DISS. ETH No. 27288

Lysine Acylation using Conjugating Enzymes for the Modification of Recombinant Proteins

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

Presented by
Raphael Hofmann

Master of Science in Biology, ETH Zurich
Born on 15.07.1990
Citizen of Rapperswil-Jona and Eschenbach

Accepted on the recommendation of
Prof. Dr. Jeffrey W. Bode, examiner
Prof. Dr. Donald Hilvert, co-examiner

2021

Acknowledgements

I thank my supervisor and mentor, Prof. Dr. Jeffrey W. Bode, for accepting me as a group member and for providing me with the opportunity to work on these exciting projects. I could not have hoped for a better environment to conduct my PhD studies and to learn and grow as a researcher. I truly appreciate the advice, guidance and unbound support I enjoyed at all times.

I am thankful to Prof. Dr. Donald Hilvert for agreeing to be my co-examiner, for taking the time to read this dissertation, and for providing valuable comments. I am grateful for the teaching, research experience and mentoring I received from Prof. Hilvert throughout my ETH studies.

I thank Jakob Farnung, Mikail Levasseur, Gaku Akimoto, Haewon Song, Mamiko Ninomiya, Kateryana Tolmachova and Philipp Schilling for proofreading parts of this thesis. Their feedback and comments made this thesis so much better.

Thank you to Mario Kessinger and Dr. Vijaya Pattabiraman for fantastic help whenever I needed something, and for all the work that goes on behind the scenes.

I am indebted to my closest collaborator, Gaku Akimoto, whose originality, hard work and input propelled this thesis. A special thank you also goes to Dr. Thomas Wucherpfennig who started the SUMO wrestling in the Bode Group and who was vital to the initiation of this work.

I thank my many collaborators, whose creativity and efforts allowed to move these projects further than I could have done it myself; Prof. Dr. Cathleen Zeymer who helped with solving the X-ray structure; Mikail Levasseur who is the expert on the protein cages and who acquired the beautiful TEM images of the modified cages; Mamiko Ninomiya and Cecilie Egholm who majorly contributed to the interleukin experiments; Dr. Dominik Schauenburg who I worked with on the KAT projects; Dr. Bertran Rubi and Martin Köhler who helped with sequencing of the peptide library, and Christian Altorfer who assisted with library synthesis. I also thank the two fantastic semester students, Nikoletta Piperidou and Sara Marie Duke, for lending their help and enthusiasm.

The entire Bode Group, present and former, deserves a big thank you for the superb atmosphere and help along the way. It was a pleasure working with everyone. Besides support in research, the lunches, Coop breaks, after-hour beers, debates, Sola training, parties and hikes were a treat that made the time in Zürich unforgettable. A special thank you goes to Jakob Farnung for his incredible chemistry, biology, literature and Politico knowledge, and for the countless discussions and input; Dr. Dmitry Mazunin, Dr. Chris White and Dr. Kevin Neumann for being great mentors and scientific role models; Yi-Chung Dzeng for being a terrific lab mate and helping me in uncountable ways; Chalupat (Ice) Jindakun for being a wonderful housemate and for showing me the best Thai restaurants in Zürich; and Anne Schuhmacher for always-fun shopping hauls and organizing group trips in the Swiss Alps.

Research at ETH is made possible by fantastic facilities and analytical support. I thank the mass spectrometry teams of MoBiAs and the Functional Genomics Center Zürich, the NMR team, and the HCI shop team for excellent support that made my work so much easier.

I am thankful for financial support of my doctoral studies by the Scholarship Fund of the Swiss Chemical Industry (SSCI).

And finally, I thank my parents, Monika and Karl, and my family for always supporting me outside of the lab and throughout my studies. My deepest gratitude and love go out to Julu for standing by me throughout the years.

Publications and Copyright Permission Notes

Part of this dissertation is reproduced with permission from:

Hofmann, R.; Akimoto, G.; Wucherpfennig, T. G.; Zeymer, C.; Bode, J. W. Lysine acylation using conjugating enzymes for site-specific modification and ubiquitination of recombinant proteins. *Nat. Chem.* **2020**, *12*, 1008–1015.

Table of Contents

Abstract

Proteins and enzymes perform important functions of life. Further expanding and regulating their function, proteins are dynamically decorated with posttranslational modifications (PTMs). Because of the large number of modifications and pathways involved, it is challenging to isolate a protein with a specific modification from natural sources. Besides native PTMs, artificial modifications can augment the properties and efficacy of proteins for their use in pharmaceutical applications and research. To determine the characteristics of natively existing modifications and to create modified protein products, efficient strategies for their preparation are required.

Protein modification in a controlled manner is a difficult task because of the richness of functional groups and chemical environments that the protein structure harbors. While chemical modification of proteins generally offers chemoselectivity for a specific functional group, enzymatic approaches frequently proceed with high site specificity and operate under mild reaction conditions. Such chemoenzymatic methods rely on pathways that operate on proteins, and substrate analogs that can be utilized by these enzymes. Many existing protocols, however, are restricted to modification at the protein termini, rely on non-peptidic metabolites, or require large recognition domains.

This dissertation describes the development and application of lysine acylation using conjugating enzymes (LACE), a chemoenzymatic strategy to site-specifically modify folded proteins at internal lysine residues. LACE relies on a minimal genetically encoded tag (four residues) recognized by the E2 small ubiquitin-like modifier-conjugating enzyme Ubc9 (Ube2I), and peptide or protein thioesters. Together, this approach obviates the need for E1 and E3 enzymes, enabling isopeptide formation with just Ubc9 in a programmable manner. Our studies demonstrate that LACE accepts a series of functionalized thioester probes, ranging from synthetic molecules to entire protein domains. The short tag size allowed the modification of diverse substrates, ranging from monomeric to multimeric proteins, and combination of the method with existing strategies for one-pot dual modifications. Importantly, LACE enabled the site-specific installation of native ubiquitin and ISG15 in a programmable manner from entirely recombinant sources in one step.

In a complementary approach, this dissertation describes the development of a workflow towards the identification of sequence-specific redox reactivity of cysteine residues. To this end, a solid-supported combinatorial peptide library was prepared, and a high-throughput screen combined with semi-automated peptide sequencing was established for the identification and characterization of library hits.

Zusammenfassung

Proteine und Enzyme erfüllen wichtige Funktionen des Lebens. Proteine werden dynamisch mit posttranslationalen Modifikationen (PTMs) dekoriert, was zur Erweiterung und Regulierung ihrer Funktion führt. Aufgrund der grossen Anzahl von Modifikationen und Pfaden ist es schwierig, ein Protein mit einer spezifischen Modifikation aus natürlichen Quellen zu isolieren. Neben nativen PTMs können künstliche Modifikationen die Eigenschaften und Wirksamkeit von Proteinen für ihre Verwendung in pharmazeutischen Anwendungen und der Forschung verbessern. Um nativ vorkommende Modifikationen zu charakterisieren und um modifizierte Proteinprodukte herzustellen, sind effiziente Strategien für deren Herstellung erforderlich.

Kontrollierte Modifikation von Proteinen ist eine schwierige Aufgabe aufgrund der Reichhaltigkeit der Proteinstruktur an funktionellen Gruppen und chemischen Umgebungen. Während chemische Proteinmodifikation im Allgemeinen Selektivität für eine bestimmte funktionelle Gruppe bietet, laufen enzymatische Ansätze häufig mit hoher Ortsspezifität und unter milden Reaktionsbedingungen ab. Solche chemoenzymatischen Verfahren beruhen auf Reaktionspfaden die auf Proteine wirken, und Substratanaloga, welche von diesen Enzymen verwendet werden können. Viele existierende Protokolle sind jedoch auf die Modifikation von Proteinenden beschränkt, greifen auf nicht-peptidische Metaboliten zurück, oder erfordern grosse Erkennungsdomänen.

Diese Dissertation beschreibt die Entwicklung und Anwendung der Lysin-Acylierung mittels konjugierender Enzyme (LACE), einer chemoenzymatischen Strategie zur ortsspezifischen Modifizierung gefalteter Proteine an internen Lysinresten. LACE beruht auf einem minimalen genetisch codierten Tag (vier Reste) der vom E2 SUMO-konjugierenden Enzym Ubc9 (Ube2I) erkannt wird, und Peptid- oder Proteinthioestern. Zusammengenommen macht dieser Ansatz E1- und E3-Enzyme überflüssig, und ermöglicht programmierbare Isopeptidbildung einfach mit Ubc9. Unsere Studien zeigen, dass LACE eine breite Palette funktionalisierter Thioestersonden akzeptiert, die synthetische Moleküle bis hin zu ganzen Proteindomänen beinhalten können. Der kurze Tag ermöglichte es, verschiedene monomere und multimere Proteinsubstrate zu modifizieren, sowie die Methode mit bestehenden Strategien zu kombinieren um Eintopf-Doppelmodifikationen durchzuführen. Insbesondere ermöglichte LACE die ortsspezifische Installation von nativem Ubiquitin und ISG15 auf programmierbare Weise aus vollständig rekombinanten Quellen in einem Schritt.

In einem komplementären Ansatz beschreibt diese Dissertation die Entwicklung eines Ablaufs zur möglichen Identifizierung von sequenzspezifischer Redoxreaktivität von Cysteinresten. Zu diesem Zweck wurde eine Peptidbibliothek mittels kombinatorischer Festphasensynthese hergestellt. Zur Identifizierung und Charakterisierung von Bibliothekstreffern wurde ein Hochdurchsatz-Screen kombiniert mit halb-automatisierter Peptidsequenzierung etabliert.

Symbols and Abbreviations

$[\alpha]_D$	specific optical rotation at wavelength of sodium D line
βME	2-mercaptoethanol
A, Ala	L-alanine
Ac	acetyl
ADC	antibody-drug conjugate
AET	2-aminoethanethiol
Alloc	allyloxycarbonyl
AMP	adenosine monophosphate
aq.	aqueous
Ar	aryl
ATP	adenosine triphosphate
BirA	*Escherichia coli* biotin ligase
Bis-Tris	bis-(2-hydroxyethyl)-amino-tris(hydroxymethyl)methane
Boc	*tert*-butoxycarbonyl
C, Cys	L-cysteine
calc.	calculated
Cam	carbamidomethyl
CHCA	α-cyano-4-hydroxycinnamic acid
CoA	coenzyme A
D, Asp	L-aspartic acid
Dap	2,3-diaminopropionic acid
DBCO	dibenzocyclooctyne
DBHDA	2,5-dibromohexanediamide
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
Dha	dehydroalanine
DIC	*N,N′*-diisopropylcarbodiimide
DIPEA	*N,N*-diisopropylethylamine
DMAP	4-dimethylaminopyiridine
DMF	*N,N*-dimethylformamide
DMSO	dimethyl sulfoxide
DTT	1,4-dithiothreitol
DUB	deubiquitinase
E, Glu	L-glutamic acid
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid

ESI	electrospray ionization
Et	ethyl
et al.	*et alia*
equiv	equivalent
EWG	electron withdrawing group
F, Phe	L-phenylalanine
Fab	antigen-binding fragment
FBS	fetal bovine serum
Fmoc	9-fluorenylmethoxycarbonyl
G, Gly	glycine
GFP	green fluorescent protein
Gnd	guanidine
h	hour(s)
H, His	L-histidine
HATU	1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxide hexafluorophosphate
HC	heavy chain
HCTU	2-(6-Chloro-1-H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HECT	homology to E6AP C-terminus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMBA	4-hydroxymethylbenzoic acid
HRP	horseradish peroxidase
I, Ile	L-isoleucine
IL	interleukin
IPTG	isopropyl-β-D-1-thiogalactopyranoside
IR	infrared
ISG15	interferon-stimulated gene 15
J	coupling constant
K, Lys	L-lysine
KAT	potassium acyltrifluoroborate
L, Leu	L-leucine
LACE	lysine acylation using conjugating enzymes
LB	lysogeny broth
LC	light chain
LC	liquid chromatography
M, Met	L-methionine

MALDI	matrix-assisted laser desorption/ionization
MBP	maltose-binding protein
Me	methyl
Mes	2-mercaptoethanesulfonate
MES	2-(*N*-morpholino)ethanesulfonic acid
MMP	matrix metalloprotease
min	minute(s)
MOPS	3-(*N*-morpholino)propanesulfonic acid
MPAA	4-mercaptophenylacetic acid
MS	mass spectrometry
MW	molecular weight
MWCO	molecular weight cutoff
Mxe GyrA	*Mycobacterium xenopi* DNA gyrase subunit A
N, Asn	L-asparagine
NCL	native chemical ligation
NHS	*N*-hydroxysuccinimide
Ni-NTA	nickel nitrilotriacetic acid
NMM	4-methylmorpholine
NMR	nuclear magnetic resonance
obs.	observed
P, Pro	L-proline
Pbf	2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl
PBS	phosphate-buffered saline
PEG	polyethylene glycol
PMSF	phenylmethylsulfonyl fluoride
PTM	posttranslational modification
Q, Gln	L-glutamine
R	general organic substituent, *Restgruppe*
R, Arg	L-arginine
r.m.s.	root-mean-square
RanGAP1	Ran GTPase-activating protein 1
RBR	RING-between-RING
RING	really interesting new gene
RNase A	bovine pancreatic ribonuclease
RP–HPLC	reversed-phase high-performance liquid chromatography
rt	room temperature
S, Ser	L-serine

Sae	SUMO activating enzyme
SDS–PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SENP	sentrin-specific protease
SIM	SUMO-interacting motif
SPPS	solid-phase peptide synthesis
SrtA	sortase A
SUMO	small ubiquitin-like modifier protein
T, Thr	L-threonine
$t_{1/2}$	half-life
T4L	T4 lysozyme
tBu	*tert*-butyl
TCEP	tris-(2-carboxyethyl)phosphine
TEM	transmission electron microscopy
TEV	tobacco etch virus
Tf	triflyl
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIM	trifluoroborate iminium
TIPS	triisopropyl silane
TLC	thin layer chromatography
TNFα	tumor necrosis factor α
TOF	time-of-flight
Tris	tris(hydroxymethyl)aminomethane
Trt	trityl
U, Sec	L-selenocysteine
UAA	unnatural amino acid
Ub	ubiquitin
Ubl	ubiquitin-like protein
Ufm1	ubiquitin-fold modifier 1
UV	ultraviolet
V, Val	L-valine
wt	wild type
W, Trp	L-tryptophan
X, Xaa	any α-amino acid
Y, Tyr	L-tyrosine

Meinen Eltern gewidmet für all die Unterstützung.

CHAPTER 1

Introduction to Posttranslational Modification of Proteins

1. Background

1.1. Protein structure and function

Proteins are a major class of biomolecules. Proteins are polyamides, also called polypeptides, composed of the 21 canonical α-amino acids (Figure 1). With the exception of glycine, proteinogenic amino acids carry side chains with diverse functional groups, and are almost exclusively present in natural proteins in the L-form (corresponding to (*S*)-α-amino acids in the absolute configuration, with the exception of (*R*)-cysteine). Most proteins in nature are synthesized according to the central dogma of molecular biology,[1] comprising transcription of the underlying genetic code from the DNA level to RNA, followed by ribosomal protein translation.[2] The length of the polypeptide chain constituting a protein can vary from a few amino acids, in which case it is more commonly referred to as oligopeptide or peptide, up to thousands of amino acids.

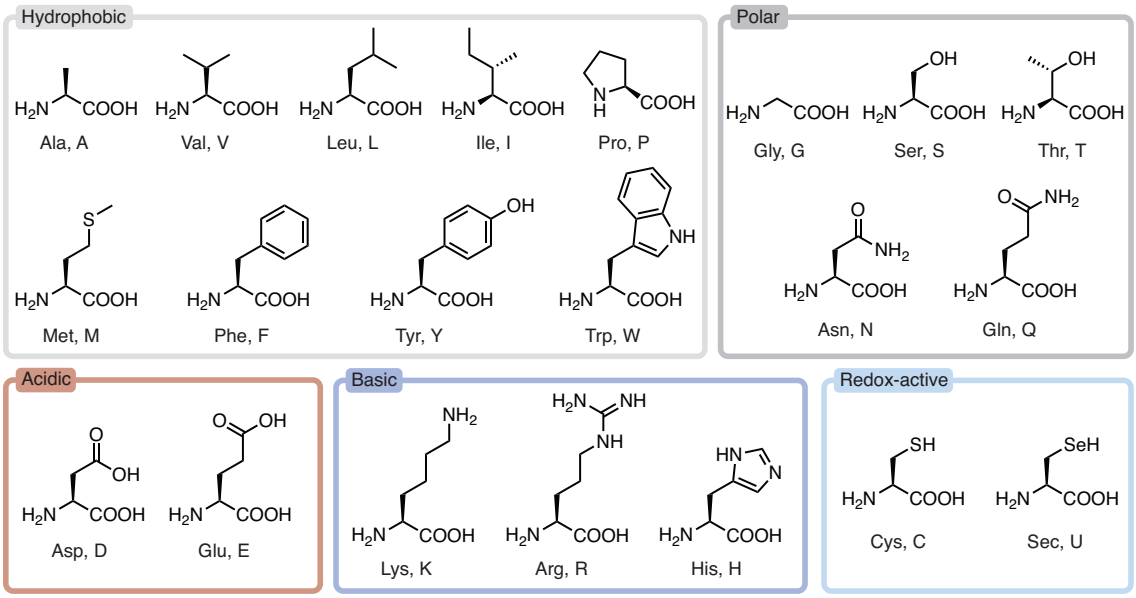


Figure 1. Chemical structures of the 21 proteinogenic amino acids in eukaryotes with the respective three- and one-letter abbreviations. The amino acids are grouped based on the chemical property of their side chain.

Protein function is dictated by the three-dimensional structure that the linear polypeptides adopt (Figure 2).[3] Following ribosomal translation, the linear primary structure hierarchically folds into α-helix and β-sheet secondary structures, which undergo further interaction with each other to arrive at the folded tertiary structure of the protein. Non-covalent or covalent association of folded proteins affords higher-order molecular machines and assemblies, collectively called quaternary structure.

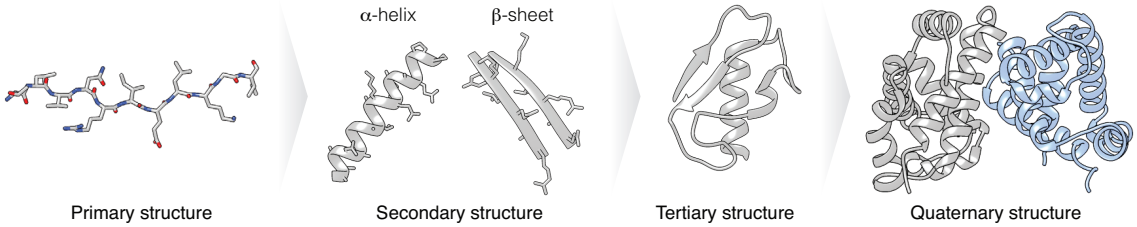


Figure 2. Hierarchical levels of protein structure. (Protein images were generated from PDB entries 1a00, 4lqt and 5f6e.)

Protein folding is dictated by a complicated energy landscape – a result of hydration and dehydration, and supramolecular interactions such as hydrophobic, polar or ionic contacts between the different protein regions over the course of the folding pathway (Figure 3).[4] This energy landscape has been coined 'folding funnel', in which, according to Anfinsen's dogma,[5] an unfolded protein samples many high-energy states and gradually adopts secondary and tertiary structures as it progresses towards a native, folded structure representing a thermodynamically stable state. The 'molten globule' state is a set of folding intermediates that are characterized by adoption of secondary structures, with overall greater flexibility compared to the native fold.[6] The folding pathway may be influenced and guided in some cases by chaperons or by co-translational protein folding.[7]

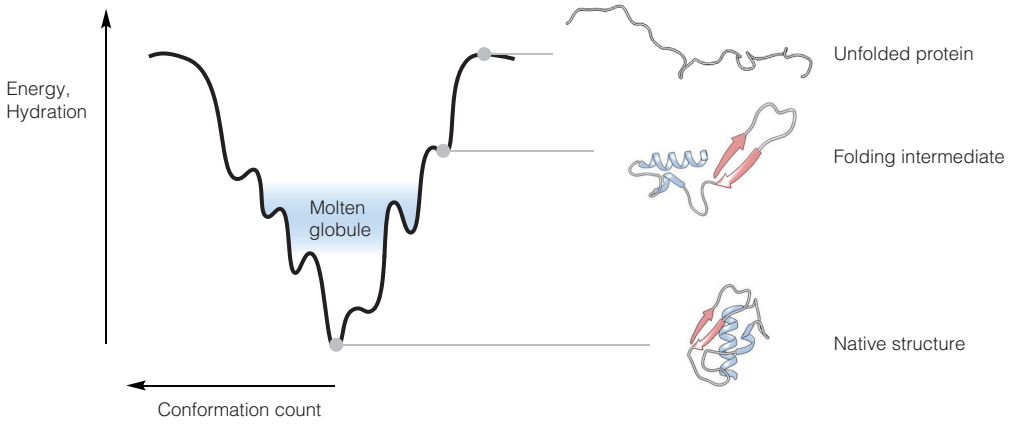


Figure 3. 'Folding funnel' used to describe the energy landscape for the folding pathways of a protein. (Protein images were generated from PDB entries 1xq8 and 5f6e.)

Proteins perform many of life's functions (Figure 4). These functions, to name a few, include catalysis of reactions of the cellular metabolism as enzymes, signal transduction across membranes as receptors and channels, maintenance of cellular structure as scaffold and structural proteins, signaling between cells as hormones, or exerting force as motor proteins.

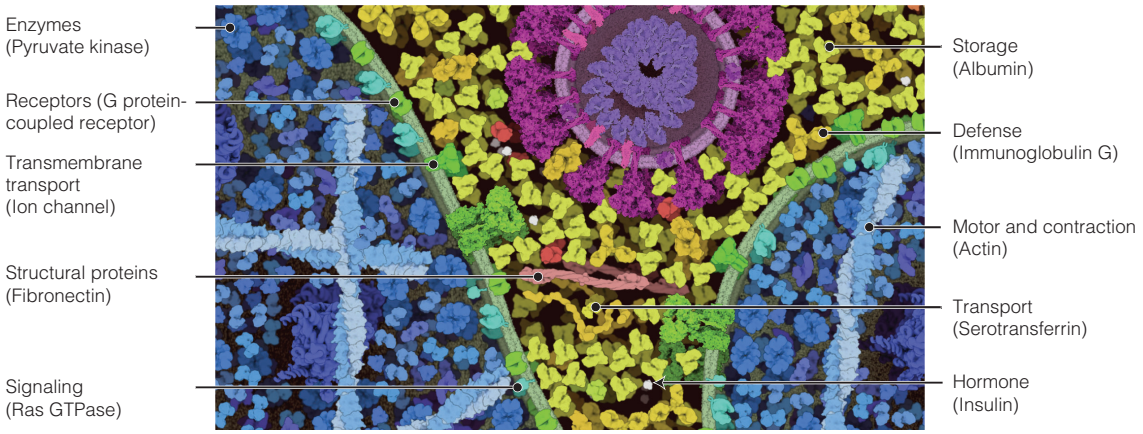


Figure 4. Illustration of cells (blue) surrounded by blood serum (yellow) and a coronavirus (red), with a focus on proteins and their functions. Proteins are drawn to scale. The image was generated using CellPAINT WebGL.[8]

2. Modification of Proteins

2.1. Importance and applications

Because of the plethora of structures and shapes that proteins adopt and the functions they accomplish, there is a great interest in utilizing proteins, for example, for therapy in pharmaceutical applications, biocatalysis in industrial processes, imaging to enable study of biological processes inside cells, microarrays and immobilization for high-throughput screening, and protein-based materials. To augment the function of proteins for a particular application, they are often tailored and modified to generate bioconjugates.

2.2. Requirements

Protein modification reactions should proceed under mild reaction conditions, at or near ambient temperature, and under physiological aqueous conditions to preserve the intricate three-dimensional structure of folded proteins. Additionally, the reactions should exhibit high selectivity for a given protein site to afford homogeneous products with a defined number of modifications. Lastly, protein modification reactions should proceed with a high reaction rate because of the inherently low concentration of the high-molecular weight substrates.

2.2.1. Rate of bioconjugation reactions

Proteins and other biomacromolecules are often only soluble at low micromolar concentrations. In order to achieve complete or near-complete reaction conversion within useful reaction times, a high second-order rate constant is required (Figure 5a). For instance, a bimolecular reaction between reactants at 50 μM concentration has to proceed with a rate upwards of 1–10 M^{-1} s^{-1} to reach full conversion within 24 hours.

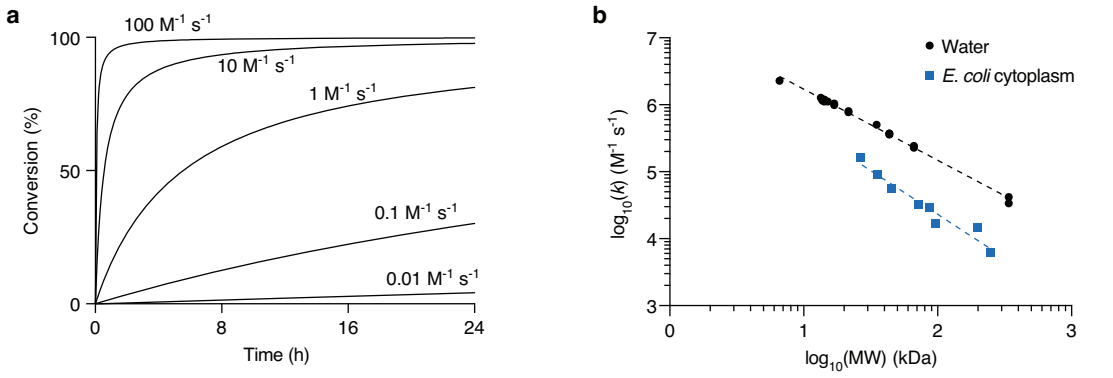


Figure 5. Rate of bioconjugation reactions. (**a**) Simulation of reaction conversion over time for bimolecular reactions between reactants at 50 μM concentration with indicated bimolecular rate constants. Adapted from Dirksen *et al.*[9] (**b**) Diffusion-controlled bimolecular rate constant for protein–protein conjugation, estimated for various proteins based on experimentally determined translational diffusion coeffcients[10,11] in water or *E. coli* cytoplasm at 20 °C. For the collision model, proteins were assumed as spheres with a minimal radius sufficient to contain the protein, and 5 Å^2 reactive surface patches per reactant (kept constant for all protein sizes) that result in productive conjugation upon collision with each other. Linear trends are shown as dashed lines, with y = 1.061x + 7.295 (R^2 = 0.996) and y = 1.305 + 6.975 (R^2 = 0.946) for water and *E. coli* cytoplasm, respectively. Underlying data, calculations and references are shown in Table 3 and Table 4 of the Appendix.

The highest achievable reaction rate is limited by diffusion. Especially for conjugation of two macromolecules, for which diffusion becomes measurably slower as the molecules become bigger, this can have an impact on the upper possible limit of the reaction rate. Using experimentally determined translational diffusion coefficients[10,11] and simplified diffusion and collision models,[12,13,14] the upper limit of the bimolecular rate constant for diffusion-controlled conjugation between two equally sized proteins can be estimated (Figure 5b; see also Table 3 and Table 4 of the Appendix for details). Protein–protein conjugation of two proteins with a molecular weight of 10–100 kDa can reach 10^5–10^6 M^{-1} s^{-1} in water according to this model. In the cytoplasm of *Escherichia coli*, the reaction rate may fall by up to an order of magnitude due to slower diffusion. The rate constant of protein–protein ligations between very large proteins may therefore never exceed 10^5 M^{-1} s^{-1}. This estimation agrees with experimentally determined rates for some of the fastest bimolecular processes observed with biomacromolecules that simply rely on collision.[15,16] While protein modification reactions do not have to achieve such high rates to be useful, it is nevertheless worth keeping in mind the potential upper limit for protein conjugation rates.

2.3. Chemical methods

2.3.1. Natural residues

Several residues can be targeted selectively by chemical strategies for protein modification.[17,18] Among natural residues, cysteines are relatively rare and exhibit unique reactivity which can be harnessed for chemoselective protein modification (Figure 6).[19] The thiol group of cysteines is generally the most nucleophilic functional group in proteins, which can be harnessed for polar reactions with electrophiles such as maleimide,[20] α-halocarbonyl compounds, methylsulfone oxadiazoles,[21] or α,β-unsaturated carbonyl groups.[22] Under photoirradiation, radical-promoted thiol-ene[23] and thiol-yne[24] reactions between cysteines and alkenes or alkynes, respectively, can be performed. Taking advantage of the redox reactivity of cysteines, protein modification via formation of mixed disulfides has been extensively employed,[25] which offers the added benefit that the modification can be selectively removed using reducing agents.[26] Organometallic reagents[27] and improved variants thereof with increased water-solubility have also been described.[28] Lastly, the residue dehydroalanine (Dha), which can be generated from cysteine,[29] serves either as an electrophile or undergoes carbon-carbon bond formation with alkyl radicals.[30,31]

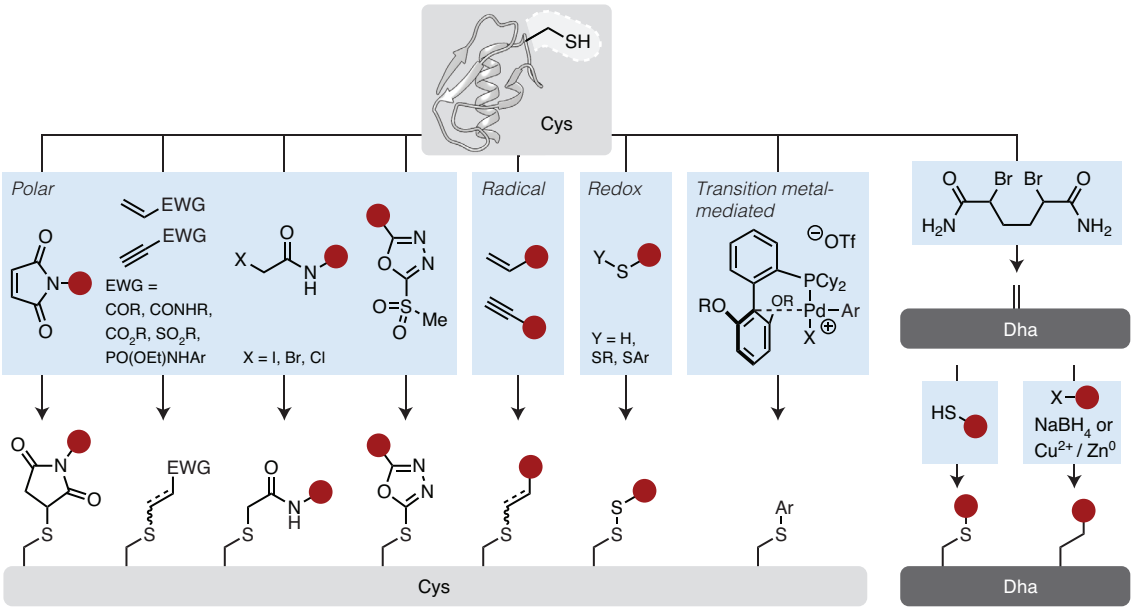


Figure 6. Chemical methods for cysteine modification.

A popular approach especially for the generation of antibody-drug conjugates (ADCs) is the reduction and rebridging of disulfide bonds (Figure 7).[32–34] This strategy has the advantage that the structural role of the disulfide bond is preserved and does not require engineering of proteins to introduce surface-exposed cysteines.[35] In immunoglobulin G, the intermolecular

disulfide bond in particular can be selectively reduced and modified to obtain homogeneous ADCs.[36]

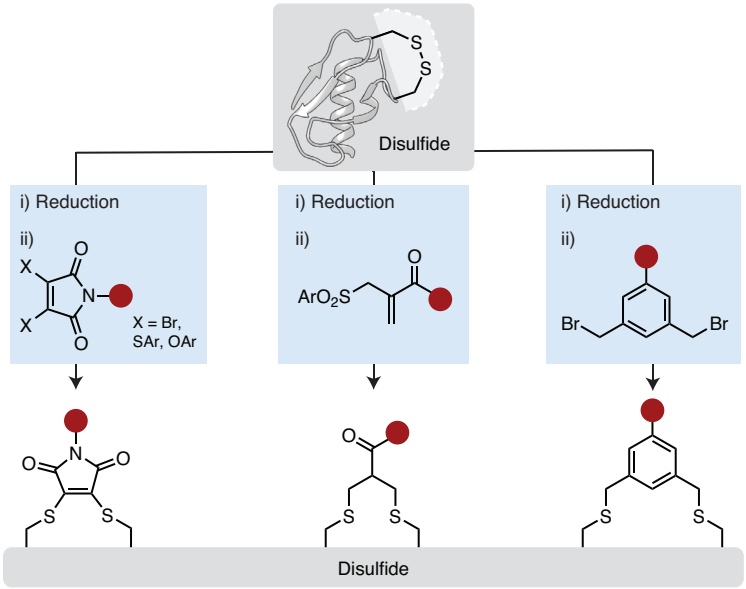


Figure 7. Chemical methods for disulfide rebridging.

Lysines are the second-most nucleophilic residue after cysteine and therefore a popular choice for protein modification. It can be targeted for example by acylation or by reaction with iso(thio)cyanates (Figure 8). However, often multiple surface-exposed lysine residues are present in any given protein which generally leads to heterogeneous labeling of proteins. Another common strategy is reductive amination. A benefit of reductive amination compared to acylation is that the secondary amine product preserves the net positive charge, which has been shown to maintain therapeutic activity in the case of labeled insulin.[37] Instead of the ε-amine of lysines, the α-amine of the protein N-terminus offers a unique place of modification and often serves as a distinct site of modification in a protein to generate homogeneous conjugates. For example, reductive amination can occur selectively with the N-terminal α-amine.[38] Lysines have also been shown to be labeled with very reactive Michael acceptors such as sulfonyl acrylates.[39] Various condensation strategies have been reported, for example orthophtaladehyde-amine condensation[40] and iminoboronate formation.[41] More exotic methods include condensation followed by 6π-aza-cyclization and autooxidation.[42]

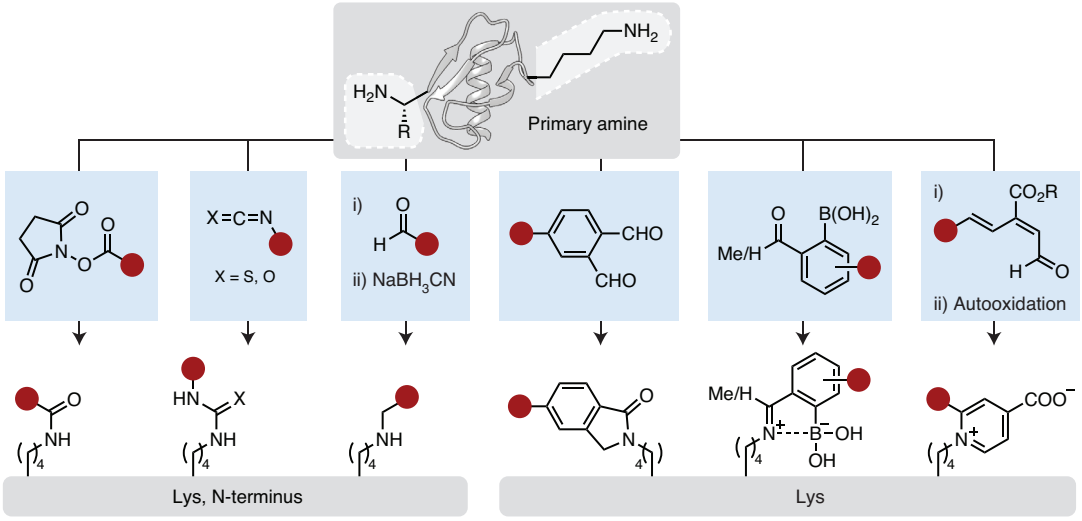


Figure 8. Chemical methods for modification of primary amines.

The phenol functional group in tyrosine offers unique reactivity compared to other side chains (Figure 9). Tyrosine labeling was achieved already at the beginning of the 20th century using diazonium salts, and newer procedures have been described more recently.[43,44] Other possibilities for tyrosine functionalization are an ene-like reaction with triazoline-diones[45,46] and Mannich-type reactions.[47,48] Oxidative tyrosine aniline coupling[49] and tyrosine allylation also belong to the toolbox for tyrosine modification.[50,51] A frequent problem of tyrosine labeling strategies is cross-reactivity with tryptophan residues.[52]

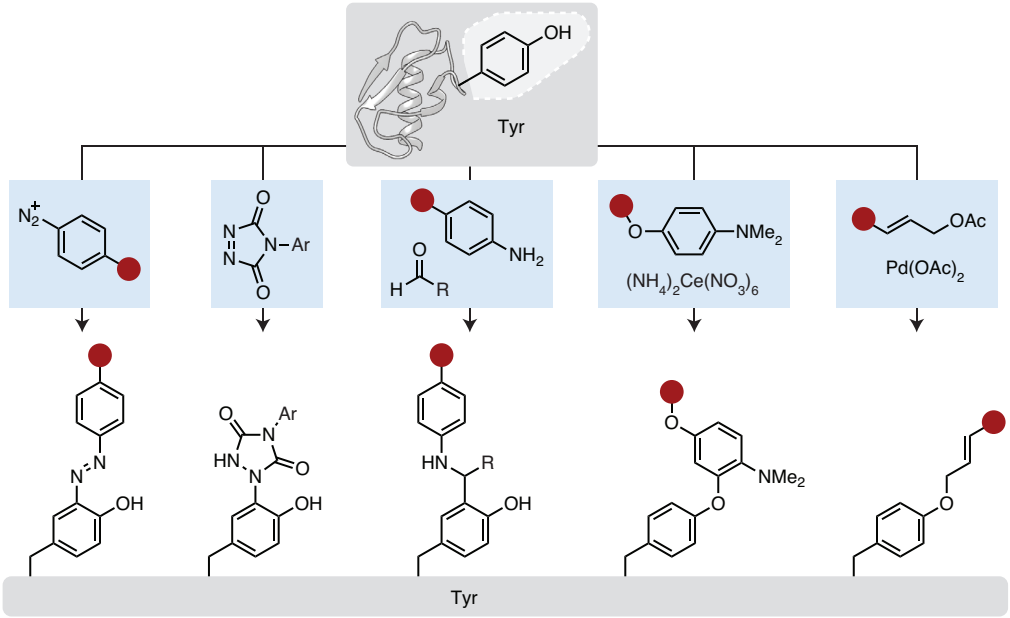


Figure 9. Chemical methods for tyrosine modification.

Several efficient methods have been developed that take advantage of the unique reactivity of the N-terminus (Figure 10).[53] These methods include imidazolidinone formation with various

N-terminal residues in the presence 2-pyridinecarboxyaldehydes[54] or oxidative couplings of *o*-aminophenols[55] and catechols.[56] The latter have been shown to be especially effective with N-terminal prolines – the only secondary amine in a natural protein. General modification of the α-amine has also been achieved by biomimetic transamination, for example by using pyridoxal 5'-phosphate[57] or *N*-methylpyridinium-4-carboxaldehyde (Rapoport's salt).[58] The resulting carbonyl group can subsequently be used for condensation reactions. Likewise, oxidation of N-terminal serine or threonine residues affords aldehyde residues for subsequent oxime or hydrazone ligations.[59]

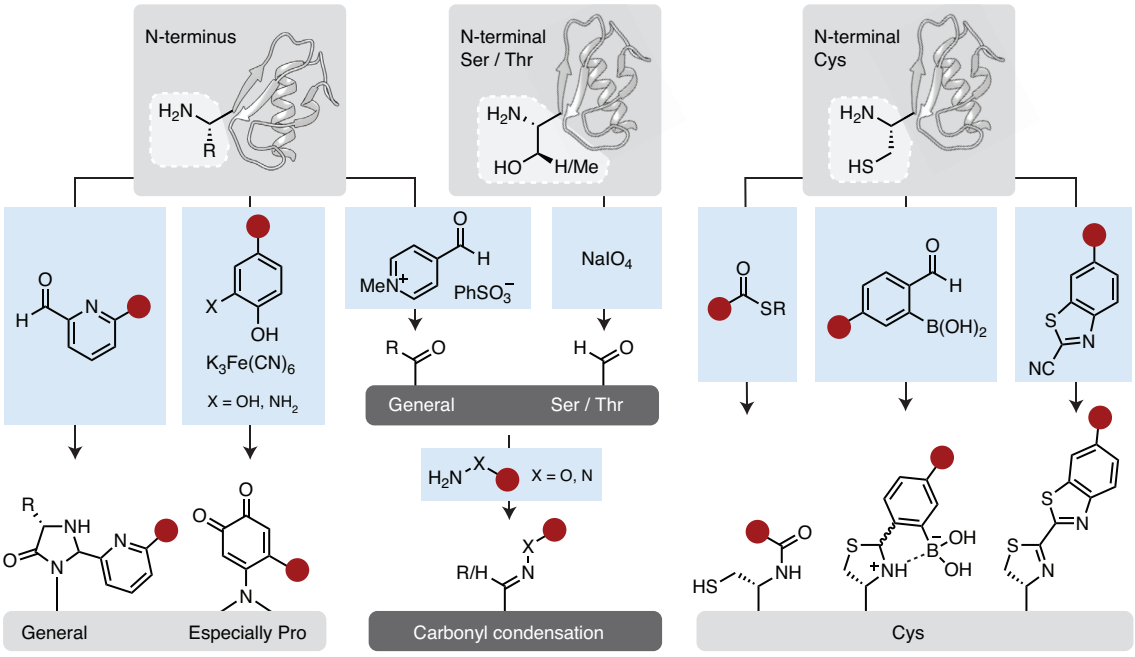


Figure 10. Chemical methods for modification of N-terminal residues.

Unique reactivity is afforded by N-terminal cysteines (Figure 10). N-Terminal cysteines undergo native chemical ligation (NCL) with thioesters.[60] Secondly, thiazolidine formation with aldehydes has been reported,[61] which is especially fast and affords a stable product using benzaldehyde with *o*-boronic acid substituents to give thiazolidine boronates.[62,63] Lastly, 2-cyanobenzothiazole condensation with N-terminal cysteines, an analogous reaction to the last step of D-luciferin synthesis from D-cysteine in fireflies, occurs rapidly and has even been applied to the modification of proteins on cell surfaces.[64] A common limitation of these strategies is the lack of direct recombinant methods to access proteins with free N-terminal cysteines. Instead, protease treatment of fusion proteins is frequently required to reveal the desired N-terminus.[65]

2.3.2. Unnatural residues

To afford orthogonal reactivity over the canonical amino acids, strategies for the incorporation of unnatural amino acids (UAA) have been developed. Common techniques are genetic code expansion techniques for example by amber codon suppression,[66] or the use of an artificial ribozyme called flexizyme to perform tRNA aminoacylation with a broad range of UAAs for *in vitro* translation.[67] However, UAA incorporation requires advanced genetic techniques and can result in reduced expression levels. A common side product in the case of amber codon suppression is the generation of truncated protein. Even with genetic code expansion, a second modification step using a bioorthogonal reaction is often required to introduce a desired moiety. Relying on canonical residues only is therefore preferred.

2.4. Enzymatic methods

Enzymes are powerful tools for protein modification[68] due to the mild reaction conditions under which they can operate, as well as the sequence- and domain-specificity that is afforded by the large reaction interface between the substrate and the enzyme. Compared to chemical methods that commonly exhibit only chemoselectivity for one type of functional group, enzymatic methods can exhibit dramatically increased specificity for modification of single sites in a protein, and for modification of a specific protein in a mixture.

2.4.1. Transpeptidases

Several proteases and transpeptidases have been discovered that shuffle parts of proteins. These enzymes have been developed into broadly used tools. Common to this strategy is the exploitation of an acyl enzyme intermediate, which can be intercepted with various nucleophiles (Figure 11). One example is the transpeptidase sortase A from *Staphylococcus aureus* (Figure 11),[69] which recognizes a C-terminal LPETG motif and forms an acyl intermediate with a protein substrate bearing this tag. In the presence of a suitable nucleophile, typically a protein or peptide bearing a poly-glycine stretch at the N-terminus, the sortase is displaced and the desired conjugate is obtained. For protein modification, the poly-glycine nucleophile can be functionalized with a group of interest. Other examples of transpeptidases that operate via an analogous mechanism include butelase 1 (ref. [70]) and the asparaginyl endopeptidase OaAEP1.[71] In a related strategy, proteases have been engineered which exhibit an increased propensity to catalyze transamidation rather than hydrolysis, for example subtiligase[72] and trypsiligase.[73]

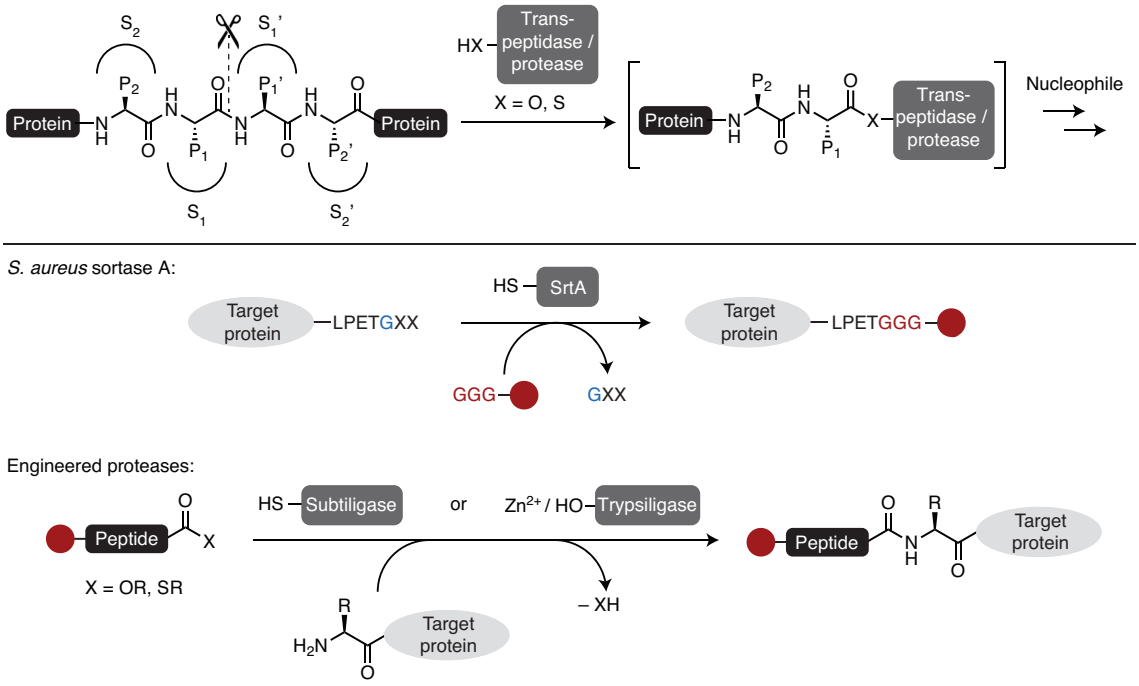


Figure 11. Protein modification using transpeptidases and engineered proteases. Top: Schechter and Berger nomenclature[74] for protease and transpeptidase substrates. The scissile bond is indicated. The acyl enzyme intermediate is resolved by attack of a nucleophile and can be exploited for protein modification. Bottom: Protein modification strategies based on sortase A from *S. aureus*, or the engineered proteases subtiligase and trypsiligase.

A potential drawback of both transpeptidases and engineered proteases is that the products can be a substrate for the enzyme as well, rendering the reaction reversible. To achieve useful conversion, C-terminally activated substrates bearing esters or thioesters have been used to increase the rate of acyl enzyme intermediate formation, and to drive the equilibrium towards the thermodynamically more stable amide product (Figure 11). An alternative strategy relies on Le Chatelier's principle in which the incoming nucleophile is supplied in excess to favor product formation.

2.4.2. Chemoenzymatic modification

Proteins are dynamically decorated with posttranslational modifications (PTMs) in enzymatic processes inside and outside of cells.[75] Frequently found PTMs of proteins include methylation, acetylation, phosphorylation, GlcNAcylation, lipidation or ubiquitination.[76] Some of these enzymatic processes possess intrinsic sequence specificity. Chemoenzymatic strategies for protein modification rely on substrate analogs that are accepted by these enzymes, which are then installed in a protein of interest bearing a specific domain or recognition sequence for this enzyme (Figure 12).[77,78,79]

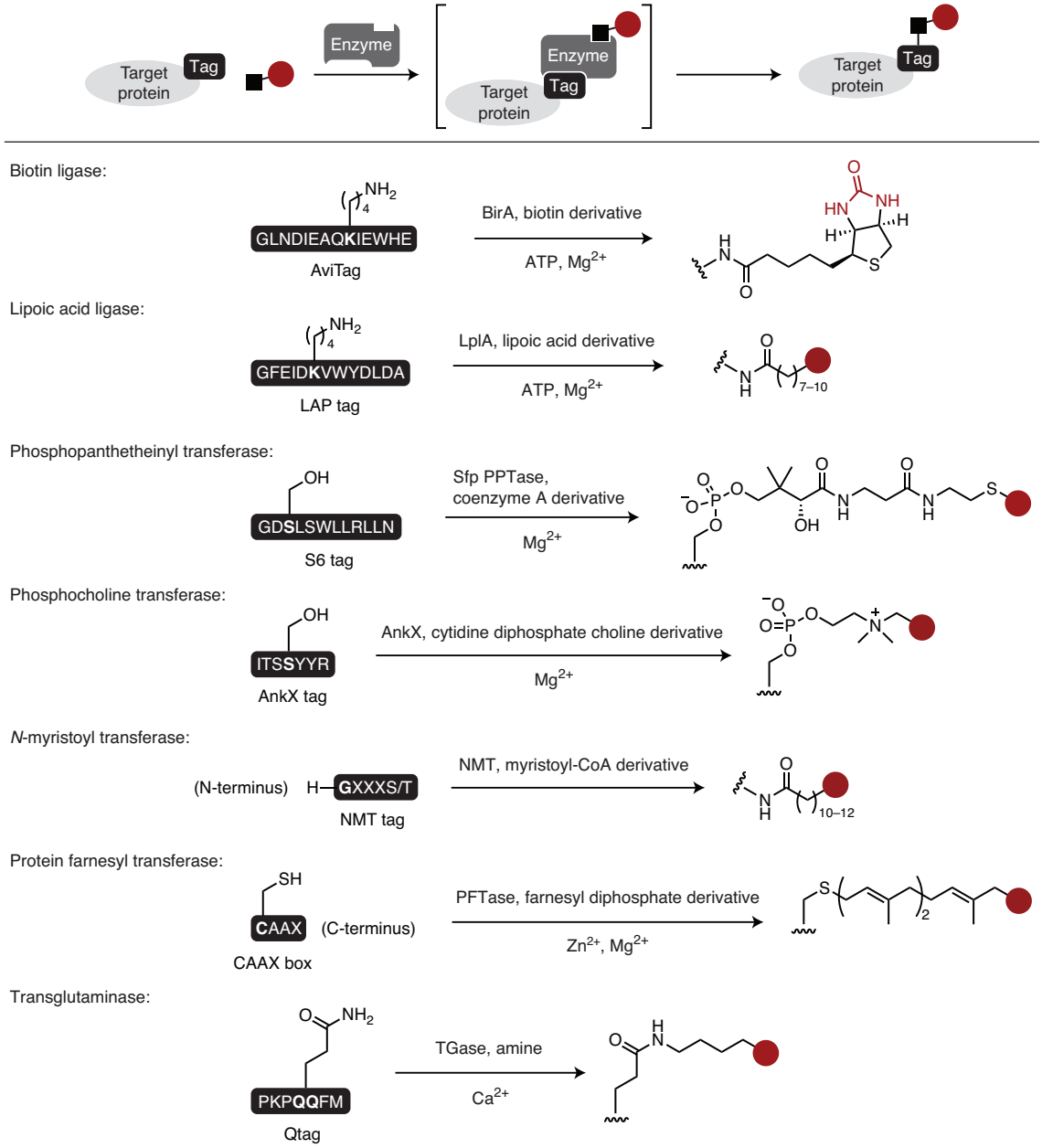


Figure 12. Chemoenzymatic protein modification. Top: General strategy. Bottom: Examples of enzymes with their respective metabolites and recognition tags.

Enzymes that have been applied to transfer metabolite analogs include biotin[80] and lipoic acid ligases,[81] phosphopantetheinyl transferase,[82] phosphocholine transferase,[83,84] *N*-myristoyl transferase,[85,86] farnesyl transferase,[87] and transglutaminase (Figure 12).[88] A bottleneck of these strategies can be probe synthesis, intolerance of metabolite analogs by the enzymes resulting in slower rates of labeling, and remnant metabolite which may impact the biophysical properties of the modified protein.

2.4.3. Self-labeling and autocatalytic domains

A uniquely simple method for protein modification relies on self-labeling and autocatalytic domains. By genetically fusing self-labeling enzyme variants to a target protein that irreversibly react with a covalent ligand or inhibitor, small molecules can be linked to a target protein (Figure 13). Examples of self-labeling protein domains include SNAP-tag,[89] Halo tag (haloalkane dehalogenase)[90] and cutinase.[91] Catalytic antibodies[92] capable of performing aldolase reactions can undergo condensation with 1,3-diketone derivatives[93] or amide bond formation with β-lactam derivatives.[94]

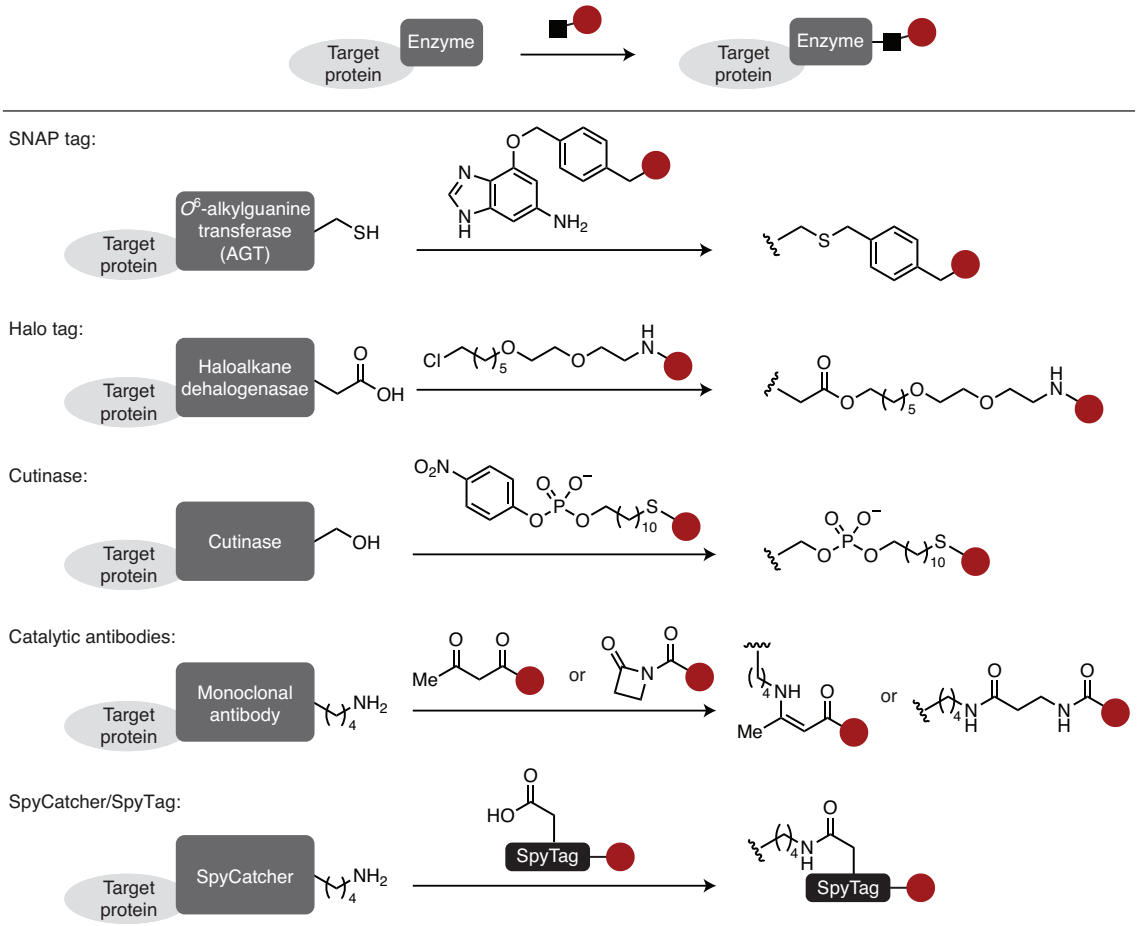


Figure 13. Self-labeling enzymes and protein domains. Top: General strategy. Bottom: Examples of self-labeling enzymes and protein domains with their respective metabolite for covalent modification of an active site residue.

The SpyCatcher/SpyTag system is another self-labeling strategy (Figure 13),[95,96] which was derived from an immunoglobulin-like collagen adhesin domain (CnaB2). This domain, from the fibronectin binding protein of *Streptococcus pyogenes*, spontaneously forms an intramolecular isopeptide bond between a lysine and an aspartic acid residue.[97] Based on this protein domain, a bimolecular system consisting of SpyCatcher and SpyTag has been engineered which undergoes rapid association and formation of the native isopeptide for spontaneous

bioconjugation. The SpyCatcher/SpyTag system exhibits a bimolecular rate constant of 5.5 x 10^5 M^{-1} s^{-1} (ref. [98]) and has therefore nearly reached the diffusion limit (see Section 2.2.1).

SpyCatcher is unique among the self-labeling domains as its reacting partner, SpyTag, is fully peptidic, and protein modification can be achieved from entirely recombinant starting materials. However, common to all strategies that rely on self-labeling domains is the requirement for genetic fusion of a large reacting domain to a target protein. Reliance on a specific acceptor domain results in a sizeable ligation scar in the product.

An elegant way around the large remnant domain in the modified products is offered by inteins (Figure 14).[99,100] These autocatalytic domains are capable of performing protein splicing, which involves cleavage and formation of a peptide bond and results in the excision of the intein domain from the product. Intein fusion proteins undergo an intramolecular *S*-to-*N* acyl shift catalyzed by the intein, and the resulting thioester can be subsequently trapped with β-aminothiols such as N-terminal cysteines to introduce a moiety of choice. Similarly, naturally occurring or engineered split inteins,[101–103] consisting of an N- and C-terminal intein fragment that bind with high affinity to reconstitute the active intein fold, can be used. An advantage of split inteins is that no N-terminal cysteine is required, which cannot be generally accessed by recombinant means.[65] Additionally, some naturally occurring split inteins exhibit markedly increased reaction rates compared to full-length inteins, which has enabled *in vivo* protein trans-splicing.[104]

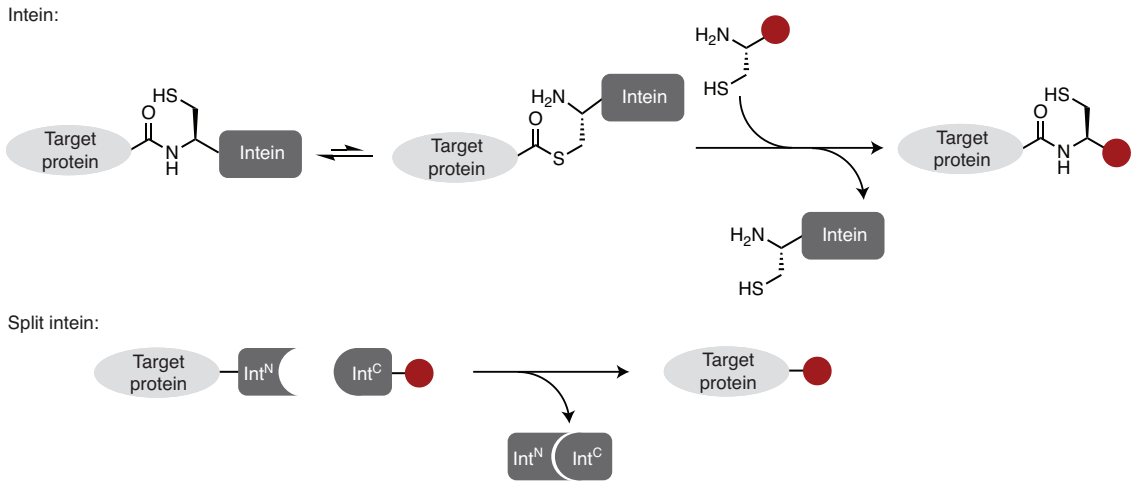


Figure 14. Intein and split intein (Int^N and Int^C) strategies for protein transamidation.

2.5. Branching out: Internal protein modification with peptidic substrates

Protein modification at internal sites has several advantages over labeling of protein termini, such as allowing for multiple labels and affording a flexible site of modification, without having to use circular permutation or internal cleavage by proteases to generate new termini (Figure 15). Very few techniques, however, exist that allow for protein modification at internal sites

in a general manner. This is particularly true for protein labeling with peptidic substrates to obtain branched engineered protein conjugates.

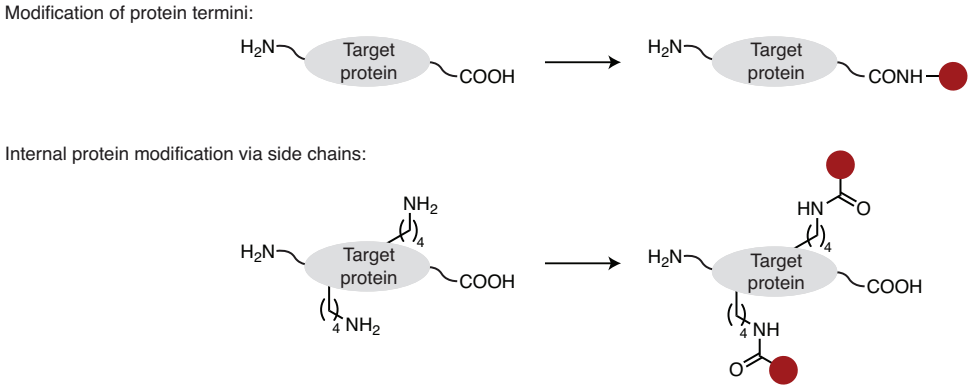


Figure 15. Comparison of protein modification via termini, or internally via side chains.

Chemoenzymatic strategies (see Section 2.4.2) generally operate on amino acid side chains, but a metabolite residue in the resulting bioconjugates may not always be desired. While transpeptidases and engineered proteases do not rely on metabolites and can operate on fully peptidic substrates (see Section 2.4.1), they are generally restricted to modification at protein termini. Similarly, autocatalytic and self-labeling domains such as inteins and SpyTag/SpyCatcher (see Section 2.4.3), do not allow for general internal labeling in this manner. The SpyCatcher/SpyTag system formally forms an isopeptide bond, but the relatively large size of approximately 150 residues of the autocatalytic domain effectively limits application of this domain to C- or N-terminal fusions with a protein of interest. The size of the remaining domain has since been reduced by engineering tripartite systems,[105–107] but these currently proceed with reduced reactivity and conversion. In this sense, the high reaction rate of SpyCatcher/SpyTag has to be traded off with the size of the reactive domain.

One example of an enzymatic process that allows for side chain labeling with peptidic substrates relies on transglutaminase (see Figure 12).[108] Transglutaminase can catalyze the transamidation between glutamine side chains and various primary amine substrates, including the ε-amine of lysines.[109,110] However, modification by transglutaminase has been shown to exhibit substrate-dependent selectivity and reactivity.[111] Additionally, dehydration side products are observed.

Besides transglutaminase, isopeptide activity of sortase has been reported. One such function was derived from sortase-mediated isopeptide conjugation of pilin domains in certain Gram-positive bacteria (Figure 16).[112] By using a sortase from *Corynebacterium diphteriae* together with a truncated version of the pilin-derived SpaA acceptor domain, isopeptide labeling of SpaA was achieved.[113] Similarly, non-canonical activity of *S. aureus* sortase A with a pilin domain has been reported.[114] However, with 20 and 165 residues for pilin and SpaA, respectively,

these specific acceptor domains are relatively big. In analogy to the fusion of SpyCatcher domains to a protein of interest, direct isopeptide labeling of a general substrate, without the need for a discrete acceptor domain, has not been achieved using sortases. Furthermore, isopeptide reactivity of these sortase processes is relatively low.[115]

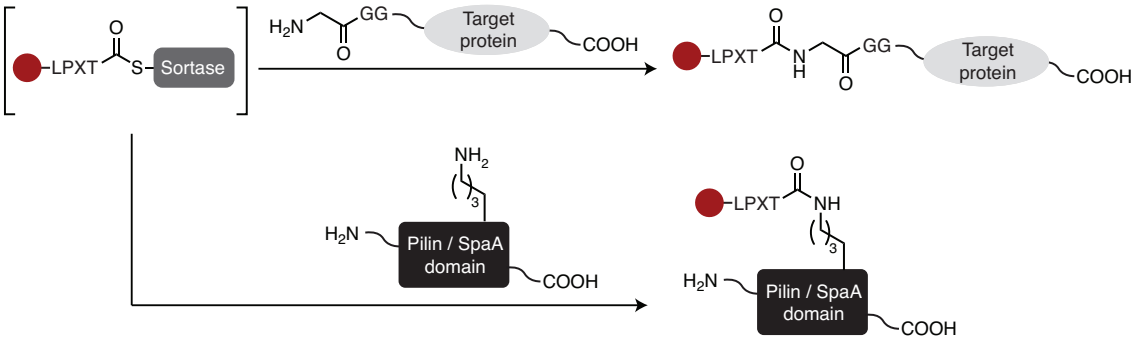


Figure 16. Comparison of N-terminal amidation activity (top) and isopeptide formation with lysine side chains (bottom) by sortases.

3. Conclusions and project goals

Based on these considerations, the ideal protein conjugation method relies on natural amino acids only for facile access to starting materials, requires minimal or no changes to the substrate, and exhibits high site specificity under mild reaction conditions. Furthermore, the site of modification should be flexible with respect to the protein scaffold, which necessitates the possibility to label internal sites and precludes reactions with the protein termini. In terms of moiety to be attached to the protein, a broad probe tolerance and easy synthetic or recombinant access to derivatives is desired. Lastly, the protein conjugation method should exhibit sufficiently high reactivity to allow for conjugation of macromolecules under highly dilute conditions.

In Chapter 2, we describe the development of a chemoenzymatic method for site-specific isopeptide labeling of folded proteins, which we term lysine acylation using conjugating enzymes (LACE).[116] In Chapter 3, applications of LACE are discussed for the modification and conjugation of recombinant proteins. In a complementary approach, Chapter 4 describes the development of a high-throughput workflow towards the identification of sequence-specific redox reactivity of cysteine residues.

4. References

(1) Crick, F. Central dogma of molecular biology. *Nature* **1970**, *227*, 561–563.

(2) Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **2002**, *108*, 557–572.

(3) Petsko, G. A.; Ringe, D. *Protein structure and function*. New Science Press: London, UK, 2004.

(4) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys.* **2008**, *37*, 289–316.

(5) Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230.

(6) Christensen, H.; Pain, R.H. Molten globule intermediates and protein folding. *Eur. Biophys. J.* **1991**, *19*, 221229.

