 125.1 (s, 1C), 122.0 (s, 2C), 119.0 (s, 1C), 110.5 (s, 1C), 78.1 (s, 1C), 57.7 (s, 1C), 55.8 (s, 1C), 54.9 (s, 1C), 36.3 (s, 1C), 28.1(s, 3C).

HRMS-ESI (m/z): calc. for $C_{22}H_{28}N_3O_6^+$ ([M+H]$^+$): 430.1973; found: 430.1967.

(*S,E*)-2-amino-3-(4-((3-(hydroxymethyl)-4-methoxyphenyl)diazenyl)phenyl)propanoic acid (68,m1OH): *N*-Boc deprotection of **67** according to standard procedure H.

Deprotection of **67** (0.79 g, 1.84 mmol, 1.0 eq) employing 4 M HCl/dioxane (2.75 mL, 11 mmol, 6.0 eq) diluted with dry dioxane (3.25 mL) furnished the HCl-salt **68** as an orange solid (0.66 g, 1.80 mmol, quant.) which was used without further purification.

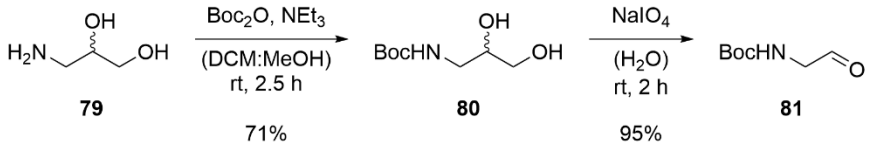
^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.44 (s, br, 3H), 7.98 (dt, *J* = 2.2, 1.0 Hz, 1H), 7.88 – 7.79 (m, 3H), 7.48 (d, *J* = 8.4 Hz, 2H), 7.17 (d, *J* = 8.8 Hz, 1H), 5.28 (s, br, 1H), 4.55 (s, 2H), 4.24 (t, *J* = 6.4 Hz, 1H), 3.89 (s, 3H), 3.22 (d, *J* = 6.4 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 170.3 (s, 1C), 158.8 (s, 1C), 151.4 (s, 1C), 145.9 (s, 1C), 137.8 (s, 1C), 131.7 (s, 1C), 130.6 (s, 2C), 125.2 (s, 1C), 122.4 (s, 2C), 119.1 (s, 1C), 110.5 (s, 1C), 57.7 (s, 1C), 55.8 (s, 1C), 53.0 (s, 1C), 35.5 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{17}H_{20}N_3O_4^+$ ([M+H]$^+$): 330.1448; found: 330.1444.

9.5 Synthetic Procedures of Electrophilic Warheads and Fluorogenic Probes

9.5.1 Synthesis of Moderately Electrophilic Warhead G-VME



***tert*-butyl (2,3-dihydroxypropyl)carbamate (80):**

To commercially available (±)-3-amino-1,2-propanediol (**79**) (10.2 g, 112 mmol, 1.0 eq) and NEt_3 (1.87 mL, 13.4 mmol, 12 mol-%) dissolved in DCM (100 mL) and MeOH (20 mL)was added a solution of Boc_2O (29.3 g, 134 mmol, 1.2 eq) in DCM (165 mL). The reaction mixture was stirred at rt for 2.5 h until complete consumption of the starting material was observed via TLC. Then, the organic solvent was removed under reduced pressure and the crude resuspended in aqueous 10% (w/v) $KHSO_4$ (200 mL). The aqueous mixture was extracted with EtOAc (6 × 100 mL) and the combined organic layers washed with aqueous 10% (w/v) $KHSO_4$ (1 × 200 mL) and brine (1 × 200 mL), dried over Na_2SO_4, filtered and concentrated. Purification via flash column chromatography over silica gel (DCM/MeOH gradient, 97:3 → 9:1) furnished **80** as a white solid (15.2 g, 80.0 mmol, 71%).

R_f = 0.10 (DCM:EtOAc, 1:1) [ninhydrin].

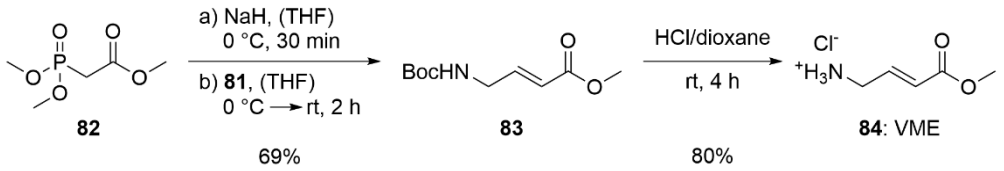
^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] 6.57 (s, 1H), 4.60 (d, *J* = 5.0 Hz, 1H), 4.45 (t, *J* = 5.8 Hz, 1H), 3.49 – 3.39 (m, 2H), 3.27 (ddd, *J* = 5.8, 4.2, 1.0 Hz, 2H), 3.08 – 2.98 (m, 1H), 2.84 (ddd, *J* = 13.4, 6.7, 5.5 Hz, 1H), 1.37 (s, 9H).

***tert*-butyl (2-oxoethyl)carbamate (81):**

For oxidative cleavage, the vicinal diol **80** (1.0 g, 5.2 mmol, 1.0 eq) was dissolved in H_2O (9 mL) and solid $NaIO_4$ (1.3 g, 6.3 mmol, 1.2 eq) was added. The reaction mixture was subsequently stirred at rt for 2 h in the dark before being extracted with DCM (6 × 30 mL). The combined organic layers were washed with brine (1 × 60 mL), dried over Na_2SO_4, filtered and concentrated to yield **81** (0.79 g, 4.9 mmol, 95%) as colorless oil that was stored at -20 °C until further use.

R_f = 0.47 (Pent:EtOAc, 1:1) [ninhydrin].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 9.45 (s, 1H), 7.19 (t, *J* = 5.6 Hz, 1H), 3.73 (d, *J* = 5.6 Hz, 2H), 1.39 (s, 9H).



methyl (*E*)-4-((*tert*-butoxycarbonyl)amino)but-2-enoate (83): The HWE reaction was performed according to standard procedure K.

The commercially available olefination reagent **82** (2.4 mL, 14.7 mmol, 1.15 eq) was reacted with NaH (0.59 g, 14.7 mmol, 1.15 eq) in dry THF (65 mL), stirring at 0 °C for 30 min. Then, a solution of aldehyde **81** (2.0 g, 12.8 mmol, 1.00 eq) in dry THF (50 mL) was added and the reaction mixture stirred at 0° C for 30 min before being warmed to ambient temperature.

After stirring at rt for 1.5 h, the crude product was purified via flash column chromatography over silica gel (DCM/EtOAc gradient, 49:1 → 23:2) to exclusively yield the *E*-isomer of the α,β-unsaturated ester **83** as a colorless oil (1.9 g, 8.84 mmol, 69%).

R_f = 0.39 (DCM:EtOAc, 10:1) [UV, ninhydrin].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.17 (t, *J* = 5.8 Hz, 1H), 6.82 (dt, *J* = 15.7, 4.7 Hz, 1H), 5.85 (dt, *J* = 15.7, 2.0 Hz, 1H), 3.75 – 3.70 (m, 2H), 3.65 (s, 3H), 1.38 (s, 9H).

^{13}C NMR (101 MHz, $CDCl_3$, 300 K): δ[ppm] = 166.7 (s, 1C), 155.7 (s, 1C), 145.2 (s, 1C), 121.1 (s, 1C), 80.1 (s, 1C), 51.8 (s, 1C), 41.5 (s, 1C), 28.5 (s, 3C).

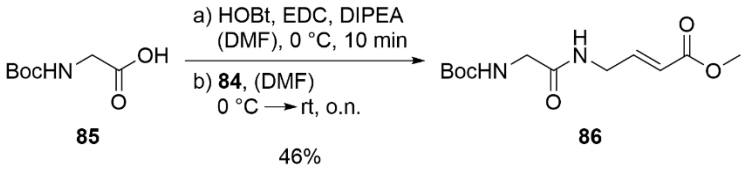
LRMS-ESI (m/z): calc. for $C_{10}H_{17}NNaO_4^+$ ([M+Na]$^+$): 238.1; found: 238.1.

methyl (*E*)-4-aminobut-2-enoate (84, VME): *N*-Boc deprotection of **83** according to standard procedure H.

Deprotection of **83** (0.48 g, 2.22 mmol, 1.0 eq) employing 4 M HCl/dioxane (3.3 mL, 13.3 mmol, 6.0 eq) diluted with dry dioxane (2.2 mL) furnished the HCl-salt **84** after stirring at rt for 2.5 h as an orange solid (0.27 g, 1.79 mmol, 80%), which was used without further purification.

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.49 (s, br, 3H), 6.87 (dt, *J* = 15.9, 5.7 Hz, 1H), 6.18 (dt, *J* = 15.9, 1.8 Hz, 1H), 3.69 (s, 3H), 3.66 (dd, *J* = 5.7, 1.8 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 165.4 (s, 1C), 140.7 (s, 1C), 123.4 (s, 1C), 51.7 (s, 1C), 39.1 (s, 1C).



methyl (*E*)-4-(2-((*tert*-butoxycarbonyl)amino)acetamido)but-2-enoate (86): Amide coupling of **85** and **84** by standard procedure L.

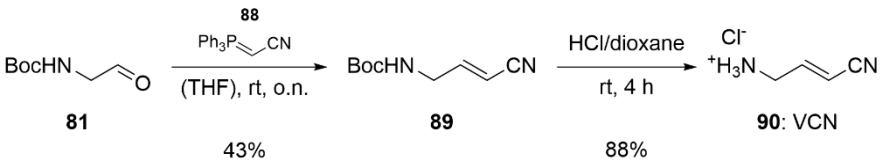
Commercially available *N*-Boc glycine (**85**) (1.29 g, 7.38 mmol, 1.0 eq) was reacted with EDC×HCl (1.56 g, 8.11 mmol, 1.1 eq), HOBt (1.24 g, 8.11 mmol, 1.1 eq) and DIPEA (2.63 mL, 15.5 mmol, 2.1 eq) in dry DMF (17 mL), stirring at 0 °C for 10 min. Next, a solution of **84** (1.12 g, 7.38 mmol, 1.0 eq) in dry DMF (10 mL) was added and the reaction mixture stirred o.n. at rt. The crude product was purified via flash column chromatography over silica gel (DCM/EtOAc gradient, 13:7 → 2:3) to furnish the desired product **86** as a colorless oil (931 mg, 3.42 mmol, 46%).

R_f = 0.16 (DCM:EtOAc, 1:1) [UV, ninhydrin].

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.08 (t, *J* = 5.9 Hz, 1H), 6.99 (t, *J* = 6.1 Hz, 1H), 6.83 (dt, *J* = 15.8, 4.6 Hz, 1H), 5.90 (dt, *J* = 15.8, 2.0 Hz, 1H), 3.87 (ddd, *J* = 5.9, 4.6, 2.0 Hz, 2H), 3.65 (s, 3H), 3.55 (d, *J* = 6.1 Hz, 2H), 1.39 (s, 9H).

^{13}C NMR (75 MHz, DMSO-d_6, 300 K): δ[ppm] = 169.5 (s, 1C), 166.0 (s, 1C), 155.8 (s, 1C), 146.2 (s, 1C), 119.8 (s, 1C), 78.1 (s, 1C), 51.3 (s, 1C), 43.3 (s, 1C), 39.2 (s, 1C), 28.2 (s, 3C).

9.5.2 Synthesis of Moderately Electrophilic Warhead G-VCN



***tert*-butyl (*E*)-(3-cyanoallyl)carbamate (89):**

To the commercially available Wittig reagent **88** (1.3 g, 4.3 mmol, 1.5 eq) in dry THF (20 mL) was added *N*-Boc glycinal (**81**) (0.46 g, 2.9 mmol, 1.0 eq) and the reaction mixture was stirred o.n. at rt until complete conversion was observed via TLC. The mixture was diluted with water and THF evaporated under reduced pressure. The aqueous mix was subsequently extracted with EtOAc (3 × 60 mL) and the combined organic layers washed with a saturated solution of aqueous $NaHCO_3$ (1 × 60 mL), aqueous 10% (w/v) $KHSO_4$ (1 × 60 mL) and brine (1 × 60 mL). The crude product was dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:1) yielded a mixture of the *E*- and *Z*-isomer (3:1) of **89** as a colorless oil (0.23 g, 1.3 mmol, 43%).

R_f = 0.54 (Pent:EtOAc, 2.5:1) [UV, ninhydrin].

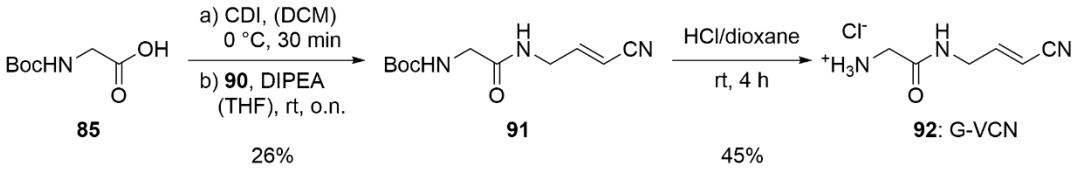
^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.18 (d, *J* = 6.1 Hz, 1H), 6.83 (dt, *J* = 16.4, 4.6 Hz, 1H), 5.64 (dt, *J* = 16.4, 2.1 Hz, 1H), 3.76 – 3.69 (m, 2H), 1.38 (s, 9H).

(*E*)-4-aminobut-2-enenitrile (90, VCN): *N*-Boc deprotection of **89** according to standard procedure H.

Deprotection of **89** (0.98g, 5.4 mmol, 1.0 eq) employing 4 M HCl/dioxane (5.0 mL, 20 mmol, 3.7 eq) diluted with dry dioxane (5.0 mL) furnished the HCl-salt **90** after stirring at rt for 5 h as a beige-brown solid (0.56 g, 4.7 mmol, 88%), which was used without further purification.

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.50 (s, br, 3H), 6.86 (dt, *J* = 16.5, 5.8 Hz, 1H), 6.06 (dt, *J* = 16.5, 1.8 Hz, 1H), 3.66 (dd, *J* = 5.8, 1.8 Hz, 2H).

Note that the product is extremely labile towards nucleophiles and was obtained as a mixture of the different isomers and the chloro-adduct, *E*:*Z*:adduct (3:1:1).



***tert*-butyl (*E*)-(2-((3-cyanoallyl)amino)-2-oxoethyl)carbamate (91):**

Commercially available *N*-Boc glycine (**85**) (0.40 g, 2.3 mmol, 1.0 eq) was dissolved in dry DCM (10 mL) and cooled to 0 °C using an ice bath. Then CDI (0.37 g, 2.8 mmol, 1.2 eq) was added and the reaction mixture stirred at 0 °C for 30 min. The HCl-salt **90** (0.27 g, 2.3 mmol, 1.0 eq) and DIPEA (0.78 mL, 4.6 mmol, 2.0 eq) were then suspended in a 3:2-mixture of dry THF:DCM (5.0 mL) and subsequently added to the cooled solution. The dark brown reaction mixture was allowed to warm to ambient and stirred o.n. at rt.

Next, the organic solvent was removed under reduced pressure and the crude resuspended in water (20 mL) and EtOAc (20 mL). The organic phase was separated and the aqueous layer extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with aqueous 10% (w/v) $KHSO_4$ (1 × 25 mL) and brine (1 × 25 mL), dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica. Purification via flash column chromatography over silica gel (DCM/MeOH gradient, 49:1 → 93:7) furnished **91** as a colorless solid (0.14 g, 0.60 mmol, 26%).

R_f = 0.14 (DCM:EtOAc, 7:3) [UV, ninhydrin].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.13 (t, *J* = 5.9 Hz, 1H), 7.04 (t, *J* = 6.0 Hz, 1H), 6.86 (dt, *J* = 16.3, 4.2 Hz, 1H), 5.69 (dt, *J* = 16.3, 2.2 Hz, 1H), 3.86 (ddd, *J* = 5.9, 4.2, 2.2 Hz, 2H), 3.54 (d, *J* = 6.0 Hz, 2H), 1.39 (s, 9H).

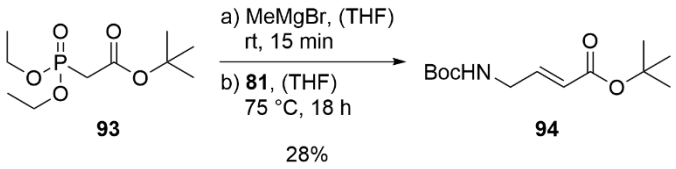
^{13}C NMR (75 MHz, DMSO-d_6, 300 K): δ[ppm] = 169.7 (s, 1C), 155.9 (s, 1C), 153.2 (s, 1C), 117.7 (s, 1C), 98.5 (s, 1C), 78.2 (s, 1C), 43.4 (s, 1C), 39.9 (s, 1C), 28.1 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{11}H_{17}N_3NaO_3^+$ ([M+Na]$^+$): 262.1; found: 262.1.

(*E*)-2-amino-N-(3-cyanoallyl)acetamide (92, G-VCN): *N*-Boc deprotection of **91** according to standard procedure H.

Deprotection of **83** (144 mg, 0.60 mmol, 1.0 eq) employing 4 M HCl/dioxane (0.540 mL, 2.2 mmol, 3.7 eq) diluted with dry dioxane (1 mL) furnished the HCl-salt **91** after stirring at rt for 5 h. Since the product is extremely labile towards nucleophiles, a mixture containing the chloro-adduct (> 50%) was obtained and thus no yield was determined.

9.5.3 Synthesis Towards Moderately Electrophilic Warhead G-VtBuE



***tert*-butyl (*E*)-4-((*tert*-butoxycarbonyl)amino)but-2-enoate (94):**

The synthesis of **94** was carried out similar to known literature procedures.[332]

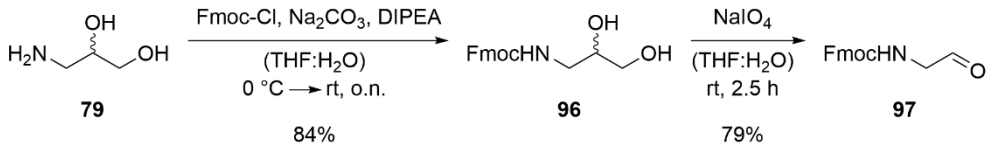
To the commercially available olefination reagent **93** (1.2 mL, 5.14 mmol, 1.15 eq) in dry THF (30 mL) was added a 3 M solution of MeMgBr in Et_2O (1.8 mL, 5.36 mmol, 1.20 eq). The clear solution was stirred at rt for 15 min before adding a solution of *N*-Boc glycinal (**81**) (0.72 g, 4.47 mmol, 1.0 eq) dissolved in dry THF (5 mL). The reaction mixture was then refluxed at 74 °C for 18 h until complete conversion was observed via TLC. The mixture was diluted with water and THF evaporated under reduced pressure. The aqueous mix was subsequently extracted with Et_2O (3 × 40 mL) and the combined organic layers washed with a saturated solution of aqueous $NaHCO_3$ (1 × 40 mL), aqueous 10 % (w/v) $KHSO_4$ (1 × 40 mL) and brine (1 × 40 mL). The crude product was dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica. Purification via flash column chromatography over silica gel (DCM/EtOAc gradient, 99:1 → 47:3) yielded the desired product **94** as a colorless oil (0.32 g, 1.25 mmol, 28 %).

R_f = 0.54 (DCM:EtOAc, 9:1) [UV, ninhydrin].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.14 (t, *J* = 6.0 Hz, 1H), 6.70 (dt, *J* = 15.7, 4.7 Hz, 1H), 5.73 (dt, *J* = 15.7, 2.0 Hz, 1H), 3.74 – 3.66 (m, 2H), 1.43 (s, 9H), 1.38 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{13}H_{234}NNaO_4^+$ ([M+Na]$^+$): 280.2; found: 280.1.

Note that subsequent, selective *N*-Boc deprotection of **83** using dry HCl in dioxane according to known literature procedures[334] resulted in unselective removal the *N*-Boc protecting group and the *tert*-butyl ester.



(9H-fluoren-9-yl)methyl (2,3-dihydroxypropyl)carbamate (96):

To commercially available (±)-3-amino-1,2-propanediol (**79**) (2.1 g, 22 mmol, 1.0 eq) and DIPEA (3.9 mL, 22 mmol, 1.0 eq) dissolved in THF (55 mL) was added a solution of Na_2CO_3 (2.3 g, 22 mmol, 1.0 eq) in water (55 mL). The white suspension was then cooled to 0 °C before adding solid Fmoc-Cl (6.0 g, 23 mmol, 1.05 mmol) portion wise. The reaction mixture was allowed to warm to ambient and stirred o.n. at rt. Then, the organic solvent was removed under reduced pressure and the aqueous mixture was extracted with EtOAc (3 × 100 mL). The combined organic layers washed with aqueous 10% (w/v) $KHSO_4$ (1 × 100 mL) and brine (1 × 100 mL), dried over Na_2SO_4, filtered and concentrated. Purification via flash column chromatography over silica gel (DCM/EtOAc gradient, 7:3 → 3:7) furnished **96** as a white solid (5.8 g, 18 mmol, 84%).

R_f = 0.19 (DCM:EtOAc, 3:7) [CAM, UV].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] 7.89 (dt, *J* = 7.5, 1.0 Hz, 2H), 7.71 (d, *J* = 7.5 Hz, 2H), 7.42 (td, *J* = 7.5, 1.0 Hz, 2H), 7.33 (td, *J* = 7.5, 1.2 Hz, 2H), 7.17 (t, *J* = 5.7 Hz, 1H), 4.66 (d, *J* = 5.0 Hz, 1H), 4.47 (t, *J* = 5.7 Hz, 1H), 4.27 (d, *J* = 7.7 Hz, 2H), 4.20 (t, *J* = 7.0 Hz, 1H), 3.52 – 3.45 (m, 1H), 3.29 (td, *J* = 5.7, 2.7 Hz, 2H), 3.11 (dt, *J* = 13.3, 5.7 Hz, 1H), 2.92 (ddd, *J* = 13.3, 6.8, 5.7 Hz, 1H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 158.2 (s, 1C), 142.6 (s, 1C), 139.4 (s, 2C), 137.4 (s, 2C), 128.9 (s, 2C), 127.3 (s, 2C), 121.4 (s, 2C), 120.1 (s, 2C), 109.8 (s, 1C), 71.0 (s, 1C), 63.7 (s, 1C), 43.8 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{18}H_{20}NO_4^+$ ([M+H]$^+$): 314.1; found: 314.1.

(9H-fluoren-9-yl)methyl (2-oxoethyl)carbamate (97):

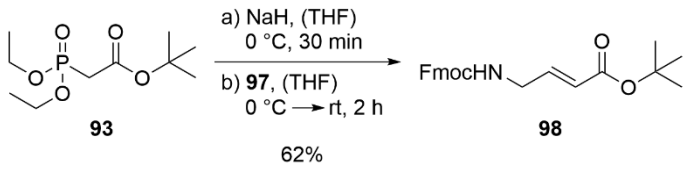
For oxidative cleavage, the vicinal diol **96** (2.0 g, 6.4 mmol, 1.0 eq) was dissolved in THF (35 mL) and a solution of $NaIO_4$ (1.6 g, 7.7 mmol, 1.2 eq) in water (18 mL) was added. The reaction mixture was subsequently stirred at rt for 2.5 h in the dark before being extracted with DCM (3 × 100 mL). The combined organic layers were washed with brine (1 × 100 mL), dried over Na_2SO_4, filtered and concentrated. The crude product was subsequently dry-loaded onto silica and purified via flash column chromatography over silica gel (Pent/EtOAc gradient, 1:1 → 1:2.5) to yield **97** (1.4 g, 5.0 mmol, 79%) as colorless oil which was stored at -20 °C until further use.

R_f = 0.34 (DCM:EtOAc, 9:1) [UV, CAM].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 9.46 (s, 1H), 7.90 (d, *J* = 7.5 Hz, 2H), 7.70 (m, 3H), 7.42 (t, *J* = 7.5 Hz, 2H), 7.34 (td, *J* = 7.4, 1.2 Hz, 2H), 4.34 (d, *J* = 6.9 Hz, 2H), 4.24 (t, *J* = 6.9 Hz, 1H), 3.83 (d, *J* = 5.7 Hz, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 200.2 (s, 1C), 156.6 (s, 1C), 143.8 (s, 2C), 140.7 (s, 2C), 127.6 (s, 2C), 127.1 (s, 2C), 125.2 (s, 2C), 120.1 (s, 2C), 65.7 (s, 1C), 50.5 (s, 1C), 46.6 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{19}H_{18}N_2O_3$ ([M+MeCN]): 322.1; found: 322.1



***tert*-butyl (E)-4-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)but-2-enoate (98):** The HWE reaction was performed according to standard procedure K.

The commercially available olefination reagent **93** (1.4 mL, 6.1 mmol, 1.20 eq) was reacted with NaH (0.23 g, 5.8 mmol, 1.15 eq) in dry THF (27 mL), stirring at 0 °C for 30 min. Then, a solution of aldehyde **97** (1.4 g, 5.0 mmol, 1.0 eq) in dry THF (20 mL) was added and the reaction mixture stirred at 0° C for 30 min before being warmed to ambient temperature. After stirring at rt for 1.5 h, the crude product was purified via flash column chromatography over silica gel (DCM/EtOAc gradient, 49:1 → 23:2) to yield the desired *E*-isomer of **98** as a colorless oil (1.2 g, 3.1 mmol, 62%).

R_f = 0.48 (DCM:EtOAc, 23:2) [UV, ninhydrin].

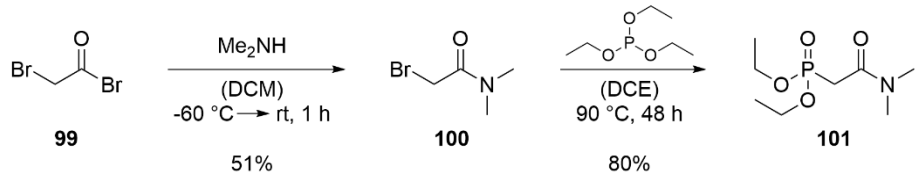
^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.89 (dd, *J* = 7.5, 3.9 Hz, 2H), 7.70 (d, *J* = 7.4 Hz, 2H), 7.63 (t, *J* = 5.8 Hz, 1H), 7.42 (t, *J* = 7.5 Hz, 2H), 7.38 – 7.30 (m, 2H), 6.72 (dt, *J* = 15.8, 4.6 Hz, 1H), 5.75 (d, *J* = 15.8 Hz, 1H), 4.33 (d, *J* = 6.9 Hz, 2H), 4.23 (t, *J* = 6.9 Hz, 1H), 3.81 – 3.73 (m, 2H), 1.43 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 164.8 (s, 1C), 156.1 (s, 1C), 145.1 (s, 1C), 143.8 (s, 2C), 140.7 (s, 2C), 127.6 (s, 2C), 127.1 (s, 2C), 125.1 (s, 2C), 121.7 (s, 1C), 120.2 (s, 2C), 79.8 (s, 1C), 65.5 (s, 1C), 46.7 (s, 1C), 40.9 (s, 1C), 27.8 (s, 3C).

LRMS-ESI (m/z): calc. for $C_{23}H_{25}NNaO_4^+$ ([M+Na]$^+$): 402.2; found: 402.2

Note that subsequent *N*-Fmoc deprotection of **98** using DBU or piperidine yielded complex, inseparable product mixtures, presumably due to Michael addition of nucleophilic amine bases. The use of aqueous NaOH in THF resulted in unselective deprotection of both *N*-Fmoc and the *tert*-butyl ester.

9.5.4 Synthesis of Moderately Electrophilic Warhead G-V(D)MA



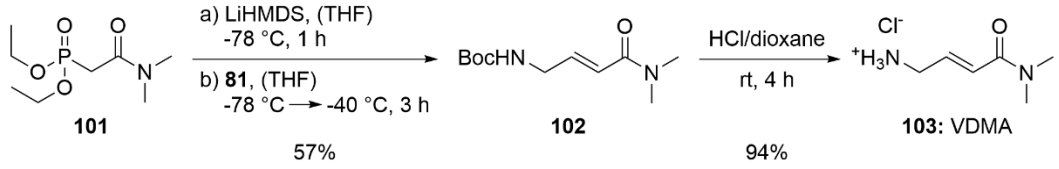
2-bromo-*N,N*-dimethylacetamide (100):

Commercially available bromoacetyl bromide (**99**) (1.00 mL, 13.5 mmol, 1.0 eq) was dissolved in dry DCM (15 mL) and cooled to -60 °C using a cooling bath (EtOAc in liquid N_2). To this, a 2 M solution of Me_2NH in dry THF (13.5 mL, 27.0 mmol, 2.0 eq) was added dropwise and the reaction subsequently stirred for 1 h while being allowed to warm to ambient temperature. The mixture was then quenched with water (10 mL) and the organic solvents removed under reduced pressure. The crude aqueous mixture was extracted with DCM (4 × 30 mL) and the combined organic layers washed with water (1 × 50 mL), a saturated solution of aqueous $NaHCO_3$ (1 × 50 mL) and brine (1 × 50 mL), dried over Na_2SO_4, filtered and concentrated. Purification by passing the crude through a short plug of silica (Pent/EtOAc gradient, 1:1 → 1:2.5) furnished **100** as a solid (1.14 g, 6.86 mmol, 51%).

R_f = 0.31 (Pent:EtOAc, 1:3) [ninhydrin].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 4.10 (s, 2H), 3.02 (s, 3H), 2.84 (s, 3H)
LRMS-ESI (m/z): calc. for $C_4H_9BrNO^+$ ([M+H]$^+$): 166.0; found: 166.1.

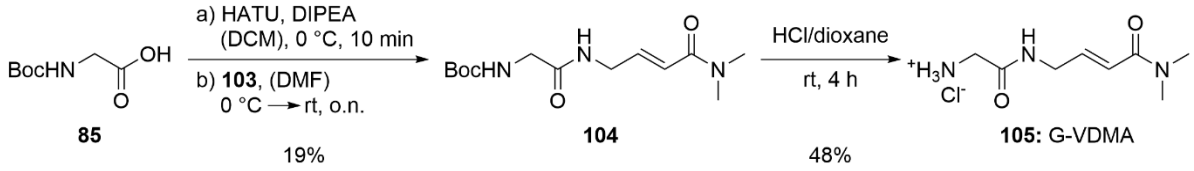
diethyl (2-(dimethylamino)-2-oxoethyl)phosphonate (101):
To **100** (1.14 g, 6.86 mmol, 1.0 eq) dissolved in dry DCE (3 mL) was added triethyl phosphite (1.76 mL, 10.3 mmol, 1.5 eq) and the resulting mixture was heated to reflux. The reaction was stirred at 90 °C for 48 h until complete consumption of the starting material was observed via TLC. Then, the organic solvent was removed *in vacuo* and the crude product purified via flash column chromatography over silica gel (DCM/MeOH gradient, 97:3 → 9:1). The desired olefination reagent **101** was obtained as a white solid (1.23 g, 5.51 mmol, 80 %).
R_f = 0.39 (DCM:MeOH, 19:1) [ninhydrin].
^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 4.02 (dq, $^2J_{H\text{-}P}$ = 8.1, 7.1 Hz, 4H), 3.10 (d, $^1J_{H\text{-}P}$ = 21.6 Hz, 2H), 3.03 (d, $^4J_{H\text{-}P}$ = 0.5 Hz, 3H), 2.82 (d, $^4J_{H\text{-}P}$ = 1.4 Hz, 3H), 1.22 (td, J = 7.1, $^3J_{H\text{-}P}$ = 0.5 Hz, 6H).
LRMS-ESI (m/z): calc. for $C_8H_{19}NO_4P^+$ ([M+H]$^+$): 224.1; found: 224.2.



***tert*-butyl (*E*)-(4-(dimethylamino)-4-oxobut-2-en-1-yl)carbamate (102):**
To a heat-dried *Schlenk*-flask under an argon atmosphere was added the olefination reagent **101** (410 mg, 1.8 mmol, 1.2 eq) dissolved in dry THF (7.5 mL). The solution was cooled to -78 °C using a cooling bath (EtOAc in liquid N_2). Then a 1 M solution of LiHMDS in THF (1.84 mL, 1.8 mmol, 1.2 eq) was added dropwise, turning the reaction mixture into a yellow solution that was subsequently stirred at -78 °C for 1 h. Next, the aldehyde **81** (244 mg, 1.5 mmol, 1.0 eq) was dissolved in dry THF (3 mL) and added dropwise to the cooled reaction. The reaction mixture was then stirred at -78 °C for 1 h before being allowed to warm to -40 °C. After stirring at -40 °C for 2 h, the reaction was quenched with a saturated solution of aqueous $NaHCO_3$ (4 mL) and THF was evaporated under reduced pressure. The crude aqueous mix was extracted with Et_2O (3 × 30 mL) and the combined organic layers washed with a saturated solution of aqueous $NaHCO_3$ (1 × 30 mL), aqueous 10% (w/v) $KHSO_4$ (1 × 30 mL) and brine (1 × 30 mL). The crude solution was dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica. Purification via flash column chromatography over silica gel (DCM/MeOH gradient, 49:1 → 24:1) furnished a 1:1.5-mixture (*E*:*Z*) of the α,β-unsaturated amide **102** as a colorless oil (200 mg, 0.88 mmol, 57%).
R_f = 0.32 (DCM:MeOH, 19:1) [UV, ninhydrin].
^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.13 (s, 1H), 6.52 (dt, J = 15.2, 5.0 Hz, 1H), 6.42 (dt, J = 15.2, 1.7 Hz, 1H), 3.73 – 3.66 (m, 2H), 3.01 (s, 3H), 2.86 (s, 3H), 1.39 (s, 9H).

(*E*)-4-amino-*N*,*N*-dimethylbut-2-enamide (103, VDMA): *N*-Boc deprotection of **102** according to standard procedure H.
Deprotection of **102** (0.2 g, 0.88 mmol, 1.0 eq) using 4 M HCl/dioxane (0.8 mL, 3.2 mmol, 3.6 eq) diluted with dry dioxane (0.8 mL) furnished a 1:1.5-mixture (*E*:*Z*) of the HCl-salt **103** (0.14 g, 0.82 mmol, 94%) as a beige powder, which was used without further purification.