(7) Thommen, M.; Holtkamp, W.; Rodnina, M. V. Co-translational protein folding: Progress and methods. *Curr. Opin. Struct. Biol.* **2017**, *42*, 83–89.

(8) Gardner, A.; Autin, L.; Barbaro, B.; Olson, A. J.; Goodsell, D. S. CellPAINT: interactive illustration of dynamic mesoscale cellular environments. *IEEE Comput. Graph.* **2018**, *38*, 51–66.

(9) Dirksen, A.; Dawson, P. E. Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjug. Chem.* **2008**, *19*, 2543–2548.

(10) Nauman, J. V.; Campbell, P. G.; Lanni, F.; Anderson, J. L. Diffusion of insulin-like growth factor-I and ribonuclease through fibrin gels. *Biophys. J.* **2007**, *92*, 4444–4450.

(11) Kumar, M.; Mommer, M. S.; Sourjik, V. Mobility of cytoplasmic, membrane, and DNA-binding proteins in *Escherichia coli*. *Biophys. J.* **2010**, *98*, 552–559.

(12) Berg, O. G. Orientation constraints in diffusion-limited macromolecular association. The role of surface diffusion as a rate-enhancing mechanism. *Biophys. J.* **1985**, *47*, 1–14.

(13) Northrup, S. H.; Erickson, H. P. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3338–3342.

(14) Schreiber, G.; Haran, G.; Zhou, H.-X. Fundamental aspects of protein-protein association kinetics. *Chem. Rev*. **2009**, *109*, 839–860.

(15) Qureshi, M. H.; Yeung, J. C.; Wu, S.-C.; Wong, S.-L. Development and characterization of a series of soluble tetrameric and monomeric streptavidin muteins with differential biotin binding affinities. *J. Biol. Chem.* **2001**, *276*, 46422–46428.

(16) Schreiber, G.; Fersht, A. R. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* **1996**, *3*, 427–431.

(17) Boutureira, O.; Bernardes, G. J. L. Advances in chemical protein modification. *Chem. Rev.* **2015**, *115*, 2174–2195.

(18) deGruyter, J. N.; Malins, L. R.; Baran, P. S. Residue-specific peptide modification: a chemist's guide. *Biochemistry* **2017**, *56*, 3863–3873.

(19) Gunnoo, S. B.; Madder, A. Chemical protein modification through cysteine. *ChemBioChem* **2016**, *17*, 529–553.

(20) Ravasco, J. M. J. M.; Faustino, H.; Trindade, A.; Gois, P. M. P. Bioconjugation with maleimides: A useful tool for chemical biology. *Chem. Eur. J.* **2019**, *25*, 43–59.

(21) Toda, N.; Asano, S.; Barbas, C. F. Rapid, stable, chemoselective labeling of thiols with Julia–Kocieński-like reagents: a serum-stable alternative to maleimide-based protein conjugation. *Angew. Chem. Int. Ed.* **2013**, *52*, 12592–12596.

(22) Bernardim, B.; Cal, P. M. S. D.; Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Albuquerque, I. S.; Perkins, E.; Corzana, F.; Burtoloso, A. C. B.; Jiménez-Osés, G.; Bernardes, G. J. L. Stoichiometric and irreversible cysteine-selective protein modification using carbonylacrylic reagents. *Nat. Commun.* **2016**, *7*, 13128.

(23) Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiöld, L.; Liu, C. A direct method for site-specific protein acetylation. *Angew. Chem. Int. Ed.* **2011**, *50*, 9611–9614.

(24) Conte, M. L.; Staderini, S.; Marra, A.; Sanchez-Navarro, M.; Davis, B. G.; Dondoni, A. Multi-molecule reaction of serum albumin can occur through thiol-yne coupling. *Chem. Commun.* **2011**, *47*, 11086–11088.

(25) Chalker, J. M.; Bernardes, G. J. L.; Davis, B. G. A “tag-and-modify” approach to site-selective protein modification. *Acc. Chem. Res.* **2011**, *44*, 730–741.

(26) Fierz, B.; Chatterjee, C.; McGinty, R. K.; Bar-Dagan, M.; Raleigh, D. P.; Muir, T. W. Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat. Chem. Biol.* **2011**, *7*, 113–119.

(27) Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. Organometallic palladium reagents for cysteine bioconjugation. *Nature* **2016**, *526*, 687–691.

(28) Rojas, A. J.; Pentelute, B. L.; Buchwald, S. L. Water-soluble palladium reagents for cysteine S-arylation under ambient aqueous conditions. *Org. Lett.* **2017**, *19*, 4263–4266.

(29) Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernández-González, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2011**, *2*, 1666–1676.

(30) Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J. L.; Wiewiora, R.; Ng, W.-L.; Raj, R.; Faulkner, S.; Vallée, M. R. J.; Phanumartwiwath, A.; Coleman, O. D.; Thézénas, M.-L.; Khan, M.; Galan, S. R. G.; Lercher, L.; Schombs, M. W.; Gerstberger, S.; Palm-Espling, M. E.; Baldwin, A. J.; Kessler, B. M.; Claridge, T. D. W.; Mohammed, S.; Davis, B. G. Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **2016**, *354*, aag1465.

(31) Hofmann, R.; Bode, J. W. A radical approach to posttranslational mutagenesis. *Science* **2016**, *354*, 553–554.

(32) Morais, M.; Nunes, J. P. M.; Karu, K.; Forte, N.; Benni, I.; Smith, M. E. B.; Caddick, S.; Chudasama, V.; Baker, J. R. Optimisation of the dibromomaleimide (DBM) platform for native antibody conjugation by accelerated post-conjugation hydrolysis. *Org. Biomol. Chem.* **2017**, *15*, 2947–2952.

(33) Wang, T.; Riegger, A.; Lamla, M.; Wiese, S.; Oeckl, P.; Otto, M.; Wu, Y.; Fischer, S.; Barth, H.; Kuan, S. L.; Weil, T. Water-soluble allyl sulfones for dual site-specific labelling of proteins and cyclic peptides. *Chem. Sci.* **2016**, *7*, 3234–3239.

(34) Ramos-Tomillero, I.; Perez-Chacon, G.; Somovilla-Crespo, B.; Sanchez-Madrid, F.; Domínguez, J. M.; Cuevas, C.; Zapata, J. M.; Rodríguez, H.; Albericio, F. Bioconjugation through mesitylene thiol alkylation. *Bioconjug. Chem.* **2018**, *29*, 1199–1208.

(35) Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; Lu, Y.; Meng, Y. G.; Ng, C.; Yang, J.; Lee, C. C.; Duenas, E.; Gorrell, J.; Katta, V.; Kim, A.; McDorman, K.; Flagella, K.; Venook, R.; Ross, S.; Spencer, S. D.; Wong, W. L.; Lowman, H. B.; Vandlen, R.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Mallet, W. Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* **2008**, *26*, 925–932.

(36) Sun, M. M. C.; Beam, K. S.; Cerveny, C. G.; Hamblett, K. J.; Blackmore, R. S.; Torgov, M. Y.; Handley, F. G. M.; Ihle, N. C.; Senter, P. D.; Alley, S. C. Reduction–alkylation strategies for the modification of specific monoclonal antibody disulfides. *Bioconjug. Chem.* **2005**, *16*, 1282–1290.

(37) Chen, D.; Disotuar, M. M; Xiong, X.; Wang, Y.; Chou, D H.-C. Selective N-terminal functionalization of native peptides and proteins. C*hem. Sci.* **2017**, *8*, 2717–2722.

(38) Sayers, C. T.; Mantovani, G.; Ryan, S. M.; Randev, R. K.; Keiper, O.; Leszczyszyn, O. I.; Blindauer, C.; Brayden, D. J.; Haddleton, D. M. Site-specific N-terminus conjugation of poly(mPEG1100) methacrylates to salmon calcitonin: Synthesis and preliminary biological evaluation. *Soft Matter* **2009**, *5*, 3038–3046.

(39) Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jiménez-Osés, G.; Bernardes, G. J. L. Chemo- and regioselective lysine modification on native proteins. *J. Am. Chem. Soc*. **2018**, *140*, 4004–4017.

(40) Tung, C. L.; Wong, C. T. T.; Fung, E. Y. M.; Li, X. Traceless and chemoselective amine bioconjugation via phthalimidine formation in native protein modification. *Org. Lett.* **2016**, *18*, 2600–2603.

(41) Cal, P. M. S. D.; Vicente, J. B.; Pires, E.; Coelho, A. V.; Veiros, L. F.; Cordeiro, C.; Gois, P. M. P. Iminoboronates: A new strategy for reversible protein modification. *J. Am. Chem. Soc.* **2012**, *134*, 10299–10305.

(42) Tanaka, K.; Fukase, K.; Katsumura, S. Exploring a unique reactivity of 6π-azaelectrocyclization to enzyme inhibition, natural products synthesis, and molecular imaging: an approach to chemical biology by synthetic chemists. *Synlett.* **2011**, *15*, 2115–2139.

(43) Schlick, T. L.; Ding, Z.; Kovacs, E. W.; Francis, M. B. Dual-surface modification of the Tobacco mosaic virus. *J. Am. Chem. Soc.* **2005**, *127*, 3718–3723.

(44) Jones, M. W.; Mantovani, G.; Blindauer, C. A.; Ryan, S. M.; Wang, X.; Brayden, D. J.; Haddleton, D. M. Direct peptide bioconjugation/PEGylation at tyrosine with linear and branched polymeric diazonium salts. *J. Am. Chem. Soc.* **2012**, *134*, 7406–413.

(45) Ban, H.; Gavrilyuk, J.; Barbas, C. F. Tyrosine bioconjugation through aqueous ene-type reactions: a click-like reaction for tyrosine. *J. Am. Chem. Soc.* **2010**, *132*, 1523–1525.

(46) Ban, H.; Nagano, M.; Gavrilyuk, J.; Hakamata, W.; Inokuma, T.; Barbas, C. F. Facile and stabile linkages through tyrosine: bioconjugation strategies with the tyrosine-click reaction. *Bioconjugate Chem.* **2013**, *24*, 520–532.

(47) Joshi, N. S.; Whitaker, L. R.; Francis, M. B. A three-component Mannich-type reaction for selective tyrosine bioconjugation. *J. Am. Chem. Soc.* **2004**, *126*, 15942–5943.

(48) Lorenzi, M.; Puppo, C.; Lebrun, R.; Lignon, S.; Roubaud, V.; Martinho, M.; Mileo, E.; Tordo, P.; Marque, S. R. A.; Gontero, B.; Guigliarelli, B.; Belle, V. Tyrosine-targeted spin labeling and EPR spectroscopy: an alternative strategy for studying structural transitions in proteins. *Angew. Chem. Int. Ed.* **2011**, *50*, 9108–9111.

(49) Seim, K. L.; Obermeyer, A. C.; Francis, M. B. Oxidative modification of native protein residues using Cerium(IV) ammonium nitrate. *J. Am. Chem. Soc.* **2011**, *133*, 16970–16976.

(50) Tilley, S. D.; Francis, M. B. Tyrosine-selective protein alkylation using π-allylpalladium complexes. *J. Am. Chem. Soc.* **2006**, *128*, 1080–1081.

(51) Chen, S., Li, X., Ma, H. New approach for local structure analysis of the tyrosine domain in proteins by using a site-specific and polarity-sensitive fluorescent probe. *ChemBioChem* **2009**, *10*, 1200–1207.

(52) McFarland, J. M.; Joshi, N. S.; Francis, M. B. Characterization of a three-component coupling reaction on proteins by isotopic labeling and nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **2008**, *130*, 7639–7644.

(53) Rosen, C. B.; Francis, M. B. Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **2017**, *13*, 697–705.

(54) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **2015**, *11*, 326–331.

(55) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. N-terminal modification of proteins with *o*-aminophenols. *J. Am. Chem. Soc.* **2014**. *136*, 9572–9579.

(56) Brauer, D. D.; Hartman, E. C.; Bader, D. L. V.; Merz, Z. N.; Tullman-Ercek, D.; Francis, M. B. Systematic engineering of a protein nanocage for high-yield, site-specific modification. *J. Am. Chem. Soc.* **2019**, *141*, 3875–3884.

(57) Scheck, R. A.; Dedeo, M. T.; Iavarone, A. T.; Francis, M. B. Optimization of a biomimetic transamination reaction. *J. Am. Chem. Soc.* **2008**, *130*, 11762–11770.

(58) Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M.; Weng, C.-H.; Iavarone, A. T.; Francis, M. B. Site-specific protein transamination using *N*-methylpyridinium-4-carboxaldehyde. *J. Am. Chem. Soc.* **2013**, 135, 17223–17229.

(59) Geoghegan, K. F.; Stroh, J. G. Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconjugate Chem.* **1992**, *3*, 138–146.

(60) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 776–779.

(61) Zhang, L.; Tam, J. P. Thiazolidine formation as a general and site-specific conjugation method for synthetic peptides and proteins. *Anal. Biochem.* **1996**, *233*, 87–93.

(62) Bandyopadhyay, A.; Cambray, S,; Gao, J. Fast and selective labeling of N-terminal cysteines at neutral pH via thiazolidino boronate formation. *Chem. Sci.* **2016**, *7*, 4589–4593.

(63) Faustino, H.; Silva, M. J. S. A.; Veiros, L. F.; Bernardes, G. J. L.; Gois, P. M. P. Iminoboronates are efficient intermediates for selective, rapid and reversible N-terminal cysteine functionalisation. *Chem. Sci.* **2016**, *7*, 5052–5058.

(64) Ren, H.; Xiao, F.; Zhan, K.; Kim, Y.; Xie, H.; Xia, Z.; Rao, J. A biocompatible condensation reaction for the labeling of terminal cysteine residues on proteins. *Angew Chem. Int. Ed.* **2009**, *48*, 9658–9662.

(65) Gentle, I. E.; Souza, D. P. D.; Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjugate Chem.* **2004**, *15*, 658–663.

(66) Wang, L.; Xie, J.; Schultz, P. G. Expanding the genetic code. *Biophysics Biomol. Struct.* **2006**, *35*, 225–249.

(67) Ohuchi, M.; Murakami, H.; Suga, H. The flexizyme system: A highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr. Opin. Chem. Biol.* **2007**, *11*, 537–542.

(68) Rashidian, M.; Dozier, J. K.; Distefano, M. D. Enzymatic labeling of proteins: Techniques and approaches. *Bioconjugate. Chem.* **2013**, *24*, 1277–1294.

(69) Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L. Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **2013**, *8*, 1787–1799.

(70) Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **2014**, *10*, 732–738.

(71) Harris, K. S.; Durek, T.; Kaas, Q.; Poth, A. G.; Gilding, E. K.; Conlan, B. F.; Saska, I.; Daly, N. L.; Weerden, N. L. van der; Craik, D. J.; Anderson, M. A. Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* **2015**, *6*, 10199.

(72) Jackson, D.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A designed peptide ligase for total synthesis of Ribonuclease A with unnatural catalytic residues. *Science* **1994**, *266*, 243–247.

(73) Liebscher, S.; Schöpfel, M.; Aumüller, T.; Sharkhuukhen, A.; Pech, A.; Höss, E.; Parthier, C.; Jahreis, G.; Stubbs, M. T.; Bordusa, F. N-Terminal protein modification by substrate-activated reverse proteolysis. *Angew. Chem. Int. Ed.* **2014**, *53*, 3024–3028.

(74) Schechter, I.; Berger, A. On the size of the active site in proteases. *Biochem. Biophys. Res. Com.* **1967**, *27*, 157–162.

(75) Walsh C. *Posttranslational modification of proteins: expanding nature's inventory*; Roberts and Company Publishers: Englewood, CO, 2006.

(76) Allis, C. D.; Caparros, M.-L.; Jenuwein, T.; Reinberg, D.; Lachlan, M. *Epigenetics, 2nd Edition*; Cold Spring Harbor Laboratory Press: New York, NY, 2015.

(77) Milczek, E. M. Commercial applications for enzyme-mediated protein conjugation: New developments in enzymatic processes to deliver functionalized proteins on the commercial scale. *Chem. Rev.* **2017**, *118*, 119–141.

(78) Rabuka, D. Chemoenzymatic methods for site-specific protein modification. *Curr. Opin. Chem. Biol.* **2010**, *14*, 790–796.

(79) Lotze, J.; Reinhardt, U.; Seitz, O.; Beck-Sickinger, A. G. Peptide-tags for site-specific protein labelling *in vitro* and *in vivo*. *Mol. Biosyst.* **2016**, *12*, 1731–1745.

(80) Howarth, M.; Ting, A. Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat. Protoc.* **2008**, *3*, 534–545.

(81) Fernández-Suárez, M.; Baruah, H.; Martínez-Hernández, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **2007**, *25*, 1483–1487.

(82) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15815–15820.

(83) Heller, K.; Ochtrop, P.; Albers, M. F.; Zauner, F. B.; Itzen, A.; Hedberg, C. Covalent protein labeling by enzymatic phosphocholination. *Angew. Chem. Int. Ed.* **2015**, *54*, 10327–10330.

(84) Ochtrop, P.; Ernst, S.; Itzen, A.; Hedberg, C. Exploring the substrate scope of the bacterial phosphocholine transferase AnkX for versatile protein functionalization. *ChemBioChem* **2019**, *20*, 2336–2340.

(85) Heal, W. P.; Wickramasinghe, S. R.; Bowyer, P. W.; Holder, A. A.; Smith, D. F.; Leatherbarrow, R. J.; Tate, E. W. Site-specific N-terminal labelling of proteins *in vitro* and *in vivo* using *N*-myristoyl transferase and bioorthogonal ligation chemistry. *Chem. Commun.* **2007**, *0*, 480–482.

(86) Heal, W. P.; Wickramasinghe, S. R.; Leatherbarrow, R. J.; Tate, E. W. *N*-myristoyl transferase-mediated protein labelling *in vivo*. *Org. Biomol. Chem.* **2008**, *6*, 2308–2315.

(87) Duckworth, B. P.; Zhang, Z.; Hosokawa, A.; Distefano, M. D. Selective labeling of proteins by using protein farnesyltransferase. *ChemBioChem* **2007**, *8*, 98–105.

(88) Lin, C.-W.; Ting, A. Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins *in vitro* and on the surface of living cells. *J. Am. Chem. Soc.* **2006**, *128*, 4542–4543.

(89) Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.* **2003**, *21*, 86–89.

(90) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V. HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **2008**, *3*, 373–382.

(91) Bonasio, R.; Carman, C. V.; Kim, E.; Sage, P. T.; Love, K. R.; Mempel, T. R.; Springer, T. A.; Andrian, U. H. von. Specific and covalent labeling of a membrane protein with organic fluorochromes and quantum dots. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14753–14758.

(92) Hilvert, D. Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **2000**, *69*, 751–793.

(93) Rader, C.; Sinha, S. C.; Popkov, M.; Lerner, R. A.; Barbas, C. F. Chemically programmed monoclonal antibodies for cancer therapy: adaptor immunotherapy based on a covalent antibody catalyst. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5396–5400.

(94) Gavrilyuk, J. I.; Wuellner, U.; Barbas, C. F. β-Lactam-based approach for the chemical programming of aldolase antibody 38C2. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 1421–1424.

(95) Zakeri, B.; Howarth, M. Spontaneous intermolecular amide bond formation between side chains for irreversible peptide targeting. *J. Am. Chem. Soc.* **2010**, *132*, 4526–4527.

(96) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E690–E697.

(97) Hagan, R. M.; Björnsson, R.; McMahon, S. A.; Schomburg, B.; Braithwaite, V.; Bühl, M.; Naismith, J. H.; Schwarz-Linek, U. NMR spectroscopic and theoretical analysis of a spontaneously formed Lys–Asp isopeptide bond. *Angew. Chem. Int. Ed.* **2010**, *49*, 8421–8425.

(98) Keeble, A. H.; Turkki, P.; Stokes, S.; Anuar, I. N. A. K.; Rahikainen, R.; Hytönen, V. P.; Howarth, M. Approaching infinite affinity through engineering of peptide–protein interaction. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 26523–26533.

(99) Muralidharan, V.; Muir, T. W. Protein ligation: An enabling technology for the biophysical analysis of proteins. *Nat. Methods.* **2006**, *3*, 429–438.

(100) Shah, N. H.; Muir, T. W. Inteins: Nature's gift to protein chemists. *Chem. Sci.* **2013**, *5*, 446–461.

(101) Lockless, S. W.; Muir, T. W. Traceless protein splicing utilizing evolved split inteins. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10999–11004.

(102) Shah, N. H.; Dann, G. P.; Vila-Perelló, M.; Liu, Z.; Muir, T. W. Ultrafast protein splicing is common among cyanobacterial split inteins: Implications for protein engineering. *J. Am. Chem. Soc.* **2012**, *134*, 11338–11341.

(103) Vila-Perelló, M.; Liu, Z.; Shah, N. H.; Willis, J. A.; Idoyaga, J.; Muir, T. W. Streamlined expressed protein ligation using split inteins. *J. Am. Chem. Soc.* **2013**, *135*, 286–292.

(104) David, Y.; Vila-Perelló, M.; Verma, S.; Muir, T. W. Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. *Nat. Chem.* **2015**, *7*, 394–402.

(105) Fierer, J. O.; Veggiani, G.; Howarth, M. SpyLigase peptide–peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1176–E1181.

(106) Buldun, C. M.; Jean, J. X.; Bedford, M. R.; Howarth, M. SnoopLigase catalyzes peptide–peptide locking and enables solid-phase conjugate isolation. *J. Am. Chem. Soc.* **2018**, *140*, 3008–3018.

(107) Siegmund, V.; Piater, B.; Zakeri, B.; Eichhorn, T.; Fischer, F.; Deutsch, C.; Becker, S.; Toleikis, L.; Hock, B.; Betz, U. A. K.; Kolmar, H. Spontaneous isopeptide bond formation as a powerful tool for engineering site-specific antibody-drug conjugates. *Sci. Rep.* **2016**, *6*, 39291.

(108) Sato, H.; Ikeda, M.; Suzuki, K.; Hirayama, K. Site-specific modification of interleukin-2 by the combined use of genetic engineering techniques and transglutaminase. *Biochemistry* **1996**, *35*, 13072–13080.

(109) Jeger, S.; Zimmermann, K.; Blanc, A.; Grünberg, J.; Honer, M.; Hunziker, P.; Struthers, H.; Schibli, R. Site-specific and stoichiometric modification of antibodies by bacterial transglutaminase. *Angew. Chem. Int. Ed.* **2010**, *49*, 9995–9997.

(110) Dennler, P.; Chiotellis, A.; Fischer, E.; Brégeon, D.; Belmant, C.; Gauthier, L.; Lhospice, F.; Romagne, F.; Schibli, R. Transglutaminase-based chemo-enzymatic conjugation approach yields homogeneous antibody–drug conjugates. *Bioconjugate Chem.* **2014**, *25*, 569–578.

(111) Farias, S. E.; Strop, P.; Delaria, K.; Casas, M. G.; Dorywalska, M.; Shelton, D. L.; Pons, J.; Rajpal, A. Mass spectrometric characterization of transglutaminase based site-specific antibody–drug conjugates. *Bioconjugate Chem*. **2014**, *25*, 240–250.

(112) Ton-That, H.; Schneewind, O. Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol. Microbiol.* **2003**, *50*, 1429–1438.

(113) McConnell, S. A.; Amer, B. R.; Muroski, J.; Fu, J.; Chang, C.; Loo, R. R. O.; Loo, J. A.; Osipiuk, J.; Ton-That, H.; Clubb, R. T. Protein labeling via a specific lysine-isopeptide bond using the pilin polymerizing sortase from *Corynebacterium diphtheriae*. *J. Am. Chem. Soc.* **2018**, *140*, 8420–8423.

(114) Bellucci, J. J.; Bhattacharyya, J.; Chilkoti, A. A noncanonical function of sortase enables site-specific conjugation of small molecules to lysine residues in proteins. *Angew. Chem. Int. Ed.* **2015**, *54*, 441–445.

(115) Sue, C. K.; McConnell, S. A.; Ellis-Guardiola, K.; Muroski, J. M.; McAllister, R. A.; Yu, J.; Alvarez, A. I.; Chang, C.; Loo, R. R. O.; Loo, J. A.; Ton-That, H.; Clubb, R. T. Kinetics and optimization of the lysine–isopeptide bond forming sortase enzyme from *Corynebacterium diphtheriae*. *Bioconjugate Chem.* **2020**, *31*, 1624–1634.

(116) Hofmann, R.; Akimoto, G.; Wucherpfennig, T. G.; Zeymer, C.; Bode, J. W. Lysine acylation using conjugating enzymes for site-specific modification and ubiquitination of recombinant proteins. *Nat. Chem.* **2020**, *12*, 1008–1015.

CHAPTER 2

Lysine Acylation using Conjugating Enzymes (LACE)

Portions of the work described in this Chapter have been performed in collaboration with Dr. Thomas Wucherpfennig (Bode Group, ETH Zürich) who performed initial experiments, Sara Marie Duke (visiting student, University of Texas at Austin) who contributed to T4L labeling experiments, and Prof. Dr. Cathleen Zeymer (ETH Zürich, current address: TU Munich) who contributed to structural analyses by X-ray crystallography.

1. Introduction

1.1. SUMO and ubiquitin-like proteins

Ubiquitin (Ub) and ubiquitin-like proteins (Ubls) constitute a family of almost 20 eukaryotic proteins that are posttranslationally conjugated to other proteins,[117] with critical functions in many cellular processes including signal transduction,[118] proteostasis,[119] transcription,[120] DNA repair[121] and cell-cycle control.[122] The diversity of Ubl signaling is further increased by modification of substrates with multiple individual Ubls at different lysine residues, and the formation of linear and branched Ubl chains when preinstalled Ubls themselves are modified by the Ubl machinery. For example, polyubiquitination is a hallmark modification that targets substrate proteins to the proteasome for degradation.[123–125]

The 76-residue ubiquitin protein is one of the most evolutionarily conserved eukaryotic proteins.[126] Since its discovery, a growing family of Ubls have been described,[127] including small ubiquitin-like modifier proteins (SUMO),[128] Nedd8,[129] ubiquitin-fold modifier 1 (Ufm1)[130] and interferon-stimulated gene 15 (ISG15).[131] A common sequence feature of most Ubls is a C-terminal Gly-Gly motif in the mature form of the protein (Figure 17a), while the overall amino acid identity is generally low between the different Ubls (Figure 17b).

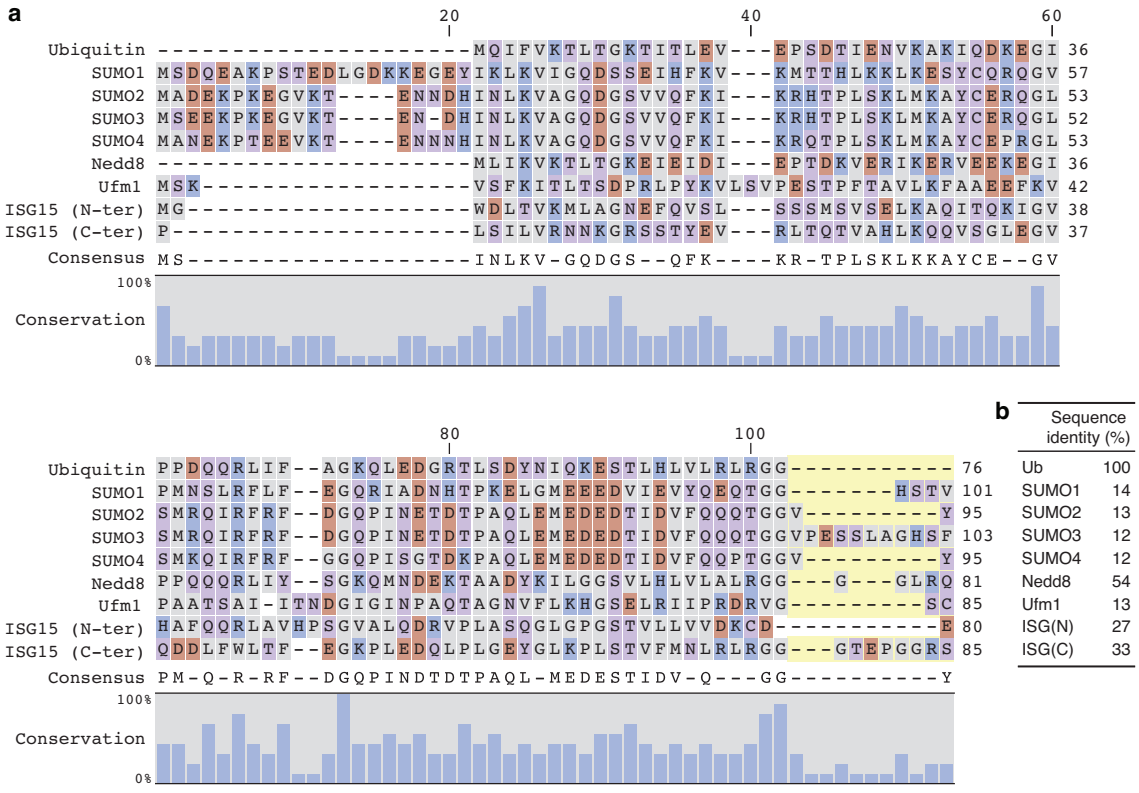


Figure 17. Ubiquitin-like proteins. (**a**) Sequence alignment of human ubiquitin and ubiquitin-like proteins (Ubls). The N- and C-terminal domains of ISG15 are listed individually. C-terminal residues belonging to propeptides, which are removed during Ubl maturation, are underlaid in yellow. The sequence alignment was generated using CLC Genomics Workbench 12.0.3. (**b**) Sequence identity with ubiquitin, including

propeptides. Note: Naming of SUMO2 and SUMO3 has been applied inconsistently throughout the literature. We use the nomenclature of the UniProt database, which follows the seminal publication.[132]

Despite low sequence identity, Ubls share a common β-grasp fold (Figure 18a).[133] Mapping the amino acid conservation of Ubls onto the structure of ubiquitin shows that most conserved residues appear to have a structural role, in particular glycine residues that mark the beginning or end of secondary structures (Figure 18b). Other conserved residues are predominantly located within the β-sheet. In contrast to small molecule PTMs such as acetylation or phosphorylation, the attachment of a Ubl protein introduces large new surfaces on a target protein. As a consequence, modification by the diverse Ubls can sterically block further modification or existing binding sites, or act as signaling hubs for the many cellular processes in which they are active.[117]

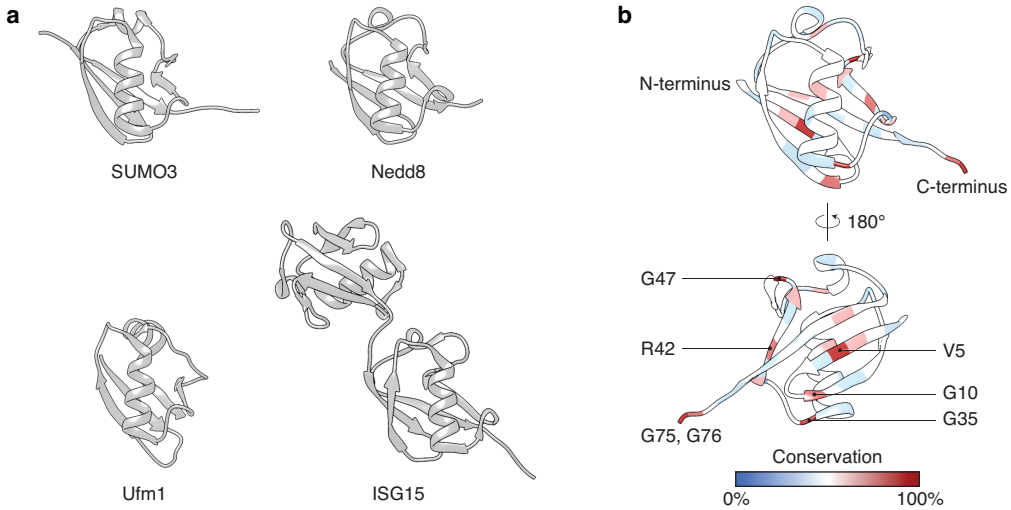


Figure 18. Structure and conservation of Ubls. (**a**) Structures of four Ubls (PDB entries 1u4a, 1ndd, 5ia7 and 5chw). (**b**) Structure of ubiquitin (PDB entry 1ubq), colored by conservation of residues based on the sequences listed in Figure 17a. Seven highly conserved residues are labeled.

1.2. The Ubl conjugation pathway

Conjugation of Ubls to a target protein involves the formation of an isopeptide bond between the C-terminus of the Ubl and an acceptor lysine of the substrate. The pathway generally relies on the consecutive action of three enzymes – E1 activating enzymes, E2 conjugating enzymes, and E3 ligases (Figure 19).[134,135] Ubls that are expressed as a C-terminal fusion with a propeptide undergo maturation by proteases or deubiquitinases (DUBs)[136,137] to reveal a C-terminal glycine residue before they can be conjugated to a target protein. Modification of proteins with Ubls can be a dynamic and reversible process. Removal of Ubl is also performed by DUBs. In an ATP- and magnesium-dependent step, E1 activating enzymes catalyze adenylation of the Ubl C-terminus with subsequent release of pyrophosphate. This high-energy intermediate is then trapped by an E1 active site cysteine to form an E1-Ubl thioester intermediate (E1~Ubl) with concomitant release of AMP.[138] The approximately ten known E1 activating enzymes play a

crucial role in isoform selectivity by activating specific Ubl isoforms and coordinate their use in downstream pathways with cognate E2 and E3 enzymes.[139] Charging of the E1 enzyme with a Ubl results in a conformational change that allows the E1~Ubl thioester intermediate to bind a cognate E2 conjugating enzyme from a group of about 40 E2 enzymes. Upon binding of the E2, the Ubl is then transferred to an E2 active site cysteine via transthioesterification.[140]

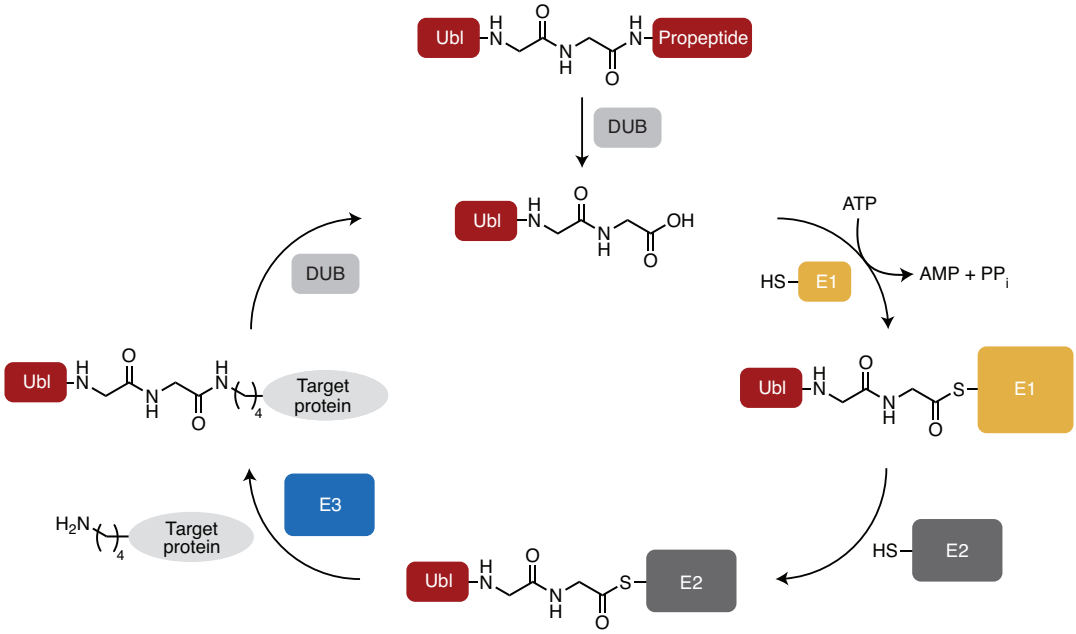


Figure 19. Mechanism of Ubl conjugation by the consecutive action of E1 activating enzymes, E2 conjugating enzymes, and E3 ligases. Deubiquitinases (DUBs) and proteases cleave Ubl conjugates and propeptides.

Transfer of the Ubl from the E2~Ubl thioester intermediate to a target protein is aided by E3 ligases that contribute to selectivity and efficiency of Ubl transfer (Figure 20).[141] Specific modification of a cellular substrate with a given Ubl is achieved by a great diversity of E3 ligases, with over 600 known E3s in humans.[142] Despite the large number of E3 ligases, most of them fall into three general categories. Really interesting new gene (RING) E3 ligases act as scaffolds between a cognate E2~Ubl and a specific target protein and mediate direct aminolysis of the E2~Ubl thioester intermediate by the acceptor lysine (Figure 20a).[141,143] Compared to RING E3 ligases, the homology to E6AP C-terminus (HECT)[144,145] and the RING-between-RING (RBR) [146,147] family E3 ligases themselves have an active site cysteine. The Ubl is transferred from the E2 to the active site cysteine of HECT- or RBR-type E3s, followed by transfer from the E3~Ubl thioester intermediate to the substrate (Figure 20b).

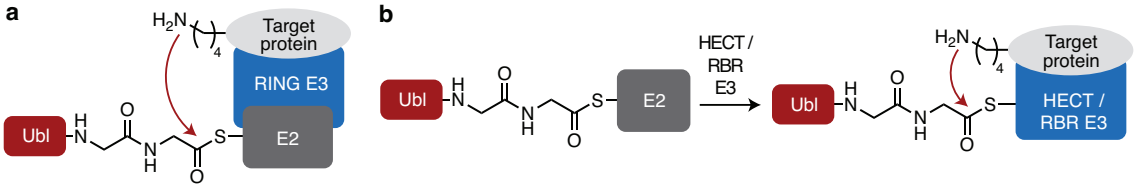


Figure 20. Mechanism of RING-type (**a**) as well as HECT- and RBR-type (**b**) E3 ligases.

1.2.1. SUMOylation machinery and the SUMO isoforms

SUMO activation is performed by a single E1 activating enzyme, a heterodimeric enzyme consisting of SUMO activating enzyme subunit 1 (Sae1) and Sae2.[148,149] Likewise, only a single highly conserved E2 enzyme, Ubc9 (Ube2I), is involved in the SUMOylation machinery.[150] Remarkably, Ubc9 is able to directly bind and modify lysine residues that are located within a short consensus SUMOylation motif (see also Section 2.1).[151] Recently, several other SUMOylation motifs have been described, including a phosphorylation-dependent SUMOylation motif[152,153] as well as a reverse consensus motif.[154–156] This direct interaction of Ubc9 with target proteins containing a consensus SUMOylation motif enables SUMOylation of certain substrates, even in the absence of E3 ligases.[151] The specificity and efficiency of SUMOylation is enhanced by E3 ligases and is thought to be the predominant mechanism for substrate selection *in vivo*.[157] Alternatively, hydrophobic patches called SUMO-interacting motifs (SIMs) in target proteins can bind to SUMO of the Ubc9~SUMO thioester intermediate and promote SUMO transfer in this manner.[158]

All four SUMO isoforms harbor an unstructured N-terminal extension of the core Ubl fold, and require maturation of the C-terminal propeptide for conjugation (see Figure 17a). Propeptide cleavage and deSUMOylation is performed by SUMO-specific proteases, including sentrin-specific proteases (SENPs)[159] and deSUMOylating isopeptidases.[160,161] SUMO4 contains a proline residue (P90) in its C-terminal region, which renders it resistant to maturation by SENPs.[162] As a result, SUMO4 is not generally found to be conjugated to target proteins, and the role of SUMO4 and its importance is poorly understood.[163,164] Among the four SUMO isoforms, SUMO2 and SUMO3 are very similar, with 96% sequence identity. These two isoforms are commonly referred to as SUMO2/3 and appear to confer the same function in most contexts. SUMO2/3 is generally more abundant inside cells and conjugation of these isoforms is more dynamic than that of SUMO1.[165,166] A further distinguishing feature of SUMO2/3 is the presence of a consensus SUMOylation motif at K11 in the N-terminal extension, resulting in the formation of SUMO chains.[167,168]

1.3. E2 conjugating enzymes, central actors in Ubl conjugation

E2 conjugating enzymes are only approximately twice the size of the Ubl which they are involved in transferring, but they have a central role in both substrate selection as well as mediating transfer of the Ubl to the acceptor lysine.[169] For instance, despite the seemingly dominant role of E3 ligases in target selection, E2 enzymes have been shown to function together with E3 ligases in Ubl transfer and substrate specificity[170] and in some cases exhibit strong intrinsic selectivity.[151,171]

E2 enzymes share a conserved fold consisting of four α-helices and a four-stranded β-sheet (Figure 21). Important regions for the E2 domain fold are a conserved proline-rich loop (PXXPP, Loop 4) and an interacting tryptophan residue, which together are thought to stabilize the active site and the relative orientation of the surrounding loops (Loops 4–7).[172] Another hallmark of E2 enzymes is the gateway residue located in Loop 8, which forms the opening to the active site cleft. The gateway residue is most often an aspartic acid or serine (Figure 21), which is frequently followed by a neighboring proline residue (see Figure 23).[169] It is presumed that E2 enzymes with an aspartic acid as gateway residue are constitutively active,[169,173] whereas E2 enzymes that harbor a serine residue at this position may be controlled by phosphorylation.[174,175]

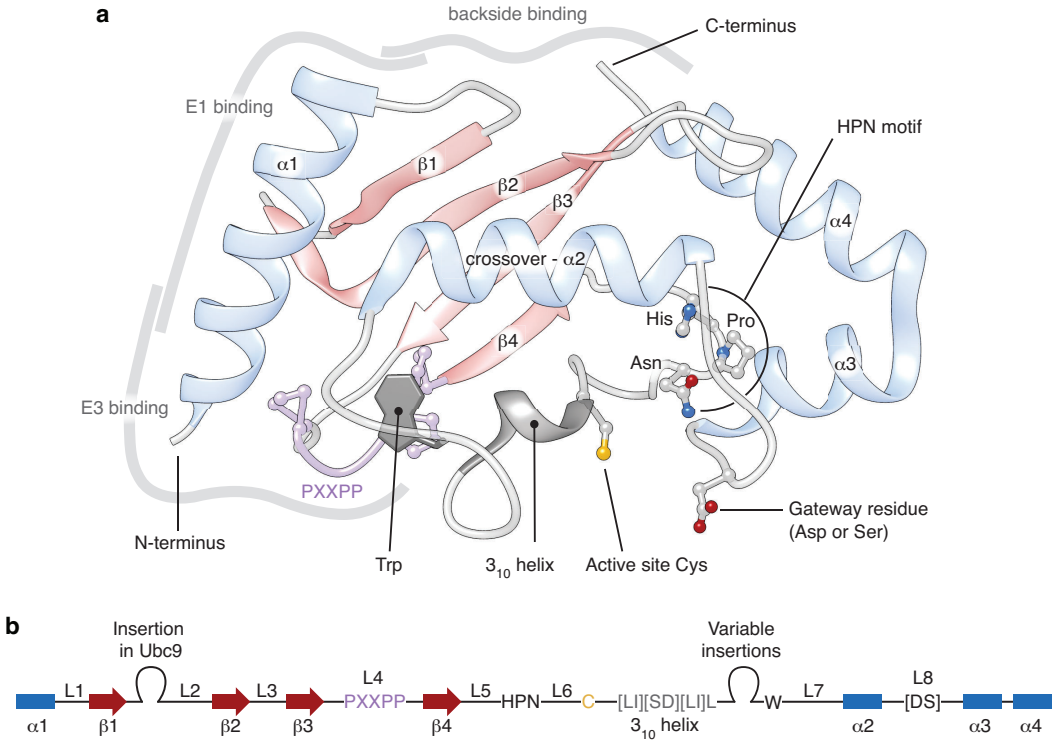


Figure 21. General architecture of E2 conjugating enzymes (**a**), and schematic representation of primary and secondary structure elements (**b**). Conserved, as well as structurally and catalytically important residues are highlighted. α-Helices (blue) are labeled as α1–α4, β-sheets (red) as β1–β4, and loops as L1–L8. Binding surfaces are indicated with grey lines. The Figure is based on the structure of Ubc9 (PDB entry 5d2m) and is adapted from Stewart *et al.*,[169] Michelle *et al.*[172] and Winn *et al.*[176]

Central to the activity of E2 enzymes is an asparagine residue in a highly conserved His-Pro-Asn (HPN) motif, which is positioned in close proximity to the active site cysteine (Figure 21). While the functional importance of the asparagine residue for efficient aminolysis of the E2~Ubl thioester intermediate by the acceptor lysine is well established, the exact mechanism by which this residue operates has not been determined conclusively. One hypothesis is that the asparagine residue in the HPN motif is positioned such that it can form a hydrogen bond to the thioester carbonyl for oxyanion stabilization of the tetrahedral intermediate during aminolysis.[177] Alternatively, a computational study offered an entropy-based explanation and predicted that the asparagine residue actually increases the barrier to product formation compared to an alanine residue, but that the hydrogen bond helps stabilize the motion of the electrophile prior to attack by the lysine residue.[178] Lastly, a structural and biochemical approach suggested that the importance of the asparagine residue stems primarily from its structural role in stabilizing the active site loop, rather than having an active role in catalysis.[179]

The active site cysteine of E2 conjugating enzymes has been found to exhibit a pK_a of approximately 10.5, which is significantly elevated compared to that of free cysteine with a pK_a of approximately 8.3.[180] It has been proposed that the high pK_a of the active site cysteine prevents E2 enzymes from reacting spontaneously with cellular electrophiles, and interaction with E1~Ubl intermediates may activate the E2 for controlled transthioesterification.[180] The pool of E2 enzymes inside cells is mostly present in the charged form as E2~Ubl thioester.[181] Solution studies of the E2~Ubl thioester intermediate showed that this thioester conjugate is generally flexible and that the Ubl samples different orientations relative to the E2.[182] This flexibility is thought to prevent efficient activation of the thioester carbonyl for aminolysis in the resting state of E2~Ubl thioester intermediate, a state that has been termed the 'open conformation' (Figure 22). Indeed, the propensity of certain E2~Ubl thioester intermediates to adopt a 'closed conformation' correlates with increased aminolysis reactivity with lysine residues.[183] The E2 active site itself is devoid of residues that could perform in acid-base catalysis. Instead, the main contribution to catalysis is thought to stem from pK_a suppression and desolvation of the acceptor lysine upon binding of the target protein.[184]

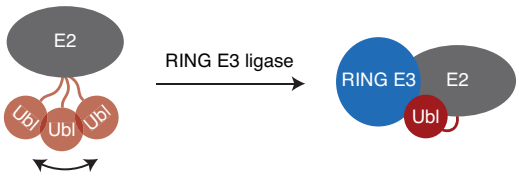


Figure 22. Schematic representation of the open (left) and closed (right) conformation of the E2~Ubl thioester intermediate. The closed conformation can be stabilized by binding of a RING E3 ligase. Adapted from Stewart *et al.*[169]

The canonical binding sites for E1 and E3 enzymes on E2 conjugating enzymes are overlapping and located near the N-terminus and α1-helix (Figure 21), with the result that E1 and E3 binding are mutually exclusive.[185,186] Additional interactions with E3 ligases[187] or Ubls[188] via an E2 backside binding site (Figure 21) have been shown to be important for some Ubl transfer mechanisms, for example Ubl chain formation.[189] Binding of a RING E3 ligase induces a closed conformation, in which the Ubl folds back onto the E2 crossover α2-helix, the C-terminal tail is locked in the E2 active site cleft and the thioester carbonyl is positioned for nucleophilic attack by the acceptor lysine in the active site (Figure 22).[173] Interactions between the RING domain and the C-terminal Ubl tail promote formation of this closed conformation.[190] Additionally, a conserved 'linchpin' residue of the E3 RING domain, typically arginine,[191] acts as a hydrogen bond donor for a backbone carbonyl group in Loop 7 of the E2 and one or more backbone groups of the Ubl tail, resulting in allosteric modulation of the E2 active site and re-positioning of the HPN motif.[192] These conformational changes are thought to reconfigure the active site of conjugating enzymes for efficient aminolysis and transfer of the Ubl.

The importance of the 'housekeeping residues' of E2 conjugating enzymes that have structural or functional roles can be illustrated by looking at the sequence conservation. Figure 23 shows the structure of the SUMO conjugating enzyme Ubc9 in which the residues are colored by amino acid conservation. From this representation, it becomes clear that the housekeeping residues (see Figure 21) are highly conserved, whereas other residues show moderate to low sequence conservation.

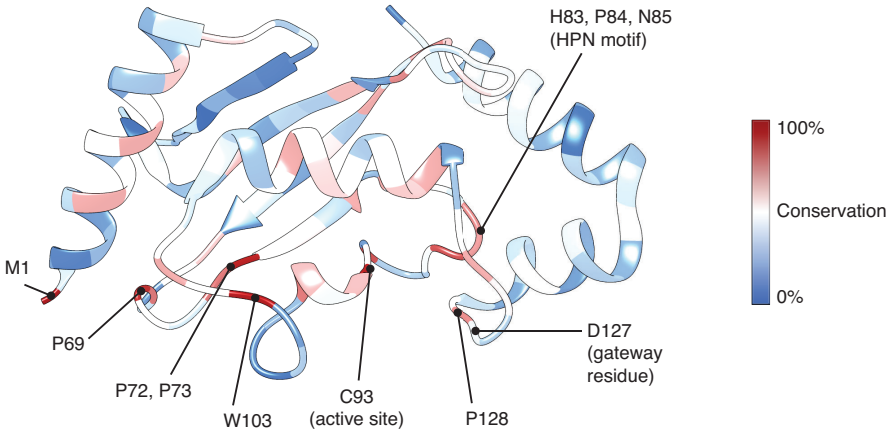


Figure 23. Conservation of E2 enzymes**.** Structure of Ubc9 (PDB entry 5d2m), colored by conservation of residues based on the sequence alignment in Figure 100 of the Appendix. Highly conserved residues are labeled. In addition, D127 of Ubc9 is marked, even though the position is less conserved and several homologs contain a serine at this position.

1.4. Protein labeling methods utilizing components of the Ubl pathway

The Ubl system is arguably the most prominent eukaryotic pathway for protein conjugation at internal residues. However, only a few protein engineering and labeling methods have been described that rely on parts of the Ubl system (Figure 24). Madden *et al.* showed that biotinylated peptides derived from the C-terminus of ubiquitin (LRLRGG) could be activated by E1 activating enzymes in an ATP-dependent fashion and were transferred to tubulin in the presence of its specific E3 ligase parkin (Figure 24a).[193] Similarly, the authors observed biotinylation of cell lysates with similar patterns as the one observed with native ubiquitination, and the product distribution was largely dominated by E3 ligases. Chang *et al.* engineered a DUB which favors aminolysis of the acyl intermediate over hydrolysis and is capable of ubiquitinating itself at a specific lysine residue if presented with a Ub-D77 substrate (Figure 24b).[194] Lastly, certain HECT E3 ligases were shown to react directly with C-terminal ubiquitin thioesters, without the need for ATP, E1 and E2 enzymes, to directly ubiquitinate natural substrates of these E3 ligase and effect the formation of polyubiquitin chains (Figure 24c).[195] While these methods can be used to modify certain specific substrates, no general method has been developed utilizing components from the Ubl pathway.[196]

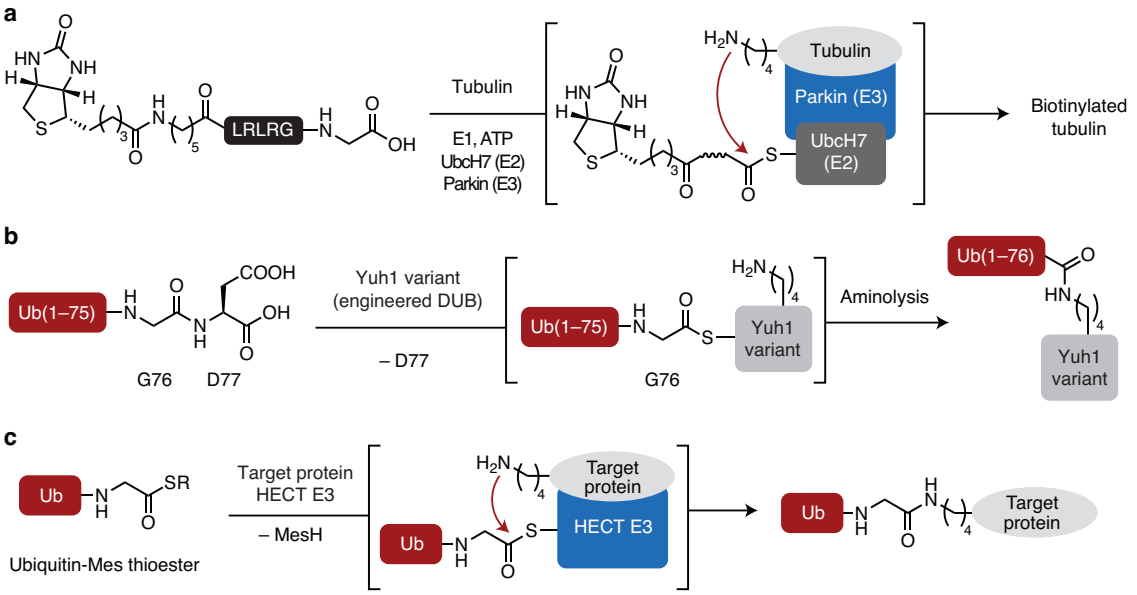


Figure 24. Protein engineering using components of the ubiquitination pathway. (**a**) Biotinylation using ubiquitin-derived synthetic peptides with the canonical ubiquitination pathway as reported by Madden *et al.*[193] (**b**) Self-ubiquitination of an engineered DUB as described by Chang *et al.*[194] (**c**) Chemically activated ubiquitin thioesters for HECT E3 ligase-mediated transfer to target proteins as shown by Park *et al.*[195] Mes, 2-mercaptoethanesulfonate.

2. Harnessing the SUMOylation pathway for protein modification

2.1. Sequence-specific SUMOylation by Ubc9 in the absence of E3 ligases

At the outset of our investigations, we were intrigued by the relative simplicity of the SUMOylation pathway,[157] for which a clear consensus motif around SUMOylated lysine residues has emerged from biochemical[197] and proteomics[153] studies (Figure 25a, see also Section 1.2.1). The sequence specificity arises from Ubc9, the only known SUMO-conjugating enzyme. Ubc9 recognizes and modifies lysines embedded in a consensus SUMOylation motif, even in the absence of E3 ligases.[198] This is exemplified in Figure 25b,c, which shows SUMOylation of the model substrate Ran GTPase-activating protein 1 (RanGAP1) in the presence of the SUMO activating enzyme Sae1/Sae2 (E1), ATP and Ubc9. Although E3 enzymes can significantly contribute to the specificity and efficiency of SUMOylation, the reaction proceeds even in the absence of an E3 ligase. This sequence specificity and ability to directly engage substrates in an E3-independent manner is a unique feature of Ubc9 among the approximately 40 E2-conjugating enzymes.[169,199]

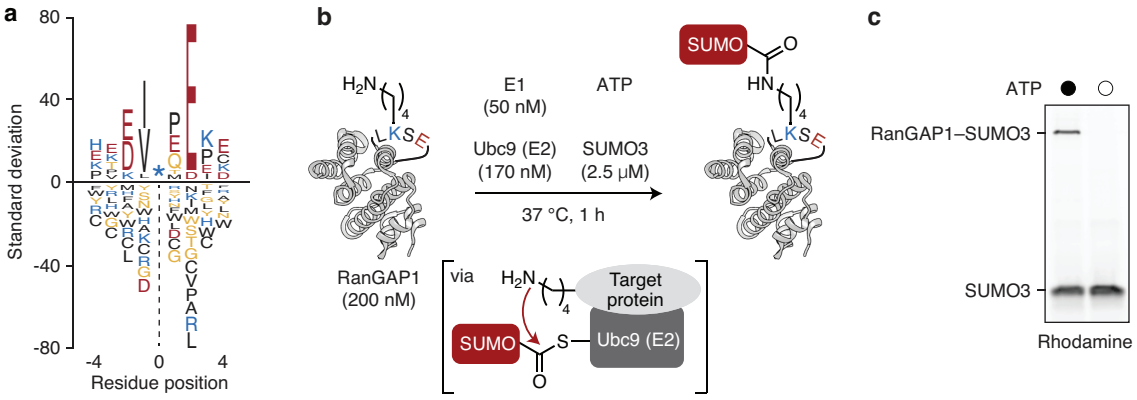


Figure 25. Ubc9 effects SUMOylation of acceptor lysines embedded in a consensus SUMOylation motif in the absence of E3 ligases. (**a**) Consensus SUMOylation motif, adapted from Hendriks *et al.*[153] (**b**) SUMOylation of RanGAP1 with rhodamine-labeled SUMO3 using E1, ATP and Ubc9. The SUMOylation motif (LKSE) and the interaction between Ubc9 and the target protein are shown schematically. (The protein image is based on PDB entry 2grn.) (**c**) Reaction analysis of **b** by in-gel rhodamine fluorescence after 1 h in the presence (filled circle) or absence (hollow circle) of ATP.

2.2. Thioesters as acyl donors for Ubc9

To circumvent the need for ATP-dependent activation by E1 enzymes, we sought to employ short peptide thioesters as acyl donors that would serve as a minimal analog of the thioester-linked E1 intermediate (Ubl~E1) (Figure 26).

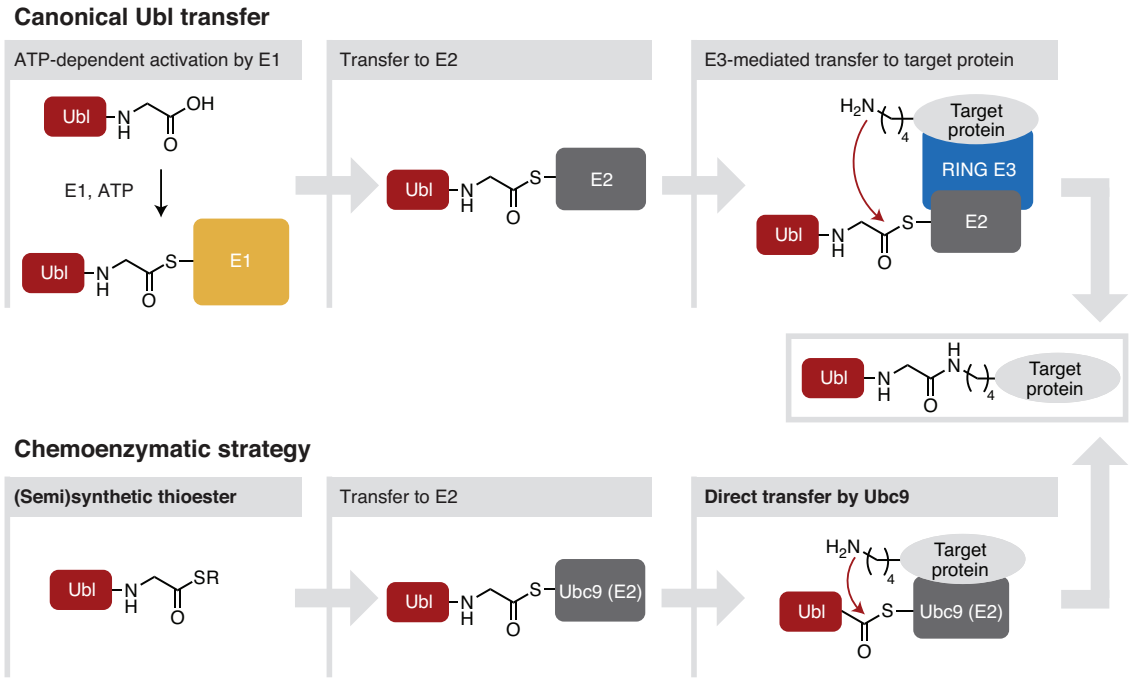


Figure 26. General strategy for Ubc9-mediated isopeptide labeling without the need for E1 and E3 enzymes.

With this chemoenzymatic approach, the canonical SUMOylation pathway, consisting of the consecutive action of three enzymes, would be simplified to a process that relies on just the E2 enzyme Ubc9. Furthermore, ATP as a cofactor would not be required due to the chemical preactivation of the acyl donors as thioesters.

As a first test, we prepared synthetic peptide thioesters with the cysteine derivative *N*-acetyl-L-cysteine-*N*-methylamide (Ac-Cys-NHMe (**1**), see Experimental Part for a column-free gram-scale synthesis), intended to mimic the active site cysteine of E1. For the peptide sequence, we chose QQQTGG (**2a**), corresponding to the last six residues of the flexible C-terminal tail of SUMO2/3. As additional leaving groups, we tested the commercially available 2-mercaptoethanesulfonate (Mes) and *N*-acetylcysteamine in thioesters **2b** and **2c**, respectively. We investigated the ability of these peptide thioesters to serve as acyl donors in place of the Ubl~E1 intermediate in Ubc9-mediated labeling reactions of RanGAP1 (Figure 27a).

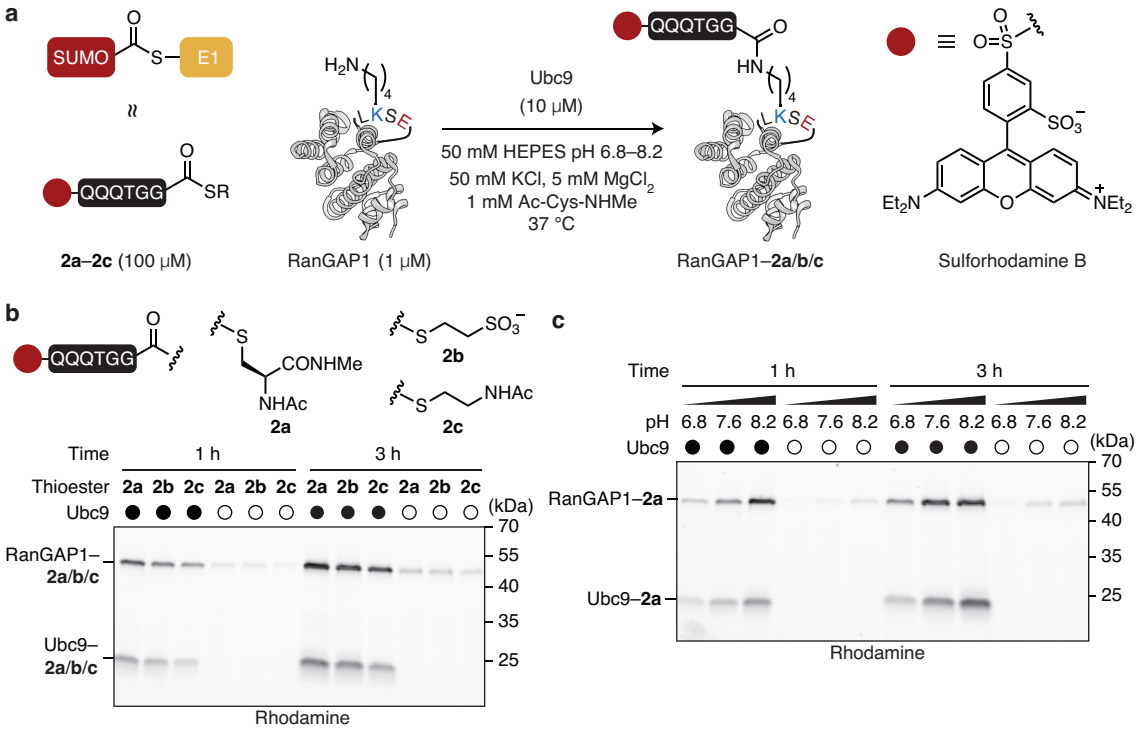


Figure 27. Synthetic peptide thioesters as analog of the E1~Ubl thioester intermediate. (**a**) RanGAP1 labeling with synthetic peptide thioesters **2a**–**2c**. For reaction monitoring, the peptides were modified with sulforhodamine B at the N-terminus. (**b**,**c**) In-gel rhodamine fluorescence analysis of reactions in the presence (filled circle) or absence (hollow circle) of Ubc9 and the indicated thioester (**b**), and of reactions with thioester **2a** at varying reaction pH (**c**).

We were pleased to find that all thioesters were transferred in a Ubc9–dependent manner (Figure 27b). Thioester **2a** exhibited the highest reactivity and specificity, followed by thioester **2b**. Performing the reaction at varying pH from pH 6.8–8.2 showed a strong pH-dependence of the labeling efficiency (Figure 27c). Based on these results, we chose Ac-Cys-NHMe as the preferred leaving group, and a reaction pH of 7.6 as a suitable compromise between reactivity and specificity for further reaction development.

Given that Ubc9 and the thioester had to be used in relatively large excess compared to the substrate RanGAP1 – 10 and 100 equivalents, respectively – we sought to further optimize the acyl donor. We screened thioesters corresponding to **2a** with peptide sequences taken from the other two SUMO isoforms, SUMO1 (**3**) and SUMO4 (**4**), as well as ubiquitin (**5**) in labeling reactions of RanGAP1 (Figure 28).

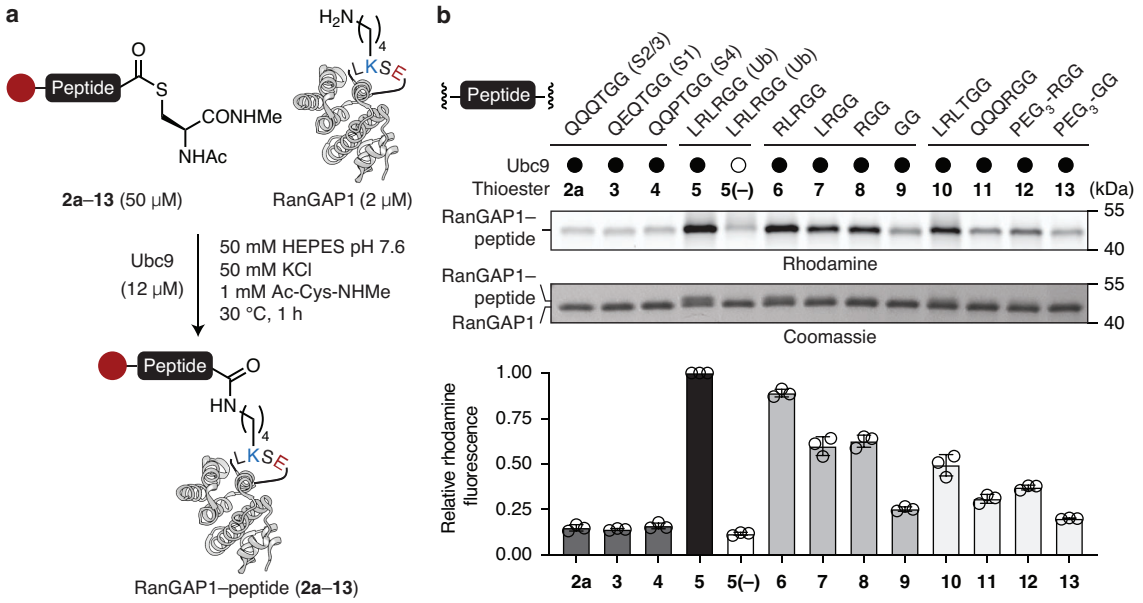


Figure 28. Sequence optimization of the acyl donor. (**a**) RanGAP1 labeling with synthetic peptide thioesters **2a**–**13**. (**b**) In-gel fluorescence analysis (top) and Coomassie-stained sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS–PAGE) (middle) of reactions with the indicated thioesters. Quantification of RanGAP1–peptide formation relative to thioester **5** by densitometry of rhodamine fluorescence (bottom). Individual data points are shown, error bars represent s.d.

Surprisingly, the ubiquitin-derived thioester **5** exhibited the highest reactivity from the tested Ubl-derived peptide thioesters, and the labeling proceeded in a Ubc9-dependent manner (Figure 28b). In contrast, the SUMO-derived thioesters **2a**–**4** showed generally low reactivity with minimal variation between the different isoforms.

To determine the residues that contribute to the high reactivity of **5**, we synthesized truncated thioesters **6**–**9**. A stepwise reduction in reactivity was observed after each removal of an arginine (Figure 28b). The tripeptide RGG (**8**) was found to be the minimal sequence which provided acceptable conversion. The hybrid peptides **10** and **11**, combining sequence features of SUMO2/3 and ubiquitin, did not perform better than entirely ubiquitin-derived thioesters. Comparison of thioesters **8** and **9** with **12** and **13**, the latter containing short polyethylene glycol (PEG) linkers between the rhodamine and the peptide, indicates that the reduced reactivity of the truncated peptides was not due to steric effects or other unfavorable interactions between Ubc9 and the fluorophore.

Tandem mass spectrometry (MS/MS) analysis of the Ubc9-mediated RanGAP1 labeling reaction with thioester **5** showed exclusively the desired isopeptide product with the acceptor lysine located within the SUMOylation motif in RanGAP1 (K524) (Figure 29).

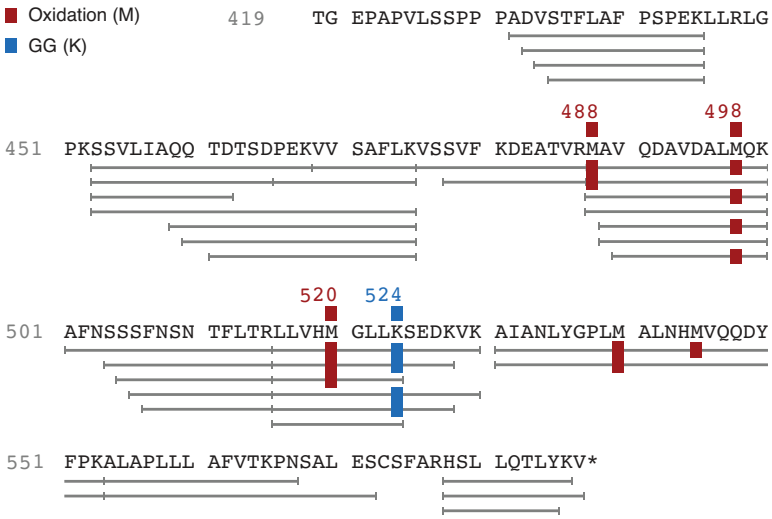


Figure 29. Observed tryptic peptide fragments (grey bars) after Ubc9-mediated labeling of RanGAP1 with peptide **5**. Methionine oxidation (red) and glycylglycine modification (blue) of K524, corresponding to the modification site with peptide **5**, were observed and mapped to the sequence.

3. LACE tag for isopeptide labeling of recombinant proteins

3.1. Identification of a LACE tag

With an optimized acyl donor in hand, we tested whether non-canonical substrates of Ubc9 could be modified in the same manner as RanGAP1. In its simplest form, the consensus SUMOylation motif recognized by Ubc9 consists of the four residues ΨKXD/E, where Ψ is a hydrophobic residue, K is the lysine to be modified, X is a non-conserved residue and D/E is an acidic residue.[157] We tested three 11-residue-long sequences that all contained consensus SUMOylation motifs – taken from the commonly SUMOylated proteins[154] RanGAP1, IκBα and PML – by fusing them to the C-terminus of superfolder green fluorescent protein (GFP)[200] (Figure 30a,b). All variants were found to be labeled in a Ubc9-dependent manner, whereby the sequence taken from PML provided the best reactivity (Figure 30c,d).

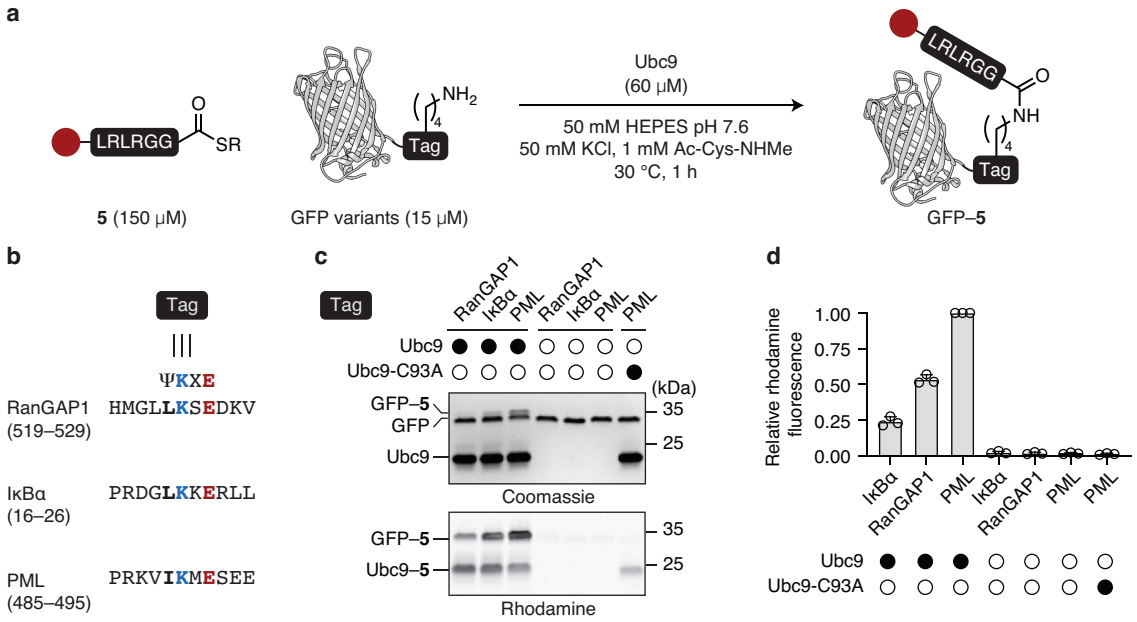


Figure 30. Identification of a LACE tag. (**a**) Labeling of GFP variants containing C-terminal peptide tags with thioester **5**. (The protein image is based on PDB entry 4lqt.) (**b**) Sequences of the tested peptide tags. The consensus SUMOylation motif is highlighted. (**c**) Coomassie-stained SDS–PAGE (top) and in-gel fluorescence (bottom) analysis of reactions with the indicated GFP variants. Control reactions were performed with the active site mutant Ubc9-C93A or without Ubc9. (**d**) Quantification of GFP–**5** formation relative to GFP-PML by densitometry of rhodamine fluorescence (bottom). Individual data points are shown, error bars represent s.d.

To further optimize the acceptor tag, we investigated an extended SUMOylation motif, consisting of several acidic residues downstream of the core motif, which has been reported to increase the SUMOylation specificity and efficiency by interacting with a basic patch in Ubc9 (Figure 31a).[201]

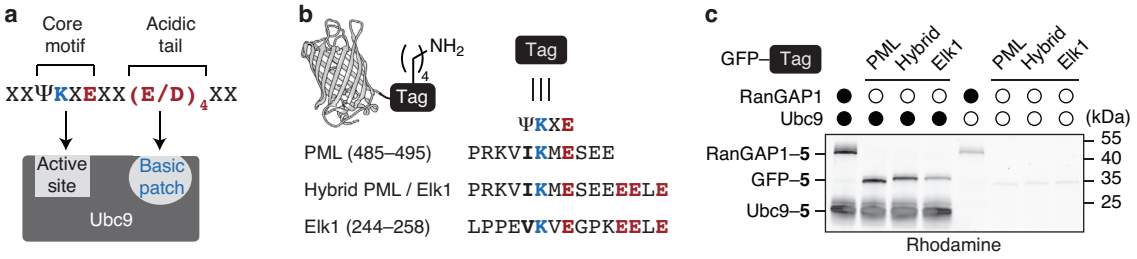


Figure 31. Extended SUMOylation motif. (**a**) Interactions of Ubc9 with the core motif and the acidic tail of the extended motif, adapted from Yang *et al.*[201] (**b**) Sequences of the tested extended peptide tags. The consensus SUMOylation motif and the extension (EELE) are highlighted. (**c**) In-gel fluorescence analysis of reactions with the indicated GFP variants or RanGAP1 (1 μM) after 2 h in the presence or absence of Ubc9 (50 μM) and peptide thioester **5** (75 μM).

To test this possibility, we prepared a GFP variant with a 15-residue long C-terminal tag taken from Elk1 (residues 244–258) (Figure 31b) for which the extended SUMOylation motif has been shown to have an effect in canonical SUMOylation.[201] Additionally, a hybrid tag between PML and the C-terminal extension of Elk1 was tested. However, no increase in labeling efficiency was observed with the hybrid tag, and GFP-Elk1 was less reactive than GFP-PML (Figure 31c). We therefore chose PML (from here on called 'LACE tag') and its GFP construct GFP-PML (from here on called GFP-LACE$_C$ for 'C-terminal') for further studies.

3.2. Identification of useful reaction conditions

With a proficient LACE tag in hand, we wanted to establish practical reaction conditions for complete and specific isopeptide labeling. We first investigated the stability of Ubc9 and its ability to refold spontaneously from a denatured state (Figure 32).

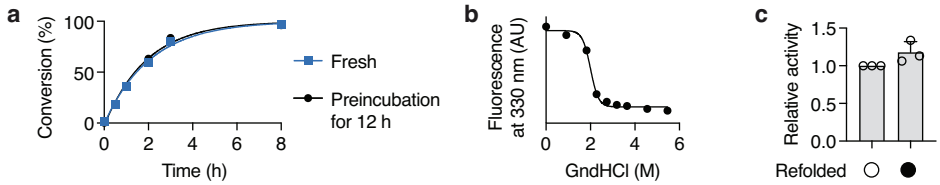


Figure 32. Ubc9 stability and refolding. (**a**) Conversion over time of GFP-LACE$_C$ labeling reactions (15 μM) with thioester **5** (150 μM) using fresh Ubc9 or Ubc9 which has been preincubated at 30 °C for 12 h (60 μM), as determined by densitometry of in-gel rhodamine fluorescence (GFP–**5**). (**b**) Chemical denaturation curve of Ubc9 in the presence of GndHCl as determined by tryptophan fluorescence at 330 nm. (**c**) Relative activity of refolded Ubc9 (10 μM) (filled circle) compared to a fresh recombinant aliquot (hollow circle) during RanGAP1 labeling (1 μM) reactions in the presence thioester **2a** (100 μM). Individual data points are shown, error bars represent s.d.

To test whether Ubc9 can remain active over the course of long reaction times, we compared the activity of a Ubc9 batch which has been preincubated at 30 °C for 12 hours prior to the addition of GFP-LACE$_C$ and thioester **5** to a freshly thawed aliquot of Ubc9 (Figure 32a). The two batches showed indistinguishable reactivity, indicating that Ubc9 can readily be used for

reactions over several hours without loss of activity. Next, we determined the guanidine hydrochloride (GndHCl) concentration required for unfolding of Ubc9, which was found to be 2.0 M (Figure 32b). Refolding after chemical denaturation occurred spontaneously during dialysis (see Experimental Part for details) and the activity was identical to a fresh recombinant stock of enzyme (Figure 32c). Together, these results indicate that prolonged reaction times are possible with Ubc9 and that the enzyme exhibits a thermodynamically stable tertiary structure.

Next, we varied the reactant concentrations and performed time course studies with GFP-LACE$_C$ and thioester **5** to determine useful reaction conditions that would provide complete and specific labeling in a given time (Figure 33).

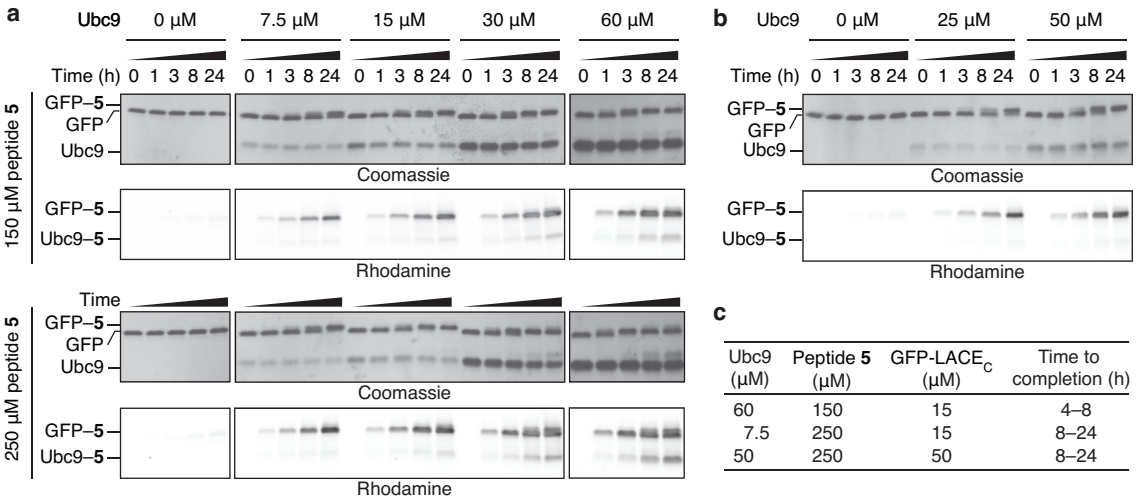


Figure 33. Reaction optimization. (**a**,**b**) SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of GFP-LACE$_C$ labeling reactions (15 μM) with either 150 or 250 μM thioester **5** (**a**), or of GFP-LACE$_C$ labeling reactions (50 μM) with 250 μM thioester **5** (**b**). The Ubc9 concentration used is indicated for each time course. (**c**) Summary of useful LACE conditions for complete and specific labeling of GFP-LACE$_C$ with thioester **5** within 8–24 h.

Specific and complete modification of GFP-LACE$_C$ was achieved by incubating 15 μM GFP-LACE$_C$ in the presence of 60 μM Ubc9 and 150 μM peptide thioester **5** for 4–8 hours (standard conditions unless otherwise mentioned) (Figure 33a, top). Full conversion could also be achieved with sub-stoichiometric amounts of Ubc9, for example by reacting 15 μM GFP-LACE$_C$ with 7.5 μM Ubc9 and 250 μM thioester **5** for 8–24 hours (Figure 33a, bottom). Lastly, higher-concentrated conditions were found which proceeded to full conversion, for instance by reacting 50 μM GFP-LACE$_C$ in the presence of 50 μM Ubc9 and 250 μM thioester **5** for 8–24 hours (Figure 33b). The three described reaction conditions are summarized in Figure 33c.

3.3. Specificity and minimal sequence requirements of the LACE tag

We were pleased to find that full conversion (>95% by gel densitometry) was achieved when reacting 15 μM of GFP-LACE$_C$ with 60 μM Ubc9 and 150 μM thioester **5** at pH 7.6 and 30 °C (standard conditions) for 4–8 hours (Figure 34a,b). Under these conditions, the reaction proceeded with an apparent first-order reaction profile and a half-life of approximately 1.3 hours (Figure 34b). The reaction was strictly dependent on Ubc9 and its active site residue C93, as well as the acceptor lysine K6 of the tag (Figure 34c).

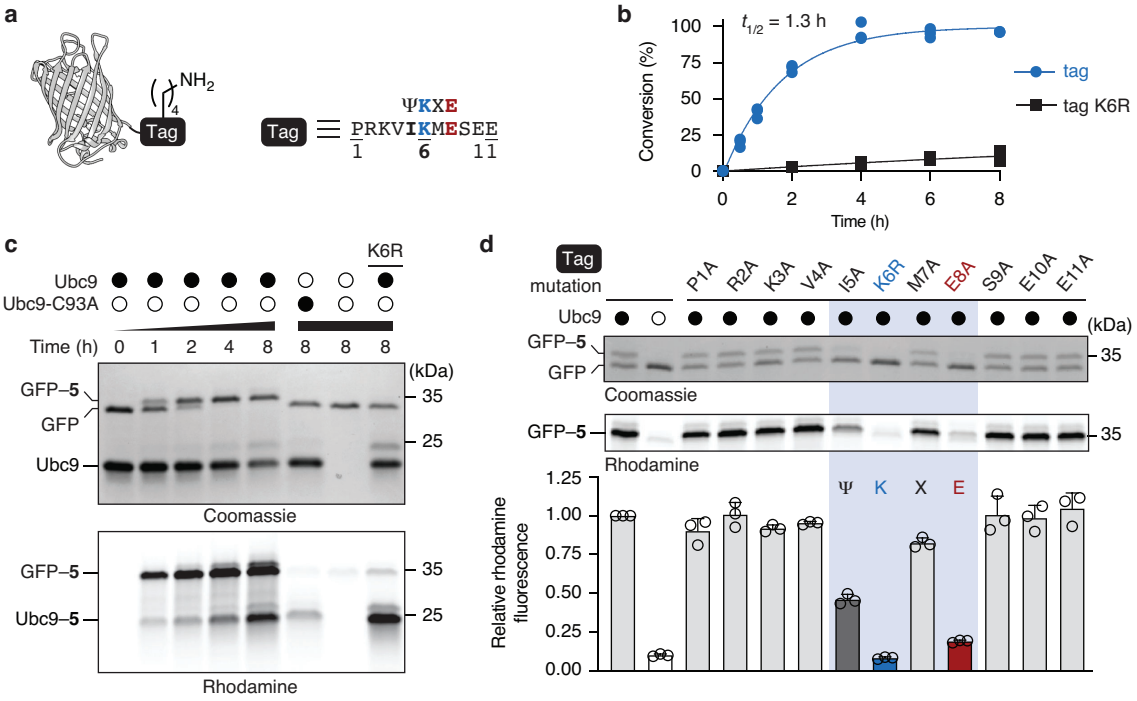


Figure 34. Specificity and minimal sequence requirements of the LACE tag. (**a**) GFP-LACE$_C$ with the LACE tag sequence. (**b**) Conversion over time of GFP-LACE$_C$ labeling reactions with thioester **5** using the competent substrate or the acceptor lysine mutant K6R, as determined by densitometry of in-gel rhodamine fluorescence (GFP–**5**). The apparent reaction half-life ($t_{1/2}$) is given. Individual data points are shown ($n = 3$). (**c**) Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of reactions at indicated times. Control reactions with the active site mutant Ubc9-C93A or without Ubc9 were performed. (**d**) Mutational analysis of the LACE tag. Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (middle) of labeling reactions of GFP-LACE$_C$ variants with the indicated point mutations and thioester **5**. For comparison, reactions were quenched after 1 h before complete labeling was achieved. Quantification of GFP–**5** formation relative to the unmutated LACE tag by densitometry of rhodamine fluorescence (bottom). Individual data points are shown, error bars represent s.d.

Mutational analysis of the tag confirmed that the consensus motif was crucial for efficient labeling (Figure 34d). The reactivity was reduced when I5 (hydrophobic residue) or E8 (acidic residue) was mutated to alanine. All other amino acids could be individually mutated to alanine without loss of activity, indicating the importance of the core SUMOylation motif within the LACE tag. This suggests that a minimal LACE tag IKXE should be sufficient (see Section 3.5).

3.4. Tag positioning throughout the protein sequence and dual-labeling

Isopeptide labeling offers the possibility of positioning the tag throughout the protein sequence. Apart from the C-terminus, we also incorporated the LACE tag into GFP internally (GFP-LACE$_I$) or fused it to the N-terminus (GFP-LACE$_N$) (Figure 35). Modification of GFP-LACE$_I$ proceeded with similar efficiency as that achieved with GFP-LACE$_C$, whereas the reaction half-life approximately doubled with GFP-LACE$_N$.

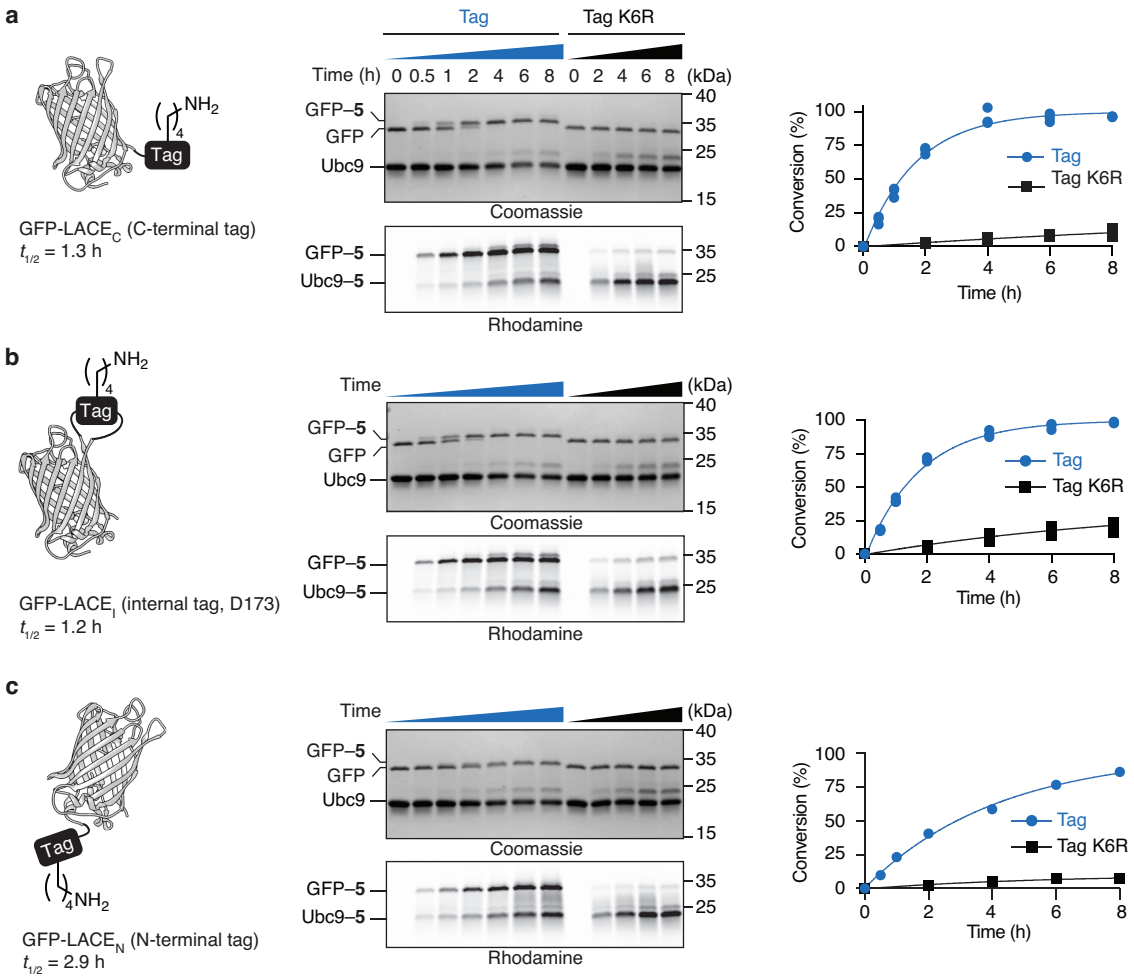


Figure 35. LACE tag location screen within the GFP scaffold. (**a**–**c**) SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of GFP-LACE$_C$ (**a**), GFP-LACE$_I$ (**b**) and GFP-LACE$_N$ (**c**) (15 μM) labeling reactions with thioester **5** (150 μM) and Ubc9 (60 μM) at indicated time points. Reactions were carried out with the competent substrates or with the acceptor lysine mutant K6R. Quantification of GFP–**5** formation over time as determined by densitometry of rhodamine fluorescence is shown for each variant, as well as the apparent $t_{1/2}$. Individual data points are shown (n = 3 for GFP-LACE$_C$ and GFP-LACE$_I$, n = 1 for GFP-LACE$_N$).

Compared with strategies to modify protein N- or C-termini, isopeptide labeling directly enables conjugation at multiple sites. To test multiple labeling of a protein, we prepared a GFP variant containing two LACE tags, with a LACE tag placed internally and a tag^{2M} fused to the C-terminus (GFP-LACE$_{I,C}^{2M}$) (Figure 36a). Tag^{2M} contained two mutations relative to the PML-

derived tag variant. On the one hand, the mutation K3A was introduced because this change did not negatively affect the reactivity as seen from the alanine screen (Figure 34d), removing any ambiguity as to the exact modification site. On the other hand, M7Q was introduced because glutamine is more frequently found at the variable position X in proteomics analysis of SUMOylated proteins.[153] Near complete conversion to the doubly modified product was observed within 8 hours using the same standard conditions as those employed for single-tagged GFP variants (Figure 36b,c).

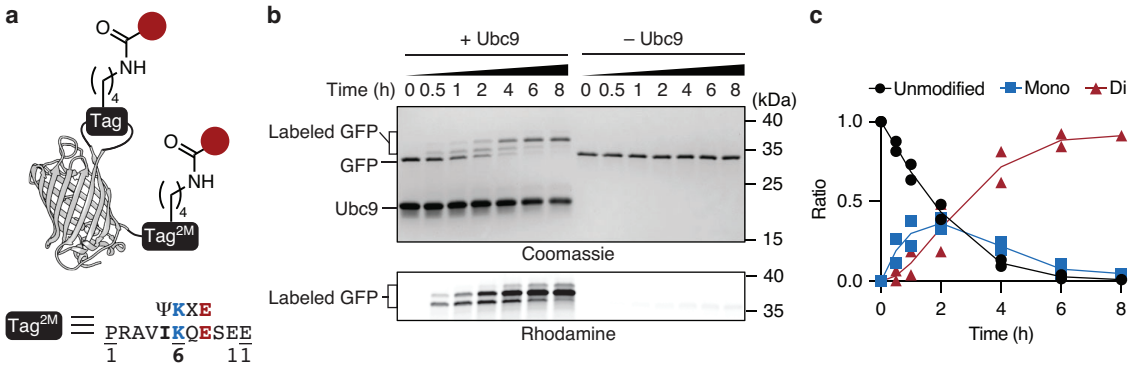


Figure 36. Dual-labeling of GFP-LACE$_{I,C}^{2M}$. (**a**) A slightly modified tag^{2M} was used at the C-terminus which contains two mutations compared to the PML-derived LACE tag. (**b**) Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of GFP-LACE$_{I,C}^{2M}$ labeling reactions with thioester **5** in the presence or absence of Ubc9. (**c**) Ratio of products over time as determined by densitometry of Coomassie-stained bands. Individual data points are shown.

3.5. Minimal LACE tags

The mutational analysis (Figure 34d) indicates that all residues outside of the core SUMOylation motif IKXE are dispensable. To test if the minimal sequence IKXE would be sufficient for recognition and modification by Ubc9, we tested labeling of GFP substrates with a truncated minimal LACE tag IKQE, either at the C-terminus (GFP-LACE$_C^{minimal}$) or internally (GFP-LACE$_I^{minimal}$) (Figure 37). Full conversion was achieved in both cases with reaction half-lives of approximately 3 hours.

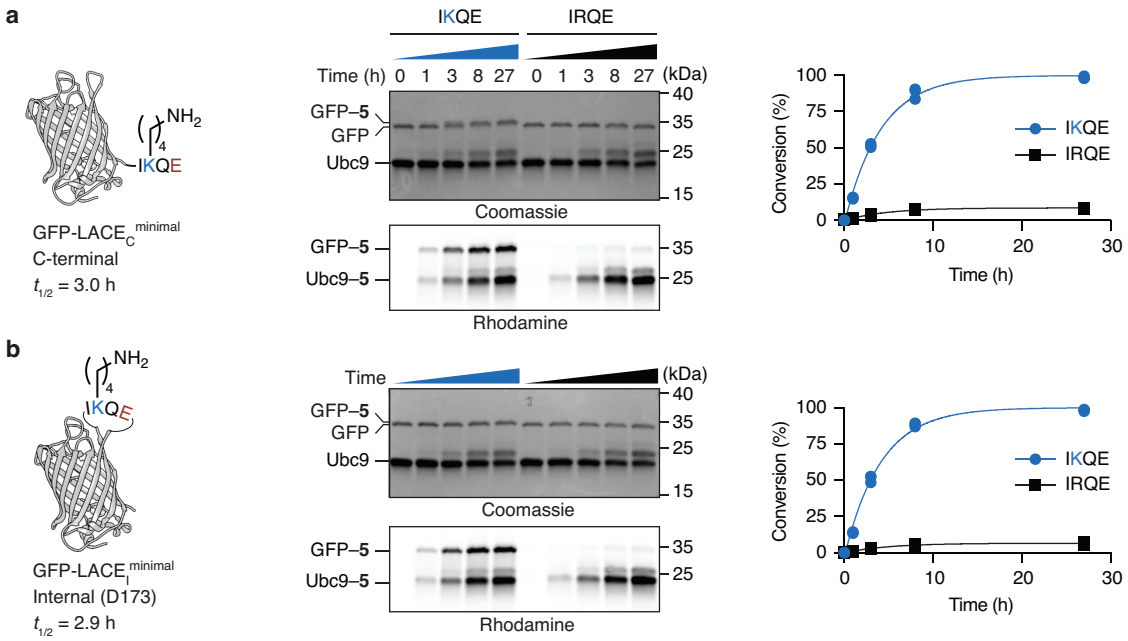


Figure 37. Labeling of GFP variants with a minimal LACE tag. (**a**,**b**) SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of GFP labeling reactions with a minimal LACE tag (IKQE) at the C-terminus (GFP-LACE$_C^{minimal}$) (**a**) or internally (GFP-LACE$_I^{minimal}$) (**b**) with thioester **5** at indicated time points. Reactions were carried out with the competent substrates or with the acceptor lysine mutant (IRQE). Quantification of GFP–**5** formation over time as determined by densitometry of rhodamine fluorescence is shown for each variant, as well as the apparent $t_{1/2}$. Individual data points are shown (n = 3).

3.6. Generality of the LACE tag

To determine whether the LACE tag would function in protein contexts other than GFP, we prepared a variant of the IG27 domain of the human muscle protein titin containing a tag^{2M} (titin-LACE$_C^{2M}$) and T4 lysozyme (T4L) with a LACE tag (T4L-LACE$_C$), both fused to the C-terminus. Modification of these substrates proceeded with excellent specificity and conversion, with a reaction half-life of approximately 1.3 hours for titin-LACE$_C^{2M}$ and 3.3 hours in the case of T4L-LACE$_C$ (Figure 38).

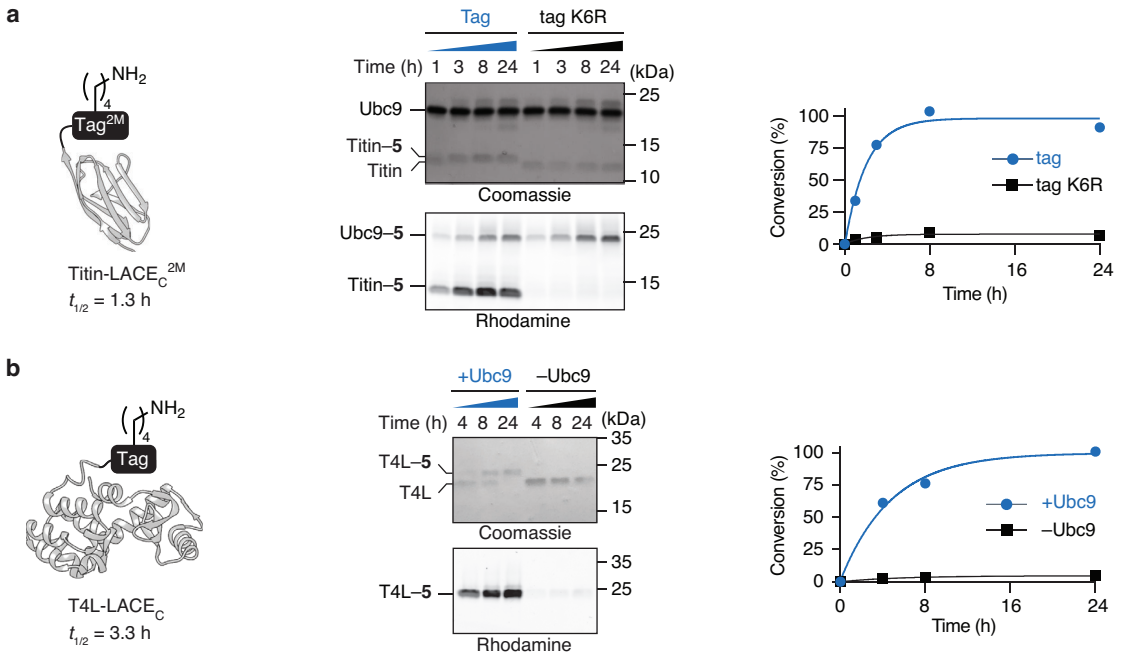


Figure 38. Labeling of different protein substrates. (**a,b**) SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of titin IG27 labeling reactions with a tag^{2M} at the C-terminus (titin-LACE$_C^{2M}$) (**a**) and of T4 lysozyme (T4L) with a C-terminal LACE tag (T4L-LACE$_C$) (**b**) with thioester **5** at indicated time points. Reactions were carried out with the competent substrate or with the acceptor lysine mutant K6R (**a**), or in the presence and absence of Ubc9 (**b**). Quantification of GFP–**5** formation over time as determined by densitometry of rhodamine fluorescence is shown for each variant, as well as the apparent $t_{1/2}$. (**b**) Because of band overlap between T4L-LACE$_C$ and Ubc9, His$_6$-tagged Ubc9 was removed by reverse nickel nitrilotriacetic acid (Ni-NTA) purification prior to analysis. (The protein images are based on PDB entries 1tit and 1lyd.)

4. Mechanistic studies

4.1. Self-labeling of Ubc9 and thioester hydrolysis

During development of the method, it became apparent that Ubc9 underwent self-labeling with peptide thioesters over the course of the reaction. Given that the Ubc9–peptide side product was stable under reducing and denaturing conditions, a covalent isopeptide bond is likely formed. A possible route for this self-labeling could be spontaneous, non-catalyzed aminolysis of the thioesters. Non-enzymatic protein acylation, for example in the presence of cellular levels of acetyl or acyl-coenzyme A (CoA), has been described.[202–204] This pathway is increasingly appreciated to play a signaling role, for example under carbon stress.[205] Non-enzymatic acylation has been shown to be especially important when a surface-exposed cysteine undergoes transthioesterification, followed by acyl transfer to a proximal lysine residue.[206–208] A second pathway could be Ubc9-mediated self-labeling.[209,210] A third possibility involves interception of the Ubc9~thioester intermediate by a nearby nucleophilic active site residue. Although the latter has not been described for Ubc9, approximately 25% of all E2 enzymes undergo such an autoinhibitory modification in which a lysine residue, located at a conserved +5 position relative to the active site cysteine, intercepts the thioester intermediate of the catalytic center.[211]

We incubated Ubc9 with thioester **5** in the absence of a substrate to induce self-labeling, and analyzed the products by MS/MS (Figure 39). All lysine and cysteine residues that were covered by identifiable tryptic peptides were associated with at least one modified tryptic peptide, with the exception of K146. This result indicates that no single residue is responsible for the self-labeling process, and that Ubc9–**5** is likely heterogeneous.

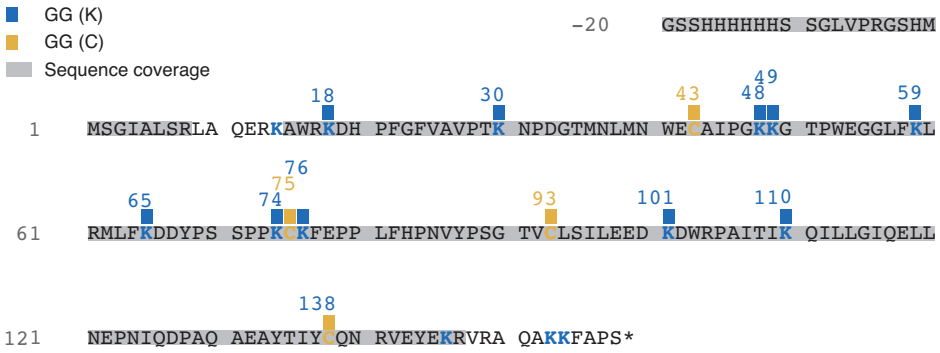


Figure 39. MS/MS analysis of Ubc9–**5**. Glycylglycine modifications on lysines (blue) and cysteines (yellow) were observed and mapped to the covered sequence. Full MS/MS results with the observed tryptic peptides are shown in Figure 101 of the Appendix.

Because the MS/MS analysis cannot be used to quantify the relative labeling amount at the observed positions directly, we tested two residues in a hypothesis-driven approach (Figure 40a). On the one hand, Ubc9 contains a second surface exposed cysteine C138 besides the active site residue C93, which might catalyze acylation of proximal lysine residues. On the other hand, human Ubc9 has been shown to undergo autoSUMOylation at K14.[209,210] No tryptic peptide fragments covering this residue could be observed in the MS/MS analysis (Figure 39). To test the contribution of these sites, Ubc9-C138A and Ubc9-K14R variants were prepared and analyzed for their reactivity in GFP-LACE$_C$ labeling reaction as well as Ubc9 self-labeling.

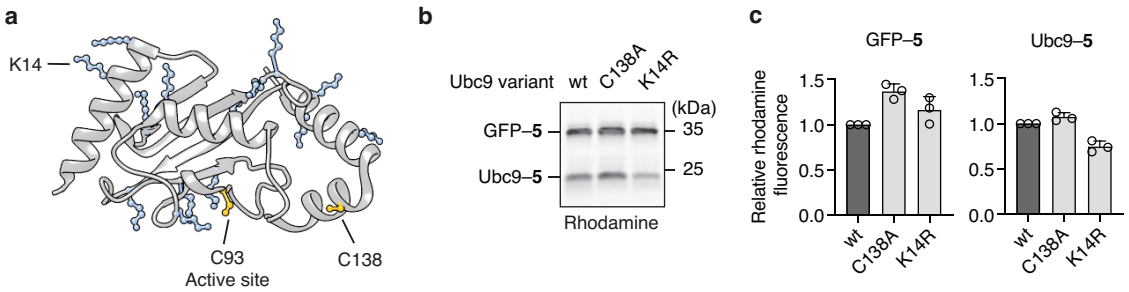


Figure 40. Self-labeling of Ubc9. (The protein image is based on PDB entry 5d2m.) (**a**) Lysines (blue) and the two surface-exposed cysteines (C93 active site and C138, yellow) are highlighted in the Ubc9 structure. K14 in helix a1 has been shown to be SUMOylated[209,210] and is annotated. (**b**) In-gel fluorescence analysis after 3 h of GFP-LACE$_C$ (2 μM) labeling reaction in the presence of thioester **5** (50 μM) and the indicated Ubc9 variants (25 μM). (**c**) Quantification of GFP–**5** and Ubc9–**5** formation relative to wt Ubc9 by densitometry of rhodamine fluorescence. Individual data points are shown, error bars represent s.d.

Both variants displayed comparable activity to wild type Ubc9 in GFP-LACE$_C$ labeling reactions (Figure 40b,c). Looking at self-labeling, Ubc9-K14R showed a 25% reduction, indicating that this site is responsible for a significant portion of this side-reaction, whereas self-labeling of Ubc9-C138A was indistinguishable from the wild type. This implies that the *S*-to-*N* relay pathway is not effective with C138, presumably because of the lack of proximal lysine residues (Figure 40a).

Because no single residue was identified as a major site of self-labeling, and because the self-labeling did not interfere with complete conversion and purification of the product, we opted to use wild type Ubc9 for standard LACE reactions. The ease of removal is exemplified by the T4L-LACE$_C$ labeling analysis, in which the self-labeled side-product was removed from the reaction mixture together with His$_6$-Ubc9 by reverse Ni-NTA purification (Figure 38b).

The C-terminal peptide thioesters employed are also susceptible to hydrolysis, which proceeds with a zero-order rate constant of approximately 1.5 x 10^{-8} M s^{-1} in reaction buffer as determined by liquid chromatography MS (LC–MS) analysis. This rate translates to approximately 4% hydrolysis per hour when starting with 150 μM thioester. Because of Ubc9 self-labeling and thioester hydrolysis, an excess of thioester should be employed if complete conversion of the substrate is desired. This requirement is reflected in all our optimized conditions for complete labeling of proteins by LACE (Figure 33c).

4.2. Influence of buffer composition on LACE

To inform labeling of protein substrates with specific buffer requirements, we investigated the influence of several commonly used buffer additives on the LACE process (Figure 41).

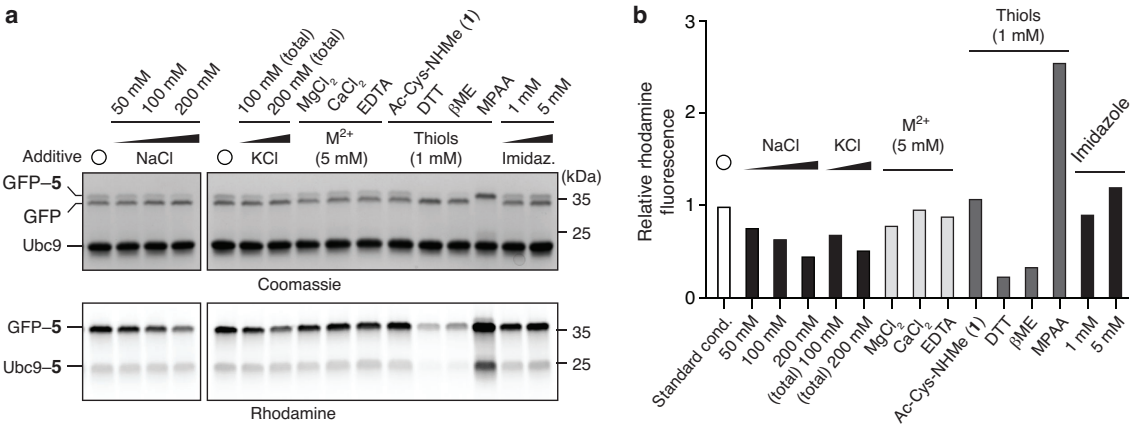


Figure 41. Buffer additive screen. (**a**) Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of GFP-LACE$_C$ labeling reactions with thioester **5** in standard conditions (hollow circle) or with indicated additives after 1 h at 25 °C. Standard conditions: 50 mM HEPES pH 7.6, 50 mM KCl. (**b**) Quantification of GFP–**5** formation relative to standard conditions by densitometry of rhodamine fluorescence.

Increasing concentration of NaCl or KCl in the reaction buffer resulted in decreasing labeling efficiency, with a reduction of up to 50% in the presence of 200 mM NaCl. Commonly used divalent ions such as $MgCl_2$ and $CaCl_2$ at 5 mM concentration did not have a pronounced effect on the reactivity, and sequestration of divalent ions by 5 mM ethylenediaminetetraacetic acid (EDTA) demonstrated that divalent ions are not required.

Looking at thiol additives, we found that 1,4-dithiothreitol (DTT) and 2-mercaptoethanol (βME) were not compatible with LACE when used at 1 mM concentration. In contrast, the cysteine derivative Ac-Cys-NHMe (**1**), which we also used for activation of the peptide thioesters such as thioesters **2a** and **3**–**13**, was well tolerated. Strikingly, addition of the aromatic thiol 4-mercaptophenylacetic acid (MPAA) at 1 mM concentration resulted in clean and complete conversion after 1 hour reaction time, greatly accelerating the reaction compared to the standard conditions.

Lastly, we tested addition of imidazole at low millimolar concentration, which may be present as a contaminant in protein samples that were isolated by Ni-NTA affinity purification. No detrimental effect on labeling was observed, and even led to a small increase in labeling at the higher concentration (5 mM).

Taken together, LACE appears to have a broad tolerance with respect to salt additives, without a strict requirement for a particular ion, but operates best in low ionic strength buffers. This is in contrast to canonical SUMOylation or ubiquitination, which require ATP-Mg^{2+}-dependent activation of Ubls by E1 enzymes.[212]

The incompatibility of LACE with the commonly used reducing agents DTT and βME is notable. The mechanism by which these additives hamper labeling was not further investigated. One possibility could be that these thiols undergo transthioesterification with the Ac-Cys-NHMe-activated peptide thioesters. The exchanged thioesters may be unreactive as acyl donors of Ubc9. Alternatively, these thiols might favor the reverse reaction of E2~Ubl thioester formation. Lastly, β-hydroxy-functionalized thiols such as βME and DTT have been shown to cause formal hydrolysis of thioesters via intramolecular displacement of the carboxylate (Figure 42),[213] which may result in non-productive consumption of the activated acyl donor in the reaction mixture.

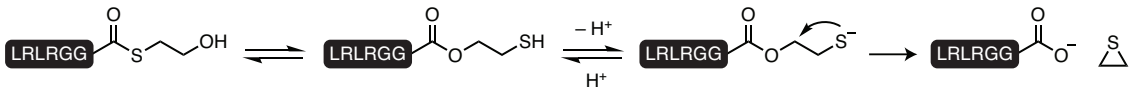


Figure 42. Formal thioester hydrolysis by β-hydroxy-functionalized thiols, adapted from Gates *et al.*[213]

4.3. Nucleophilic catalysis

Following up on the findings from the buffer additive, we sought to further investigate the increase in reactivity in the presence of MPAA (Figure 43). While the increased labeling with MPAA at 1 mM concentration was reproduced, higher concentration of the aromatic thiol (10 mM) abolished the positive effect and resulted in a reversion of the GFP-LACE$_C$ labeling rate similar to the ones observed under standard conditions in the absence of a thiol additive. Additionally, a heterogeneous mixture of self-labeled Ubc9 was observed in the presence of 10 mM MPAA, and the labeling specificity was decreased as seen from a control reaction with the acceptor lysine mutant K6R (Figure 43a).

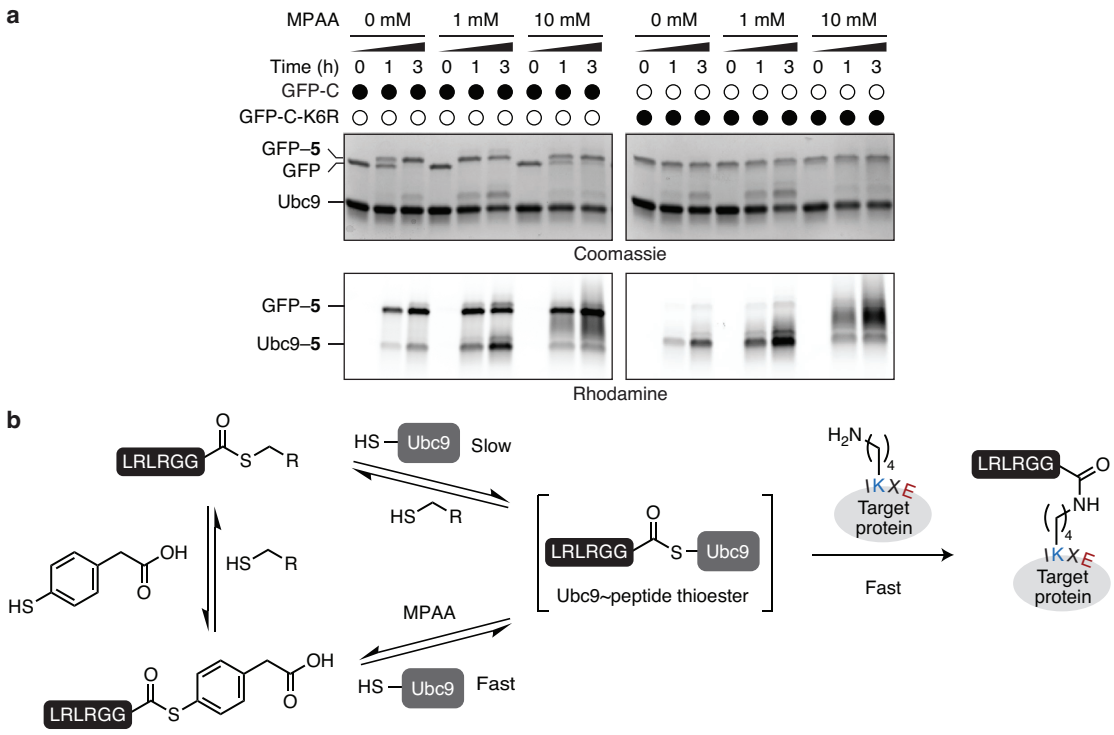


Figure 43. Nucleophilic catalysis. (**a**) Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of GFP-LACE$_C$ labeling reactions with thioester **5** in the presence of indicated amounts of MPAA. Control reactions with the acceptor lysine mutant K6R were performed. (**b**) Proposed model for MPAA-catalyzed LACE reaction.

The increase in reactivity with moderate amounts of MPAA can be rationalized by a two-step model of the LACE reaction, which, under standard conditions, involves slow charging of Ubc9 with the alkyl thioester followed by fast transfer of the peptide to the acceptor lysine of the target protein (Figure 43b). MPAA can act as a nucleophilic catalyst by undergoing a thiol exchange with the alkyl thioester, which results in the *in situ* formation of an aromatic thioester. This high-energy intermediate results in increased formation of the Ubc9~thioester intermediate, thus accelerating the rate-limiting step. An analogous mechanism has been well appreciated to operate in NCL.[60] Although less pronounced, the slight increase in reactivity in the presence of

imidazole observed during the additive screen (Figure 41) may be due to a similar mechanism and involve the formation of an acyl imidazolium intermediate.[214] High concentrations of MPAA likely negate the positive effect on the reaction rate either by favoring the reverse reaction of Ubc9~peptide thioester formation, or by causing increased hydrolysis of the activated acyl donor via the aromatic thioester intermediate.

4.4. Reaction profile

We wanted to verify the proposed two-step reaction profile, involving slow formation of the Ubc9~peptide thioester formation followed by fast transfer of the peptide to the target protein. The enhanced efficiency of ubiquitin-derived thioesters relative to SUMO derivatives also prompted us to compare their reaction profiles (Figure 44).

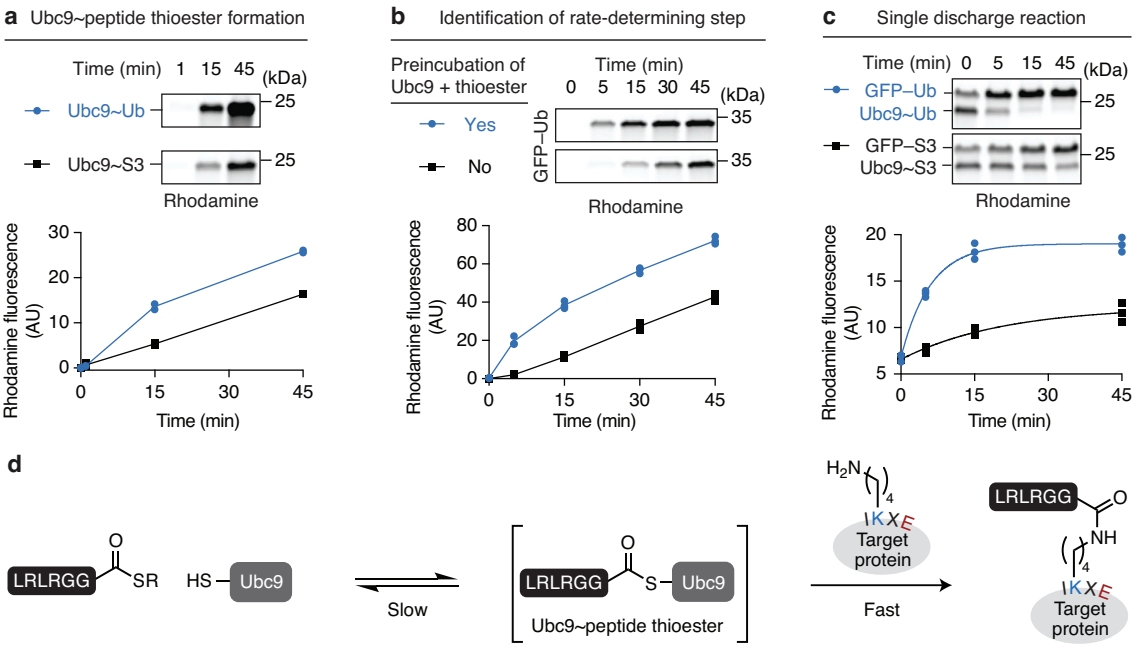


Figure 44. Reaction profile of LACE. (**a**) Non-reducing SDS–PAGE and in-gel fluorescence analysis (top) of Ubc9 thioester formation with peptide thioesters **5** (Ub) or **2a** (S3) at indicated times. Quantification of Ubc9~peptide formation by densitometry of rhodamine fluorescence (bottom). (**b**) SDS–PAGE and in-gel fluorescence analysis (top) of GFP-LACE$_C$ labeling reactions with thioester **5** (Ub) at indicated times, with or without preincubation of Ubc9 and thioester before addition of GFP-LACE$_C$. Quantification of GFP–**5** product formation (GFP–Ub) by densitometry of rhodamine fluorescence (bottom). (**c**) SDS–PAGE and in-gel fluorescence analysis (top) of single-discharge reactions with preformed Ubc9~peptide thioester, with either Ub (**5**) or S3 (**2a**) sequence, and GFP-LACE$_C$ as the acceptor at indicated times. Quantification of GFP–peptide formation by densitometry of rhodamine fluorescence (bottom). (**a**–**c**) Individual data points are shown (n = 2 in **a**, n = 3 in **b** and **c**). The Ubc9 variant C138A was used for mechanistic studies to exclude reaction with the surface-exposed non-catalytic C138. (**d**) Proposed reaction profile of LACE.

Looking at the individual reaction steps, the first event – Ubc9~peptide thioester formation – proceeded faster with the ubiquitin-derived thioester **5** than with SUMO2/3-derived thioester **2a** (Figure 44a). From preincubation experiments, we determined that the first step was indeed the

rate-limiting step of the overall LACE reaction under the standard conditions. Specifically, preincubation of Ubc9 with peptide thioester **5** resulted in an initial burst of product formation after addition of the substrate. This is in contrast to a standard reaction, which appears to exhibit a slight lag phase during the first 15 minutes before reaching a steady reaction rate (Figure 44b). Interestingly, the second step – transfer of the peptide from Ubc9 to the acceptor lysine of the target protein – was also faster with the ubiquitin-derived thioester compared to the SUMO derivative as determined from single-discharge assays with preformed Ubc9~peptide thioester intermediate (Figure 44c). The determined reaction profile (Figure 44d) is in agreement with the nucleophilic catalysis model with MPAA.

4.5. Assessing the uniqueness of ubiquitin-derived sequences as acyl donors for Ubc9

With the finding that Ubc9 both loads and transfers the ubiquitin-derived thioester **5** faster than the SUMO2/3-derived variant **2a** (Figure 44), we wanted to compare the reactivity of the ubiquitin-derived thioester to additional peptides, beyond simply further Ubl variants.

Several peptides have been discovered from a phage-displayed library that could be activated by the SUMO activating enzyme Sae1/Sae2, of which some were shown to be passed down to Ubc9 and in part to RanGAP1.[215] We prepared the respective rhodamine-functionalized peptide thioesters **14**–**17** and compared them to thioesters **5** (ubiquitin-derived) and **2a** (SUMO-derived) in a Ubc9-mediated LACE reaction with GFP-LACE$_C$. Analogously, thioester **18**, which was prepared with a sequence selected for activation by the ubiquitin activating enzyme Ube1,[216] was tested in the same assay (Figure 45). However, none of the tested thioesters were transferred by Ubc9 to the same degree as thioester **5**, and most sequences conferred similar reactivity as the SUMO2/3-derived thioester **2a**.

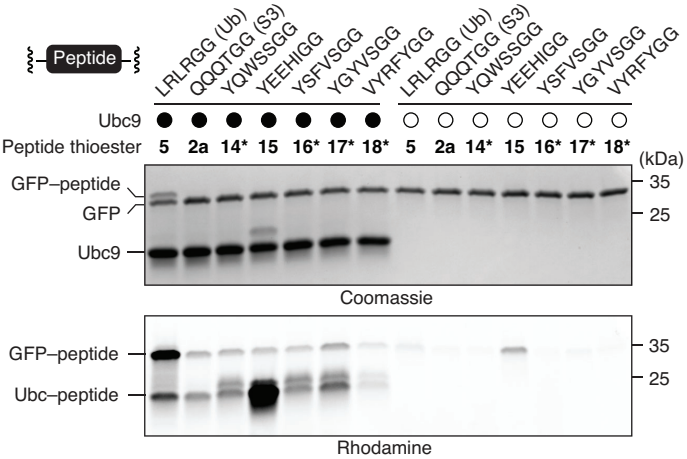


Figure 45. Comparison of the ubiquitin-derived acyl donor to reported sequences. Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of GFP-LACE$_C$ labeling reactions in the presence of the indicated rhodamine-functionalized peptide thioesters **5** (Ub), **2a** (S3), and **14**–**18**. Peptides marked with an asterisk were not fully soluble at the concentration used (150 μM). For comparison, reactions were quenched after 45 min before complete labeling was achieved.

Thioester **15** stands out for the high level of observed Ubc9 self-labeling and a high background reaction with GFP-LACE$_C$ that is independent from Ubc9. The reason for this high non-specific reactivity was not investigated. One possibility might be the presence of a histidine residue close to the C-terminal thioester. The histidine residue might undergo an intramolecular or intermolecular nucleophilic attack of the thioester, resulting in the formation of a more reactive acyl imidazolium intermediate.

The general lack of reactivity of thioesters **14–18** may reflect their intended use as an inhibitor of the Ubl cascade. Indeed, they were selected mainly for their activation by and ability to bind to E1 activating enzymes, rather than their ability to be passed on in the Ubl cascade.[215,216] In summary, the ubiquitin-derived thioester, besides outperforming all tested Ubl variants as a substrate for Ubc9, is therefore also a better acyl donor compared to engineered peptides which have been previously reported to be processed by the SUMO and ubiquitin pathways.

5. Structural analysis by X-ray crystallography

5.1. Preparation of X-ray substrate

We pursued structural studies to investigate whether the ubiquitin-derived peptide undergoes specific, favorable interactions with Ubc9 that could explain the higher reactivity compared to SUMO-derived peptides (Figure 46).

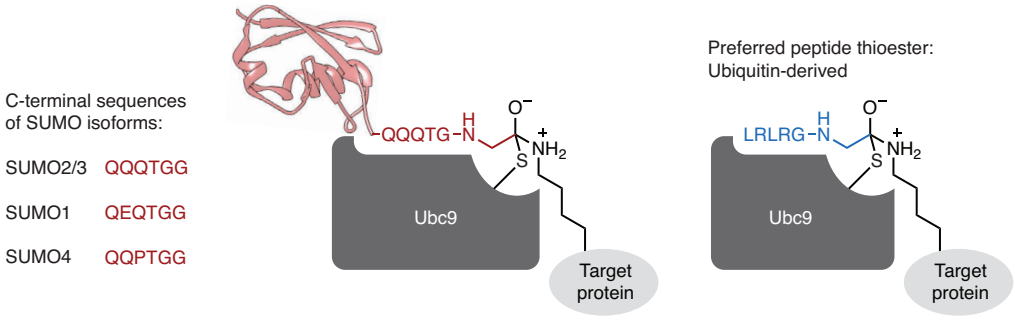
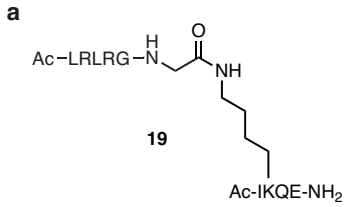


Figure 46. The distinctly different C-terminal ubiquitin sequence is the preferred acyl donor for Ubc9 over canonical sequences derived from SUMO isoforms.

Attempts to co-crystallize ligand **19**, consisting of a consensus SUMOylation motif (IKQE) that is linked to the ubiquitin-derived C-terminal hexapeptide via an isopeptide bond (Figure 47a), with a previously reported triple-mutant X-ray construct of Ubc9 (Ubc9-K48A-K49A-E54A, from here on called Ubc9-3A)[217] only yielded apo crystals, even in the presence of 5–10 equivalents of the ligand (Figure 47b).



b

Ubc9-3A (mg/mL)	Ligand **19** (equiv)	Result
15	10	Apo crystals
25	5	Apo crystals

Figure 47. Attempts at co-crystallization of ligand **19** and Ubc9-3A. (**a**) Noncovalent isopeptide ligand **19**. (**b**) Crystallizations were set up in two concentration regimes and screened against a panel of crystallization precipitants.

Instead, covalent ligands were investigated to prepare the respective complexes with Ubc9. Following similar conditions to those reported for disulfide formation between a ubiquitin variant (ubiquitin-G76C) and the E2 enzyme Ubc1,[218] we attempted disulfide formation between peptide **20** and a variant of the Ubc9 X-ray construct in which the surface-exposed cysteine C138 was mutated to alanine (Ubc9-3A-C138A). This mutation was introduced to favor disulfide formation at the active site cysteine. However, the reaction did not proceed to completion as judged by matrix-assisted laser desorption/ionization mass spectrometry (MALDI–MS) analysis, and precipitation of unmodified or modified Ubc9 was observed (Figure 48a).

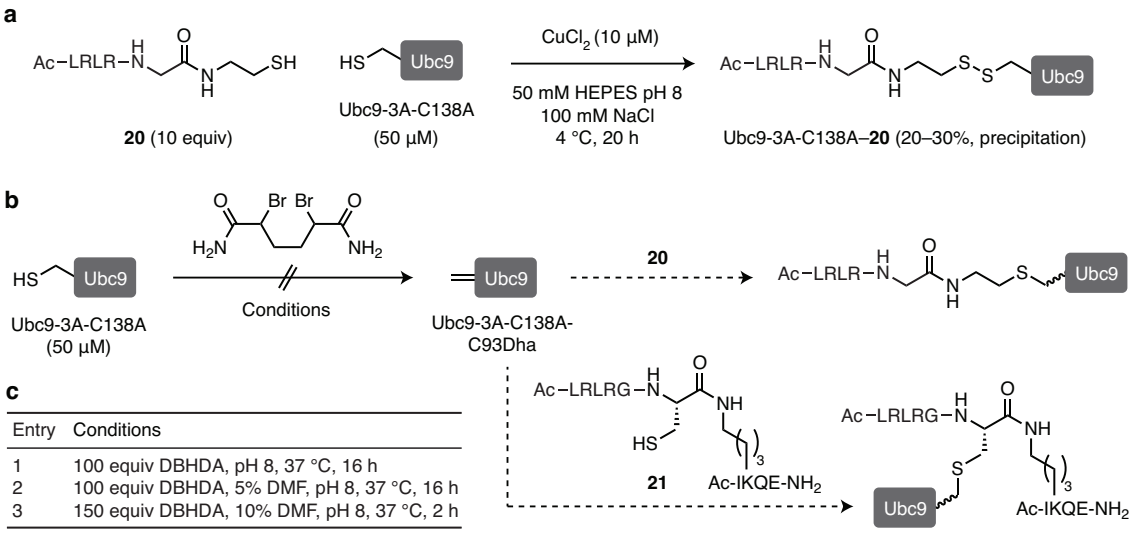


Figure 48. Attempts at covalent complex formation for X-ray analysis. (**a**) Disulfide formation between ligand **20** and Ubc9-3A-C138A, conditions adapted from Merkley *et al.*[218] (**b**) Dha formation at the active site C93 using DBHDA under three different conditions (listed in **c**). The Ubc9-3A-C138A-C93Dha was intended to act as a Michael acceptor for ligands **20** and **21**.

Next, we investigated Dha formation at the active site C93 of Ubc9 using 2,5-dibromohexanediamide (DBHDA),[29] which could act as a Michael acceptor for ligand **20** or the cysteine-derivative of ligand **19** (**21**) (Figure 48b). However, the desired Dha-functionalized Ubc9 (Ubc9-3A-C138A-C93Dha) was not obtained, even in the presence of 100–150 equivalents of DBHDA and after prolonged reaction times (Figure 48b,c). Instead, only remaining starting material and loss thereof was observed as judged by MALDI–MS.

We then tested disulfide formation between ligand **21** and Ubc9-3A-C138A (Figure 49a). Spontaneous and nearly complete covalent complex formation was observed after 48 hours. The complex was purified by cation exchange chromatography and formation of the desired product Ubc9-3A-C138-**21** was confirmed by electrospray ionization mass spectrometry (ESI–MS) (Figure 49b–e).

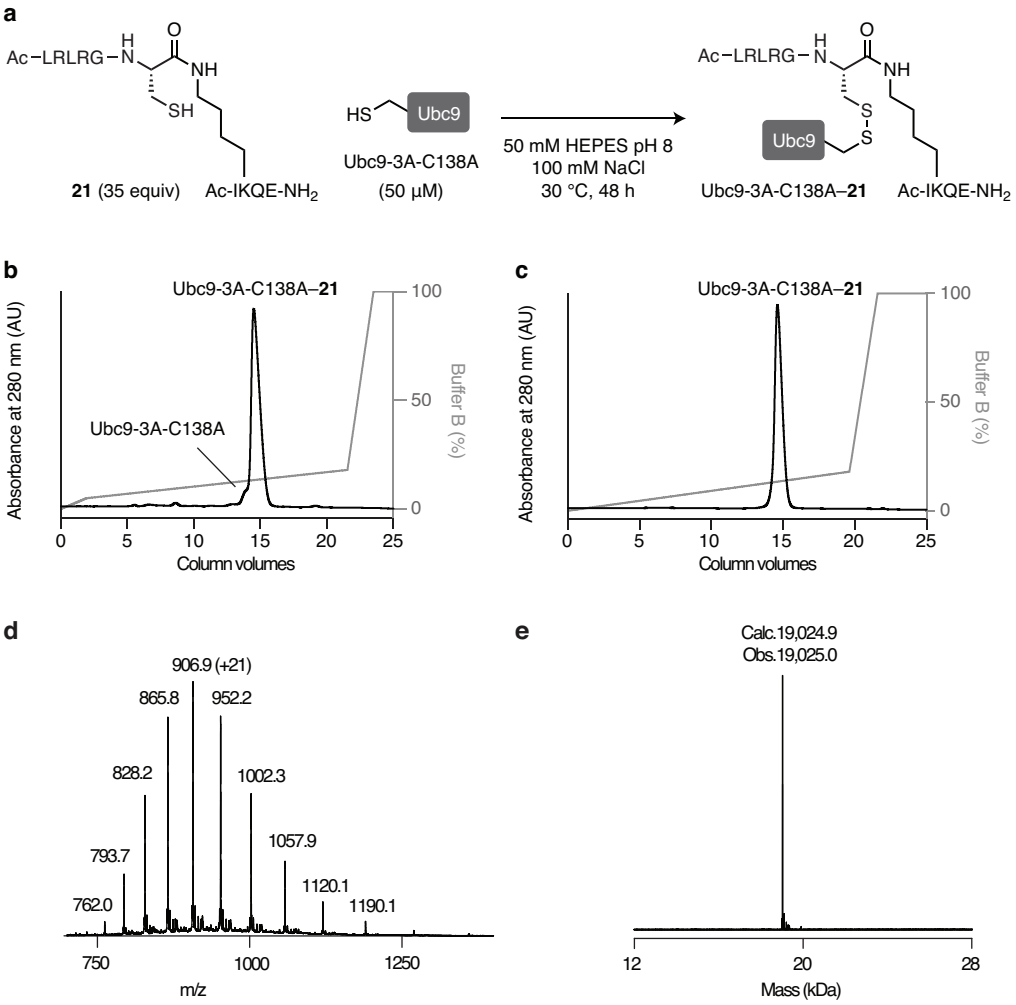


Figure 49. Disulfide-bonded complex formation. (**a**) Spontaneous disulfide formation between the active site C93 of Ubc9-3A-C138A with the isopeptide ligand **21**. (**b**,**c**) Cation exchange chromatograms of the reaction mixture (**b**) and of the purified conjugate (**c**) (see Experimental Part for details). (**d**,**e**) ESI–MS analysis of the purified disulfide complex. Shown are m/z (**d**) and the deconvoluted data (**e**).