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.10 (s, br, 3H), 6.76 (dt, *J* = 15.3, 1.6 Hz, 1H), 6.59 (dt, *J* = 15.3, 5.8 Hz, 1H), 3.64 (d, *J* = 5.8 Hz, 2H), 3.06 (s, 3H), 2.89 (s, 3H).



***tert*-butyl (*E*)-(2-((4-(dimethylamino)-4-oxobut-2-en-1-yl)amino)-2-oxoethyl)carbamate (104):**

Commercially available *N*-Boc protected glycine (**85**) (0.13 g, 0.75 mmol, 1.0 eq) and DIPEA (0.20 mL, 1.1 mmol, 1.5 eq) were dissolved in dry DCM (10 mL) and the solution was cooled to 0 °C using an ice bath. To this cooled solution was added HATU (0.34 g, 0.90 mmol, 1.2 eq) and the reaction stirred at 0 °C for 10 min. After the addition of **103** (0.14 g, 0.82 mmol, 1.1 eq) dissolved in dry DMF (3 mL), the cooled reaction was allowed to warm to ambient temperature and stirred o.n. at rt. The mixture was diluted with an aqueous solution of 10% (w/v) LiCl (40 mL) and the organic phase separated. The aqueous layer was extracted with DCM (3 × 15 mL) and the combined organic layers subsequently washed with an aqueous solution of 10% (w/v) LiCl (1 × 40 mL), water (1 × 40 mL) and brine (1 × 20 mL). The crude product was dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica. Purification via flash column chromatography over silica gel (DCM/MeOH gradient, 47:3 → 23:2) gave **104** as a solid (41 mg, 0.14 mmol, 19%).

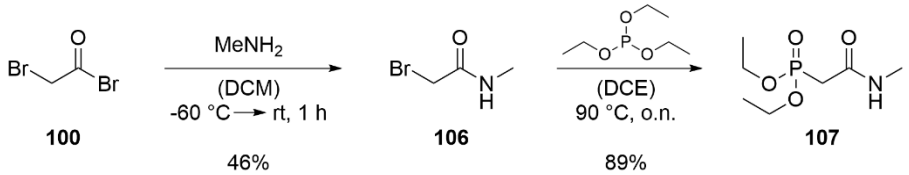
^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.05 (t, *J* = 5.8 Hz, 1H), 7.01 (t, *J* = 6.0 Hz, 1H), 6.55 (dt, *J* = 15.1, 4.6 Hz, 1H), 6.46 (dt, *J* = 15.1, 1.7 Hz, 1H), 3.88 – 3.82 (m, 2H), 3.53 (d, *J* = 6.0 Hz, 2H), 3.03 (s, 3H), 2.86 (s, 3H), 1.38 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{13}H_{24}N_3O_4^+$ ($[M+H]^+$): 286.2; found: 286.2.

(*E*)-4-(2-aminoacetamido)-*N*,*N*-dimethylbut-2-enamide (105, G-VDMA): *N*-Boc deprotection of **104** according to standard procedure H.

Deprotection of **104** (36 mg, 0.13 mmol, 1.0 eq) employing 4 M HCl/dioxane (0.1 mL, 0.40 mmol, 3.2 eq) diluted with dry dioxane (0.3 mL) yielded the HCl-salt **105** as a white solid (13 mg, 0.060 mmol, 48%), which was used without further purification.

^{1}H NMR (500 MHz, CD_3OD, 300 K): δ[ppm] = 5.86 (dt, *J* = 15.3, 5.4 Hz, 1H), 5.74 (dt, *J* = 15.3, 1.7 Hz, 1H), 3.24 (dd, *J* = 5.4, 1.7 Hz, 2H), 2.91 (s, 2H), 2.32 (s, 3H), 2.18 (s, 3H).



2-bromo-*N*-methylacetamide (106):

Commercially available bromoacetyl bromide (**99**) (1.67 mL, 22.5 mmol, 1.0 eq) was dissolved in dry DCM (25 mL) and cooled to -60 °C using a cooling bath (EtOAc in liquid N_2). To this, a 2 M solution of $MeNH_2$ in THF (22.5 mL, 45.0 mmol, 2.0 eq) was added dropwise and the reaction subsequently stirred for 1 h while being allowed to warm to ambient. The mixture was then quenched with water (15 mL) and the organic solvents removed under reduced pressure.

The crude aqueous mixture was extracted with DCM (4 × 50 mL) and the combined organic layers washed with aqueous 10% (w/v) $KHSO_4$ (1 × 70 mL), saturated aqueous $NaHCO_3$ (1 × 70 mL) and brine (1 × 70 mL), dried over Na_2SO_4, filtered and concentrated. Purification by passing the crude through a short plug of silica (Pent/EtOAc gradient, 1:1 → 1:2.5) furnished **106** as a solid (1.59 g, 10.5 mmol, 46%).

R_f = 0.21 (Pent:EtOAc, 1:1) [ninhydrin].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.18 (s, 1H), 3.83 (s, 2H), 2.61 (d, J = 4.7 Hz, 3H).

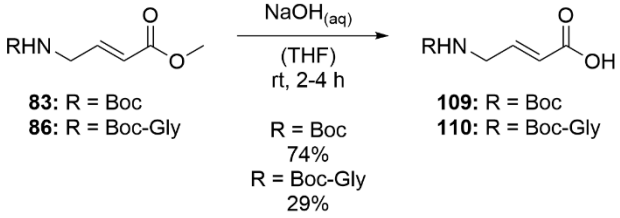
Diethyl (2-(methylamino)-2-oxoethyl)phosphonate (107):

To **106** (1.59 g, 10.5 mmol, 1.0 eq) dissolved in dry DCE (5 mL) was added triethyl phosphite (2.70 mL, 15.7 mmol, 1.5 eq) and the resulting mixture was heated to reflux. The reaction was stirred at 90 °C for 4 d until complete consumption of the starting material was observed via TLC. Then, the organic solvent was removed *in vacuo* and the crude product purified via flash column chromatography over silica gel (DCM/MeOH gradient, 99:1 → 19:1). The desired olefination reagent **107** was obtained as a white solid (1.96 g, 9.38 mmol, 89%).

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.87 (q, J = 4.8 Hz, 1H), 4.08 – 3.95 (m, 4H), 2.81 (d, $^1J_{H\text{-}P}$ = 21.3 Hz, 2H), 2.58 (dd, J = 4.8, $^4J_{H\text{-}P}$ = 1.5 Hz, 3H), 1.22 (t, $^3J_{H\text{-}P}$ = 7.1 Hz, 6H).

Note that subsequent HWE reactions using strong bases such as NaH or LiHMDS only resulted in the formation of the undesired *Z*-isomer. Also, the application of mild tertiary amine bases in the presence of lithium halides[336] or Lewis acids[337] such as $Zn(OTf)_2$ used in metal-promoted HWE reactions did not give the desired result.

9.5.5 Synthesis of Lysine-Based ncAAs (G-)VAK



(*E*)-4-((*tert*-butoxycarbonyl)amino)but-2-enoic acid (109): Hydrolysis of the methyl ester **83** by standard procedure F.

To **83** (0.55 g, 2.53 mmol, 1.0 eq) were added THF (30 mL) and 1 M aqueous NaOH (20 mL, 20.0 mmol, 8.0 eq) and the reaction mixture was stirred at rt for 4 h until complete hydrolysis of the methyl ester was observed via TLC. The desired carboxylic acid **109** was obtained as colorless foam (0.38 g, 1.88 mmol, 74%) after purification via flash column chromatography (DCM/EtOAc gradient, 9:1 → 4:1 + 1 vol% AcOH).

R_f = 0.60 (DCM:EtOAc, 1:1+ 1 vol% AcOH) [UV, ninhydrin].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 12.26 (s, 1H), 7.16 (t, J = 5.8 Hz, 1H), 6.73 (dt, J = 15.6, 4.7 Hz, 1H), 5.76 (dt, J = 15.6, 2.0 Hz, 1H), 3.74 – 3.66 (m, 2H), 1.38 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 166.9 (s, 1C), 155.5 (s, 1C), 145.9 (s, 1C), 121.0 (s, 1C), 78.0 (s, 1C), 40.7 (s, 1C), 28.2 (s, 3C).

LRMS-ESI (m/z): calc. for $C_9H_{15}NNaO_4^+$ ([M+Na]$^+$): 224.1; found: 224.1.

(*E*)-4-(2-((*tert*-butoxycarbonyl)amino)acetamido)but-2-enoic acid (110): Hydrolysis of the methyl ester **86** by standard procedure F.

To **86** (1.42 g, 5.22 mmol, 1.0 eq) were added THF (60 mL) and 1 M aqueous NaOH (40 mL, 40.0 mmol, 7.7 eq) and the reaction mixture was stirred at rt for 2 h until complete hydrolysis of the methyl ester was observed via TLC.

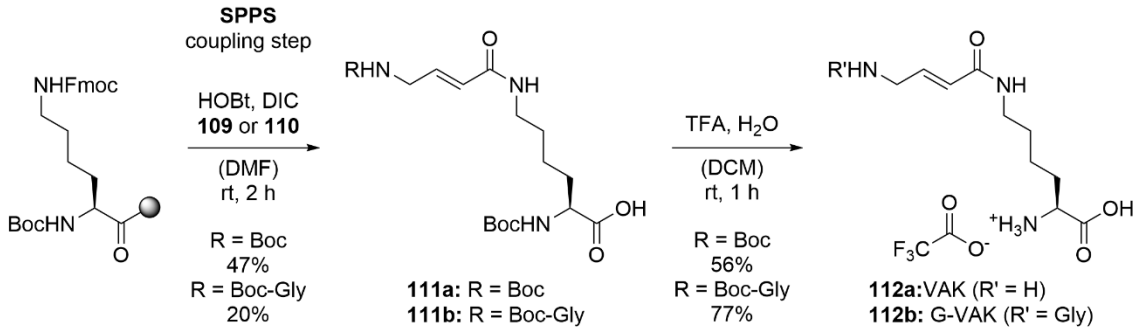
The desired carboxylic acid **110** was obtained as a colorless foam (0.38 g, 1.49 mmol, 29%) after purification via flash column chromatography (DCM/MeOH gradient, 49:1 → 93:7 + 1 vol% AcOH).

***R*$_f$** = 0.22 (DCM:MeOH, 19:1+ 1 vol% AcOH) [UV, ninhydrin].

^{1}H NMR (500 MHz, DMSO-d$_6$, 300 K): *δ*[ppm] = 12.22 (br. s, 1H), 8.06 (t, *J* = 5.8 Hz, 1H), 6.98 (t, *J* = 6.2 Hz, 1H), 6.74 (dt, *J* = 15.7, 4.6 Hz, 1H), 5.80 (dt, *J* = 15.7, 2.0 Hz, 1H), 3.85 (ddd, *J* = 6.3, 4.6, 2.0 Hz, 2H), 3.55 (d, *J* = 6.2 Hz, 2H), 1.38 (s, 9H).

^{13}C NMR (101 MHz, CD$_3$OD, 300 K): *δ*[ppm] = 172.7 (s, 1C), 169.5 (s, 1C), 158.4 (s, 1C), 145.9 (s, 1C), 122.5 (s, 1C), 80.8 (s, 1C), 44.7 (s, 1C), 40.8 (s, 1C), 28.7 (s, 3C).

HRMS-ESI (m/z): calc. for $C_{12}H_{20}N_2NaO_5^+$ ([M+Na]$^+$): 295.1264; found: 295.1261.



(*E*)-*N*^{2}-(*tert*-butoxycarbonyl)-*N*^{6}-(4-((*tert*-butoxycarbonyl)amino)but-2-enoyl)-*L*-lysine (111a): Synthesis via SPPS according to Chapter 9.6.

Boc-Lys(Fmoc)-OH (501 mg, 1.1 mmol, 2.3 eq) was immobilized on CTC resin (394 mg, 0.47 mmol, 1.0 eq) using DIPEA (320 µL, 1.9 mmol, 4.0 eq) in DCM (6 mL). After incubation at rt for 1 h and on-resin Fmoc-deprotection, the coupling reaction mixture comprising the carboxylic acid **109** (300 mg, 0.70 mmol, 1.5 eq), DIC (100 µL, 0.65 mmol, 1.4 eq) and HOBt (126 mg, 0.93 mmol, 2.0 eq) in DMF (6 mL) was applied to the resin. Incubation at rt for 1 h and subsequent cleavage from the CTC resin furnished the desired ncAA **111a** (94.0 mg, 0.22 mmol, 47%) after purification via flash column chromatography (DCM/MeOH gradient, 49:1 → 23:2 + 1 vol% AcOH).

***R*$_f$** = 0.21 (DCM:MeOH, 19:1 + 1 vol% AcOH) [UV, ninhydrin].

^{1}H NMR (400 MHz, DMSO-d$_6$, 300 K): *δ*[ppm] = 8.00 (t, *J* = 5.6 Hz, 1H), 7.09 (t, *J* = 5.9 Hz, 1H), 6.89 (s, 1H), 6.50 (dt, *J* = 15.4, 4.9 Hz, 1H), 5.95 – 5.86 (m, 1H), 3.79 (s, 1H), 3.69 – 3.62 (m, 2H), 3.07 (q, *J* = 6.4 Hz, 2H), 1.70 – 1.48 (m, 2H), 1.44 – 1.20 (m, 22H).

^{13}C NMR (101 MHz, DMSO-d$_6$, 300 K): *δ*[ppm] = 174.4 (s, 1C), 164.5 (s, 1C), 155.5 (s, 2C*), 139.5 (s, 1C), 123.8 (s, 1C), 77.8 (s, 2C), 53.6 (s, 1C), 40.6 (s, 1C), 38.3 (s, 1C), 30.7 (s, 1C), 28.8 (s, 1C), 28.2 (s, 6C), 23.1 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{20}H_{36}N_3O_7^+$ ([M+H]$^+$): 430.2548; found: 430.2548.

(*E*)-N^2-(*tert*-butoxycarbonyl)-N^6-(4-(2-((*tert*-butoxycarbonyl)amino)acetamido)but-2-enoyl)-*L*-lysine (111b): Synthesis via SPPS according to Chapter 9.6.

Boc-Lys(Fmoc)-OH (483 mg, 1.0 mmol, 2.2 eq) was immobilized on CTC resin (380 mg, 0.45 mmol, 1.0 eq) using DIPEA (310 µL, 1.8 mmol, 4.0 eq) in DCM (6 mL). After incubation at rt for 1 h and on-resin Fmoc-deprotection, the coupling reaction mixture comprising the carboxylic acid **110** (327 mg, 0.67 mmol, 1.5 eq), DIC (97.0 µL, 0.63 mmol, 1.4 eq) and HOBt (121 mg, 0.90 mmol, 2.0 eq) in DMF (6 mL) was applied to the resin.

Incubation at rt for 1 h and subsequent cleavage from the CTC resin furnished the desired ncAA **111b** (43.0 mg, 0.088 mmol, 20%) after purification via flash column chromatography (DCM/MeOH gradient, 24:1 → 9:1 + 1 vol% AcOH).

R_f = 0.11 (DCM:MeOH, 19:1 + 1 vol% AcOH) [UV, ninhydrin].

LRMS-ESI (m/z): calc. for $C_{17}H_{31}N_4O_6^+$ ([M-Boc+H]$^+$): 387.5; found: 387.2.

(*E*)-N^6-(4-aminobut-2-enoyl)-*L*-lysine (112a, VAK): *N*-Boc deprotection of **111a** according to standard procedure G.

Deprotection of **111a** (94 mg, 0.22 mmol) with a mixture of TFA (0.40 mL) and DCM (1.5 mL) containing 5% (v/v) water, gave the TFA-salt **112a** as a white solid (56 mg, 0.12 mmol, 56%) which was used without further purification.

LRMS-ESI (m/z): calc. for $C_{10}H_{20}N_3O_3^+$ ([M+H]$^+$): 230.1; found: 230.2.

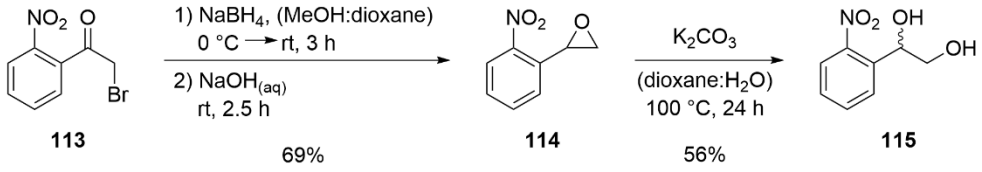
(*E*)-N^6-(4-(2-aminoacetamido)but-2-enoyl)-*L*-lysine (112b, G-VAK): *N*-Boc deprotection of **111b** according to standard procedure G.

Deprotection of **111b** (43 mg, 0.088 mmol) with a mixture of TFA (0.40 mL) and DCM (1.5 mL) containing 5% (v/v) water, gave the TFA-salt **112a** as a white solid (35 mg, 0.068 mmol, 77%) which was used without further purification.

LRMS-ESI (m/z): calc. for $C_{12}H_{23}N_4O_4^+$ ([M+H]$^+$): 287.2; found: 287.2.

Note that more detailed analytical data of both **VAK** (**111a**) and **G-VAK** (**111b**) can be found in the Master Thesis of Philipp Ruckgaber who further pursued the scale-up of both compounds and their application as tools in genetic code expansion.[338]

9.5.6 Synthesis of Masked Electrophilic Warhead G-MVK



2-(2-nitrophenyl)oxirane (114):

A solution of 2-bromo-2'-nitroacetopheonone (**113**) (38.6 g, 158 mmol, 1.0 eq) in dry dioxane (440 mL) was cooled to 0 °C using an ice bath and $NaBH_4$ (16.4 g, 432 mmol, 2.7 eq) dissolved in dry MeOH (290 mL) was added portion wise over a time course of 15 min. The reaction mixture was stirred at 0 °C until the formation of gas could no longer be observed (typically 5-15 min). The slightly yellow suspension was subsequently allowed to warm to ambient temperature and stirred at rt for 2.5 h, turning the reaction mixture green. Then, an aqueous solution of 2 M NaOH (115 mL, 230 mmol, 1.5 eq) was added and the reaction mixture stirred at rt until complete epoxide formation was observed via TLC (typically 0.5-1.5 h), changing color from turquoise, over orange and red to dusky pink. The organic solvent was removed under reduced pressure and the crude aqueous mixture extracted with EtO_2 (4 × 150 mL). The combined organic layers were washed with water (1 × 200 mL) and brine (2 × 200 mL), dried over Na_2SO_4, filtered and concentrated. Purification via flash column chromatography over silica gel (Pent/Et_2O gradient, 15:1 → 4:1) furnished the desired epoxide **114** (18.0 g, 109 mmol, 69%).

R_f = 0.55 (Pent:Et_2O, 9:1) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.14 (dd, *J* = 8.2, 1.3 Hz, 1H), 7.79 (td, *J* = 7.6, 1.3 Hz, 1H), 7.60 (ddd, *J* = 8.6, 7.6, 1.5 Hz, 1H), 7.52 (dd, *J* = 7.7, 1.5 Hz, 1H), 4.43 (dd, *J* = 4.4, 2.7 Hz, 1H), 3.26 (dd, *J* = 5.5, 4.4 Hz, 1H), 2.74 (dd, *J* = 5.5, 2.7 Hz, 1H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 147.9 (s, 1C), 134.5 (s, 1C), 134.0 (s, 1C), 129.0 (s, 1C), 126.7 (s, 1C), 124.5 (s, 1C), 50.0 (s, 2C).

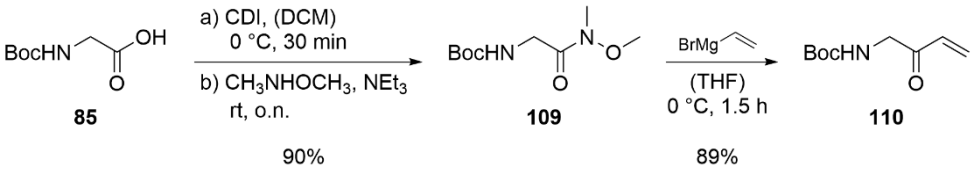
LRMS-ESI (m/z): calc. for $C_8H_8NO_3^+$ ($[M+H]^+$): 166.0; found: 166.0.

1-(2-nitrophenyl)ethane-1,2-diol (115):

For ring-opening, the epoxide **114** (18.0 g, 109 mmol, 1.0 eq) was dissolved in dioxane (140 mL) and a solution of K_2CO_3 (21.0 g, 152 mmol, 1.4 eq) in water (175 mL) was added, yielding a brown suspension. The reaction mixture was subsequently heated to 100 °C and refluxed until complete consumption of the starting material was observed via TLC (typically 16-24 h). Next, dioxane was removed under reduced pressure and the pH of the aqueous crude was adjusted to approx. 5 by the addition of 1 M HCl. The aqueous mixture was extracted with Et_2O (3 × 120 mL) and the combined organic layers washed with brine (2 × 100 mL), dried over Na_2SO_4, filtered and concentrated. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 5:1 → 1:2.5) yielded the vicinal diol **115** (11.2 g, 61.2 mmol, 56%) as a yellow solid.

R_f = 0.25 (Pent:EtOAc, 1:1) [UV, CAM].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.85 (dd, *J* = 8.1, 1.3 Hz, 1H), 7.78 (dd, *J* = 7.8, 1.5 Hz, 1H), 7.70 (td, *J* = 7.5, 1.3 Hz, 1H), 7.49 (ddd, *J* = 8.5, 7.5, 1.5 Hz, 1H), 5.61 (d, *J* = 4.8 Hz, 1H), 5.07 (dt, *J* = 6.6, 4.8 Hz, 1H), 4.89 (t, *J* = 5.8 Hz, 1H), 3.52 – 3.40 (m, 2H).



***tert*-butyl (2-(methoxy(methyl)amino)-2-oxoethyl)carbamate (109):**

Commercially available *N*-Boc glycine (**85**) (5.00 g, 28.6 mmol, 1.0 eq) was dissolved in dry DCM (100 mL) and cooled to 0 °C using an ice bath. Then, CDI (6.48 g, 40.0 mmol, 1.4 eq) was added and the suspension stirred at 0 °C for 30 min. The reaction mixture was subsequently allowed to warm to ambient temperature and *N*,*O*-dimethylhydroxylamine hydrochloride (3.90 g, 40.0 mmol, 1.4 eq) and NEt_3 (11.2 mL, 57.0 mmol, 2.0 eq) were added. After stirring o.n. at rt, the crude reaction mixture was extracted with EtO_2 (3 × 100 mL) and the combined organic layers were washed with aqueous 1 M HCl (3 × 80 mL), aqueous saturated $NaHCO_3$ (1 × 80 mL) and brine (1 × 80 mL), dried over Na_2SO_4 and filtered. The crude product was then concentrated *in vacuo* to obtain the Weinreb-amide **109** as a white solid (5.61 g, 25.7 mmol, 90%).

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 6.83 (t, *J* = 6.2 Hz, 1H), 3.82 (d, *J* = 6.2 Hz, 2H), 3.67 (s, 3H), 3.08 (s, 3H), 1.38 (s, 9H).

LRMS-ESI (m/z): calc. for $C_9H_{18}N_2NaO_4^+$ ($[M+H]^+$): 241.1; found: 241.1.

***Tert*-butyl (2-oxobut-3-en-1-yl)carbamate (110):**

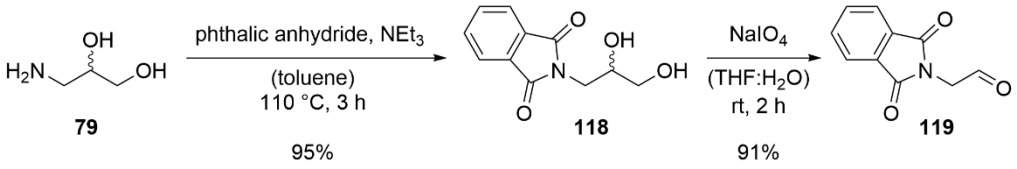
The Weinreb-amide **109** (0.50 g, 2.30 mmol, 1.0 eq) was dissolved in dry Et_2O (6 mL) and the solution was cooled to -20 °C using a cooling bath (EtOAc in liquid N_2). To this cooled mixture was added dropwise a 1 M solution of vinylmagnesium bromide in THF (9.0 mL, 9.20 mmol, 4.0 eq) and the Grignard reaction allowed to warm to 0 °C using an ice bath.

The solution was stirred at 0 °C until complete consumption of the starting material was observed via TLC (typically 1-1.5 h). Upon completion, the reaction was quenched by the addition of cold, aqueous 2 M HCl (23 mL) and the crude was subsequently extracted with Et_2O (3 × 20 mL). The combined organic layers were then washed with aqueous 2 M HCl (3 × 30 mL), aqueous saturated $NaHCO_3$ (1 × 30 mL) and brine (1 × 30 mL), dried over Na_2SO_4 and filtered. The crude product **110** (0.38 g, 2.05 mmol, 89%) was obtained as a yellow oil via concentration under reduced pressure at 25 °C to prevent polymerization.

R_f = 0.80 (DCM:EtOAc, 1:1) [ninhydrin].

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.01 (s, 1H), 6.43 (dd, *J* = 17.7, 10.4 Hz, 1H), 6.29 (dd, *J* = 17.7, 1.6 Hz, 1H), 5.92 (dd, *J* = 10.4, 1.6 Hz, 1H), 3.98 (d, *J* = 6.0 Hz, 2H), 1.38 (s, 12H).

^{13}C NMR (75 MHz, DMSO-d_6, 300 K): δ[ppm] = 196.8 (s, 1C), 156.3 (s, 1C), 134.2 (s, 1C), 129.7 (s, 1C), 78.6 (s, 1C), 48.5 (s, 1C), 28.6 (s, 3C).



2-(2,3-dihydroxypropyl)isoindoline-1,3-dione (118):

To commercially available (±)-3-amino-1,2-propanediol (**79**) (2.5 g, 27.4 mmol, 1.0 eq) and NEt_3 (0.38 mL, 2.74 mmol, 10 mol-%) dissolved in dry toluene (40 mL) was added phthalic anhydride (4.3 g, 28.8 mmol, 1.05 eq). Using a Dean-Stark apparatus, the reaction mixture was refluxed at 110 °C for 3 h until complete consumption of the starting material was observed via TLC. The organic solvent was removed under reduced pressure, the crude resuspended in aqueous 1 M HCl (50 mL) and subsequently extracted with EtOAc (5 × 80 mL). The combined organic layers were washed with aqueous saturated $NaHCO_3$ (1 × 120 mL) and brine (1 × 120 mL), dried over Na_2SO_4, filtered and concentrated *in vacuo*. Purification by recrystallization from hot EtOAc (15 mL, 100 °C → 4 °C) afforded the title compound **118** (5.76 g, 26.1 mmol, 95%) as a white solid.

R_f = 0.21 (Et_2O:EtOAc, 2:1) [UV, $KMnO_4$].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.88 – 7.81 (m, 4H), 4.90 (d, J = 5.2 Hz, 1H), 4.65 (t, J = 5.7 Hz, 1H), 3.81 (h, J = 5.5 Hz, 1H), 3.58 (d, J = 6.6 Hz, 2H), 3.44 – 3.32 (m, 2H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 168.1 (s, 2C), 134.2 (s, 2C), 131.8 (s, 2C), 122.9 (s, 2C), 68.4 (s, 1C), 64.2 (s, 1C), 41.7 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{11}H_{12}NO_4^+$ ($[M+H]^+$): 222.1; found: 222.1.

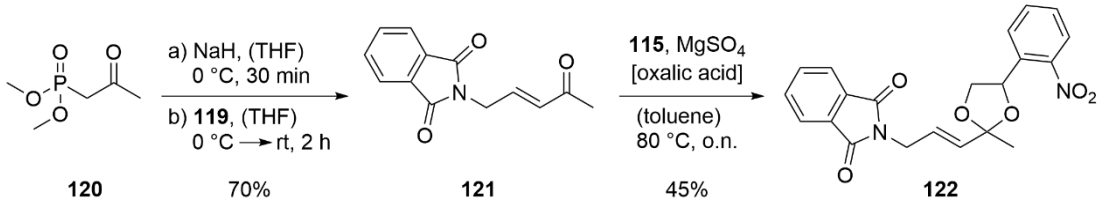
2-(1,3-dioxoisoindolin-2-yl)acetaldehyde (119):

For oxidative cleavage, the vicinal diol **118** (0.75 g, 3.4 mmol, 1.0 eq) was dissolved in THF (5 mL) and a solution of $NaIO_4$ (0.87 g, 4.1 mmol, 1.2 eq) in water (6 mL) was added. The reaction mixture was subsequently stirred at rt for 2 h in the dark before being extracted with DCM (3 × 12 mL). The combined organic layers were washed with brine (1 × 10 mL), dried over Na_2SO_4, filtered and concentrated to furnish the aldehyde **119** (0.58 g, 3.1 mmol, 91%) as a white solid which was stored at -20 °C until further use.

R_f = 0.72 (DCM:EtOAc, 1:1) [UV, $KMnO_4$].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 9.60 (s, 1H), 7.97 – 7.85 (m, 4H), 4.59 (s, 2H).

LRMS-ESI (m/z): calc. for $C_{10}H_8NO_3^+$ ($[M+H]^+$): 190.0; found: 190.1.



(*E*)-2-(4-oxopent-2-en-1-yl)isoindoline-1,3-dione (121): The HWE reaction was performed according to standard procedure K.

The commercially available olefination reagent **120** (0.51 mL, 3.72 mmol, 1.2 eq) was reacted with NaH (0.14 g, 3.41 mmol, 1.1 eq) in dry THF (24 mL) and subsequently stirred at 0 °C for 30 min. Then, a solution of aldehyde **119** (0.58 g, 3.10 mmol, 1.0 eq) in dry THF (18 mL) was added and the reaction mixture stirred at 0 °C for 30 min before being warmed to ambient temperature.

After stirring at rt for 1.5 h, the crude product was purified via flash column chromatography over silica gel (DCM/EtOAc gradient, 3:1 → 2:1) to exclusively yield the *E*-isomer of the α,β-unsaturated ketone **121** as a white, crystalline solid (0.50 g, 2.18 mmol, 70%).

R_f = 0.77 (DCM:EtOAc, 1:1) [UV, $KMnO_4$].

^{1}H NMR (400 MHz, $CDCl_3$, 300 K): δ[ppm] = 7.93 – 7.72 (m, 4H), 6.75 (dtd, *J* = 16.0, 5.2, 1.0 Hz, 1H), 6.12 (dt, *J* = 16.0, 1.5 Hz, 1H), 4.47 (dt, *J* = 5.2, 1.5 Hz, 2H), 2.25 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$, 300 K): δ[ppm] = 197.7 (s, 1C), 167.7 (s, 2C), 139.7 (s, 1C), 134.5 (s, 2C), 132.0 (s, 3C*), 123.7 (s, 2C), 38.5 (s, 1C), 27.3 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{13}H_{12}NO_3^+$ ([M+H]$^+$): 230.1; found: 230.1.

(*E*)-2-(3-(2-methyl-4-(2-nitrophenyl)-1,3-dioxolan-2-yl)allyl)isoindoline-1,3-dione (122):

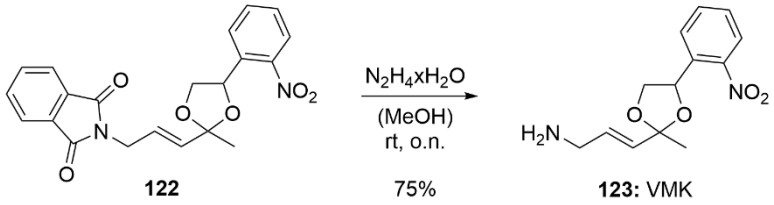
To a solution of the α,β-unsaturated ketone **121** (450 mg, 2.0 mmol, 1.0 eq) in dry toluene (20 mL) were added the vicinal diol **115** (1.08 g, 5.9 mmol, 3.0 eq), anhydrous Mg_2SO_4 (260 mg, 2.2 mmol, 1.1 eq) and catalytic amounts of oxalic acid (18.0 mg, 0.20 mmol, 10 mol-%). After stirring the reaction mixture o.n. at 80 °C, the crude was purified via flash column chromatography over silica gel (cyclohexane/EtOAc gradient, 4:1 → 3:2) to afford the desired target **122** (354 mg, 0.90 mmol, 45%) without any β,γ-homologation.

R_f = 0.58 (cyclohexane:EtOAc, 3:2) [UV, $KMnO_4$].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.13 – 8.07 (m, 1H), 7.93 – 7.78 (m, 6H), 7.62 – 7.56 (m, 1H), 5.97 – 5.91 (m, 1H), 5.64 (dt, *J* = 15.6, 1.6 Hz, 1H), 5.50 (dd, *J* = 7.5, 4.8 Hz, 1H), 4.30 (dd, *J* = 8.5, 7.5 Hz, 1H), 4.25 (dd, *J* = 5.3, 1.6 Hz, 3H), 3.78 (dd, *J* = 8.5, 4.8 Hz, 1H), 1.57 (s, 3H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 167.5 (s, 2C), 146.7 (s, 1C), 137.1 (s, 1C), 134.5 (s, 1C), 134.4 (s, 2C), 134.3 (s, 1C), 131.7 (s, 1C), 131.5 (s, 1C), 128.8 (s, 1C), 127.3 (s, 1C), 125.4 (s, 1C), 124.7 (s, 1C), 123.1 (s, 2C), 108.1 (s, 1C), 73.6 (s, 1C), 70.5 (s, 1C), 38.1 (s, 1C), 24.7 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{21}H_{18}N_2NaO_6^+$ ([M+Na]$^+$): 417.1057; found: 417.1061.