5.2. Structural analysis

The disulfide complex Ubc9-3A-C138-**21** was crystallized using the sitting drop vapor diffusion method, and X-ray-quality crystals appeared within two days (Figure 50a,b, see Experimental Part for details). The obtained crystals diffracted to 1.9 Å resolution.

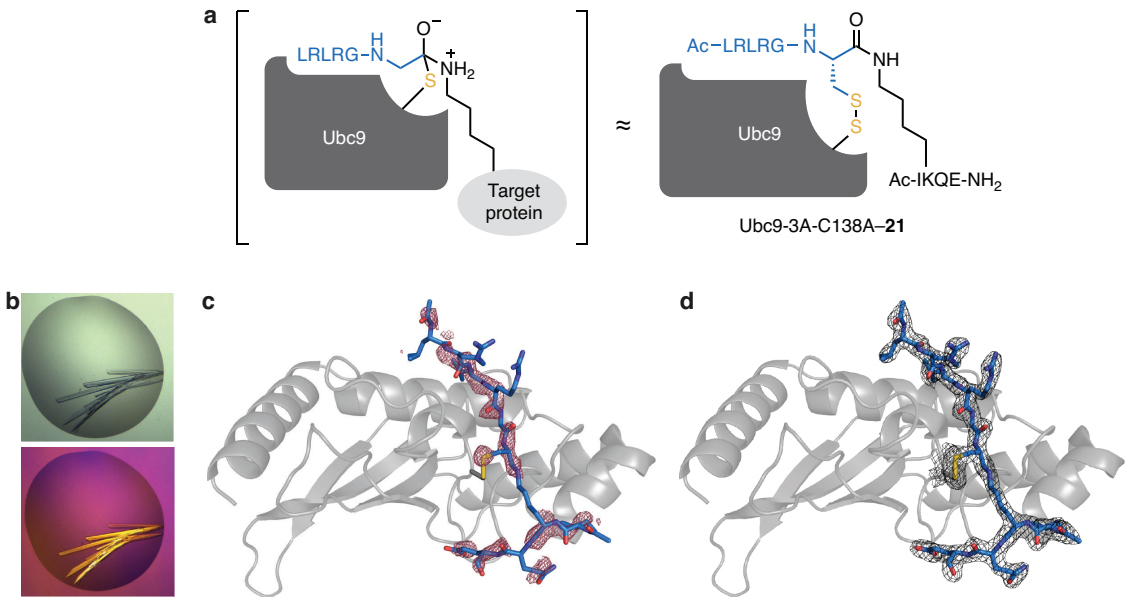


Figure 50. Structure of the disulfide complex. (**a**) Disulfide complex with isopeptide ligand **21** as a structural analog of the tetrahedral intermediate during transfer of ubiquitin-derived peptides to target lysines by Ubc9. (**b**) Brightfield (top) and polarized light (bottom) images of obtained crystals. (**c**,**d**) X-ray structure, with Ubc9 colored in grey and the isopeptide in blue. Shown are the difference electron density Fo–Fc after molecular replacement in red mesh contoured at 2.5 σ (**c**), and the refined electron density 2Fo–Fc of the peptide ligand and the disulfide linkage in black mesh contoured at 1.0 σ (**d**).

Following molecular replacement with a previously published structure of Ubc9-3A (PDB entry 5f6e)[217], features of the isopeptide ligand were clearly visible in the Fo–Fc electron density (Figure 50c). The ligand was modeled and the structure was refined accordingly (Figure 50d and Table 1).

Table 1. Data collection and refinement statistics (molecular replacement)

	Ubc9 with covalent isopeptide ligand (PDB entry 6syf)
Data collection	
Space group	$P2_1$
Cell dimensions	
α, β, γ (Å)	94.7, 38.7, 97.8
α, β, γ (°)	90.0, 118.9, 90.0
Resolution (Å)	47.00–1.90 (1.97–1.90)[a]
R_{merge}	0.062 (0.394)
I / σ_I	13.20 (1.91)

$CC_{1/2}$ (%)	99.6 (65.2)
Completeness (%)	99.17 (98.88)
Redundancy	3.31 (3.20)
Refinement	
Resolution (Å)	47.00–1.90 (1.97–1.90)
No. reflections	49319
R_{work} / R_{free}	0.213 / 0.263
No. atoms	
Protein	5316
Ligand/ion	24
Water	518
B-factors	
Protein	30.46
Ligand/ion	41.37
Water	33.33
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.95
Ramachandran statistics	
Favored (%)	97.16
Allowed (%)	2.68
Outliers (%)	0.16
Rotamer outliers (%)	1.42
Clashscore	3.27

[a] Values in parentheses are for highest-resolution shells.

The obtained structure is shown in Figure 51 as an alignment with a reported structure of SUMO2–RanGAP1 in a complex with the E2 ligase ZNF451 (PDB entry 5d2m). The comparison revealed that the isopeptide ligand was superimposable onto the corresponding portions of the SUMO2 C-terminal tail and the SUMOylation motif in RanGAP1, with a Cα-backbone r.m.s. distance of 0.6 Å between the two models across the corresponding section. The alignment showed that Ubc9 can accommodate the distinctly different ubiquitin-derived sequence in place of the SUMO C-terminal tail. Furthermore, the isopeptide adopted a closed conformation, even in the absence of an E3 ligase, which has been shown to be important for efficient transfer of Ubls from an E2 to the acceptor lysine of a target protein (see Figure 22).[199] The ubiquitin-derived isopeptide portion of the ligand is located in the active site cleft and underneath of the gateway residue D127.

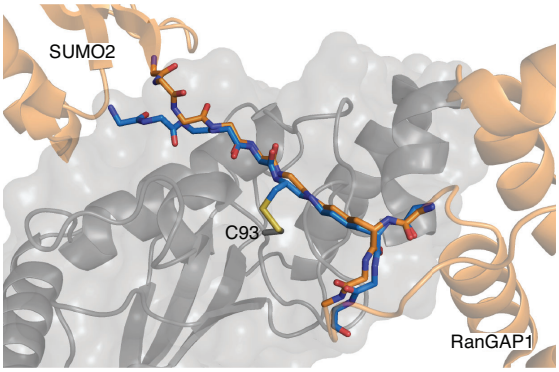


Figure 51. Alignment with a structure of Ubc9 in a complex with isopeptide-linked SUMO2–RanGAP1 (PDB entry 5d2m). The backbone of isopeptide ligand **21** (blue), the corresponding residues of SUMO2–RanGAP1 (orange) and the isopeptide and disulfide linkages are highlighted as sticks.

The SUMOylation motif IKQE interacted with the active site of Ubc9 in the expected manner, with the isoleucine contacting a hydrophobic patch of Ubc9, consisting of P128, A129, Q130 and A131.[198] Likewise, the glutamic acid residue was oriented towards S89 and T91, described to be important for recognition of the acidic residue of the SUMOylation motif.[198] Ubc9 bound the IKQE motif in an extended conformation, analogous to canonical substrates containing a consensus SUMOylation motif.[198] This conformation is consistent with proteomics studies showing that consensus SUMOylation is predominantly found in disordered regions.[153]

It is possible that the covalent disulfide complex with the isopeptide ligand does not precisely reflect the conformation of the Ubc9~peptide thioester intermediate or the tetrahedral intermediate of the *S*-to-*N* acyl shift. Nevertheless, a weak hydrogen bond of approximately 3.0 Å length between the heavy atoms of the isopeptide carbonyl group and N85 of the HPN motif within Ubc9 was observed (Figure 52). The observed orientation indicates that the ubiquitin-derived peptide can adopt a conformation that allows for activation of the C-terminal carbonyl group by Ubc9 to facilitate transthioesterification or aminolysis.

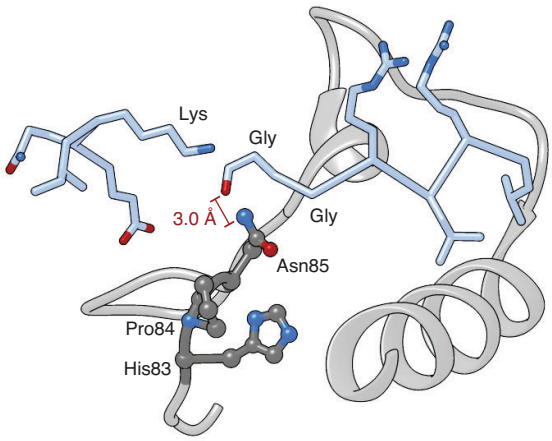


Figure 52. Positioning of the isopeptide carbonyl group (peptide in blue) relative to the HPN motif of Ubc9. The HPN motif, consisting of His83, Pr84 and Asn85, is highlighted in dark grey. The heavy atom distance

between the nitrogen atom of the Asn85 primary amide and the carbonyl oxygen of the isopeptide bond is indicated.

5.3. Structure validation

Further analysis of the structure did not reveal any strong interactions between the ubiquitin-derived isopeptide portion of the ligand and Ubc9, such as direct hydrogen bonds or ionic interactions. Nevertheless, the observed orientation was consistent with the alternating sequence of leucines and arginines, in which the leucine residues are buried at the interface and the arginines are solvent-exposed (Figure 53a).

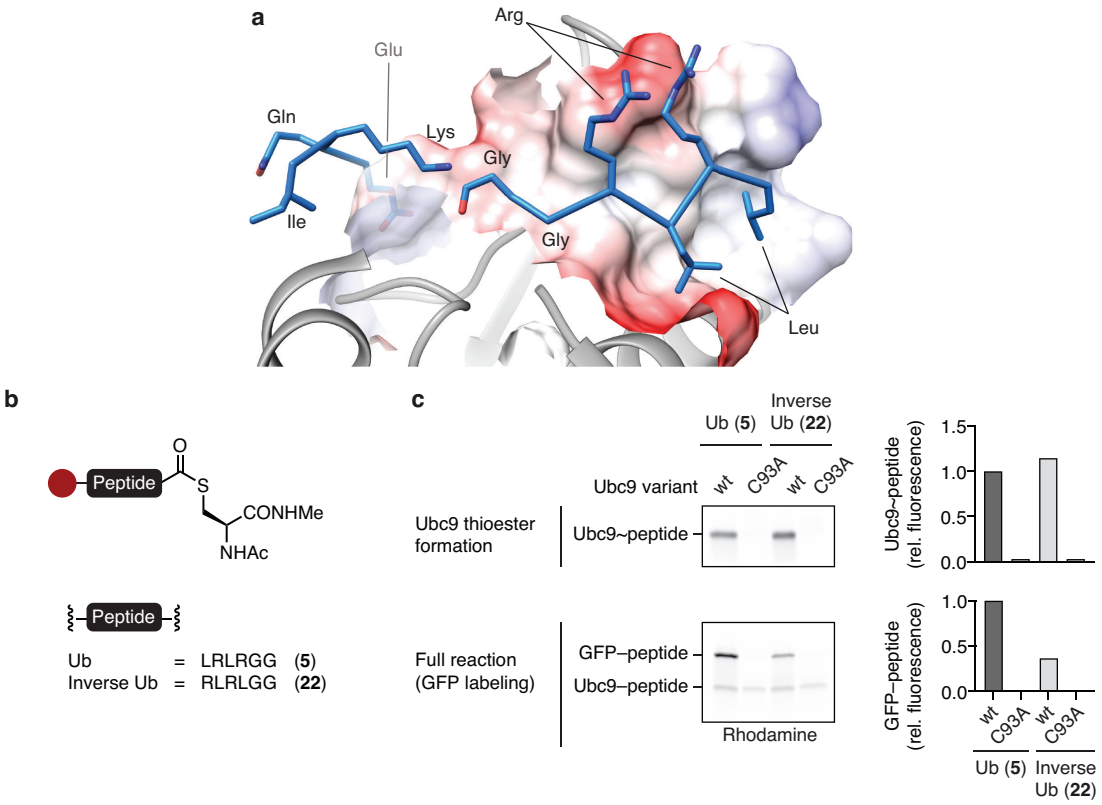


Figure 53. Interaction of the ubiquitin-derived peptide with Ubc9. (**a**) Positioning of the isopeptide ligand on Ubc9. The binding surface is shown in Coulombic coloring. (**b**) Structure and peptide sequences of ubiquitin-derived thioester **5** and the 'inverse' thioesters **22**. (**c**) In-gel fluorescence analysis of Ubc9~peptide thioester formation under non-reducing SDS–PAGE conditions (top) and of GFP-LACE$_C$ labeling reactions analyzed under reducing SDS–PAGE conditions (bottom). Reactions were carried out with the indicated Ubc9 variants and thioesters. The corresponding quantification by in-gel rhodamine fluorescence of Ubc9~peptide thioester formation (top) and GFP–peptide formation (bottom) is shown on the right. The mutation C138A was present in all Ubc9 variants to exclude reaction with the surface-exposed non-catalytic C138.

The extended β-sheet-like structure of the ubiquitin-derived peptide inspired us to test the effect of a thioester in LACE reactions in which the register of leucines and arginines is shifted (**22**, RLRL instead of LRLR) (Figure 53b,c). Looking at charging of the thioester onto Ubc9, no appreciable difference in Ubc9~peptide thioester formation was observed compared to the

standard thioester **5**. The rate of GFP-LACE$_C$ labeling in a full reaction was markedly reduced with thioester **22** compared to **5** (Figure 53c).

Since the 'inverse' peptide **22** showed lower reactivity in the overall labeling reaction of GFP-LACE$_C$ but not in Ubc9~peptide thioester formation, it is tempting to speculate that switching the register of leucines and arginines resulted in an altered orientation of the C-terminal carbonyl group of the peptide in the Ubc9 active site, resulting in slower *S*-to-*N* acyl transfer. However, further structural investigations are necessary to support this hypothesis and to conclusively determine the effect of the register switch. It is also possible that the flexibility afforded by the two glycine residues enables peptide **22** to adopt a partially proficient conformation relative to the HPN motif, which could explain the incomplete reduction in labeling efficiency.

The binding surface of Ubc9 surrounding the ubiquitin-derived peptide is enriched in acidic residues as can be seen from the coulombic surface representation in Figure 54. This might indicate that favorable ionic interactions between Ubc9 and the basic ubiquitin-derived peptide contribute to the higher reactivity compared to SUMO-derived sequences. This hypothesis is in agreement with the observation that high ionic strength buffers reduce the reactivity of LACE as observed in the buffer screen (see Figure 41), as well as the stepwise reduction in reactivity after each removal of an arginine residue in the thioester truncation study (see Figure 28b).

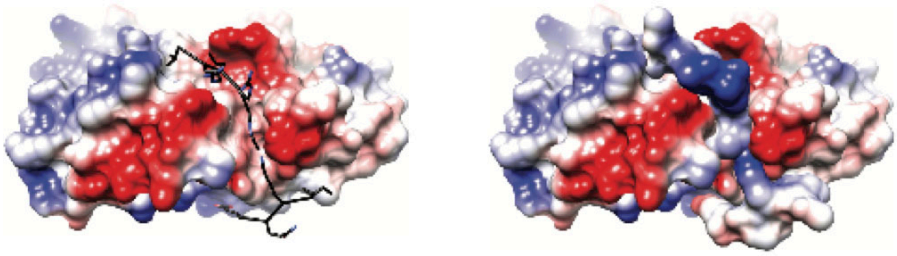


Figure 54. Structure of the disulfide complex in coulombic surface representation, without (left) and with (right) surface of the isopeptide ligand (black) shown.

To validate the above analyses, we assessed the effect of Ubc9 point mutations at the binding interface of Ubc9 and the isopeptide ligand (Figure 55). We tested the reactivity of the Ubc9 variants with ubiquitin-derived thioester **5** and the SUMO-derivative **2a** in Ubc9~peptide thioester formation and GFP-LACE$_C$ labeling reactions. Mutation of the acidic residues D102, E118 and E122, which form the presumed acidic binding surface of Ubc9 for the ubiquitin-derived sequence, to alanine or arginine residues had a neutral or even positive effect on both the rate of Ubc9~peptide thioester formation and the full labeling reaction of GFP-LACE$_C$. Generally, a similar reactivity change of the Ubc9 variants was observed with both ubiquitin-derived thioester **5** and the SUMO derivative **2a**. With these results, the ionic interaction between Ubc9 and the ubiquitin-derived peptide, proposed in Figure 54, could not be asserted. It is however possible that a single amino acid change was not sufficient to abolish the favorable interaction.

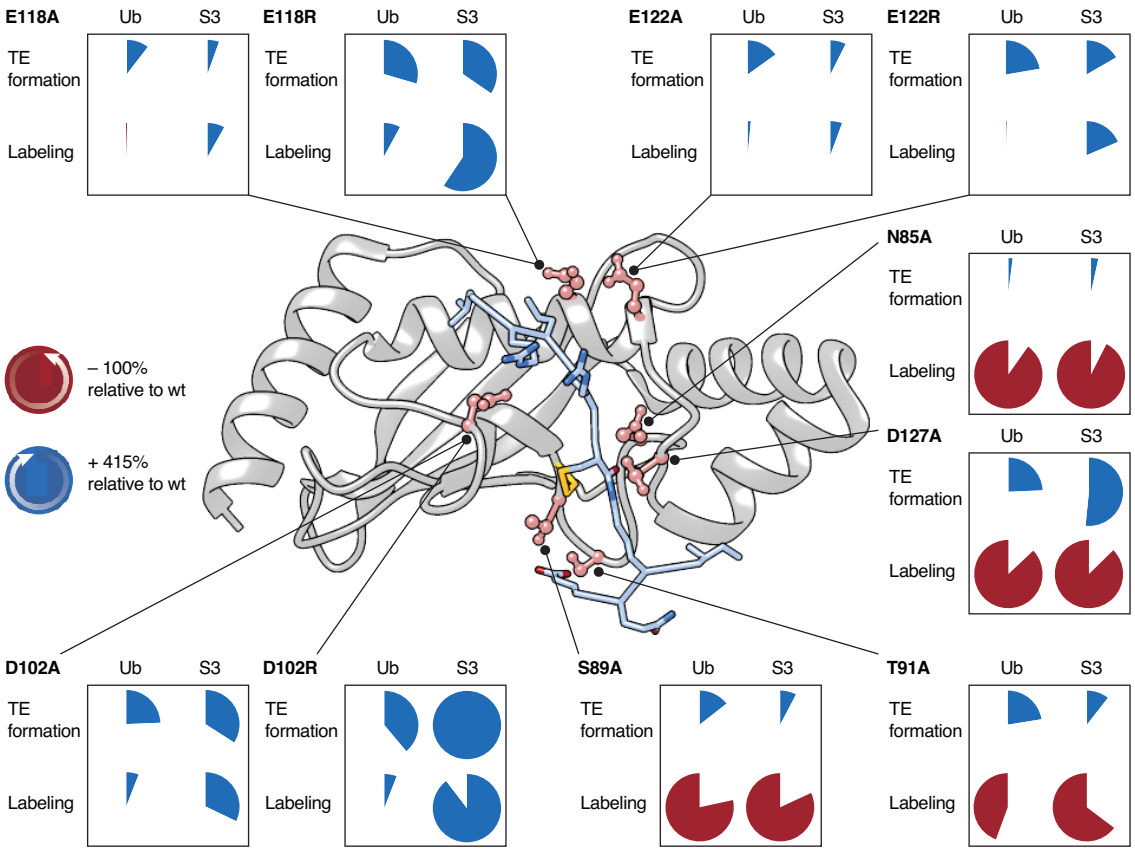


Figure 55. Ubc9 mutagenesis analysis. Residues that were mutated are highlighted in red, and the isopeptide ligand is shown in blue. Variants were assessed for Ubc9~peptide thioester formation (TE formation) using thioesters **5** (Ub) and **2a** (S3), and for labeling of GFP-LACE$_C$ (Labeling). The results are summarized relative to Ubc9-C138A (wt). To exclude reaction with the surface-exposed non-catalytic C138, the mutation C138A was included in all variants.

As expected, alanine mutation of two housekeeping residues, N85 as part of the HPN motif as well as the gateway residue D127, almost completely abolished GFP-LACE$_C$ labeling, whereas a neutral or slightly beneficial effect on thioester formation was observed. This result shows that N85 and D127 are not required for transthioesterification with peptide thioesters, but contribute significantly to efficient aminolysis of the Ubc9~thioester intermediate by the acceptor lysine within the LACE tag. Lastly, mutation of S89 and T91 to alanine had a negligible effect on thioester formation, but greatly reduced labeling of the substrate GFP-LACE$_C$. This finding is in agreement with their reported role in binding of the acidic residue within the SUMOylation motif of substrates,[198] and thus also of the LACE tag.

Overall, the reactivity of the Ubc9 mutants with thioester **5** qualitatively mirrors that observed with thioester **2a**. This indicates that the studied single-point mutations reflect their contribution on general LACE reactivity, rather than sequence-specific interactions with the peptide. As such, we could not conclusively determine residues that contribute to the high reactivity of Ubc9 with thioester **5**. Nevertheless, the observed effects support the notion that LACE reactions are dictated by interactions similar to the ones observed in the static crystal

structure. Residues close to the ubiquitin-derived peptide in the crystal structure indeed had the largest effect on thioester formation, for example residues D102, E118 and D127, whereas housekeeping residues and residues in contact with the LACE tag IKQE were required for efficient aminolysis and transfer of the thioester to the acceptor lysine.

6. Discussion

In this Chapter, we described the development of a chemoenzymatic method for internal isopeptide labeling using the conjugating enzyme Ubc9 – a component of the natural SUMOylation pathway. We identified peptide thioesters as efficient acyl donors for Ubc9 and determined their structure-activity relationship. By taking advantage of the known sequence specificity of Ubc9 for the modification of lysines embedded in a consensus SUMOylation motif,[153,197] we identified the minimal sequence requirements of the acceptor tag and developed reaction conditions that allow for complete, specific and irreversible protein modification of unnatural substrates of Ubc9 bearing such a LACE tag.

The identified minimal acceptor tag (four to eleven residues) allowed us to direct protein modification to terminal as well as internal parts of a protein. Additionally, isopeptide labeling allowed us to modify a protein substrate bearing two acceptor tags. While LACE tags embedded in secondary structures have limited reactivity, the absence of a specific secondary structure requirement enables straightforward incorporation of LACE tags into unstructured regions, loops and tags fused to the N- or C-termini. Placement of the tag in different protein locations and contexts underlines the generality of the method. These findings should expand the scope of available substrates for general internal isopeptide labeling.

The surprising proficiency of ubiquitin-derived peptide thioesters as acyl donors for the SUMO-conjugating enzyme Ubc9 can be attributed to the fact that no E1 enzyme – thought to be the gatekeeper for isoform selectivity[139] – is required. Compared to the ubiquitin-derived acyl donor, peptide thioesters derived from SUMO isoforms showed low reactivity, with little difference in reactivity between the isoforms. This finding is in contrast to canonical SUMOylation, where distinct conjugation dynamics for the four SUMO isoforms have been well described.[162,165,166] These differences between SUMO isoform conjugation dynamics therefore likely stem from interactions with the entire SUMO fold, or from the action of E1 and E3 enzymes and SUMO isopeptidases, rather than their intrinsic reactivity with Ubc9.

Additionally, several artificial peptide acyl donors bearing sequences that were identified by phage display conferred comparable reactivity to the SUMO-derived thioesters (see Figure 45). These findings suggest that Ubc9 does not engage in any specific interactions with the SUMO C-terminus under the LACE conditions, and provide potential insights into the evolution of E2 conjugating enzymes. It is tempting to speculate that an ancestral E2 conjugating enzyme has evolved to accept ubiquitin peptides. The intrinsic preference for ubiquitin-derived sequences in

the E2 active site cleft may have been preserved during divergent evolution which gave rise to the various Ubl systems, including SUMO. Instead, isoform selectivity was realized by specific interaction between E2 enzymes and cognate E1 activating enzymes, and recognition of cognate Ubls by E1 enzymes.[139] In this background, E2 conjugating enzymes may not have been under selection pressure to evolve intrinsic selectivity for a particular Ubl C-terminus, and other E2 enzymes might exhibit a preference for ubiquitin sequences as well.

Our reaction profile and optimization studies revealed that the first of two steps – charging of Ubc9 with the acyl donor – is the rate-limiting step of LACE. This observation is in agreement with the generally elevated pK_a of the active site cysteine reported for free E2 enzymes, which may require interaction with E1 enzymes to tune the nucleophilicity for efficient transthioesterification.[180] The low reactivity of the active site cysteine may also explain why Dha formation by bis-alkylation and elimination using DBHDA to prepare covalent conjugates for X-ray crystallography was not successful (see Figure 48b,c). Acceleration of the first step is therefore a key requirement to increase the efficiency of LACE.

To this end, further optimization of the interaction between the acyl donor and Ubc9 should be possible. Our screen for acyl donor sequences included several hypothesis-driven candidates taken from SUMO isoforms and ubiquitin, and included a comparison to literature-known peptides that are utilized by components of the Ubl pathway. Further optimization by an unbiased screen might be possible. Likewise, the leaving group, which is currently a simple alkyl thioester or cysteine derivative, may be optimized. We find that, from the tested leaving groups, the cysteine derivative Ac-Cys-NHMe is preferred. However, commercially available alkyl thioesters such as Mes can be used instead with just slightly less reactivity. While the cysteine derivative was intended to mimic the active site cysteine of the E1~Ubl thioester intermediate, the enhanced reactivity could also be explained by the intrinsic electrophilicity of the respective thioesters. Indeed, NCL kinetics studies showed that cysteine-derived thioesters (pK_a 8.3–8.5, ref. [219]) exhibit improved ligation efficiency compared to Mes-derived thioesters (pK_a 9.5, ref. [220]).[219] Good correspondence between the NCL rate and the pK_a of the departing thiols has been observed.[221,222] In line with this notion, we describe the use of nucleophilic catalysis using MPAA (pK_a 6.6, ref. [223]) as one approach to further increase the efficiency of the first step.

In terms of reaction rate, LACE compares favorably to first generation sortases that are used in stoichiometric amounts for labeling of protein termini.[224,225] Additionally, sortase reactions usually require high concentrations of the nucleophile because of their reversibility to drive the equilibrium to the product side, and to suppress hydrolysis side products. The irreversibility of the LACE reaction with an activated acyl donor allows for this process to reach completion without requiring a large excess of one of the reaction components. Comparing LACE to chemoenzymatic methods for internal peptidic labeling, such as non-canonical activity of sortases[113–115] or

SpyCatcher/SpyTag,[95,96,105–107] both the reactivity and the acceptor tag size of LACE represent a significant improvement.

Lastly, our structural analyses revealed the molecular details for the surprising ability of Ubc9 to accept the distinctly different C-terminal ubiquitin sequence in place of SUMO. The ubiquitin sequence appears to adopt a similar conformation as the native SUMO substrates. LACE labeling with ubiquitin-derived thioesters benefits from the known interactions and active site residues through which Ubc9 mediates aminolysis with canonical substrates. Ubc9 mutagenesis studies showed that single amino acid substitutions greatly affect the reactivity. The observed plasticity of the Ubc9 reactivity, combined with the compact tertiary structure and facile handling of Ubc9, indicates that this enzyme may be amenable to directed evolution or engineering approaches to obtain variants with higher reactivity or altered specificity. Given our mechanistic studies, the rate of LACE reactions may be improved by finding Ubc9 variants that have a more nucleophilic active site cysteine to accelerate Ubc9~peptide thioester formation with activated acyl donors – currently the rate-limiting step under our standard conditions. Alternatively, variants could be identified that exhibit increased affinity for the acyl donors, thereby promoting transthioesterification and charging of the Ubc9.

7. References

(117) Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **2009**, *458*, 422–429.

(118) Welchman, R. L.; Gordon, C.; Mayer, R. J. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 599–609.

(119) Hochstrasser, M. *Ubiquitin-dependent protein degradation*; Academic Press: Cambridge, MA, 2019; Vol. 619.

(120) Hay, R. T. Role of ubiquitin-like proteins in transcriptional regulation. In *The Histone Code and Beyond*; Berger, S. L., Nakanishi, O., Haendler, B., Ed.; Ernst Schering Research Foundation Workshop; Springer: Berlin, Heidelberg, Germany, 2006; Vol. 57, pp 173–192.

(121) Ulrich, H. D. Two-way communications between ubiquitin-like modifiers and DNA. *Nat. Struct. Mol. Biol.* **2014**, *21*, 317–324.

(122) Dasso, M. Emerging roles of the SUMO pathway in mitosis. *Cell Div.* **2008**, *3*, 5.

(123) Bard, J. A. M.; Goodall, E. A.; Greene, E. R.; Jonsson, E.; Dong, K. C.; Martin, A. Structure and function of the 26S proteasome. *Annu. Rev. Biochem.* **2018**, *87*, 1–28.

(124) Haakonsen, D. L.; Rape, M. Branching out: Improved signaling by heterotypic ubiquitin chains. *Trends Cell. Biol.* **2019**, *29*, 704–716.

(125) Yau, R.; Rape, M. The increasing complexity of the ubiquitin code. *Nat. Cell Biol.* **2016**, *18*, 579–586.

(126) Sharp, P. M.; Li, W.-H. Molecular evolution of ubiquitin genes. *Trends. Ecol. Evol.* **1987**, *2*, 328–332.

(127) Veen, A. G. van der; Ploegh, H. L. Ubiquitin-like proteins. *Annu. Rev. Biochem.* **2012**, *81*, 323–357.

(128) Matunis, M. J.; Coutavas, E.; Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* **1996**, *135*, 1457–1470.

(129) Kamitani, T.; Kito, K.; Nguyen, H. P.; Yeh, E. T. H. Characterization of NEDD8, a developmentally down-regulated ubiquitin-like protein. *J. Biol. Chem.* **1997**, *272*, 28557–28562.

(130) Komatsu, M.; Chiba, T.; Tatsumi, K.; Iemura, S.; Tanida, I.; Okazaki, N.; Ueno, T.; Kominami, E.; Natsume, T.; Tanaka, K. A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. *EMBO J.* **2004**, *23*, 1977–1986.

(131) Loeb, K. R.; Haas, A. L. The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins. *J. Biological. Chem.* **1992**, *267*, 7806–7813.

(132) Lapenta, V.; Chiurazzi, P.; van der Spek, P.; Pizzuti, A.; Hanaoka, F.; Brahe, C.: SMT3A, a human homologue of the *S. cerevisiae* SMT3 gene, maps to chromosome 21 qter and defines a novel gene family. *Genomics* 1**997**, *40*, 362–366.

(133) Burroughs, A. M.; Balaji, S.; Iyer, L. M.; Aravind, L. Small but versatile: The extraordinary functional and structural diversity of the β-grasp fold. *Biol. Direct* **2007**, *2*, 18.

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

(135) Schulman, B. A. Twists and turns in ubiquitin-like protein conjugation cascades. *Protein Sci.* **2011**, *20*, 1941–1954.

(136) Mevissen, T. E. T.; Komander, D. Mechanisms of deubiquitinase specificity and regulation. *Annu. Rev. Biochem.* **2016**, *86*, 1–33.

(137) Clague, M. J.; Urbé, S.; Komander, D. Breaking the chains: Deubiquitylating enzyme specificity begets function. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 338–352.

(138) Cappadocia, L.; Lima, C. D. Ubiquitin-like protein conjugation: Structures, chemistry, and mechanism. *Chem. Rev.* **2017**, *118*, 889–918.

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

(140) Olsen, S. K.; Lima, C. D. Structure of a ubiquitin E1-E2 complex: Insights to E1-E2 thioester transfer. *Mol. Cell.* **2013**, *49*, 884–896.

(141) Berndsen, C. E.; Wolberger, C. New insights into ubiquitin E3 ligase mechanism. *Nat. Struct. Mol. Biol.* **2014**, *21*, 301–307.

(142) Li, W.; Bengtson, M. H.; Ulbrich, A.; Matsuda, A.; Reddy, V. A.; Orth, A.; Chanda, S. K.; Batalov, S.; Joazeiro, C. A. P. Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PLoS One* **2008**, *3*, e1487.

(143) Buetow, L.; Huang, D. T. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 626–642.

(144) Sluimer, J.; Distel, B. Regulating the human HECT E3 ligases. *Cell Mol. Life Sci.* **2018**, *75*, 3121–3141.

(145) Scheffner, M.; Kumar, S. Mammalian HECT ubiquitin-protein ligases: Biological and pathophysiological aspects. *Biochim. Biophys. Acta* **2014**, *1843*, 61–74.

(146) Smit, J. J.; Sixma, T. K. RBR E3-ligases at work. *EMBO Rep.* **2014**, *15*, 142–154.

(147) Spratt, D. E.; Walden, H.; Shaw, G. S. RBR E3 ubiquitin ligases: New structures, new insights, new questions. *Biochem. J.* **2014**, *458*, 421–437.

(148) Desterro, J. M. P.; Rodriguez, M. S.; Kemp, G. D.; Hay, R. T. Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *J. Biol. Chem.* **1999**, *274*, 10618–10624.

(149) Streich, F. C. Jr.; Lima, C. D. Structural and functional insights to ubiquitin-like protein conjugation. *Annu. Rev. Biophys.* **2014**, *43*, 357–379.

(150) Lee, G. W.; Melchior, F.; Matunis, M. J.; Mahajan, R.; Tian, Q.; Anderson, P. Modification of Ran GTPase-activating protein by the small ubiquitin-related modifier SUMO-1 requires Ubc9, an E2-type ubiquitin-conjugating enzyme homologue. *J. Biol. Chem.* **1998**, *273*, 6503–6507.

(151) Flotho, A.; Melchior, F. SUMOylation: A regulatory protein modification in health and disease. *Annu. Rev. Biochem.* **2013**, *82*, 357–385.

(152) Hietakangas, V.; Anckar, J.; Blomster, H. A.; Fujimoto, M.; Palvimo, J. J.; Nakai, A.; Sistonen, L. PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 45–50.

(153) Hendriks, I. A.; Lyon, D.; Young, C.; Jensen, L. J.; Vertegaal, A. C. O.; Nielsen, M. L. Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat. Struct. Mol. Biol.* **2017**, *24*, 325–336.

(154) Hendriks, I. A.; Vertegaal, A. C. O. A Comprehensive compilation of SUMO proteomics. *Nat. Rev. Mol. Cell. Bio.* **2016**, *17*, 581–595.

(155) Lamoliatte, F.; Caron, D.; Durette, C.; Mahrouche, L.; Maroui, M. A.; Caron-Lizotte, O.; Bonneil, E.; Chelbi-Alix, M. K.; Thibault, P. Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. *Nat. Commun.* **2014**, *5*, 5409.

(156) Hendriks, I. A.; D'Souza, R. C. J.; Yang, B.; Vries, M. V.; Mann, M.; Vertegaal, A. C. O. Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat. Struct. Mol. Biol.* **2014**, *21*, 927–936.

(157) Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **2004**, *73*, 355–382.

(158) Hecker, C.-M.; Rabiller, M.; Haglund, K.; Bayer, P.; Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* **2006**, *281*, 16117–16127.

(159) Reverter, D.; Lima, C. D. Structural basis for SENP2 protease interactions with SUMO precursors and conjugated substrates. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1060–1068.

(160) Shin, E. J.; Shin, H. M.; Nam, E.; Kim, W. S.; Kim, J.; Oh, B.; Yun, Y. DeSUMOylating isopeptidase: A second class of SUMO protease. *EMBO Rep.* **2012**, *13*, 339–346.

(161) Schulz, S.; Chachami, G.; Kozaczkiewicz, L.; Winter, U.; Stankovic-Valentin, N.; Haas, P.; Hofmann, K.; Urlaub, H.; Ovaa, H.; Wittbrodt, J.; Meulmeester, E.; Melchior, F. Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. *EMBO Rep.* **2012**, *13*, 930–938.

(162) Owerbach, D.; McKay, E. M.; Yeh, E. T. H.; Gabbay, K. H.; Bohren, K. M. A proline-90 residue unique to SUMO-4 prevents maturation and SUMOylation. *Biochem. Bioph. Res. Commun.* **2005**, *337*, 517–520.

(163) Guo, D.; Han, J.; Adam, B.-L.; Colburn, N. H.; Wang, M.-H.; Dong, Z.; Eizirik, D. L.; She, J.-X.; Wang, C.-Y. Proteomic analysis of SUMO4 substrates in HEK293 cells under serum starvation-induced stress. *Biochem. Bioph. Res. Commun.* **2005**, *337*, 1308–1318.

(164) Guo, D.; Li, M.; Zhang, Y.; Yang, P.; Eckenrode, S.; Hopkins, D.; Zheng, W.; Purohit, S.; Podolsky, R. H.; Muir, A.; Wang, J.; Dong, Z.; Brusko, T.; Atkinson, M.; Pozzilli, P.; Zeidler, A.; Raffel, L. J.; Jacob, C. O.; Park, Y.; Serrano-Rios, M.; Larrad, M. T. M.; Zhang, Z.; Garchon, H.-J.; Bach, J.-F.; Rotter, J. I.; She, J.-X.; Wang, C.-Y. A functional variant of SUMO4, a new IκBα modifier, is associated with type 1 diabetes. *Nat. Genet.* **2004**, *36*, 837–841.

(165) Saitoh, H.; Hinchey, J. Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* **2000**, *275*, 6252–6258.

(166) Ayaydin, F.; Dasso, M. Distinct *in vivo* dynamics of vertebrate SUMO paralogues. *Mol. Biol. Cell* **2004**, *15*, 5208–5218.

(167) Tatham, M. H.; Jaffray, E.; Vaughan, O. A.; Desterro, J. M. P.; Botting, C. H.; Naismith, J. H.; Hay, R. T. Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. *J. Biol. Chem.* **2001**, *276*, 35368–35374.

(168) Matic, I.; Hagen, M. van; Schimmel, J.; Macek, B.; Ogg, S. C.; Tatham, M. H.; Hay, R. T.; Lamond, A. I.; Mann, M.; Vertegaal, A. C. O. *In vivo* identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an *in vitro* to *in vivo* strategy. *Mol. Cell Proteomics* **2008**, *7*, 132–144.

(169) Stewart, M. D.; Ritterhoff, T.; Klevit, R. E.; Brzovic, P. S. E2 enzymes: More than just middle men. *Cell Res.* **2016**, *26*, 423–440.

(170) Ye, Y.; Rape, M. Building ubiquitin chains: E2 enzymes at work. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 755–764.

(171) David, Y.; Ziv, T.; Admon, A.; Navon, A. The E2 ubiquitin-conjugating enzymes direct polyubiquitination to preferred lysines. *J. Biol. Chem.* **2010**, *285*, 8595–8604.

(172) Michelle, C.; Vourc'h, P.; Mignon, L.; Andres, C. R. What was the set of ubiquitin and ubiquitin-like conjugating enzymes in the eukaryote common ancestor? *J. Mol. Evol.* **2009**, *68*, 616–628.

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

(174) Valimberti, I.; Tiberti, M.; Lambrughi, M.; Sarcevic, B.; Papaleo, E. E2 superfamily of ubiquitin-conjugating enzymes: constitutively active or activated through phosphorylation in the catalytic cleft. *Sci. Rep.* **2015**, *5*, 14849.

(175) Shchebet, A.; Karpiuk, O.; Kremmer, E.; Eick, D.; Johnsen, S. A. Phosphorylation by cyclin-dependent kinase-9 controls ubiquitin-conjugating enzyme-2A function. *Cell Cycle* **2014**, *11*, 2122–2127.

(176) Winn, P. J.; Religa, T. L.; Battey, J. N. D.; Banerjee, A.; Wade, R. C. Determinants of functionality in the ubiquitin conjugating enzyme family. *Structure* **2004**, *12*, 1563–1574.

(177) Wu, P.-Y.; Hanlon, M.; Eddins, M.; Tsui, C.; Rogers, R. S.; Jensen, J. P.; Matunis, M. J.; Weissman, A. M.; Weisman, A. M.; Weissman, A. M.; Wolberger, C.; Wolberger, C. P.;

Pickart, C. M. A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J.* **2003**, *22*, 5241–5250.

(178) Jones, W. M.; Davis, A. G.; Wilson, R. H.; Elliott, K. L.; Sumner, I. A Conserved asparagine in a ubiquitin-conjugating enzyme positions the substrate for nucleophilic attack. *J. Comput. Chem.* **2019**, *40*, 1969–1977.

(179) Berndsen, C. E.; Wiener, R.; Yu, I. W.; Ringel, A. E.; Wolberger, C. A conserved asparagine has a structural role in ubiquitin-conjugating enzymes. *Nat. Chem. Biol.* **2013**, *9*, 154–156.

(180) Tolbert, B. S.; Tajc, S. G.; Webb, H.; Snyder, J.; Nielsen, J. E.; Miller, B. L.; Basavappa, R. The active site cysteine of ubiquitin-conjugating enzymes has a significantly elevated pK_a: Functional implications. *Biochemistry* **2005**, *44*, 16385–16391.

(181) Jin, J.; Li, X.; Gygi, S. P.; Harper, J. W. Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. *Nature* **2007**, *447*, 1135–1138.

(182) Pruneda, J. N.; Stoll, K. E.; Bolton, L. J.; Brzovic, P. S.; Klevit, R. E. Ubiquitin in motion: Structural studies of the ubiquitin-conjugating enzyme~ubiquitin conjugate. *Biochemistry* **2011**, *50*, 1624–1633.

(183) Hamilton, K. S.; Ellison, M. J.; Barber, K. R.; Williams, R. S.; Huzil, J. T.; McKenna, S.; Ptak, C.; Glover, M.; Shaw, G. S. Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. *Structure* **2001**, *9*, 897–904.

(184) Yunus, A. A.; Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nat. Struct. Mol. Biol.* **2006**, *13*, 491–499.

(185) Huang, L.; Kinnucan, E.; Wang, G.; Beaudenon, S.; Howley, P. M.; Huibregtse, J. M.; Pavletich, N. P. Structure of an E6AP-UbcH7 complex: Insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **1999**, *286*, 1321–1326.

(186) Eletr, Z. M.; Huang, D. T.; Duda, D. M.; Schulman, B. A.; Kuhlman, B. E2 conjugating enzymes must disengage from their E1 enzymes before E3-dependent ubiquitin and ubiquitin-like transfer. *Nat. Struct. Mol. Biol.* **2005**, *12*, 933–934.

(187) Das, R.; Mariano, J.; Tsai, Y. C.; Kalathur, R. C.; Kostova, Z.; Li, J.; Tarasov, S. G.; McFeeters, R. L.; Altieri, A. S.; Ji, X.; Byrd, R. A.; Weissman, A. M. Allosteric activation of E2-RING finger-mediated ubiquitylation by a structurally defined specific E2-binding region of Gp78. *Mol. Cell* **2009**, *34*, 674–685.

(188) Knipscheer, P.; Dijk, W. J. van; Olsen, J. V.; Mann, M.; Sixma, T. K. Noncovalent interaction between Ubc9 and SUMO promotes SUMO chain formation. *EMBO J.* **2007**, *26*, 2797–2807.

(189) Nguyen, L.; Plafker, K. S.; Starnes, A.; Cook, M.; Klevit, R. E.; Plafker, S. M. The ubiquitin-conjugating enzyme, UbcM2, is restricted to monoubiquitylation by a two-fold mechanism that involves backside residues of E2 and Lys48 of ubiquitin. *Biochemistry* **2014**, *53*, 4004–4014.

(190) Dou, H.; Buetow, L.; Sibbet, G. J.; Cameron, K.; Huang, D. T. BIRC7–E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat. Struct. Mol. Biol.* **2012**, *19*, 876–883.

(191) Özkan, E.; Yu, H.; Deisenhofer, J. Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18890–18895.

(192) Pruneda, J. N.; Littlefield, P. J.; Soss, S. E.; Nordquist, K. A.; Chazin, W. J.; Brzovic, P. S.; Klevit, R. E. Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-Box ligases. *Mol. Cell* **2012**, *47*, 933–942.

(193) Madden, M. M.; Song, W.; Martell, P. G.; Ren, Y.; Feng, J.; Lin, Q. Substrate properties of ubiquitin carboxyl-terminally derived peptide probes for protein ubiquitination. *Biochemistry* **2008**, *47*, 3636–3644.

(194) Chang, L. H.; Strieter, E. R. Reprogramming a deubiquitinase into a transamidase. *ACS Chem. Biol.* **2018**, *13*, 2808–2818.

(195) Park, S.; Krist, D. T.; Statsyuk, A. V. Protein ubiquitination and formation of polyubiquitin chains without ATP, E1 and E2 enzymes. *Chem. Sci.* **2014**, *6*, 1770–1779.

(196) Zhao, B.; Tsai, Y. C.; Jin, B.; Wang, B.; Wang, Y.; Zhou, H.; Carpenter, T.; Weissman, A. M.; Yin, J. Protein engineering in the ubiquitin system: Tools for discovery and beyond. *Pharmacol. Rev.* **2020**, *72*, 380–413.

(197) Rodriguez, M. S.; Dargemont, C.; Hay, R. T. SUMO-1 conjugation *in vivo* requires both a consensus modification motif and nuclear targeting. *J. Biol. Chem.* **2001**, *276*, 12654–12659.

(198) Bernier-Villamor, V.; Sampson, D. A.; Matunis, M. J.; Lima, C. D. Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* **2002**, *108*, 345–356.

(199) Gareau, J. R.; Lima, C. D. The SUMO pathway: Emerging mechanisms that shape specificity, conjugation and recognition. *Nat. Rev. Mol. Cell. Bio.* **2010**, *11*, 861–871.

(200) Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.

(201) Yang, S.-H. H.; Galanis, A.; Witty, J.; Sharrocks, A. D. An extended consensus motif enhances the specificity of substrate modification by SUMO. *EMBO J.* **2006**, *25*, 5083–5093.

(202) Zheng, Q.; Maksimovic, I.; Upad, A.; David, Y. Non-enzymatic covalent modifications: A new link between metabolism and epigenetics. *Protein Cell* **2020**, *11*, 401–416.

(203) Wagner, G. R.; Hirschey, M. D. Nonenzymatic protein acylation as a carbon stress regulated by sirtuin deacylases. *Mol. Cell.* **2014**, *54*, 5–16.

(204) Kulkarni, R. A.; Worth, A. J.; Zengeya, T. T.; Shrimp, J. H.; Garlick, J. M.; Roberts, A. M.; Montgomery, D. C.; Sourbier, C.; Gibbs, B. K.; Mesaros, C.; Tsai, Y. C.; Das, S.; Chan, K. C.; Zhou, M.; Andresson, T.; Weissman, A. M.; Linehan, W. M.; Blair, I. A.; Snyder, N. W.; Meier, J. L. Discovering targets of non-enzymatic acylation by thioester reactivity profiling. *Cell Chem. Biol.* **2017**, *24*, 231–242.

(205) Trub, A. G.; Hirschey, M. D. Reactive acyl-CoA species modify proteins and induce carbon stress. *Trends. Biochem. Sci.* **2018**, *43*, 369–379.

(206) James, A. M.; Smith, A. C.; Smith, C. L.; Robinson, A. J.; Murphy, M. P. Proximal cysteines that enhance lysine *N*-acetylation of cytosolic proteins in mice are less conserved in longer-living species. *Cell Rep.* **2018**, *24*, 1445–1455.

(207) James, A. M.; Hoogewijs, K.; Logan, A.; Hall, A. R.; Ding, S.; Fearnley, I. M.; Murphy, M. P. Non-enzymatic *N*-acetylation of lysine residues by AcetylCoA often occurs via a proximal *S*-acetylated thiol intermediate sensitive to glyoxalase II. *Cell Rep.* **2017**, *18*, 2105–2112.

(208) James, A. M.; Smith, C. L.; Smith, A. C.; Robinson, A. J.; Hoogewijs, K.; Murphy, M. P. The causes and consequences of nonenzymatic protein acylation. *Trends Biochem. Sci.* **2018**, *43*, 921–932.

(209) Ho, C.-W.; Chen, H.-T.; Hwang, J. Ubc9 autoSUMOylation negatively regulates SUMOylation of septins in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **2011**, *286*, 21826–21834.

(210) Knipscheer, P.; Flotho, A.; Klug, H.; Olsen, J. V.; Dijk, W. J. van; Fish, A.; Johnson, E. S.; Mann, M.; Sixma, T. K.; Pichler, A. Ubc9 SUMOylation regulates SUMO target discrimination. *Mol. Cell* **2008**, *31*, 371–382.

(211) Liess, A. K. L.; Kucerova, A.; Schweimer, K.; Yu, L.; Roumeliotis, T. I.; Diebold, M.; Dybkov, O.; Sotriffer, C.; Urlaub, H.; Choudhary, J. S.; Mansfeld, J.; Lorenz, S. Autoinhibition mechanism of the ubiquitin-conjugating enzyme Ube2S by autoubiquitination. *Structure* **2019**, *27*, 1195–1210.

(212) Hann, Z. S.; Ji, C.; Olsen, S. K.; Lu, X.; Lux, M. C.; Tan, D. S.; Lima, C. D. Structural basis for adenylation and thioester bond formation in the ubiquitin E1. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 15475–15484.

(213) Gates, Z. P.; Stephan, J. R.; Lee, D. J.; Kent, S. B. H. Rapid formal hydrolysis of peptide-α-thioesters. *Chem. Commun.* **2012**, *49*, 786–788.

(214) Sakamoto, K.; Tsuda, S.; Mochizuki, M.; Nohara, Y.; Nishio, H.; Yoshiya, T. Imidazole-aided native chemical ligation: imidazole as a one-pot desulfurization-amenable non-thiol-type alternative to 4-mercaptophenylacetic acid. *Chem. Eur. J.* **2016**, *22*, 17940–17944.

(215) Zhao, B.; Villhauer, E. B.; Bhuripanyo, K.; Kiyokawa, H.; Schindelin, H.; Yin, J. SUMO-mimicking peptides inhibiting protein SUMOylation. *ChemBioChem* **2014**, *15*, 2662–2666.

(216) Zhao, B.; Choi, C. H. J.; Bhuripanyo, K.; Villhauer, E. B.; Zhang, K.; Schindelin, H.; Yin, J. Inhibiting the protein ubiquitination cascade by ubiquitin-mimicking short peptides. *Org. Lett.* **2012**, *14*, 5760–5763.

(217) Hewitt, W. M.; Lountos, G. T.; Zlotkowski, K.; Dahlhauser, S. D.; Saunders, L. B.; Needle, D.; Tropea, J. E.; Zhan, C.; Wei, G.; Ma, B.; Nussinov, R.; Waugh, D. S.; Schneekloth, J. S. Insights into the allosteric inhibition of the SUMO E2 enzyme Ubc9. *Angew. Chem. Int. Ed.* **2016**, *55*, 5703–5707.

(218) Merkley, N.; Barber, K. R.; Shaw, G. S. Ubiquitin manipulation by an E2 conjugating enzyme using a novel covalent intermediate. *J. Biol. Chem.* **2005**, *280*, 31732–31738.

(219) Tsuda, S.; Yoshiya, T.; Mochizuki, M.; Nishiuchi, Y. Synthesis of cysteine-rich peptides by native chemical ligation without use of exogenous thiols. *Org. Lett.* **2015**, *17*, 1806–1809.

(220) Danehy, J. P.; Noel, C. J. The relative nucleophilic character of several mercaptans toward ethylene oxide. *J. Am. Chem. Soc.* **1960**,*82*, 2511–2515.

(221) Agouridas, V.; Mahdi, O. E.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O. Native chemical ligation and extended methods: Mechanisms, catalysis, scope, and limitations. *Chem. Rev.* **2019**, *119*, 7328–7443.

(222) Hupe, D. J.; Jencks, W. P. Nonlinear structure-reactivity correlations. acyl transfer between sulfur and oxygen nucleophiles. *J. Am. Chem. Soc.* **1977**, *99*, 451–464.

(223) DeCollo, T. V.; Lees, W. J. Effects of aromatic thiols on thiol–disulfide interchange reactions that occur during protein folding. *J. Org. Chem.* **2001**, *66*, 4244–4249.

(224) Antos, J. M.; Truttmann, M. C.; Ploegh, H. L. Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **2016**, *38*, 111–118.

(225) Theile, C. S.; Witte, M. D.; Blom, A. E. M.; Kundrat, L.; Ploegh, H. L.; Guimaraes, C. P. Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **2013**, *8*, 1800–1807.

CHAPTER 3

Protein Functionalization and Conjugation by LACE

Portions of the work described in this Chapter have been performed in collaborations.

Dominik Schauenburg (Bode Group, ETH Zürich) developed and prepared TIM pentafluorophenyl ester **38** and contributed to the development of TIM thioester **39**. Dr. Régis Boehringer (Bode Group, ETH Zürich) developed and prepared DBCO thioester **30**. The work is presented in Sections 3.3.1 and 3.3.2.

Mamiko Ninomiya (Bode Group, ETH Zürich) and Cecilie Egholm (Boyman Group, UZH) conceived the project on the IL-4–IL-13 conjugate. M. Ninomiya developed and prepared the synthetic IL-4 and IL-13 variants, performed and monitored the KAT ligation, and purified the IL-4–IL-13 conjugate. C. Egholm performed pSTAT assays. The work is presented in Section 3.3.2.4.

Gaku Akimoto (Bode Group, ETH Zürich) designed and performed monoubiquitination reactions of α-synuclein and SUMO2, which are presented in Section 4.1.1.2.

Mikail Levasseur (Hilvert Group, ETH Zürich) designed and prepared the protein cage, contributed to the design of the thioester probes for cage labeling, and performed analyses by size-exclusion chromatography and transmission electron microscopy. The work is presented in Section 5. We also thank M. Levasseur for the help with flow cytometry analysis of the immunoconjugate in Section 2.3 and for a gift of cell lines.

1. Introduction

In this Chapter, the application of LACE for the functionalization and conjugation of proteins is described. Section 2 discusses the use of LACE with other chemoenzymatic methods for the one-pot heterobifunctionalization of proteins. In Section 3, we describe the introduction of functionalized thioesters for labeling of proteins with biochemical probes and bioorthogonal handles. In Section 4, we demonstrate that LACE can be used to install full-length ubiquitin and ISG15 onto recombinant proteins under native conditions, and build on this strategy to perform general protein–protein conjugation. In Section 5, we discuss the application of LACE for the post-assembly modification of multimeric proteins.

2. Heterobifunctionalization

2.1. Compatibility with other chemoenzymatic methods

LACE provides access to isopeptide conjugates at internal protein sites. We have shown that the small recognition sequence required for modification by Ubc9, which can be placed throughout the protein sequence (see Figure 35), allows for the dual-labeling of GFP-LACE$_{I,C}^{2M}$ for site-specific homobifunctionalization (see Figure 36). The flexibility in terms of LACE tag location also opens up the possibility to introduce additional tags that are employed in other commonly used chemoenzymatic strategies. We have shown that LACE can operate under a range of conditions as determined from the buffer additive screen (see Figure 41), potentially allowing for adjustments to the reaction conditions to match the needs of a second process. The goal of this part of the project was to provide a proof of concept that LACE can be used in combination with two other chemoenzymatic methods, SpyTag/SpyCatcher and sortase, for the specific generation of heterobifunctionalized proteins in a one-pot fashion.

2.2. Dual-isopeptide labeling with SpyCatcher/SpyTag

We first assessed whether LACE could function in combination with the autocatalytic SpyCatcher domain.[96] By installing a LACE tag in SpyCatcher, this variant could serve as a module for dual-isopeptide conjugation. The N-terminal region of SpyCatcher has been shown to be dispensable for its ligation activity with a SpyTag, and was found to be disordered in crystal structures.[226] SpyCatcher variants that contain an AviTag at the N-terminus for chemoenzymatic biotinylation have also been generated.[227] After inspecting this region, we chose to install the minimal LACE tag IKSE by introducing two mutations in the original sequence (LSSE) (Figure 56a). We compared the ligation efficiency of the variant to unmodified SpyCatcher in ligation reactions with a SpyTag fused to maltose-binding protein (SpyTagMBP).[96] We were pleased to find that both SpyCatcher variants showed complete consumption when incubated with

1.0 equivalents of SpyTagMBP after 1 hour, without any noticeable difference in their reactivity (Figure 56b).

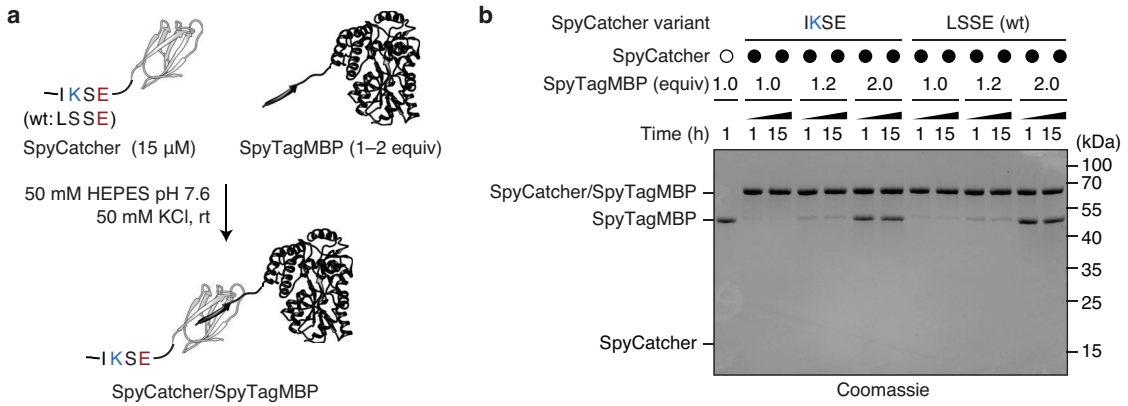


Figure 56. SpyCatcher/SpyTag reactivity of a SpyCatcher variant containing a minimal LACE tag. (**a**) The reactivity of a SpyCatcher variant containing the minimal tag IKSE was compared to the unmodified SpyCatcher (LSSE) in reactions with SpyTagMBP. (Protein images are based on PDB entries 4mli and 1mpd.) (**b**) Coomassie-stained SDS–PAGE analysis of reactions with the indicated SpyCatcher variant and varying amounts of SpyTagMBP after 1 and 15 h.

Next, we attempted to modify the SpyCatcher variant (IKSE) with thioester **5** in the presence of Ubc9. However, rapid precipitation was observed, and analysis by SDS–PAGE showed heterogeneous and non-specific labeling (Figure 57a). Similarly, control variants with an acceptor lysine mutation (IRSE) or the unmodified SpyCatcher were non-specifically labeled with thioester **5**. We attributed these side products to the spontaneous reaction between peptide thioesters and the active site lysine of SpyCatcher, which is poised to undergo a condensation reaction with an aspartic acid residue in SpyTag for isopeptide bond formation (Figure 57b).

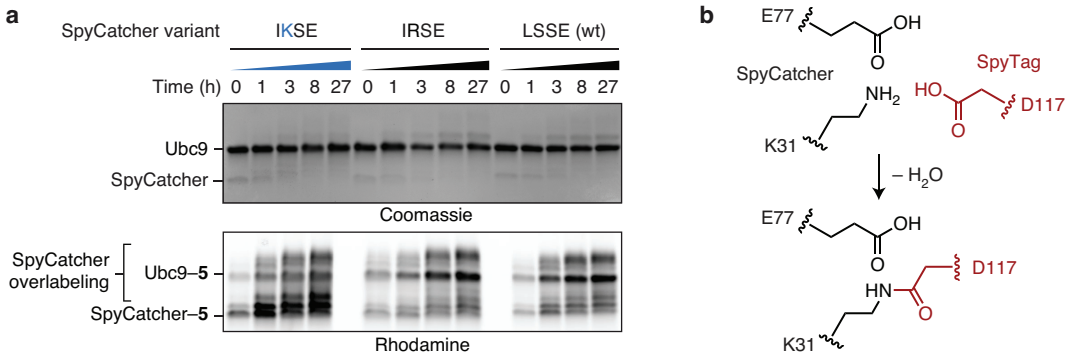


Figure 57. SpyCatcher labeling by Ubc9. (**a**) Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of labeling reactions with thioester **5** and indicated SpyCatcher variants at various time points. (**b**) SpyCatcher active site residues K31 and E77 for isopeptide formation with the SpyTag residue D117 (red). Residue numbers are based on the collagen adhesin domain (CnaB2) from *S. pyogenes*, adapted from Zakeri *et al.*[96]

Instead, we opted to react the SpyCatcher variant first with SpyTagMBP before initiating the LACE reaction. Indeed, this sequential one-pot protocol afforded the desired tripartite isopeptide conjugate (Figure 58).

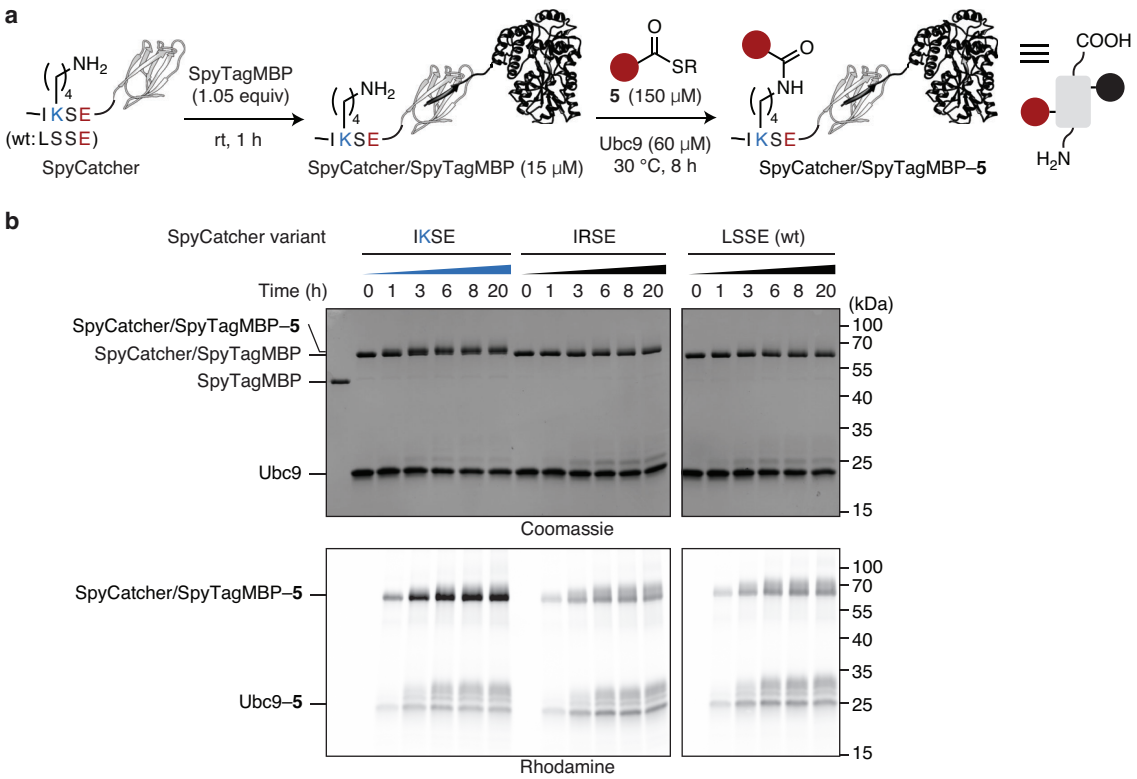


Figure 58. SpyCatcher/SpyTagMBP labeling by Ubc9. (**a**) Reaction of SpyCatcher variants with SpyTagMBP, followed by LACE labeling of the *in situ* generated SpyCatcher/SpyTagMBP with thioester **5**. (**b**) Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of SpyCatcher/SpyTagMBP labeling reactions with thioester **5** at indicated time points. Reactions were carried out with the competent variant (IKSE), or the acceptor lysine mutant (IRSE) and unmodified SpyCatcher (wt).

Since both the reaction with SpyTagMBP and thioester **5** involved modification of side chains, the dual-modified SpyCatcher domain still possesses a free N- and C-terminus. It is therefore straightforward to envision genetic fusion of this SpyCatcher variant to a protein of interest as a modular domain for dual-modification.

2.3. Dual-labeling with sortase

As a second example, we investigated the use of LACE together with the transpeptidase sortase A (SrtA). We prepared a HER2-specific antigen-binding fragment (Fab) (trastuzumab Fab) in which the light chain was equipped with a C-terminal LACE tag^{2M} and the heavy chain C-terminus with the sortag LPETGG (Figure 59a).[228] This Fab was allowed to react with thioester **5** under standard conditions in the presence of Ubc9 and 2 mM of a coumarin-modified triglycine (**23**) (Figure 59b). After 6 hours, the temperature was lowered to 4 °C and the sortase variant

SrtA^{7M} (ref. [229]) was added. This sequential addition was used to suppress hydrolysis of the sortase–acyl intermediate, which was observed when both enzymes were used at 30 °C. Following another 6 hours at 4 °C, analysis by reducing SDS–PAGE showed specific labeling of the respective Fab chains with either thioester **5** or the sortase-reactive probe **23** (Figure 59c).

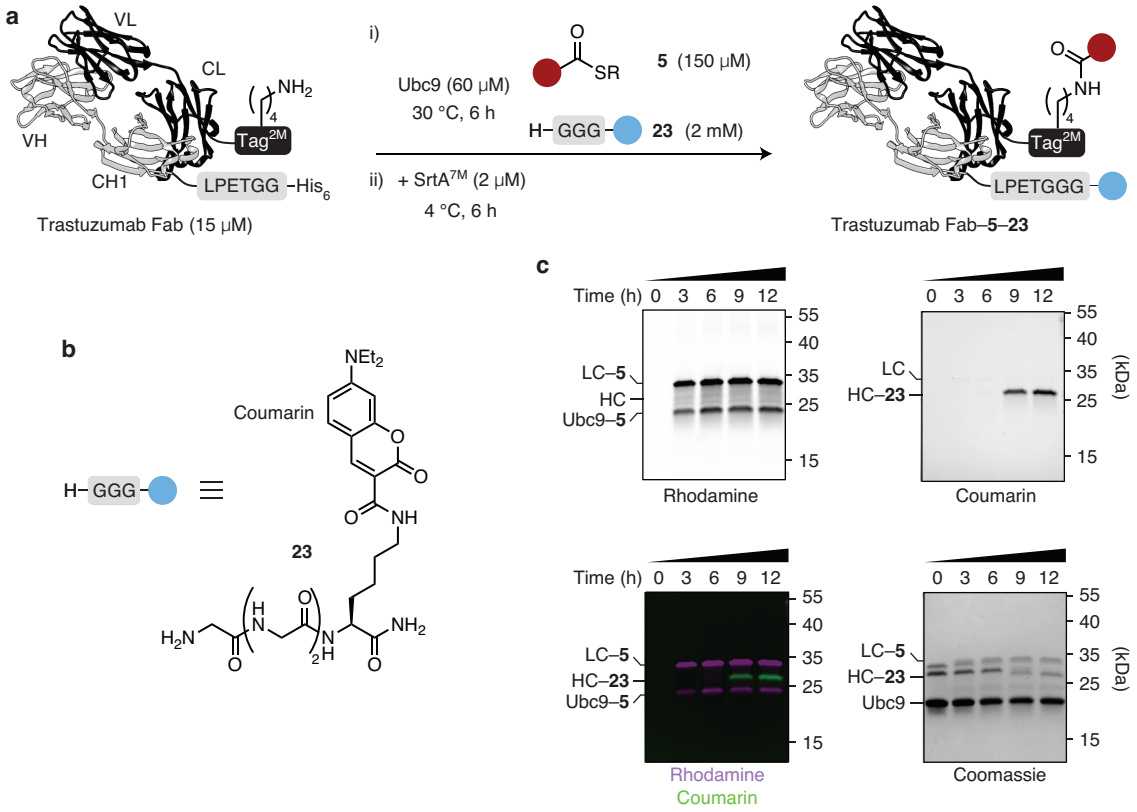


Figure 59. One-pot dual-labeling of trastuzumab Fab. (**a**) Labeling of trastuzumab Fab with a LACE tag on the light chain (LC) and the sortag LPETGG on the heavy chain (HC) with thioester **5** (red) and glycine coumarin probe **23** (light blue) in the presence of Ubc9 and SrtA^{7M}. (The protein image is based on PDB entry 1n8z.) (**b**) Structure of the sortase-reactive coumarin-glycine probe **23**. (**c**) Reducing SDS–PAGE analysis of the reaction at indicated times, visualized by in-gel fluorescence to detect the presence of rhodamine from thioester **5** and coumarin from sortase probe **23**, and by Coomassie stain.

Following direct purification of the reaction mixture by size-exclusion chromatography to isolate the Fab fraction, the intact Fab products were analyzed by ESI–MS (Figure 60a). Similarly, the individual chains were characterized by reducing the interchain disulfide bond prior to ESI–MS analysis (Figure 60b). In both MS analyses, a small amount of unmodified light chain was observed, indicating that the LACE reaction did not proceed to completion. Nevertheless, we were pleased to observe clean and near quantitatively dual-modified Fab.

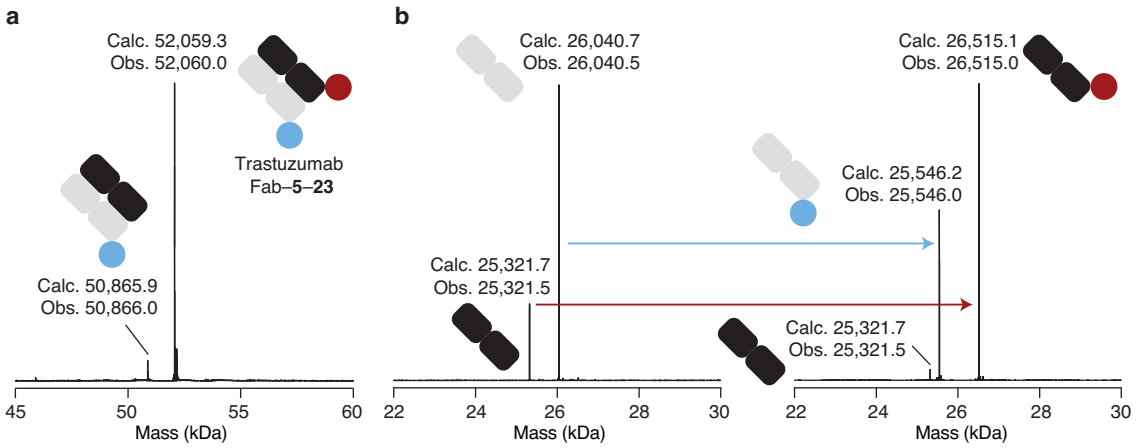


Figure 60. MS characterization. (**a**) Deconvoluted ESI–MS of the intact Fab products (calc., calculated; obs., observed). (**b**) Deconvoluted ESI–MS of unmodified (left) and modified (right) Fab after treatment of the sample with DTT to reduce the interchain disulfide bond. Signals correspond to the light chain (black), heavy chain (grey), and the coumarin- and rhodamine-modified products (light blue and red, respectively).

We confirmed by flow cytometry that the heterofunctionalized immunoconjugate retained the expected specificity for the HER2-positive cell lines SKOV3 and SKBR3, whereas the HER2-negative cell line MCF7 was not recognized by trastuzumab Fab–**5**–**23** (Figure 61). SKOV3 and SKBR3 both showed a several fold higher signal in the presence of the labeled immunoconjugate compared to free rhodamine, whereas MCF7 did not show fluorescence above background.

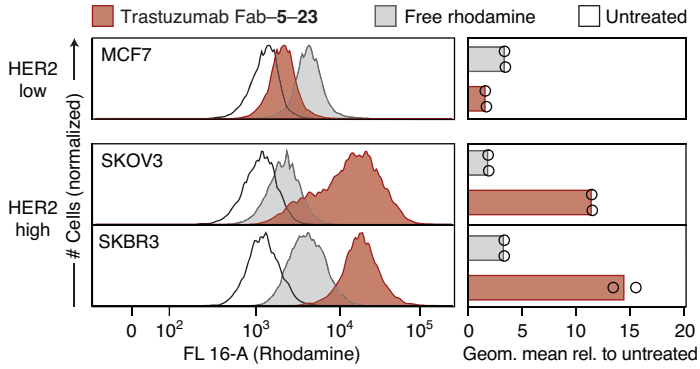


Figure 61. Flow cytometry analysis using the HER2-negative cell line MCF7, and the HER2-positive cell lines SKOV3 and SKBR3. Left: Histogram of singlet live cells versus rhodamine fluorescence (FL 16-A) of samples incubated with dual-modified trastuzumab Fab–**5**–**23** or free rhodamine, or untreated cells. Right: Fold increase in fluorescence (geometric mean) of samples incubated with dual-modified Fab or free rhodamine over untreated cells. Individual data points are shown.

3. Thioesters with functional moieties

3.1. LACE for the introduction of small molecule probes

Using peptide thioesters with a rhodamine at the N-terminus has allowed us to monitor LACE reactions directly by in-gel fluorescence or HER2 binding with the trastuzumab immunoconjugate by flow cytometry (see Figure 61). The goal of this part of the project was to explore other biochemical and biophysical probes that could be introduced into peptide thioesters for transfer by Ubc9 (Figure 62a). Furthermore, the introduction of bioorthogonal handles for subsequent conjugation with a second macromolecule, such as a polymer or a second protein bearing the corresponding reaction handle, is discussed (Figure 62b).

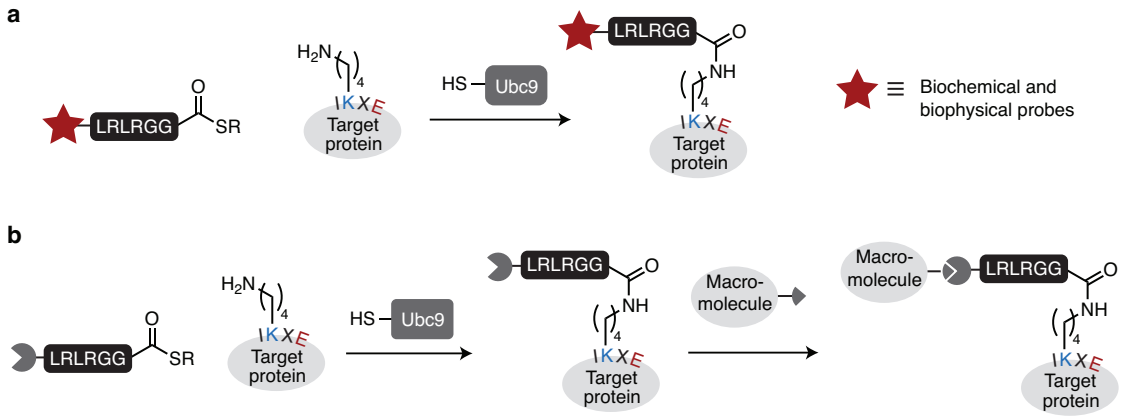


Figure 62. LACE for the installation of small molecule probes (**a**) and bioorthogonal handles for subsequent bioconjugation (**b**) on recombinant proteins.

Functionalized peptide thioesters are attractive probes in this respect due to their ease of synthesis. Peptide synthesis offers a modular synthetic approach, and several efficient protocols for the preparation of C-terminal peptide thioesters by Fmoc-based solid-phase peptide synthesis (SPPS) have been described.[230–233] A versatile route proceeds via resin-supported synthesis of acyl hydrazides, which can be converted to the thioester by *in situ* activation with sodium nitrite and thiolysis.[230]

With these SPPS tools available, there are two general strategies to access functionalized thioesters (Figure 63). On the one hand, functional groups can be directly installed on the resin-supported peptide hydrazide using standard coupling procedures.[234] Secondly, because the ubiquitin-derived hexapeptide thioester is devoid of nucleophilic groups besides the N-terminal α-amine, solution coupling can be employed to introduce the functional moiety onto the N-terminus of preformed peptide thioesters.

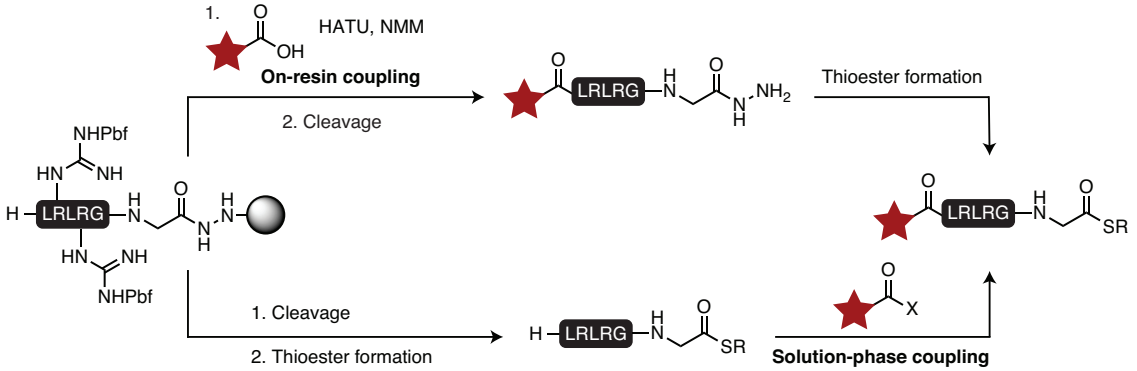


Figure 63. On-resin coupling (top) or solution-phase coupling (bottom) strategies for the preparation of ubiquitin-derived peptide thioesters with N-terminal functional moieties.

3.2. Biotin affinity handles

As a first test case for the tolerance of Ubc9 towards probes other than rhodamine, we investigated the transfer of biotin. A previously reported method for biotinylation of recombinant proteins relies on a 15-residue AviTag, which is recognized and modified by the *E. coli* biotin ligase (BirA).[235,236] Biotinylation by BirA was used for example for the high-affinity pull-down of a protein of interest from cell lysate,[237] the detection of phage particles,[238] or for imaging of biotinylated proteins on the cell surface of mammalian cells.[80]

We prepared biotin thioester **24** or biotin-PEG thioester **25** (Figure 64a) by standard amide bond coupling on resin, followed by cleavage and thioester formation (see Experimental Part for details). The biotin thioesters were tested for their reactivity in Ubc9-mediated labeling reactions with titin-LACE$_C^{2M}$ under standard conditions. Gratifyingly, complete and clean conversion was observed with a reaction half-life of 1.8 and 2.4 hours, respectively (Figure 64b,c).

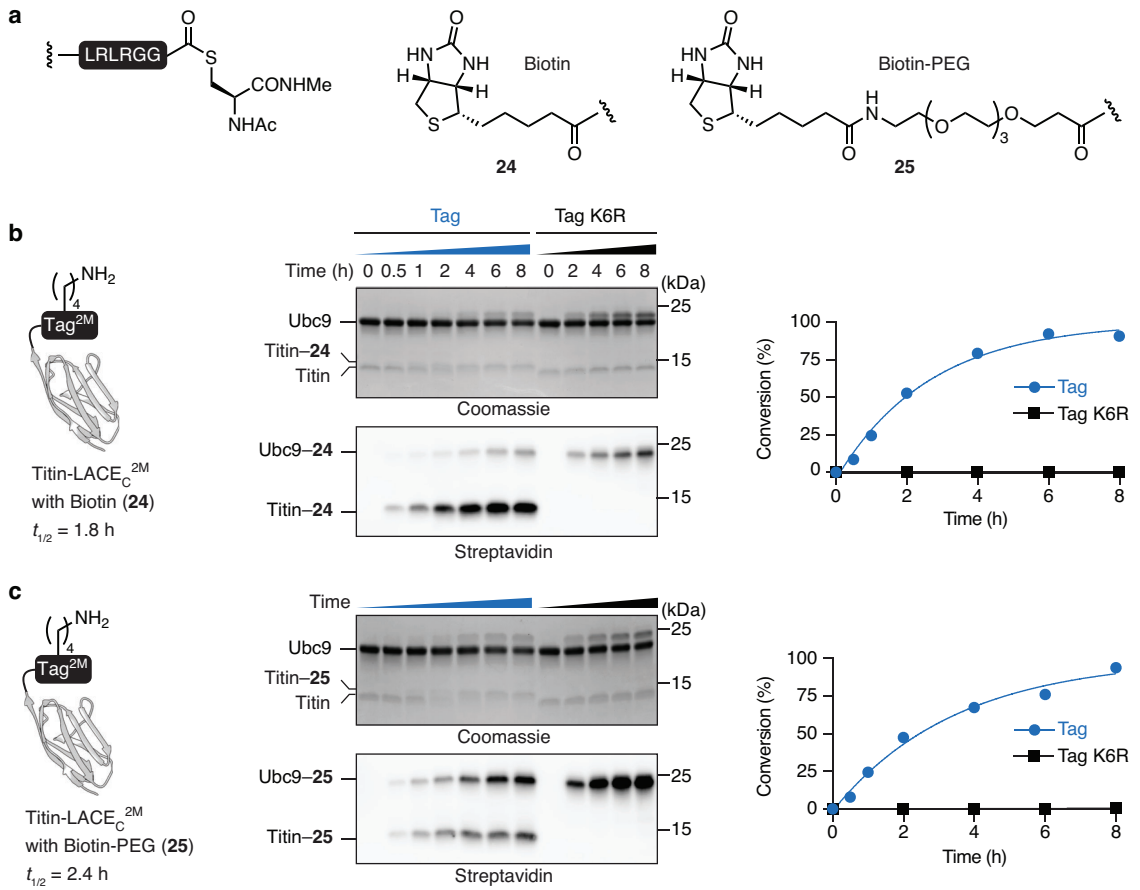


Figure 64. Installation of affinity handles. (**a**) Structures of biotin thioester **24** and biotin-PEG thioester **25**. (**b**,**c**) Coomassie-stained SDS–PAGE analysis (top) and western blot using streptavidin (bottom) of titin-LACE$_C^{2M}$ labeling reactions with thioester **24** (**b**) and thioester **25** (**c**) at indicated time points. Reactions were carried out with the competent substrates or the acceptor lysine mutant K6R. Quantification of titin–peptide formation over time as determined by densitometry of the streptavidin blot is shown for each variant, as well as the apparent $t_{1/2}$.

Besides modification of purified proteins, the AviTag has also been shown to be modified in bacterial[238] and eukaryotic cells[239,240] in the presence of overexpressed BirA, or on the surface of mammalian cells with BirA added to the medium.[80] *In vivo* application represents an advantage of the AviTag over the LACE tag, which is currently restricted to *in vitro* modification. However, the reactivity of BirA is strongly influenced by the substrate concentration and buffer conditions. BirA reactivity is negatively affected by buffer components such as glycerol and salt, and requires ATP and up to 10 mM Mg^{2+} at a reaction pH above pH 8 for useful conversion.[241] With the smaller tag size, the tolerance towards various buffer conditions, and without the need for a specific cofactor or buffer additive, biotinylation by LACE therefore represents a useful addition to the toolbox.

3.3. Bioorthogonal handles

3.3.1. Azides and alkynes

A widely applied bioorthogonal reaction for bioconjugation is the 1,3-dipolar cycloaddition between azides and alkynes,[242] the prototypical click reaction.[243] The cycloaddition of azides and terminal alkynes occurs under copper(I)-catalyzed conditions with high specificity and useful reaction rates for bioconjugation.[244,245] Alternatively, strained alkynes such as cyclooctyne groups were found to undergo rapid cycloaddition with azides under copper-free conditions.[246–248]

Here, we assessed introduction of azide and alkyne click handles to proteins by LACE. The azide thioester **28** could be prepared by solution coupling of the *N*-hydroxysuccinimide (NHS) ester **27** to the pre-formed thioester **26** (Figure 65a). Analogously, dibenzocyclooctyne (DBCO) thioester **30** was obtained by allowing the pre-formed thioester **29** to react with a commercially available DBCO-NHS ester (Figure 65b).

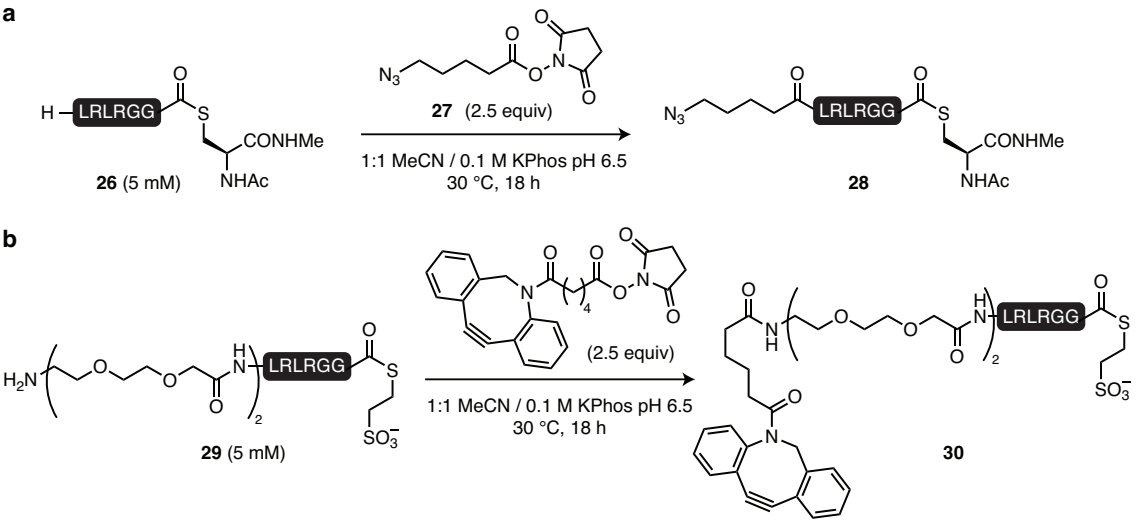


Figure 65. Synthesis of azide thioester **28** (**a**) and DBCO thioester **30** (**b**).

With the click-functionalized thioesters in hand, we tested the Ubc9-mediated transfer of DBCO thioester **30** to GFP-LACE$_C$. Efficient and clean labeling was observed within approximately 3 hours under standard conditions (Figure 66). For a successful application of azide thioester **28** transfer by Ubc9, see Section 5.2.

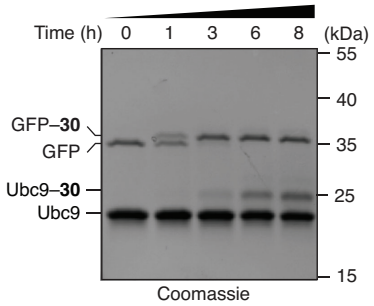


Figure 66. Coomassie-stained SDS–PAGE analysis of GFP-LACE$_C$ labeling with DBCO thioester **30** at indicated time points.

Introduction of the requisite azide or alkyne functional groups into proteins has been achieved previously, for example, by UAA incorporation using genetic code expansion techniques[249,250] or by metabolic incorporation of methionine analogs.[251,252] Alternatively, several chemoenzymatic methods have been used for the posttranslational modification of proteins with click handles, for instance using lipoic acid ligase,[253] or the *N*-myristoyl[86] and farnesyl transferases.[254] However, chemical synthesis of metabolite analogs is not always trivial, and tolerance of these substrate analog for transfer by the enzyme can become a bottleneck. In comparison, the facile preparation of the click-functionalized peptide thioesters **28** and **29**, their efficient transfer by Ubc9, and the reliance solely on canonical amino acids makes LACE an attractive choice for labeling of recombinant proteins with click handles.

3.3.2. KATs and hydroxylamines

3.3.2.1. The potassium acyltrifluoroborate (KAT) ligation

Over the past years, the Bode group has developed an amide-forming ligation between potassium acyltrifluoroborates (KATs)[255] and hydroxylamines (Figure 67).[256] The ligation proceeds with a second order rate constant of up to 20 M^{-1} s^{-1} under aqueous acidic conditions without the need for any additive or catalyst.[257,258]

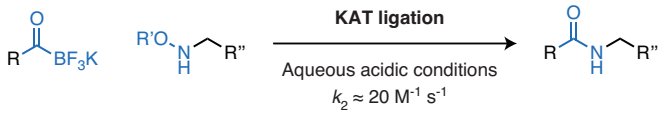


Figure 67. Potassium acyltrifluoroborate (KAT) ligation.

The KAT ligation has been applied to the synthesis of biocompatible PEG hydrogels[259–261] or the post-polymerization functionalization and conjugation of polymers.[262] KAT groups have also been incorporated into small synthetic peptides.[263] While hydroxylamines have been successfully incorporated into peptides[257] and synthetic proteins[264] by SPPS, a current limitation is the lack of suitable methods for the direct introduction of KAT moieties and hydroxylamines into recombinant proteins for bioconjugation reactions.

For hydroxylamines, White *et al.*[265] reported a two-step method in which a cysteine-reactive methylsulfonephenyl-oxadiazole reagent was used to install hydroxylamines on recombinant proteins, which then underwent efficient modification with PEG-KATs at near equimolar ratios for site-specific PEGylation (Figure 68). Similarly, hydroxylamines were installed on a surface-exposed cysteine using maleimides for subsequent radiolabeling with ^{18}F-containing KAT reagents.[266]

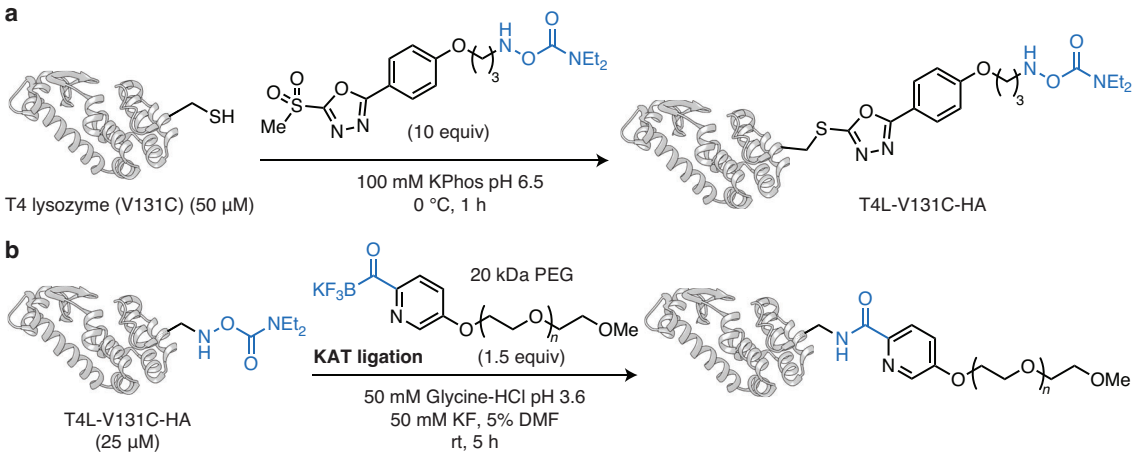


Figure 68. Protein PEGylation using the KAT ligation as reported by White *et al.*[265] (**a**) Installation of hydroxylamine via a surface-exposed cysteine on recombinant proteins. (**b**) PEGylation using near-equimolar KAT-PEG-reagents under dilute aqueous conditions.

A drawback of both of these approaches for the installation of hydroxylamines is the requirement for degassed buffers and treatment of the cysteine-containing protein with reducing agents prior to labeling. Installation of KAT groups on proteins has not been reported to date. In this part of the project, we discuss the use of LACE to functionalize recombinant proteins with hydroxylamines and KATs for subsequent bioconjugation applications.

3.3.2.2. Hydroxylamines and comparison of photolabile protecting groups

For the synthesis of hydroxylamine-bearing peptide thioesters, we installed *O*-diethylcarbamoyl-hydroxylamines at the N-terminus of the resin-supported peptide hydrazides by standard SPPS coupling as previously reported for other hydroxylamines.[257,264] By using *N*-protected hydroxylamines to prevent decomposition of the hydroxylamine during the oxidative step of the thioester formation procedure,[230] hydroxylamine thioesters **31a** and **31b** were obtained, bearing either a photolabile 2,2-bis(2-nitrophenyl)ethyl carbamoyl (**31a**) or 1-(2-nitrophenyl)ethyl carbamoyl (**31b**) group (Figure 69a).

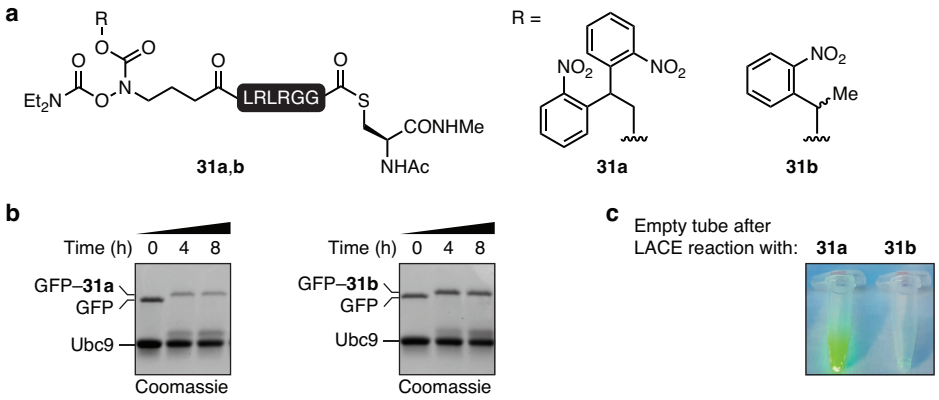


Figure 69. (**a**) Structures of hydroxylamine thioesters **31a** and **31b**. (**b**) Coomassie-stained SDS–PAGE analysis of GFP-LACE$_C$ labeling reactions with the indicated thioesters over time. (**c**) Image of the reaction tubes under a handheld UV lamp (365 nm) after completion of the time-course and removal of the sample solution.

We compared the two hydroxylamine thioesters in GFP-LACE$_C$ labeling reactions. Although complete labeling of GFP-LACE$_C$ was observed with both variants after 8 hours, the GFP-LACE$_C$ conjugate with thioester **31a** appeared to be unstable in solution over the course of the reaction as judged by the reduced SDS–PAGE band intensity (Figure 69b). Inspection of the tubes after completion of the reaction and removal of the solutions showed strong GFP fluorescence with this sample, indicating loss of GFP–**31a** due to adsorption. This is likely due to the increased hydrophobicity of the bis-2-nitrophenyl group in **31a** compared to the photolabile group **31b**.

Repeating the labeling reactions with GFP-LACE$_I$, we noticed that this variant showed improved stability with thioester **31a** compared to GFP-LACE$_C$, and likewise underwent complete labeling with both variants within 8 hours (Figure 70, top). With these GFP–hydroxylamine conjugates in hand, we next attempted photodeprotection and PEGylation by KAT ligation (Figure 70, bottom). The samples were desalted to remove excess peptide thioester and exchanged to an acidic KAT ligation buffer.[265] Following ultraviolet (UV) light irradiation at 365 nm for 1 hour in the presence of 2-pyridyl PEG-KAT **32**,[265] the desired GFP–PEG conjugate was observed. Somewhat to our disappointment, no further improvement of the conversion was observed after 15 hours, even though the ligation was performed with an excess of PEG-KAT. Similar experiments with other GFP and titin variants did not afford fully PEGylated protein conjugates either.

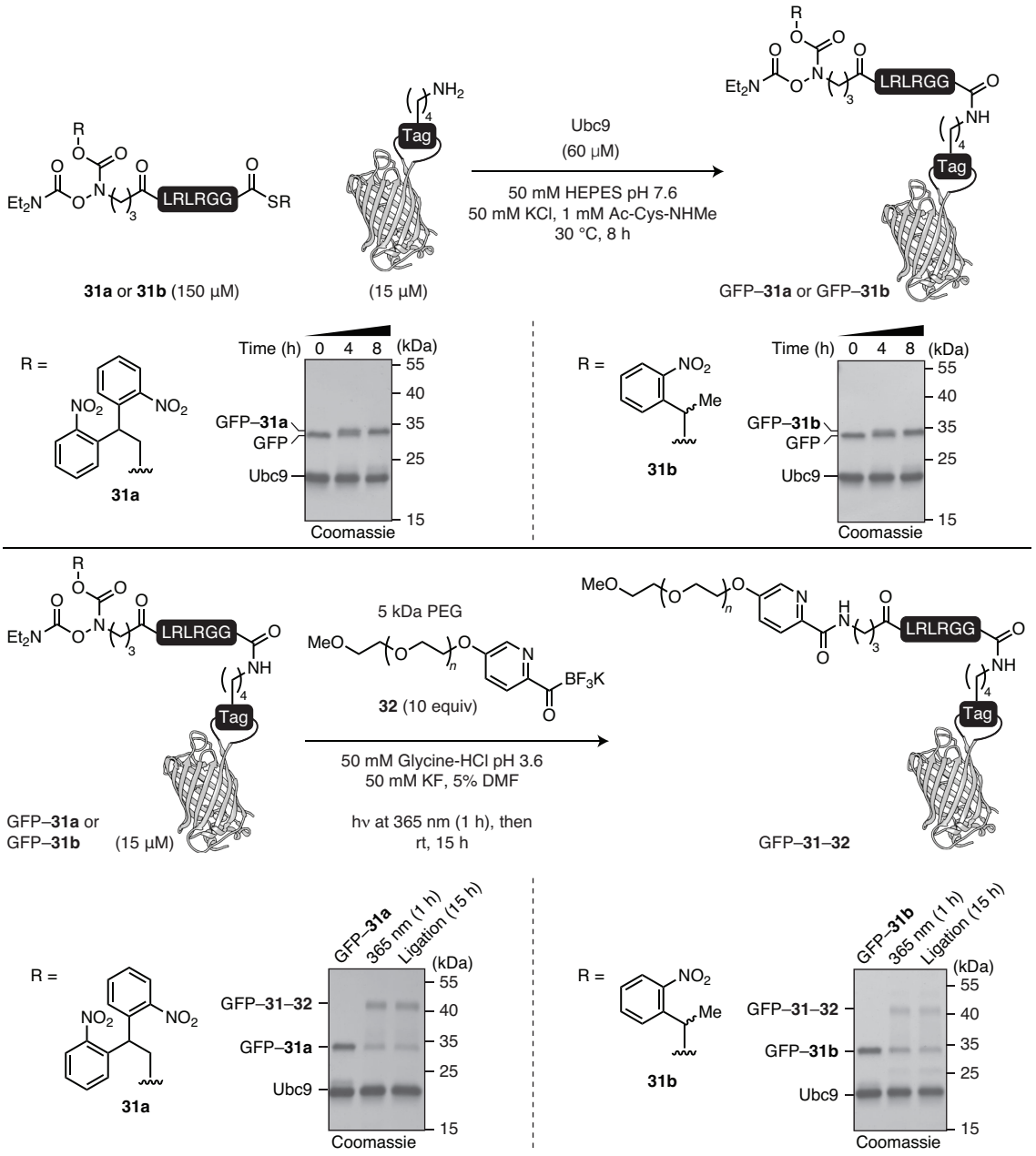


Figure 70. Top: GFP-LACE$_I$ labeling with hydroxylamine thioesters **31a** and **31b**. Coomassie-stained SDS–PAGE analysis at indicated times is shown for each variant. Bottom: *In situ* photodeprotection and KAT ligation with 5 kDa PEG-KAT (**32**). Coomassie-stained SDS–PAGE analyses of the GFP–hydroxylamine intermediate, and of the reactions after 1 h UV irradiation at 365 nm and 15 h ligation in the presence of PEG-KAT are shown.

We and others in our group occasionally observed that *N*-carbamoyl-protected and unprotected linear hydroxylamines underwent slow decomposition, likely to an aldehyde species. This decomposition occurs especially under neutral or basic aqueous conditions. We therefore attribute the incomplete PEGylation to an apparent instability of the hydroxylamines, either during the LACE reaction or subsequently during photodeprotection and KAT ligation. This is in contrast to previous reports that relied on cysteine-reactive hydroxylamine probes, where full consumption

of the installed hydroxylamines has been observed.[265] Compared to cysteine labeling, one difference is the prolonged reaction time as well as the higher basicity of the reaction buffer during the LACE reaction which could explain the observed hydroxylamine instability.

Nevertheless, these results represent the first installation of hydroxylamines to recombinant proteins, without the need for cysteine conjugation, for subsequent PEGylation by KAT ligation at micromolar concentrations. Building on these results, future directions should include the development of peptide thioesters with more stable hydroxylamine variants and improvement of the LACE reaction rate.

3.3.2.3. KATs

To introduce KAT functionalities into peptide thioesters, we first considered an S_N2-type reaction between thiols and benzyl bromide to install trifluoroborate iminium (TIM) **33** (Figure 71).[262] TIMs have been shown to readily hydrolyze to the free KAT under aqueous neutral to basic in the presence of a potassium source.[267,268] We attempted to prepare peptide thioester **35a** containing an N-terminal cysteine. However, the cyclic peptide **35b** was observed as the major product after thioesterification of hydrazide **34**, stemming from intramolecular NCL. Next, we tried to prevent the irreversible amide formation by protecting the N-terminal α-amine with an acetyl group (**36**). However, cyclic thioester **37b** and dimerized thioester **37c** were observed as the major products instead of the linear thioester **37a**.

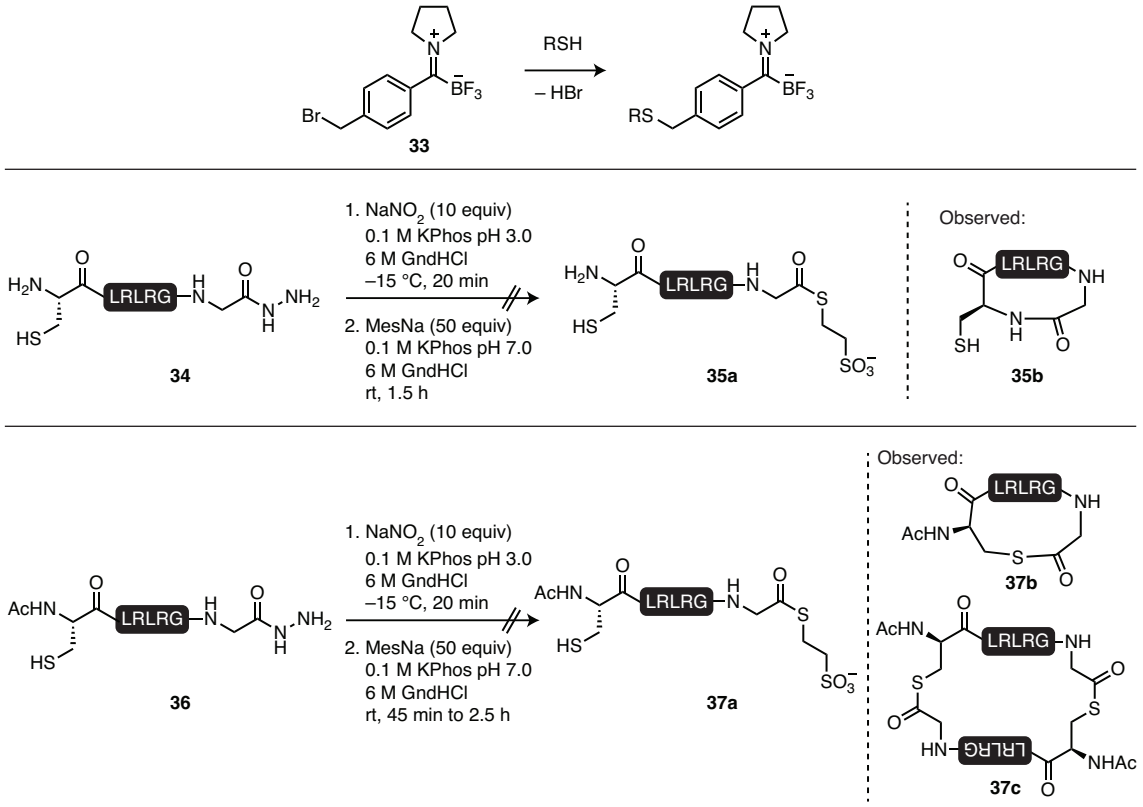


Figure 71. Attempts at preparing cysteine-containing thioesters **35a** and **37a** for subsequent installation of KAT functionalities.

Next, we investigated solution-phase coupling of TIM pentafluorophenyl ester **38** with the free N-terminus of thioester **26**. Clean conversion to the desired TIM thioester **39** was observed.

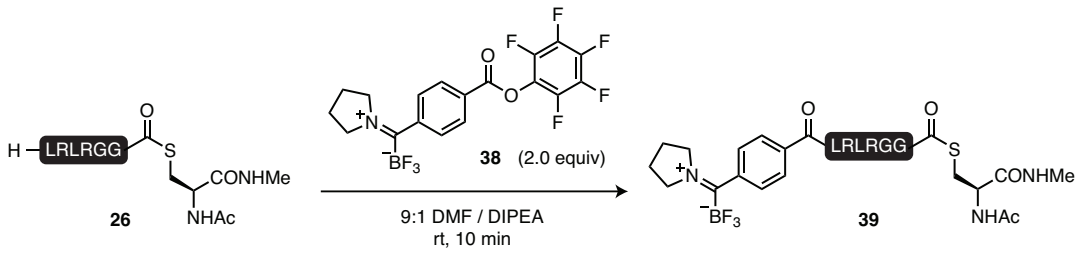


Figure 72. Synthesis of TIM thioester **39**.

With the TIM thioester in hand, we tested the Ubc9-mediated transfer of thioester **39** to GFP-LACE$_C$ (Figure 73). We were pleased to observe that the reaction proceeded to completion under standard conditions. To test the subsequent KAT ligation, we removed the excess thioester and exchanged the reaction buffer to the KAT ligation buffer by desalting. Upon addition of the rhodamine hydroxylamine **40**,[262] a second band shift was observed by SDS–PAGE to give a clean new GFP product. In-gel fluorescence confirmed the incorporation of rhodamine.

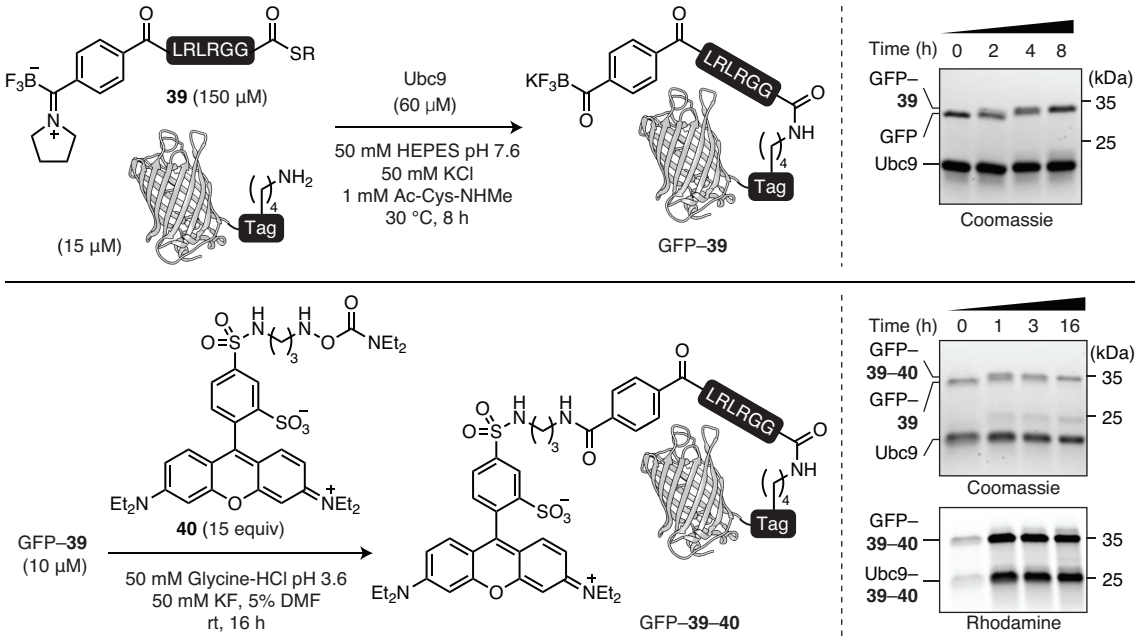


Figure 73. Top: GFP-LACE$_C$ labeling with TIM thioesters **39** with concomitant hydrolysis of the iminium to the KAT under the standard LACE reaction conditions. Bottom: Ligation of GFP–**39** with rhodamine hydroxylamine **40**. For each step, Coomassie-stained SDS–PAGE analyses and in-gel fluorescence (bottom) are shown at indicated times.

Generally, the GFP–KAT conjugates were more stable than GFP–hydroxylamine products, both in terms of product solubility and reactivity in KAT ligations. Moreover, the TIM appeared to conveniently hydrolyze under the LACE conditions to reform the KAT, as shown by the complete consumption of GFP–**39** in the presence of an excess of hydroxylamine **40**. LACE can therefore be used to directly install KAT groups on fully recombinant proteins, significantly broadening the scope of biomolecules that can be used in KAT bioconjugations.

3.3.2.4. Application of KAT transfer by LACE to the conjugation of cytokines

As a test case for the combination of LACE and KAT ligation, we investigated protein–protein conjugation between interleukin (IL)-4 and IL-13 for the control of neutrophils in psoriasis disease. IL-4 and IL-13 are two related cytokines that regulate inflammation.[269] Neutrophils are one of the first cell types recruited to a site of inflammation and have an important function in warding off bacterial and fungal infections.[270] However, misregulation of neutrophils has also been implicated in various autoimmune diseases.[271] Neutrophils induce the formation of mast cell and neutrophil extracellular traps (NETs) in the chronic systemic inflammatory disorder psoriasis, causing an inflammation-boosting loop and disease.[272–274]

Recently, it has been described that IL-4 signaling reduces the expansion and migration of neutrophils.[271] IL-4 signaling can proceed via two receptor complexes, type I and type II IL-4 receptor (IL-4R),[275] both of which lead to phosphorylation of signal transducer and activator of transcription 6 (STAT6) (Figure 74a).[271] The type I IL-4R is formed by heterodimerization between

IL-4Rα and the γ-chain (γc), which is commonly expressed in hematopoietic cells including B and T cells.[276] In contrast, the type II IL-4R is formed between IL-4Rα and IL-13Rα1, and is mostly found in non-hematopoietic cells, macrophages and neutrophils.[271,275,276] Type II IL-4R signaling can also be induced by IL-13, by engaging IL-13Rα1 with high affinity followed by binding of IL-4Rα.[275] We therefore envisioned that an IL-4–IL-13 conjugate could specifically engage the type II IL-4R found in neutrophils by favoring interaction with both IL-4Rα and IL-13Rα1 (Figure 74b), without inducing type I IL-4R signaling in other hematopoietic cells.

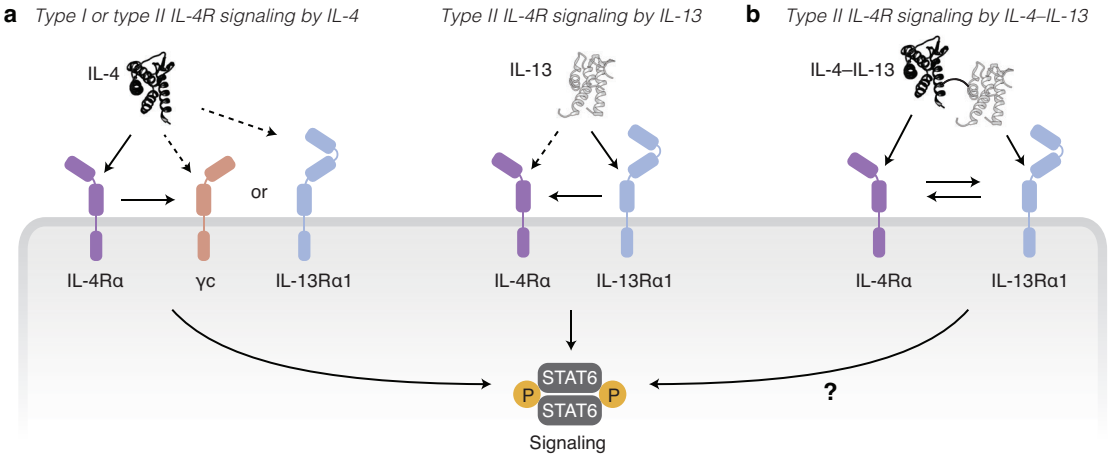


Figure 74. IL-4 receptor signaling. (**a**) Il-4 and IL-13 signaling via type I and type II IL-4R. (**b**) Proposed IL-4 signaling specifically via type II IL-4R by an IL-4–IL-13 conjugate. (Protein images are based on PDB entries 1hzi and 3bpo.)

Inspection of the IL–receptor interactions[275] led us to install an isopeptide-linked LACE tag^{2M} on the sidechain of IL-13-R66K (IL-13-LACE^{2M}), and a hydroxylamine in IL-4 by substituting Q116 with an ornithine-hydroxylamine variant with a photolabile protecting group[264] (IL-4-HA-PG) to create the linkage between the two cytokines. Both ILs were prepared synthetically and folded (M. Ninomiya, unpublished results).

With these materials in hand, we tested labeling of IL-13-LACE^{2M} with TIM thioester **39** (Figure 75). To preserve the native disulfide bonds of IL-13-LACE^{2M}, we omitted reducing agents such as the thiol Ac-Cys-NHMe (**1**) in the reaction buffer. LACE reactions over several hours resulted in loss of product due to precipitation, whereas the starting material by itself was stable in the reaction buffer. We therefore opted to reduce the reaction time to 2 hours under otherwise standard conditions. Direct purification of the reaction mixture by preparative reversed-phase high-performance liquid chromatography (RP–HPLC) afforded a mixture of starting material and IL-13-LACE^{2M}–**39** (approximately 3:1 as judged by MALDI–MS). Following exchange of the solvent to the KAT ligation buffer, we initiated the KAT ligation by UV irradiation in the presence of IL-4-HA-PG. Reaction monitoring by analytical RP–HPLC showed consumption of the KAT-bearing IL-13-LACE^{2M}–**39** and formation of the desired conjugate IL-4–IL-13 (Figure 75a,b).

Following purification of IL-4–IL-13 by RP–HPLC, formation of the desired conjugate was confirmed by SDS–PAGE analysis (Figure 75c) and MALDI–MS.

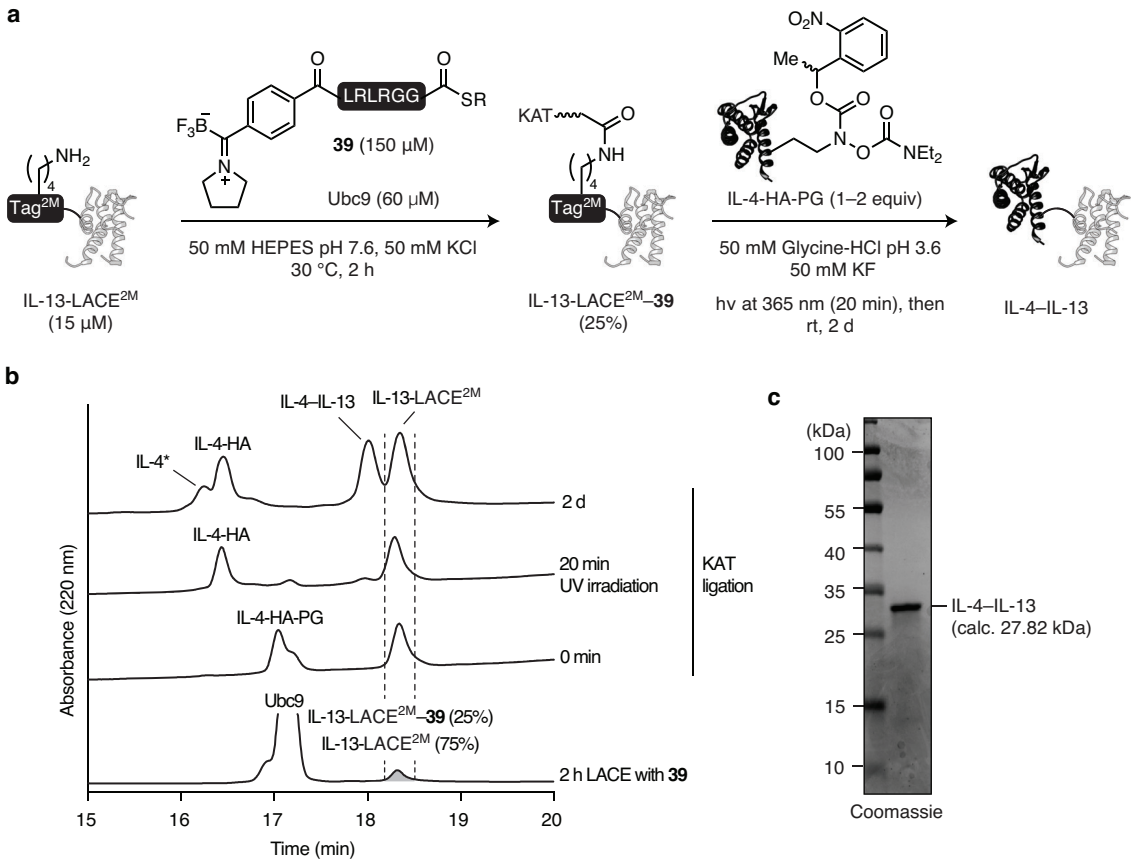


Figure 75. Preparation of IL-4–IL-13 conjugate. (**a**) LACE reaction of IL-13-LACE^{2M} with TIM thioester **39**, followed by KAT ligation with IL-4-HA. (**b**) Analytical RP–HPLC analysis of the crude LACE reaction after 2 h (bottom), and of the KAT ligation at indicated time points (top). IL-4-HA-PG was converted to IL-4-HA by removal of the photolabile protecting group under UV light (365 nm). An aldehyde decomposition product of IL-4-HA (IL-4*) was observed by MALDI–MS after prolonged reaction time. (**c**) Coomassie-stained SDS–PAGE analysis of the purified IL-4–IL-13 conjugate. Analytical RP–HPLC: Courtesy of M. Ninomiya.

The prepared conjugate was then tested in STAT phosphorylation (pSTAT) assays. Gratifyingly, selective activation of IL-4 signaling via pSTAT6 was observed in neutrophils with the IL-4–IL-13 conjugate, whereas T cells and B cells showed at least two orders of magnitude less sensitivity. In contrast, all cell types tested were sensitive to IL-4 alone.

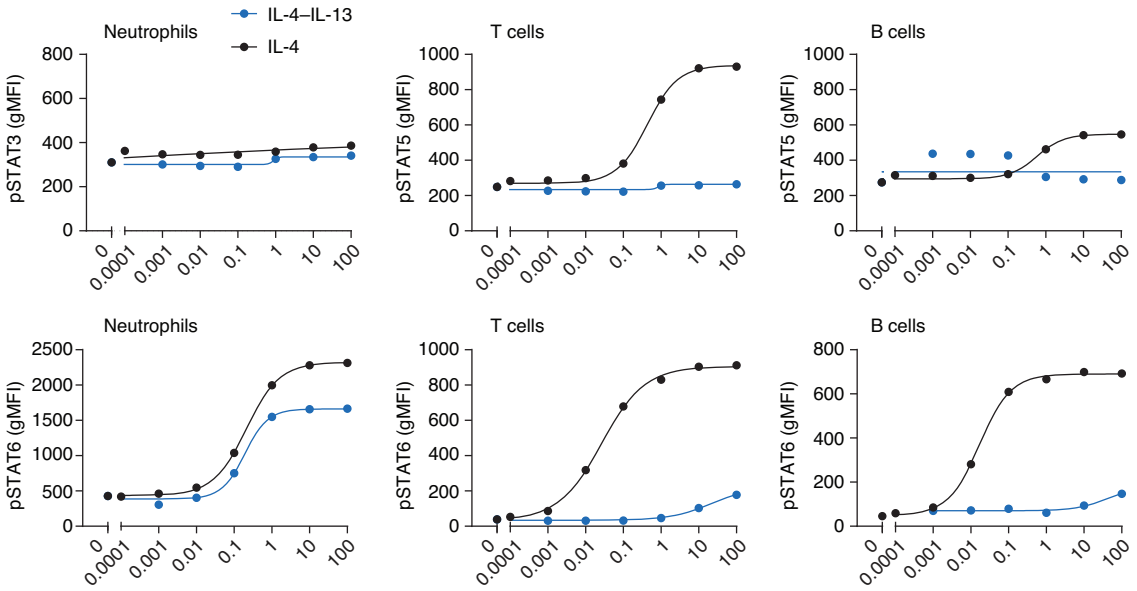


Figure 76. pSTAT assays with IL-4–IL-13 and IL-4 in neutrophils, T cells and B cells. gMFI, geometric mean fluorescence intensity. pSTAT assays: Courtesy of C. Egholm.

Taken together, these results suggest that IL-4–IL-13 conjugates open up a potential new avenue for the selective IL-4 signaling induction in neutrophils. The distinct geometry of the prepared conjugate would not be possible by direct genetic fusion and recombinant techniques, and underscores the value of LACE-mediated transfer of KATs to a specific site in a protein.

4. Protein thioesters

Synthetic peptide thioesters allowed us to conveniently introduce biochemical probes and bioorthogonal handles using LACE, with a broad tolerance for functional groups and moieties. However, thioesters are also accessible via various recombinant pathways,[277,278] for example by thiolysis of the corresponding intein fusion (Figure 77a).[100,279,280] The goal of this part of the project was to explore the use of recombinant thioesters to achieve direct protein–protein conjugation, beyond the introduction of small molecule probes (Figure 77b).

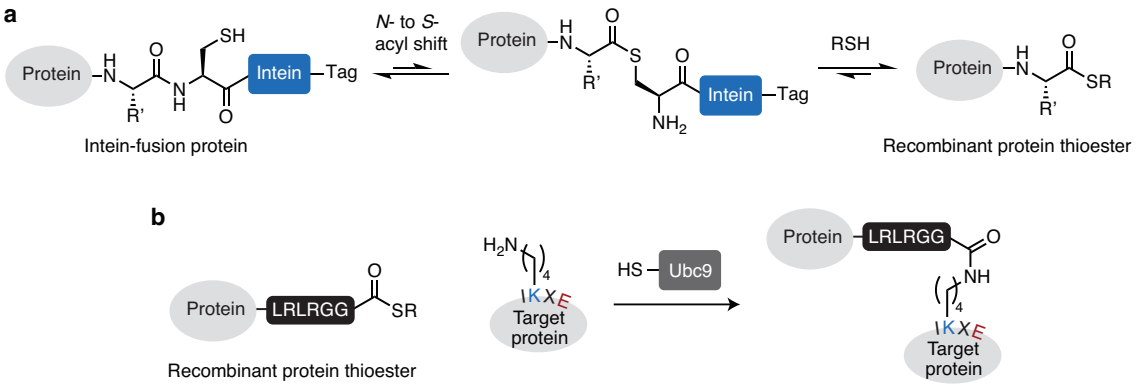


Figure 77. Recombinant thioesters as substrates for LACE. (**a**) Generation of recombinant protein thioesters from the corresponding C-terminal intein fusion. (**b**) Recombinant protein thioesters containing a C-terminal, ubiquitin-derived recognition sequences for Ubc9-mediated protein–protein conjugation with target proteins containing a LACE tag.

4.1. Site-specific installation of Ubls onto target proteins

Because Ubc9 prefers the ubiquitin-derived C-terminal hexapeptide sequence as an acyl donor, the obvious choice was to explore monoubiquitination of protein substrates containing a LACE tag. Methods to prepare ubiquitinated proteins to study their molecular and biochemical effects are very sparse.[281] Specific monoubiquitination has been challenging to achieve enzymatically due to the need for a defined cascade of E1, E2 and E3 enzymes, as well as the potential for undesired ubiquitin chain formation. Using LACE, one could circumvent the need for both the E1 and E3 enzymes. Even if the canonical pathway is known, it can be challenging to produce requisite E3 enzymes recombinantly, in contrast to the E2 enzyme Ubc9 which can be expressed in good yields (see Experimental Part). The two main approaches to date to obtain site-specifically monoubiquitinated proteins are total chemical synthesis[282] and semisynthesis.[283]

Total chemical synthesis of ubiquitinated proteins by SPPS relies on the use of orthogonal protecting groups to elongate the ubiquitin chain via a pre-defined isopeptide bond (Figure 78a).[282] The size of possible products using this strategy is however restricted, due to the limited size of peptide fragments accessible by SPPS. More commonly used semisynthetic methods include the use of recombinantly prepared ubiquitin thioesters from the respective intein fusion, followed by NCL and desulfurization with δ-mercaptolysine-bearing substrates (Figure 78b).

Target proteins bearing the requisite unnatural residue δ-mercaptolysine at a specific position can be accessed by amber codon suppression[284] or prepared by SPPS.[285,286] A limitation of both of these approaches is that the final products have to be refolded, narrowing the scope of ubiquitinated proteins that can be studied in this way.

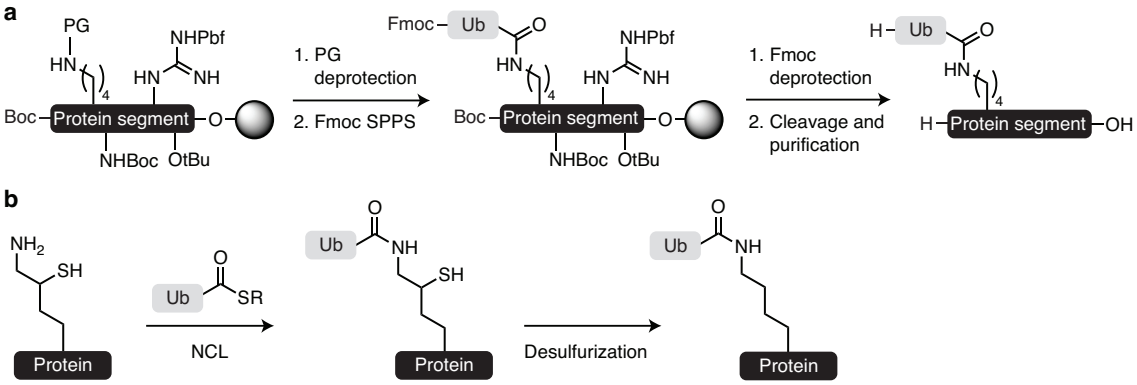


Figure 78. Chemical strategies for the generation of monoubiquitinated proteins, including synthesis by SPPS using orthogonal protecting group strategies (**a**) or NCL between C-terminal ubiquitin thioesters and the UAA δ-mercaptolysine (**b**).

Besides generating authentic ubiquitin conjugates, close chemical analogs have been described, often with the advantage that they are easier to access. NCL between an isopeptide-bonded cysteine and a Ub(1–75) thioester followed by desulfurization circumvents the need for δ-mercaptolysine, but results in a G76A mutation in ubiquitin (Figure 79a).[287] Other examples involve the use of 1,3-dichloroacetone to bridge cysteine-modified target proteins and ubiquitin variants (Figure 79b),[288] or the formation of a disulfide bond (Figure 79c).[26] Although these methods afford an isostere of the native isopeptide bond, the products have been shown to be functionally equivalent in many cases.[26,288]

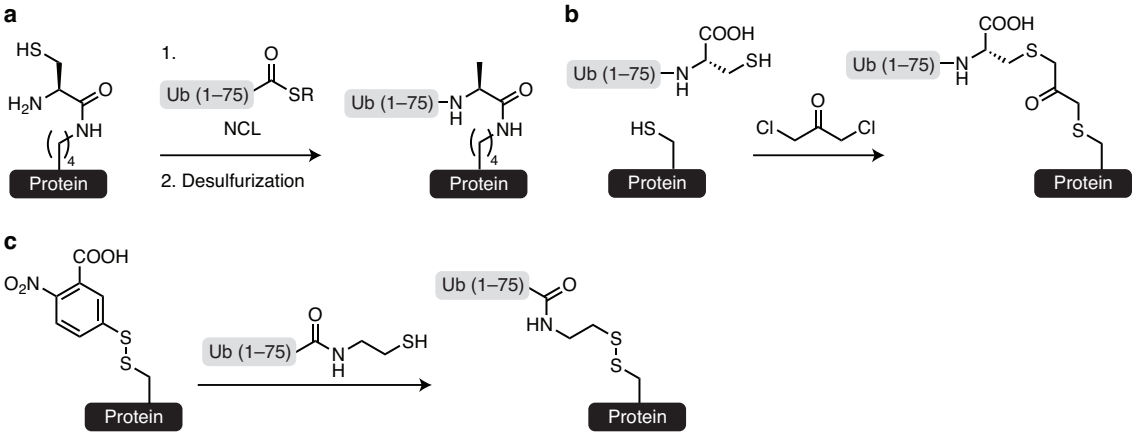


Figure 79. Chemical strategies for the generation of isosteres and close analogs of ubiquitin-conjugates. (**a**) NCL between an isopeptide-linked N-terminal cysteine and a ubiquitin thioester. (**b**,**c**) Bis-thio-acetone linker (**b**) and disulfide bond (**c**) as an analog of the lysine-ubiquitin isopeptide bond.

Recently, two reports described the use of sortase for the conjugation of ubiquitin and SUMO derivatives to substrates bearing a glycylglycine moiety (Figure 80).[289,290] The requisite glycylglycine moiety for recognition by sortase can be introduced either by SPPS[289] or by amber codon suppression with the unnatural residue AzGGK, in which the N-terminus of glycylglycine is masked by an azide moiety (Figure 80).[290] The latter requires conversion of the azido-group to the amine by Staudinger reduction, but offers the added benefit of temporal control over onset of the sortase reaction. A limitation of these methods is the requirement for several mutations in the C-terminal region of the Ubl. Nevertheless, these methods increase the scope of proteins that can be ubiquitinated because the sortase-mediated modification can occur on folded proteins, and has also been demonstrated *in vivo*.[290]

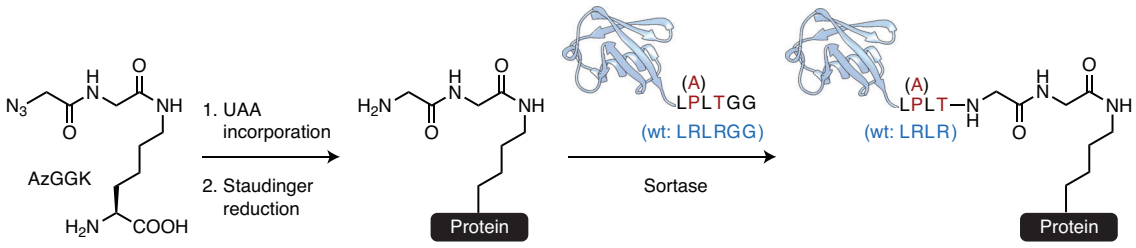


Figure 80. Sortase-mediated conjugation of ubiquitin variants onto glycylglycine-modified substrate proteins, adapted from Fottner *et al.*[290]

4.1.1. Ubiquitination and ISG15ylation of proteins by LACE

4.1.1.1. Proof of concept

We prepared full-length ubiquitin-Mes thioester **41** (Ub-Mes) by thiolysis of the corresponding *Mycobacterium xenopi* DNA gyrase subunit A (Mxe GyrA) intein fusion.[291] We were pleased to find that reacting GFP-LACE$_I$ with Ub-Mes under standard conditions resulted in 82% conversion to the monoubiquitinated GFP after 24 hours (Figure 81). Ubiquitination occurred specifically at the LACE tag and was strictly dependent on Ubc9.

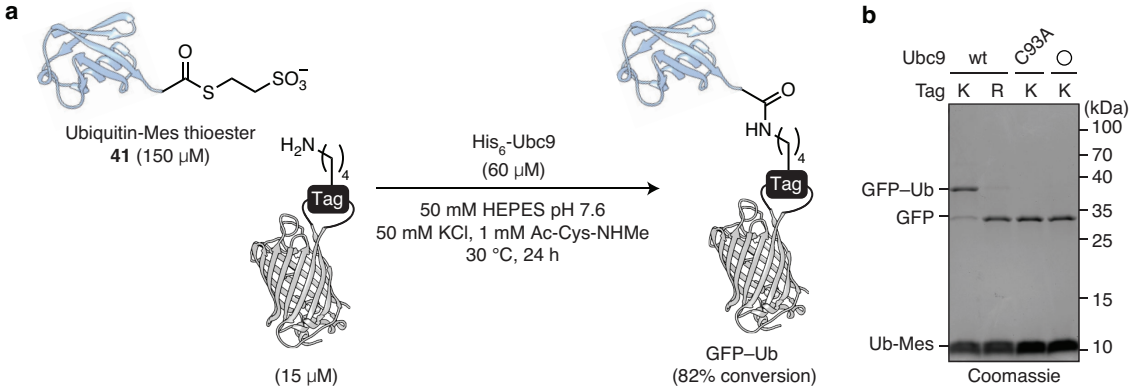


Figure 81. GFP ubiquitination. (**a**) Labeling of GFP-LACE$_I$ with full length ubiquitin-Mes thioester **41**. (**b**) Coomassie-stained SDS–PAGE analysis of reactions with the competent substrate (K) or the acceptor tag

mutant K6R (R). Reactions were carried out in in the presence or absence (hollow circle) of wt His_6-Ubc9 or the active site mutant C93A. His_6-Ubc9 was removed by reverse Ni-NTA purification prior to analysis.

4.1.1.2. Monoubiquitination of SUMO2 and α-synuclein

LACE only requires two specific residues near a desired acceptor lysine, as seen from the mutational analysis of the tag (see Figure 34d) as well as the labeling of substrates containing a minimal LACE tag (see Figure 37 and Figure 58). Apart from this sequence requirement, Ubc9 does not otherwise rely on distinct structural features of the substrate, such as a specific secondary structure or recognition domain. We therefore hypothesized that monoubiquitination of target proteins might be achieved with a minimal four-residue LACE tag (Figure 82a).