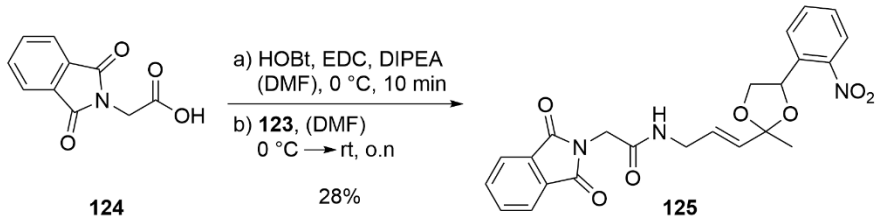
(*E*)-3-(2-methyl-4-(2-nitrophenyl)-1,3-dioxolan-2-yl)prop-2-en-1-amine (123, VMK):

The photocaged ketone **122** (0.48 g, 1.2 mmol, 1.0 eq) was dissolved in dry MeOH (19 mL) and hydrazine hydrate (0.61 mL, 12 mmol, 10 eq) was added dropwise. The yellow solution was stirred o.n. at rt before quenching with aqueous 5% (v/v) H_2O_2. The organic solvent was evaporated under reduced pressure and the crude subsequently basified using aqueous 2 M NaOH (15 mL). The aqueous mixture was extracted with DCM (3 × 10 mL) and the combined organic layers washed with aqueous 2 M NaOH (2 × 10 mL) to then extract them with aqueous 10% (w/v) citric acid (3 × 15 mL). Basification of the combined aqueous phases by aqueous 2 M NaOH allowed for anew extraction of the product with DCM (3 × 15 mL). The combined organic phases were finally dried over Na_2SO_4, filtered and concentrated to yield the free amine **123** (0.24 g, 0.90 mmol, 75%) as a yellow oil which was used without further purification.

R_f = 0.16 (DCM:MeOH, 17:3) [UV, $KMnO_4$].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.11 (dd, *J* = 8.2, 1.2 Hz, 1H), 7.90 – 7.79 (m, 4H), 7.63 – 7.57 (m, 1H), 5.94 (dq, *J* = 15.5, 5.1 Hz, 1H), 5.60 (dt, *J* = 15.5, 1.7 Hz, 1H), 5.50 (dd, *J* = 7.6, 4.7 Hz, 1H), 4.35 (dd, *J* = 8.5, 7.6 Hz, 1H), 3.81 – 3.76 (m, 1H), 3.20 (ddd, *J* = 5.1, 3.1, 1.7 Hz, 2H), 1.61 (s, 3H).

HRMS-ESI (m/z): calc. for $C_{13}H_{17}N_2O_4^+$ ([M+H]$^+$): 265.1183; found: 265.1187.



(*E*)-2-(1,3-dioxoisoindolin-2-yl)-N-(3-(2-methyl-4-(2-nitrophenyl)-1,3-dioxolan-2-yl)allyl)-acetamide (125): Amide coupling of **124** and **123** by standard procedure L.

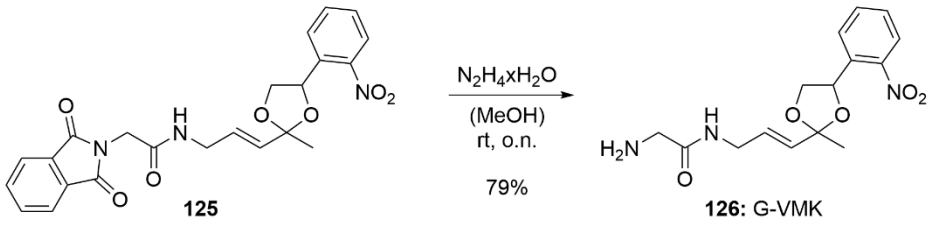
Commercially available *N*-phthaloyl glycine (**124**) (0.37 g, 1.80 mmol, 1.1 eq) was reacted with EDC×HCl (0.35 mg, 1.80 mmol, 1.1 eq), HOBt (0.28 mg, 1.80 mmol, 1.1 eq) and DIPEA (0.59 mL, 3.44 mmol, 2.1 eq) in dry DMF (4 mL), stirring at 0 °C for 10 min. Next, the photocaged amine **123** (0.43 g, 1.64 mmol, 1.0 eq) was added and the reaction mixture stirred o.n. at rt. The crude product was purified via flash column chromatography over silica gel (DCM/MeOH gradient, 19:1 → 17:3) to give the desired product **125** as a yellow solid (0.21 g, 0.456 mmol, 28%).

R_f = 0.27 (DCM:EtOAc, 4:1) [UV, $KMnO_4$].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.44 (t, *J* = 5.7 Hz, 1H), 8.12 (dd, *J* = 8.3, 1.2 Hz, 1H), 7.93 – 7.86 (m, 5H), 7.83 (td, *J* = 7.4, 1.2 Hz, 1H), 7.61 (ddd, *J* = 8.3, 7.4, 1.7 Hz, 1H), 5.83 (dt, *J* = 15.5, 5.1 Hz, 1H), 5.65 (dt, *J* = 15.5, 1.7 Hz, 1H), 5.52 (dd, *J* = 7.5, 4.8 Hz, 1H), 4.35 (dd, *J* = 8.6, 7.5 Hz, 1H), 4.24 (s, 2H), 3.83 – 3.76 (m, 3H), 1.61 (s, 3H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 167.6 (s, 2C), 165.9 (s, 1C), 146.8 (s, 1C), 137.1 (s, 1C), 134.5 (s, 2C), 134.3 (s, 1C), 131.8 (s, 2C), 130.3 (s, 1C), 128.8 (s, 1C), 127.6 (s, 1C), 127.3 (s, 1C), 124.7 (s, 1C), 123.2 (s, 2C), 108.3 (s, 1C), 73.6 (s, 1C), 70.6 (s, 1C), 39.9 (s, 1C), 39.1 (s, 1C), 24.9 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{23}H_{21}N_3NaO_7^+$ ([M+Na]$^+$): 474.1; found: 474.1.



(*E*)-2-amino-N-(3-(2-methyl-4-(2-nitrophenyl)-1,3-dioxolan-2-yl)allyl)acetamide (126, G-VMK):

The photocaged ketone **125** (0.15 mg, 0.33 mmol, 1.0 eq) was dissolved in dry MeOH (6 mL) and hydrazine hydrate (0.16 mL, 8.3 mmol, 10 eq) was added dropwise. The yellow solution was stirred o.n. at rt before quenching with aqueous 5% (v/v) H_2O_2.

The organic solvent was evaporated under reduced pressure and the crude subsequently basified using aqueous 2 M NaOH (8 mL). The aqueous mixture was extracted with DCM (3 × 5 mL) and the combined organic layers washed with aqueous 2 M NaOH (2 × 5 mL) to then extract them with aqueous 10% (w/v) citric acid (3 × 8 mL). Basification of the combined aqueous phases by aqueous 2 M NaOH allowed for anew extraction of the product with DCM (3 × 8 mL). The combined organic phases were finally dried over Na_2SO_4, filtered and concentrated to yield the free amine **126** (82 mg, 0.26 mmol, 79%) as a yellow oil which was used without further purification.

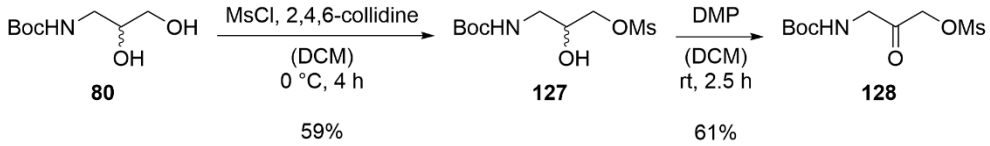
R_f = 0.17 (DCM:MeOH, 9:1) [UV, $KMnO_4$].

^{1}H NMR (400 MHz, DMOS-d_6, 300 K): *δ*[ppm] = 8.14 – 8.04 (m, 2H), 7.89 – 7.79 (m, 2H), 7.60 (td, *J* = 7.6, 1.7 Hz, 1H), 5.88 – 5.81 (m, 1H), 5.61 (d, *J* = 15.7 Hz, 1H), 5.50 (dd, *J* = 7.5, 4.9 Hz, 1H), 4.34 (t, *J* = 8.0 Hz, 1H), 3.85 – 3.76 (m, 3H), 3.16 (s, 2H), 1.60 (s, 3H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): *δ*[ppm] = 172.0 (s, 1C), 146.8 (s, 1C), 137.1 (s, 1C), 134.3 (s, 1C), 130.1 (s, 1C), 128.8 (s, 1C), 128.2 (s, 1C), 127.3 (s, 1C), 124.7 (s, 1C), 108.3 (s, 1C), 73.5 (s, 1C), 70.6 (s, 1C), 44.3 (s, 1C), 39.2 (s, 1C), 24.9 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{15}H_{20}N_3O_5^+$ ($[M+H]^+$): 322.1; found: 322.2.

9.5.7 Synthesis Towards Masked Electrophilic Warhead G-ClMK



3-((*tert*-butoxycarbonyl)amino)-2-hydroxypropyl methanesulfonate (127):

N-Boc protected (±)-3-amino-1,2-propanediol (**80**) (0.50 g, 2.62 mmol, 1.0 eq) and collidine (3.5 mL, 26.2 mmol, 10 eq) were dissolved in dry DCM (50 mL) and subsequently cooled to 0 °C using an ice bath. To this, MsCl (0.22 mL, 2.88 mmol, 1.1 eq) was added and the reaction mixture stirred at 0 °C for 4 h before the addition of water (50 mL). The crude mixture was acidified with aqueous 10% (w/v) $KHSO_4$ (50 mL) and the organic phase separated. The aqueous phase was extracted with DCM (3 × 50 mL) and the combined organic layers washed with aqueous 10% (w/v) $KHSO_4$ (1 × 80 mL) and brine (1 × 80 mL), dried over Na_2SO_4, filtered and concentrated. Purification via flash column chromatography over silica gel (Pent/EtOAc gradient, 1:1 → 1:1.5) furnished the selectively mesylated diol **127** (0.42 g, 1.54 mmol, 59%).

R_f = 0.26 (cyclohexane:EtOAc, 1:2.5) [ninhydrin].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): *δ*[ppm] = 6.83 (t, *J* = 6.0 Hz, 1H), 5.28 (d, *J* = 5.4 Hz, 1H), 4.13 (dd, *J* = 10.5, 3.5 Hz, 1H), 4.05 – 3.97 (m, 1H), 3.77 – 3.68 (m, 1H), 3.16 (s, 3H), 2.99 (hept, *J* = 6.9 Hz, 2H), 1.38 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): *δ*[ppm] = 155.7 (s, 1C), 77.8 (s, 1C), 72.1 (s, 1C), 67.6 (s, 1C), 42.7 (s, 1C), 36.7 (s, 1C), 28.2 (s, 3C).

HRMS-ESI (m/z): calc. for $C_9H_{19}NNaO_6S^+$ ($[M+Na]^+$): 290.0825; found: 290.0829.

3-((*tert*-butoxycarbonyl)amino)-2-oxopropyl methanesulfonate (128):

To the mesylated diol **127** (2.0 g, 7.37 mmol, 1.0 eq) in dry DCM (20 mL) was added a solution of Dess-Martin periodinane (3.8 g, 8.85 mmol, 1.2 eq) in dry DCM (20 mL) and the reaction mixture was stirred at rt until full conversion was observed via TLC (typically 2-4 h).

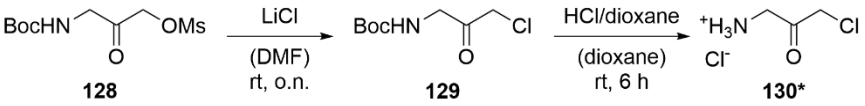
The reaction was subsequently quenched with water (60 mL) and the organic phase separated. The aqueous phase was extracted with DCM (3 × 30 mL) and the combined organic layers were washed with brine (1 × 40 mL), dried over Na_2SO_4 and filtered. The crude was concentrated, dry-loaded onto silica and purified by flash column chromatography (Pent/EtOAc gradient, 2:1 → 1:2) to yield the desired product **128** (1.2 g, 4.48 mmol, 61%) as a colorless oil.

R_f = 0.28 (cyclohexane:EtOAc, 1:1) [ninhydrin].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.13 (t, J = 5.8 Hz, 1H), 5.02 (s, 2H), 3.87 (d, J = 5.8 Hz, 2H), 3.23 (s, 3H), 1.39 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 199.9 (s, 1C), 155.8 (s, 1C), 78.4 (s, 1C), 71.1 (s, 1C), 46.9 (s, 1C), 37.3 (s, 1C), 28.1 (s, 3C).

HRMS-ESI (m/z): calc. for $C_9H_{18}NO_6S^+$ ([M+H]$^+$): 268.0849; found: 268.0852.



***tert*-butyl (3-chloro-2-oxopropyl)carbamate (129):**

To the mesylate **128** (0.13 mg, 0.39 mmol, 1.0 eq) in dry DMF (4 mL) was added anhydrous LiCl (0.17 g, 3.9 mmol, 10 eq) and the reaction mixture was stirred at rt until complete consumption of the starting material was observed (typically stirred o.n.). The crude mixture was diluted with aqueous 10% (w/v) LiCl (25 mL) and subsequently extracted with DCM (3 × 10 mL). The combined organic layers were then washed with water (3 × 50 mL) and brine (1 × 20 mL), dried over Na_2SO_4 and filtered. The crude was concentrated, dry-loaded onto silica and purified by flash column chromatography (Pent/EtOAc gradient, 2:1 → 1:1) to afford the α-chloro ketone **129** (60 mg, 0.29 mmol, 74%) as an orange oil.

R_f = 0.88 (Pent:EtOAc, 1:1) [ninhydrin].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.12 (t, J = 6.0 Hz, 1H), 4.52 (s, 2H), 3.89 (d, J = 6.0 Hz, 2H), 1.38 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 199.1 (s, 1C), 155.8 (s, 1C), 78.3 (s, 1C), 47.7 (s, 1C), 47.2 (s, 1C), 28.1 (s, 3C).

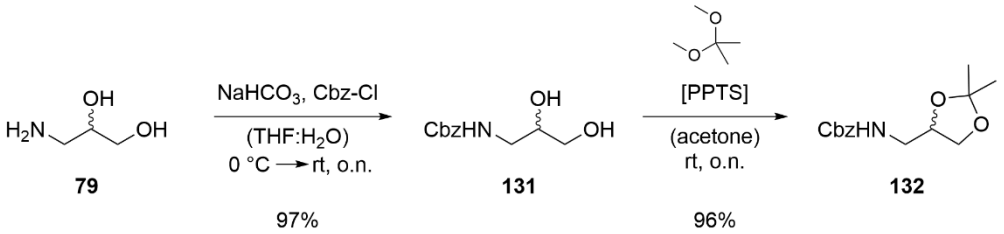
LRMS-ESI (m/z): calc. for $C_3H_7ClNO^+$ ([M-Boc+H]$^+$): 108.1; found: 108.1.

1-amino-3-chloropropan-2-one (130): *N*-Boc deprotection of **129** according to standard procedure H.

Deprotection of **129** (0.29 g, 1.4 mmol, 1.0 eq) employing 4 M HCl/dioxane (2.1 mL, 8.4 mmol, 6.0 eq) diluted with dry dioxane (1.4 mL) furnished the HCl-salt **130** as an impure mixture of product and side products, which could neither be identified via LR-MS nor by NMR. Consequently, no yield was determined.

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.35 (s, br, 3H), 4.64 (s, 2H), 4.01 (q, J = 5.6 Hz, 1H).

Note that the subsequent amide coupling to *N*-phthaloyl glycine yielded an intractable reaction mixture from which the desired product could not be isolated, presumably due to the α-chloro ketone's susceptibility towards nucleophiles.



benzyl (2,3-dihydroxypropyl)carbamate (131):

To commercially available (±)-3-amino-1,2-propanediol (**79**) (2.73 g, 30 mmol, 1.0 eq) in THF (15 mL) was added a solution of $NaHCO_3$ (5.04 g, 60 mmol, 2.0 eq) dissolved in water (15 mL). The biphasic system was cooled to 0 °C using an ice bath and Cbz-Cl (4.20 mL, 30 mmol, 1.0 eq) was added dropwise. The reaction mixture was allowed to warm to ambient temperature and subsequently stirred vigorously at rt. If needed, additional Cbz-Cl (typically 0.30 eq) was added until complete conversion was observed via TLC. Then, the organic solvent was removed under reduced pressure and the crude acidified using aqueous 1 M HCl. The aqueous mixture was extracted with EtOAc (3 × 50 mL) and the combined organic layers washed with aqueous 1 M HCl (2 × 50 mL) and brine (1 × 50 mL), dried over Na_2SO_4, filtered and concentrated. Purification via flash column chromatography over silica gel (DCM/MeOH, 19:1) gave the desired diol **131** as a colorless oil (6.55 g, 29 mmol, 97%).

R_f = 0.08 (DCM:MeOH, 9:1) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.42 – 7.28 (m, 5H), 7.10 (t, *J* = 5.8 Hz, 1H), 5.01 (s, 2H), 4.66 (d, *J* = 5.2 Hz, 1H), 4.49 (t, *J* = 5.7 Hz, 1H), 3.48 (dp, *J* = 6.7, 5.2 Hz, 1H), 3.29 (td, *J* = 5.7, 0.9 Hz, 2H), 3.11 (dt, *J* = 13.4, 5.7 Hz, 1H), 2.97 – 2.88 (m, 1H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 156.3 (s, 1C), 137.2 (s, 1C), 128.3 (s, 3C), 127.7 (s, 2C), 70.5 (s, 1C), 65.2 (s, 1C), 63.8 (s, 1C), 43.9 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{11}H_{15}NNaO_4^+$ ([M+Na]$^+$): 248.0893; found: 248.0900.

benzyl ((2,2-dimethyl-1,3-dioxolan-4-yl)methyl)carbamate (132):

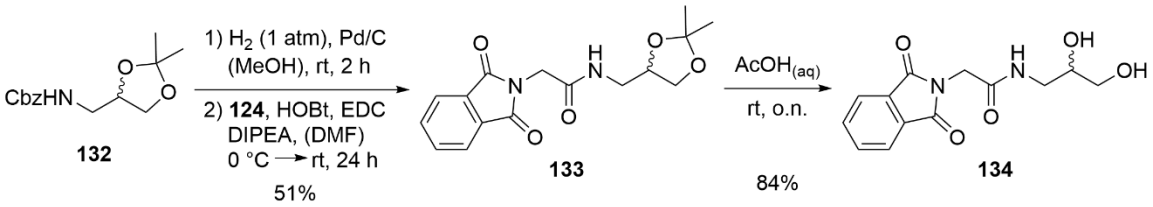
To the diol **131** (7.27 g, 32.3 mmol, 1.0 eq) dissolved in acetone (120 mL) was added 2,2-dimethoxypropane (12.6 mL, 103 mmol, 3.2 eq) and catalytic amounts of PPTS (73.5 mg, 9 mol-%). The reaction mixture was stirred o.n. at rt and acetone was removed *in vacuo*. The crude was subsequently purified by flash column chromatography (cyclohexane/EtOAc gradient, 5:1 → 2.5:1) to yield the fully protected compound **132** (8.23 g, 31.1 mmol, 96%) as a colorless oil.

R_f = 0.19 (cyclohexane:EtOAc, 1:1) [UV, CAM].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.42 – 7.28 (m, 6H), 5.07 – 5.00 (m, 2H), 4.06 (p, *J* = 5.9 Hz, 1H), 3.94 (dd, *J* = 8.3, 6.2 Hz, 1H), 3.62 (dd, *J* = 8.3, 5.8 Hz, 1H), 3.11 (qt, *J* = 13.7, 5.9 Hz, 2H), 1.31 (s, 3H), 1.24 (s, 3H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 156.3 (s, 1C), 137.1 (s, 1C), 128.3 (s, 2C), 127.8 (s, 1C), 127.7 (s, 2C), 108.4 (s, 1C), 74.3 (s, 1C), 66.5 (s, 1C), 65.3 (s, 1C), 43.1 (s, 1C), 26.8 (s, 1C), 25.3 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{14}H_{19}NNaO_4^+$ ([M+Na]$^+$): 288.1206; found: 288.1207.



***N*-((2,2-dimethyl-1,3-dioxolan-4-yl)methyl)-2-(1,3-dioxoisoindolin-2-yl)acetamide (133):** *N*-Cbz deprotection and subsequent amide coupling with **124** by standard procedure L.

To Pd/C (loading 10 wt%, 73 mg) was added a solution of **132** (0.90 g, 3.63 mmol, 1.0 eq) in MeOH (30 mL) and the argon atmosphere subsequently replaced with H_2 two times. The reaction mixture was stirred vigorously at rt for 2 h and the catalyst subsequently filtered through a sand/cotton wool plug. The solvent was removed under reduced pressure (not below 100 mbar) at 25 °C to give the volatile free amine containing minimal amounts of residual MeOH, which was used without further purification in the next step.

Commercially available *N*-phthaloyl glycine (**124**) (0.74 g, 3.63 mmol, 1.0 eq) was reacted with EDC×HCl (0.77 g, 3.99 mmol, 1.1 eq), HOBt (0.54 g, 3.99 mmol, 1.1 eq) and DIPEA (1.3 mL, 7.62 mmol, 2.1 eq) in dry DMF (5 mL), stirring at 0 °C for 10 min. Next, a solution of previously obtained free amine (3.63 mmol, 1.0 eq) was added and the reaction mixture stirred o.n. at rt. The crude product was purified via flash column chromatography over silica gel (DCM/EtOAc gradient, 4:1 → 2:3) to furnish the desired product **133** as a white solid (0.59 g, 1.86 mmol, 51%).

R_f = 0.35 (DCM:EtOAc, 3:2) [UV, $KMnO_4$].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.39 (t, *J* = 5.9 Hz, 1H), 7.94 – 7.84 (m, 4H), 4.21 (s, 2H), 4.06 (p, *J* = 5.9 Hz, 1H), 3.94 (dd, *J* = 8.3, 6.2 Hz, 1H), 3.56 (dd, *J* = 8.3, 6.2 Hz, 1H), 3.20 (m, 2H), 1.34 (s, 3H), 1.26 (s, 3H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 168.0 (s, 1C), 166.8 (s, 2C), 135.0 (s, 2C), 132.2 (s, 2C), 123.7 (s, 2C), 108.9 (s, 1C), 74.7 (s, 1C), 66.9 (s, 1C), 41.8 (s, 1C), 40.5 (s, 1C), 27.2 (s, 1C), 25.9 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{16}H_{18}N_2NaO_5^+$ ([M+Na]$^+$): 341.1108; found: 341.1111.

***N*-(2,3-dihydroxypropyl)-2-(1,3-dioxoisoindolin-2-yl)acetamide (134):**

The acetonide **133** (0.19 g, 0.61 mmol, 1.0 eq) was dissolved in aqueous 80% (v/v) AcOH (6.3 mL) and stirred o.n. at rt until full conversion was observed via TLC. AcOH was co-evaporated with toluene and the crude diluted with brine (20 mL) before extraction with a 3:1 mixture of $CHCl_3$:IPA (6 × 10 mL). The combined organic layers were then washed with brine (1 × 20 mL), dried over Na_2SO_4 and filtered. The crude was concentrated and dried *in vacuo* to obtain the vicinal diol **134** (0.14 g, 0.51 mmol, 84%) as a slightly yellow solid.

R_f = 0.35 (DCM:MeOH, 9:1) [UV, $KMnO_4$].

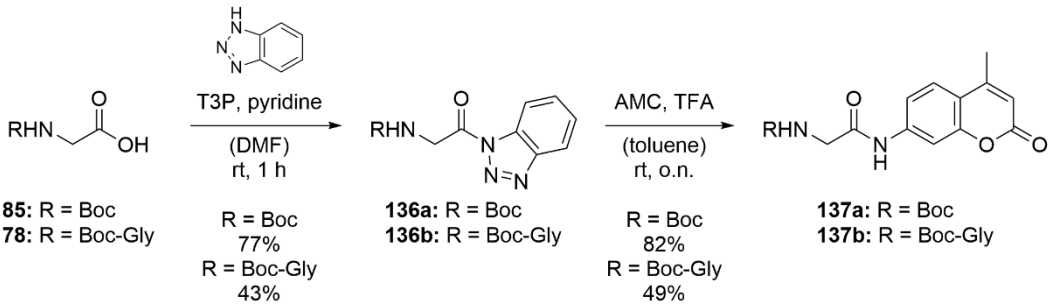
^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.21 (t, *J* = 5.7 Hz, 1H), 7.94 – 7.84 (m, 4H), 4.77 (d, *J* = 5.2 Hz, 1H), 4.51 (t, *J* = 5.8 Hz, 1H), 4.22 (s, 2H), 3.49 (dp, *J* = 6.8, 5.2 Hz, 1H), 3.28 (td, *J* = 5.8, 2.1 Hz, 1H), 3.26 – 3.19 (m, 1H), 2.98 (ddd, *J* = 13.4, 6.8, 5.3 Hz, 1H).

^{13}C NMR (101 MHz, DMSO-d_6, 300 K): δ[ppm] = 167.5 (s, 1C), 166.2 (s, 2C), 134.6 (s, 2C), 131.7 (s, 2C), 123.2 (s, 2C), 70.2 (s, 1C), 63.6 (s, 1C), 42.3 (s, 1C), 40.0 (s, 1C).

HRMS-ESI (m/z): calc. for $C_{13}H_{15}N_2O_5^+$ ([M+H]$^+$): 279.0975; found: 279.0974.

Note that subsequent, selective mesylation of the primary alcohol failed due to solubility issues of **134** in DCM. Different polar, aprotic solvents could be tested instead.

9.5.8 Synthesis of Fluorogenic Probes G-AMC and GG-AMC



***tert*-butyl (2-((2-(1H-benzo[d][1,2,3]triazol-1-yl)-2-oxoethyl)amino)-2-oxoethyl)carbamate (136a):**

The synthesis of **136a** was carried out similar to known literature procedures.[343]

To commercially available *N*-Boc glycine (**85**) (0.70 g, 4.0 mmol, 1.0 eq), benzotriazole (0.57 g, 4.8 mmol, 1.2 eq) and pyridine (0.97 mL, 12 mmol, 3.0 eq) in dry DMF (5 mL) was added dropwise a 50% solution of propanephosphonic acid anhydride (T3P) in DMF (3.8 g, 6.0 mmol, 1.5 eq). The reaction mixture was stirred at rt for 1.5 h and subsequently poured into ice water. The white precipitate was filtered, dried and redissolved in EtOAc. The organic phase was then dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was finally redissolved in minimal amounts of DCM and carefully layered with six times the volume of pentane. After being cooled o.n. at 4 °C, the precipitate was filtered and dried to furnish the desired product **136a** (0.85 g, 3.1 mmol, 77%) as a white solid.

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.30 – 8.26 (m, 1H), 8.23 (d, *J* = 8.2 Hz, 1H), 7.80 (ddd, *J* = 8.2, 7.0, 1.0 Hz, 1H), 7.63 (ddd, *J* = 8.2, 7.0, 1.0 Hz, 1H), 7.52 (t, *J* = 6.0 Hz, 1H), 4.81 (d, *J* = 6.0 Hz, 2H), 1.43 (s, 9H).

***tert*-butyl (2-((2-((2-(1H-benzo[d][1,2,3]triazol-1-yl)-2-oxoethyl)amino)-2-oxoethyl)-amino)-2-oxoethyl)carbamate (136b):**

The synthesis of **136b** was carried out similar to known literature procedures.[343]

To *N*-Boc-Gly-Gly-OH (**78**) (0.50 g, 2.2 mmol, 1.0 eq), benzotriazole (0.31 g, 2.6 mmol, 1.2 eq) and pyridine (0.52 mL, 6.5 mmol, 3.0 eq) in dry DMF (2 mL) was added dropwise a 50% solution of propanephosphonic acid anhydride (T3P) in DMF (2.1 g, 3.2 mmol, 1.5 eq). The reaction mixture was stirred at rt for 1.5 h and subsequently poured into ice water. The white precipitate was filtered, dried and redissolved in EtOAc. The organic phase was then dried over Na_2SO_4, filtered and concentrated *in vacuo*. The crude was finally redissolved in minimal amounts of DCM and carefully layered with six times the volume of pentane. After being cooled o.n. at 4 °C, the precipitate was filtered and dried to furnish the desired product **136b** (0.32 g, 0.95 mmol, 43%) as a white solid.

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 8.49 (t, *J* = 5.7 Hz, 1H), 8.28 (dt, *J* = 8.3, 1.0 Hz, 1H), 8.21 (dt, *J* = 8.3, 1.0 Hz, 1H), 7.81 (ddd, *J* = 8.3, 7.2, 1.0 Hz, 1H), 7.64 (ddd, *J* = 8.3, 7.2, 1.0 Hz, 1H), 7.06 (t, *J* = 6.1 Hz, 1H), 4.96 (d, *J* = 5.7 Hz, 2H), 3.69 (d, *J* = 6.1 Hz, 2H), 1.39 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{15}H_{19}N_5NaO_4^+$ ($[M+Na]^+$): 356.1; found: 356.1.

***tert*-butyl (2-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-2-oxoethyl)carbamate (137a):**

The synthesis of **137a** was carried out similar to known literature procedures.[343]

To a solution of **136a** (0.85 g, 3.1 mmol, 1.0 eq) and 7-amino-4-methylcoumarin (0.54 g, 3.1 mmol, 1.0 eq) in toluene (61 mL), TFA (0.23 mL, 3.1 mmol, 1.0 eq) was added dropwise and the reaction mixture stirred o.n. at rt.

Next, toluene was evaporated under reduced pressure and the crude redissolved in EtOAc (300 mL). After washing with aqueous saturated $NaHCO_3$ (1 × 100 mL) and brine (1 × 100 mL), the organic phase was dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica. Purification via flash column chromatography (Pent/EtOAc gradient, 2.5:1 → 1:5) afforded the title compound **137a** (0.83 g, 2.5 mmol, 82%) as a slightly yellow solid.

R_f = 0.20 (Pent:EtOAc, 1:1) [UV, ninhydrin].

^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ[ppm] = 10.38 (s, 1H), 7.74 (d, *J* = 2.1 Hz, 1H), 7.72 (d, *J* = 8.7 Hz, 1H), 7.48 (dd, *J* = 8.7, 2.1 Hz, 1H), 7.10 (t, *J* = 6.1 Hz, 1H), 6.26 (q, *J* = 1.3 Hz, 1H), 3.77 (d, *J* = 6.1 Hz, 2H), 2.40 (d, *J* = 1.3 Hz, 3H), 1.40 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{17}H_{21}N_2O_5^+$ ($[M+H]^+$): 333.1; found: 333.1.

Note that the extraction was troublesome due to solubility issues and a small amount of MeOH was typically added. Different polar solvent systems such as a 3:1 mixture of EtOAc:THF could also be tested instead.

***tert*-butyl (2-((2-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-2-oxoethyl)amino)-2-oxo-ethyl)carbamate (137b):**

The synthesis of **137b** was carried out similar to known literature procedures.[343]

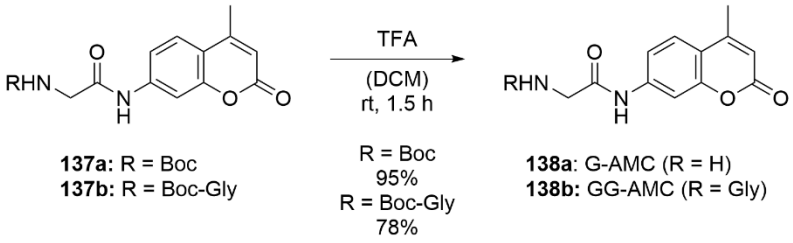
To a solution of **136b** (0.13 g, 0.38 mmol, 1.0 eq) and 7-amino-4-methylcoumarin (67 mg, 0.38 mmol, 1.0 eq) in toluene (7.6 mL), TFA (29 µL, 0.38 mmol, 1.0 eq) was added dropwise and the reaction mixture stirred o.n. at rt. Next, toluene was evaporated under reduced pressure and the crude redissolved in EtOAc (100 mL). After washing with aqueous saturated $NaHCO_3$ (1 × 40 mL) and brine (1 × 40 mL), the organic phase was dried over Na_2SO_4, filtered, concentrated and dry-loaded onto silica. Purification via flash column chromatography (DCM/MeOH gradient, 24:1 → 23:2) gave the target compound **137b** (72 mg, 0.19 mmol, 49%) as a slightly yellow solid.

R_f = 0.43 (DCM:MeOH, 9:1) [UV, ninhydrin].

^{1}H NMR (400 MHz, DMSO-d_6, 300 K): δ[ppm] = 10.31 (s, 1H), 8.20 (t, *J* = 5.8 Hz, 1H), 7.76 (d, *J* = 2.0 Hz, 1H), 7.73 (d, *J* = 8.6 Hz, 1H), 7.49 (dd, *J* = 8.6, 2.0 Hz, 1H), 7.08 (t, *J* = 6.0 Hz, 1H), 6.27 (d, *J* = 1.4 Hz, 1H), 3.95 (d, *J* = 5.8 Hz, 2H), 3.61 (d, *J* = 6.0 Hz, 2H), 2.40 (d, *J* = 1.4 Hz, 3H), 1.39 (s, 9H).

LRMS-ESI (m/z): calc. for $C_{19}H_{23}N_3NaO_6^+$ ($[M+Na]^+$): 412.1; found: 412.1.

Note that the extraction was troublesome due to solubility issues and a small amount of MeOH was typically added. Different polar solvent systems such as a 3:1 mixture of EtOAc:THF could also be tested instead.



2-amino-*N*-(4-methyl-2-oxo-2H-chromen-7-yl)acetamide (138a, G-AMC): *N*-Boc deprotection of **137a** according to standard procedure G.

Deprotection of **137a** (0.83 g, 2.51 mmol) with excess TFA (10 mL) in DCM (40 mL) gave the TFA-salt **138a** as an off-white solid (0.83 g, 2.39 mmol, 95%) which was used without further purification.

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 7.81 – 7.74 (m, 2H), 7.45 (dd, *J* = 8.6, 2.1 Hz, 1H), 6.30 (d, *J* = 1.3 Hz, 1H), 3.84 (s, 2H), 2.41 (d, *J* = 1.3 Hz, 3H).

^{13}C NMR (75 MHz, DMSO-d_6, 300 K): δ[ppm] = 165.8 (s, 1C), 159.9 (s, 1C), 153.7 (s, 1C), 153.1 (s, 1C), 141.3 (s, 1C), 126.3 (s, 1C), 115.6 (s, 1C), 115.2 (s, 1C), 112.7 (s, 1C), 105.8 (s, 1C), 41.3 (s, 1C), 18.0 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{12}H_{13}N_2O_3^+$ ([M+H]$^+$): 233.1; found: 233.1.

2-amino-*N*-(2-((4-methyl-2-oxo-2H-chromen-7-yl)amino)-2-oxoethyl)acetamide (138b, GG-AMC): *N*-Boc deprotection of **137b** according to standard procedure G.

Deprotection of **137b** (72 mg, 0.19 mmol) with excess TFA (2 mL) in DCM (3 mL) gave the TFA-salt **138b** as an off-white solid (60 mg, 0.15 mmol, 78%) which was used without further purification.

^{1}H NMR (300 MHz, DMSO-d_6, 300 K): δ[ppm] = 10.58 (s, 1H), 8.79 (t, *J* = 5.7 Hz, 1H), 8.08 (s, br, 3H), 7.77 (d, *J* = 2.0 Hz, 1H), 7.73 (d, *J* = 8.7 Hz, 1H), 7.46 (dd, *J* = 8.7, 2.0 Hz, 1H), 6.27 (t, *J* = 1.3 Hz, 1H), 4.06 (d, *J* = 5.7 Hz, 2H), 3.67 (s, 2H), 2.40 (d, *J* = 1.3 Hz, 3H).

^{13}C NMR (75 MHz, DMSO-d_6, 300 K): δ[ppm] = 168.0 (s, 1C), 166.6 (s, 1C), 160.0 (s, 1C), 153.7 (s, 1C), 153.1 (s, 1C), 142.0 (s, 1C), 126.1 (s, 1C), 115.2 (s, 2C*), 112.4 (s, 1C), 105.6 (s, 1C), 42.8 (s, 1C), 40.1 (s, 1C), 18.0 (s, 1C).

LRMS-ESI (m/z): calc. for $C_{14}H_{16}N_3O_4^+$ ([M+H]$^+$): 290.1; found: 290.1.

9.6 Solid Phase Peptide Synthesis

The peptides **61**, KETWV, **112a** and **b** were synthesized via solid phase peptide synthesis (SPPS). SPPS was performed in plastic syringes equipped with a frit for small scale preparations (max. 1 g) according to standard Fmoc-strategies for solid phase synthesis using CTC-resin. Reaction vessels were rolled at rt during each individual step for the indicated time.

Loading of CTC Resin

The CTC resin (1.2-1.3 mmol/g maximal loading capacity, 1.0 eq) was suspended in DCM (30 mL/g resin) and left to swell at rt for 5 min in a syringe equipped with a frit Then, the Fmoc-protected amino acid (2.3 eq) and DIPEA (4.0 eq) dissolved in anhydrous DCM (30 mL/g resin) were applied to the resin and the reaction mixture was rolled at rt for 1 h. The resin was subsequently washed with DCM (5×) and DMF (5×) to then proceed with the on-resin Fmoc-deprotection.

On-Resin Fmoc-Deprotection

The resin was then incubated with 20 % (v/v) piperidine in DMF, rolling at rt for 20 min. This step was repeated a second time using fresh 20 % (v/v) piperidine in DMF and the resin was subsequently washed with DMF (10×).

On-Resin Coupling of Amino Acids and Electrophilic Warheads

To couple amino acids or electrophilic warheads, the desired Fmoc-protected amino acid or electrophilic warhead (2.0 eq or 1.5 eq, respectively) was reacted with DIC (1.9 eq), HOBt×H_2O (2.0 eq) and DIPEA (2.0 eq or none, respectively) in anhydrous DMF and stirred at rt for 20 min before being applied to the resin.

The resulting reaction mixture was rolled at rt until completion, checking the reaction progress by TLC and LC-MS. After completion, the resin was washed with DMF (5×) to proceed with the on-resin Fmoc-deprotection or cleavage from the CTC-resin.

Cleavage From the CTC Resin

The resin was first washed with DCM (5×) and subsequently incubated with 20 % (v/v) HFIP in DCM at rt for 15 min. The filtrate was collected and the cleavage procedure repeated once more. Afterwards, the resin was washed with DCM (4×), combining the organic washes with the filtrate. The solvent was removed under reduced pressure to give the desired peptide bearing side-chain and *N*-terminal protecting groups. The crude product was typically purified by column chromatography with a DCM/MeOH gradient containing 0.5 % (v/v) AcOH.

9.7 Biology – General Remarks

All experiments were performed under sterile conditions using nitrile gloves. Benches were wiped using 80% EtOH in H_2O prior to experiments, which were performed close to a Bunsen burner flame.

9.7.1 Assay, Protein Purification and Storage Buffers

All protein purification and storage buffers were supplemented with 0.5 mM TCEP prior to purifying amber suppressed proteins and/or proteins bearing cysteine mutations. The pH was determined at 4 °C if not otherwise indicated.

Table | 9.1: Composition of assay, protein purification and storage buffers.

buffer	concentration/quantity of components
Affibody storage buffer	20 mM Tris-HCl (pH 7.5 at rt), 100 mM NaCl
Cam assay buffer	10 mM Tris-HCl (pH 7.5 at rt)
Cam storage buffer	50 mM Tris-HCl (pH 8.0), 150 mM NaCl
ClpP assay buffer	25 mM HEPES (pH 7.6), 200 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 10 % (v/v) glycerol
ClpP storage buffer	20 mM HEPES (pH 7.0), 100 mM NaCl
GST elution buffer	20 mM Tris-HCl (pH 7.5 at rt), 100 mM NaCl, 2 mM DTT, 2 mM reduced GSH
GST lysis buffer	20 mM Tris-HCl (pH 7.5 at rt), 100 mM NaCl, 2 mM DTT, 100 mM PMSF

GST storage buffer	20 mM Tris-HCl (pH 7.5 at rt), 100 mM NaCl
GST wash buffer	20 mM Tris-HCl (pH 7.5 at rt), 100 mM NaCl, 2 mM DTT
His Cam elution buffer	20 mM Tris-HCl (pH 8.0), 300 mM imidazole (pH 8.0), 300 mM NaCl, 10 % (v/v) glycerol, 5 mM $CaCl_2$
His Cam lysis buffer	20 mM Tris-HCl (pH 8.0), 30 mM imidazole (pH 8.0), 300 mM NaCl, 0.5 mM PMSF, 10 % (v/v) glycerol, 5 mM $CaCl_2$
His Cam washing buffer	20 mM Tris-HCl (pH 8.0), 30 mM imidazole (pH 8.0), 300 mM NaCl, 10 % (v/v) glycerol, 5 mM $CaCl_2$
His standard elution buffer	20 mM Tris-HCl (pH 8.0), 300 mM imidazole (pH 8.0), 300 mM NaCl
His standard lysis buffer	20 mM Tris-HCl (pH 8.0), 30 mM imidazole (pH 8.0), 300 mM NaCl, 0.5 mM PMSF
His standard washing buffer	20 mM Tris-HCl (pH 8.0), 30 mM imidazole (pH 8.0), 300 mM NaCl
His standard storage buffer	20 mM Tris-HCl (pH 8.0), 300 mM NaCl
MID1sc10 assay buffer	40 mM HEPES (pH 8.0), 50 mM NaCl, 50 µM $ZnSO_4$
MID1sc10 storage buffer	40 mM HEPES (pH 8.0), 150 mM NaCl
Pull-down (affibody) buffer	20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 2 mM DTT, 0.2 % (v/v) Tween20
Pull-down (nb) buffer	50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM TCEP, 0.2 % (v/v) Tween20
Strep elution buffer	100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin
Strep ClpP lysis buffer	100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA
Strep lysis buffer	100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, 0.5 mM PMSF
Strep washing buffer	100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA
PDZ3 assay buffer	10 mM phosphate (pH 7.4)

9.7.2 Buffers and Media

All buffers and media were stored at 4 °C or room temperature if not indicated otherwise.

Table | 9.2: Composition of general buffers.

buffer	concentration/quantity of components
DNA loading dye (10×)	40 % (w/v) sucrose, 0.15% (w/v) orange G, 0.05 % (w/v) xylene cyanole, 0.05 % (w/v) bromophenol blue

Hybridization buffer (SLIM buffer, 5×)	125 mM Tris-HCl (pH 9.0), 750 mM NaCl, 100 mM (EDTA)
Q5 reaction buffer (5×)	100 mM Tris-HCl (pH 8.8), 50 mM KCl, 5 mM $MgCl_2$, 50 mM ammonium sulfate
Laemmli loading buffer (4×)	240 mM Tris-HCl (pH 6.8), 5 % (v/v) BME, 40 % (v/v) glycerol, 8.4 % (w/v) SDS, 0. 4% (v/v) BPB
SDS-PAGE MES running buffer (20×)	1 M MES (pH 7.3), 1 M Tris (pH 7.3), 2 % (w/v) SDS, 20 mM EDTA
SDS-PAGE resolving buffer (4×)	0.5 M Tris-HCl (pH 6.8), 0.4 % (w/v) SDS
SDS-PAGE stacking buffer (4×)	1.5 M Tris-HCl (pH 8.8), 0.4 % (w/v) SDS
TBS buffer (10×)	200 mM Tris-HCl (pH 7.6), 1.5 M NaCl
TBS-T buffer	20 mM Tris-HCl (pH 7.6), 150 mM NaCl, 0.1 % (v/v) Tween20
Western blot transfer buffer	48 mM Tris (pH 8.9 – 9.4), 39 mM glycine, 1.3 mM SDS, 20 % (v/v) MeOH

Table | 9.3: Composition of media.

medium	**concentration/quantity of components**
2× YT medium	16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl (pH 7.0)
Autoinduction (AI) medium[349]	0.25% aspartate; 0.5% glycerol; 4% (v/v) 18 amino acid mix (25×); 2% (v/v) M solution (50×); 40 µg/mL leucine; 0.05% arabinose; 2 mM $MgSO_4$; 0.05% glucose; 0.02% (v/v) trace metal mix (5000×)
LB agar	10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 15 g/L agar-agar
SOC medium	20 g/L tryptone, 5 g/L yeast extract, 0.96 g/L $MgCl_2$, 0.5 g NaCl, 0.186 g/L KCl, 20 mM glucose, stored at -20 °C

9.7.3 *E. coli* strains

Table9.4 | Bacterial strains used within this work.

strain	**genotype**	**purpose**	**supplier**
NEB® 10-beta	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14-ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	cloning	NEB

K12 (BW25113)	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514*	protein expression, amber suppression	former GE Healthcare
K12 dnfsA (JW0835) – part of the KEIO collection[266, 350] (BW25113 derivative)	*F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, ΔnfsA752::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514*	amber suppression	Horizon Discovery
K12 dnfsB (JW0567) – part of the KEIO collection[266, 350] (BW25113 derivative)	*F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), ΔnfnB787::kan, λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514*	amber suppression	Horizon Discovery
K12 (dnfsA + dnfsB)	JW0835 derivative	amber suppression	M. von Wrisberg (Lang lab)
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	OP expression	NEB

9.7.4 Oligonucleotides

Oligonucleotides were designed using the software *SnapGene 5.06* by GSL Biotech LLC (now Dotmatics) and the browser-based tool *Tm Calculator v1.16.4* by NEB. All oligonucleotides were purchased desalted and lyophilized from Microsynth and Sigma Aldrich. New genes were ordered as strings supplied by Thermo or Twist Bioscience.

9.7.5 Plasmids

The plasmids used in this work were either already available within the lab or cloned during this project. The following tables list all plasmids used within this work.

Synthetases

The pEVOL plasmids listed below harbor a PylT copy under a constitutive promotor and two copies of the respective *Mm* PylRS, one under a constitutively active GlnS and the second under an araBAD promotor in which the overexpression is induced by arabinose.

Table 9.5 | List of plasmids bearing the respective PylRS used within this work.

designation	plasmid	resistance
MW-S59	pEVOL_*Mm*-wt-PylRS	Cam
MW-S62	pEVOL_*Mm*-AzoA-PylRS	Cam
MW-S63	pEVOL_*Mm*-AzoB-PylRS	Cam
MW-S64	pEVOL_*Mm*-AzoC-PylRS	Cam
KK4	pEVOL_*Mm*-KK4-PylRS	Cam

KK10	pEVOL_*Mm*-KK10-PylRS	Cam
KK11	pEVOL_*Mm*-KK11-PylRS	Cam
KK13	pEVOL_*Mm*-KK13-PylRS	Cam

Target Proteins

The pBAD plasmids listed below harbor the POI under an araBAD promotor in which the overexpression was induced by arabinose. The pET29b(+) plasmids listed below harbor the POI under a T7 promotor in which the overexpression was induced by IPTG.

Table 9.6 | List of plasmids bearing the respective POI used within this work.

designation	plasmid	resistance
SM14	pBAD_sfGFP-N150TAG-His_6	Amp
KK-P73	pBAD_wt-Cam-His_6	Amp
KK-P78	pBAD_Cam-M76TAG-E83C-His_6	Amp
KK-P62	pBAD_wt-affibody-His_6	Amp
KK-P65	pBAD_affibody-Q26C-S33TAG-His_6	Amp
KK-P106	pBAD_GST-GGGS-Z domain	Amp
KK-P107	pBAD_wt-affibody-GSWGS-His_6	Amp
KK-P108	pBAD_affibody-Q26C-S33TAG-GSWGS-His_6	Amp
MW-7	pBAD_His_6-Throm-Rap80 7A-R97TAG-TEV-StrepII	Amp
MW-10	pBAD_His_6-Throm-Rap80 7A-R97TAG-E104C-TEV-StrepII	Amp
KK-P109	pBAD_His_6-Throm-Rap80-R97TAG-E104C-TEV-StrepII	Amp
KK-P169	pBAD_His_6-TEV-wt-PDZ3	Amp
KK-P170	pBAD_wt-PDZ3-His_6	Amp
KK-P171	pBAD_PDZ3-E395TAG-A402C-His_6	Amp
KK-P172	pBAD_PDZ3-E395C-A402TAG-His_6	Amp
KK-P175	pBAD_pelB-nb(min)-Y116TAG-His_6	Amp
KK-P176	pBAD_pelB-nb(max)-Y37TAG-His_6	Amp
KK-P178	pBAD_pelB-wt-nb(max)-His_6	Amp
KK-P83	pBAD_AltTPase-D32TAG-D39C-GGGS-StrepII	Amp
KK-P84	pBAD_AltTPase-D32C-D39TAG-GGGS-StrepII	Amp
KK-P120	pBAD_His_6-MBP-TEV-wt-MID1sc10	Amp
KK-P121	pBAD_His_6-TEV-wt-MID1sc10	Amp
KK-P122	pBAD_wt-MID1sc10-His_6	Amp
KK-P128	pBAD_wt-MID1sc10-TEV-His_6	Amp
KK-P129	pBAD_wt-MID1sc10-CPD-His_6	Amp

KK-P131	pBAD_MID1sc10-E26C-N33TAG-CPD-His_6	Amp
KK-P164	pBAD_His_6-MBP-TEV-MID1sc10-Y41TAG-Y90C	Amp
KK-P168	pBAD_MID1sc10-Y41TAG-Y90C-CPD-His_6	Amp
KK-P11	pBAD_ClpP-N141TAG-His_6	Amp
KK-P12	pBAD_ClpP-N141TAG-T134C-His_6	Amp
KK-P13	pBAD_ClpP-N141TAG-E137C-His_6	Amp
KK-P14	pBAD_ClpP-N141TAG-K145C-His_6	Amp
KK-P15	pBAD_ClpP-R152TAG-His_6	Amp
KK-P16	pBAD_ClpP-R152TAG-N141C-His_6	Amp
KK-P17	pBAD_ClpP-R152TAG-K145C-His_6	Amp
MW-P1	pBAD_wt-ClpP-StrepII	Amp
KK-P34	pBAD_ClpP-T134C-N141C-StrepII	Amp
KK-P35	pBAD_ClpP-E137C-N141C-StrepII	Amp
KK-P36	pBAD_ClpP-N141C-K145C-StrepII	Amp
KK-P37	pBAD_ClpP-N141C-R152C-StrepII	Amp
KK-P38	pBAD_ClpP-K145C-R152C-StrepII	Amp
KK-P54	pBAD_ClpP-T134C-StrepII	Amp
KK-P55	pBAD_ClpP-E137C-StrepII	Amp
KK-P57	pBAD_ClpP-K145C-StrepII	Amp
KK-P58	pBAD_ClpP-R152C-StrepII	Amp
KK-P59	pBAD_ClpP-N141TAG-R152C-StrepII	Amp
KK-P60	pBAD_ClpP-K145TAG-R152C-StrepII	Amp
KK-P9	pBAD_ClpP-N141C-R152TAG-StrepII	Amp
KK-P10	pBAD_ClpP-K145C-R152TAG-StrepII	Amp
KK-P110	pBAD_wt-OP-His_6	Amp
TE-OP2	pET29b(+)_wt-OP-His_6	Kan
KK-P140	pET29b(+)_sfGFP-N150TAG-His_6	Kan
KK-P144	pET29b(+)_OP-G51TAG-E58C-His_6	Kan
KK-P145	pET29b(+)_OP-G51C-E58TAG-His_6	Kan
KK-P146	pET29b(+)_OP-Q55TAG-S62C-His_6	Kan
KK-P147	pET29b(+)_OP-Q55C-S62TAG-His_6	Kan

The pRSF plasmids listed below harbor the POI under a T7 promotor in which the overexpression was induced by IPTG, as well as the PylT under a constitutive promotor and one copy of the respective *Mm* PylRS under a constitutive promotor.

Table 9.7 | List of pRSF-vectors from this work, which are bearing the respective PylRS besides the POI.

designation	plasmid	resistance
OA157	pRSF_sfGFP-N150TAG-His_6_*Mb*-wt-PylRS	Kan
pRSF-KK6	pRSF_sfGFP-N150TAG-His_6_*Mm*-wt-PylRS	Kan

9.7.6 Stock Solutions

All stock solutions were dissolved in H_2O and sterilized via filtration using a 0.2 µm sterile membrane filter and stored at 4 °C if not indicated otherwise.

Table | 9.8: Composition of stock solutions.

stock	concentration/quantity of components
18 amino acid mix (25×)[349]	5 mg/mL of each of the following amino acids: glutamic acid sodium salt, aspartic acid, lysine×HCl, arginine×HCl, alanine, histidine×HCl×H_2O, proline, glycine, valine, threonine, serine, leucine, glutamine, asparagine×H_2O, isoleucine, phenylalanine, tryptophan and methionine
APS-solution (10%)	10% (w/v) ammonium persulfate, stored at -20 °C
Arabinose solution (20%)	20% (w/v) arabinose
Aspartate solution (5%)	5% (w/v) aspartic acid (pH 7.5)
M solution (50×)	1.25 M Na_2HPO_4; 1.25 M KH_2PO_4; 2.5 M NH_4Cl; 0.25 M Na_2SO_4, stored at 60 °C
Trace metal mix (5000×)[349]	20 mM $CaCl_2$×$2H_2O$, 10 mM $MnCl_2$×$4H_2O$, 10 mM $ZnSO_4$×$7H_2O$, 2 mM $CoCl_2$×$6H_2O$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4×$2H_2O$, 2 mM Na_2SeO_3, 2 mM H_3BO_3, 50 mM $FeCl_3$

All antibiotics were dissolved and sterilized via filtration using a 0.2 µm sterile membrane filter and stored at -20 °C.

Table 9.9 | Antibiotic stock solutions and respective working concentrations.

antibiotic	stock concentration [mg/mL]	working concentration [µg/mL]	solvent
Ampicillin (Amp)	100	100	H_2O
Chloramphenicol (Cam)	50	50	96 % (v/v) EtOH
Kanamycin (Kan)	50	50	H_2O
Tetracycline (Tet)	12.5	12.5	70 % (v/v) EtOH

9.8 Microbiology Procedures

9.8.1 Preparation of Chemically Competent *E. coli* Cells

The desired strain of chemically competent *E. coli* was inoculated into 50 mL 2× YT medium and incubated o.n. at 37 °C and 200 rpm. The pre-culture was then diluted into 100 mL 2× YT to an OD_{600} of 0.04 and subsequently incubated at 37 °C and 200 rpm until an OD_{600} of 0.5-0.6 was reached. The culture was chilled on ice for 30 min and harvested via centrifugation (4000× g, 10 min, 4 °C). The obtained pellet was carefully resuspended in 40 mL ice-cold 100 mM $CaCl_2$ and harvested via centrifugation (4000× g, 10 min, 4 °C). Resuspension in 40 mL ice-cold 100 mM $CaCl_2$ was repeated and the resulting cell suspension chilled on ice for 30 min. Cells were harvested (4000× g, 10 min, 4 °C) and resuspended in 5 mL 100 mM $CaCl_2$. After the addition of 1 mL 80 % (v/v) glycerol the cell suspension was incubated on ice for 5 min, aliquoted into precooled 1.5 mL Eppendorf tubes (100 µL/tube) and immediately flash frozen using liquid nitrogen. Chemically competent *E. coli cells* were stored at -80 °C until further use. For the quality control of each batch of competent cells, the absence of contamination was assessed by cultivation on LB agar plates containing Amp, Cam, Kan, Tet or no antibiotics.

9.8.2 Heat-Shock Transformation of Chemically Competent *E. coli* Cells

For heat-shock transformation 20-100 ng of the desired plasmid DNA were added to freshly thawed, chemically competent *E. coli* cells. In the case of co-transformations equimolar amounts of each plasmid DNA was added. The mixture was gently tapped for homogenization and incubated on ice for 30 min. Afterwards the cells were heat-shocked for 45 s at 42 °C (water bath) and incubated on ice for 5 min before the addition of 700 µL SOC medium. The rescued cells were incubated at 37 °C, shaking at 200 rpm for 1 h. Finally, cells were either inoculated in fresh 2× YT medium with full strength antibiotic(s) or plated onto LB agar plates supplemented with the corresponding antibiotic(s).

9.9 Molecular Biology Procedures

9.9.1 Polymerase Chain Reaction (PCR)

PCRs were applied to amplify double stranded DNA used in restriction digests and Gibson assembly. Point mutations, deletions and small insertions were introduced via PCR using the site-directed ligase independent mutagenesis (SLIM) approach.[351] In general, PCRs were exerted in a total volume of 50 µL containing 200 µM dNTP mix (NEB), 500 nM of forward and reverse primer, approx. 20 ng template and 0.5 µL Q5 polymerase (NEB) in 1x Q5 reaction buffer. Annealing temperatures (T_a) were calculated using the browser-based tool *Tm Calculator v1.16.4* by NEB.

Table 9.10 | Standard protocol for PCR using Q5 polymerase.

step	temperature [°C]	time [s]	cycles
Initial denaturation	98	60	1
Denaturation	98	20	35
Annealing	T_a (-3 for SLIM)	30	
Elongation	72	30 per kb	
Final elongation	72	300	1

Large pieces of double stranded DNA (> 5 kb) for subsequent restriction digest or Gibson assembly were oftentimes amplified using a modified touchdown PCR protocol.

Table 9.11: Touchdown PCR protocol using Q5 polymerase.

step	temperature [°C]	time [s]	cycles
Initial denaturation	98	60	1
Denaturation	98	20	10
Annealing	T_a (+5, -0.5/cycle)	30	
Elongation	72	30 per kb	
Denaturation	98	20	25
Annealing	T_a	30	
Elongation	72	30 per kb	
Final elongation	72	300	1

Afterwards, PCRs were kept at 12 °C until further use.

9.9.2 Molecular Cloning

New genes were cloned into respective backbones using standard restriction cloning protocols or Gibson assembly (enzymes, master mix solutions and protocols by NEB were applied as recommended). Point mutations, deletions and small insertions were introduced using site-directed ligase independent mutagenesis (SLIM).[351]

9.9.3 DNA Isolation and Purification

Plasmid DNA was isolated from *E. coli* o.n. cultures grown in 8 mL 2× YT medium using the *peqGOLD Plasmid Miniprep Kit I* (VWR) or the *ZR Plasmid Miniprep – Classic* (Zymo Research) according to the manufacturer's instructions. Plasmid DNA was eluted with 30-50 µL nuclease free water.

Double stranded DNA from agarose gels was purified via gel extraction using the *Monarch® DNA gel extraction kit* (NEB) or the *Zymoclean™ Gel DNA Recovery Kit* (Zymo Research) according to the manufacturer's instructions. Double stranded DNA was eluted with 30-50 µL nuclease free water.

Standard PCR purification was performed using a *PCR purification kit* by Jena Bioscience or the *DNA Clean & Concentrator™-5* (Zymo Research) according to the manufacturer's instructions. Double stranded DNA was eluted with 30-50 µL nuclease free water.

All DNA concentrations and the purity were measured photometrically at the wavelengths 230 nm, 260 nm and 280 nm using a *NanoPhotometer® N60* (Implen). Purified plasmids and DNA were stored at -20 °C until further use.

9.9.4 Agarose Gel Electrophoresis

Size-dependent separation of DNA samples was performed using agarose gel electrophoresis. To obtain 1% agarose gels, solid agarose was dissolved in the respective amount of 1× TAE buffer using a microwave followed by the addition of *SERVA DNA Stain Clear G* (Serva). DNA samples were mixed with DNA loading dye (10×) and electrophoreses run with constant voltage of 140 V for 25-35 min using the *High range ladder* (Jena Biosciences) as size standard.

9.9.5 DNA Sequencing

Sanger sequencing of plasmid DNA was performed to verify cloning. To this end, 7.5 µL plasmid DNA sample (50 – 100 ng/µL) was mixed with 2.5 µL of a 10 µM sequencing primer solution and sent to Genewiz using the overnight premix service.

9.10 Protein Biochemistry Procedures

Cell growth during protein expressions was checked on a *Ultrospec® 10 Cell Density Meter* (Biochrom), measuring the density of cell suspensions at λ = 600 nm. The *Ultrasonics™ Sonifier™ SFX250 Cell Disruptor* (Branson) was used for ultrasonic homogenisation of cell suspensions. Protein concentrations were either measured on the spectrophotometer *NanoPhotometer® N60* (Implen) using the Protein UV mode (absorption at λ = 280 nm) or via Biuret reaction on a *Varioskan LUX Multimode Microplate Reader* using the *Pierce™ BCA Protein Assay Kit*, both supplied by ThermoFisher.

It should be noted, that affinity chromatography purification of POIs amber suppressed with azo-ncAAs has to be performed manually using the respective affinity beads. Purification using prepacked *HiTrap* columns (Cytiva) on the Äkta led to precipitation of residual azo-ncAA contained in the lysate and thus blockage of the column.

9.10.1 Standard Protein Expression and Purification of sfGFP-N149TAG-His_6

Chemically competent *E. coli* K12 or K12 (dnfsA + dnfsB) were transformed with pBAD_sfGFPN149TAG-His_6 and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. The cells were rescued by incubation in 700 µL SOC medium at 37 °C, shaking at 200 rpm for 1 h. Afterwards the cell suspension was incubated in 50 mL 2× YT medium containing ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL) at 37 °C, shaking at 200 rpm. For K12 (dnfsA + dnfsB), kanamycin (25 µg/mL) was additionally added to the medium. The overnight culture was diluted into fresh autoinducing medium supplemented with ampicillin (50 µg/mL) and chloramphenicol (25 µg/mL) to an OD_{600} of 0.04 and subsequently cultured at 37 °C, shaking at 200 rpm until OD_{600} = 0.3. Then, the ncAA was added to a final concentration of typically 1-2 mM (0.2 mM for tripeptide). For azo-ncAAs, 2 % (v/v) Tween20 was also added to the cell suspension to enhance the solubility of azo-ncAAs. Protein expression was automatically induced at an OD_{600} around 0.6 by arabinose contained in the autoinduction medium and cultures were incubated shaking o.n. (typically 16-18 h) at 37 °C. The next day, 0.5 mL samples were harvested by centrifugation (4,000×g, 10 min, 4 °C) and analyzed by 15% SDS-PAGE.

For protein purification, cells were harvested by centrifugation (4,000×g, 10 min, 4 °C), the supernatant was discarded and the cell pellet resuspended in His standard lysis buffer (20 mM Tris-HCl pH 8.0, 30 mM imidazole pH 8.0, 300 mM NaCl, 0.5 mM PMSF). The cell suspension was incubated on ice for 30 min and then lysed on ice via sonication. After centrifugation at 12,000×g, 20 min, 4 °C, the cleared cell lysate was added to washed *Ni^{2+}-NTA agarose* (Jena Bioscience) and incubated with agitation for 1 h at 4 °C. The mixture was then given onto a plastic column, washed with 6 CV of His standard washing buffer (20 mM Tris-HCl pH 8.0, 30 mM imidazole pH 8.0, 300 mM NaCl) followed by elution of the protein with His standard elution buffer (20 mM Tris-HCl pH 8.0, 300 mM imidazole pH 8.0, 300 mM NaCl). The process of protein purification was checked by 15% SDS-PAGE. Purified proteins were concentrated using *Amicon® Ultra* centrifugal filter units with 10 kDa MWCO (Millipore), rebuffered into His standard storage buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl) and stored at -80 °C until further use.

9.10.2 Expression and Purification of Calmodulin-His_6 Variants

Transformation of *E. coli* K12 (dnfsA + dnfsB) was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_Cam-M73TAG-E83C-His_6 and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression of Cam variants was also performed according to the standard protein expression procedure described in Chapter 9.10.1.

For protein purification, harvested cells were resuspended in His Cam standard lysis buffer (20 mM Tris-HCl pH 8.0, 30 mM imidazole pH 8.0, 300 mM NaCl, 10 % (v/v) glycerol, 5 mM $CaCl_2$, 0.5 mM PMSF, 0.5 mM TCEP). The cell suspension was incubated on ice for 30 min and then lysed on ice via sonication. After centrifugation at 12,000×g, 20 min, 4 °C, the cleared cell lysate was added to washed *Ni Sepharose™ High Performance* (Cytiva) and incubated with agitation for 1 h at 4 °C. The mixture was then given onto a plastic column, washed with 6 CV of His Cam standard washing buffer (20 mM Tris-HCl pH 8.0, 30 mM imidazole pH 8.0, 300 mM NaCl, 10 % (v/v) glycerol, 5 mM $CaCl_2$, 0.5 mM TCEP) followed by elution of the protein with His Cam standard elution buffer (20 mM Tris-HCl pH 8.0, 300 mM imidazole pH 8.0, 300 mM NaCl, 10 % (v/v) glycerol, 5 mM $CaCl_2$, 0.5 mM TCEP). Elution fractions were concentrated using *Amicon® Ultra* centrifugal filter units with 10 kDa MWCO (Millipore), rebuffered into Cam storage buffer containing EDTA (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP) and subsequently purified via SEC using a Superdex® Increase 75 10/300 GL (Cytiva) with Cam storage buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM TCEP). The process of protein purification was checked by 15% SDS-PAGE. Fractions containing Cam were pooled together, concentrated employing *Amicon® Ultra* centrifugal filter units with 10 kDa MWCO (Millipore) and stored at -80 °C until further use.

9.10.3 Expression and Purification of Affibody-His_6 Variants

Transformation of *E. coli* K12 or K12 (dnfsA + dnfsB) was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_affibody-His_6 bearing the respective affibody mutant, and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression of affibody variants was performed similar to the standard protein expression procedure described in Chapter 9.10.1. After 6 h of protein expression, 0.5 mL samples were harvested by centrifugation (4,000×g, 10 min, 4 °C) and analyzed by 15% SDS-PAGE. Consecutively, anti-His Western Blot analysis was performed.

Protein purification via affinity chromatography of affibody variants was performed according to the standard protein purification procedure described in Chapter 9.10.1. Elution fractions were concentrated using *Amicon® Ultra* centrifugal filter units with 3 kDa MWCO (Millipore), rebuffered into His standard storage buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.5 mM TCEP) and subsequently purified via SEC using a *Superdex® Increase 75 10/300 GL* (Cytiva) with affibody storage buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.5 mM TCEP). The process of protein purification was checked by 15% SDS-PAGE. Fractions containing affibody were pooled together, concentrated employing *Amicon® Ultra* centrifugal filter units with 3 kDa MWCO (Millipore) and stored at -80 °C until further use.

9.10.4 Expression and Purification of GST-Z-Domain

Chemically competent *E. coli* K12 were transformed with pBAD_GST-GGGS-Z-domain and rescued by incubation in 700 µL SOC medium at 37 °C, shaking at 200 rpm for 1 h. Afterwards the cell suspension was inoculated in 50 mL 2×YT medium containing ampicillin (100 µg/mL) at 37 °C, shaking at 200 rpm. The overnight culture was diluted into fresh 2×YT medium supplemented with ampicillin (50 µg/mL) to an OD_{600} of 0.04 and subsequently cultured at 37 °C. Protein expression was induced at an OD_{600} around 0.6 by the addition of 0.02% (v/v) arabinose and the culture was incubated shaking o.n. (typically 16-18 h) at 37 °C.

The next day, cells were harvested by centrifugation (4,000×g, 10 min, 4 °C), the supernatant was discarded and the cell pellet resuspended in GST lysis buffer (20 mM Tris-HCl pH 7.5 at rt, 100 mM NaCl, 2 mM DTT, 100 mM PMSF). The cell suspension was incubated on ice for 30 min and then lysed on ice via sonication. After centrifugation at 12,000×g, 20 min, 4 °C, the cleared cell lysate was added to washed *Glutathione Sepharose® 4B* (Cytiva) and incubated with agitation for 1 h at 4 °C. The mixture was then given onto a plastic column, washed with 6 CV of GST washing buffer (20 mM Tris-HCl pH 7.5 at rt, 100 mM NaCl, 2 mM DTT) followed by elution of the protein with GST elution buffer (20 mM Tris-HCl pH 7.5 at rt, 100 mM NaCl, 2 mM DTT, 2 mM reduced GSH). Elution fractions were pooled and subsequently purified via SEC using a *Superdex® Increase 75 10/300 GL* (Cytiva) with GST storage buffer (20 mM Tris-HCl pH 7.5 at rt, 100 mM NaCl). The process of protein purification was checked by 15% SDS-PAGE. Fractions containing GST-GGGS-Z-domain were pooled together, concentrated employing *Amicon® Ultra* centrifugal filter units with 10 kDa MWCO (Millipore) and stored at -80 °C until further use.

9.10.5 Expression and Purification of His_6-Throm-Rap80-TEV-StrepII Variants

Transformation of *E. coli* K12 or K12 (dnfsA + dnfsB) was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_His_6-Throm-Rap80(81-124)-TEV-StrepII bearing the respective Rap80 variant or mutant, and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression of Rap80 variants was performed similar to the standard protein expression procedure described in Chapter 9.10.1. After 5 h of protein expression, 0.5 mL samples were harvested by centrifugation (4,000×g, 10 min, 4 °C) and analyzed by 15% SDS-PAGE. Consecutively, anti-His Western Blot analysis was performed.