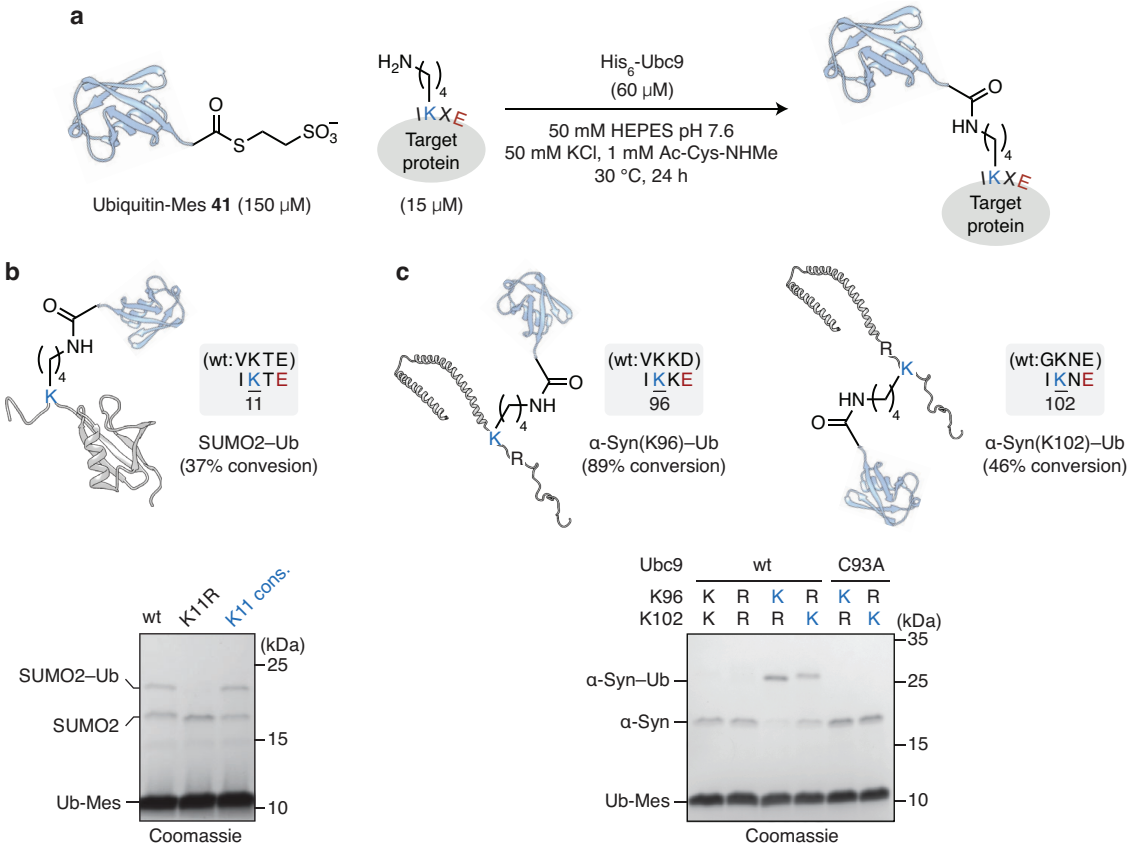


Figure 82. Monoubiquitination by LACE. (**a**) Protein substrates with a minimal internal tag IKXE were reacted with full-length Ub-Mes thioester **41**. (**b**) Coomassie-stained SDS–PAGE analysis of reactions with wt SUMO2, the acceptor mutant K11R or the consensus motif variant of K11 (^{10}IKTE^{13}) in the presence of His_6-Ubc9-R13A. (**c**) Coomassie-stained SDS–PAGE analysis of reactions with either wt α-synuclein, the double mutant K96R and K102R, or the competent substrates (blue) with a consensus motif at either K96 (^{95}IKKE^{98}) or K102 (^{101}IKNE^{104}) and the concomitant lysine mutations K102R and K96R, respectively. Control reactions with the competent substrates and the Ubc9 active site mutant C93A were performed. In all samples, His_6-Ubc9 was removed by reverse Ni-NTA purification before analysis.

As a first example, we tested ubiquitination of SUMO2. SUMO2 has been shown to be ubiquitinated at K11 to form mixed SUMO-ubiquitin chains.[292–294] Reacting wild type SUMO2 with Ub-Mes **41** in the presence of Ubc9 resulted in formation of the ubiquitinated product, although

with low efficiency. We hypothesized that a known interaction between the backside of Ubc9 and SUMO[295] might interfere with the LACE reaction. By using the R13A mutant of Ubc9 to suppress the backside binding of SUMO2,[296] the yield could be improved to afford the SUMO–ubiquitin conjugate with approximately 23% conversion. Further improvement of the conversion to 37% was achieved when a single mutation, V10I, was introduced in SUMO2 to generate an optimal LACE tag (Figure 82b). Reaction with the K11R mutant of SUMO2 did not result in product formation, indicating that the ubiquitination occurred with high specificity.

We also investigated modification of α-synuclein, which has been shown to be ubiquitinated and SUMOylated with functional consequences for fibrillization, Lewy body formation and the pathogenesis of Parkinson's disease.[297–299] α-Synuclein has been predominantly found to be mono- and diubiquitinated in a heterogeneous fashion,[300,301] including at K10, K12, K21, K23, K32, K34, K43 and K96.[302,303] Endogenous ubiquitination of α-synuclein is therefore not suitable for ubiquitinating a particular residue to study its effect. We aimed to generate site-specifically monoubiquitinated α-synuclein at either K96 or K102. Of these sites, proteomics analyses have shown that K96 is the major ubiquitination site in rat brain,[304] and K96 ubiquitination inhibits fibrillization but promotes oligomerization of α-synuclein.[305] K102 has not been shown to be ubiquitinated, but the surrounding sequence only requires one mutation to install an ideal minimal LACE tag, thereby representing a good test case for the programmability the method.

Test reactions of wild type α-synuclein yielded only trace amounts of ubiquitination at K96 and K102. Aiming to direct Ubc9 towards modification of K96, we introduced the mutations V95I and D98E to provide an optimal LACE tag around K96, and K102R to suppress modification of this site. With this substrate, we observed 89% conversion to the monoubiquitinated product at K96 (Figure 82c). This increase in reactivity is notable, given the moderate change in the hydrophobic and acidic character of the two mutated residues. Likewise, we could direct site-specific ubiquitination to K102 with 46% conversion by introducing the mutations K96R and G101I to install an optimal LACE tag. Both reactions were strictly dependent on active Ubc9 (Figure 82c).

4.1.1.3. ISG15ylation of α-synuclein

ISG15 is a 17 kDa Ubl with the same C-terminus as ubiquitin,[306] making it an ideal candidate for transfer by LACE. Indeed, conjugation of ISG15-Mes thioester **42** to α-synuclein with the LACE tag at K96 proceeded to 35% after 24 hours with excellent specificity (Figure 83).

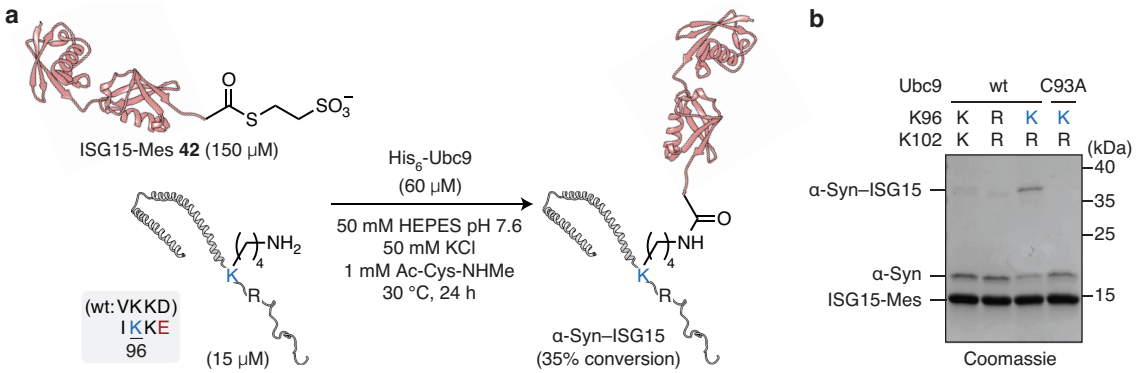


Figure 83. ISG15ylation by LACE (**a**) α-Synuclein variants were reacted with full-length ISG15-Mes thioester **42**. (**b**) Coomassie-stained SDS–PAGE analysis of reactions with either wt α-synuclein, the double mutant K96R and K102R, or the competent substrate (blue) with a consensus motif at K96 (^{95}IKKE^{98}) and concomitant lysine mutations K102R. A control reaction with the competent substrate and the Ubc9 active site mutant C93A were performed. In all samples, His$_6$-Ubc9 was removed by reverse Ni-NTA purification before analysis.

Together, these results show that Ubc9 can be efficiently directed towards ubiquitination and ISG15ylation of a particular site in a substrate by introducing modest changes around an acceptor lysine. Additionally, the programmability of this approach by the introduction of a LACE tag allowed us to prepare unnatural ubiquitin-conjugates at sites for which no canonical ubiquitination pathway is known, as exemplified by the engineered GFP substrate. The transfer of Ubls by LACE is operationally simple and can be performed in one step from entirely recombinant substrates under folded conditions, thereby expanding the scope of proteins for which monoubiquitination and ISG15ylation and its effects can be studied.

4.2. Protein–protein conjugation towards an artificial procytokine

In addition to the transfer of Ubls, we investigated whether other proteins could be installed by Ubc9. We hypothesized that general transfer of recombinant protein thioesters by Ubc9 could be achieved by including the ubiquitin-derived recognition sequence at the C-terminus of the protein.

As a test case, we investigated the transfer of affibody thioesters by Ubc9. Affibodies are derived from the B domain of the immunoglobulin-binding region of the *S. aureus* protein A and exhibit a compact three-helix bundle structure.[307,308] Screening of combinatorial libraries of surface-exposed residues within this protein domain allowed for the selection of high-affinity binders towards a range of targets.[309–311]

We aimed to generate a covalent conjugate between the soluble form of the homotrimeric tumor necrosis factor α (TNFα)[312–314] and a TNFα-specific affibody (Figure 84a).[315] TNFα is an inflammatory cytokine and is implicated in various diseases, for example rheumatoid arthritis.[316] Existing treatments include sequestration of TNFα by administering soluble variants of the TNFα receptor (TNFR) or of monoclonal antibodies.[317] Conversely, loss of TNFα sensitivity and

downregulation of TNFα signaling pathways can lead to immune evasion and tumor survival.[318–320] By generating a covalent conjugate between TNFα and Z_{TNF}, and by including a cleavable linker in the affibody, we envisioned that the conjugate would act as an artificial procytokine which would reveal the active TNFα upon cleavage of the linker (Figure 84b). As cleavable linker, we incorporated a peptide sequence that is specifically cleaved by the matrix metalloprotease 1 (MMP1),[321] a protease that has been found to be upregulated in various cancer tissues.[322–324] A corresponding affibody thioester (Z_{TNF}-Mes, **43**) could be prepared analogously to Ubl thioesters. Likewise, a TNFα variant with an N-terminal LACE tag^{3M} (TNFα-LACE$_N^{3M}$) was prepared successfully (see Experimental Part for details) (Figure 84a).

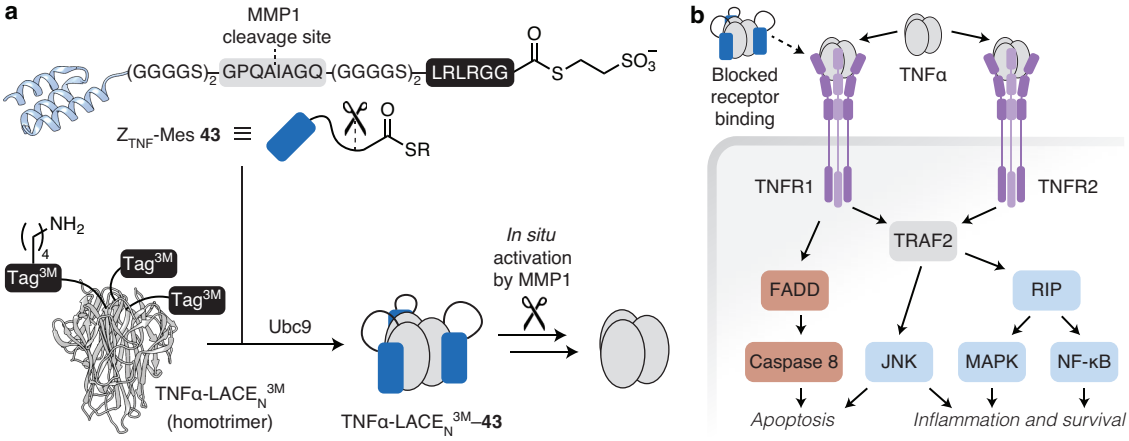


Figure 84. Strategy for an artificial procytokine. (**a**) Ubc9-mediated conjugation between TNFα-LACE$_N^{3M}$ and the affibody thioester Z_{TNF}-Mes **43** to form the conjugate TNFα-LACE$_N^{3M}$–**43**. TNFα-LACE$_N^{3M}$ contains the N-terminal LACE tag^{3M} (PRAVIKQESAE), with the additional mutation E10A compared to tag^{2M}. The Z_{TNF}-Mes construct contains a recognition sequence for cleavage by MMP1. (Protein images are based on PDB entries 1tnf and 2mzw.) (**b**) TNFα signaling pathway via TNFR1/2, resulting in apoptosis or inflammatory responses, adapted from Wu *et al.*[313] TNFα-LACE$_N^{3M}$–**43** is blocked from binding to TNFR1/2 until the linker is cleaved by MMP1, allowing for dissociation of Z_{TNF}.

After TNFα-LACE$_N^{3M}$ labeling with Z_{TNF}-Mes **43** for 6 hours, analysis by size-exclusion chromatography showed a higher-molecular weight product which was specifically formed in the presence of Ubc9 but was not observed in the absence of Ubc9 (Fraction 1, Figure 85a). SDS–PAGE analysis showed that the peak contained a mixture of unmodified TNFα-LACE$_N^{3M}$ and the conjugate TNFα-LACE$_N^{3M}$–**43**, as well as some Ubc9 and Ubc9–**43** which coeluted with the products. Additionally, a new species was formed (Fraction 3 and 3' in Figure 85a with the corresponding SDS–PAGE in Figure 85b), which is most consistent with the formation of monomeric TNFα-LACE$_N^{3M}$ due to noncovalent interaction with the Z_{TNF} affibody. Importantly, labeling of TNFα-LACE$_N^{3M}$ by Ubc9 did not cause more disassembly of the TNFα trimer than was observed in the reaction without Ubc9. MS/MS analysis of the product fraction confirmed that TNFα-LACE$_N^{3M}$ underwent specific isopeptide labeling at the acceptor lysine of the LACE tag (Figure 85c), without any other detectable modification sites.

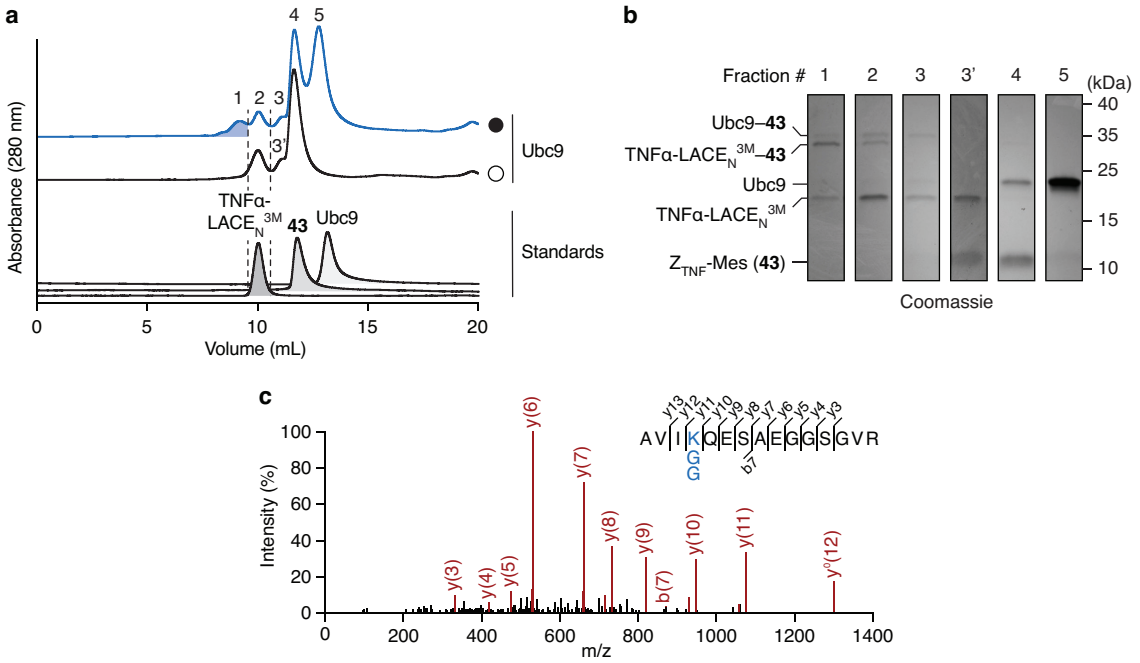


Figure 85. TNFα-LACE$_N^{3M}$ labeling with Z_{TNF}-Mes thioester **43**. (**a**) Size-exclusion chromatography (SEC) analysis of reactions after 6 h in presence (filled circle) or absence (hollow circle) of Ubc9 (top). The elution profiles of purified standards are shown as reference (bottom). (**b**) Coomassie-stained SDS–PAGE analysis of indicated fractions from **a**. (**c**) MS/MS result of tryptic peptides from fraction 1, showing the identified isopeptide modification at the acceptor lysine of the LACE tag^{3M}.

With these results, we have shown that Ubc9 can transfer protein thioesters other than Ubls, and expanded the range of substrates to multimeric proteins. By demonstrating the preparation of LACE-tag^{3M} modified TNFα and its labeling with a TNFα-specific affibody, we provide a starting point for the study of the biological activity of the conjugate and its ability to act as a MMP1-activatable procytokine. Compared to direct genetic fusion, the use of LACE for the installation of the affibody may enable a modular strategy for the preparation of several conjugates, and permit the synthetic incorporation of various cleavable linkers.[325,326]

5. Post-assembly functionalization of protein cages by LACE

Protein cages are self-assembled structures that form a nanoscale compartment with a defined structure and geometry. Such cage-forming proteins have been derived from viral capsids,[56,327] natural enzymes[328] and carrier proteins,[329] or were generated by design to afford completely artificial structures.[330] These nanocontainers have been used to confine various macromolecules such as proteins[331] and nucleic acids,[332] and were applied to the sequestration of toxic proteins,[333] storage of genetic information,[334,335] or as nanoreactors that harbor enzymatic cascades.[327,329,336,337]

These applications can often be augmented by decorating the protein cages at the surface or lumenal side with additional functionalities. While direct genetic fusion of additional domains or displayed peptides to the cage protein is possible,[338] this is often challenging due to the high

spatial confinement and the susceptibility of the multimeric structures to subtle structural changes.[339] In contrast, post-assembly modification of protein cages by chemical[56] or chemoenzymatic[340–342] methods enables straight-forward tuning of the degree of functionalization, and direct incorporation of non-proteinogenic groups. For example, carbohydrates were installed on the surface of a protein cage using glycosyltransferases[343] or on a viral capsid using a click reaction,[344] the latter resulting in a multivalent binder of influenza A virus.

Modification of the homotrimeric TNFα-LACE$_N^{3M}$ has already offered a glimpse into the application of LACE for multimeric substrates (see Section 4.2). Here, we investigated modification of the engineered lumazine synthase variant AaLS-13.[328] AaLS-13 forms a cage with icosahedral symmetry with a diameter of approximately 40 nm, and is composed of 360 subunits that are arranged in 72 pentamers.[345] Post-assembly modification of this highly multimeric structure poses a unique substrate challenge for LACE due to the local concentration of substrate subunits in any given cage particle.

5.1. Modification with small molecule probes

Introduction of a C-terminal LACE tag^{3M} on AaLS-13 (AaLS-13-LACE$_C^{3M}$) was achieved using an analogous design as was used for installation of a C-terminal peptide tag for cytoplasmic glycosylation.[343] Similarly, a recognition sequence for modification with an asparaginyl endopeptidase has been installed at the C-terminus of AaLS-13.[346] We first examined modification of AaLS-13-LACE$_C^{3M}$ cages with rhodamine thioester **5** (Figure 86a). To maintain the integrity of the cages, we included 200 mM sodium chloride and 5 mM EDTA in the reaction buffer, and thiol additives were omitted (from here on called cage buffer, see Figure 86a and the Experimental Part for detailed reaction conditions). Labeling tests with the rhodamine-functionalized thioester resulted in precipitation of the product during prolonged labeling, likely due to the hydrophobicity of the dye. We therefore lowered the amount of thioester and Ubc9 compared to the standard reaction concentrations to 120 μM and 30 μM, respectively. We terminated the reaction after 1 hour by desalting to remove excess thioester, followed by purification of the reaction mixture by size-exclusion chromatography to isolate the intact cage fraction (Figure 86b). We were pleased to find that intact cages were obtained as seen by transmission electron microscopy (TEM), with approximately 10% of the subunits modified with the rhodamine peptide **5** as judged by SDS–PAGE analysis (Figure 86c).

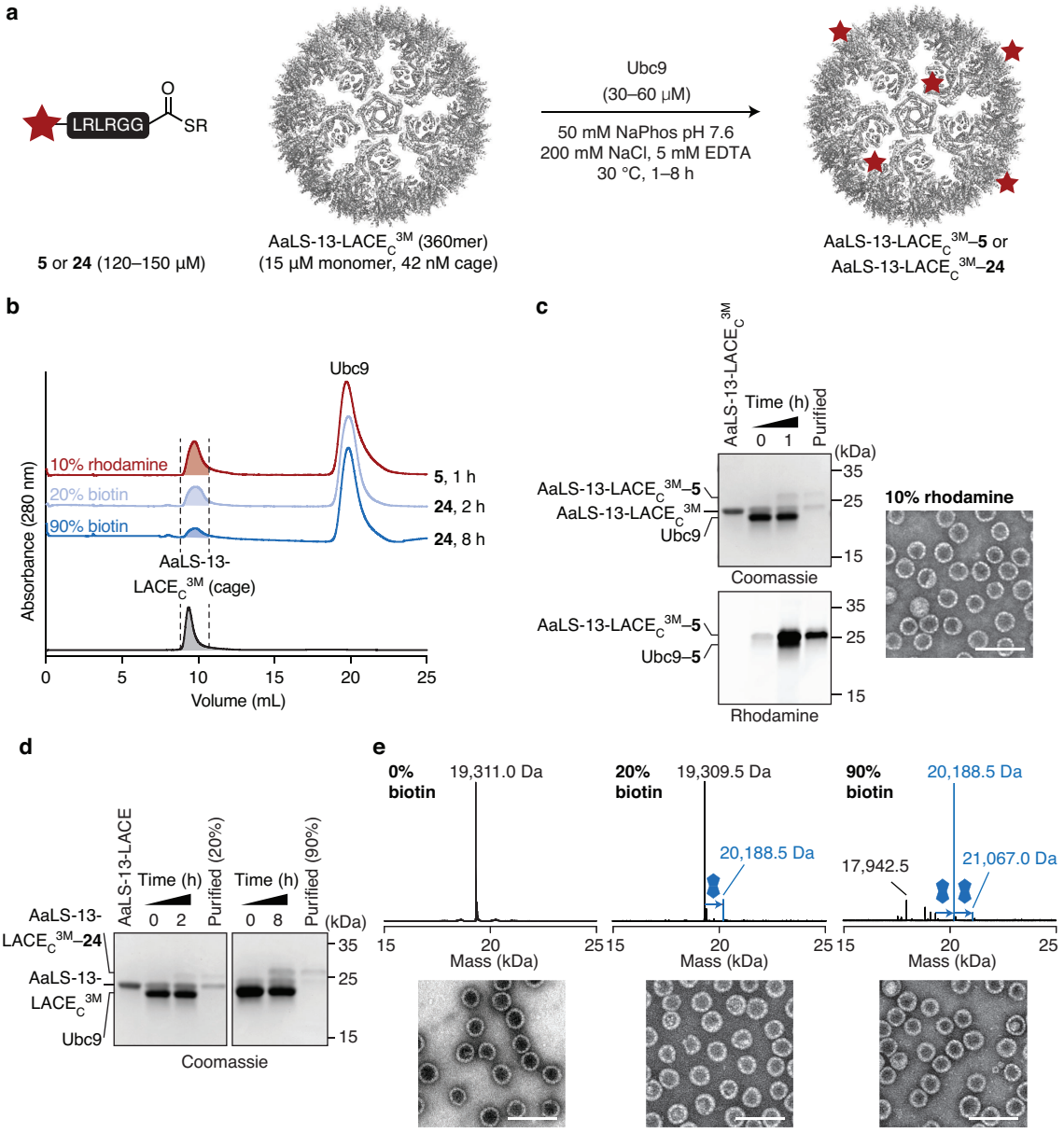


Figure 86. AaLS-13-LACE$_C^{3M}$ cage labeling with small molecule probes. (**a**) Ubc9-mediated labeling of AaLS-13-LACE$_C^{3M}$ with rhodamine thioester **5** or biotin thioester **24**. (The protein image is based on PDB entry 5mq7.) (**b**) SEC analysis of reactions in the presence of 30 μM Ubc9 with 120 μM thioester **5** (1 h, red) or 120 μM thioester **24** (2 h, light blue), or 60 μM Ubc9 with 150 μM thioester **24** (8 h, blue) (top). The elution profile of purified AaLS-13-LACE$_C^{3M}$ cages is shown as reference (bottom). (**c**) Coomassie-stained SDS–PAGE analysis (top) and in-gel fluorescence (bottom) of labeling reactions with thioester **5** and of SEC-purified labeled cages, and TEM micrograph of purified labeled cages. (**d**) Coomassie-stained SDS–PAGE analysis of labeling reactions with thioester **24** and of SEC-purified labeled cages. (**e**) Deconvoluted ESI–MS of AaLS-13-LACE$_C^{3M}$ cages (0% biotin, calc. 19,309.8 Da) and of purified cages after labeling with thioester **24** (20% and 90% biotin; truncated, calc. 17,942.4 Da; one modification, calc. 20,188.9 Da; two modifications, calc. 21,068.0 Da). The addition of biotin peptide **24** is highlighted with a blue symbol. (**c**,**e**) Corresponding TEM micrographs are shown for each variant. The scale bar (white) corresponds to 100 nm. SEC and TEM: Courtesy of M. Levasseur.

Next, we attempted to form cages with different degrees of labeling with biotin thioester **24**. To achieve a low degree of labeling, we performed the reaction for 2 hours, under otherwise

identical conditions as those employed for the rhodamine thioester. Secondly, we performed the reaction under standard conditions for 8 hours with the goal of achieving a higher degree of labeling. In both cases, purification of the reaction mixtures by size-exclusion chromatography afforded intact cages as judged by TEM, and analysis by SDS–PAGE and ESI–MS showed approximately 20% and 90% modification with the biotin peptide **24** for the low- and high-level labeling conditions, respectively (Figure 86b,d,e). In the sample with high modification, a small degree of non-specific overlabeling was observed by ESI–MS. The reason for the non-specific overlabeling was not further investigated experimentally. Inspection of the protein sequence revealed a potential second sequence motif besides the LACE tag^{3M} that has a small probability to be recognized by Ubc9 (^{68}RKED^{71}), which is also located in a surface-exposed loop. However, further experiments are needed to verify whether this site was responsible for the small amount of observed overlabeling. In the same sample, a truncated side product was present with an observed mass closely matching cleavage within the LACE tag, likely caused by trypsin or another protease (PRAVIK*QESAE, cleavage site marked with an asterisk) (Figure 86e). Since similar degradation has not been observed with other substrates bearing a LACE tag, we attribute this side product to an artefact during sample preparation or purification, rather than an inherent instability of the LACE tag^{3M}.

5.2. Display of peptides and protein domains

Simply elongating the ubiquitin-derived peptide thioester sequences N-terminally should represent a general strategy for the transfer of peptides by LACE (Figure 87a). To investigate display of peptides on the surface of AaLS-13-LACE$_C^{3M}$, we prepared thioester **44** by SPPS, which carries the peptide SP94 as an N-terminal extension of the ubiquitin-derived sequence separated by short flexible linker (Figure 87b). SP94 is a hepatocellular carcinoma-targeting peptide that was identified by phage display,[347] which has been installed on the surface of the protein cage ferritin for tumor-specific drug delivery.[338] Labeling of AaLS-13-LACE$_C^{3M}$ under standard conditions in cage buffer for 6 hours resulted in approximately 50% conversion to AaLS-13-LACE$_C^{3M}$–**44** as judged by SDS–PAGE analysis (Figure 87c). ESI–MS analysis confirmed formation of the desired conjugate, and cage integrity was verified by TEM (Figure 87d). With these results, we provide a proof of concept for the facile installation of peptides by Ubc9-mediated transfer, directly from elongated sequences of the ubiquitin-derived peptide thioester.

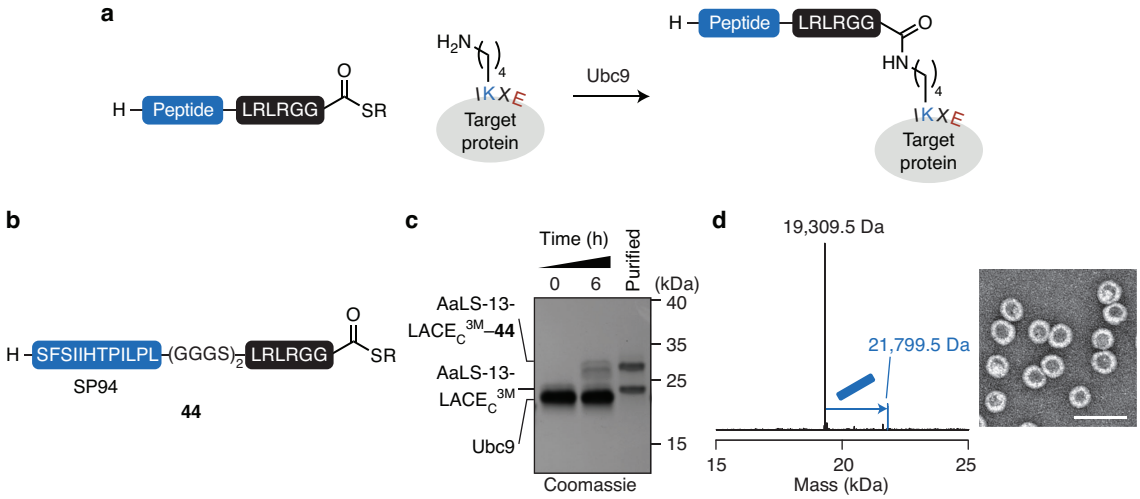


Figure 87. AaLS-13-LACE$_C^{3M}$ cage labeling for the display of peptides with free N-termini. (**a**) General strategy for the display of peptides with free N-termini. (**b**) Structure of peptide thioester **44**. (**c**) Coomassie-stained SDS–PAGE analysis of AaLS-13-LACE$_C^{3M}$ labeling reactions with thioester **44** and of SEC-purified labeled cages. (**d**) Left: Deconvoluted ESI–MS of purified cages after labeling with thioester **44** (unmodified, calc. 19,309.8 Da; modified, calc. 21,798.7 Da). The addition of the SP94 peptide **44** is highlighted in blue. Right: Corresponding TEM micrograph. The scale bar (white) corresponds to 100 nm. TEM: Courtesy of M. Levasseur.

Although such elongated peptides can be readily prepared by SPPS, a general limitation of this approach is that peptides can only be displayed N-terminally due to the isopeptide linkage via the C-terminal end. To circumvent this restraint, we explored a two-step approach, in which an azide is installed by LACE for subsequent click ligation with DBCO-functionalized peptides that bear a free C-terminus (Figure 88a). As a test case, we chose to display DBCO peptide **45** (Figure 88b). This sequence is derived from the C-end rule for peptides that bind to the cell surface receptor Neuropilin-1 for endocytosis.[348] Recently, the furin cleavage product of the Spike protein of SARS-CoV-2, the causative agent of COVID-19, has been shown to bind Neuropilin-1 via the same mechanism.[349,350] Because of the strict requirement for a free C-terminus for binding to Neuropilin-1, DBCO peptide **45** is a suitable candidate to test its display using this two-step approach.

We reacted AaLS-13-LACE$_C^{3M}$ with the azide thioester **28** in cage buffer under standard conditions for 8 hours. Following isolation of the cage fraction by size-exclusion chromatography (Figure 88c), partial azide functionalization was obtained as judged by SDS–PAGE (Figure 88d).

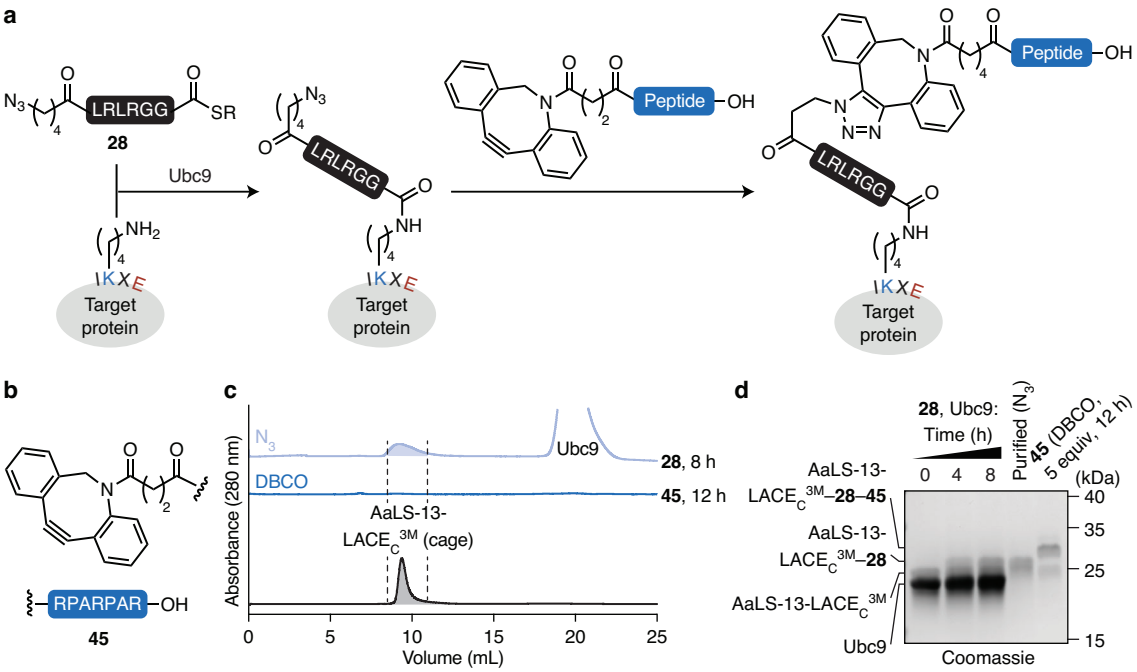


Figure 88. AaLS-13-LACE$_C^{3M}$ cage labeling for the display of peptides with free C-termini. (**a**) General strategy for the display of peptides with free C-termini using a two-step approach via copper-free click reaction. (**b**) Structure of DBCO peptide **45**. (**c**) SEC analysis of the AaLS-13-LACE$_C^{3M}$ reaction with azide thioester **28** (8 h, light blue) and of the subsequent click reaction with DBCO peptide **45** (12 h, blue) (top). The elution profile of purified AaLS-13-LACE$_C^{3M}$ cages is shown as reference (bottom). (**d**) Coomassie-stained SDS–PAGE analysis of the AaLS-13-LACE$_C^{3M}$ labeling reaction with azide thioester **28**, of the SEC-purified labeled cages (N_3), and of the subsequent click reaction with DBCO peptide **45**. SEC: Courtesy of M. Levasseur.

We then treated the labeled cage fraction with 5 equivalents of DBCO peptide **45** relative to AaLS-13-LACE$_C^{3M}$ substrate. Complete labeling of the azide-functionalized fraction to the AaLS-13-LACE$_C^{3M}$–**28**–**45** conjugate was observed as judged by SDS–PAGE analysis which showed a second band shift (Figure 88d). However, purification of the reaction mixture by size-exclusion chromatography did not afford any product (Figure 88c), likely due to precipitation of the product during the click reaction.

A relatively high degree of azide functionalization was obtained from the LACE-mediated first step, which consequently resulted in an equally high degree of labeling with an excess of peptide **45**. One reason for the loss of product could be the hydrophobicity of the DBCO-derived triazole product which is formed upon click ligation. Lower degree of labeling or more water-soluble alkynes should be explored to improve the outcome. Nevertheless, the described concept should be generalizable for the display of peptides with free C-termini.

Lastly, we investigated the transfer of a HER2-specific affibody[351,352] to AaLS-13-LACE$_C^{3M}$ for potential targeting of the cages to HER2-positive cell types (Figure 89a). The corresponding affibody thioester Z_{HER2}-Mes (**46**) was prepared recombinantly and reacted with AaLS-13-LACE$_C^{3M}$ in the presence of Ubc9. We observed approximately 5% labeling within 3 hours, which was improved slightly by increasing the reaction time to 8 hours (Figure 89b,c). Following

purification of the reaction mixture by size-exclusion chromatography, we confirmed by ESI–MS that the desired conjugate was obtained, and TEM analysis revealed intact cage particles.

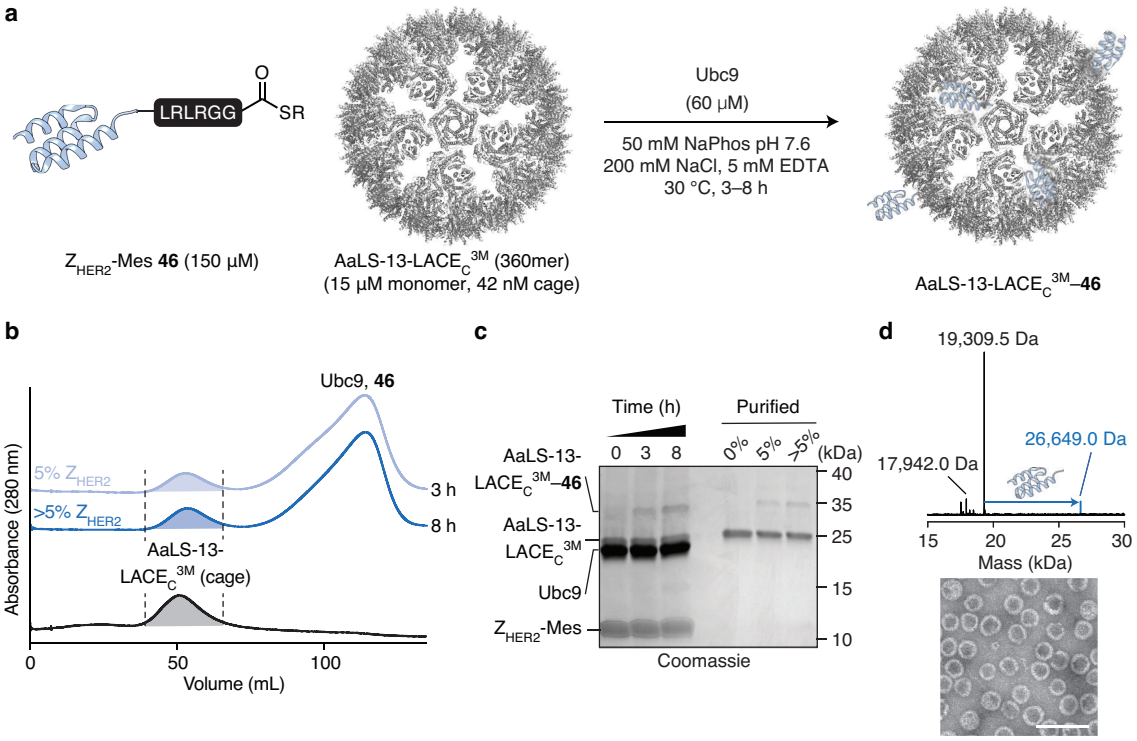


Figure 89. AaLS-13-LACE$_C^{3M}$ cage labeling with an affibody domain. (**a**) Ubc9-mediated labeling of AaLS-13-LACE$_C^{3M}$ with Z$_{HER2}$-Mes thioester **46**. (**b**) SEC analysis of the reaction after 3 h (light blue) and 8 h (blue) (top). The elution profile of purified AaLS-13-LACE$_C^{3M}$ cages is shown as reference (bottom). (**c**) Coomassie-stained SDS–PAGE analysis of the reaction at indicated times, and of the SEC-purified cages with indicated degrees of labeling. (**d**) Top: Deconvoluted ESI–MS of purified cages after labeling with thioester **46** (truncated, calc. 17,942.4 Da; unmodified, calc. 19,309.8 Da; modified, calc. 26,650.1 Da). The addition of the Z$_{HER2}$-domain **46** is highlighted in blue. Bottom: Corresponding TEM micrograph. The scale bar (white) corresponds to 100 nm. SEC and TEM: Courtesy of M. Levasseur.

With these results, we showed that the multimeric AaLS-13 cage tolerates introduction of a C-terminal LACE tag^{3M} by genetic fusion. Pre-assembled cages of this variant were successfully labeled with thioesters bearing small molecules, peptides and an affibody domain. Compared to direct genetic fusion, this post-assembly functionalization allowed us to tune the degree of labeling by adjusting the reaction duration as well as the amount of thioester and Ubc9 used. It is noteworthy that the reaction buffer contained 200 mM of sodium chloride without any thiol additives, which was preferred to maintain integrity of the cages. Despite the non-optimal buffer for LACE due to its high ionic strength (see Figure 41), labeling with biotin thioester **24** or azide thioester **28** proceeded to near completion, under otherwise the same conditions as those used for monomeric substrates. In this regard, it is interesting to note that the intact 360mer cage particle was present at a concentration of just 42 nM under the employed monomer substrate concentration of 15 µM. Further studies are required to determine the effect of this type of LACE

tag presentation on the reactivity with Ubc9, such as whether the multivalency of the tag has a positive effect or whether the high local concentration of the tag has a negative impact due to reduced encounter frequency.

6. Discussion

In this Chapter, we explored applications of LACE for the modification of diverse protein substrates with a range of small molecules, peptides and protein domains. We showed that the chemoenzymatic Ubc9 process is compatible with two existing technologies, sortase for terminal protein labeling and SpyTag/SpyCatcher for labeling of domain fusion proteins via an isopeptide bond. In this manner, we have prepared site-specifically heterobifunctionalized proteins in one step.

We also studied transfer of various small molecules and functional moieties by LACE. Similar reactivity and rate of transfer of the modified thioesters was observed, irrespective of the moiety or linker attached to the thioester, underlining the broad substrate tolerance of Ubc9. The ability of Ubc9 to transfer a broad range of probes, combined with the facile and modular synthesis of the requisite functional thioesters by SPPS, are a distinguishing feature of LACE. In comparison, several chemoenzymatic processes, such as biotin ligase or lipoic acid ligase, show reduced reactivity with certain modified substrates, and probe synthesis can become a bottleneck.

By transferring bioorthogonal handles such as click functional groups and KAT ligation handles, we performed two-step protein modifications. Introduction of click functional groups into recombinant proteins has been developed extensively, for example by amber codon suppression or via various chemoenzymatic methods. Conversely, the introduction of hydroxylamines has previously relied on modification of a surface-exposed cysteine, and labeling of recombinant proteins with KAT has not been reported to date. Here, we demonstrate for the first time introduction of KATs into recombinant proteins. We apply this to the generation of an amide-linked cytokine dimer. The achieved relative orientation of the two cytokines would not be accessible via direct genetic techniques and enabled selective receptor targeting.

Given that Ubc9 prefers ubiquitin-derived peptide thioesters as acyl donors, this finding led us to test whether Ubc9 was able to transfer full-length ubiquitin. We successfully prepared monoubiquitinated proteins from entirely recombinant starting materials, a significant improvement over previously reported. Previous synthetic and semisynthetic methods required a tradeoff between accessibility and the type of linkage (native versus nonnative). We achieved site-specific ubiquitination and ISG15ylation in a programmable manner in one step to afford a native isopeptide linkage. Furthermore, installation of the Ubls proceeded under folded conditions, which removes the need for refolding and potentially broadens the scope of proteins that can be studied. Recent advances were made by using sortase to attach Ubl variants to substrates bearing an isopeptide-linked glycylglycine moiety (see Figure 80).[289,290] These methods have

been shown to also work *in vivo*, but require mutations in the C-terminus of the Ubl as well as incorporation of an unnatural residue for recognition by sortase. LACE is complementary to these methods as it allows for conjugation of wild type ubiquitin and only relies on natural amino acids, but depends on the presence of a LACE tag in the substrate. We show that ubiquitination can be directed to a particular site in a substrate by introducing modest changes around an acceptor lysine, and that ubiquitination by LACE proceeds site-specifically on some wild type substrates bearing a consensus SUMOylation motif. Attractive targets in this system are proteins that are ubiquitinated and SUMOylated at the same lysine residue, a common mode of regulation and crosstalk between the two Ubl systems.[353] Such substrates require minimal or no changes to be ubiquitinated by Ubc9 at biologically relevant sites.

We demonstrated that, besides ubiquitin and ISG15, Ubc9 is able to transfer affibody domains that carry a C-terminal ubiquitin-derived sequence for charging of the thioester. We applied this to the conjugation of a TNFα-specific affibody to TNFα towards the generation of an artificial procytokine. Given the distinctly different fold of ubiquitin and affibodies, and the fact that synthetic peptides itself are sufficient, this indicates that LACE could be used for general protein–protein conjugation between a protein thioester and an acceptor protein carrying a LACE tag. Taken together, we show that the rate of transfer by Ubc9 is sufficient for the conjugation of entire proteins under dilute and folded conditions.

Lastly, we explored the use of LACE for the post-functionalization of a protein cages. This highly multimeric protein served as a challenging substrate to validate our method. In terms of potential applications, lumazine synthase has been used as an antigen delivery platform[354] and as an HIV immunogen.[355] Recently, antigen display on a protein cage has been applied for the generation of a vaccine candidate against SARS-CoV-2.[356] Compared to recombinant fusion techniques, post-assembly functionalization of pre-assembled AaLS-13-LACE$_C^{3M}$ would represent a modular strategy for antigen display directly from peptidic thioesters.

7. References

(226) Li, L.; Fierer, J. O.; Rapoport, T. A.; Howarth, M. Structural analysis and optimization of the covalent association between SpyCatcher and a peptide tag. *J. Mol. Biol.* **2014**, *426*, 309–317.

(227) Veggiani, G.; Nakamura, T.; Brenner, M. D.; Gayet, R. V.; Yan, J.; Robinson, C. V.; Howarth, M. Programmable polyproteams built using twin peptide superglues. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1202–1207.

(228) Harmand, T. J.; Bousbaine, D.; Chan, A.; Zhang, X.; Liu, D. R.; Tam, J. P.; Ploegh, H. L. One-pot dual labeling of IgG 1 and preparation of C-to-C fusion proteins through a combination of sortase A and butelase 1. *Bioconjug. Chem.* **2018**, *29*, 3245–3249.

(229) Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11399–11404.

(230) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **2013**, *8*, 2483–2495.

(231) Flood, D. T.; Hintzen, J. C. J.; Bird, M. J.; Cistrone, P. A.; Chen, J. S.; Dawson, P. E. Leveraging the Knorr pyrazole synthesis for the facile generation of thioester surrogates for use in native chemical ligation. *Angew. Chem. Int. Ed.* **2018**, *57*, 11634–11639.

(232) Yan, B.; Shi, W.; Ye, L.; Liu, L. Acyl donors for native chemical ligation. *Curr. Opin. Chem. Biol.* **2018**, *46*, 33–40.

(233) Huang, Y.-C.; Fang, G.-M.; Liu, L. Chemical synthesis of proteins using hydrazide intermediates. *Natl. Sci. Rev.* **2016**, *3*, 107–116.

(234) El-Faham, A.; Albericio, F. Peptide coupling reagents, more than a letter soup. *Chem. Rev.* **2011**, *111*, 6557–6602.

(235) Beckett, D.; Kovaleva, E.; Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **1999**, *8*, 921–929.

(236) Fairhead, M.; Howarth, M. Site-specific biotinylation of purified proteins using BirA. In *Site-Specific Protein Labeling: Methods and Protocols*; Gautier, A. Hinner, M. J., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2015; Vol. 1266, pp 171–184.

(237) Mito, Y.; Henikoff, J. G.; Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **2005**, *37*, 1090–1097.

(238) Edgar, R.; McKinstry, M.; Hwang, J.; Oppenheim, A. B.; Fekete, R. A.; Giulian, G.; Merril, C.; Nagashima, K.; Adhya, S. High-sensitivity bacterial detection using biotin-tagged phage and quantum-dot nanocomplexes. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4841–4845.

(239) Athavankar, S.; Peterson, B. R. Control of gene expression with small molecules: biotin-mediated acylation of targeted lysine residues in recombinant yeast. *Chem. Biol.* **2003**, *10*, 1245–1253.

(240) Yang, J.; Jaramillo, A.; Shi, R.; Kwok, W. W.; Mohanakumar, T. *In vivo* biotinylation of the major histocompatibility complex (MHC) class II/peptide complex by coexpression of BirA enzyme for the generation of MHC class II/tetramers. *Hum. Immunol.* **2004**, *65*, 692–699.

(241) Cull, M. G.; Schatz, P. J. Biotinylation of proteins *in vivo* and *in vitro* using small peptide tags. In *Applications of Chimeric Genes and Hybrid Proteins Part A: Gene Expression and Protein*; Thorner, J., Emr, S. D., Abelson, J. N., Ed.; Methods in Enzymology; Academic Press: Cambridge, MA, 2000; Vol. 326, pp 430–440.

(242) Huisgen, R. 1,3-Dipolar cycloadditions. Past and future. *Angew. Chem. Int. Ed.* **1963**, *2*, 565–598.

(243) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click chemistry: Diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.

(244) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.

(245) Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-Triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **2002**, *67*, 3057–3064.

(246) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A strain-promoted [3 + 2] azide–alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.

(247) Sletten, E. M.; Bertozzi, C. R. A hydrophilic azacyclooctyne for Cu-free click chemistry. *Org. Lett.* **2008**, *10*, 3097–3099.

(248) Agard, N. J.; Baskin, J. M.; Prescher, J. A.; Lo, A.; Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **2006**, *1*, 644–648.

(249) Tamshen, K.; Wang, Y.; Jamieson, S. M. F.; Perry, J. K.; Maynard, H. D. Genetic code expansion enables site-specific PEGylation of a human growth hormone receptor antagonist through click chemistry. *Bioconjugate Chem.* **2020**, *31*, 2179–2190.

(250) Lee, K. J.; Kang, D.; Park, H.-S. Site-specific labeling of proteins using unnatural amino acids. *Mol. Cells* **2019**, *42*, 386–396.

(251) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9482–9487.

(252) Beatty, K. E.; Liu, J. C.; Xie, F.; Dieterich, D. C.; Schuman, E. M.; Wang, Q.; Tirrell, D. A. Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew. Chem. Int. Ed.* **2006**, *45*, 7364–7367.

(253) Yao, J. Z.; Uttamapinant, C.; Poloukhtine, A.; Baskin, J. M.; Codelli, J. A.; Sletten, E. M.; Bertozzi, C. R.; Popik, V. V.; Ting, A. Y. Fluorophore targeting to cellular proteins via enzyme-mediated azide ligation and strain-promoted cycloaddition. *J. Am. Chem. Soc.* **2012**, *134*, 3720–3728.

(254) Wollack, J. W.; Silverman, J. M.; Petzold, C. J.; Mougous, J. D.; Distefano, M. D. A Minimalist substrate for enzymatic peptide and protein conjugation. *ChemBioChem* **2009**, *10*, 2934–2943.

(255) Wu, D.; Taguchi, J.; Tanriver, M.; Bode, J. W. Synthesis of acylboron compounds. *Angew. Chem. Int. Ed.* **2020**, *59*, 16847–16858.

(256) Dumas, A. M.; Molander, G. A.; Bode, J. W. Amide-forming ligation of acyltrifluoroborates and hydroxylamines in water. *Angew. Chem. Int. Ed.* **2012**, *51*, 5683–5686.

(257) Noda, H.; Erős, G.; Bode, J. W. Rapid ligations with equimolar reactants in water with the potassium acyltrifluoroborate (KAT) amide formation. *J. Am. Chem. Soc.* **2014**, *136*, 5611–5614.

(258) Saito, F.; Noda, H.; Bode, J. W. Critical evaluation and rate constants of chemoselective ligation reactions for stoichiometric conjugations in water. *ACS Chem. Biol.* **2015**, *10*, 1026–1033.

(259) Mazunin, D.; Broguiere, N.; Zenobi-Wong, M.; Bode, J. W. Synthesis of biocompatible PEG hydrogels by pH-sensitive potassium acyltrifluoroborate (KAT) amide ligations. *ACS Biomater. Sci. Eng.* **2015**, *1*, 456–462.

(260) Mazunin, D.; Bode, J. W. Potassium acyltrifluoroborate (KAT) ligations are orthogonal to thiol-Michael and SPAAC reactions: Covalent dual immobilization of proteins onto synthetic PEG hydrogels. *Helv. Chim. Acta* **2017**, *100*, e1600311.

(261) Song, H.; Wu, D.; Mazunin, D.; Liu, S. M.; Sato, Y.; Broguiere, N.; Zenobi-Wong, M.; Bode, J. W. Post-assembly photomasking of potassium acyltrifluoroborates (KATs) for two-photon 3D patterning of PEG-hydrogels. *Helv. Chim. Acta* **2020,** *103*, e2000172.

(262) Schauenburg, D.; Divandari, M.; Neumann, K.; Spiegel, C. A.; Hackett, T.; Dzeng, Y.; Spencer, N. D.; Bode, J. W. Synthesis of polymers containing potassium acyltrifluoroborates (KATs) and post-polymerization ligation and conjugation. *Angew. Chem. Int. Ed.* **2020**, *59*, 14656–14663.

(263) Liu, S. M.; Mazunin, D.; Pattabiraman, V. R.; Bode, J. W. Synthesis of bifunctional potassium acyltrifluoroborates. *Org. Lett.* **2016**, *18*, 5336–5339.

(264) Boross, G. N.; Schauenburg, D.; Bode, J. W. Chemoselective derivatization of folded synthetic insulin variants with potassium acyltrifluoroborates (KATs). *Helv. Chim. Acta* **2019**, *102*, e1800214.

(265) White, C. J.; Bode, J. W. PEGylation and dimerization of expressed proteins under near equimolar conditions with potassium 2-pyridyl acyltrifluoroborates. *ACS Cent. Sci.* **2018**, *4*, 197–206.

(266) Chiotellis, A.; Ahmed, H.; Betzel, T.; Tanriver, M.; White, C. J.; Song, H.; Ros, S. D.; Schibli, R.; Bode, J. W.; Ametamey, S. M. Chemoselective ^{18}F-incorporation into pyridyl acyltrifluoroborates for rapid radiolabelling of peptides and proteins at room temperature. *Chem. Commun.* **2019**, *56*, 723–726.

(267) Schuhmacher, A.; Shiro, T.; Ryan, S. J.; Bode, J. W. Synthesis of secondary and tertiary amides without coupling agents from amines and potassium acyltrifluoroborates (KATs). *Chem. Sci.* **2020**, *11*, 7609–7614.

(268) Shiro, T.; Schuhmacher, A.; Jackl, M. K.; Bode, J. W. Facile synthesis of α-aminoboronic acids from amines and potassium acyltrifluoroborates (KATs) via trifluoroborate-iminiums (TIMs). *Chem. Sci.* **2018**, *9*, 5191–5196.

(269) Junttila, I. S. Tuning the cytokine responses: An update on interleukin (IL)-4 and IL-13 receptor complexes. *Front. Immunol.* **2018**, *9*, 888.

(270) Malech, H. L.; Gallin, J. I. Neutrophils in human diseases. *N. Engl. J. Med.* **1987**, *317*, 687–694.

(271) Woytschak, J.; Keller, N.; Krieg, C.; Impellizzieri, D.; Thompson, R. W.; Wynn, T. A.; Zinkernagel, A. S.; Boyman, O. Type 2 interleukin-4 receptor signaling in neutrophils antagonizes their expansion and migration during infection and inflammation. *Immunity* **2016**, *45*, 172–184.

(272) Chiang, C.-C.; Cheng, W.-J.; Korinek, M.; Lin, C.-Y.; Hwang, T.-L. Neutrophils in psoriasis. *Front. Immunol.* **2019**, *10*, 2376.

(273) Lin, A. M.; Rubin, C. J.; Khandpur, R.; Wang, J. Y.; Riblett, M.; Yalavarthi, S.; Villanueva, E. C.; Shah, P.; Kaplan, M. J.; Bruce, A. T. Mast cells and neutrophils release IL-17 through extracellular trap formation in psoriasis. *J. Immunol.* **2011**, *187*, 490–500.

(274) Terui, T.; Ozawa, M.; Tagami, H. Role of Neutrophils in induction of acute inflammation in T-cell-mediated immune dermatosis, psoriasis: A neutrophil-associated inflammation-boosting loop. *Exp. Dermatol.* **2000**, *9*, 1–10.

(275) LaPorte, S. L.; Juo, Z. S.; Vaclavikova, J.; Colf, L. A.; Qi, X.; Heller, N. M.; Keegan, A. D.; Garcia, K. C. Molecular and structural basis of cytokine receptor pleiotropy in the interleukin-4/13 system. *Cell* **2008**, *132*, 259–272.

(276) Nelms, K.; Keegan, A. D.; Zamorano, J.; Ryan, J. J.; Paul, W. E. The IL-4 receptor: Signaling mechanisms and biologic functions. *Annu. Rev. Immunol.* **1999**, *17*, 701–738.

(277) Komiya, C.; Shigenaga, A.; Tsukimoto, J.; Ueda, M.; Morisaki, T.; Inokuma, T.; Itoh, K.; Otaka, A. Traceless synthesis of protein thioesters using enzyme-mediated hydrazinolysis and subsequent self-editing of the cysteinyl prolyl sequence. *Chem. Commun.* **2019**, *55*, 7029–7032.

(278) Qiao, Y.; Yu, G.; Kratch, K. C.; Wang, X. A.; Wang, W. W.; Leeuwon, S. Z.; Xu, S.; Morse, J. S.; Liu, W. R. Expressed protein ligation without intein. *J. Am. Chem. Soc.* **2020**, *142*, 7047–7054.

(279) Thompson, R. E.; Muir, T. W. Chemoenzymatic semisynthesis of proteins. *Chem. Rev.* **2019**, *120*, 3051–3126.

(280) Muir, T. W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **2003**, *72*, 249–289.

(281) Faggiano, S.; Pastore, A. The challenge of producing ubiquitinated proteins for structural studies. *Cells* **2014**, *3*, 639–656.

(282) Kumar, K. S. A.; Spasser, L.; Erlich, L. A.; Bavikar, S. N.; Brik, A. Total chemical synthesis of di-ubiquitin chains. *Angew. Chem. Int. Ed.* **2010**, *49*, 9126–9131.

(283) Jbara, M.; Sun, H.; Kamnesky, G.; Brik, A. Chemical chromatin ubiquitylation. *Curr. Opin. Chem. Biol.* **2018**, *45*, 18–26.

(284) Virdee, S.; Kapadnis, P. B.; Elliott, T.; Lang, K.; Madrzak, J.; Nguyen, D. P.; Riechmann, L.; Chin, J. W. Traceless and site-specific ubiquitination of recombinant proteins. *J. Am. Chem. Soc.* **2011**, *133*, 10708–10711.

(285) Morgan, M. T.; Haj-Yahya, M.; Ringel, A. E.; Bandi, P.; Brik, A.; Wolberger, C. Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **2016**, *351*, 725–728.

(286) Ajish Kumar, K. S.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A. Highly efficient and chemoselective peptide ubiquitylation. *Angew. Chem. Int. Ed.* **2009**, *48*, 8090–8094.

(287) McGinty, R. K.; Köhn, M.; Chatterjee, C.; Chiang, K. P.; Pratt, M. R.; Muir, T. W. Structure–activity analysis of semisynthetic nucleosomes: Mechanistic insights into the stimulation of Dot1L by ubiquitylated histone H2B. *ACS Chem. Biol.* **2009**, *4*, 958–968.

(288) Lewis, Y. E.; Abeywardana, T.; Lin, Y. H.; Galesic, A.; Pratt, M. R. Synthesis of a bis-thio-acetone (BTA) analogue of the lysine isopeptide bond and its application to investigate the effects of ubiquitination and SUMOylation on α-synuclein aggregation and toxicity. *ACS Chem. Biol.* **2016**, *11*, 931–942.

(289) Pawale, V. S.; Yadav, P.; Roy, R. P. Facile one-step assembly of bona fide SUMO conjugates by chemoenzymatic ligation. *ChemBioChem* **2018**, *19*, 1137–1141.

(290) Fottner, M.; Brunner, A.-D.; Bittl, V.; Horn-Ghetko, D.; Jussupow, A.; Kaila, V. R. I.; Bremm, A.; Lang, K. Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat. Chem. Biol.* **2019**, *15*, 276–284

(291) Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705–6710.

(292) Udeshi, N. D.; Mani, D. R.; Eisenhaure, T.; Mertins, P.; Jaffe, J. D.; Clauser, K. R.; Hacohen, N.; Carr, S. A. Methods for quantification *of in vivo* changes in protein ubiquitination following proteasome and deubiquitinase inhibition. *Mol. Cell. Proteomics* **2012**, *11*, 148–159.

(293) Kim, W.; Bennett, E. J.; Huttlin, E. L.; Guo, A.; Li, J.; Possemato, A.; Sowa, M. E.; Rad, R.; Rush, J.; Comb, M. J.; Harper, J. W.; Gygi, S. P. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell.* **2011**, *44*, 325–340.

(294) Tatham, M. H.; Geoffroy, M.-C.; Shen, L.; Plechanovova, A.; Hattersley, N.; Jaffray, E. G.; Palvimo, J. J.; Hay, R. T. RNF4 Is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat. Cell. Biol.* **2008**, *10*, 538–546.

(295) Klug, H.; Xaver, M.; Chaugule, V. K.; Koidl, S.; Mittler, G.; Klein, F.; Pichler, A. Ubc9 sumoylation controls SUMO chain formation and meiotic synapsis in *Saccharomyces cerevisiae*. *Mol. Cell.* **2013**,*50*, 625–636.

(296) Capili, A. D.; Lima, C. D. Structure and analysis of a complex between SUMO and Ubc9 illustrates features of a conserved E2-Ubl interaction. *J. Mol. Biol.* **2007**, *369*, 608–618.

(297) Haj-Yahya, M.; Fauvet, B.; Herman-Bachinsky, Y.; Hejjaoui, M.; Bavikar, S. N.; Karthikeyan, S. V.; Ciechanover, A.; Lashuel, H. A.; Brik, A. Synthetic polyubiquitinated α-synuclein reveals important insights into the roles of the ubiquitin chain in regulating its pathophysiology. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17726–17731.

(298) Hejjaoui, M.; Haj-Yahya, M.; Kumar, K. S. A.; Brik, A.; Lashuel, H. A. Towards elucidation of the role of ubiquitination in the pathogenesis of Parkinson's disease with semisynthetic ubiquitinated α-synuclein. *Angew. Chem. Int. Ed.* **2011**, *50*, 405–409.

(299) Rott, R.; Szargel, R.; Shani, V.; Hamza, H.; Savyon, M.; Elghani, F. A.; Bandopadhyay, R.; Engelender, S. SUMOylation and ubiquitination reciprocally regulate α-synuclein degradation and pathological aggregation. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 13176–13181.

(300) Anderson, J. P.; Walker, D. E.; Goldstein, J. M.; Laat, R. de; Banducci, K.; Caccavello, R. J.; Barbour, R.; Huang, J.; Kling, K.; Lee, M.; Diep, L.; Keim, P. S.; Shen, X.; Chataway, T.; Schlossmacher, M. G.; Seubert, P.; Schenk, D.; Sinha, S.; Gai, W. P.; Chilcote, T. J. Phosphorylation of Ser-129 is the dominant pathological modification of α-synuclein in familial and sporadic Lewy body disease. *J. Biol. Chem.* **2006**, *281*, 29739–29752.

(301) Hasegawa, M.; Fujiwara, H.; Nonaka, T.; Wakabayashi, K.; Takahashi, H.; Lee, V. M.-Y.; Trojanowski, J. Q.; Mann, D.; Iwatsubo, T. Phosphorylated α-synuclein is ubiquitinated in α-synucleinopathy lesions. *J. Biol. Chem.* **2002**, *277*, 49071–49076.

(302) Nonaka, T.; Iwatsubo, T.; Hasegawa, M. Ubiquitination of α-synuclein. *Biochemistry* **2005**, *44*, 361–368.

(303) Lee, J. T.; Wheeler, T. C.; Li, L.; Chin, L.-S. Ubiquitination of α-synuclein by Siah-1 promotes α-synuclein aggregation and apoptotic cell death. *Hum. Mol. Genet.* **2008**, *17*, 906–917.

(304) Na, C. H.; Jones, D. R.; Yang, Y.; Wang, X.; Xu, Y.; Peng, J. Synaptic protein ubiquitination in rat brain revealed by antibody-based ubiquitome analysis. *J. Proteome. Res.* **2012**, *11*, 4722–4732.

(305) Meier, F.; Abeywardana, T.; Dhall, A.; Marotta, N. P.; Varkey, J.; Langen, R.; Chatterjee, C.; Pratt, M. R. Semisynthetic, site-specific ubiquitin modification of α-synuclein reveals differential effects on aggregation. *J. Am. Chem. Soc.* **2012**, *134*, 5468–5471.

(306) Dzimianski, J. V.; Scholte, F. E. M.; Bergeron, É.; Pegan, S. D. ISG15: It's complicated. *J. Mol. Biol.* **2019**, *431*, 4203–4216.

(307) Nilsson, B.; Moks, T.; Jansson, B.; Abrahmsén, L.; Elmblad, A.; Holmgren, E.; Henrichson, C.; Jones, T. A.; Uhlén, M. A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng. Des. Sel.* **1987**, *1*, 107–113.

(308) Nord, K.; Nilsson, J.; Nilsson, B.; Uhlén, M.; Nygren, P.-Å. A combinatorial library of an α-helical bacterial receptor domain. *Protein Eng. Des. Sel.* **1995**, *8*, 601–608.

(309) Löfblom, J.; Feldwisch, J.; Tolmachev, V.; Carlsson, J.; Ståhl, S.; Frejd, F. Y. Affibody molecules: Engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS Lett.* **2010**, *584*, 2670–2680.

(310) Ståhl, S.; Gräslund, T.; Karlström, A. E.; Frejd, F. Y.; Nygren, P.-Å.; Löfblom, J. Affibody molecules in biotechnological and medical applications. *Trends Biotechnol.* **2017**, *35*, 691–712.

(311) Nygren, P. Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J.* **2008**, *275*, 2668–2676.

(312) Strieter, R. M.; Kunkel, S. L.; Bone, R. C. Role of tumor necrosis factor-α in disease states and inflammation. *Crit. Care Med.* **1993**, *21*, S447.

(313) Wu, Y.; Zhou, B. P. TNF-α/NF-κB/Snail pathway in cancer cell migration and invasion. *Brit. J. Cancer* **2010**, *102*, 639–644.

(314) Eck, M. J.; Sprang, S. R. The structure of tumor necrosis factor-α at 2.6 Å resolution. *J. Biol. Chem.* **1990**, *264*, 17595–17605.

(315) Jonsson, A.; Wållberg, H.; Herne, N.; Ståhl, S.; Frejd, F. Y. Generation of tumour-necrosis-factor-α-specific affibody molecules capable of blocking receptor binding *in vitro*. *Biotechnol. Appl. Bioc.* **2009**, *54*, 93–103.

(316) Feldmann, M. Development of anti-TNF therapy for rheumatoid arthritis. *Nat. Rev. Immunol.* **2002**, *2*, 364–371.

(317) Caporali, R.; Pallavicini, F. B.; Filippini, M.; Gorla, R.; Marchesoni, A.; Favalli, E. G.; Sarzi-Puttini, P.; Atzeni, F.; Montecucco, C. Treatment of rheumatoid arthritis with anti-TNF-alpha agents: A reappraisal. *Autoimmun. Rev.* **2009**, *8*, 274–280.

(318) Bertazza, L.; Mocellin, S. Tumor necrosis factor (TNF) biology and cell death. *Front. Biosci.* **2008**, *13*, 2736.

(319) Kearney, C. J.; Vervoort, S. J.; Hogg, S. J.; Ramsbottom, K. M.; Freeman, A. J.; Lalaoui, N.; Pijpers, L.; Michie, J.; Brown, K. K.; Knight, D. A.; Sutton, V.; Beavis, P. A.; Voskoboinik, I.; Darcy, P. K.; Silke, J.; Trapani, J. A.; Johnstone, R. W.; Oliaro, J. Tumor immune evasion arises through loss of TNF sensitivity. *Sci. Immunol.* **2018**, *3*, eaar3451.

(320) Töpfer, K.; Kempe, S.; Müller, N.; Schmitz, M.; Bachmann, M.; Cartellieri, M.; Schackert, G.; Temme, A. Tumor evasion from T cell surveillance. *J. Biomed. Biotechnol.* **2011**, *2011*, 918471.

(321) Sandersjöö, L.; Jonsson, A.; Löfblom, J. A new prodrug form of affibody molecules (pro-affibody) is selectively activated by cancer-associated proteases. *Cell Mol. Life Sci.* **2015**, *72*, 1405–1415.

(322) Juncker-Jensen, A.; Deryugina, E. I.; Rimann, I.; Zajac, E.; Kupriyanova, T. A.; Engelholm, L. H.; Quigley, J. P. Tumor MMP-1 activates endothelial PAR1 to facilitate vascular intravasation and metastatic dissemination. *Cancer. Res.* **2013**, *73*, 4196–4211.

(323) Murray, G. I.; Duncan, M. E.; O'Neil, P.; Melvin, W. T.; Fothergill, J. E. Matrix metalloproteinase–1 is associated with poor prognosis in colorectal cancer. *Nat. Med.* **1996**, *2*, 461–462.

(324) Schütz, A.; Schneidenbach, D.; Aust, G.; Tannapfel, A.; Steinert, M.; Wittekind, C. Differential expression and activity status of MMP-1, MMP-2 and MMP-9 in tumor and stromal cells of squamous cell carcinomas of the lung. *Tumor Biol.* **2002**, *23*, 179–184.

(325) Maluch, I.; Czarna, J.; Drag, M. Applications of unnatural amino acids in protease probes. *Chem. Asian. J.* **2019**, *14*, 4103–4113.

(326) Pomplun, S.; Shugrue, C. R.; Schmitt, A. M.; Schissel, C. K.; Farquhar, C. E.; Pentelute, B. L. Secondary amino alcohols: Traceless cleavable linkers for use in affinity capture and release. *Angew. Chem. Int. Ed.* **2020**, *59*, 11566–11572.

(327) Jordan, P. C.; Patterson, D. P.; Saboda, K. N.; Edwards, E. J.; Miettinen, H. M.; Basu, G.; Thielges, M. C.; Douglas, T. Self-assembling biomolecular catalysts for hydrogen production. *Nat. Chem.* **2016**, *8*, 179–185.

(328) Azuma, Y.; Edwardson, T. G. W.; Hilvert, D. Tailoring lumazine synthase assemblies for bionanotechnology. *Chem. Soc. Rev.* **2018**, *47*, 3543–3557.

(329) Tetter, S.; Hilvert, D. Enzyme encapsulation by a ferritin cage. *Angew. Chem. Int. Ed.* **2017**, *56*, 14933–14936.

(330) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **2012**, *336*, 1171–1174.

(331) Zschoche, R.; Hilvert, D. Diffusion-limited cargo loading of an engineered protein container. *J. Am. Chem. Soc.* **2015**, *137*, 16121–16132.

(332) Azuma, Y.; Edwardson, T. G. W.; Terasaka, N.; Hilvert, D. Modular protein cages for size-selective RNA packaging *in vivo*. *J. Am. Chem. Soc.* **2018**, *140*, 566–569.

(333) Wörsdörfer, B.; Woycechowsky, K. J.; Hilvert, D. Directed evolution of a protein container. *Science* **2011**, *331*, 589–592.

(334) Terasaka, N.; Azuma, Y.; Hilvert, D. Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 5432–5437.

(335) Edwardson, T. G. W.; Hilvert, D. Virus-inspired function in engineered protein cages. *J. Am. Chem. Soc.* **2019**, *141*, 9432–9443.

(336) Azuma, Y.; Bader, D. L. V.; Hilvert, D. Substrate sorting by a supercharged nanoreactor. *J. Am. Chem. Soc.* **2018**, *140*, 860–863.

(337) Azuma, Y.; Zschoche, R.; Tinzl, M.; Hilvert, D. Quantitative packaging of active enzymes into a protein cage. *Angew. Chem. Int. Ed.* **2015**, *55*, 1531–1534.

(338) Jiang, B.; Zhang, R.; Zhang, J.; Hou, Y.; Chen, X.; Zhou, M.; Tian, X.; Hao, C.; Fan, K.; Yan, X. GRP78-targeted ferritin nanocaged ultra-high dose of doxorubicin for hepatocellular carcinoma therapy. *Theranostics* **2019**, *9*, 2167–2182.

(339) Robinson, S. A.; Hartman, E. C.; Ikwuagwu, B. C.; Francis, M. B.; Tullman-Ercek, D. Engineering a virus-like particle to display peptide insertions using an apparent fitness landscape. *Biomacromolecules* **2020**, *21*, 4194–4204.

(340) Lobba, M. J.; Fellmann, C.; Marmelstein, A. M.; Maza, J. C.; Kissman, E. N.; Robinson, S. A.; Staahl, B. T.; Urnes, C.; Lew, R. J.; Mogilevsky, C. S.; Doudna, J. A.; Francis, M. B. Site-specific bioconjugation through enzyme-catalyzed tyrosine–cysteine bond formation. *ACS Cent. Sci.* **2020**, *6*, 1564–1571.

(341) Marmelstein, A. M.; Lobba, M. J.; Mogilevsky, C. S.; Maza, J. C.; Brauer, D. D.; Francis, M. B. Tyrosinase-mediated oxidative coupling of tyrosine tags on peptides and proteins. *J. Am. Chem. Soc.* **2020**, *142*, 5078–5086.

(342) Liu, M.; Zhu, Y.; Wu, T.; Cheng, J.; Liu, Y. Nanobody-ferritin conjugate for targeted photodynamic therapy. *Chem. Eur. J.* **2020**, *26*, 7442–7450.

(343) Tytgat, H. L. P.; Lin, C.; Levasseur, M. D.; Tomek, M. B.; Rutschmann, C.; Mock, J.; Liebscher, N.; Terasaka, N.; Azuma, Y.; Wetter, M.; Bachmann, M. F.; Hilvert, D.; Aebi, M.; Keys, T. G. Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. *Nat. Commun.* **2019**, *10*, 5403.

(344) Lauster, D.; Klenk, S.; Ludwig, K.; Nojoumi, S.; Behren, S.; Adam, L.; Stadtmüller, M.; Saenger, S.; Zimmler, S.; Hönzke, K.; Yao, L.; Hoffmann, U.; Bardua, M.; Hamann, A.; Witzenrath, M.; Sander, L. E.; Wolff, T.; Hocke, A. C.; Hippenstiel, S.; Carlo, S. D.; Neudecker, J.; Osterrieder, K.; Budisa, N.; Netz, R. R.; Böttcher, C.; Liese, S.; Herrmann, A.; Hackenberger, C. P. R. Phage capsid nanoparticles with defined ligand arrangement block influenza virus entry. *Nat. Nanotechnol.* **2020**, *15*, 373–379.

(345) Sasaki, E.; Böhringer, D.; Waterbeemd, M. van de; Leibundgut, M.; Zschoche, R.; Heck, A. J. R.; Ban, N.; Hilvert, D. Structure and assembly of scalable porous protein cages. *Nat. Commun.* **2017**, *8*, 14663.

(346) Tang, T. M. S.; Cardella, D.; Lander, A. J.; Li, X.; Escudero, J. S.; Tsai, Y.-H.; Luk, L. Y. P. Use of an asparaginyl endopeptidase for chemo-enzymatic peptide and protein labeling. *Chem. Sci.* **2020**, *11*, 5881–5888.

(347) Lo, A.; Lin, C.-T.; Wu, H.-C. Hepatocellular carcinoma cell-specific peptide ligand for targeted drug delivery. *Mol. Cancer Ther.* **2008**, *7*, 579–589.

(348) Pang, H.-B.; Braun, G. B.; Friman, T.; Aza-Blanc, P.; Ruidiaz, M. E.; Sugahara, K. N.; Teesalu, T.; Ruoslahti, E. An Endocytosis pathway initiated through Neuropilin-1 and regulated by nutrient availability. *Nat. Commun.* **2014**, *5*, 4904.

(349) Daly, J. L.; Simonetti, B.; Klein, K.; Chen, K.-E.; Williamson, M. K.; Antón-Plágaro, C.; Shoemark, D. K.; Simón-Gracia, L.; Bauer, M.; Hollandi, R.; Greber, U. F.; Horvath, P.; Sessions, R. B.; Helenius, A.; Hiscox, J. A.; Teesalu, T.; Matthews, D. A.; Davidson, A. D.; Collins, B. M.; Cullen, P. J.; Yamauchi, Y. Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **2020**, *370*, 861–865.

(350) Cantuti-Castelvetri, L.; Ojha, R.; Pedro, L. D.; Djannatian, M.; Franz, J.; Kuivanen, S.; Meer, F. van der; Kallio, K.; Kaya, T.; Anastasina, M.; Smura, T.; Levanov, L.; Szirovicza, L.; Tobi, A.; Kallio-Kokko, H.; Österlund, P.; Joensuu, M.; Meunier, F. A.; Butcher, S. J.; Winkler, M. S.; Mollenhauer, B.; Helenius, A.; Gokce, O.; Teesalu, T.; Hepojoki, J.; Vapalahti, O.; Stadelmann, C.; Balistreri, G.; Simons, M. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **2020**, *370*, 856–860.

(351) Orlova, A.; Magnusson, M.; Eriksson, T. L. J.; Nilsson, M.; Larsson, B.; Höidén-Guthenberg, I.; Widström, C.; Carlsson, J.; Tolmachev, V.; Ståhl, S.; Nilsson, F. Y. Tumor imaging using a picomolar affinity HER2 binding affibody molecule. *Cancer Res.* **2006**, *66*, 4339–4348.

(352) Eigenbrot, C.; Ultsch, M.; Dubnovitsky, A.; Abrahmsén, L.; Härd, T. Structural basis for high-affinity HER2 receptor binding by an engineered protein. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15039–15044.

(353) Denuc, A.; Marfany, G. SUMO and ubiquitin paths converge. *Biochem. Soc. T.* **2010**, *38*, 34–39.

(354) Ra, J.-S.; Shin, H.-H.; Kang, S.; Do, Y. Lumazine synthase protein cage nanoparticles as antigen delivery nanoplatforms for dendritic cell-based vaccine development. *Clin. Exp. Vaccine Res.* **2014**, *3*, 227–234.

(355) Jardine, J.; Julien, J.-P.; Menis, S.; Ota, T.; Kalyuzhniy, O.; McGuire, A.; Sok, D.; Huang, P.-S.; MacPherson, S.; Jones, M.; Nieusma, T.; Mathison, J.; Baker, D.; Ward, A. B.; Burton, D. R.; Stamatatos, L.; Nemazee, D.; Wilson, I. A.; Schief, W. R. Rational HIV immunogen design to target specific germline B cell receptors. *Science* **2013**, *340*, 711–716.

(356) Walls, A. C.; Fiala, B.; Schäfer, A.; Wrenn, S.; Pham, M. N.; Murphy, M.; Tse, L. V.; Shehata, L.; O'Connor, M. A.; Chen, C.; Navarro, M. J.; Miranda, M. C.; Pettie, D.; Ravichandran, R.; Kraft, J. C.; Ogohara, C.; Palser, A.; Chalk, S.; Lee, E.-C.; Guerriero, K.; Kepl, E.; Chow, C. M.; Sydeman, C.; Hodge, E. A.; Brown, B.; Fuller, J. T.; Dinnon, K. H.; Gralinski, L. E.; Leist, S. R.; Gully, K. L.; Lewis, T. B.; Guttman, M.; Chu, H. Y.; Lee, K. K.; Fuller, D. H.; Baric, R. S.; Kellam, P.; Carter, L.; Pepper, M.; Sheahan, T. P.; Veesler, D.; King, N. P. Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* **2020**, *183*, 1367–1382.

CHAPTER 4

Library Screening Workflow Towards a Redox-Switchable Disulfide Tag

Work described in this Chapter has been performed in collaboration with Martin Köhler (Zenobi Group, ETH Zürich) and Dr. Bertran Rubi (MoBiAs, ETH Zürich) who contributed to MALDI-based peptide sequencing, and Christian Altorfer (apprentice, ETH Zürich) who contributed to one-bead one peptide library synthesis. We thank Dr. Erich Brunner (UZH) for access to the fluorescence-activated bead sorter and for the technical support.

1. Introduction

1.1. Advantages and limitations of cysteine modification techniques

A number of protein modification strategies take advantage of the high nucleophilicity and low natural abundance of cysteine residues, which often react rapidly and with high chemoselectivity (see Figure 6). The reactivity of cysteine residues can be further tuned by the chemical environment provided by the protein structure or by adjacent residues, which has been exploited for sequence-specific modification. A classic example for a sequence-specific reaction with cysteine residues is the venerable FlAsH tag, a tetracysteine motif that specifically reacts with organobisarsenic thioesters.[357] Another strategy relies on a cysteine residue embedded in a four-residue motif called a 'π-clamp' which undergoes selective nucleophilic aromatic substitution reactions with perfluroaromatic reagents, whereas other free cysteines are unreactive.[358]

A common limitation of these protein modification strategies is the susceptibility of free cysteine residues towards oxidation, which can complicate production and handling of the substrates. Additionally, introducing non-native cysteines at surface-exposed positions of a protein may cause scrambling of native disulfide bonds and requires careful optimization.[35]

1.2. Strategy for a redox-switchable disulfide tag

To circumvent the need for free cysteine residues, disulfide reduction and rebridging strategies have been developed (see Figure 7). Disulfide rebridging is especially useful for the modification of monoclonal antibodies which contain an easily reducible interchain disulfide bond, enabling the preparation of homogeneous ADCs in the presence of several other intrachain disulfide bonds.[36]

We envisioned a redox-switchable disulfide tag based on vicinal cysteine residues that could serve as a general platform for labeling of a protein of interest via disulfide rebridging (Figure 90). The disulfide tag should spontaneously form an intramolecular disulfide bond upon expression and purification, which would protect native disulfide bonds and simplify handling of the protein substrate. To modify the protein of interest, the disulfide tag would be selectively reduced under mild conditions, followed by site-specific labeling of the disulfide tag with rebridging reagents, Ideally, conditions could be found that leave native disulfide bonds intact.

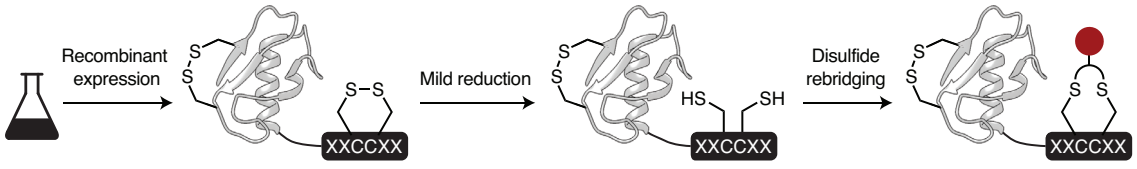


Figure 90. Strategy for a redox-switchable disulfide tag for protein modification. The disulfide tag XXCCXX, in which X is a proteinogenic amino acid, should form an internal disulfide bond during recombinant expression and purification to protect potential native disulfide bonds, and to prevent further oxidation of the

thiol group. The disulfide tag should be selectively reducible in the presence of native disulfide bonds for subsequent disulfide rebridging.

A range of disulfide rebridging reagents have been developed, which, depending on the reagent, give rise to distinct bridge compositions, lengths and geometries (see Figure 7). Popular reagents include dibromomaleimides,[32] allyl sulfones,[33] divinylsulfonamides,[359] as well as divinyltriazines[360] and -pyrimidines.[361] Lastly, several halocarbons have been used for cysteine rebridging to give mesitylene-, isobytylene-, or oxetane-linked disulfides.[34,362,363]

1.3. Modulation of the redox reactivity of cysteines

The redox reactivity of cysteines is strongly influenced by their chemical and geometric environment. Analysis of amino acid patterns surrounding disulfide bonds in proteins revealed that weakly hydrophilic and aromatic residues are enriched, whereas aliphatic and hydrophobic amino acids are less commonly located near disulfide bonds.[364] The nucleophilicity and rate of thiol-disulfide exchange reactions have been found to be influenced by neighboring charged residues that modulate the pK_a of the thiol functional group,[365] with a lower pK_a of the thiol functional group correlating with faster disulfide formation.[366]

Besides tuning of the chemical environment, the geometry and ring size of disulfide-bonded cysteine-rich peptides can influence their redox reactivity.[367] The four-residue CPPC tag, resembling the hallmark motif of protein disulfide isomerases[368] and thioredoxins,[369] has been shown to form parallel homodimers specifically, rather than oligomers or cyclic monomers. [370] Similar motifs with spacers longer than two residues between the cysteines were found to form the cyclic monomer exclusively.[370] The CXC motif has been reported to form exclusively twin homodimer and closed monomer oxidation products, but not mixed disulfides, likely because the neighboring cysteine residue can displace the mixed disulfide to form the cyclic monomer product.[371] Both the CPPC and CXC motifs have been used to generate cysteine-rich peptide libraries with defined disulfide bond patterns.[370,371] A hinge-type tetracysteine tag was derived from a naturally occurring β-hairpin motif, which enabled highly specific dimerization of proteins bearing this tag.[372,373] Lastly, a cysteine-rich tag has been developed (^{1}CISTCC^6), which selectively undergoes disulfide bond formation between Cys1 and Cys6. [374] Introduction of this tag into recombinant proteins enabled dual protein modification by first labeling the free Cys5, followed by reduction and modification of the disulfide bonded cysteines.[374]

1.3.1. Vicinal disulfides

We were intrigued by the redox chemistry of vicinal cysteines. The eight-membered ring formed upon intramolecular disulfide formation between vicinal cysteines is the smallest possible ring size for cysteine-based disulfides. However, disulfide formation between vicinal cysteines is not the most difficult to form. Intramolecular disulfide formation between vicinal cysteines occurs

at an intermediate rate compared to cysteines that are separated by short, intervening peptide spacers, and occurs for example faster than CPPC.[367] The vicinal cysteine motif is relatively rare in natural proteins, and is not known to have a general redox function in enzymes. The most common role of this disulfide motif in natural proteins is to provide a hydrophobic patch at the floor of carbohydrate binding pockets.[375] Vicinal disulfides can occur in the *cis*- and *trans*-conformation in protein structures.[375] Computational studies identified an S···C=O $n \rightarrow \pi^*$ interaction between a sulfur atom of the disulfide and the carbonyl of the connecting amide bond.[376] This non-covalent interaction is particularly strong in the *trans*-conformation and may contribute to the stabilization of secondary structures.[376]

2. A one-bead one-peptide library for disulfide tag screening

2.1. Synthesis of a one-bead one-peptide combinatorial library

The goal of this project was to investigate the relatively unexplored vicinal cysteine motif as a reactive handle for site-specific protein modification. Because of the intermediate redox reactivity of vicinal cysteines,[367] we sought to develop an unbiased screen to select for unique redox function of this motif by modulating the chemical environment. To this end, we prepared a one-bead one-peptide combinatorial library (Figure 91).[377] By splitting a resin batch into individual reaction vessels followed by coupling of one type of amino acid to each aliquot, each bead is homogeneously modified with the respective residue (Figure 91a). After repeating the process of splitting, coupling and mixing, all possible sequences are obtained in a combinatorial fashion. Importantly, each bead is homogeneously modified with only one library member (Figure 91b). This enables the property of one bead to be correlated with one library member during library screening.

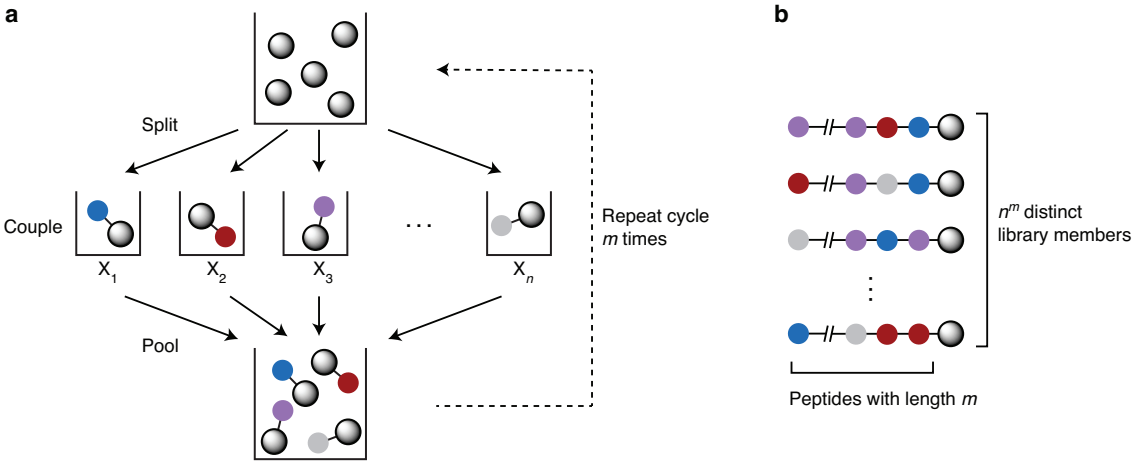


Figure 91. Split and mix peptide synthesis. (**a**) For split and mix peptide synthesis, the resin batch is split into aliquots. Each resin aliquot is elongated with a different amino acid (X_1–X_n), and the cycle of pooling and redistribution can be repeated to incorporate several randomized positions (*m*). (**b**) At the end of the

synthesis, each resin bead carries peptides with one particular sequence. The resulting one-bead one-peptide library contains n^m members.

Our one-bead one-peptide combinatorial library of vicinal cysteine peptides had the general structure XXCCXX, attached via a short flexible linker GGGRG to a solid support (Figure 92a). As solid support, we opted for a water-compatible Tentagel resin which enables on-bead screening of the library in aqueous buffer.[378,379] Several aspects of the library were designed such that the peptide hits could be analyzed by MALDI–MS/MS *de novo* peptide sequencing (see Section 2.3). At the variable positions X, we incorporated all canonical amino acids, except isoleucine which is isobaric to leucine, and glutamine, which is nearly isobaric to lysine. Additionally, we omitted cysteines, apart from the fixed positions, as well as methionine and tryptophan, which are prone to oxidation and may complicate MS/MS spectra. An arginine residue was incorporated in the linker near the C-terminus to favor formation of y-ions[380] during MALDI–MS/MS analysis. We limited ourselves to four randomized positions since the resulting library comprising 15^4 = 50,625 members would correspond to approximately 100 mg of resin and could be conveniently prepared in ten-fold excess. As a linker, we used the acid stable 4-hydroxymethylbenzoic acid (HMBA),[381] which allows deprotection of the amino acid side chains on resin. Additionally, a convenient cleavage procedure using ammonia vapor has been reported (Figure 92b),[379] which is ideal for high-throughput screening.

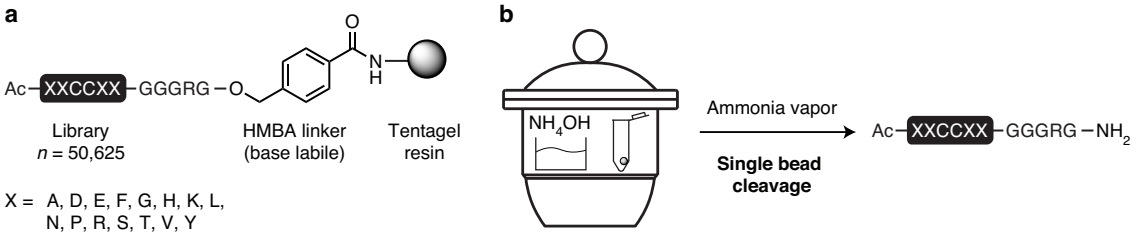


Figure 92. One-bead one-peptide library for disulfide tag screening. (**a**) Structure of the disulfide tag library. (**b**) Procedure for peptide cleavage from single beads. Aminolysis of the HMBA linker by ammonia vapor affords C-terminally amidated peptides. Following cleavage, the released peptides can be dissolved in MS-compatible solvent for subsequent analyses. Adapted from Giudicessi *et al.*[379]

To check the quality of the library, we alkylated a portion of the library with iodoacetamide, cleaved the peptides of eight single beads using the ammonia vapor method, and analyzed the eluted peptides by MALDI–MS (see Experimental Part for details) (Figure 93a). Clean spectra with mostly one major species were observed. In several of the samples, a side product was present with a mass that best corresponds to the formation of 2,3-diaminopropionic acid (Dap) from carbamidomethyl (Cam)-modified cysteine. This side product was likely formed during linker cleavage by base-promoted elimination of Cam-modified cysteine to Dha, followed by ammonia addition to give Dap (Figure 93b).

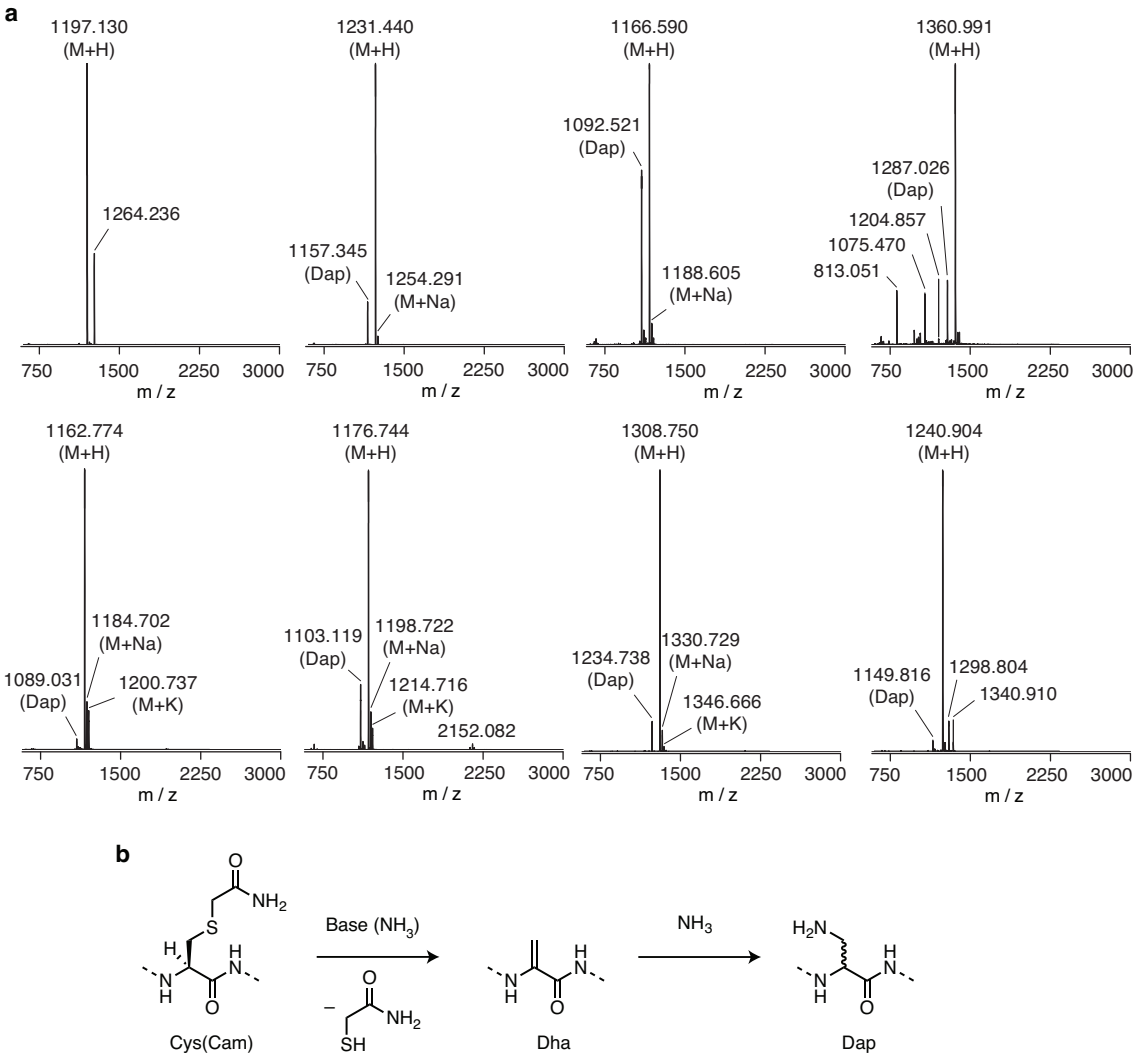


Figure 93. Library quality control. (**a**) MALDI–MS analysis of eight library members after cleavage and elution from single beads. Observed m/z spectra are shown for each library member, with the main signal (M+H) and further adducts (M+Na and M+K) annotated. (**b**) A signal at –74 m/z relative to M+H was observed in several library members, which is consistent with base-promoted elimination of Cys(Cam) to Dha followed by addition of ammonia to form Dap during linker cleavage.

2.2. Library screening assay

Compared to solution-phase libraries, use of a resin-supported library enabled us to perform a multi-step reaction and screening assay, simply by adding and draining reagent solutions (Figure 94a). We developed a screening procedure that included counter-selection against tags that do not rapidly and spontaneously oxidize to form an intramolecular disulfide bond, followed by identification of library hits that can be reduced again under mild conditions. To screen for spontaneous oxidation, the resin-supported library was fully reduced at the start of the assay and was subsequently kept in suspension in aqueous buffer at pH 7.4 in a vessel that was open to air. After 15 minutes, non-disulfide-bonded library members were stained with a cysteine-reactive red dye (rhodamine chloroacetamide **46**, see Figure 94b) and capped with

iodoacetamide. Subsequently, mild reduction was performed by treating the library suspension with 0.5 mM tris-(2-carboxyethyl)phosphine (TCEP) for 15 min, followed by staining of the reduced beads with a cysteine-reactive green dye (fluorescein chloroacetamide **47**, see Figure 94c). Capping of remaining cysteine residues with iodoacetamide was performed at each staining step and at the end of the assay to facilitate sequencing by MALDI–MS/MS. Using this screen, we hoped to identify library hits that exhibit the desired properties for the disulfide tag, specifically fast self-oxidation to form an internal disulfide and facile reduction for selective disulfide rebridging in native proteins (see Figure 90).

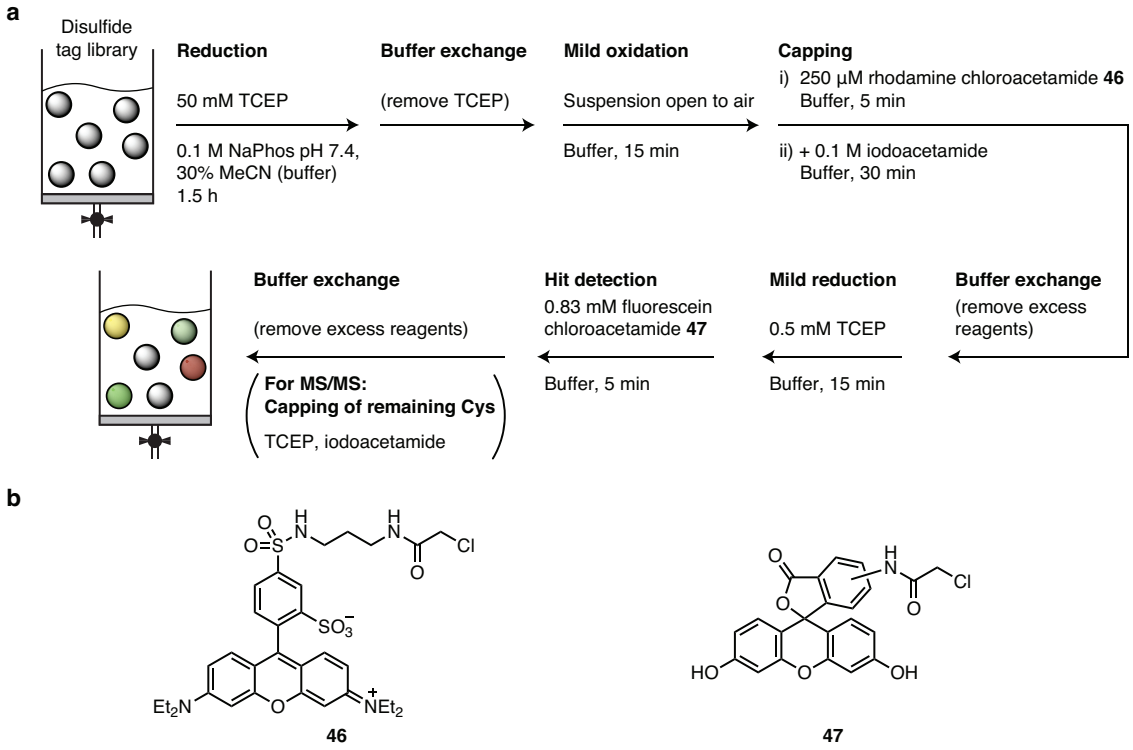


Figure 94. Disulfide tag library screening workflow. (**a**) Screening of the disulfide tag library was performed on resin as a suspension in buffer. Following complete reduction, spontaneous disulfide formation by exposure to air (open tube) was allowed to proceed for 15 min. Library members that did not form unreactive disulfide species were counter-selected using cysteine-reactive rhodamine chloroacetamide **46** and were capped with iodoacetamide. Following treatment with 0.5 mM TCEP for 15 min, library members that were reduced under these conditions were stained using fluorescein chloroacetamide **47**. Complete cysteine capping with iodoacetamide was performed at the end of the screen to facilitate peptide sequencing by MS/MS. (**b**) Structures of sulforhodamine B chloroacetamide **46** and 5(6)-aminofluorescein chloroacetamide **47**.

2.2.1. Fluorescence-activated bead sorting

Observation of the screened library under the microscope revealed beads with different staining patterns (Figure 95a). The staining patterns included strongly red (rhodamine) and green (fluorescein) beads, intermediate forms such as yellow beads (mixture of both dyes), as well as weakly stained or unstained beads. We employed fluorescence-activated bead sorting to isolate

beads that showed high green fluorescence and low red fluorescence in a high-throughput manner (Figure 95b).[382] Selected beads were automatically dispensed individually into 96-well plates for subsequent analysis. In line with the diverse staining pattern observed by eye, fluorescence-activated sorting showed a wide range of signals and enabled isolation of the desired population. As expected, control beads in which mild reduction before fluorescein addition was omitted displayed weak green signal (No TCEP, Figure 95c). Beads from library screens in which staining with either rhodamine or fluorescein was omitted clustered along the expected axes (Figure 95c).

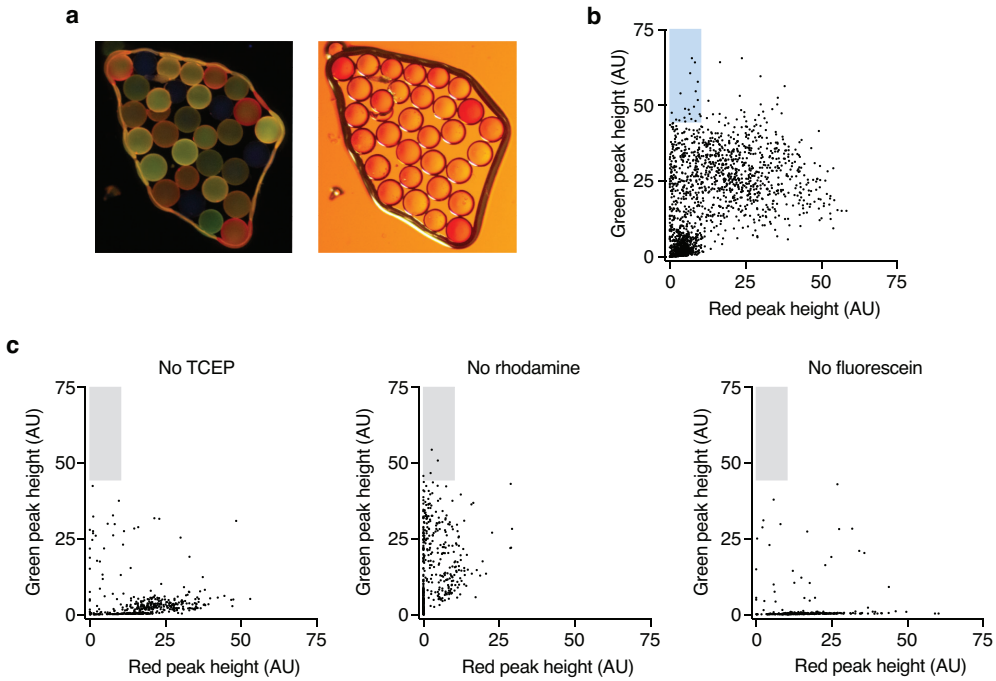


Figure 95. Fluorescence-activated bead sorting. (**a**) Images (inverted microscope) of beads from the one-bead one-peptide library after the disulfide tag screen, irradiated with a handheld UV lamp (365 nm) (left) and under white light (right). (**b**) Fluorescence-activated bead sorting of the library using red versus green fluorescence (peak height across the entire object) (representative result). Events that showed high green fluorescence and low red fluorescence were sorted (blue box, representative gate). (**c**) Sorting results of control beads in which mild reduction was omitted prior to fluorescein staining (no TCEP), or in which either no rhodamine or no fluorescein dye was added during the library screen. Grey boxes correspond to the gate in **b**.

2.3. Semi-automated peptide sequencing by tandem MALDI–MS/MS (TOF–TOF)

To sequence the library hits, we sought to perform tandem time-of-flight MALDI–MS/MS (TOF–TOF) and to analyze the MS/MS data using the Mascot server (Matrix Science)[383] with a custom data base containing all possible sequences of the library. To validate the approach, we re-synthesized three library members of known sequence by standard SPPS with Cam-capped cysteine residues and subjected the synthetic standards (**48**–**50**) to the outlined sequencing

workflow. We were pleased to find that, in all three cases, the sequence was correctly identified with high confidence (Figure 96).

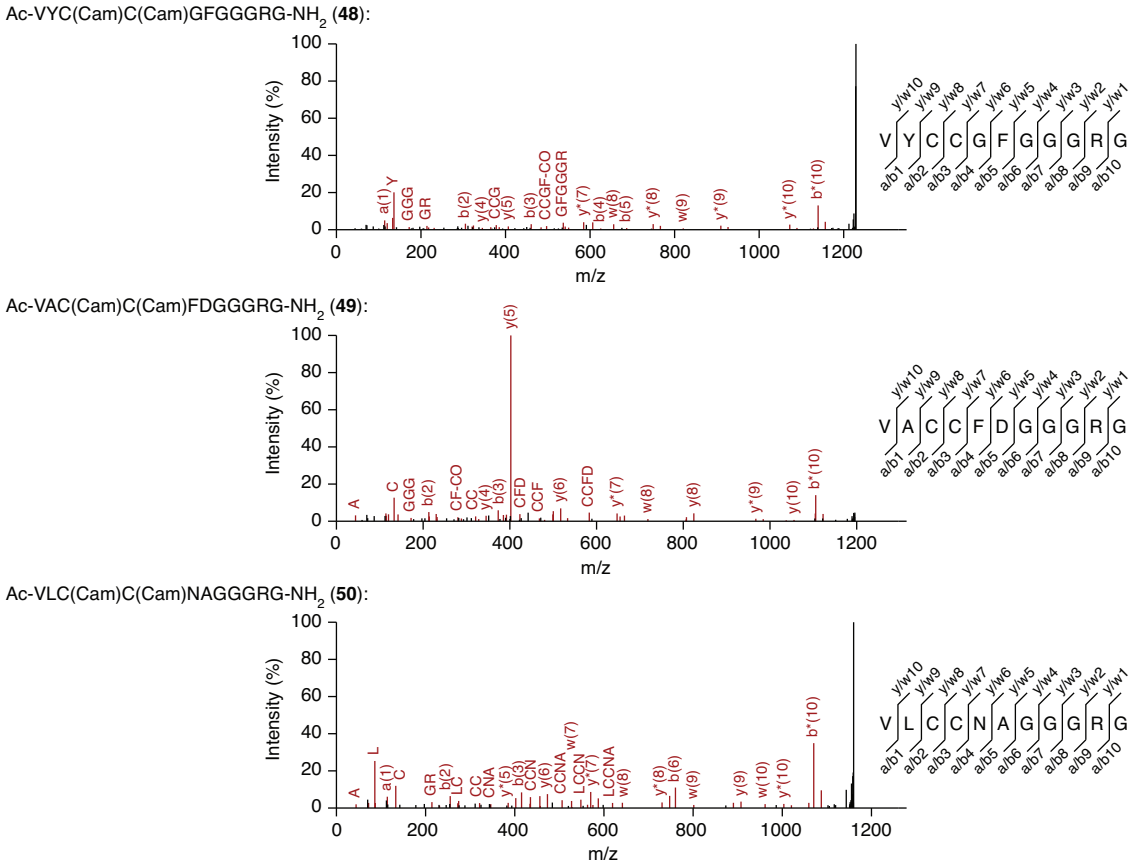


Figure 96. MS/MS sequencing results of the peptide standards **48**–**50**. The sequence hit from the custom data base with the highest confidence is shown on the right, and corresponding ions are annotated in red in the m/z spectra.

Overall, we sorted around 60,000 library members, which corresponds roughly to one library size (50,625 members) (Table 2). From these sorting events, approximately 2% were selected and gated as hits, of which a total of 970 library members were successfully obtained as isolated beads for sequencing.

Table 2. Library sorting and sequencing statistics.

Sorting		Peptide sequencing	
Analyzed events	Number of gated hits	Attempts	Successful
Library screening			
60,079	1247 (2.1%)	970	411 (42.4%)
Rejected beads			
–	–	192	126 (65.6%)

Following peptide cleavage in the 96-well format and manual spotting of the eluted samples onto a MALDI target plate, data acquisition and most of the data analysis were performed in an automated fashion. We identified the underlying peptide sequence of the library hits with high confidence in approximately 42% of the cases for which sequencing was attempted (411 sequences) (see Section 5 of the Appendix for a complete list). A custom Python script (see Experimental Part and Section 6 of the Appendix) was used to automate the analysis of the *de novo* peptide sequencing results. To this end, we took advantage of the fact that the eluted peptide samples from single beads were often obtained as mixtures with different cysteine modifications. The overall extent of labeling with fluorescein chloroacetamide **47** was low, even in the brightest library hits. The most prevalent cysteine modifications in the mixture were Cam and its side products Dha and Dap (see Figure 93). If present, multiple redundant MS/MS spectra were automatically acquired per sample, based on the differently modified species, which allowed us to independently verify the sequencing results in these cases. By comparing the sequences from the independent spectra and by scoring the sequencing results, we were able to identify true hits with high confidence with minimal user input. Additionally, the script included a test for potential contamination with multiple beads. This was achieved by assessing the relationship between the different parent ions associated with a sample. Samples with low confidence or that may have contained multiple beads were flagged for manual inspection. Lastly, the general properties of each library hit were determined and summarized automatically, including the numbers of charged, polar and apolar residues, and the net charge.

To test for a potential bias in the results that may be caused by sequencing success rather than the library screen and sort, we also isolated 192 individual beads which were rejected during fluorescence-activated bead sorting. By performing MALDI–MS/MS sequencing and data analysis analogous to that used for library hits, we successfully identified approximately 66% of these sequences with high confidence (Table 2).

3. Preliminary consensus motif

We analyzed the library hit sequences for the presence of a potential consensus motif by comparison to the expected distribution in the entire library (Figure 97a). Clear enrichment for glutamic acid and aspartic acid was found in the library hits, whereas apolar residues were depleted. A small enrichment for lysines was also observed. However, no clear consensus motif was observed besides the general enrichment based on the amino acid properties.

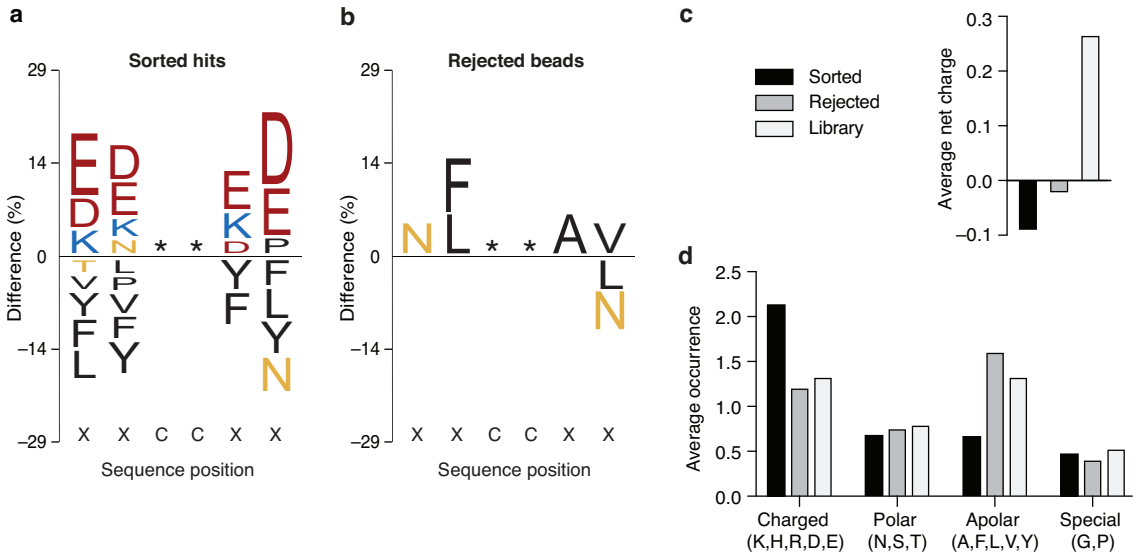


Figure 97. Sequence analysis of library hits. (**a**,**b**) iceLogo representation[384] of the library hits (**a**) and of the rejected beads (**b**).The percentage difference in the particular subset compared to the entire library is shown, with enriched residues above the x-axis and depleted residues below. The height of the amino acid letter corresponds to the associated value. Only changes with P value ≤ 0.05 are shown. (**c**) Average net charge of the library hits (sorted), the rejected beads, and of the entire library. (**d**) Average occurrence of amino acids with indicated properties across the sequence XXCCXX. Bar colors are the same as in panel **c**.

To exclude that the observed enrichment was simply due to the higher sequencing success of peptides containing these residues, we also analyzed the sequences from rejected beads in the same way (Figure 97b). In this set, no strong enrichment was observed. The slight over-representation of some hydrophobic residues may be a result of the fact that library hits containing more polar and charged residues were removed from this set of rejected beads during the sort. Overall, the comparable sequencing success (see Table 2), the lack of a strong enrichment in the set of rejected beads, and the different sequence pattern observed compared to the sorted hits, indicates that the enrichment in the library hits was not caused by sequencing bias.

The enrichment of negatively charged residues is also evident when comparing the average net charge of the library hits with that of rejected sequences or the entire library (Figure 97c). In this representation, a potential sequencing bias for negatively charged residues appears to be present, given that both the sorted and rejected beads showed a similar trend. However, other factors for the over-representation cannot be excluded, for example, different aggregation and solution behavior during the sort that may prevent successful isolation of single beads. Overall, the average occurrence of charged residues is almost twice as high as in the rejected sequences and the overall library, whereas apolar residues were two fold less prevalent (Figure 97d).

3.1. Test on GFP model substrates

We tested the redox behavior of several disulfide tags by fusing them to the N-terminus of GFP (CysTag-GFP variants) (Figure 98a). In the absence of a clear consensus motif, we

prepared eleven CysTag-GFP variants with sequences taken from the brightest-green beads (Figure 98b). This retrospective analysis was possible because library hits were individually analyzed in parallel and could be traced back to a specific sorting event. Additionally, we prepared three CysTag-GFP variants as negative controls (Neg1–3, Figure 98b) that contained the most depleted residues in the library hits (LYCCFN, see Figure 97a) and two other combinations of depleted residues (FFCCYY and YVCCFL).

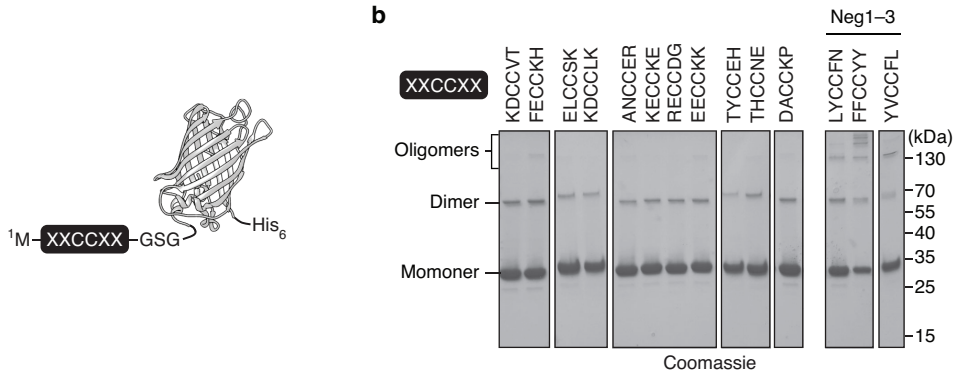


Figure 98. Expression and purification of CysTag-GFP variants. (**a**) CysTag-GFP variants containing vicinal disulfide tags (XXCCXX) fused to the N-terminus via a short flexible linker (GSG). The CysTags were installed directly after the start codon (Met1). (**b**) Coomassie-stained, non-reducing SDS–PAGE analysis of the CysTag-GFP variants, after lysis and purification by Ni-NTA affinity chromatography (C-terminal His_6-tag) under non-reducing conditions. Neg1–3: As negative controls, three CysTag-GFP variants were prepared that contained tags composed of residues which were depleted in the library hits (see Figure 97a).

Following recombinant expression and purification of the CysTag-GFP variants via a C-terminal His_6-tag under non-reducing conditions, the variants were analyzed for their propensity to form intermolecular disulfide bonds. All CysTag-GFP variants were mostly obtained as monomers as judged by non-reducing SDS–PAGE analysis, with a minor dimer-component present (Figure 98b). A similar pattern of monomer and dimer formation was observed for the negative control variants. Additionally, the negative control variants formed small amounts of tetramer and oligomer species (Neg1–3, Figure 98b).

Next, we performed a competition experiment for disulfide labeling between the CysTag-GFP variants and bovine pancreatic ribonuclease (RNase A),[385] which contains four native disulfide bonds (Figure 99a). Equal mixtures of the CysTag-GFP variants and RNase A were treated with 1 equivalent of TCEP for 15 min, followed by labeling of free cysteines with a rhodamine-functionalized methylsulfone oxadiazole reagent (**48**, Figure 99b). Exclusive labeling of the CysTag-GFP variants was observed, with little variation between the variants. Notably, the negative control variants were stained to a similar extent. A control reaction in the absence of GFP showed that the native disulfide bonds of RNase A alone were resistant to reduction and labeling under these conditions (no GFP, Figure 99a).

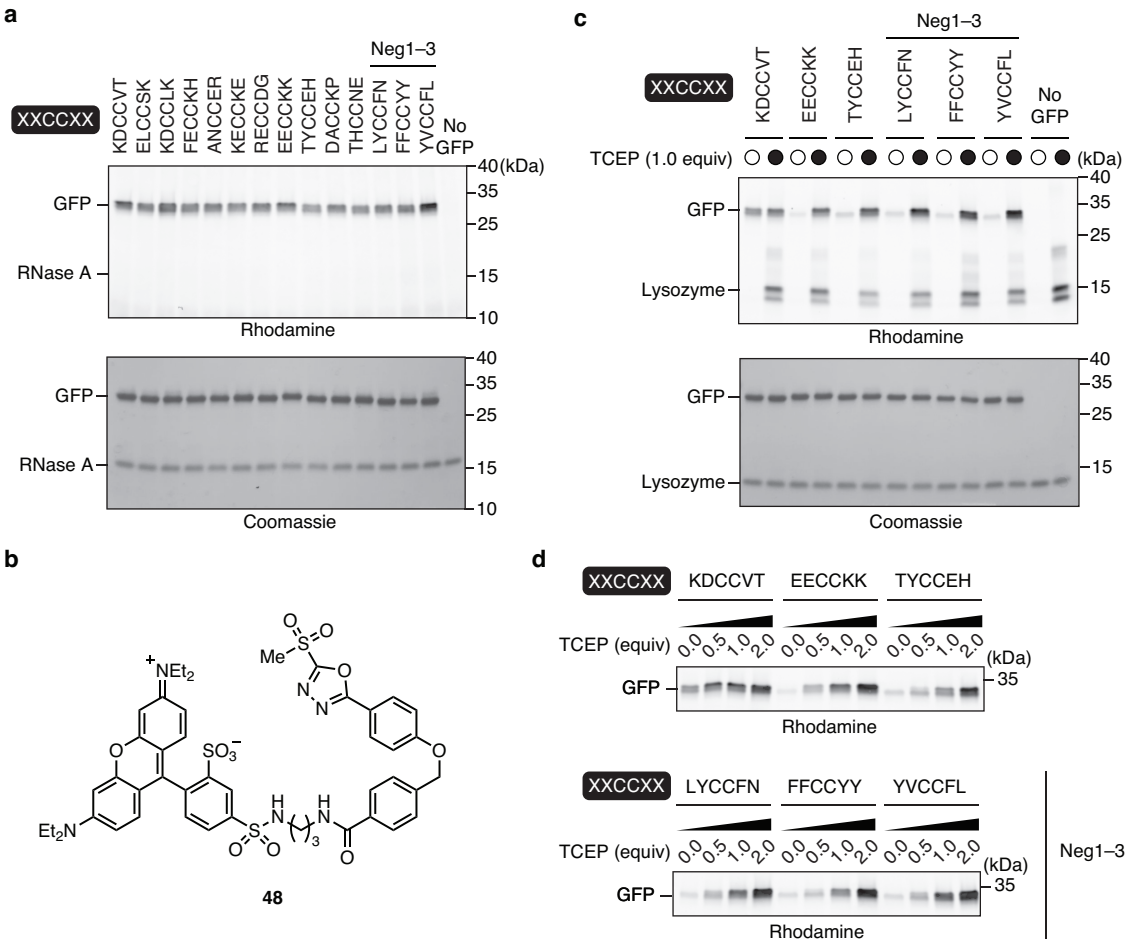


Figure 99. Redox reactivity tests of CysTag-GFP variants. (**a**) Competition experiment between indicated CysTag-GFP variants and RNase A (25 μM each). A control reaction without GFP variant was performed. Shown are in-gel fluorescence (top) and Coomassie-stained SDS–PAGE analyses (bottom) after treatment of the mixture with TCEP (1 equivalent) for 15 min at room temperature (rt), followed by labeling with the cysteine-reactive rhodamine probe **48** (4 equivalents) for 2 h at 0 °C. (**b**) Structure of the cysteine-reactive rhodamine probe **48**. (**c**) Competition experiment between indicated CysTag-GFP variants and chicken egg lysozyme (25 μM each). A control reaction without GFP variant was performed. Shown are in-gel fluorescence (top) and Coomassie-stained SDS–PAGE analyses (bottom) after incubation of the mixture in the presence (filled circle) or absence (hollow circle) of TCEP (1 equivalent) for 15 min at rt and subsequent labeling with the cysteine-reactive rhodamine probe **48** (4 equivalents) for 2 h at 0 °C. (**d**) In-gel fluorescence analyses of indicated CysTag-GFP variants (25 μM) after treatment with indicated amounts of TCEP (0–2 equivalents) for 15 min at rt, followed by labeling with the cysteine-reactive rhodamine probe **48** (4 equivalents) for 2 h at 0 °C.

As a more stringent test, we compared the redox reactivity of the CysTag-GFP variants to that of chicken egg lysozyme. Lysozyme contains four native disulfide bonds, of which one disulfide bond is particularly solvent-exposed and has been selectively reduced with TCEP and rebridged using allyl sulfone reagents.[33] Following an analogous competition procedure as before, we found that both the CysTag-GFP variants and lysozyme were stained to varying degrees after treatment with 1 equivalent of TCEP followed by rhodamine labeling with **48** (Figure 99c). Omission of TCEP before rhodamine labeling resulted in decreased CysTag-GFP-labeling,

although the labeling was not completely suppressed. This indicates that a small portion of the vicinal cysteine tag was present in the reduced form. In contrast, lysozyme was completely resistant to rhodamine staining in the absence of TCEP. While the CysTag-GFP variants generally were stained stronger than lysozyme, the selectivity was not complete. Lastly, no clear difference in redox reactivity between CysTag-GFP variants and the negative controls was observed (Neg1–3, Figure 99c). This was further confirmed by performing CysTag-GFP labeling reactions with rhodamine dye **48** after treatment of the substrates with varying amounts of TCEP (Figure 99d). CysTag-GFP variants and the negative controls exhibited largely similar staining patterns in response to increasing amounts of TCEP. The largest difference in labeling was observed for samples lacking TCEP. This result indicates that the level of auto-oxidation of the vicinal disulfides varied somewhat between the variants, whereas all tags were easily reduced by TCEP. Among the variants that were tested, the CysTag-GFP variant EECCKK showed the best response to TCEP with minimal staining before reduction (Figure 99d).

4. Discussion

In this project, we developed a workflow to screen for sequence-specific reactivity of peptides. To this end, we prepared a solid-supported combinatorial peptide library and developed a semi-automated peptide sequencing workflow to analyze library hits. We used this system to explore the redox reactivity of vicinal cysteines, which we identified as a potential candidate for redox-switchable disulfide tags for site-specific protein labeling via disulfide rebridging.

The solid-supported library proved to be ideally suited for the screening of redox reactivity, since immobilization of the library members on resin enabled multi-step reaction and rapid switching between conditions. A potential disadvantage of solid-supported libraries is the smaller throughput compared to solution-phase libraries. Screening in solution-phase enables testing of much larger libraries, and analysis of hits can be performed by LC–MS/MS analysis of complex mixtures. Nevertheless, an added benefit of the solid-supported library was that we were able to retrospectively analyze and compare the hit sequences with their properties in the fluorescence-activated bead sort to find sequences corresponding to the brightest beads in the screen. This direct association between a sorting event and hit identification is not possible in solution-phase. Development of a scripted data analysis and validation pipeline helped to increase the throughput of hit identification by MALDI–MS/MS.

The identified hit sequences were enriched in charged residues, especially glutamic and aspartic acid. This finding is somewhat unexpected, given that the amino acid pattern surrounding native disulfide bonds has been found to be enriched in weakly hydrophilic and aromatic residues.[364] Acidic residues in the vicinity of cysteines are expected to increase the pK_a of the thiol group, and hence decrease the amount of thiolate at a given reaction pH. This should result in slower disulfide exchange and oxidation.[365,366]

A potential bias may have been introduced by the choice of fluorescein for hit detection, which undergoes a red shift in increasingly apolar environments.[386] As a result of this solvatochromism, beads that were rich in hydrophilic residues may have appeared greener than library members that were highly hydrophobic. Indeed, control beads in which rhodamine labeling was omitted showed a relatively strong dispersion in the red channel, potentially caused by the solvatochromism of fluorescein (see Figure 95c). Although hit detection with sulforhodamine B was initially considered, tests showed that rhodamine-labeled peptides resulted in MS/MS fragmentation series of poor quality that were not suitable for peptide sequencing. A different dye should be considered in future library screens for hit detection by fluorescence-activated bead sorting.

While no clear consensus motif was found from the library screen, the tested vicinal cysteine tags exhibited many of the desired characteristics of a redox-switchable disulfide tag. The tested CysTag-GFP variants were mostly obtained as monomers following recombinant expression and purification under non-reducing conditions, which suggests that vicinal cysteines do not exhibit a strong propensity to form homodimers or oligomers. The tags that were selected as negative controls for the library hits were more prone to oligomerization. One reason for this result might be the hydrophobic nature of these negative control tags, which may result in increased aggregation and interaction between the disulfide tags. The CysTag-GFP variants were found to predominantly form intramolecular disulfide bonds, as evidenced by the low level of reactivity with cysteine-reactive dyes in the absence of pre-treatment with a reducing agent. Lastly, the tested CysTag-GFP variants exhibited moderate selectivity in competition reactions with chicken egg lysozyme for disulfide labeling, but could be selectively reduced in the presence of RNase A.

5. References

(357) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998**, *281*, 269–272.

(358) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Voorhis, T. V.; Pentelute, B. L. π-Clamp-mediated cysteine conjugation. *Nat. Chem.* **2016**, *8*, 120–128.

(359) Li, Z.; Huang, R.; Xu, H.; Chen, J.; Zhan, Y.; Zhou, X.; Chen, H.; Jiang, B. Divinylsulfonamides as specific linkers for stapling disulfide bonds in peptides. *Org. Lett.* **2017**, *19*, 4972–4975.

(360) Counsell, A. J.; Walsh, S. J.; Robertson, N. S.; Sore, H. F.; Spring, D. R. Efficient and selective antibody modification with functionalised divinyltriazines. *Org. Biomol. Chem.* **2020**, *18*, 4739–4743.

(361) Walsh, S. J.; Omarjee, S.; Galloway, W. R. J. D.; Kwan, T. T.-L.; Sore, H. F.; Parker, J. S.; Hyvönen, M.; Carroll, J. S.; Spring, D. R. A general approach for the site-selective modification of native proteins, enabling the generation of stable and functional antibody–drug conjugates. *Chem. Sci.* **2018**, *10*, 694–700.

(362) Sun, S.; Akkapeddi, P.; Marques, M. C.; Martínez-Sáez, N.; Torres, V. M.; Cordeiro, C.; Boutureira, O.; Bernardes, G. J. L. One-pot stapling of interchain disulfides of antibodies using an isobutylene motif. *Org. Biomol. Chem.* **2018**, *17*, 2005–2012.

(363) Martínez-Sáez, N.; Sun, S.; Oldrini, D.; Sormanni, P.; Boutureira, O.; Carboni, F.; Compañón, I.; Deery, M. J.; Vendruscolo, M.; Corzana, F.; Adamo, R.; Bernardes, G. J. L. Oxetane grafts installed site-selectively on native disulfides to enhance protein stability and activity *in vivo*. *Angew. Chem. Int. Ed.* **2017**, *56*, 14963–14967.

(364) Marques, J. R. F.; Fonseca, R. R. da; Drury, B.; Melo, A. Amino acid patterns around disulfide bonds. *Int. J. Mol. Sci.* **2010**, *11*, 4673–4686.

(365) Wu, C.; Belenda, C.; Leroux, J.; Gauthier, M. A. Interplay of chemical microenvironment and redox environment on thiol–disulfide exchange kinetics. *Chem. Eur. J.* **2011**, *17*, 10064–10070.

(366) Bermejo-Velasco, D.; Azémar, A.; Oommen, O. P.; Hilborn, J.; Varghese, O. P. Modulating thiol pK_a promotes disulfide formation at physiological pH: An elegant strategy to design disulfide cross-linked hyaluronic acid hydrogels. *Biomacromolecules* **2019**, *20*, 1412–1420.

(367) Zhang, R.; Snyder, G. H. Dependence of formation of small disulfide loops in two-cysteine peptides on the number and types of intervening amino acids. *J. Biol. Chem.* **1989**, *264*, 18472–18479.

(368) Fass, D.; Thorpe, C. Chemistry and enzymology of disulfide cross-linking in proteins. *Chem. Rev.* **2017**, *118*, 1169–1198.

(369) Cheng, Z.; Zhang, J.; Ballou, D. P.; Williams, C. H. Reactivity of thioredoxin as a protein thiol-disulfide oxidoreductase. *Chem. Rev.* **2011**, *111*, 5768–5783.

(370) Lu, S.; Wu, Y.; Li, J.; Meng, X.; Hu, C.; Zhao, Y.; Wu, C. Directed disulfide pairing and folding of peptides for the *de novo* development of multicyclic peptide libraries. *J. Am. Chem. Soc.* **2020**, *142*, 16285–16291.

(371) Wu, C.; Leroux, J.-C.; Gauthier, M. A. Twin disulfides for orthogonal disulfide pairing and the directed folding of multicyclic peptides. *Nat. Chem.* **2012**, *4*, 1044–1049.

(372) Schrimpf, A.; Hempel, F.; Li, A.; Linne, U.; Maier, U. G.; Reetz, M. T.; Geyer, A. Hinge-type dimerization of proteins by a tetracysteine peptide of high pairing specificity. *Biochemistry* **2018**, *57*, 3658–3664.

(373) Schrimpf, A.; Linne, U.; Geyer, A. Eight at one stroke – a synthetic tetra-disulfide peptide epitope. *Org. Biomol. Chem.* **2017**, *15*, 2512–2521.

(374) Agrawalla, B. K.; Wang, T.; Riegger, A.; Domogalla, M. P.; Steinbrink, K.; Dörfler, T.; Chen, X.; Boldt, F.; Lamla, M.; Michaelis, J.; Kuan, S. L.; Weil, T. Chemoselective dual labeling of native and recombinant proteins. *Bioconjug. Chem.* **2017**, *29*, 29–34.

(375) Richardson, J. S.; Videau, L. L.; Williams, C. J.; Richardson, D. C. Broad analysis of vicinal disulfides: Occurrences, conformations with *cis* or with *trans* peptides, and functional roles including sugar binding. *J. Mol. Biol.* **2017**, *429*, 1321–1335.

(376) Kilgore, H. R.; Raines, R. T. N $\rightarrow\pi^*$ Interactions modulate the properties of cysteine residues and disulfide bonds in proteins. *J. Am. Chem. Soc.* **2018**, *140*, 17606–17611.

(377) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, *354*, 82–84.

(378) Quarrell, R.; Claridge, T. D.; Weaver, G. W.; Lowe, G. Structure and properties of TentaGel resin beads: Implications for combinatorial library chemistry. *Mol. Divers.* **1996**, *1*, 223–232.

(379) Giudicessi, S. L.; Gurevich-Messina, J. M.; Martínez-Ceron, M. C.; Erra-Balsells, R.; Albericio, F.; Cascone, O.; Camperi, S. A. Friendly strategy to prepare encoded one bead–one compound cyclic peptide library. *ACS Comb. Sci.* **2013**, *15*, 525–529.

(380) Roepstorff, P.; Fohlman, J. Letter to the Editors: Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass. Spectrom.* **1984**, *11*, 601–601.