For protein purification, cells were harvested by centrifugation (4,000×g, 10 min, 4 °C), the supernatant was discarded and the cell pellet resuspended in Strep lysis buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM PMSF, 0.5 mM TCEP). The cell suspension was incubated on ice for 30 min and then lysed on ice via sonication.

After centrifugation at 12,000×g, 20 min, 4 °C, the cleared cell lysate was added to washed *PureCube Magnetic HiCap StrepTactin®* (Cube Biotech) and incubated with agitation for 1 h at 4 °C. The mixture was then given onto a plastic column, washed with 6 CV of Strep washing buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP) followed by elution of the protein with Strep elution buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin, 0.5 mM TCEP). The process of protein purification was checked by 15% SDS-PAGE.

9.10.6 Expression and Purification of anti-GFP Nanobody-His_6 Variants

Transformation of *E. coli* K12 was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_pelB-nb-His_6 bearing the respective mutant and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression and purification via affinity-chromatography of nb variants was performed analog to the standard protein expression procedure described in Chapter 9.10.1. Elution fractions were concentrated using *Amicon® Ultra* centrifugal filter units with 10 kDa MWCO (Millipore), rebuffered into His standard storage buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl) and stored at -80 °C until further use.

9.10.7 Expression of AltTPase-StrepII Variants

Transformation of *E. coli* K12 was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_AltTPase-StrepII bearing the respective AltTPase mutant, and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression of AltTPase variants was performed analog to the standard protein expression procedure described in Chapter 9.10.1. After o.n. expression at 37 °C, 0.5 mL samples were harvested by centrifugation (4,000×g, 10 min, 4 °C) and analyzed by 15% SDS-PAGE. Consecutively, anti-Strep Western Blot analysis was performed.

9.10.8 Expression and Purification of PDZ3-His_6 Variants and His_6-TEV-PDZ3

Transformation of *E. coli* K12 (dnfsA + dnfsB) was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_PDZ3-His_6 bearing the respective mutant and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression and purification via affinity-chromatography of PDZ3 variants was also performed according to the standard protein expression procedure described in Chapter 9.10.1. To the combined elution fractions of His_6-TEV-PDZ3 was added TEV protease to a final concentration of 0.05 mg/mL and incubated at rt for 1 h under constant agitation. Progress of His_6-tag cleavage was checked via LC-MS and upon completion diluted (1:20) with His standard storage buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.5 mM TCEP). The digested protein mix was then added to washed *Ni^{2+}-NTA agarose* (Jena Bioscience) and incubated with agitation o.n. at 4 °C. The next day, the crude protein slurry was transferred onto a plastic column and the flow-through concentrated using *Amicon® Ultra* centrifugal filter units with 3 kDa MWCO (Millipore).

All PDZ3 variants were subsequently purified via SEC using a *Superdex® Increase 75 10/300 GL* (Cytiva) with PDZ3 storage buffer (10 mM phosphate pH 7.4). The process of protein purification was checked by 15% SDS-PAGE. Fractions containing PDZ3 were pooled together, concentrated employing *Amicon® Ultra* centrifugal filter units with 3 kDa MWCO (Millipore) and stored at -80 °C until further use.

It should be noted that a different buffer featuring a higher salt-content (e.g. 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM TCEP) would be recommended for potential future purifications via SEC to ensure the suppression of ion repulsion and thus proper peak shape and elution of the protein.

9.10.9 Expression and Purification of MID1sc10-His_6, MID1sc10-TEV-His_6, His_6-(MBP)-TEV-MID1sc10 and His_6-MBP-TEV-MID1sc10 Variants

Transformation of *E. coli* K12 or *E. coli* K12 (dnfsA + dnfsB) was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_MID1sc10-TEV-His_6, pBAD_His_6-(MBP)-TEV-MID1sc10 or pBAD_His_6-MBP-TEV-MID1sc10 bearing the respective mutant and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression of MID1sc10 variants was performed according to the standard protein expression procedure described in Chapter 9.10.1. After expression of MID1sc10 variants o.n., 0.5 mL samples were harvested by centrifugation (4,000×g, 10 min, 4 °C) and analyzed by 15% SDS-PAGE. Consecutively, anti-His Western Blot analysis was performed. Protein purification via affinity-chromatography of MID1sc10 variants was performed according to the standard protein expression procedure described in Chapter 9.10.1.

For further processing of MID1sc10-TEV-His_6, His_6-TEV-MID1sc10 and His_6-MBP-TEV-MID1sc10 variants, TEV protease was added to the combined elution fractions to a final concentration of 0.05 mg/mL. The cleavage mix was subsequently incubated at rt for 1 h under constant agitation. Progress of TEV cleavage was checked via LC-MS and upon completion diluted (1:20) with His standard storage buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl). The digested protein mix was then added to washed *Ni^{2+}-NTA agarose* (Jena Bioscience) and incubated with agitation o.n. at 4 °C. The next day, the crude protein slurry was transferred onto a plastic column and the flow-through was concentrated using *Amicon® Ultra* centrifugal filter units with 3 kDa or MWCO (Millipore).

All wt MID1sc10 variants were subsequently purified via SEC using a *Superdex® Increase 75 10/300 GL* (Cytiva) with MID1sc10 storage buffer (40 mM HEPES pH 8.0, 150 mM NaCl). The process of protein purification was checked by 15% SDS-PAGE. Fractions containing MID1sc10 were pooled together, concentrated employing *Amicon® Ultra* centrifugal filter units with 3 kDa MWCO (Millipore) and stored at -80 °C until further use.

9.10.10 Expression and Purification of MID1sc10-CPD-His_6 Variants

Transformation of *E. coli* K12 or *E. coli* K12 (dnfsA + dnfsB) was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_MID1sc10-CPD-His_6 or pBAD_His_6-MBP-TEV-MID1sc10 bearing the respective mutant and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression of MID1sc10 variants was performed according to the standard protein expression procedure described in Chapter 9.10.1. After expression of MID1sc10 variants o.n., 0.5 mL samples were harvested by centrifugation (4,000×g, 10 min, 4 °C) and analyzed by 15% SDS-PAGE. Consecutively, anti-His Western Blot analysis was performed. Protein purification via affinity-chromatography of MID1sc10 variants was performed according to the standard protein expression procedure described in Chapter 9.10.1.

To induce auto cleavage of CPD in MID1sc10-CPD-His_6 variants, the combined elution fractions were diluted (1:10) with His standard storage buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl). Then, $InsP_6$ was added to a final concentration of 1 mM and the resulting cleavage mix was incubated at rt for 1.5 h under constant agitation.

Progress of cleavage was checked via LC-MS and upon completion added to washed *Ni^{2+}-NTA agarose* (Jena Bioscience) and incubated with agitation at 4 °C for 1 h. The crude protein slurry was subsequently transferred onto a plastic column and the flow-through was concentrated using *Amicon® Ultra* centrifugal filter units with 3 kDa or MWCO (Millipore).

MID1sc10 variants were subsequently purified via SEC using a *Superdex® Increase 75 10/300 GL* (Cytiva) with MID1sc10 storage buffer (40 mM HEPES pH 8.0, 150 mM NaCl, 0.5 mM TCEP). The process of protein purification was checked by 15% SDS-PAGE. Fractions containing MID1sc10 were pooled together, concentrated employing *Amicon® Ultra* centrifugal filter units with 3 kDa MWCO (Millipore) and stored at -80 °C until further use.

9.10.11 Expression of OP-His_6 Variants

Chemically competent *E. coli* BL21 (DE3) were transformed with pET29b(+)_OP-His_6 bearing the respective mutant and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant (at least 100 ng of each plasmid). The cells were rescued by incubation in 1 mL SOC medium at 37 °C, shaking at 200 rpm for 1 h. Afterwards the cell suspension was incubated in 50 mL 2× YT medium containing kanamycin (25 µg/mL) and chloramphenicol (25 µg/mL) at 37 °C, shaking at 200 rpm.

The overnight cultures were diluted into fresh autoinducing medium without arabinose, but supplemented with kanamycin (17 µg/mL) and chloramphenicol (17 µg/mL) to an OD_{600} of 0.04 and subsequently cultured at 37 °C, shaking at 200 rpm until OD_{600} = 0.3. Then, the ncAA was added to a final concentration of typically 1-2 mM. For azo-ncAAs, 2% (v/v) Tween20 was also added to the cell suspension to enhance the solubility of azo-ncAAs. Simultaneously, translation of the 2nd PylRS gene was manually induced by the addition of 0.05% (v/v) arabinose. After further incubation at 37 °C, protein expression was induced at an OD_{600} of 0.6 by the addition of 0.5 mM IPTG. Cultures were then incubated shaking o.n. (typically 16-18 h) at 37 °C. The next day, 0.5 mL samples were harvested by centrifugation (4,000×g, 10 min, 4 °C) and analyzed by 15% SDS-PAGE. Consecutively, anti-Strep Western Blot analysis was performed.

9.10.12 Expression and Purification of ClpP-StrepII Variants

Transformation of *E. coli* K12 or K12 (dnfsA + dnfsB) was performed according to the standard protein expression procedure described in Chapter 9.10.1, using pBAD_ClpP-StrepII bearing the respective ClpP variant or mutant, and pEVOL_PylRS-pylT bearing the desired *Mm*PylRS variant. Expression of ClpP variants was performed similar to the standard protein expression procedure described in Chapter 9.10.1. After the addition of the respective ncAAs, cultures were incubated shaking o.n. (typically 16-18 h) at 30 °C. The next day, 0.5 mL samples were harvested by centrifugation (4,000×g, 10 min, 4 °C) and analyzed by 15% SDS-PAGE. Consecutively, anti-Strep Western Blot analysis was performed.

For protein purification, cells were harvested by centrifugation (4,000×g, 10 min, 4 °C), the supernatant was discarded and the cell pellet resuspended in Strep ClpP lysis buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP). The cell suspension was incubated on ice for 30 min and then lysed on ice via sonication. After centrifugation at 12,000×g, 20 min, 4 °C, the cleared cell lysate was added to washed *Strep-Tactin® Sepharose®* (IBA Lifesciences) and incubated with agitation for 1 h at 4 °C. The mixture was then given onto a plastic column, washed with 6 CV of Strep washing buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP) followed by elution of the protein with Strep elution buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin, 0.5 mM TCEP).

Elution fractions were concentrated using *Amicon® Ultra* centrifugal filter units with 10 kDa MWCO (Millipore) and subsequently purified with ClpP storage buffer (20 mM HEPES pH 8.0, 150 mM NaCl, 0.5 mM TCEP) via SEC using a *Superdex® HiLoad 200 16/600 GL* (Cytiva) to gain sufficient separation of heptamer and functional tetradecamer. The process of protein purification was checked by 15% SDS-PAGE. Fractions containing ClpP were pooled together, concentrated employing *Amicon® Ultra* centrifugal filter units with 10 kDa MWCO (Millipore) and stored at -80 °C until further use.

9.11 Protein Characterization Methods

9.11.1 SDS-PAGE

For SDS-PAGE analysis of POIs amber suppressed with azo-ncAAs, 0.5 mL of culture was pelletized and resuspended in 200 µL Laemmli loading buffer (1×). For SDS-PAGE analysis of any other protein expression, 0.5 mL of culture was pelletized and resuspended in 200 µL Laemmli loading buffer (1×) normalized according to OD_{600} values (i.e. per OD = 1, 50 µl of Laemmli loading buffer (1×). All samples were boiled at 95 °C for 10 min, pelleted again (17,000×g, 10 min) and the supernatant was subsequently loaded onto 15% SDS-PAGE gels.

SDS-PAGE gels were stained with *Quick Coomassie Stain* (Protein Ark) and subsequently destained with deionized water. The molecular weight of the proteins was typically determined by comparison with the *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific) (3.5-5 µL).

9.11.2 Western Blot Analysis

For Western Blot Analysis, the SDS-PAGE gel was transferred to an *Amersham™ Protran® Premium 0.2* NC membrane (Cytiva) using the *PerfectBlue™ Semi-Dry Electroblotter* by VWR. Afterwards the membrane was incubated for 1 h at rt in a suspension of 5% (w/v) skim milk in TBS-T buffer. After disposal, 1% (w/v) skim milk TBS-T buffer and His-HRP antibody (1:5000) or Strep-MAB-Classic antibody (HRP conjugate by IBA Lifesciences, 1:20000) were added and the membrane was incubated at 4 °C, rolling o.n. or for 1 h, respectively. After disposal of the antibody containing buffer solution, the membrane was washed with 5 mL fresh TBS-T buffer (1×) for 5 min at rt (five times). For Western Blot chemiluminescence imaging, the membrane was either incubated with *Amersham ECL Prime Western Blotting Detection Reagent* (Cytiva) or *Immobilon® Forte* (Millipore) and subsequently imaged on a *Fusion Solo S* (Vilber Lourmat) or *Amersham ImageQuant 800* (Cytiva).

It should be noted, that for Strep-antibody detection, BSA is typically used as blocking reagent instead of skim milk. However, no difference was observed when blocking with skim milk instead.

9.11.3 Protein-MS

Purified protein samples were diluted to a final concentration of 0.1 mg/mL and analyzed by a *6130 Quadrupole* mass spectrometer (Agilent Technologies) with a MWD detector, coupled to a *1260 Infinity* HPLC-system (Agilent Technologies) and equipped with a *Jupiter* C4 (5 µm, 300 Å, 150×2.0 mm) column. Mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile) were applied for gradient elution (5 to 90 % B). Samples were analyzed by UV absorbance at 193, 254 and 280 nm together with the positive ESI-mode. The protein mass was obtained by deconvolution using the deconvolution tool of the software *OpenLab ChemStation* (Agilent). The obtained molecular weight of proteins was compared to the theoretical mass calculated using the browser-based tool *ProtParam* (Expasy, Swiss Bioinformatics Resource Portal).

High resolution protein mass spectra were measured on an *LTQ-FT Ultra* mass spectrometer (Thermo Fisher Scientific) coupled to a *Dionex UltiMate 3000* HPLC Instrument (Thermo Fisher Scientific) by the group of Prof. Sieber at the Technical University Munich. Spectra were acquired in positive ESI-mode and deconvoluted with the *Thermo Xcalibur Xtract algorithm* (Thermo Fisher Scientific) using basic deconvolution default settings.

9.11.4 Circular Dichroism Spectroscopy

CD measurements were performed on a Jasco CD spectrometer *J-810*. All measurements were performed using a Hellma *QS High Precision* quartz cuvette (1 mm pathlength). The temperature was set to 25 °C and the nitrogen flow rate kept between 3.5-4.0 l/min. All Cam samples were dissolved in 10 mM Tris buffer (pH 7.5, 0.2-0-3 mg/L), while all PDZ3 samples were dissolved in PDZ3 10 mM phosphate buffer (pH 7.4, 0.2-0-3 mg/L). Spectra reported are the averages of five scans each. A scan speed of 20 nm/min, with a 2.0 nm bandwidth and a 4 s response time, was used.

9.12 Protein Interaction and Activity Assays

9.12.1 Pull-Down Assay

All samples were finally resuspended in 50 µL Laemmli loading buffer (1×) and analyzed via SDS-PAGE.

GST-Z-domain and Affibody Variants

Pull-down assays were performed similar to known literature procedures.[283]

In short, 30 µg of GST-tagged affibody interactor Z-domain were incubated for 1 h at 4 °C with 30 µL 50% slurry of *Glutathione Sepharose 4B* (GE Healthcare) equilibrated with pull-down (affibody) buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM DTT, 0.2% (v/v) Tween20). Subsequently the beads were washed four times with 60 µL pull-down (affibody) buffer (4,000×g, 2 min, 4 °C) before adding between 1-10 µg of the corresponding affibody variant to a total volume of 50 µL. After incubation at 4 °C for 1.5 h the beads were washed six times with 50 µL pull-down (affibody) buffer.

Note that the concentration of amber suppressed affibody variants was estimated via densitometry and might not be accurate, since the obtained proteins were not entirely pure (even after purification via SEC). Dark-adapted samples were kept in the dark at 37 °C for 1 h prior to applying them to the pull-down mixture.

GST-EGFP and EGFP-Nanobody Variants

Pull-down assays were performed similar to known literature procedures.[288]

In short, 5 µg of GST-tagged EGFP were incubated for 1 h at 4 °C with 25 µL 50% slurry of *Glutathione Sepharose 4B* (GE Healthcare) equilibrated with pull-down (nb) buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM TCEP, 0.2% (v/v) Tween20). Subsequently the beads were washed four times with 50 µL pull-down (nb) buffer (4,000×g, 2 min, 4 °C) before adding between 3.5 µg of the corresponding nanobody variant to a total volume of 50 µL. After incubation at 4 °C for 1 h the beads were washed six times with 50 µL pull-down (nb) buffer.

Note that the concentration of nanobody variants was estimated via Biuret reaction using the *Pierce™ BCA Protein Assay Kit* (Thermo Scientific). Dark-adapted samples were kept in the dark at 37 °C for 1 h prior to applying them to the pull-down mixture.

9.12.2 Fluorogenic Peptidase Assay

*In vitro Sa*ClpP peptidase activity was measured by monitoring the cleavage of the fluorogenic substrate *N*-Ac-Ala-hArg-(*S*)-2-aminooctanoic acid-7-amino-4-carbamoylmethyl-coumarin (*N*-Ac-Ala-hArg-2-Aoc-ACC, provided by T. Gronauer, former PhD student in the group of Prof. Sieber at TUM), similar to known literature procedures.[297]

In short, 99 µL of the respective *Sa*ClpP cysteine mutant (final concentration: 1 µM) in ClpP assay buffer (25 mM HEPES pH 7.6, 200 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 10% (v/v) glycerol) were added to a black PS flat bottom 96-well plate (*Greiner*) and incubated at 32 °C for 15 min. The enzymatic reaction was initiated by addition of 1 µL *N*-Ac-Ala-hArg-2-Aoc-ACC (20 mM in DMSO, final concentration: 200 µM) and the fluorescence (λ_{ex} = 380 nm, λ_{em} = 430 nm) monitored using an *InfiniteM200Pro* plate reader (Tecan). The initial slope (0-540 s) of the fluorescence over time signal was calculated by linear regression fitting using *GraphPad Prism 9.5.0* (GraphPad Software). The wt *Sa*ClpP-StrepII control sample was normalized to 100% activity and the residual activity of variants bearing the respective cysteine mutation(s) was determined. Data were recorded in triplicates.

9.12.3 Intrinsic Tryptophan Fluorescence Quenching Assay

Intrinsic tryptophan fluorescence quenching assays were performed similar to known literature procedures.[158]

Fluorescence quenching of the pentapeptide KETWV, which contains a single tryptophan residue, was measured to assess the binding affinity between PDZ3 variants and the peptide ligand. The respective protein concentration was varied (0-50 µM) while keeping the pentapeptide concentration constant (15 µM). To this end, 60 µL pentapeptide (final concentration: 15 µM) in PDZ3 assay buffer (10 mM phosphate, pH 7.4) were added to a Hellma *QS High Precision* fluorescence quartz cuvette (3 mm pathlength) and incubated at 21 °C for 10 min. The respective protein in PDZ3 assay buffer was titrated (final concentrations: 3, 5, 10, 14, 20, 30, 40 and 50 µM) and the fluorescence (λ_{ex} = 250 nm, λ_{em} = 325 nm) monitored using a *FP-8350 Spectrofluorometer* (Jasco). Binding affinity constants (K_d values) were obtained by fitting the data to the function “Specific binding with Hill slope“ using *GraphPad Prism 9.5.0* (GraphPad Software).

I Literature

1. N. Ankenbruck, T. Courtney, Y. Naro, A. Deiters, Optochemical Control of Biological Processes in Cells and Animals. *Angewandte Chemie (International ed. in English)* **57**, (2018), doi:10.1002/anie.201700171.
2. C. P. O'Banion, D. S. Lawrence, Optogenetics: A Primer for Chemists. *ChemBioChem* **19**, (2018), doi:10.1002/cbic.201800013.
3. F. Zhang *et al.*, The Microbial Opsin Family of Optogenetic Tools. *Cell* **147**, (2011), doi:10.1016/j.cell.2011.12.004.
4. D. Strickland *et al.*, Rationally improving LOV domain–based photoswitches. *Nature Methods* **7**, (2010), doi:10.1038/nmeth.1473.
5. J. Von Lintig, P. D. Kiser, M. Golczak, K. Palczewski, The biochemical and structural basis for trans-to-cis isomerization of retinoids in the chemistry of vision. *Trends in Biochemical Sciences* **35**, (2010), doi:10.1016/j.tibs.2010.01.005.
6. A. Losi, W. Gärtner, Old Chromophores, New Photoactivation Paradigms, Trendy Applications: Flavins in Blue Light-Sensing Photoreceptors†. *Photochemistry and Photobiology* **87**, (2011), doi:10.1111/j.1751-1097.2011.00913.x.
7. K. S. Conrad, C. C. Manahan, B. R. Crane, Photochemistry of flavoprotein light sensors. *Nature Chemical Biology* **10**, (2014), doi:10.1038/nchembio.1633.
8. F. Crick, Thinking about the Brain. *Scientific American* **241**, (1979), doi.
9. D. Oesterhelt, W. Stoeckenius, Rhodopsin-like Protein from the Purple Membrane of Halobacterium halobium. *Nature New Biology* **233**, (1971), doi:10.1038/newbio233149a0.
10. A. Matsuno-Yagi, Y. Mukohata, Two possible roles of bacteriorhodopsin; a comparative study of strains of Halobacterium halobium differing in pigmentation. *Biochemical and Biophysical Research Communications* **78**, (1977), doi:10.1016/0006-291X(77)91245-1.
11. G. Nagel *et al.*, Channelrhodopsin-1: A Light-Gated Proton Channel in Green Algae. *Science* **296**, (2002), doi:10.1126/science.1072068.
12. E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience* **8**, (2005), doi:10.1038/nn1525.
13. K. Deisseroth, Optogenetics. *Nature Methods* **8**, (2011), doi:10.1038/nmeth.f.324.
14. E. Bamberg, W. Gärtner, D. Trauner, Introduction: Optogenetics and Photopharmacology. *Chemical Reviews* **118**, (2018), doi:10.1021/acs.chemrev.8b00483.
15. H. Tian *et al.*, Video-based pooled screening yields improved far-red genetically encoded voltage indicators. *Nature Methods*, (2023), doi:10.1038/s41592-022-01743-5.
16. Y. I. Wu *et al.*, A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* **461**, (2009), doi:10.1038/nature08241.
17. C. Renicke, D. Schuster, S. Usherenko, L.-O. Essen, C. Taxis, A LOV2 Domain-Based Optogenetic Tool to Control Protein Degradation and Cellular Function. *Chemistry & Biology* **20**, (2013), doi:10.1016/j.chembiol.2013.03.005.
18. H. Yumerefendi *et al.*, Light-induced nuclear export reveals rapid dynamics of epigenetic modifications. *Nature Chemical Biology* **12**, (2016), doi:10.1038/nchembio.2068.
19. E. Mills, X. Chen, E. Pham, S. Wong, K. Truong, Engineering a Photoactivated Caspase-7 for Rapid Induction of Apoptosis. *ACS Synthetic Biology* **1**, (2012), doi:10.1021/sb200008j.
20. L. B. Motta-Mena *et al.*, An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nature Chemical Biology* **10**, (2014), doi:10.1038/nchembio.1430.
21. A. A. Gil *et al.*, Optogenetic control of protein binding using light-switchable nanobodies. *Nature Communications* **11**, (2020), doi:10.1038/s41467-020-17836-8.
22. T. Fehrentz, M. Schönberger, D. Trauner, Optochemical Genetics. *Angewandte Chemie International Edition* **50**, (2011), doi:10.1002/anie.201103236.

23. J. H. Kaplan, B. Forbush, III, J. F. Hoffman, Rapid photolytic release of adenosine 5'-triphosphate from a protected analog: utilization by the sodium:potassium pump of human red blood cell ghosts. *Biochemistry* **17**, (1978), doi:10.1021/bi00603a020.
24. J. Engels, E. J. Schlaeger, Synthesis, structure, and reactivity of adenosine cyclic 3',5'-phosphate-benzyltriesters. *Journal of Medicinal Chemistry* **20**, (1977), doi:10.1021/jm00217a008.
25. C. Brieke, F. Rohrbach, A. Gottschalk, G. Mayer, A. Heckel, Light-Controlled Tools. *Angewandte Chemie International Edition* **51**, (2012), doi:10.1002/anie.201202134.
26. J. M. Govan, D. D. Young, M. O. Lively, A. Deiters, Optically triggered immune response through photocaged oligonucleotides. *Tetrahedron Letters* **56**, (2015), doi:10.1016/j.tetlet.2015.01.165.
27. A. Korman *et al.*, Light-controlled twister ribozyme with single-molecule detection resolves RNA function in time and space. *Proceedings of the National Academy of Sciences* **117**, (2020), doi:10.1073/pnas.2003425117.
28. D. Lovatt *et al.*, Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nature Methods* **11**, (2014), doi:10.1038/nmeth.2804.
29. Y. Liu *et al.*, Very fast CRISPR on demand. *Science* **368**, (2020), doi:10.1126/science.aay8204.
30. J. Pietruszka *et al.*, Photocaged Carbohydrates: Versatile Tools for Controlling Gene Expression by Light. *Synthesis* **49**, (2016), doi:10.1055/s-0035-1562617.
31. J. Mos, A. Jakob, J. Becker-Baldus, A. Heckel, C. Glaubitz, Light-Induced Uncaging for Time-Resolved Observations of Biochemical Reactions by MAS NMR Spectroscopy. *Chemistry – A European Journal* **26**, (2020), doi:10.1002/chem.202000770.
32. T. Courtney, A. Deiters, Recent advances in the optical control of protein function through genetic code expansion. *Current Opinion in Chemical Biology* **46**, (2018), doi:10.1016/j.cbpa.2018.07.011.
33. M. J. Hansen *et al.*, Photoactivation of MDM2 Inhibitors: Controlling Protein–Protein Interaction with Light. *Journal of the American Chemical Society* **140**, (2018), doi:10.1021/jacs.8b04870.
34. D. Bliman, J. R. Nilsson, P. Kettunen, J. Andréasson, M. Grøtli, A Caged Ret Kinase Inhibitor and its Effect on Motoneuron Development in Zebrafish Embryos. *Scientific Reports* **5**, (2015), doi:10.1038/srep13109.
35. N. Ieda, S. Yamada, M. Kawaguchi, N. Miyata, H. Nakagawa, (7-Diethylaminocoumarin-4-yl)methyl ester of suberoylanilide hydroxamic acid as a caged inhibitor for photocontrol of histone deacetylase activity. *Bioorganic & Medicinal Chemistry* **24**, (2016), doi:10.1016/j.bmc.2016.04.042.
36. R. S. Givens *et al.*, The Photo-Favorskii Reaction of p-Hydroxyphenacyl Compounds Is Initiated by Water-Assisted, Adiabatic Extrusion of a Triplet Biradical. *Journal of the American Chemical Society* **130**, (2008), doi:10.1021/ja7109579.
37. B. Hellrung, Y. Kamdzhilov, M. Schwörer, J. Wirz, Photorelease of Alcohols from 2-Nitrobenzyl Ethers Proceeds via Hemiacetals and May Be Further Retarded by Buffers Intercepting the Primary aci-Nitro Intermediates. *Journal of the American Chemical Society* **127**, (2005), doi:10.1021/ja052300s.
38. P. Klán *et al.*, Photoremovable Protecting Groups in Chemistry and Biology: Reaction Mechanisms and Efficacy. *Chemical Reviews* **113**, (2013), doi:10.1021/cr300177k.
39. A. Gandioso *et al.*, Development of Green/Red-Absorbing Chromophores Based on a Coumarin Scaffold That Are Useful as Caging Groups. *The Journal of Organic Chemistry* **82**, (2017), doi:10.1021/acs.joc.7b00788.
40. J. A. Peterson *et al.*, Family of BODIPY Photocages Cleaved by Single Photons of Visible/Near-Infrared Light. *Journal of the American Chemical Society* **140**, (2018), doi:10.1021/jacs.8b04040.
41. Y. Chitose *et al.*, Design and Synthesis of a Caged Carboxylic Acid with a Donor−π–Donor Coumarin Structure: One-photon and Two-photon Uncaging Reactions Using Visible and Near-Infrared Lights. *Organic Letters* **19**, (2017), doi:10.1021/acs.orglett.7b00957.
42. Y. Becker *et al.*, Selective Modification for Red-Shifted Excitability: A Small Change in Structure, a Huge Change in Photochemistry. *Chemistry – A European Journal* **27**, (2021), doi:10.1002/chem.202003672.

43. A. Bardhan, A. Deiters, Development of photolabile protecting groups and their application to the optochemical control of cell signaling. *Current Opinion in Structural Biology* **57**, (2019), doi:10.1016/j.sbi.2019.03.028.
44. G. Alachouzos, A. M. Schulte, A. Mondal, W. Szymanski, B. L. Feringa, Computational Design, Synthesis, and Photochemistry of Cy7-PPG, an Efficient NIR-Activated Photolabile Protecting Group for Therapeutic Applications**. *Angewandte Chemie International Edition* **61**, (2022), doi:10.1002/anie.202201308.
45. L. Fournier *et al.*, Coumarinylmethyl Caging Groups with Redshifted Absorption. *Chemistry – A European Journal* **19**, (2013), doi:10.1002/chem.201302630.
46. T. Slanina *et al.*, In Search of the Perfect Photocage: Structure–Reactivity Relationships in meso-Methyl BODIPY Photoremovable Protecting Groups. *Journal of the American Chemical Society* **139**, (2017), doi:10.1021/jacs.7b08532.
47. M. W. H. Hoorens, W. Szymanski, Reversible, Spatial and Temporal Control over Protein Activity Using Light. *Trends in Biochemical Sciences* **43**, (2018), doi:10.1016/j.tibs.2018.05.004.
48. A. Goulet-Hanssens, F. Eisenreich, S. Hecht, Enlightening Materials with Photoswitches. *Advanced Materials* **32**, (2020), doi:10.1002/adma.201905966.
49. J. Volarić, W. Szymanski, N. A. Simeth, B. L. Feringa, Molecular photoswitches in aqueous environments. *Chemical Society Reviews* **50**, (2021), doi:10.1039/d0cs00547a.
50. I. M. Welleman, M. W. H. Hoorens, B. L. Feringa, H. H. Boersma, W. Szymański, Photoresponsive molecular tools for emerging applications of light in medicine. *Chemical Science* **11**, (2020), doi:10.1039/d0sc04187d.
51. S. E. Braslavsky, Glossary of terms used in photochemistry, 3rd edition (IUPAC Recommendations 2006). **79**, (2007), doi:doi:10.1351/pac200779030293.
52. D. Villarón, S. J. Wezenberg, Stiff-Stilbene Photoswitches: From Fundamental Studies to Emergent Applications. *Angewandte Chemie International Edition* **59**, (2020), doi:10.1002/anie.202001031.
53. S. Wiedbrauk, H. Dube, Hemithioindigo—an emerging photoswitch. *Tetrahedron Letters* **56**, (2015), doi:10.1016/j.tetlet.2015.05.022.
54. T. Cordes *et al.*, Hemithioindigo-based photoswitches as ultrafast light trigger in chromopeptides. *Chemical Physics Letters* **428**, (2006), doi:10.1016/j.cplett.2006.07.043.
55. M. W. H. Hoorens *et al.*, Iminothioindoxyl as a molecular photoswitch with 100 nm band separation in the visible range. *Nature Communications* **10**, (2019), doi:10.1038/s41467-019-10251-8.
56. M. Medved' *et al.*, Tailoring the optical and dynamic properties of iminothioindoxyl photoswitches through acidochromism. *Chemical Science* **12**, (2021), doi:10.1039/d0sc07000a.
57. A. Bafana, S. S. Devi, T. Chakrabarti, Azo dyes: past, present and the future. *Environmental Reviews* **19**, (2011), doi:10.1139/a11-018.
58. J. Bieth, S. M. Vratsanos, N. Wassermann, B. F. Erlanger, Photoregulation of Biological Activity by Photochromic Reagents, II. Inhibitors of Acetylcholinesterase. *Proceedings of the National Academy of Sciences* **64**, (1969), doi:10.1073/pnas.64.3.1103.
59. W. A. Velema, W. Szymanski, B. L. Feringa, Photopharmacology: Beyond Proof of Principle. *Journal of the American Chemical Society* **136**, (2014), doi:10.1021/ja413063e.
60. J. Broichhagen, J. A. Frank, D. Trauner, A Roadmap to Success in Photopharmacology. *Accounts of Chemical Research* **48**, (2015), doi:10.1021/acs.accounts.5b00129.
61. M. Pilz Da Cunha *et al.*, An Untethered Magnetic- and Light-Responsive Rotary Gripper: Shedding Light on Photoresponsive Liquid Crystal Actuators. *Advanced Optical Materials* **7**, (2019), doi:10.1002/adom.201801643.
62. A. K. Saydjari, P. Weis, S. Wu, Spanning the Solar Spectrum: Azopolymer Solar Thermal Fuels for Simultaneous UV and Visible Light Storage. *Advanced Energy Materials* **7**, (2017), doi:10.1002/aenm.201601622.
63. A. D. Wong, T. M. Güngör, E. R. Gillies, Multiresponsive Azobenzene End-Cap for Self-Immolative Polymers. *ACS Macro Letters* **3**, (2014), doi:10.1021/mz500613d.
64. H. Zhou *et al.*, Photoswitching of glass transition temperatures of azobenzene-containing polymers induces reversible solid-to-liquid transitions. *Nature Chemistry* **9**, (2017), doi:10.1038/nchem.2625.

65. G. S. Hartley, The Cis-form of Azobenzene. *Nature* **140**, (1937), doi:10.1038/140281a0.
66. M. Reimann, E. Teichmann, S. Hecht, M. Kaupp, Solving the Azobenzene Entropy Puzzle: Direct Evidence for Multi-State Reactivity. *The Journal of Physical Chemistry Letters* **13**, (2022), doi:10.1021/acs.jpclett.2c02838.
67. H. M. D. Bandara, S. C. Burdette, Photoisomerization in different classes of azobenzene. *Chem. Soc. Rev.* **41**, (2012), doi:10.1039/c1cs15179g.
68. J. Garcia-Amorós, M. Díaz-Lobo, S. Nonell, D. Velasco, Fastest Thermal Isomerization of an Azobenzene for Nanosecond Photoswitching Applications under Physiological Conditions. *Angewandte Chemie International Edition* **51**, (2012), doi:10.1002/anie.201207602.
69. M. Dong, A. Babalhavaeji, S. Samanta, A. A. Beharry, G. A. Woolley, Red-Shifting Azobenzene Photoswitches for in Vivo Use. *Accounts of Chemical Research* **48**, (2015), doi:10.1021/acs.accounts.5b00270.
70. S. Samanta *et al.*, Photoswitching Azo Compounds in Vivo with Red Light. *Journal of the American Chemical Society* **135**, (2013), doi:10.1021/ja402220t.
71. D. Bléger, J. Schwarz, A. M. Brouwer, S. Hecht, o-Fluoroazobenzenes as Readily Synthesized Photoswitches Offering Nearly Quantitative Two-Way Isomerization with Visible Light. *Journal of the American Chemical Society* **134**, (2012), doi:10.1021/ja310323y.
72. R. Siewertsen *et al.*, Highly Efficient Reversible Z−E Photoisomerization of a Bridged Azobenzene with Visible Light through Resolved $S_1(n\pi^*)$ Absorption Bands. *Journal of the American Chemical Society* **131**, (2009), doi:10.1021/ja906547d.
73. M. Hammerich *et al.*, Heterodiazocines: Synthesis and Photochromic Properties, Trans to Cis Switching within the Bio-optical Window. *Journal of the American Chemical Society* **138**, (2016), doi:10.1021/jacs.6b05846.
74. P. Lentes *et al.*, Nitrogen Bridged Diazocines: Photochromes Switching within the Near-Infrared Region with High Quantum Yields in Organic Solvents and in Water. *Journal of the American Chemical Society* **141**, (2019), doi:10.1021/jacs.9b06104.
75. W. Moormann *et al.*, Efficient Conversion of Light to Chemical Energy: Directional, Chiral Photoswitches with Very High Quantum Yields. *Angewandte Chemie International Edition* **59**, (2020), doi:10.1002/anie.202005361.
76. M. Dong, A. Babalhavaeji, M. J. Hansen, L. Kálmán, G. A. Woolley, Red, far-red, and near infrared photoswitches based on azonium ions. *Chemical Communications* **51**, (2015), doi:10.1039/c5cc02804c.
77. M. Dong *et al.*, Near-Infrared Photoswitching of Azobenzenes under Physiological Conditions. *Journal of the American Chemical Society* **139**, (2017), doi:10.1021/jacs.7b06471.
78. C. E. Weston, R. D. Richardson, P. R. Haycock, A. J. P. White, M. J. Fuchter, Arylazopyrazoles: Azoheteroarene Photoswitches Offering Quantitative Isomerization and Long Thermal Half-Lives. *Journal of the American Chemical Society* **136**, (2014), doi:10.1021/ja505444d.
79. C. Knie *et al.*, ortho-Fluoroazobenzenes: Visible Light Switches with Very Long-Lived Z Isomers. *Chemistry - A European Journal* **20**, (2014), doi:10.1002/chem.201404649.
80. D. B. Konrad *et al.*, Computational Design and Synthesis of a Deeply Red-Shifted and Bistable Azobenzene. *Journal of the American Chemical Society* **142**, (2020), doi:10.1021/jacs.9b10430.
81. L. N. Lameijer *et al.*, General Principles for the Design of Visible-Light-Responsive Photoswitches: Tetra-ortho-Chloro-Azobenzenes. *Angewandte Chemie International Edition* **59**, (2020), doi:10.1002/anie.202008700.
82. J. Volarić *et al.*, Design and Synthesis of Visible-Light-Responsive Azobenzene Building Blocks for Chemical Biology. *The Journal of Organic Chemistry* **87**, (2022), doi:10.1021/acs.joc.2c01777.
83. H. Asanuma, T. Ito, M. Komiyama, Photo-responsive oligonucleotides carrying azobenzene in the side-chains. *Tetrahedron Letters* **39**, (1998), doi:10.1016/S0040-4039(98)02022-X.
84. H. Asanuma *et al.*, Synthesis of azobenzene-tethered DNA for reversible photo-regulation of DNA functions: hybridization and transcription. *Nature Protocols* **2**, (2007), doi:10.1038/nprot.2006.465.
85. T. Stafforst, D. Hilvert, Modulating PNA/DNA Hybridization by Light. *Angewandte Chemie International Edition* **49**, (2010), doi:10.1002/anie.201004548.

86. A. Bergen *et al.*, Photodependent Melting of Unmodified DNA Using a Photosensitive Intercalator: A New and Generic Tool for Photoreversible Assembly of DNA Nanostructures at Constant Temperature. *Nano Letters* **16**, (2016), doi:10.1021/acs.nanolett.5b04762.
87. N. A. Simeth *et al.*, Rational design of a photoswitchable DNA glue enabling high regulatory function and supramolecular chirality transfer. *Chemical Science* **12**, (2021), doi:10.1039/d1sc02194j.
88. Y. Kamiya *et al.*, Synthetic Gene Involving Azobenzene-Tethered T7 Promoter for the Photocontrol of Gene Expression by Visible Light. *ACS Synthetic Biology* **4**, (2015), doi:10.1021/sb5001092.
89. S. Ogasawara, Duration Control of Protein Expression in Vivo by Light-Mediated Reversible Activation of Translation. *ACS Chemical Biology* **12**, (2017), doi:10.1021/acschembio.6b00684.
90. T. Ko *et al.*, Optical Control of Translation with a Puromycin Photoswitch. *Journal of the American Chemical Society* **144**, (2022), doi:10.1021/jacs.2c07374.
91. L. Chen, Y. Liu, W. Guo, Z. Liu, Light responsive nucleic acid for biomedical application. *Exploration* **2**, (2022), doi:10.1002/exp.20210099.
92. D. K. Prusty, V. Adam, R. M. Zadegan, S. Irsen, M. Famulok, Supramolecular aptamer nano-constructs for receptor-mediated targeting and light-triggered release of chemotherapeutics into cancer cells. *Nature Communications* **9**, (2018), doi:10.1038/s41467-018-02929-2.
93. C. Pernpeintner *et al.*, Light-Controlled Membrane Mechanics and Shape Transitions of Photoswitchable Lipid Vesicles. *Langmuir* **33**, (2017), doi:10.1021/acs.langmuir.7b01020.
94. P. Urban *et al.*, Light-Controlled Lipid Interaction and Membrane Organization in Photolipid Bilayer Vesicles. *Langmuir* **34**, (2018), doi:10.1021/acs.langmuir.8b03241.
95. N. Chander *et al.*, Optimized Photoactivatable Lipid Nanoparticles Enable Red Light Triggered Drug Release. *Small* **17**, (2021), doi:10.1002/smll.202008198.
96. H. Xiong *et al.*, Optical control of neuronal activities with photoswitchable nanovesicles. *Nano Research* **16**, (2023), doi:10.1007/s12274-022-4853-x.
97. N. Jiménez-Rojo *et al.* (Cold Spring Harbor Laboratory, 2022), (10.1101/2022.02.14.480333).
98. J. A. Frank *et al.*, Photoswitchable diacylglycerols enable optical control of protein kinase C. *Nature Chemical Biology* **12**, (2016), doi:10.1038/nchembio.2141.
99. R. Tei, J. Morstein, A. Shemet, D. Trauner, J. M. Baskin, Optical Control of Phosphatidic Acid Signaling. *ACS Central Science* **7**, (2021), doi:10.1021/acscentsci.1c00444.
100. M. Kol *et al.*, Optical manipulation of sphingolipid biosynthesis using photoswitchable ceramides. *eLife* **8**, (2019), doi:10.7554/elife.43230.
101. J. Morstein *et al.*, Photoswitchable Isoprenoid Lipids Enable Optical Control of Peptide Lipidation. *ACS Chemical Biology* **17**, (2022), doi:10.1021/acschembio.2c00645.
102. K. Hüll, J. Morstein, D. Trauner, In Vivo Photopharmacology. *Chemical Reviews* **118**, (2018), doi:10.1021/acs.chemrev.8b00037.
103. A. Mourot *et al.*, Tuning Photochromic Ion Channel Blockers. *ACS Chemical Neuroscience* **2**, (2011), doi:10.1021/cn200037p.
104. I. Tochitsky *et al.*, How Azobenzene Photoswitches Restore Visual Responses to the Blind Retina. *Neuron* **92**, (2016), doi:10.1016/j.neuron.2016.08.038.
105. I. Tochitsky, J. Trautman, N. Gallerani, J. G. Malis, R. H. Kramer, Restoring visual function to the blind retina with a potent, safe and long-lasting photoswitch. *Scientific Reports* **7**, (2017), doi:10.1038/srep45487.
106. J. B. Trads *et al.*, Sign Inversion in Photopharmacology: Incorporation of Cyclic Azobenzenes in Photoswitchable Potassium Channel Blockers and Openers. *Angewandte Chemie International Edition* **58**, (2019), doi:10.1002/anie.201905790.
107. E. R. Thapaliya, J. Zhao, G. C. R. Ellis-Davies, Locked-Azobenzene: Testing the Scope of a Unique Photoswitchable Scaffold for Cell Physiology. *ACS Chemical Neuroscience* **10**, (2019), doi:10.1021/acschemneuro.8b00734.
108. D. Yang *et al.*, G protein-coupled receptors: structure- and function-based drug discovery. *Signal Transduction and Targeted Therapy* **6**, (2021), doi:10.1038/s41392-020-00435-w.
109. J. Morstein *et al.*, Optical control of sphingosine-1-phosphate formation and function. *Nature Chemical Biology* **15**, (2019), doi:10.1038/s41589-019-0269-7.

110. J. Morstein *et al.*, Optical Control of Lysophosphatidic Acid Signaling. *Journal of the American Chemical Society* **142**, (2020), doi:10.1021/jacs.0c02154.
111. P. Donthamsetti *et al.*, Selective Photoswitchable Allosteric Agonist of a G Protein-Coupled Receptor. *Journal of the American Chemical Society* **143**, (2021), doi:10.1021/jacs.1c02586.
112. K. Hüll *et al.*, Optical Control of Adenosine-Mediated Pain Modulation. *Bioconjugate Chemistry* **32**, (2021), doi:10.1021/acs.bioconjchem.1c00387.
113. R. C. Sarott *et al.*, Optical Control of Cannabinoid Receptor 2-Mediated Ca2+ Release Enabled by Synthesis of Photoswitchable Probes. *Journal of the American Chemical Society* **143**, (2021), doi:10.1021/jacs.0c08926.
114. P. Leippe, J. Koehler Leman, D. Trauner, Specificity and Speed: Tethered Photopharmacology. *Biochemistry* **56**, (2017), doi:10.1021/acs.biochem.7b00687.
115. J. Broichhagen, D. Trauner, The in vivo chemistry of photoswitched tethered ligands. *Current Opinion in Chemical Biology* **21**, (2014), doi:10.1016/j.cbpa.2014.07.008.
116. Y.-H. Tsai, S. Essig, J. R. James, K. Lang, J. W. Chin, Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nature chemistry* **7**, (2015), doi:10.1038/nchem.2253.
117. M. Banghart, K. Borges, E. Isacoff, D. Trauner, R. H. Kramer, Light-activated ion channels for remote control of neuronal firing. *Nature Neuroscience* **7**, (2004), doi:10.1038/nn1356.
118. M. Volgraf *et al.*, Allosteric control of an ionotropic glutamate receptor with an optical switch. *Nature Chemical Biology* **2**, (2006), doi:10.1038/nchembio756.
119. P. Gorostiza *et al.*, Mechanisms of photoswitch conjugation and light activation of an ionotropic glutamate receptor. *Proceedings of the National Academy of Sciences* **104**, (2007), doi:10.1073/pnas.0701274104.
120. S. Szobota *et al.*, Remote Control of Neuronal Activity with a Light-Gated Glutamate Receptor. *Neuron* **54**, (2007), doi:10.1016/j.neuron.2007.05.010.
121. N. Caporale *et al.*, LiGluR Restores Visual Responses in Rodent Models of Inherited Blindness. *Molecular Therapy* **19**, (2011), doi:10.1038/mt.2011.103.
122. S. Berlin *et al.*, A family of photoswitchable NMDA receptors. *eLife* **5**, (2016), doi:10.7554/elife.12040.
123. V. M. Krishnamurthy, V. Semetey, P. J. Bracher, N. Shen, G. M. Whitesides, Dependence of Effective Molarity on Linker Length for an Intramolecular Protein−Ligand System. *Journal of the American Chemical Society* **129**, (2007), doi:10.1021/ja066780e.
124. J. Broichhagen *et al.*, Orthogonal Optical Control of a G Protein-Coupled Receptor with a SNAP-Tethered Photochromic Ligand. *ACS Central Science* **1**, (2015), doi:10.1021/acscentsci.5b00260.
125. M. H. Berry *et al.*, Restoration of patterned vision with an engineered photoactivatable G protein-coupled receptor. *Nature Communications* **8**, (2017), doi:10.1038/s41467-017-01990-7.
126. A. Acosta-Ruiz *et al.*, Branched Photoswitchable Tethered Ligands Enable Ultra-efficient Optical Control and Detection of G Protein-Coupled Receptors In Vivo. *Neuron* **105**, (2020), doi:10.1016/j.neuron.2019.10.036.
127. A. Holt *et al.* (Cold Spring Harbor Laboratory, 2022), (10.1101/2022.04.07.487476).
128. J. Morstein *et al.*, Photoswitchable Serotonins for Optical Control of the 5-HT2A Receptor. *Angewandte Chemie International Edition* **61**, (2022), doi:10.1002/anie.202117094.
129. P. Donthamsetti *et al.*, Cell specific photoswitchable agonist for reversible control of endogenous dopamine receptors. *Nature Communications* **12**, (2021), doi:10.1038/s41467-021-25003-w.
130. T. K. Mukhopadhyay, J. Morstein, D. Trauner, Photopharmacological control of cell signaling with photoswitchable lipids. *Current Opinion in Pharmacology* **63**, (2022), doi:10.1016/j.coph.2022.102202.
131. J. Morstein *et al.*, Optical control of the nuclear bile acid receptor FXR with a photohormone. *Chemical Science* **11**, (2020), doi:10.1039/c9sc02911g.
132. M. Wegener, M. J. Hansen, A. J. M. Driessen, W. Szymanski, B. L. Feringa, Photocontrol of Antibacterial Activity: Shifting from UV to Red Light Activation. *Journal of the American Chemical Society* **139**, (2017), doi:10.1021/jacs.7b09281.

133. G. Testolin *et al.*, Optical Modulation of Antibiotic Resistance by Photoswitchable Cystobactamids. *Chemistry – A European Journal* **28**, (2022), doi:10.1002/chem.202201297.
134. M. Borowiak *et al.*, Photoswitchable Inhibitors of Microtubule Dynamics Optically Control Mitosis and Cell Death. *Cell* **162**, (2015), doi:10.1016/j.cell.2015.06.049.
135. C. Chen *et al.*, Photo-triggered micelles: simultaneous activation and release of microtubule inhibitors for on-demand chemotherapy. *Biomaterials Science* **6**, (2018), doi:10.1039/c7bm01053b.
136. J. Zenker *et al.*, A microtubule-organizing center directing intracellular transport in the early mouse embryo. *Science* **357**, (2017), doi:10.1126/science.aam9335.
137. D. Kolarski *et al.*, Reversible modulation of circadian time with chronophotopharmacology. *Nature Communications* **12**, (2021), doi:10.1038/s41467-021-23301-x.
138. P. Pfaff, K. T. G. Samarasinghe, C. M. Crews, E. M. Carreira, Reversible Spatiotemporal Control of Induced Protein Degradation by Bistable PhotoPROTACs. *ACS Central Science* **5**, (2019), doi:10.1021/acscentsci.9b00713.
139. M. Reynders *et al.*, PHOTACs enable optical control of protein degradation. *Science Advances* **6**, (2020), doi:10.1126/sciadv.aay5064.
140. Q. Zhang *et al.*, Light-mediated multi-target protein degradation using arylazopyrazole photoswitchable PROTACs (AP-PROTACs). *Chemical Communications*, (2022), doi:10.1039/d2cc03092f.
141. S.-L. Dong *et al.*, A Photocontrolled β-Hairpin Peptide. *Chemistry - A European Journal* **12**, (2006), doi:10.1002/chem.200500986.
142. S. Jurt, A. Aemissegger, P. Güntert, O. Zerbe, D. Hilvert, A Photoswitchable Miniprotein Based on the Sequence of Avian Pancreatic Polypeptide. *Angewandte Chemie* **118**, (2006), doi:10.1002/ange.200602084.
143. J. Broichhagen *et al.*, Optical Control of Insulin Secretion Using an Incretin Switch. *Angewandte Chemie International Edition* **54**, (2015), doi:10.1002/anie.201506384.
144. C. Hoppmann, P. Schmieder, N. Heinrich, M. Beyermann, Photoswitchable Click Amino Acids: Light Control of Conformation and Bioactivity. *ChemBioChem* **12**, (2011), doi:10.1002/cbic.201100578.
145. N. S. A. Crone *et al.*, Azobenzene-Based Amino Acids for the Photocontrol of Coiled-Coil Peptides. *Bioconjugate Chemistry*, (2023), doi:10.1021/acs.bioconjchem.2c00534.
146. J. M. Torner, P. S. Arora, Conformational control in a photoswitchable coiled coil. *Chemical Communications* **57**, (2021), doi:10.1039/d0cc08318f.
147. D. C. Burns, F. Zhang, G. A. Woolley, Synthesis of 3,3'-bis(sulfonato)-4,4'-bis(chloroacetamido)azobenzene and cysteine cross-linking for photo-control of protein conformation and activity. *Nature protocols* **2**, (2007), doi:10.1038/nprot.2007.21.
148. J. R. Kumita, O. S. Smart, G. A. Woolley, Photo-control of helix content in a short peptide. *Proceedings of the National Academy of Sciences* **97**, (2000), doi:10.1073/pnas.97.8.3803.
149. D. G. Flint, J. R. Kumita, O. S. Smart, G. A. Woolley, Using an Azobenzene Cross-Linker to Either Increase or Decrease Peptide Helix Content upon Trans-to-Cis Photoisomerization. *Chemistry & Biology* **9**, (2002), doi:10.1016/s1074-5521(02)00109-6.
150. F. Zhang, O. Sadovski, G. A. Woolley, Synthesis and Characterization of a Long, Rigid Photoswitchable Cross-Linker for Promoting Peptide and Protein Conformational Change. *ChemBioChem* **9**, (2008), doi:10.1002/cbic.200800196.
151. J. Volarić, S. Thallmair, B. L. Feringa, W. Szymanski, Photoswitchable, Water-Soluble Bisazobenzene Cross-Linkers with Enhanced Properties for Biological Applications. *ChemPhotoChem* **6**, (2022), doi:10.1002/cptc.202200170.
152. B. Schierling *et al.*, Controlling the enzymatic activity of a restriction enzyme by light. *Proceedings of the National Academy of Sciences* **107**, (2010), doi:10.1073/pnas.0909444107.
153. K. M. Blacklock, B. J. Yachnin, G. A. Woolley, S. D. Khare, Computational Design of a Photocontrolled Cytosine Deaminase. *Journal of the American Chemical Society* **140**, (2018), doi:10.1021/jacs.7b08709.
154. L. Guerrero *et al.*, Photochemical Regulation of DNA-Binding Specificity of MyoD. *Angewandte Chemie International Edition* **44**, (2005), doi:10.1002/anie.200502666.

155. F. Zhang, K. A. Timm, K. M. Arndt, G. A. Woolley, Photocontrol of coiled-coil proteins in living cells. *Angewandte Chemie (International ed. in English)* **49**, (2010), doi:10.1002/anie.201000909.
156. A. Myrhammar, D. Rosik, A. E. Karlström, Photocontrolled Reversible Binding between the Protein A-Derived Z Domain and Immunoglobulin G. *Bioconjugate chemistry* **31**, (2020), doi:10.1021/acs.bioconjchem.9b00786.
157. B. Buchli *et al.*, Kinetic response of a photoperturbed allosteric protein. *Proceedings of the National Academy of Sciences* **110**, (2013), doi:10.1073/pnas.1306323110.
158. O. Bozovic, B. Jankovic, P. Hamm, Sensing the allosteric force. *Nature Communications* **11**, (2020), doi:10.1038/s41467-020-19689-7.
159. B. Jankovic *et al.*, Photocontrolling Protein-Peptide Interactions: From Minimal Perturbation to Complete Unbinding. *Journal of the American Chemical Society* **141**, (2019), doi:10.1021/jacs.9b03222.
160. T. Ueda, K. Murayama, T. Yamamoto, S. Kimura, Y. Imanishi, Photo-regulation of hydrolysis activity of semisynthetic mutant phospholipases A2 replaced by non-natural aromatic amino acids. *Journal of the Chemical Society, Perkin Transactions 1*, (1994), doi:10.1039/p19940000225.
161. M. V. Rodnina, W. Wintermeyer, Fidelity of Aminoacyl-tRNA Selection on the Ribosome: Kinetic and Structural Mechanisms. *Annual Review of Biochemistry* **70**, (2001), doi:10.1146/annurev.biochem.70.1.415.
162. J. Doherty, M. Guo, in *Encyclopedia of Cell Biology,* R. A. Bradshaw, P. D. Stahl, Eds. (Academic Press, Waltham, 2016), pp. 309-340.
163. J. Lehmann, A. Libchaber, Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. *RNA* **14**, (2008), doi:10.1261/rna.1029808.
164. T. M. Schmeing, V. Ramakrishnan, What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, (2009), doi:10.1038/nature08403.
165. J. C. M. van Hest, D. A. Tirrell, Efficient introduction of alkene functionality into proteins in vivo. *FEBS Letters* **428**, (1998), doi:10.1016/S0014-5793(98)00489-X.
166. D. Belin, P. Puigbò, Why Is the UAG (Amber) Stop Codon Almost Absent in Highly Expressed Bacterial Genes? *Life* **12**, (2022), doi:10.3390/life12030431.
167. L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the Genetic Code of *Escherichia coli*. *Science* **292**, (2001), doi:10.1126/science.1060077.
168. D. La Torre, J. W. Chin, Reprogramming the genetic code. *Nature reviews. Genetics*, (2020), doi:10.1038/s41576-020-00307-7.
169. L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, A New Functional Suppressor tRNA/Aminoacyl−tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *Journal of the American Chemical Society* **122**, (2000), doi:10.1021/ja000595y.
170. B. Hao *et al.*, A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase. *Science* **296**, (2002), doi:10.1126/science.1069556.
171. T. Yanagisawa, R. Ishii, R. Fukunaga, O. Nureki, S. Yokoyama, Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon Methanosarcina mazei. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* **62**, (2006), doi:10.1107/s1744309106036700.
172. H.-S. Park *et al.*, Expanding the Genetic Code of Escherichia coli with Phosphoserine. *Science* **333**, (2011), doi:10.1126/science.1207203.
173. M. S. Zhang *et al.*, Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nature Methods* **14**, (2017), doi:10.1038/nmeth.4302.
174. D. T. Rogerson *et al.*, Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nature Chemical Biology* **11**, (2015), doi:10.1038/nchembio.1823.
175. V. Beránek *et al.*, Genetically Encoded Protein Phosphorylation in Mammalian Cells. *Cell Chemical Biology* **25**, (2018), doi:10.1016/j.chembiol.2018.05.013.
176. J. Andrews, Q. Gan, C. Fan, "Not-so-popular" orthogonal pairs in genetic code expansion. *Protein Science* **32**, (2023), doi:10.1002/pro.4559.
177. J. S. Italia *et al.*, An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nature Chemical Biology* **13**, (2017), doi:10.1038/nchembio.2312.

178. J. S. Italia, C. Latour, C. J. J. Wrobel, A. Chatterjee, Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. *Cell Chemical Biology* **25**, (2018), doi:10.1016/j.chembiol.2018.07.002.
179. T. Suzuki *et al.*, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nature chemical biology* **13**, (2017), doi:10.1038/nchembio.2497.
180. B. Meineke, J. Heimgärtner, L. Lafranchi, S. J. Elsässer, Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. *ACS chemical biology* **13**, (2018), doi:10.1021/acschembio.8b00571.
181. J. C. W. Willis, J. W. Chin, Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nature chemistry* **10**, (2018), doi:10.1038/s41557-018-0052-5.
182. D. L. Dunkelmann, J. C. W. Willis, A. T. Beattie, J. W. Chin, Engineered triply orthogonal pyrrolysyl–tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nature Chemistry* **12**, (2020), doi:10.1038/s41557-020-0472-x.
183. H.-K. Jiang *et al.*, Split aminoacyl-tRNA synthetases for proximity-induced stop codon suppression. *Proceedings of the National Academy of Sciences* **120**, (2023), doi:10.1073/pnas.2219758120.
184. W. Ding *et al.*, Chimeric design of pyrrolysyl-tRNA synthetase/tRNA pairs and canonical synthetase/tRNA pairs for genetic code expansion. *Nature Communications* **11**, (2020), doi:10.1038/s41467-020-16898-y.
185. H. Zhao *et al.*, Directed-evolution of translation system for efficient unnatural amino acids incorporation and generalizable synthetic auxotroph construction. *Nature Communications* **12**, (2021), doi:10.1038/s41467-021-27399-x.
186. A. J. Rovner *et al.*, Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, (2015), doi:10.1038/nature14095.
187. D. J. Mandell *et al.*, Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, (2015), doi:10.1038/nature14121.
188. J. A. Fok, C. Mayer, Genetic-Code-Expansion Strategies for Vaccine Development. *ChemBioChem* **21**, (2020), doi:10.1002/cbic.202000343.
189. L. Davis, J. W. Chin, Designer proteins: applications of genetic code expansion in cell biology. *Nature Reviews Molecular Cell Biology* **13**, (2012), doi:10.1038/nrm3286.
190. K. M. Esvelt, J. C. Carlson, D. R. Liu, A system for the continuous directed evolution of biomolecules. *Nature* **472**, (2011), doi:10.1038/nature09929.
191. D. I. Bryson *et al.*, Continuous directed evolution of aminoacyl-tRNA synthetases. *Nature chemical biology* **13**, (2017), doi:10.1038/nchembio.2474.
192. A. J. Link *et al.*, Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proceedings of the National Academy of Sciences* **103**, (2006), doi:10.1073/pnas.0601167103.
193. A. E. Lin, Q. Lin, Rapid Identification of Functional Pyrrolysyl-tRNA Synthetases via Fluorescence-Activated Cell Sorting. *International journal of molecular sciences* **20**, (2018), doi:10.3390/ijms20010029.
194. A. Hohl *et al.*, Engineering a Polyspecific Pyrrolysyl-tRNA Synthetase by a High Throughput FACS Screen. *Scientific reports* **9**, (2019), doi:10.1038/s41598-019-48357-0.
195. M. D. Bartoschek *et al.*, Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells. *Nucleic Acids Research* **49**, (2021), doi:10.1093/nar/gkab132.
196. R. Serfling *et al.*, Designer tRNAs for efficient incorporation of non-canonical amino acids by the pyrrolysine system in mammalian cells. *Nucleic acids research* **46**, (2018), doi:10.1093/nar/gkx1156.
197. W. Brown *et al.*, Chemically Acylated tRNAs are Functional in Zebrafish Embryos. *Journal of the American Chemical Society*, (2023), doi:10.1021/jacs.2c11452.
198. K. Wang, H. Neumann, S. Y. Peak-Chew, J. W. Chin, Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature Biotechnology* **25**, (2007), doi:10.1038/nbt1314.
199. T. Mukai *et al.*, Codon reassignment in the Escherichia coli genetic code. *Nucleic Acids Research* **38**, (2010), doi:10.1093/nar/gkq707.

200. D. B. F. Johnson *et al.*, RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nature Chemical Biology* **7**, (2011), doi:10.1038/nchembio.657.
201. M. J. Lajoie *et al.*, Genomically Recoded Organisms Expand Biological Functions. *Science* **342**, (2013), doi:10.1126/science.1241459.
202. T. Mukai *et al.*, Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon. *Scientific Reports* **5**, (2015), doi:10.1038/srep09699.
203. W. H. Schmied, S. J. Elsässer, C. Uttamapinant, J. W. Chin, Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1. *Journal of the American Chemical Society* **136**, (2014), doi:10.1021/ja5069728.
204. N. Muranaka, T. Hohsaka, M. Sisido, Photoswitching of peroxidase activity by position-specific incorporation of a photoisomerizable non-natural amino acid into horseradish peroxidase. *FEBS Letters* **510**, (2002), doi:10.1016/s0014-5793(01)03211-2.
205. J. C. Anderson *et al.*, An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Sciences* **101**, (2004), doi:10.1073/pnas.0401517101.
206. K. Wang, W. H. Schmied, J. W. Chin, Reprogramming the Genetic Code: From Triplet to Quadruplet Codes. *Angewandte Chemie International Edition* **51**, (2012), doi:10.1002/anie.201105016.
207. H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, (2010), doi:10.1038/nature08817.
208. K. Wang *et al.*, Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nature Chemistry* **6**, (2014), doi:10.1038/nchem.1919.
209. D. De La Torre, J. W. Chin, Reprogramming the genetic code. *Nature Reviews Genetics* **22**, (2021), doi:10.1038/s41576-020-00307-7.
210. D. L. Dunkelmann, S. B. Oehm, A. T. Beattie, J. W. Chin, A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design. *Nature Chemistry* **13**, (2021), doi:10.1038/s41557-021-00764-5.
211. J. Fredens *et al.*, Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, (2019), doi:10.1038/s41586-019-1192-5.
212. M. Spinck *et al.*, Genetically programmed cell-based synthesis of non-natural peptide and depsipeptide macrocycles. *Nature Chemistry*, (2022), doi:10.1038/s41557-022-01082-0.
213. S. Ishida, N. Terasaka, T. Katoh, H. Suga, An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. *Nature chemical biology* **16**, (2020), doi:10.1038/s41589-020-0500-6.
214. H. Saito, H. Suga, A Ribozyme Exclusively Aminoacylates the 3‘-Hydroxyl Group of the tRNA Terminal Adenosine. *Journal of the American Chemical Society* **123**, (2001), doi:10.1021/ja015756s.
215. J. Lee *et al.*, Expanding the limits of the second genetic code with ribozymes. *Nature communications* **10**, (2019), doi:10.1038/s41467-019-12916-w.
216. J. Morimoto, Y. Hayashi, K. Iwasaki, H. Suga, Flexizymes: Their Evolutionary History and the Origin of Catalytic Function. *Accounts of Chemical Research* **44**, (2011), doi:10.1021/ar2000953.
217. J. N. Coronado, P. Ngo, E. V. Anslyn, A. D. Ellington, Chemical insights into flexizyme-mediated tRNA acylation. *Cell Chemical Biology* **29**, (2022), doi:10.1016/j.chembiol.2022.03.012.
218. H. Xiong *et al.*, Dual Genetic Encoding of Acetyl-lysine and Non-deacetylatable Thioacetyl-lysine Mediated by Flexizyme. *Angewandte Chemie (International ed. in English)* **55**, (2016), doi:10.1002/anie.201511750.
219. C. Tsiamantas *et al.*, Ribosomal Incorporation of Aromatic Oligoamides as Peptide Sidechain Appendages. *Angewandte Chemie (International ed. in English)* **59**, (2020), doi:10.1002/anie.201914654.
220. T. Katoh, H. Suga, Ribosomal Elongation of Cyclic γ-Amino Acids using a Reprogrammed Genetic Code. *Journal of the American Chemical Society* **142**, (2020), doi:10.1021/jacs.9b12280.

221. T. Katoh, T. Sengoku, K. Hirata, K. Ogata, H. Suga, Ribosomal synthesis and de novo discovery of bioactive foldamer peptides containing cyclic β-amino acids. *Nature Chemistry* **12**, (2020), doi:10.1038/s41557-020-0525-1.
222. A. A. Vinogradov *et al.*, Minimal lactazole scaffold for in vitro thiopeptide bioengineering. *Nature Communications* **11**, (2020), doi:10.1038/s41467-020-16145-4.
223. K. Sakai *et al.*, Designing receptor agonists with enhanced pharmacokinetics by grafting macrocyclic peptides into fragment crystallizable regions. *Nature Biomedical Engineering*, (2022), doi:10.1038/s41551-022-00955-6.
224. W. H. Schmied *et al.*, Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* **564**, (2018), doi:10.1038/s41586-018-0773-z.
225. E. D. Carlson *et al.*, Engineered ribosomes with tethered subunits for expanding biological function. *Nature Communications* **10**, (2019), doi:10.1038/s41467-019-11427-y.
226. M. Kimoto, I. Hirao, Genetic alphabet expansion technology by creating unnatural base pairs. *Chemical Society Reviews* **49**, (2020), doi:10.1039/d0cs00457j.
227. M. Kimoto, I. Hirao, Genetic Code Engineering by Natural and Unnatural Base Pair Systems for the Site-Specific Incorporation of Non-Standard Amino Acids Into Proteins. *Front Mol Biosci* **9**, (2022), doi:10.3389/fmolb.2022.851646.
228. K. Krauskopf, K. Lang, Increasing the chemical space of proteins in living cells via genetic code expansion. *Current opinion in chemical biology* **58**, (2020), doi:10.1016/j.cbpa.2020.07.012.
229. K. Lang, J. W. Chin, Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* **114**, (2014), doi:10.1021/cr400355w.
230. H. S. Jang, S. Jana, R. J. Blizzard, J. C. Meeuwsen, R. A. Mehl, Access to Faster Eukaryotic Cell Labeling with Encoded Tetrazine Amino Acids. *Journal of the American Chemical Society* **142**, (2020), doi:10.1021/jacs.9b11520.
231. M. J. Gattner, M. Vrabel, T. Carell, Synthesis of ε-N-propionyl-, ε-N-butyryl-, and ε-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chemical Communications* **49**, (2013), doi:10.1039/C2CC37836A.
232. S. Jana *et al.*, Ultra-Fast Bioorthogonal Spin-Labeling and Distance Measurements in Mammalian Cells Using Small, Genetically Encoded Tetrazine Amino Acids. *bioRxiv*, (2023), doi:10.1101/2023.01.26.525763.
233. C. Chou, R. Uprety, L. Davis, J. W. Chin, A. Deiters, Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chemical Science* **2**, (2011), doi:10.1039/C0SC00373E.
234. H.-w. Ai, W. Shen, A. Sagi, P. R. Chen, P. G. Schultz, Probing Protein–Protein Interactions with a Genetically Encoded Photo-crosslinking Amino Acid. *ChemBioChem* **12**, (2011), doi:10.1002/cbic.201100194.
235. N. Huguenin-Dezot *et al.*, Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* **565**, (2019), doi:10.1038/s41586-018-0781-z.
236. P. Wu, The Nobel Prize in Chemistry 2022: Fulfilling Demanding Applications with Simple Reactions. *ACS Chemical Biology* **17**, (2022), doi:10.1021/acschembio.2c00788.
237. T.-A. Nguyen, M. Cigler, K. Lang, Expanding the Genetic Code to Study Protein-Protein Interactions. *Angewandte Chemie (International ed. in English)* **57**, (2018), doi:10.1002/anie.201805869.
238. N. Wang, L. Wang, Genetically encoding latent bioreactive amino acids and the development of covalent protein drugs. *Current Opinion in Chemical Biology* **66**, (2022), doi:10.1016/j.cbpa.2021.102106.
239. D. P. Nguyen *et al.*, Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells. *Journal of the American Chemical Society* **136**, (2014), doi:10.1021/ja412191m.
240. E. Arbely, J. Torres-Kolbus, A. Deiters, J. W. Chin, Photocontrol of Tyrosine Phosphorylation in Mammalian Cells via Genetic Encoding of Photocaged Tyrosine. *Journal of the American Chemical Society* **134**, (2012), doi:10.1021/ja3046958.
241. T. M. Courtney, A. Deiters, Optical control of protein phosphatase function. *Nature Communications* **10**, (2019), doi:10.1038/s41467-019-12260-z.

242. A. Ryan, J. Liu, A. Deiters, Targeted Protein Degradation through Fast Optogenetic Activation and Its Application to the Control of Cell Signaling. *Journal of the American Chemical Society* **143**, (2021), doi:10.1021/jacs.1c04324.
243. T. Bridge *et al.*, Site-Specific Encoding of Photoactivity in Antibodies Enables Light-Mediated Antibody–Antigen Binding on Live Cells. *Angewandte Chemie International Edition* **58**, (2019), doi:10.1002/anie.201908655.
244. B. Jedlitzke, Z. Yilmaz, W. Dörner, H. D. Mootz, Photobodies: Light-Activatable Single-Domain Antibody Fragments. *Angewandte Chemie International Edition* **59**, (2020), doi:10.1002/anie.201912286.
245. B. Jedlitzke, H. D. Mootz, A Light-Activatable Photocaged Variant of the Ultra-High Affinity ALFA-Tag Nanobody. *ChemBioChem* **23**, (2022), doi:10.1002/cbic.202200079.
246. J. Wang *et al.*, Time-resolved protein activation by proximal decaging in living systems. *Nature* **569**, (2019), doi:10.1038/s41586-019-1188-1.
247. M. Bose, D. Groff, J. Xie, E. Brustad, P. G. Schultz, The incorporation of a photoisomerizable amino acid into proteins in E. coli. *Journal of the American Chemical Society* **128**, (2006), doi:10.1021/ja055467u.
248. K. Nakayama, M. Endo, T. Majima, Photochemical regulation of the activity of an endonuclease BamHI using an azobenzene moiety incorporated site-selectively into the dimer interface. *Chemical Communications*, (2004), doi:10.1039/b409844g.
249. C. Hoppmann *et al.*, Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells. *Angewandte Chemie (International ed. in English)* **53**, (2014), doi:10.1002/anie.201400001.
250. C. Hoppmann, I. Maslennikov, S. Choe, L. Wang, In Situ Formation of an Azo Bridge on Proteins Controllable by Visible Light. *Journal of the American Chemical Society* **137**, (2015), doi:10.1021/jacs.5b06234.
251. J. Luo, S. Samanta, M. Convertino, N. V. Dokholyan, A. Deiters, Reversible and Tunable Photoswitching of Protein Function through Genetic Encoding of Azobenzene Amino Acids in Mammalian Cells. *Chembiochem : a European journal of chemical biology* **19**, (2018), doi:10.1002/cbic.201800226.
252. V. H. Klippenstein, Christina; Ye, Shixin; Wang, Lei; Paoletti, Pierre, Optocontrol of glutamate receptor activity by single side-chain photoisomerization. *eLife*, (2017), doi:10.7554/eLife.25808.001.
253. A. C. Kneuttinger *et al.*, Light Regulation of Enzyme Allostery through Photo-responsive Unnatural Amino Acids. *Cell chemical biology* **26**, (2019), doi:10.1016/j.chembiol.2019.08.006.
254. A. C. Kneuttinger *et al.*, Significance of the Protein Interface Configuration for Allostery in Imidazole Glycerol Phosphate Synthase. *Biochemistry* **59**, (2020), doi:10.1021/acs.biochem.0c00332.
255. T. Zheng *et al.*, Photo-regulated genetic encoding of dibenzo[c,g][1,2]diazocine on proteins via configuration switching. *Chemical Communications* **59**, (2023), doi:10.1039/d2cc06738b.
256. R. Tomar, S. Suwasia, A. R. Choudhury, S. Venkataramani, S. A. Babu, Azobenzene-based unnatural amino acid scaffolds via a Pd(II)-catalyzed C(sp3)–H arylation strategy. *Chemical Communications* **58**, (2022), doi:10.1039/d2cc04870a.
257. E. Merino, Synthesis of azobenzenes: the coloured pieces of molecular materials. *Chemical Society reviews* **40**, (2011), doi:10.1039/c0cs00183j.
258. A. A. John, C. P. Ramil, Y. Tian, G. Cheng, Q. Lin, Synthesis and Site-Specific Incorporation of Red-Shifted Azobenzene Amino Acids into Proteins. *Organic letters* **17**, (2015), doi:10.1021/acs.orglett.5b03268.
259. K. Krauskopf, Photoswitchable Unnatural Amino Acids – a Tool for Genetic Code Expansion. *Master's Thesis*, Ludwig-Maximilians-Universität München, (2017).
260. L. Gao *et al.*, A Robust, GFP-Orthogonal Photoswitchable Inhibitor Scaffold Extends Optical Control over the Microtubule Cytoskeleton. *Cell Chemical Biology* **28**, (2021), doi:10.1016/j.chembiol.2020.11.007.
261. F. Mirazizi *et al.*, Rapid and direct spectrophotometric method for kinetics studies and routine assay of peroxidase based on aniline diazo substrates. *Journal of enzyme inhibition and medicinal chemistry* **31**, (2016), doi:10.3109/14756366.2015.1103234.
262. T. A. Lutz, P. Spanner, K. T. Wanner, A general approach to substituted diphenyldiazenes. *Tetrahedron* **72**, (2016), doi:10.1016/j.tet.2016.02.011.

263. M. Schönberger, M. Althaus, M. Fronius, W. Clauss, D. Trauner, Controlling epithelial sodium channels with light using photoswitchable amilorides. *Nature chemistry* **6**, (2014), doi:10.1038/nchem.2004.
264. P. Zhang, M. Cedilote, T. P. Cleary, M. E. Pierce, Mono-nitration of aromatic compounds via their nitric acid salts. *Tetrahedron Letters* **48**, (2007), doi:10.1016/j.tetlet.2007.10.027.
265. N. G. Koch, T. Baumann, J. H. Nickling, A. Dziegielewski, N. Budisa, Engineered bacterial host for genetic encoding of physiologically stable protein nitration. *Front Mol Biosci* **9**, (2022), doi:10.3389/fmolb.2022.992748.
266. T. Baba *et al.*, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2**, (2006), doi:10.1038/msb4100050.
267. M. J. Hansen, M. M. Lerch, W. Szymanski, B. L. Feringa, Direct and Versatile Synthesis of Red-Shifted Azobenzenes. *Angewandte Chemie International Edition* **55**, (2016), doi:10.1002/anie.201607529.
268. D. B. Konrad, J. A. Frank, D. Trauner, Synthesis of Redshifted Azobenzene Photoswitches by Late-Stage Functionalization. *Chemistry (Weinheim an der Bergstrasse, Germany)* **22**, (2016), doi:10.1002/chem.201505061.
269. Q. Liu *et al.*, Palladium-catalyzed direct ortho C-X bond construction via C-H activation of azobenzenes: Synthesis of (E)-1,2-b(2,6-dibromo(chloro)-phenyl)diazene. *Tetrahedron Letters* **60**, (2019), doi:10.1016/j.tetlet.2019.05.056.
270. A. R. Hergueta, Easy Removal of Triphenylphosphine Oxide from Reaction Mixtures by Precipitation with $CaBr_2$. *Organic Process Research & Development* **26**, (2022), doi:10.1021/acs.oprd.2c00104.
271. A. J. Ross, H. L. Lang, R. F. W. Jackson, Much Improved Conditions for the Negishi Cross-Coupling of Iodoalanine Derived Zinc Reagents with Aryl Halides. *The Journal of Organic Chemistry* **75**, (2010), doi:10.1021/jo902238n.
272. A. K. Gaur *et al.*, Structure–Property Relationship for Visible Light Bidirectional Photoswitchable Azoheteroarenes and Thermal Stability of Z-Isomers. *The Journal of Organic Chemistry*, (2022), doi:10.1021/acs.joc.2c00088.
273. N. A. Simeth, S. Crespi, M. Fagnoni, B. König, Tuning the Thermal Isomerization of Phenylazoindole Photoswitches from Days to Nanoseconds. *Journal of the American Chemical Society* **140**, (2018), doi:10.1021/jacs.7b12871.
274. D. Fang *et al.*, (Hetero)arylazo-1,2,3-triazoles: "Clicked" Photoswitches for Versatile Functionalization and Electronic Decoupling. *Journal of the American Chemical Society*, (2021), doi:10.1021/jacs.1c08704.
275. P. S. Addy, S. B. Erickson, J. S. Italia, A. Chatterjee, A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *Journal of the American Chemical Society* **139**, (2017), doi:10.1021/jacs.7b05125.
276. C.-C. Cho, L. R. Blankenship, X. Ma, S. Xu, W. Liu, The Pyrrolysyl-tRNA Synthetase Activity can be Improved by a P188 Mutation that Stabilizes the Full-Length Enzyme. *Journal of Molecular Biology* **434**, (2022), doi:10.1016/j.jmb.2022.167453.
277. S. Samanta, T. M. McCormick, S. K. Schmidt, D. S. Seferos, G. A. Woolley, Robust visible light photoswitching with ortho-thiol substituted azobenzenes. *Chemical Communications* **49**, (2013), doi:10.1039/C3CC46045B.
278. J. Berry, T. K. Lindhorst, G. Despras, Sulfur and Azobenzenes, a Profitable Liaison: Straightforward Synthesis of Photoswitchable Thioglycosides with Tunable Properties. *Chemistry – A European Journal*, (2022), doi:10.1002/chem.202200354.
279. G. A. Woolley, Photocontrolling Peptide α Helices. *Accounts of Chemical Research* **38**, (2005), doi:10.1021/ar040091v.
280. J. Löfblom *et al.*, Affibody molecules: Engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS Letters* **584**, (2010), doi:10.1016/j.febslet.2010.04.014.
281. P. A. Lund, in *Advances in Microbial Physiology*. (Academic Press, 2001), vol. 44, pp. 93-140.
282. J. Wu, C. Liu, J. Chen, X. Yu, RAP80 Protein Is Important for Genomic Stability and Is Required for Stabilizing BRCA1-A Complex at DNA Damage Sites in Vivo. *Journal of Biological Chemistry* **287**, (2012), doi:10.1074/jbc.m112.351007.

283. M. Fottner *et al.*, A modular toolbox to generate complex polymeric ubiquitin architectures using orthogonal sortase enzymes. *Nature Communications* **12**, (2021), doi:10.1038/s41467-021-26812-9.
284. M. Zhang, W. Wang, Organization of Signaling Complexes by PDZ-Domain Scaffold Proteins. *Accounts of Chemical Research* **36**, (2003), doi:10.1021/ar020210b.
285. S. Muyldermans, Nanobodies: Natural Single-Domain Antibodies. *Annual Review of Biochemistry* **82**, (2013), doi:10.1146/annurev-biochem-063011-092449.
286. H. Farrants *et al.*, Chemogenetic Control of Nanobodies. *Nature methods* **17**, (2020), doi:10.1038/s41592-020-0746-7.
287. A. Kirchhofer *et al.*, Modulation of protein properties in living cells using nanobodies. *Nature Structural & Molecular Biology* **17**, (2010), doi:10.1038/nsmb.1727.
288. M. Fottner *et al.*, Site-Specific Protein Labeling and Generation of Defined Ubiquitin-Protein Conjugates Using an Asparaginyl Endopeptidase. *Journal of the American Chemical Society* **144**, (2022), doi:10.1021/jacs.2c02191.
289. M. S. Wang, M. H. Hecht, A Completely *De Novo* ATPase from Combinatorial Protein Design. *Journal of the American Chemical Society* **142**, (2020), doi:10.1021/jacs.0c02954.
290. S. H. Studer, Douglas A.; Hilvert, Donald, Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **362**, (2018), doi.
291. Y. Wei, S. Kim, D. Fela, J. Baum, M. H. Hecht, Solution structure of a *de novo* protein from a designed combinatorial library. *Proceedings of the National Academy of Sciences* **100**, (2003), doi:10.1073/pnas.1835644100.
292. B. S. Der *et al.*, Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed Protein Homodimer. *Journal of the American Chemical Society* **134**, (2012), doi:10.1021/ja208015j.
293. K. Mcluskey, C. Mottram, Jeremy, Comparative structural analysis of the caspase family with other clan CD cysteine peptidases. *Biochemical Journal* **466**, (2015), doi:10.1042/bj20141324.
294. M. Gersch, A. List, M. Groll, S. A. Sieber, Insights into Structural Network Responsible for Oligomerization and Activity of Bacterial Virulence Regulator Caseinolytic Protease P (ClpP) Protein. *Journal of Biological Chemistry* **287**, (2012), doi:10.1074/jbc.m111.336222.
295. F. Ye *et al.*, Helix unfolding/refolding characterizes the functional dynamics of Staphylococcus aureus Clp protease. *The Journal of biological chemistry* **288**, (2013), doi:10.1074/jbc.M113.452714.
296. N. Yasuike, H. Lu, P. Xia, G. A. Woolley, in *Methods in Enzymology,* A. Deiters, Ed. (Academic Press, 2019), vol. 624, pp. 129-149.
297. M. Lakemeyer *et al.*, Tailored Peptide Phenyl Esters Block ClpXP Proteolysis by an Unusual Breakdown into a Heptamer–Hexamer Assembly. *Angewandte Chemie International Edition* **58**, (2019), doi:10.1002/anie.201901056.
298. M. Schwarz, I. Hübner, S. A. Sieber, Tailored phenyl esters inhibit ClpXP and attenuate Staphylococcus aureus α-hemolysin secretion. *ChemBioChem*, (2022), doi:10.1002/cbic.202200253.
299. T. G. W. Edwardson, T. Mori, D. Hilvert, Rational Engineering of a Designed Protein Cage for siRNA Delivery. *Journal of the American Chemical Society* **140**, (2018), doi:10.1021/jacs.8b06442.
300. T. G. W. Edwardson, S. Tetter, D. Hilvert, Two-tier supramolecular encapsulation of small molecules in a protein cage. *Nature Communications* **11**, (2020), doi:10.1038/s41467-020-19112-1.
301. M. Hörner *et al.*, Spatiotemporally confined red light-controlled gene delivery at single-cell resolution using adeno-associated viral vectors. *Science Advances* **7**, (2021), doi:10.1126/sciadv.abf0797.
302. D. Hoersch, S.-H. Roh, W. Chiu, T. Kortemme, Reprogramming an ATP-driven protein machine into a light-gated nanocage. *Nature Nanotechnology* **8**, (2013), doi:10.1038/nnano.2013.242.
303. A. M. Ali, G. A. Woolley, The effect of azobenzene cross-linker position on the degree of helical peptide photo-control. *Organic & Biomolecular Chemistry* **11**, (2013), doi:10.1039/c3ob40684a.
304. R. M. Woloschuk *et al.*, Structure-based design of a photoswitchable affibody scaffold. *Protein Science* **30**, (2021), doi:10.1002/pro.4196.

305. B. Yang *et al.*, Spontaneous and specific chemical cross-linking in live cells to capture and identify protein interactions. *Nature Communications* **8**, (2017), doi:10.1038/s41467-017-02409-z.
306. A. Müller-Deku, O. Thorn-Seshold. (American Chemical Society (ACS), 2022), (10.26434/chemrxiv-2022-vpq96).
307. C. P. Janosko *et al.*, Genetic Encoding of Arylazopyrazole Phenylalanine for Optical Control of Translation. *ACS Omega* **8**, (2023), doi:10.1021/acsomega.3c03512.
308. S. Avila-Crump *et al.*, Generating Efficient Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetases for Structurally Diverse Non-Canonical Amino Acids. *ACS Chemical Biology* **17**, (2022), doi:10.1021/acschembio.2c00639.
309. D. Komander, M. Rape, The Ubiquitin Code. *Annual Review of Biochemistry* **81**, (2012), doi:10.1146/annurev-biochem-060310-170328.
310. D. R. Squair, S. Virdee, A new dawn beyond lysine ubiquitination. *Nature Chemical Biology* **18**, (2022), doi:10.1038/s41589-022-01088-2.
311. K. N. Swatek, D. Komander, Ubiquitin modifications. *Cell Research* **26**, (2016), doi:10.1038/cr.2016.39.
312. E. Oh, D. Akopian, M. Rape, Principles of Ubiquitin-Dependent Signaling. *Annual Review of Cell and Developmental Biology* **34**, (2018), doi:10.1146/annurev-cellbio-100617-062802.
313. A. Hershko, A. Ciechanover, THE UBIQUITIN SYSTEM. *Annual Review of Biochemistry* **67**, (1998), doi:10.1146/annurev.biochem.67.1.425.
314. A. G. Van Der Veen, H. L. Ploegh, Ubiquitin-Like Proteins. *Annual Review of Biochemistry* **81**, (2012), doi:10.1146/annurev-biochem-093010-153308.
315. M. J. Clague, S. Urbé, D. Komander, Breaking the chains: deubiquitylating enzyme specificity begets function. *Nature Reviews Molecular Cell Biology* **20**, (2019), doi:10.1038/s41580-019-0099-1.
316. Y. Jia, L. A. Claessens, A. C. O. Vertegaal, H. Ovaa, Chemical Tools and Biochemical Assays for SUMO Specific Proteases (SENPs). *ACS chemical biology* **14**, (2019), doi:10.1021/acschembio.9b00402.
317. M. Altun *et al.*, Activity-Based Chemical Proteomics Accelerates Inhibitor Development for Deubiquitylating Enzymes. *Chemistry & Biology* **18**, (2011), doi:10.1016/j.chembiol.2011.08.018.
318. D. S. Hewings, J. A. Flygare, M. Bogyo, I. E. Wertz, Activity-based probes for the ubiquitin conjugation-deconjugation machinery: new chemistries, new tools, and new insights. *The FEBS journal* **284**, (2017), doi:10.1111/febs.14039.
319. M. P. C. Mulder, K. F. Witting, H. Ovaa, Cracking the Ubiquitin Code: The Ubiquitin Toolbox. *Current Issues in Molecular Biology*, (2020), doi:10.21775/cimb.037.001.
320. L. C. Dang, F. D. Melandri, R. L. Stein, Kinetic and Mechanistic Studies on the Hydrolysis of Ubiquitin C-Terminal 7-Amido-4-Methylcoumarin by Deubiquitinating Enzymes. *Biochemistry* **37**, (1998), doi:10.1021/bi9723360.
321. W. Gui *et al.*, Cell-Permeable Activity-Based Ubiquitin Probes Enable Intracellular Profiling of Human Deubiquitinases. *Journal of the American Chemical Society* **140**, (2018), doi:10.1021/jacs.8b05147.
322. D. Conole, M. Mondal, J. D. Majmudar, E. W. Tate, Recent Developments in Cell Permeable Deubiquitinating Enzyme Activity-Based Probes. *Frontiers in Chemistry* **7**, (2019), doi.
323. C. Freund, D. Schwarzer, Engineered Sortases in Peptide and Protein Chemistry. *ChemBioChem* **22**, (2021), doi:10.1002/cbic.202000745.
324. B. M. Dorr, H. O. Ham, C. An, E. L. Chaikof, D. R. Liu, Reprogramming the specificity of sortase enzymes. *Proceedings of the National Academy of Sciences* **111**, (2014), doi:10.1073/pnas.1411179111.
325. M. Fottner *et al.*, Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nature Chemical Biology* **15**, (2019), doi:10.1038/s41589-019-0227-4.
326. M. Abo, E. Weerapana, A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *Journal of the American Chemical Society* **137**, (2015), doi:10.1021/jacs.5b04350.
327. M. Abo, D. W. Bak, E. Weerapana, Optimization of Caged Electrophiles for Improved Monitoring of Cysteine Reactivity in Living Cells. *Chembiochem : a European journal of chemical biology* **18**, (2017), doi:10.1002/cbic.201600524.

328. M. E. Flanagan *et al.*, Chemical and Computational Methods for the Characterization of Covalent Reactive Groups for the Prospective Design of Irreversible Inhibitors. *Journal of Medicinal Chemistry* **57**, (2014), doi:10.1021/jm501412a.
329. D. S. Allgäuer *et al.*, Quantification and Theoretical Analysis of the Electrophilicities of Michael Acceptors. *Journal of the American Chemical Society* **139**, (2017), doi:10.1021/jacs.7b05106.
330. A. Ogasawara *et al.*, Red Fluorescence Probe Targeted to Dipeptidylpeptidase-IV for Highly Sensitive Detection of Esophageal Cancer. *Bioconjugate Chemistry* **30**, (2019), doi:10.1021/acs.bioconjchem.9b00198.
331. Y. Urano *et al.*, Rapid Cancer Detection by Topically Spraying a γ-Glutamyltranspeptidase–Activated Fluorescent Probe. *Science Translational Medicine* **3**, (2011), doi:10.1126/scitranslmed.3002823.
332. T. D. W. Claridge *et al.*, Highly (E)-Selective Wadsworth−Emmons Reactions Promoted by Methylmagnesium Bromide. *Organic Letters* **10**, (2008), doi:10.1021/ol802212e.
333. F. S. Gibson, S. C. Bergmeier, H. Rapoport, Selective Removal of an N-BOC Protecting Group in the Presence of a tert-Butyl Ester and Other Acid-Sensitive Groups. *The Journal of Organic Chemistry* **59**, (1994), doi:10.1021/jo00090a045.
334. G. Han, M. Tamaki, V. J. Hruby, Fast, efficient and selective deprotection of the tert-butoxycarbonyl (Boc) group using HCl/dioxane (4 M). *The Journal of Peptide Research* **58**, (2001), doi:10.1034/j.1399-3011.2001.00935.x.
335. N. George, S. Ofori, S. Parkin, S. G. Awuah, Mild deprotection of the N-tert-butyloxycarbonyl (N-Boc) group using oxalyl chloride. *RSC Advances* **10**, (2020), doi:10.1039/d0ra04110f.
336. M. Ordóñez, E. Hernández-Fernández, M. Fernández-Zertuche, O. García-Barradas, O. Muñoz-Muñiz, Practical and Efficient Synthesis of (E)-α,β-Unsaturated Amides Bearing (S)-α-Methylbenzylamine from 2-Phosphonamides via Horner-Wadsworth-Emmons Reaction. *Synlett* **2006**, (2006), doi:10.1055/s-2006-926265.
337. P. Helquist, D. Schauer, Mild Zinc-Promoted Horner-Wadsworth-Emmons Reactions of Diprotic Phosphonate Reagents. *Synthesis* **2006**, (2006), doi:10.1055/s-2006-950292.
338. P. Ruckgaber, Synthesis and application of non-canonical amino acids to study deubiquitinases. *Master's Thesis*, Technical University Munich, (2022).
339. J. Liu, Y. Li, K. K. Deol, E. R. Strieter, Synthesis of Branched Triubiquitin Active-Site Directed Probes. *Organic letters* **21**, (2019), doi:10.1021/acs.orglett.9b02406.
340. J. Boivin, F. Carpentier, R. Jrad, An Expedient, Flexible and Convergent Access to Selectively Protected 1,5-Dicarbonyl Compounds. Applications to the Synthesis of 2,6-Disubstituted Pyridines and Thiopyridines. *Synthesis* **2006**, (2006), doi:10.1055/s-2006-926464.
341. J. W. De Leeuw, E. R. De Waard, T. Beetz, H. O. Huisman, α,β-Unsaturated acetals: Synthesis and mechanism of formation. *Recueil des Travaux Chimiques des Pays-Bas* **92**, (1973), doi:10.1002/recl.19730920915.
342. C. J. O'Donnel, S. D. Burke, Selective Mesylation of Vicinal Diols: A Systematic Case Study. *The Journal of Organic Chemistry* **63**, (1998), doi:10.1021/jo981532p.
343. G. Laconde, M. Amblard, J. Martinez, Unexpected Reactivity of N -Acyl-Benzotriazoles with Aromatic Amines in Acidic Medium (ABAA Reaction). *European Journal of Organic Chemistry* **2019**, (2019), doi:10.1002/ejoc.201801567.
344. K.-Y. Kim, B. C. Kim, H. B. Lee, H. Shin, Nucleophilic fluorination of triflates by tetrabutylammonium bifluoride. *The Journal of organic chemistry* **73**, (2008), doi:10.1021/jo8015659.
345. J. N. M. Shreeve, R. P. Singh, Recent Advances in Nucleophilic Fluorination Reactions of Organic Compounds- Using Deoxofluor and DAST. *Synthesis*, (2002), doi:10.1055/s-2002-35626.
346. L. M. Mcgregor, M. L. Jenkins, C. Kerwin, J. E. Burke, K. M. Shokat, Expanding the Scope of Electrophiles Capable of Targeting K-Ras Oncogenes. *Biochemistry* **56**, (2017), doi:10.1021/acs.biochem.7b00271.
347. M. Bottlinger, Towards the synthesis of redshifted, azobenzene-based unnatural amino acids via tetra-ortho-halogenation. *Bachelor's Thesis*, Technical University Munich, (2019).

348. A. Karmakar *et al.*, Tertiary-butoxycarbonyl (Boc) – A strategic group for N-protection/deprotection in the synthesis of various natural/unnatural N-unprotected aminoacid cyanomethyl esters. *Tetrahedron Letters* **59**, (2018), doi:10.1016/j.tetlet.2018.10.041.
349. J. T. Hammill, S. Miyake-Stoner, J. L. Hazen, J. C. Jackson, R. A. Mehl, Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nature Protocols* **2**, (2007), doi:10.1038/nprot.2007.379.
350. N. Yamamoto *et al.*, Update on the Keio collection of Escherichia coli single-gene deletion mutants. *Molecular Systems Biology* **5**, (2009), doi:10.1038/msb.2009.92.
351. J. Chiu, P. E. March, R. Lee, D. Tillett, Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Research* **32**, (2004), doi:10.1093/nar/gnh172.

II List of Abbreviations

General

Å – angstrom
aa – amino acid
aaRS – aminoacyl-tRNA synthetase
abs – absorption
AI medium – autoinduction medium
aq – aqueous
ATP – adenosin triphosphate
a.u. – arbitrary unit
azo-ncAA – azobenzene-based ncAA
bp – base pair(s)
°C – degree Celsius
calc. – calculated
CD – circular dichroism
cm – centimeter
d – day(s)
δ – chemical shift (NMR spectroscopy)
Δ – thermal relaxation
Da – Dalton
DNA – deoxyribonucleic acid
ε – molar absorption coefficient
E. coli – *Escherichia coli*
eq – equivalent(s)
ESI – electrospray ionization
φ – quantum yield
FCC – flash column chromatography
g – gram(s)
GSH – glutathione
h – hour(s)
hν – photon energy
HPLC – high-performance liquid chromatography
HR-MS – high resolution MS
HWE – Horner-Wadsworth-Emmons
Hz – Hertz
IR – infrared
J – coupling constant (NMR spectroscopy)
λ – wavelength
LC – liquid chromatography
LED – light-emitting diode(s)
LR-MS – low resolution MS
µ – micro
M – molar
Ma – *Methanomethylophilus alvus*
Mb – *Methanosarcina barkeri*
mg – milligram
min – minute(s)
mL – milliliter(s)
mM – millimolar
Mm – *Methanosarcina mazeii*
mmol – millimole(s)
mRNA – messenger RNA
MS – mass spectrometry
ncAA – non-canoncial amino acid
n.d. – not determined
NIR – near-infrared
nm – nanometer
NMR – nuclear magnetic resonance
Nu – nucleophile
o.n. – overnight
OD_{600} – optical density at 600 nm
PAGE – polyacrylamide gel electrophoresis
PCR – polymerase chain reaction
pdb – protein data bank
POI – protein of interest
ppm – part-per-million
PSS – photostationary state
Pyl – pyrrolysine
quant. – quantitative
R_f – retention factor
RNA – ribonucleic acid
RP – reversed phase
rRNA – ribosomal RNA
rt – room temperature
s - second(s)
S. aureus – *Staphylococcus aureus*
S_EAr – electrophilic aromatic substitution
SEC – size exclusion chromatography
S_N – nucleophilic substitution
S_NAr – aromatic nucleophilic substitution
SPPS – solid phase peptide synthesis
τ – relaxation time
$t_{1/2}$ – half-life (relaxation)
TLC – thin layer chromatography
tRNA -transfer RNA
trunc. – truncated
UV – ultraviolet
Vis – visible
wt – wild type

Chemicals and Reagents

AcOH – acetic acid
AMC – 7-amino-4-methylcoumarin
BME – β-mercaptoethanol
Boc – *tert*-butoxycarbonyl
BocK – N_ε-Boc-L-lysin
Boc_2O – di-*tert*-butylcarbonate
BPB – bromophenol blue
CAM – cerium ammonium molybdate
Cbz – benzyloxycarbonyl
$CDCl_3$ – deuterated chloroform
CDI – 1,1'-carbonyldiimidazole
$CHCl_3$ – chloroform
DAST – diethylaminosulfur trifluoride
DBU – 1,8-diazabicyclo[5.4.0]undec-7-ene
DCE – 1,2-dichloroethane
DCM – dichloromethane
DIC – *N*,*N*'-diisopropylcarbodiimide
DIPEA – diisopropylethylamine
DMF – dimethylformamide
DMP – Dess-Martin periodinane
DMSO – dimethylsulfoxide
DMSO-d_6 – deuterated dimethylsulfoxide
DTT – dithiotheitol
EDC – 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
Et_2O – diethyl ether
EtOH – ethanol
EtOAc – ethyl acetate
Fmoc – 9-fluorenylmethoxycarbonyl
HATU – *O*-(7-Azabenzotriazol-1-yl)-*N*,*N*,*N*',-*N*'-tetramethyluroniumhexafluorphosphat
HOBt – hydroxybenzotriazole
$InsP_6$ – inositol hexakisphosphate
IPA – isopropyl alcohol
IPTG – isopropyl-β-D-1-thiogalacto pyranoside
KOtBu – potassium *tert*-butoxide
LiHMDS – lithium-bis(trimethylsilyl)amide
MeCN – acetonitrile
MeI – methyl iodide
MeMgBr – methylmagnesium bromide
$MeNH_2$ – methylamine
Me_2NH – dimethylamine
MeCN – acetonitrile
MeOH – methanol
MeOD – deuterated methanol
Ms – methanesulfonyl, mesyl
MsCl – mesyl chloride
NCS – *N*-chlorosuccinimide
NEt_3 – triethylamine
NIS – *N*-iodosuccinimide
Pd/C – palldium on carbon
Pent - pentane
PMSF – phenylmethylsulfonyl fluoride
PPTS – pyridinium *para*-toluenesulfonate
p-TsOH – *para*-toluenesulfonic acid
SDS – sodium dodecyl sulfate
T3P – propylphosphonic anhydride
TBME – *tert*-butyl methyl ether
TCCA – trichloroisocyanuric acid
TCEP – Tris(2-carboxyethyl)phosphine
TFA – trifluoracetic acid
THF – tetrahydrofuran
TMEDA – *N*,*N*,*N*',*N*'-tetramethylethylendiamine

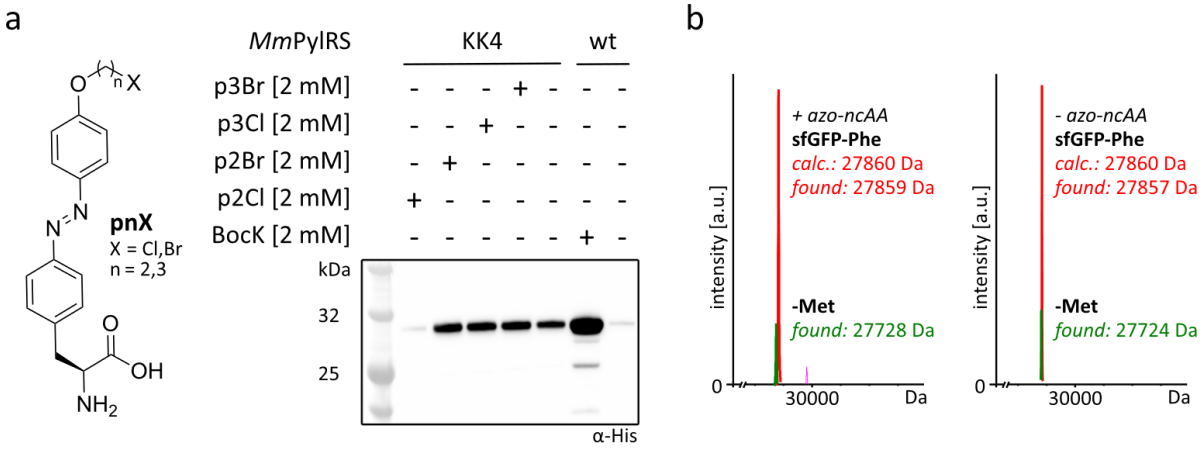
Proteins

AltTPase – alternative ATPase (synthtetic)
CPD – cysteine protease domain
Cam – calmodulin
ClpP – caseinolytic protease proteolytic subunit
diUb – di-ubiqutin
DUB – deubiquitinase
EGFP – enhanced GFP
GFP – green fluorescent protein
GST – glutathione-S-transferase
MBP – maltose binding protein
nb – nanobody
OP – artificial protein cage
PDZ3 – third PDZ-domain of the synaptic protein PSD-95
Rap80 – receptor-associated protein 80
SENP – SUMO-specific protease
sfGFP – superfolder GFP
SUMO – small ubiquitin-related modifier
TEV protease – cysteine protease from Tobacco Etch Virus
Throm – thrombin

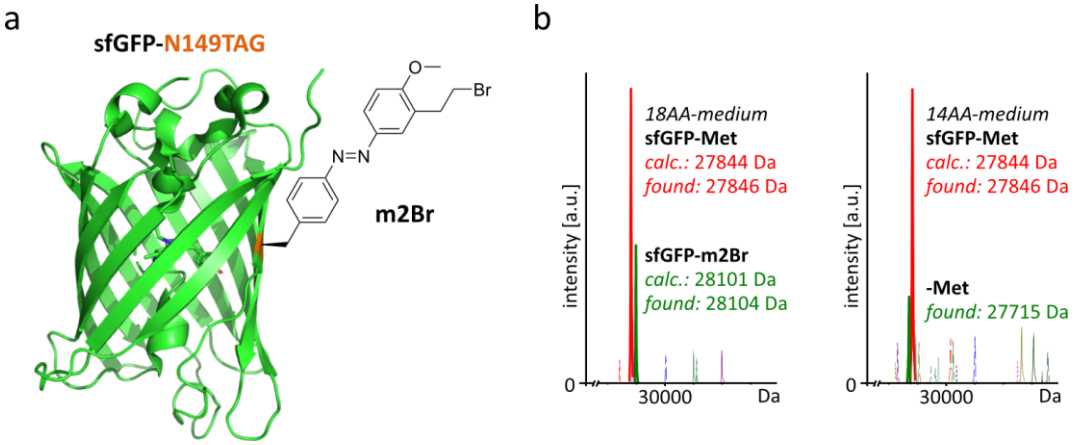
MID1sc10 – engineered metalloesterase

Ub – ubiquitin

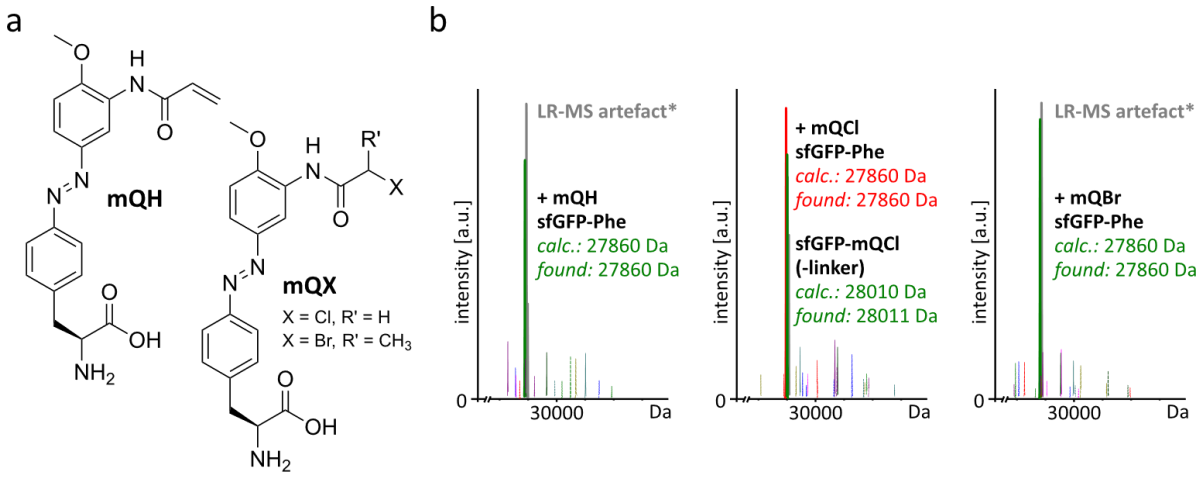
III Supplementary Figures



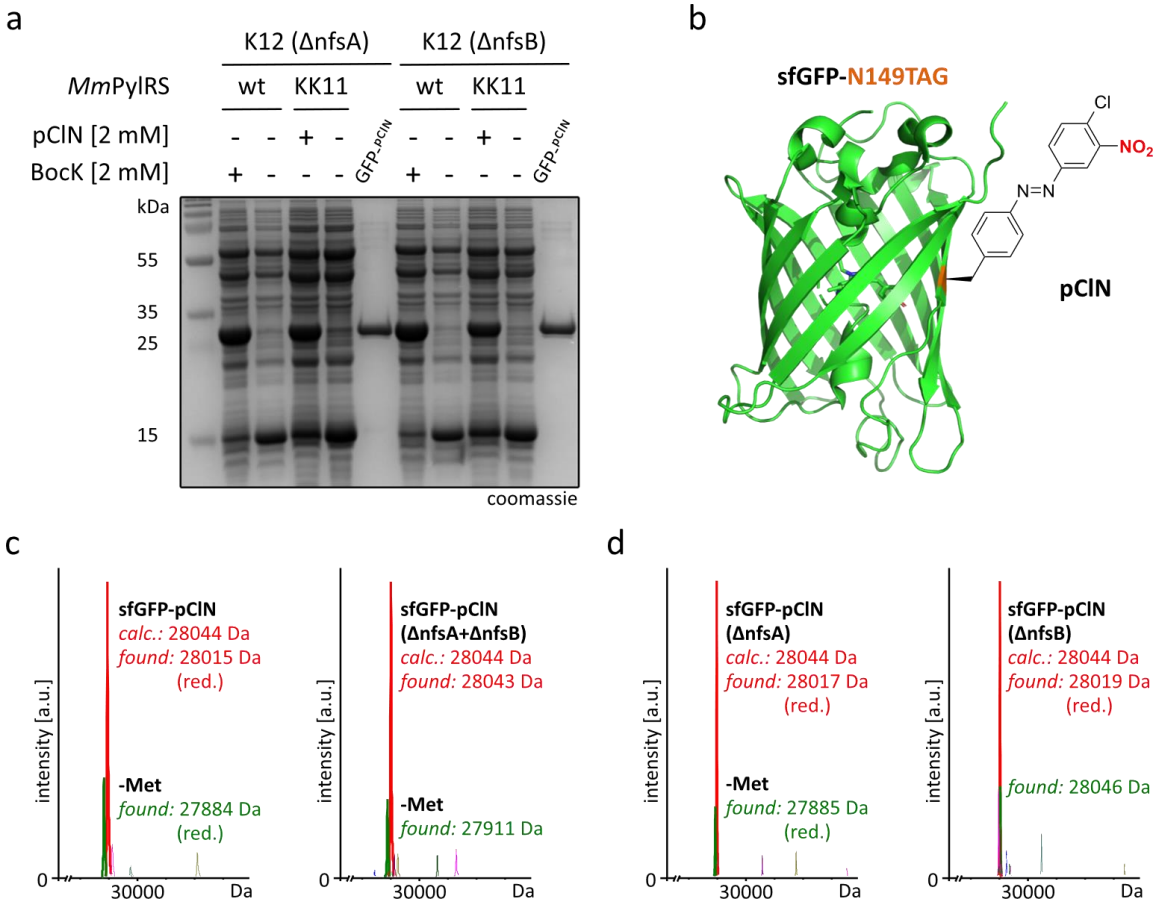
Supplementary Figure III.1 | Amber suppression of sfGFP-N149TAG-His$_6$ with pnX series in *E. coli*. (**a**) Western Blot analysis of azo-ncAA incorporation screening the PylRS variant KK4. BocK, incorporated by the wt PylRS was used as positive control. (**b**) LR-MS analysis of **p3Br**-bearing sfGFP variants amber suppressed with the *Mm*PylRS KK4. Only misincorporation of phenylalanine was observed for sfGFP-N149TAG-His$_6$ amber suppressed with (left panel) but also without any addition of **p3Br** (right panel).



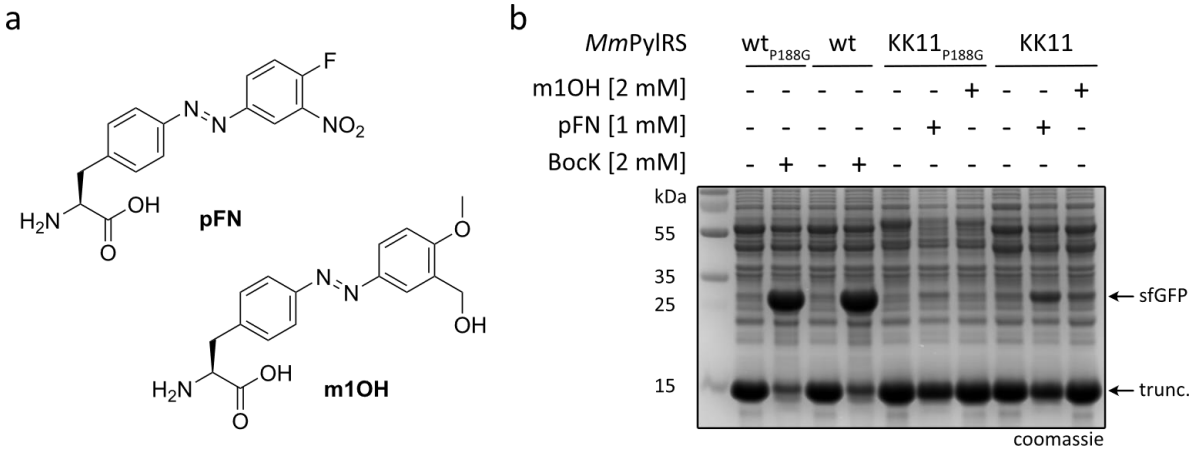
Supplementary Figure III.2 | Amber suppression of sfGFP-N149TAG-His$_6$ with m2Br in *E. coli*. (**a**) Crystal structure of sfGFP (pdb: 2b3p) with the TAG-position (orange) bearing the azo-ncAA m2Br. (**b**) LR-MS analysis of sfGFP variants amber suppressed with **m2Br** using the *Mm*PylRS AzoA. Only misincorporation of presumably methionine was observed for sfGFP-N149TAG-His$_6$ amber suppressed in AI-medium containing 18 amino acids (left panel) but also in AI-medium containing only 14 amino acids (right panel), missing Phe, Cys, Tyr, Lys, Glu and Gln.



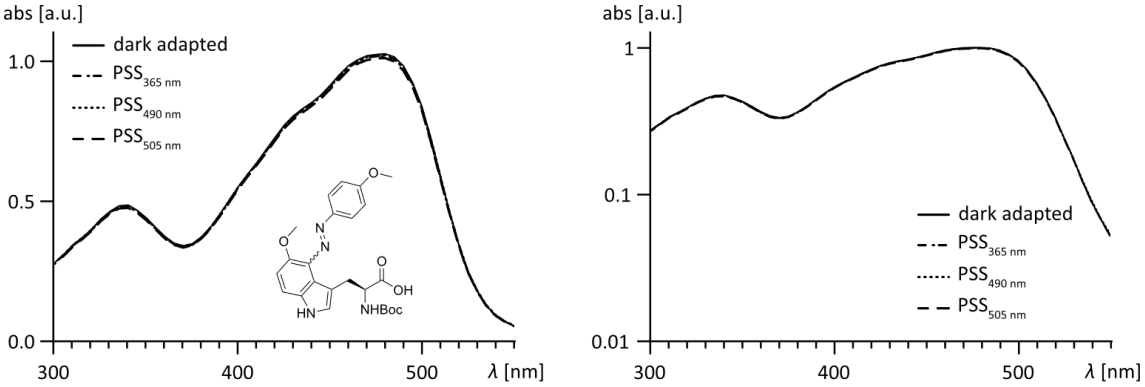
Supplementary Figure III.3 | Amber suppression of sfGFP-N149TAG-His$_6$ with mQH, mQCl and mQBr in *E. coli*. (**a**) Structural formulas of the azo-ncAAs **mQH**, **mQCl** and **mQBr**. (**b**) LR-MS analysis of sfGFP variants amber suppressed with the respective azo-ncAA using the *Mm*PylRS KK11. Misincorporation of phenylalanine was predominantly observed for sfGFP-N149TAG-His$_6$ amber suppressed with the mQR and mQX series. LR-MS analysis of **mQCl** further revealed cleavage of the aromatic amide bond, presumably by unspecific peptidases in *E. coli*.



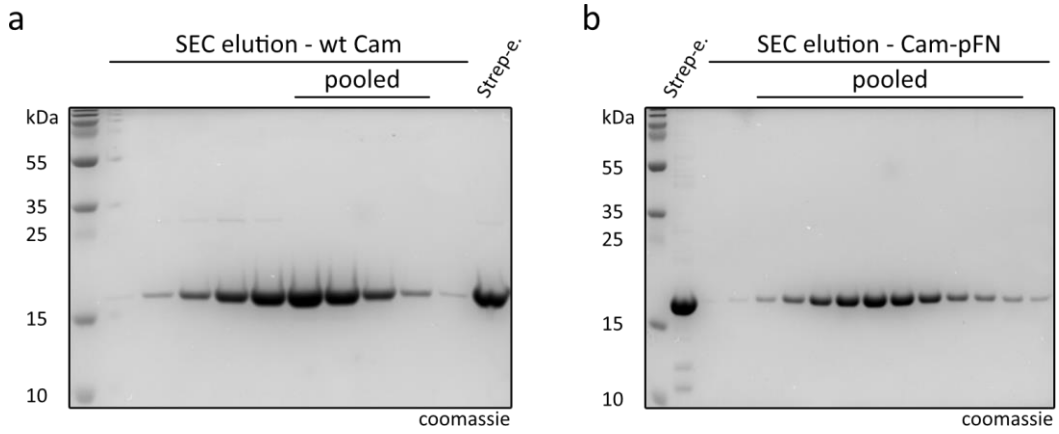
Supplementary Figure III.4 | Amber suppression of sfGFP-N149TAG-His$_6$ with azo-ncAAs pFN and pClN in different *E. coli* nitro-reductase knockout strains. (**a**) Coomassie-stained SDS-PAGE analysis of **pClN** incorporation the *E. coli* strain K12 (ΔnfsA) or K12 (ΔnfsB). The respective sfGFP variants were purified via Ni^{2+}-NTA beads and loaded onto the gel, labeled as GFP^{-pClN}. BocK was used as positive control. (**b**) Crystal structure of sfGFP (pdb: 2b3p) with the TAG-position (orange) bearing the azo-ncAA derivatives **pClN**. The reduction-sensitive nitro-moiety is shown in red. (**c, d**) LR-MS analysis of sfGFP amber suppressed with pClN using the PylRS AzoC (**c**) or KK11 (**d**) in the respective *E. coli* strain K12.



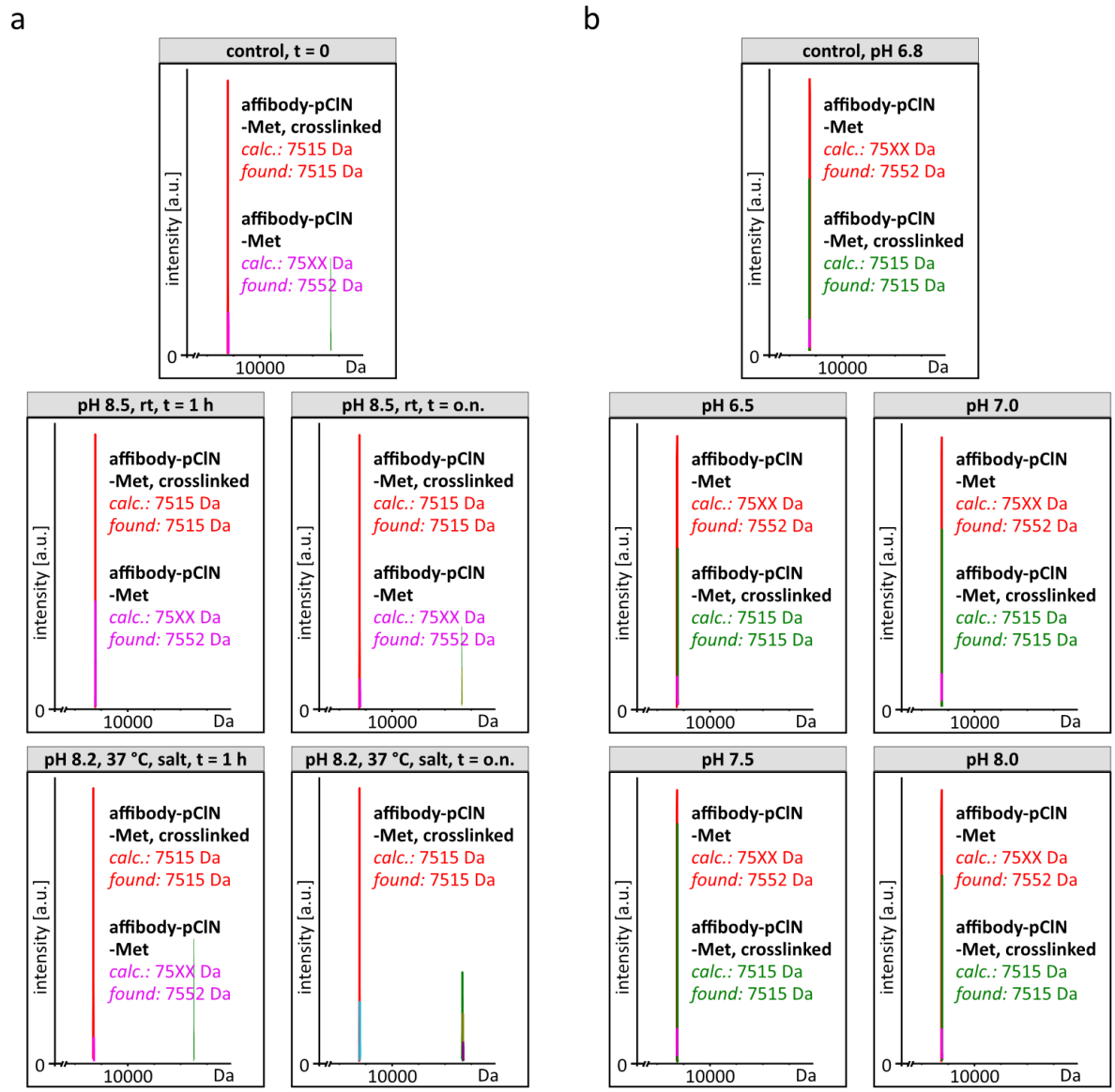
Supplementary Figure III.5 | Site-specific incorporation test for azo-ncAAs pFN and m1OH into sfGFP-N149TAG-His$_6$ in *E. coli*. (**a**) Structural formulas of the azo-ncAAs **pFN** and **m1OH**. (**b**) Coomassie-stained SDS-PAGE analysis of azo-ncAA incorporation into sfGFP-N149TAG using the PylRS variants KK11 and KK11 bearing the additional point mutation P188G. BocK, incorporated by the wild-type PylRS (with and without the point mutation P188G), was included as a positive control.



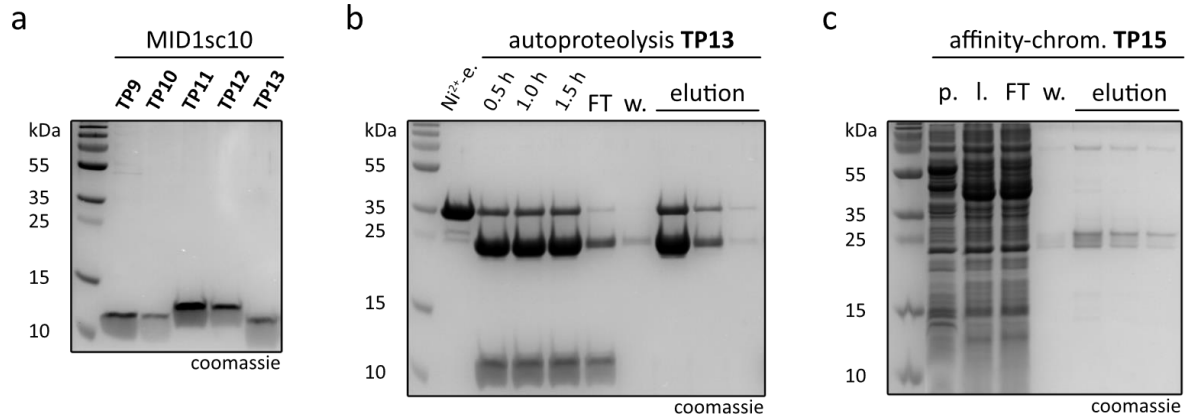
Supplementary Figure III.6 | Absorption spectrum of the model arylazoindole 58 in linear (left panel) and log scale (right panel).



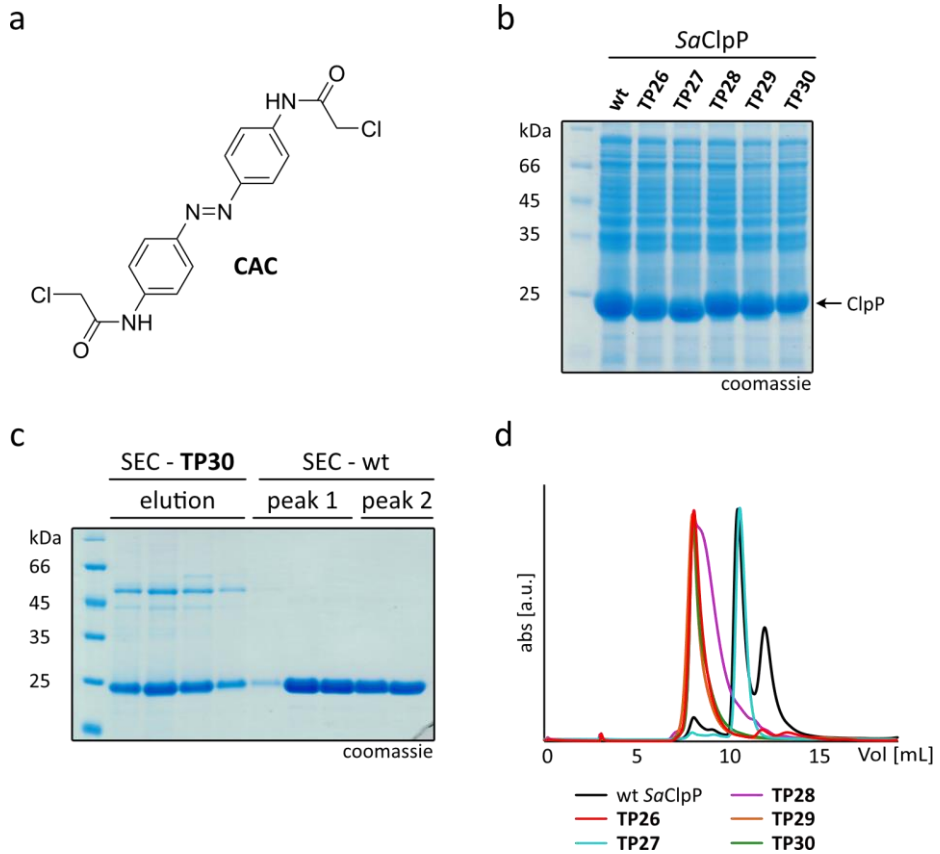
Supplementary Figure III.7 | Purification of wt Cam-His$_6$ and Cam-M76TAG-E83C-His$_6$ amber suppressed with the azo-ncAA pFN. (**a**) Coomassie-stained SDS-PAGE analysis of SEC purification. Strep-e. shows wt Cam protein after affinity chromatography and prior to SEC purification. (**b**) Coomassie-stained SDS-PAGE analysis of SEC purification. Strep-e. shows the amber suppressed Cam protein after affinity chromatography and prior to SEC purification.



Supplementary Figure III.8 | LR-MS analysis of assays to enhance intramolecular crosslinking of pClN with a proximal cysteine residue in affibody-Q26C-S33TAG-His$_6$ (TP1). (**a**) Post-translational incubation of purified affibody-Q26C-S33TAG-His$_6$ amber suppressed with **pClN** at a more basic pH and/or at higher temperatures. Proteins were incubated either at rt or 37 °C while in 20 mM Tris-HCl buffer containing 300 mM NaCl and featuring pH 8.5 at rt. "salt" denotes high-salt concentrations of 600 mM NaCl. (**b**) Amber suppression of affibody-Q26C-S33TAG-His$_6$ (**TP1**) with **pClN** in AI medium featuring different pH (6.5 – 8.0).



Supplementary Figure III.9 | Purification and autoproteolysis of wt MID1sc10-CPD-His$_6$ (TP13) and MID1sc10-E26C-N33TAG-CPD-His$_6$ (TP15) amber suppressed with the azo-ncAA pFN. (**a**) Coomassie-stained SDS-PAGE analysis of all wt MID1sc10 variants, purified via Ni-NTA and SEC. Prior to purification via SEC, **TP9**, **TP10** and **TP12** were processed via TEV-cleavage and **TP13** by initiating autoproteolysis. (**b**) Coomassie-stained SDS-PAGE analysis of the autoproteolysis process of wt MID1sc10-CPD-His$_6$ (**TP13**), initiated by the addition of 1 mM InsP$_6$. Ni^{2+}-e. shows the protein prior to autoproteolysis. 0.5, 1.0 or 1.5 h indicate that samples were taken after 0.5, 1.0 or 1.5 h of incubation with InsP$_6$. FT = flow-through, w. = wash. (**c**) Coomassie-stained SDS-PAGE analysis of **pFN**-bearing MID1sc10-E26C-N33TAG-CPD-His$_6$ (**TP15**) purification via affinity-chromatography. p. = pellet prior to sonication, l. = lysate after sonication, FT = flow-through, w. = wash.



Supplementary Figure III.10 | Expression of *Sa*ClpP-StrepII variants bearing two cysteine mutations and subsequent purification via SEC. (**a**) Structural formula of the for intramolecular crosslinking of ClpP intended azo-crosslinker **CAC**.[296] (**b**) Coomassie-stained SDS-PAGE analysis of the expression levels of the different cysteine *Sa*ClpP-StrepII mutants **TP26** (T134C-N141C), **TP27** (E137C-N141C), **TP28** (N141C-K145C), **TP29** (N141C-R152C) and **TP30** (K145C-R152C). The expression level of wt *Sa*ClpP-StrepII is shown as control. (**c**) Coomassie-stained SDS-PAGE analysis of *Sa*ClpP-StrepII purification via SEC, exemplarily shown for **TP30** (K145C-R152C) and wt as control. (**d**) SEC elution profiles for all tested *Sa*ClpP-StrepII cysteine mutants obtained from purification using a Superdex 200 (10/300) and ClpP storage buffer (20 mM HEPES pH 7.0, 100 mM NaCl, 0.5 mM TCEP).

IV Supplementary Tables

Supplementary Table IV.1 | Overview of screened reaction conditions for late-stage *tetra-ortho*-chlorination of azo-ncAAs.

catalyst	halide source	oxidant	solvent	result	remark
$Pd(OAc)_2$	NCS (5.0 eq)	-	AcOH	✓	mostly *tetra*-halogenation, but N_α-Boc deprotection
$Pd(OAc)_2$	NCS (5.0 eq)	-	DCE	✗	mono-halogenation
$Pd(PPh_3)_4$	TCCA (2.0 eq)	-	DCE	✗	*tetra*- & *penta*-halogenation (1:1)
$Pd(PPh_3)_4$	TCCA (1.5 eq)	$K_2S_2O_8$	DCE	✗	mix of multi-halogenations
$Pd(PPh_3)_4$	TCCA (1.8 eq)	$K_2S_2O_8$	DCE	✓	*tetra*- & *penta*-halogenation (2:1)
$Pd(OAc)_2$	TCCA (2.0 eq)	$K_2S_2O_8$	DCE	✗	mix of multi-halogenations
$Pd(PPh_3)_4$	NCS (5.0 eq)	$K_2S_2O_8$	DCE	✓	only *tetra*-halogenation
$Pd(OAc)_2$	NCS (5.0 eq)	$K_2S_2O_8$	DCE	✗	mono-halogenation & educt

Supplementary Table IV.2 | Mutations of respective *Ma*PylRS variants analog to the *Mm*PylRS mutants AzoA-C. AzoB wrongly contained the mutation M129S instead of the intended mutation M129A.

*Ma*PylRS	A122	M129	I142	N166	V168	H227	Y228
AzoA	T	S		V	G	I	P
AzoB	T	S		A	G	I	P
AzoC	T	A	T	A	G	I	P

Supplementary Table IV.3 | Data and calculations for the kinetic measurements of the *cis* → *trans* relaxation for the azo-ncAAs m1F and pSEtN.

t[min]	A_{kin} (**m1F**)	A_{kin} (**pSEtN**)	Q_{relax} (**m1F**)	Q_{relax} (**pSEtN**)
0	0.01692	0.1132	0	0
10	0.01836	0.1621	0.7639	16.93
20	0.01941	0.1749	1.319	21.34
30	0.02022	0.1823	1.751	23.89
40	0.02104	0.2090	2.184	33.14
50	0.02205	0.2088	2.718	33.08
60	0.02279	0.2136	3.109	34.75
70	0.02441	0.2153	3.965	35.34
80	0.02473	0.2170	4.134	35.90
90	0.02529	0.2184	4.434	36.39
100	0.02588	0.2198	4.742	36.88

110	0.02614	0.2210	4.882	37.31
120	0.02662	0.2220	5.137	37.66

Supplementary Table IV.4 | Results for the maximum reaction rate constant *k* and the minimum half-life $t_{1/2}$ of the *cis* → *trans* relaxation for the azo-ncAAs m1F and pSEtN.

azo-ncAA	k[h^{-1}]	$t_{1/2}$[h]
m1F	0.02637	26.29
pSEtN	0.2362	2.934

Supplementary Table IV.5 | Overview of target proteins and the respective TAG-(Cys) mutants from Chapter 3.3.1 which were tested for amber suppression with azo-ncAAs.

protein	mutant	amber suppression	remark
Cam	M76TAG-E83C	yes	published for azo-ncAAs[250]
affibody	Q26TAG-S33C	no	not even BocK
affibody	Q26C-S33TAG	yes	-
affibody	N43TAG-K50C	no	not even BocK
affibody	N43C-K50TAG	no	not even BocK
affibody	T18TAG-P25C	no	not even BocK
affibody	T18C-P25TAG	yes	-
affibody	D36TAG-N43C	no	not even BocK
affibody	D36C-N43TAG	yes	-
affibody	E8TAG-E15C	no	only BocK
affibody	E8C-E15TAG	yes	only **pClN**
affibody	A42TAG-K49C	no	not even BocK
affibody	A42C-K49TAG	no	not even BocK
affibody	A46TAG-D53C	no	not even BocK
affibody	A46C-D53TAG	no	not even BocK
affibody	E47TAG-A54C	no	not even BocK
affibody	E47C-A54TAG	no	not even BocK
Rap80 (7A)	R97TAG	yes	not tested for pXN series
Rap80 (7A)	R97TAG-E104C	yes	smear with **pFN**
PDZ3	E395TAG-K402C	yes	-
PDZ3	E395C-K402TAG	yes	-
nb(max)	Y37TAG	yes	-
nb(min)	Y116TAG	yes	-

Supplementary Table IV.6 | Overview of target proteins and the respective TAG-(Cys) mutants from Chapter 3.3.2 which were tested for amber suppression with azo-ncAAs.

protein	mutant	amber suppression	remark
AltTPase	E62TAG-Q69C	yes	-
AltTPase	E62C-Q69TAG	no	only BocK
AltTPase	E87TAG-H94C	no	two protein bands were observed with BocK
AltTPase	E87C-H94TAG	no	
AltTPase	D32TAG-D39C	no	only BocK
AltTPase	D32C-D39TAG	no	only BocK
AltTPase	V66TAG-H73C	no	only BocK
AltTPase	V66C-H73TAG	no	not even BocK
MID1sc10	E26TAG-N33C	yes	
MID1sc10	E26C-N33TAG	yes	
MID1sc10	L30TAG-L37C	yes	could not be purified
MID1sc10	L30C-L37TAG	yes	
MID1sc10	L37TAG-Q44C	yes	
MID1sc10	L37C-Q44TAG	yes	
MID1sc10	Y59TAG-Q66C	yes	mostly BocK
MID1sc10	Y59C-Q66TAG	yes	mostly BocK
MID1sc10	Y41TAG-Y90C	yes	could not be purified
ClpP	T134TAG	yes	-
ClpP	E137TAG	yes	but also protein detected in negative control (-ncAA)
ClpP	N141TAG	yes	-
ClpP	K145TAG	yes	-
ClpP	E148TAG	yes	-
ClpP	R152TAG	yes	-
ClpP	E156TAG	yes	-
ClpP	T134C-N141TAG	yes	-
ClpP	E137C-N141TAG	yes	-
ClpP	N141TAG-K145C	yes	-
ClpP	N141TAG-R152C	yes	-
ClpP	K145TAG-R152C	yes	-
ClpP	N141C-R152TAG	yes	-
ClpP	K145C-R152TAG	yes	-

Supplementary Table IV.7 | LR-MS analysis of purified wt MID1sc10 constructs prior and post cleavage.

construct	LR-MS prior cleavage		LR-MS post cleavage	
	calc. [Da]	obs. [Da]	calc. [Da]	obs. [Da]
TP9	55331	55195 (-Met)	10857	10858
TP10	12952	12820 (-Met)	10857	10858
TP11	11812	11680 (-Met)	-	-
TP12	12751	12619 (-Met)	11784	11652 (-Met)
TP13	34386	34252 (-Met)	10958	10826 (-Met)

Supplementary Table IV.8 | Overview of target proteins and the respective TAG-(Cys) mutants from Chapter 3.3.3 which were tested for amber suppression with azo-ncAAs.

protein	mutant	amber suppression	remark
OP	G51TAG-E58C	yes	mostly BocK, however only tested with **TriF** so far
OP	G51C-E58TAG	yes	
OP	Q55TAG-S62C	yes	
OP	Q55C-S62TAG	yes	
OP	D147TAG-V153C	yes	
OP	D146C-V153TAG	yes	
OP	L149TAG-S156C	yes	
OP	L149C-S156TAG	yes	