(381) Camperi, S. A.; Marani, M. M.; Iannucci, N. B.; Côté, S.; Albericio, F.; Cascone, O. An efficient strategy for the preparation of one-bead-one-peptide libraries on a new biocompatible solid support. *Tetrahedron Lett.* **2005**, *46*, 1561–1564.

(382) Marani, M. M.; Ceron, M. C. M.; Giudicessi, S. L.; Oliveira, E. de; Côté, S.; Erra-Balsells, R.; Albericio, F.; Cascone, O.; Camperi, S. A. Screening of one-bead-one-peptide combinatorial library using red fluorescent dyes. Presence of positive and false positive beads. *J. Comb. Chem.* **2009**, *11*, 146–150.

(383) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–3567.

(384) Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. Improved visualization of protein consensus sequences by IceLogo. *Nat. Methods* **2009**, *6*, 786–787.

(385) Sela, M.; White Jr, F. H.; Anfinsen, C. B. Reductive cleavage of disulfide bridges in ribonuclease. *Science* **1957**, *125*, 691–692.

(386) Naderi, F.; Farajtabar, A. Solvatochromism of fluorescein in aqueous aprotic solvents. *J. Mol. Liq.* **2016**, *221*, 102–107.

CHAPTER 5

Summary and Outlook

1. Summary and Outlook

Protein modification is important for a number of applications, including for biomedical research, production of protein-based pharmaceuticals, and materials sciences. An ideal protein modification reaction should rely on natural amino acids only for facile access to starting materials, require minimal or no changes to the substrate to be modified, and allow for the installation of diverse and easily accessible probes and modifications of interest under mild reaction conditions. Furthermore, the site of modification should be flexible with respect to the protein scaffold, ideally allowing internal labeling of side chains rather than protein termini.

The Ubl system is arguably the most prominent eukaryotic pathway for the generation of internal protein conjugates. It is responsible for the generation of isopeptide conjugates with, by and large, every cellular protein. This versatility is achieved by hundreds of dedicated enzymes. In Chapter 2, we describe the development of a chemoenzymatic method called LACE for site-specific isopeptide labeling of recombinant proteins at internal lysine residues. By breaking down the SUMOylation pathway, and by taking advantage of the intrinsic sequence specificity of Ubc9, we have shown that isopeptide labeling can be performed in a chemoenzymatic manner, relying only on short peptide thioesters as acyl donors, a genetically encoded recognition sequence (LACE tag), and a single enzyme (Ubc9). To our knowledge, this represents the first utilization of components of the Ubl pathway for chemoenzymatic and modification of recombinant protein in a general manner. Site-specific isopeptide conjugation between the C-terminus of peptidic probes and lysines of substrate proteins provides unique access to this product topology, with minimal changes required to the substrate. The recognition sequence for LACE represents one of the shortest tags to date for labeling of this kind.[387]

Chapter 3 describes the use of LACE for modification of diverse protein substrates with small molecules, peptides and protein domains. The facile and modular synthesis of peptide thioesters allowed us to prepare variants with biochemical probes and bioorthogonal handles. We found that the functionalized thioesters were accepted and transferred by Ubc9 with similar efficiency irrespective of the moiety attached, which underlines the broad substrate tolerance of Ubc9. We further extended the use of LACE to the transfer of recombinant protein thioesters, including ubiquitin, ISG15 and affibody domains. Site-specific installation of ubiquitin and ISG15 in a protein of interest is a particularly intriguing application of LACE, given the scarcity of existing approaches to access such products.[281]

The absence of a requisite secondary structure of the LACE tag for recognition by Ubc9 renders its incorporation into non-canonical substrates straightforward, for example by grafting the recognition sequence into a loop or fusing it to a protein terminus. The strict requirement for an extended conformation of the LACE tag may be one reason for the excellent specificity we observed. While the primary sequence motif [ILV]KX[ED] that is recognized by Ubc9 is present in approximately 30% of all proteins, only motifs present in loops, disordered or extended

conformations can be potentially recognized and modified by Ubc9. To expand the scope of this approach, it would nevertheless be desirable to develop acceptor tags that can function within secondary structures. Intriguingly, Ubc9 has been reported to SUMOylate the E2 conjugating enzyme E2-25K within the N-terminal α1-helix.[388] Even though this helix contains two consensus SUMOylation motifs based on the primary sequence, none of those are modified. Instead, it was found that a separate lysine was SUMOylated, and that mutation of specific residues in the vicinity of the modification site reduce this reactivity. This finding suggests that α-helix modification by Ubc9 proceeds with some sequence-dependence. Analyzing the sequence-requirements for α-helix modification using this system might represent a promising starting point towards LACE conjugation of α-helical acceptor substrates.

A potential limitation of LACE is the requirement for chemical activation of the acyl donor as a thioester. The need for protein thioesters currently precludes the use of LACE *in vivo*. A possible strategy to circumvent this limitation relies on enzymatic preactivation of C-terminal carboxylic acid probes using engineered E1 enzymes and ATP. This approach may provide a path to performing protein–protein conjugation inside living cells.

For prospective pharmaceutical applications of LACE conjugates, it would be interesting to determine their stability in serum and cell lysates. Given that the conjugate mimics the isopeptide bond of ubiquitin-modified proteins, the conjugate may be susceptible to cleavage by various cellular proteases, including DUBs. Many DUBs require recognition of the entire Ubl fold for efficient processing,[137] in which case the conjugates comprising just the ubiquitin C-terminus may be resistant to cleavage. On the other hand, cleavage of LACE conjugates by certain cellular proteases could be exploited as a strategy to release a conjugate upon internalization.

The wild type Ubc9 enzyme was capable of performing in a remarkably broad range of applications, including one-pot heterobifunctionalization of proteins in combination with other chemoenzymatic processes, installation of bioorthogonal handles as seen in the cytokine conjugates we generated, Ubl transfer, and functionalization of large, multimeric protein cages. Similar to the directed evolution of more proficient sortase variants,[229] further improvements of Ubc9 are conceivable. Avenues to explore in this regard could include selection of Ubc9 variants with different substrate specificities or enhanced reactivity. Nevertheless, our findings show that the rate of LACE with wild type Ubc9 is sufficiently high to facilitate conjugation of macromolecules under highly dilute conditions. By transferring protein thioesters, we demonstrate that LACE can effect protein–protein isopeptide conjugation under folded conditions from entirely recombinant starting materials.

In summary, LACE meets several of the criteria that are desired in protein modification reactions, including high site-specificity and mild reaction conditions, and sole reliance on a genetically encoded tag and natural amino acids. Additionally, the short recognition sequence requires minimal changes to the substrate and enables modification throughout the protein

sequence in a flexible and programmable manner. Lastly, we demonstrated a broad range of applications, including transfer of biochemical probes, peptides, and Ubl domains to diverse protein substrates with useful reaction rates and conversion.

In Chapter 4, we have developed a workflow to identify peptides that react sequence-specifically. To this end, we prepared a solid-supported combinatorial peptide library and developed a semi-automated peptide sequencing workflow to analyze library hits. We used this approach to explore the redox reactivity of vicinal cysteines, which we identified as potential candidates for redox-switchable disulfide tags for site-specific protein labeling via disulfide rebridging. While no clear consensus motif was obtained from a peptide library screen, vicinal cysteines displayed distinct redox reactivity that may warrant their further exploration as protein modification tags. In particular, vicinal cysteines were found to form intramolecular disulfide bonds rather than intermolecular ones when presented as surface-exposed tags on recombinant proteins, and they could be reduced rapidly using stoichiometric amounts of reducing agent. It is likely that this strategy can be adapted for identification of sequence-specific reactivities of residues other than cysteines.

2. References

(387) Fottner, M.; Lang, K. Decorating proteins with LACE. *Nat. Chem.* **2020**, *12*, 980–982.

(388) Pichler, A.; Knipscheer, P.; Oberhofer, E.; Dijk, W. J. van; Körner, R.; Olsen, J. V.; Jentsch, S.; Melchior, F.; Sixma, T. K. SUMO modification of the ubiquitin-conjugating enzyme E2-25K. *Nat. Struct. Mol. Biol.* **2005**, *12*, 264–269.

CHAPTER 6

Experimental Part

1. General methods and reagents

Synthetic methods and commercial materials

Reactions were carried out in oven-dried glassware under an atmosphere of dry N_2 using standard techniques. Thin layer chromatography (TLC) was performed on Merck TLC glass plates (0.25 mm) pre-coated with silica gel 60 F254 and visualized by UV quenching and/or potassium permanganate staining and warming with a heat gun. Solvents were removed by rotary evaporation at 40 °C under reduced pressure. Chemical reagents were purchased from Sigma Aldrich (Buchs, Switzerland), Acros Organics (Geel, Belgium) and TCI Europe (Zwijndrecht, Belgium) and used without further purification. MeCN, DMF, Et_2O, EtOH, MeOH and THF were purchased from Fisher Scientific (Geel, Belgium) and Sigma Aldrich (Buchs, Switzerland) and used as supplied (reagent or HPLC grade). Milli-Q water was obtained from a Millipore purification system. Cyclohexane, EtOAc and CH_2Cl_2 were of technical grade and distilled prior to use. Amino acid derivatives for SPPS and Fmoc-Rink-Amide-MBHA resin were purchased from Peptide International (Louisville, KY, USA) and Merck (Darmstadt, Germany). 2-Chlorotritylchloride polystyrene resin was purchased from Christof Senn Laboratories AG (Dielsdorf, Switzerland). 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) and 2-(6-Chloro-1-H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU) were purchased from Chem-Impex (Wood Dale, IL, USA). Restriction enzymes, Phusion® and Q5® High-Fidelity DNA polymerases, and PCR reagents were purchased from New England BioLabs (Ipswich, MA, USA). DNase I was obtained from Roche Diagnostics GmbH (Mannheim, Germany). pRK793 was a gift from David Waugh (Addgene plasmid #8827) and was used to express the Tobacco etch virus (TEV) protease variant His_6-TEV-S219V-Arg_7 as an MBP fusion as reported.[1] pET39b-7M SrtA was a gift from Hidde Ploegh (Addgene plasmid #51141) and was used to express $SrtA^{7M}$-His_6 as reported.[2] DNA purification kits were purchased from Fisher Scientific (Geel, Belgium) and Zymo Research (Irvine, CA, USA). Benzonase was obtained from Merck (Darmstadt, Germany). Lysozyme (22500 U/mg) was obtained from Axon Lab AG (Baden, Switzerland). Ampicillin sodium salt and kanamycin sulfate were obtained from AppliChem GmbH (Darmstadt, Germany). Ni-NTA agarose resin was obtained from Qiagen GmbH (Hilden, Germany). Dialysis tubing and devices (SnakeSkin™ dialysis tubing, or Slide-A-Lyzer™ Mini dialysis devices for samples smaller than 2 mL, both with 3.5 kDa molecular weight

1 Kapust, R. B.; Tözsér, J.; Fox, J. D.; Anderson, E. D.; Cherry, S.; Copeland, T. D.; Waugh, D. S. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **2001**, *14*, 993–1000.

2 Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L. Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **2013**, *8*, 1787–1799.

cutoff (MWCO)) were obtained from Thermo Fisher Scientific (Waltham, MA, USA). Amicon® Ultra centrifugal filters were purchased from Merck (Darmstadt, Germany), VivaSpin 500 centrifugal concentrators from Sartorius Stedim Lab (Stonehouse, UK), and PD Mini-/MidiTrap desalting columns from GE Healthcare. All buffers were prepared using Milli-Q water, pH adjusted for the temperature at which the buffer was used, and sterile-filtered (0.2 μm membrane filter). Oligonucleotide synthesis and sequencing was carried out by Microsynth AG (Balgach, Switzerland). Synthetic genes were purchased from ATG Biosynthetics (Merzhausen, Germany) and Twist Bioscience (San Francisco, CA, USA).

Peptide synthesis

SPPS was performed using the following Boc and Fmoc amino acids with suitable side-chain protecting groups: Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Cys(Cam)-OH, Fmoc-Cys(Trt)-OH, Boc-Gly-OH, Fmoc-Gly-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Alloc)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Val-OH. Before synthesis, the resin was swollen in CH_2Cl_2/DMF (1:1) for 30 min. Fmoc deprotection was performed with 20% piperidine in DMF (2 x 10 min), and in the case of ubiquitin peptide sequences (RLR or LRL) with 20% piperidine in DMF (10 min) and 2% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in DMF (5 min). Couplings were performed with Fmoc (or Boc-Gly-OH) amino acid (4.1 equiv to resin substitution), HATU (4.0 equiv) and 4-methylmorpholine (NMM, 8.0 equiv) in a minimal amount of DMF to dissolve the reagents. After premixing for 3 min, the pre-activated solution was added to the resin. Couplings were carried out twice for 45 min. Capping was performed with 20% Ac_2O and 10% NMM in DMF (10 min) after each coupling cycle. During all steps, the resin was agitated by shaking or N_2 bubbling, and the resin was washed with DMF (6 times) between each reaction step. Following SPPS, the resin was washed with CH_2Cl_2, dried under vacuum and stored at 4 °C. Global deprotection and cleavage from resin was carried out in 95% trifluoroacetic acid (TFA), 2.5% triisopropyl silane (TIPS) and 2.5% water (1 mL cleavage cocktail per 100 mg resin) at room temperature (rt) for 2 h, before the resin was removed by filtration. The cleavage cocktail was concentrated under reduced pressure and the peptide was precipitated and triturated three times with cold Et_2O. The crude peptide was dissolved in Milli-Q water with 30% MeCN and 0.1% TFA, lyophilized and stored at –20 °C until purification by RP–HPLC or thioester formation.

Reversed-phase high-performance liquid chromatography (RP–HPLC)

Unless otherwise mentioned, RP–HPLC was performed as follows: RP–HPLC was carried out on Jasco analytical and preparative instruments with simultaneous monitoring of the eluent at 220 nm, 254 nm and 301 nm at rt. The mobile phase was Milli-Q water with 0.1% TFA (solvent A) and HPLC-grade MeCN with 0.1% TFA (solvent B). Analytical RP–HPLC was performed on a

Shiseido C18 (5 μm, 4.6 mm I.D. x 250 mm) column at a flow rate of 1 mL/min, using the following method: 10% solvent B for 3 min, 10 to 95% solvent B over 14 min, and 95% solvent B for 3 min. Preparative RP–HPLC was performed on a YMC C18 (5 μm, 20 mm I.D. x 250 mm) column at a flow rate of 10 mL/min. HPLC data was acquired using EZChrom Elite (Version 3.3.2).

Gel electrophoresis

SDS–PAGE was carried out on a Mini-PROTEAN Tetra Cell system (Bio-Rad) connected to a PowerPac Basic (Bio-Rad) programmable power supply. Samples were separated on 8–16% gradient Mini-PROTEAN TGX Precast gels (Bio-Rad) for 30 min at 200 V. A 10–180 kDa pre-stained protein ladder (Thermo Fisher) was applied to at least one well of each gel (5 μL). Gels were imaged on a Bio-Rad Molecular Imager Pharos FX (TAMRA fluorescence), or Bio-Rad ChemiDoc MP Imaging System (Alexa488 and rhodamine fluorescence, Coomassie staining, and chemiluminescence) using the Image Lab Touch Software (Version 2.4.0.03). Images were cropped for illustration purposes. After in-gel fluorescence imaging, gels were stained with Coomassie (0.1% Coomassie Brilliant Blue R, 40% MeOH, 10% acetic acid) with gentle agitation for 1 h, followed by destaining (40% MeOH, 10% acetic acid).

Reducing SDS–PAGE samples

Reducing SDS–PAGE samples were treated with an equal volume of sample buffer (Laemmli 2x Concentrate, Sigma Aldrich), placed in a heat block at 95 °C for 5 min and stored at –20 °C until separation.

Reducing SDS–PAGE samples with thioesters

For SDS–PAGE samples containing thioesters, which is every LACE reaction, the sample buffer (Laemmli 2x Concentrate, Sigma Aldrich) was supplemented with 100 mM 2-aminoethanethiol hydrochloride (AET-HCl) (50 mM final concentration), unless otherwise mentioned. The sample was incubated at rt for 5 min to quench unreacted thioester prior to boiling for 5 min at 95 °C. The quenched sample was stored at –20 °C until separation.

Nonreducing SDS–PAGE samples

Non-reducing SDS–PAGE samples were treated with an equal volume of an in-house prepared buffer containing 100 mM tris(hydroxymethyl)aminomethane (Tris)-HCl pH 6.8, 2% SDS, 4 M urea, 20% glycerol and 0.004% w/v bromophenol blue (2x concentrated). The sample was not boiled and separated directly.

Western blotting

Samples were separated by SDS–PAGE and transferred to PVDF membranes using a Trans-Blot Turbo Transfer System (both Bio-Rad). Membranes were blocked for 1 h with 5% skim milk in phosphate-buffered saline (PBS) with 1% Tween 20 (PBST), washed with PBST (3 x 5 min),

and incubated at rt for 1.5 h in PBST with Pierce High Sensitivity Streptavidin-HRP (Thermo Fisher Scientific) at a dilution of 1:5000. Membranes were washed with PBST (3 x 5 min) and images were developed using Clarity Western ECL chemiluminescence substrate (Bio-Rad).

Quantification

Densitometry analysis was performed using Fiji (ImageJ Version 2.0).[3] Data was analyzed and plotted using Prism (Version 8). Reaction half times were estimated by fitting a first order exponential function to band intensities over time (determined by densitometry analysis of in-gel rhodamine fluorescence, or streptavidin-HRP chemiluminescence blot for reactions with biotin). Reaction conversion based on Coomassie-stained bands was estimated by densitometry according to the following equation:

$$\textit{Conversion to } A\ [\%] = A / (A + B + ...) * 100,$$

where A and B are intensities of bands normalized by their molecular weight. Distinct reactions were performed for replicates.

Fast protein liquid chromatography (FPLC)

Unless otherwise specified, protein chromatography was performed on an ÄKTA pure chromatography system (GE Healthcare) at 4 °C. Cation exchange chromatography was performed using the strong ion exchange column Mono S 5/50 GL, anion exchange chromatography was performed using the strong ion exchange columns Mono Q 5/50 GL and HiTrap Capto Q, and size-exclusion chromatography (SEC) was performed using a Superdex 75 Increase 10/300 GL column. All columns were purchased from GE Healthcare. Protein elution was monitored at 280 nm. FPLC data was acquired using UNICORN (Version 6.3.2.89).

Protein quantification and UV-Vis spectroscopy

Protein concentration and OD_{600} measurements were carried out on a NanoDrop 2000c UV-Vis spectrophotometer. Protein concentrations were determined by the absorption at 280 nm using extinction coefficients calculated by ProtParam[4] based on the amino acid sequence.

---

3 Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; Tinevez, J.-Y.; White, D. J.; Hartenstein, V.; Eliceiri, K.; Tomancak, P.; Cardona, A. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676–682.

4 Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Springer Protocols Handbooks; Humana Press Inc.: Totowa, NJ, 2005; pp 571–607.

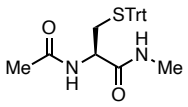
Characterization

Unless otherwise mentioned, high-resolution mass spectra (HR–MS) were obtained by the mass spectrometry service of the ETH Laboratorium für Organische Chemie on a Bruker Daltonics maXis ESI–QTOF spectrometer (ESI) or a Bruker Daltonics solariX spectrometer (MALDI–FTICR–MS) using α-cyano-4-hydroxycinnamic acid (CHCA) as matrix. Low resolution mass spectra were recorded on a Bruker Microflex Silver Edition MALDI–MS using CHCA as matrix. Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker AV-400 or AV-500 spectrometer. Chemical shifts (δ) are given in ppm relative to residual solvent peaks. Data for ^{1}H NMR are reported as follows: chemical shift (multiplicity, coupling constants where applicable, number of hydrogens). Abbreviations are as follows: s (singlet), d (doublet), dd (doublet of doublet), ddd (doublet of doublet of doublet), t (triplet), q (quartet), qd (quartet of doublet), td (triplet of doublet) and m (multiplet). Infrared (IR) spectra were recorded on a Jasco FT/IR-4100 spectrometer and only major peaks are reported in frequency of absorption (cm^{-1}). Optical rotations were measured on a Jasco P-2000 operating at the sodium D line with a 100 mm path length cell and reported as $[\alpha]^T_D$ (concentration (g/100 mL), solvent) where T was the temperature in °C.

2. Synthesis

2.1. Ac-Cys-NHMe (1)[5]

Ac-Cys(Trt)-NHMe (S1)



To a stirred suspension of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) hydrochloride (4.98 g, 26.0 mmol, 1.30 equiv) in THF/DMF (1:1, 35 mL) was added a suspension of Fmoc-Cys(Trt)-OH (11.71 g, 20.0 mmol, 1.00 equiv) and NHS (3.06 g, 26.6 mmol, 1.33 equiv) in THF (60 mL) over 10 min. The reaction was stirred at rt for 5 h before methylamine (2.74 mL of 33% w/w in EtOH, 22.0 mmol, 1.10 equiv) was added, and stirring was continued at rt for 2 h. The reaction was diluted with EtOAc (50 mL) and the combined organic phase was washed with H_2O/brine, dried over $MgSO_4$, filtered and concentrated under reduced pressure to give crude Fmoc-Cys(Trt)-NHMe as a yellow foam, which was used without further purification. To a suspension of crude Fmoc-Cys(Trt)-NHMe (approximately 20 mmol, 1 equiv) in DMF (80 mL) was added piperidine (10 mL, 100 mmol, 5 equiv) and the reaction was stirred at rt for 15 min. The reaction was diluted with EtOAc (80 mL) and the combined organic phase was washed with 0.2 M phosphate citrate buffer pH 6 (2 x 200 mL) until the aqueous phase was neutralized (pH 6–7). The organic phase was washed with H_2O/brine, dried over $MgSO_4$, filtered and concentrated to give crude H-Cys(Trt)-NHMe as an off-white foam, which was used without further purification. To a suspension of crude H-Cys(Trt)-NHMe (approximately 20 mmol, 1 equiv) in DMF (100 mL) was added NMM (6.60 mL, 60 mmol, 3 equiv) and Ac_2O (2.83 mL, 30 mmol, 1.5 equiv). The reaction was stirred at rt for 2 h. The reaction was diluted with EtOAc (500 mL) and CH_2Cl_2 (100 mL) and the combined organic phase was washed with H_2O/brine (2x), dried over $MgSO_4$, filtered and concentrated under reduced pressure. The product was recrystallized from EtOAc/cyclohexane (4:1, 100 mL), washed with EtOAc and dried under reduced pressure to give **S1** (4.96 g, 11.9 mmol, 59% overall yield) as a white solid.

^{1}H NMR (400 MHz, DMSO-d_6): δ 8.12 (d, J = 8.4 Hz, 1H), 7.87 (q, J = 4.6 Hz, 1H), 7.35 – 7.21 (m, 15H), 4.32 (td, J = 8.2, 6.0 Hz, 1H), 2.53 (d, J = 4.6 Hz, 3H), 2.39 – 2.25 (m, 2H), 1.82 (s, 3H).

^{13}C NMR (101 MHz, DMSO-d_6): δ 170.1, 169.1, 144.3, 129.1, 128.0, 126.8, 65.8, 51.5, 34.0, 25.6, 22.5.

IR (cm^{-1}, neat): ν 3281, 3053, 1742, 1635, 1532, 1442, 1372, 1237.

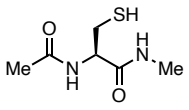
$[\alpha]^{24}_D$ = +10.9 (c = 0.542, $CHCl_3$).

HR–MS (MALDI–FTICR–MS): calc. for $C_{25}H_{26}N_2NaO_2S$ $[M+Na]^+$: 441.1607, obs.: 441.1608.

---

5 As reported in: Hofmann, R.; Akimoto, G.; Wucherpfennig, T. G.; Zeymer, C.; Bode, J. W. Lysine acylation using conjugating enzymes for site-specific modification and ubiquitination of recombinant proteins. *Nat. Chem.* **2020**, *12*, 1008–1015.

Ac-Cys-NHMe (1)



To a stirred solution of Ac-Cys(Trt)-NHMe **S1** (2.0 g, 4.8 mmol, 1.0 equiv) and TIPS (2.4 mL, 12.8 mmol, 2.5 equiv) in CH_2Cl_2 (10 mL) was added TFA (5 mL) over 15 min. The reaction was stirred at rt for 30 min, concentrated under reduced pressure, and the precipitate was triturated with cold Et_2O (2 x 30 mL). The product was dissolved in Milli-Q water with 30% MeCN and 0.1% TFA and lyophilized to give **1** (0.8 g, 4.5 mmol, 94%) as a white solid.

^{1}H NMR (400 MHz, DMSO-d_6): δ 8.06 (d, J = 8.2 Hz, 1H), 7.91 (q, J = 4.5 Hz, 1H), 4.28 (td, J = 7.9, 5.6 Hz, 1H), 2.80 – 2.60 (m, 2H), 2.58 (d, J = 4.6 Hz, 3H), 2.26 (s, 1H), 1.87 (s, 3H).

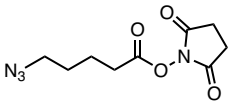
^{13}C NMR (101 MHz, DMSO-d_6): δ 170.2, 169.4, 55.1, 26.0, 25.6, 22.6.

IR (cm^{-1}, neat): ν 3282, 3118, 2974, 2933, 2554, 1631, 1541, 1372.

$[\alpha]^{27}_D$ = -20.6 (c = 0.095, $CHCl_3$).

HR–MS (ESI): calc. for $C_6H_{12}N_2NaO_2S$ $[M+Na]^+$: 199.0512, obs.: 199.0509.

2.2. 2,5-dioxopyrrolidin-1-yl 5-azidopentanoate (27)



To a stirred suspension of EDC hydrochloride (1.39 g, 7.24 mmol, 1.30 equiv) in DMF (5 mL) was added dropwise a solution of 5-azidopentanoic acid[6] (0.80 g, 5.57 mmol, 1.00 equiv) and NHS (0.85 g, 7.41 mmol, 1.33 equiv) in DMF (20 mL). The suspension was stirred at rt overnight. The reaction was diluted with EtOAc (100 mL) and the combined organic phase was washed with aq. sat. NH_4Cl (2 x 100 mL) and brine (50 mL), dried over $MgSO_4$, filtered, concentrated under reduced pressure and purified by flash column chromatography (CH_2Cl_2/MeOH 50:1) to give **27** (0.39 g, 1.62 mmol, 29%) as a colorless oil.

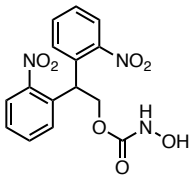
^{1}H NMR (400 MHz, $CDCl_3$): δ 3.34 (t, J = 6.6 Hz, 2H), 2.91 – 2.77 (m, 4H), 2.66 (t, J = 7.2 Hz, 2H), 1.91 – 1.79 (m, 2H), 1.79 – 1.66 (m, 2H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 169.20, 168.31, 50.97, 30.57, 28.02, 25.73, 21.98.

6 Purwin, M.; Hernández-Toribio, J.; Coderch, C.; Panchuk, R.; Skorokhyd, N.; Filipiak, K.; Pascual-Teresa, B. de; Ramos, A. Design and synthesis of novel dual-target agents for HDAC1 and CK2 inhibition. *RSC Adv.* **2016**, *6*, 66595–66608.

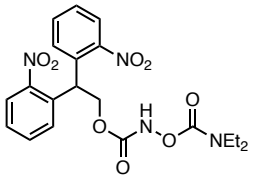
2.3. 4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)oxy)amino)butanoic acid (S2)

2,2-bis(2-nitrophenyl)ethyl hydroxycarbamate (S3):



To a stirred suspension of hydroxylamine hydrochloride (324 mg, 4.66 mmol, 2 equiv) in Et_2O (35 mL) and H_2O (7 mL) and K_2CO_3 (966 mg, 6.99 mmol, 3 equiv) at 0 °C was added 2,2-bis(2-nitrophenyl)ethyl (2,5-dioxopyrrolidin-1-yl) carbonate[7] (1.00 g, 2.33 mmol, 1 equiv). After 10 min, the reaction was allowed to reach rt while stirring for another 3 h. The reaction was diluted with aq. 1 N HCl (30 mL) and H_2O (20 mL) and the aqueous phase was extracted with Et_2O. The combined organic phase was dried over $MgSO_4$, filtered and concentrated under reduced pressure to yield crude **S3** as an off-white foam (quant.), which was used without further purification.

2,2-bis(2-nitrophenyl)ethyl ((diethylcarbamoyl)oxy)carbamate (S4)



To a stirred suspension of **S3** (243 mg, 0.70 mmol, 1.0 equiv) in dry CH_2Cl_2 (10 mL) was added NEt_3 (195 μL, 1.40 mmol, 2.0 equiv), 4-dimethylaminopyiridine (DMAP) (17 mg, 0.14 mmol, 0.2 equiv) and *N*,*N*-diethylcarbamoyl chloride (133 μL, 1.05 mmol, 1.5 equiv). The solution was stirred at rt for 24 h. The reaction was diluted with aq. sat. NH_4Cl (50 mL) and extracted with CH_2Cl_2 (3 x 50 mL). The combined organic phase was dried over $MgSO_4$, filtered, concentrated under reduced pressure and purified by flash column chromatography (EtOAc/cyclohexane 2:1) to give **S4** (285 mg, 0.64 mmol, 91%) as an off-white foam.

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.96 (dd, *J* = 8.1, 1.4 Hz, 2H), 7.93 (s, 1H), 7.58 (td, *J* = 7.6, 1.4 Hz, 2H), 7.47 – 7.36 (m, 4H), 5.74 (t, *J* = 6.7 Hz, 1H), 4.88 (d, *J* = 6.7 Hz, 2H), 3.26 (q, *J* = 7.1 Hz, 4H), 1.11 (t, *J* = 7.1 Hz, 6H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 156.82 (CO), 155.17 (CO), 149.59 (C), 133.73 (C), 133.27 (CH), 130.02 (CH), 128.59 (CH), 125.69 (CH), 66.00 (CH_2), 43.13* (CH_2), 41.57* (CH_2), 40.78 (CH), 13.87* (CH_3), 13.31* (CH_3). *Rotamers.

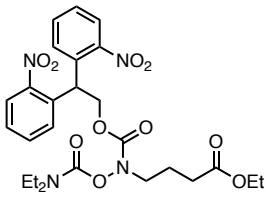
IR (cm^{-1}, thin film): ν 3669, 3237, 2976, 2360, 1726, 1523, 1347, 1148, 912, 855, 784, 737.

HR–MS (ESI): calc. for $C_{20}H_{23}N_4O_8$ $[M+H]^+$: 447.1510, obs.: 447.1507.

---

7 Kindly provided by Dr. S. Antunes (ETH Zürich, Bode Group). See also: Rohrbacher, N. F. Synthesis of cyclic peptides and proteins by α-ketoacid-hydroxylamine ligation. Doctoral dissertation, ETH Zürich, Switzerland, 2017.

ethyl 4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)oxy)amino)butanoate (S5)



To a stirred solution of **S4** (285 mg, 0.64 mmol, 1.0 equiv) in dry DMF (10 mL) at 0 °C was added K_2CO_3 (177 mg, 1.28 mmol, 2.0 equiv) and ethyl 4-bromobutanoate (138 μL, 0.96 mmol, 1.5 equiv). The reaction was stirred at 0 °C for 1 h and at rt for 20 h. Solid K_2CO_3 was removed by filtration, the filtrate was washed with H_2O (30 mL) and brine (30 mL). The organic phase was dried over $MgSO_4$, filtered, concentrated under reduced pressure and purified by flash column chromatography (EtOAc/cyclohexane 2:3) to give **S5** (337 mg, 0.60 mmol, 94%) as a yellow oil.

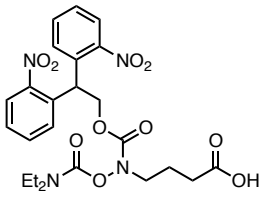
^{1}H NMR (400 MHz, $CDCl_3$): δ 7.94 (dd, *J* = 8.1, 1.4 Hz, 2H), 7.57 (td, *J* = 7.6, 1.5 Hz, 2H), 7.46 – 7.37 (m, 4H), 5.69 (t, *J* = 6.7 Hz, 1H), 4.82 (d, *J* = 6.7 Hz, 2H), 4.10 (q, *J* = 7.1 Hz, 2H), 3.63 (m, 2H), 3.24 – 3.06 (m, 4H), 2.32 (t, *J* = 7.4 Hz, 2H), 1.84 (m, 2H), 1.23 (t, *J* = 7.1 Hz, 3H), 1.01 (m, 6H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 173.00 (CO), 155.18 (CO), 153.71 (CO), 149.56 (C), 133.95 (C), 133.20 (CH), 130.19 (CH), 128.45 (CH), 125.56 (CH), 65.97 (CH_2), 60.52 (CH_2), 49.78 (CH_2), 42.98* (CH_2), 41.53* (CH_2), 40.84 (CH), 31.31 (CH_2), 22.60 (CH_2), 14.33 (CH_3), 13.91* (CH_3), 13.28* (CH_3). *Rotamers.

IR (cm^{-1}, thin film): ν 2977, 2935, 1732, 1524, 1348, 1143, 855, 786, 742.

HR–MS (ESI): calc. for $C_{26}H_{33}N_4O_{10}$ $[M+H]^+$: 561.2191, obs.: 561.2195.

4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)-oxy)amino)butanoic acid (S2)



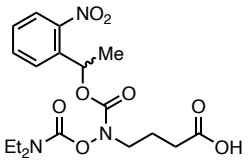
To a stirred solution of **S5** (27 mg, 48 μmol, 1.0 equiv) in THF (0.75 mL), MeOH (0.25 mL) and H_2O (0.25 mL) at 0 °C was added LiOH monohydrate (4 mg, 96 μmol, 2.0 equiv) and H_2O_2 (40 μL of a 30% solution (9.79 M), 385 μmol, 8.0 equiv). The solution was stirred at 0 °C for 2 h and at rt for 20 h. To the mixture was added aq. 1 N HCl (pH 0) and the mixture was extracted with EtOAc. The combined organic phase was dried over $MgSO_4$, filtered, concentrated under reduced pressure and purified by flash column chromatography (EtOAc/cyclohexane 3:2 with 1% AcOH) to give **S2** (10 mg, 19 μmol, 40%) as a pale yellow paste.

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.94 (dd, *J* = 8.1, 1.4 Hz, 2H), 7.58 (td, *J* = 7.6, 1.4 Hz, 2H), 7.41 (qd, *J* = 7.8, 1.4 Hz, 4H), 5.69 (t, *J* = 6.7 Hz, 1H), 4.83 (d, *J* = 6.7 Hz, 2H), 3.69 – 3.61 (m, 2H), 3.23 – 3.06 (m, 4H), 2.39 (t, *J* = 7.3 Hz, 2H), 1.83 (q, *J* = 7.0 Hz, 2H), 1.08 – 0.94 (m, 6H).

^{13}C NMR (101 MHz, $CDCl_3$): δ 178.13 (CO), 155.23 (CO), 153.79 (CO), 149.55 (C), 133.90 (C), 133.24 (CH), 130.16 (CH), 128.50 (CH), 125.58 (CH), 66.05 (CH_2), 49.64 (CH_2), 43.03* (CH_2), 41.56* (CH_2), 40.83 (CH), 30.89 (CH_2), 22.32 (CH_2), 13.90* (CH_3), 13.26* (CH_3). *Rotamers.

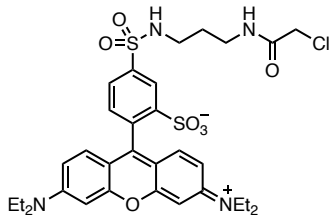
IR (cm^{-1}, thin film): ν 3676, 2982, 2901, 2357, 1708, 1523, 1420, 1347, 1143, 854, 785, 738.
HR–MS (ESI): calc. for $C_{24}H_{28}N_4NaO_{10}$ $[M+Na]^+$: 555.1698, obs.: 555.1692.

2.4. 4-(((diethylcarbamoyl)oxy)((1-(2-nitrophenyl)ethoxy)carbonyl)amino)butanoic acid (S6)



To a solution of commercially available 4-((*tert*-butoxycarbonyl)-((diethylcarbamoyl)oxy)amino)butanoic acid (318 mg, 1.00 mmol, 1.0 equiv) in dry CH_2Cl_2 (5.0 mL) was added TFA (2.5 mL). The solution was stirred at rt for 1 h and evaporated under reduced pressure to give a crude oil. To a solution of the residue in THF (5.0 mL) at 0 °C was added sat aq. $NaHCO_3$ (5.0 mL) and 1-(2-nitrophenyl)ethyl carbonochloridate[8] (343 mg, 1.5 mmol, 1.5 equiv). The mixture was stirred overnight and allowed to warm to rt. The solution was acidified by the addition of aq. 1 N HCl and extracted with EtOAc (3 x 10 mL). The combined organic phase was dried over $MgSO_4$, filtered, concentrated under reduced pressure and purified by flash column chromatography (CH_2Cl_2, 1 to 10% MeOH) to give **S6**[9] (279 mg, 0.58 mmol, 58% over two steps) as a yellow oil.
^{1}H NMR (500 MHz, $CDCl_3$): δ 7.99 (ddd, *J* = 8.2, 1.2, 0.5 Hz, 1H), 7.69 – 7.60 (m, 2H), 7.44 (ddd, *J* = 8.2, 6.8, 2.0 Hz, 1H), 6.33 (q, *J* = 6.4 Hz, 1H), 3.75 (m, 2H), 3.37 (m, 4H), 2.50 (t, *J* = 7.3 Hz, 2H), 1.95 (m, *J* = 7.0 Hz, 2H), 1.66 (d, *J* = 6.5 Hz, 3H), 1.23 (m, 6H).
^{13}C NMR (126 MHz, $CDCl_3$): δ 176.80, 154.65, 153.95, 147.36, 138.09, 133.75, 128.33, 127.06, 124.59, 70.69, 49.53, 43.21, 41.77, 30.63, 22.36, 22.29, 14.12, 13.33.
HR–MS (ESI) calc. for $C_{18}H_{25}N_3NaO_8$ $[M+Na]^+$: 434.1534, obs.: 434.1532

2.5. 5-(*N*-(3-(2-chloroacetamido)propyl)sulfamoyl)-2-(6-(diethylamino)-3-(diethyliminio)-3*H*-xanthen-9-yl)benzenesulfonate (sulforhodamine chloroacetamide 46)



To a stirred solution of 1,3-diaminopropane (4.71 mL, 56.4 mmol, 30 equiv) in dry CH_2Cl_2 (50 mL) at 0 °C was added dropwise a solution of sulforhodamine B sulfonyl chloride[10] (1.09 g, 1.88 mmol, 1 equiv) in dry CH_2Cl_2 (100 mL) over 1 h. The mixture was stirred at rt for 2 h, diluted with H_2O (150 mL) and brine (50 mL), and extracted with CH_2Cl_2 (4 x 150 mL). The combined organic phase was concentrated under

8 Baldauf, S.; Ogunkoya, A. O.; Boross, G. N.; Bode, J. W. Aspartic acid forming α-ketoacid–hydroxylamine (KAHA) ligations with (S)-4,4-difluoro-5-oxaproline. *J. Org. Chem.* **2019**, *85*, 1352–1364.

9 Kindly provided by D. Schauenburg (Bode Group, ETH Zürich).

10 Yang, H.; Vasudevan, S.; Oriakhi, C. O.; Shields, J.; Carter, R. G. Scalable synthesis of lissamine rhodamine B sulfonyl chloride and incorporation of xanthene derivatives onto polymer supports. *Synthesis* **2008**, *6*, 957–961.

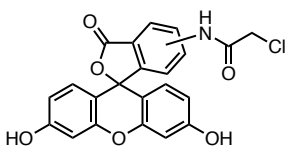
reduced pressure to yield a purple solid (quant.). The residue (1.15 g, 1.88 mmol, 1 equiv) was dissolved in dry DMF (100 mL). To the stirred solution at 0 °C was added dry NEt_3 (1.05 mL, 7.52 mmol, 4 equiv), and then dropwise 2-chloroacetyl chloride (300 μL, 3.76 mmol, 2 equiv) over 5 min. The mixture was stirred at rt for 30 min, concentrated under reduced pressure and purified by flash column chromatography (CH_2Cl_2, 10% MeOH) to give **46** (504 mg, 0.73 mmol, 39% over two steps) as a red solid.

^{1}H NMR (500 MHz, DMSO-d_6): δ 8.41 (d, *J* = 1.9 Hz, 1H), 8.25 (s, 1H), 7.93 (q, *J* = 5.5, 4.7 Hz, 2H), 7.48 (d, *J* = 7.9 Hz, 1H), 7.07 – 7.01 (m, 2H), 6.98 (d, *J* = 9.5 Hz, 2H), 6.94 (d, *J* = 2.4 Hz, 2H), 4.03 (s, 2H), 3.65 (m, *J* = 7.6, 6.8 Hz, 8H), 3.11 (q, *J* = 6.5 Hz, 2H), 2.89 (q, *J* = 6.7 Hz, 2H), 1.66 – 1.58 (m, 2H), 1.21 (t, *J* = 7.2 Hz, 12H).

^{13}C NMR (126 MHz, DMSO-d_6): δ 165.94, 157.45, 157.12, 155.03, 148.02, 141.41, 133.05, 132.72, 130.68, 126.55, 125.70, 113.66, 113.48, 95.39, 45.26, 42.63, 40.47, 36.57, 29.03, 12.47.

HR–MS (ESI) calc. for $C_{32}H_{40}ClN_4O_7S_2$ $[M+H]^+$: 691.2021, obs.: 691.2013.

2.6. 2-chloro-*N*-(3',6'-dihydroxy-3-oxo-3*H*-spiro[isobenzofuran-1,9'-xanthen]-5(6)-yl)acetamide (5(6)-(chloroacetamido)fluorescein 47, mixed isomers)



To a suspension of 5(6)-aminofluorescein[11,12] (100 mg, 0.29 mmol, 1 equiv) in dry acetone (10 mL) at 0 °C was added 2-chloroacetyl chloride (90 μL, 1.15 mmol, 4 equiv). The mixture was stirred at rt for 3.5 h, concentrated under reduced pressure and purified by flash column chromatography (CH_2Cl_2, 10 to 20% MeOH) to give **47** (63 mg, 0.15 mmol, 52%, 1:2 isomeric mixture of 5- and 6-(chloroacetamido)fluorescein) as an orange solid.

^{1}H NMR (400 MHz, MeOD-d_4): δ 8.34 (d, *J* = 2.0 Hz, 0.3H)*, 7.94 (d, *J* = 8.4 Hz, 0.7H)*, 7.87 (dd, *J* = 8.3, 2.0 Hz, 0.3H)*, 7.81 (dd, *J* = 8.4, 1.8 Hz, 0.7H)*, 7.56 (d, *J* = 1.7 Hz, 0.7H)*, 7.17 (d, *J* = 8.3 Hz, 0.3H)*, 6.67 (t, *J* = 2.0 Hz, 2H), 6.62 (d, *J* = 8.7 Hz, 2H), 6.54 (dd, *J* = 8.7, 2.4 Hz, 2H), 4.25 (s, 0.6H)*, 4.14 (s, 1.4H)*. *Isomers.

^{13}C NMR (101 MHz, MeOD-d_4): δ 171.12, 167.80 9, 161.27, 154.22, 154.00, 146.10, 141.22, 130.21, 130.12, 128.21, 126.79, 125.96, 123.44, 122.32, 116.50, 115.27, 113.75, 113.60, 111.45, 111.28, 103.53, 103.32, 43.98.

IR (cm^{-1}, thin film): ν 2988.6, 2360, 2343, 1685, 1604, 1453, 1393, 1262, 1113, 849.

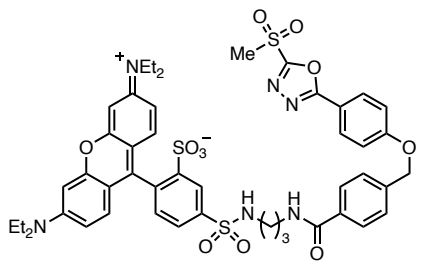
HR–MS (ESI) calc. for $C_{22}H_{15}ClNO_6$ $[M+H]^+$: 424.0582, obs.: 424.0586.

---

11 McKinney, R. M.; Spillane, J. T.; Pearce, G. W. Amino- and nitrofluorescein derivatives. *J. Org. Chem.* **1962**, *27*, 3986–3988.

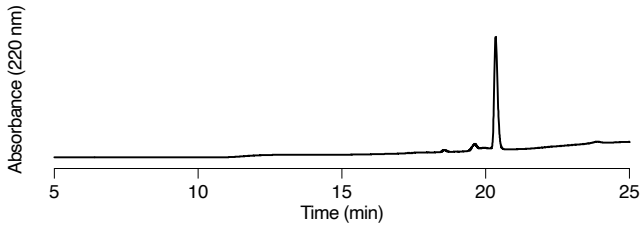
12 Tomasch, M.; Schwed, J. S.; Kuczka, K.; Santos, S. M. dos; Harder, S.; Nüsing, R. M.; Paulke, A.; Stark, H. Fluorescent human EP 3 receptor antagonists. *ACS Med. Chem. Lett.* **2012**, *3*, 774–779.

2.7. Sulforhodamine B methylsulfone oxadiazole reagent 48

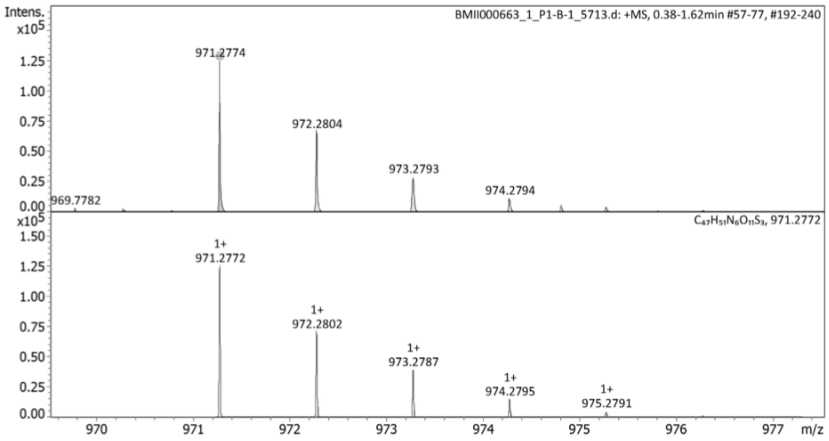


To a solution of 4-((4-(5-(methylsulfonyl)-1,3,4-oxadiazol-2-yl)phenoxy)methyl)benzoic acid[13] (43.5 mg, 116 μmol, 1.0 equiv) in dry DMF (2 mL) was added NHS (14.7 mg, 128 μmol, 1.1 equiv) and *N,N'*-diisopropylcarbodiimide (DIC) (19.8 μL, 128 μmol, 1.1 equiv). The mixture was stirred at rt for 4 h. Separately, to a stirred solution of 1,3-diaminopropane (868 μL, 10.4 mmol, 30.0 equiv) in dry CH_2Cl_2 (10 mL) at 0 °C was added dropwise a solution of sulforhodamine B sulfonyl chloride[10] (200 mg, 347 μmol, 1.0 equiv) in dry CH_2Cl_2 (10 mL) over 45 min. The rhodamine mixture was stirred at rt for 2 h, diluted with H_2O (40 mL) and brine (10 mL), and extracted with CH_2Cl_2 (4 x 100 mL). The combined organic phase was dried over $MgSO_4$ and concentrated under reduced pressure to yield a purple solid (168 mg). A portion of this residue (94.0 mg,151 μmol, 1.3 equiv) was added to the solution containing the *in situ* prepared NHS ester, and the mixture was stirred at rt for 18 h. The reaction mixture of the coupling product was quenched with aq. 1 N HCl (4 mL) and H_2O (4 mL) and extracted with CH_2Cl_2 (3 x 50 mL). The combined organic phase was concentrated under reduced pressure. The crude product was dissolved in aq. 30% MeCN with 0.1% TFA and purified by preparative RP–HPLC using a gradient of 5 to 50% solvent B over 10 min and 50 to 95% solvent B over 20 min. Fractions containing pure product were pooled and lyophilized to obtain **48** as a red solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).

[13] Kindly provided by Dr. C. White (Bode Group, ETH Zürich). Unpublished results. See footnote 22 for a related reference.



Analytical RP–HPLC of sulforhodamine B methylsulfone oxadiazole reagent 48.



HR–MS (ESI) of sulforhodamine B methylsulfone oxadiazole reagent 48. Obs. 971.2774 (top), calc. for $C_{47}H_{51}N_6O_{11}S_3$ $[M+H]^+$: 971.2772 (bottom).

3. Peptide synthesis

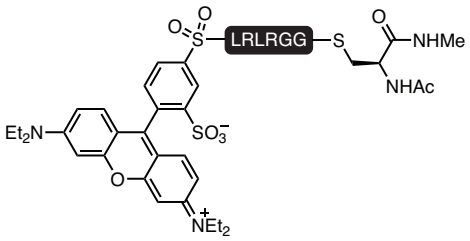
3.1. Peptide thioesters

3.1.1. Preparation of Fmoc-Gly-NHNH-2-chlrotrityl resin (S7)

For peptide thioester syntheses, 2-chlorotrityl chloride resin was loaded with hydrazine as previously described.[14] Briefly, 2-chlorotrityl chloride resin was swollen in CH_2Cl_2/DMF (1:1) for 30 min and treated with 5% hydrazine monohydrate in CH_2Cl_2/DMF (1:1) for 45 min. The resin was washed with CH_2Cl_2/DMF (1:1), capped with 5% MeOH and 10% NMM in DMF for 10 min, and washed again with CH_2Cl_2/DMF (1:1). Fmoc-Gly-OH (0.4 mmol/g resin) was immediately coupled onto the prepared 2-chlorotrityl hydrazine resin using HATU (0.4 mmol/g resin) and NMM (0.8 mmol/g resin) in DMF (4 mL/g resin) for 2.5 h. The resin was washed with DMF and unreacted hydrazine was capped with Ac_2O according to the general procedure.

3.1.2. Rhodamine thioesters 2a–c, 3–18 and 22

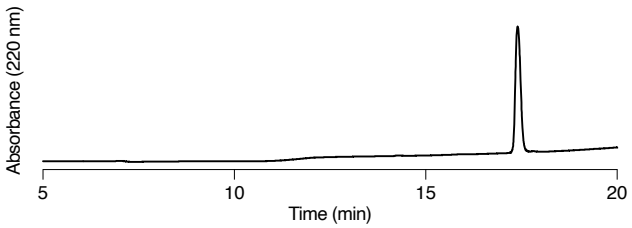
Rhodamine-LRLRGG-(Ac-Cys-NHMe) (5)



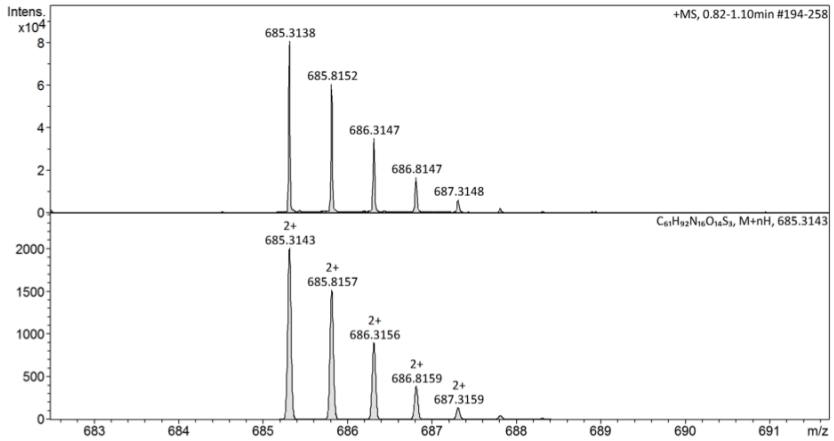
Rhodamine-LRLRGG-$NHNH_2$ (S8): Onto Fmoc-Gly-NHNH-2-chlorotrityl resin (**S7**) (500 mg resin, 0.163 mmol/g loading, 82 μmol), peptide **5** was elongated by manual SPPS according to the general procedure using Fmoc amino acids (336 μmol, 4.1 equiv), HATU (125 mg, 328 μmol, 4.0 equiv) and NMM (72 μL, 656 μmol, 8.0 equiv) in DMF (1.5 mL). N-terminal sulforhodamine B was conjugated to the resin using sulforhodamine B sulfonyl chloride[10] (188 mg, 328 μmol, 4.0 equiv) and NMM (108 μL, 984 μmol, 12.0 equiv) in DMF (5 mL) for 2 h. The resin was washed extensively with DMF after reaction with the dye. Following SPPS, global deprotection and cleavage from resin was carried out as described in the general procedure. After lyophilization, the crude peptide hydrazide **S8** was obtained as a red solid (135 mg, >95%) and stored at –20 °C.
Rhodamine-LRLRGG-(Ac-Cys-NHMe) (5): Peptide thioester formation was done following a reported procedure.[14] Briefly, a portion of the crude **S8** (7.4 mg, 6 μmol, 1 equiv) was dissolved in oxidation buffer (500 μL, 0.2 M sodium phosphates pH 3.0, 6 M GndHCl) and cooled to –15 °C

[14] Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **2013**, *8*, 2483–2495.

in an ice/NaCl bath. The solution was treated with $NaNO_2$ (120 μL of a fresh 0.5 M solution in water, 60 μmol, 10 equiv) at –15 °C for 20 min. Subsequently, a solution of Ac-Cys-NHMe (**1**) (31.7 mg, 180 μmol, 30 equiv) in thiolysis buffer (720 μL, 0.2 M sodium phosphates pH 7.0, 6 M GndHCl) was added to the reaction mixture. The pH was adjusted to pH 6.9–7.0 using aq. 1 N NaOH (approximately 100 μL) and the reaction was incubated at rt for 2 h. The reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min. Fractions containing pure product were pooled and lyophilized to obtain **5** (approximately 2.5 mg, 1.8 μmol, 30% yield) as a red solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below). Thioesters were stored as a solid at –20 °C for months without noticeable hydrolysis.

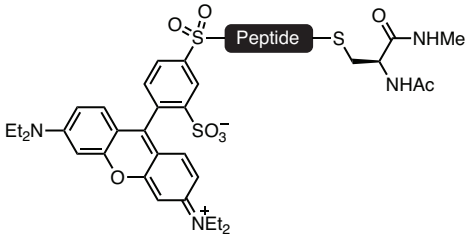


Analytical RP–HPLC of rhodamine-LRLRGG-(Ac-Cys-NHMe) (5).



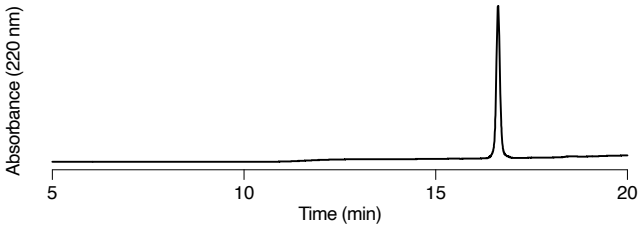
HR–MS (ESI) of rhodamine-LRLRGG-(Ac-Cys-NHMe) (5). Obs. 685.3138 (top), calc. for $C_{61}H_{94}N_{16}O_{14}S_3$ $[M+2H]^{2+}$: 685.3143 (bottom).

Rhodamine-peptide-(Ac-Cys-NHMe) thioester variants 2a, 3, 4, 6–11, 14–18, 22

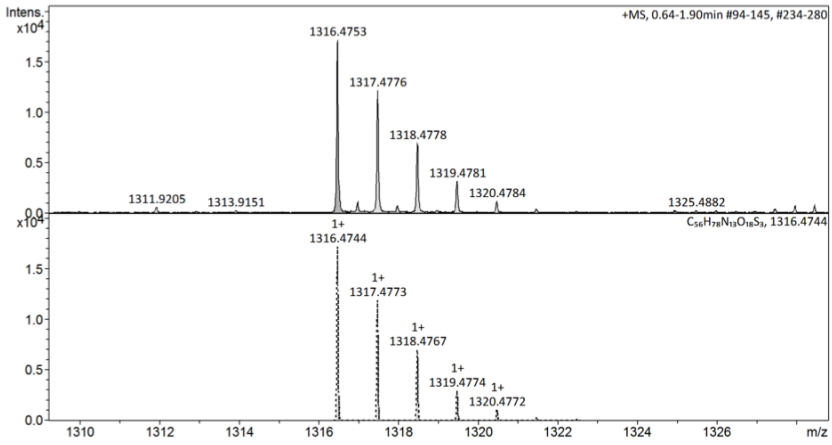


Thioester variant	peptide sequence	Thioester variant	peptide sequence
2a	QQQRGG	**11**	QQQRGG
3	QEQTGG	**14**	YQWSSGG
4	QQPTGG	**15**	YEEHIGG
6	RLRGG	**16**	YSFVSGG
7	LRGG	**17**	YGYVSGG
8	RGG	**18**	VYRFYGG
9	GG	**22**	RLRLGG
10	LRLTGG		

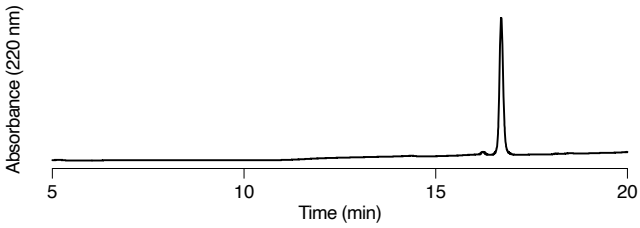
Rhodamine thioesters **2a, 3**, **4**, **6–11**, **14–18** and **22** were synthesized analogously to **5** with the respective peptide sequence in place of LRLRGG. Following conversion to the thioester, the reaction mixtures were directly purified by preparative RP–HPLC. Analytical RP–HPLC and HR–MS are shown below, together with gradients used for purification by preparative RP–HPLC.



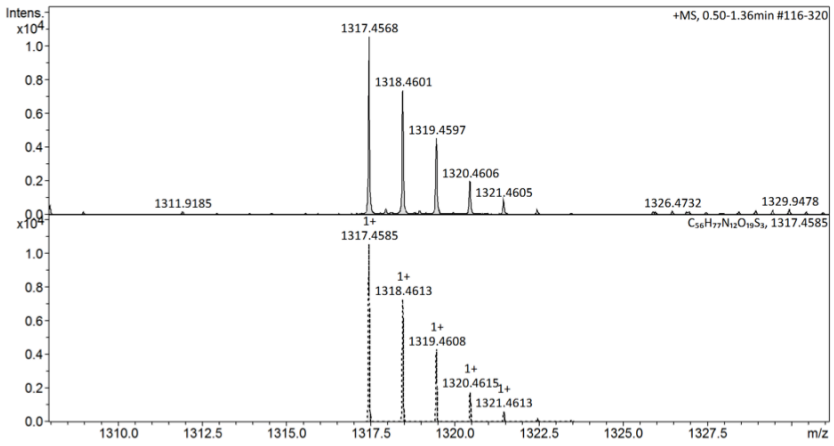
Analytical RP–HPLC of rhodamine-QQQTGG-(Ac-Cys-NHMe) (2a). The product was purified by prep. RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 60% solvent B over 30 min.



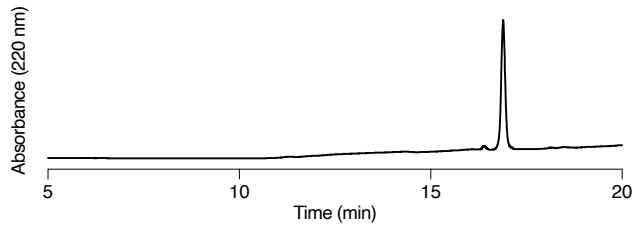
HR–MS (ESI) of rhodamine-QQQTGG-(Ac-Cys-NHMe) (2a). Obs. 1316.4753 (top), calc. for $C_{56}H_{78}N_{13}O_{18}S_3$ $[M+H]^+$: 1316.4744 (bottom).



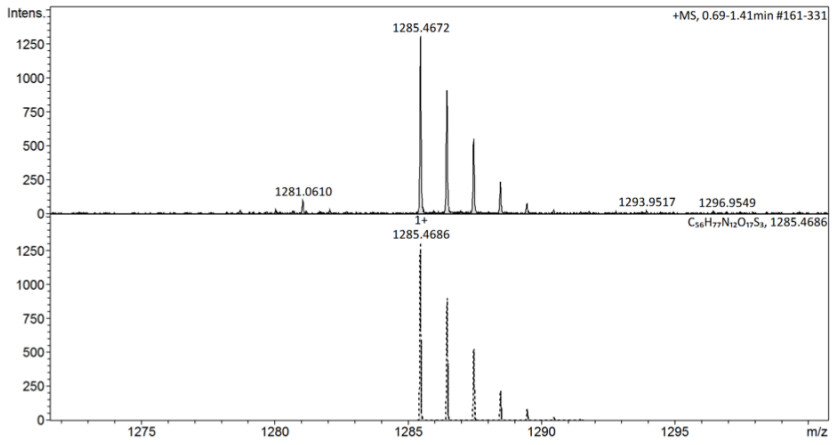
Analytical RP–HPLC of rhodamine-QEQTGG-(Ac-Cys-NHMe) (3). The product was purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 60% solvent B over 30 min.



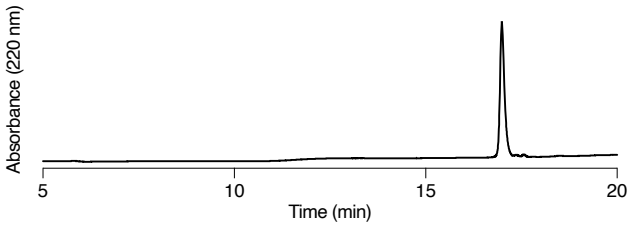
HR–MS (ESI) of rhodamine-QEQTGG-(Ac-Cys-NHMe) (3). Obs. 1317.4568 (top), calc. for $C_{56}H_{77}N_{12}O_{19}S_3$ $[M+H]^+$: 1317.4585 (bottom).



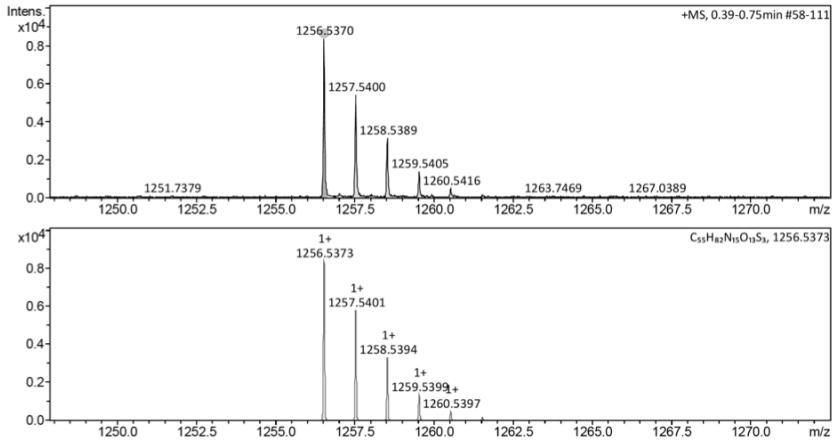
Analytical RP–HPLC of rhodamine-QQPTGG-(Ac-Cys-NHMe) (4). The product was purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 60% solvent B over 30 min.



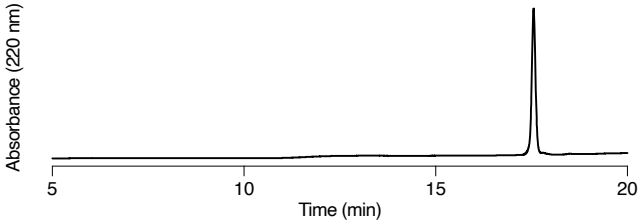
HR–MS (ESI) of rhodamine-QQPTGG-(Ac-Cys-NHMe) (4). Obs. 1285.4672 (top), calc. for $C_{56}H_{77}N_{12}O_{17}S_3$ $[M+H]^+$: 1285.4686 (bottom).



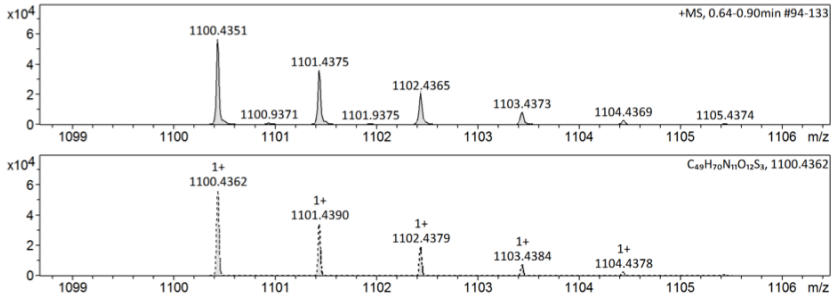
Analytical RP–HPLC of rhodamine-RLRGG-(Ac-Cys-NHMe) (6). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min.



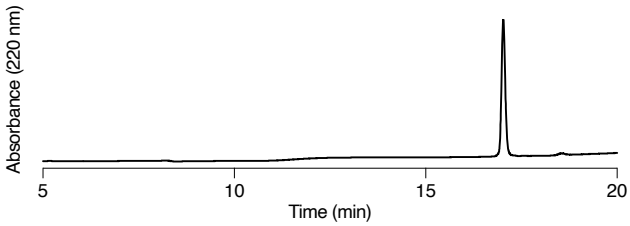
HR–MS (ESI) of rhodamine-RLRGG-(Ac-Cys-NHMe) (6). Obs. 1256.5370 (top), calc. for $C_{55}H_{82}N_{15}O_{13}S_3$ $[M+H]^+$: 1256.5373 (bottom).



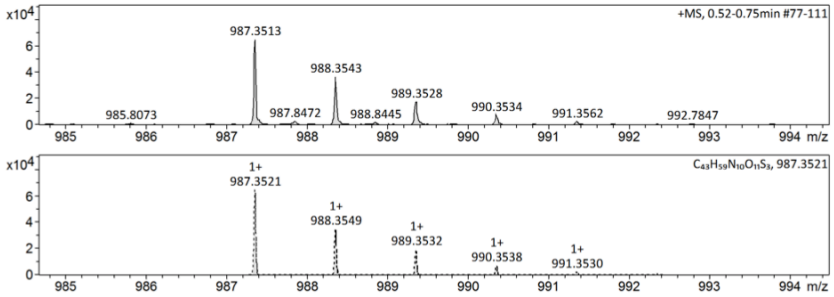
Analytical RP–HPLC of rhodamine-LRGG-(Ac-Cys-NHMe) (7). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min.



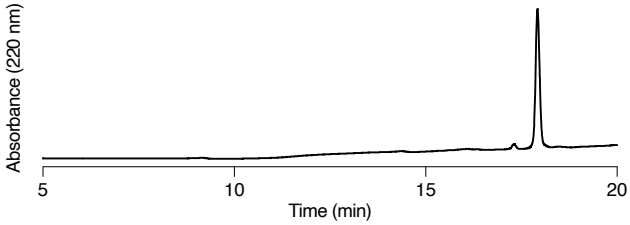
HR–MS (ESI) of rhodamine-LRGG-(Ac-Cys-NHMe) (7). Obs. 1100.4351 (top), calc. for $C_{49}H_{70}N_{11}O_{12}S_3$ $[M+H]^+$: 1100.4362 (bottom).



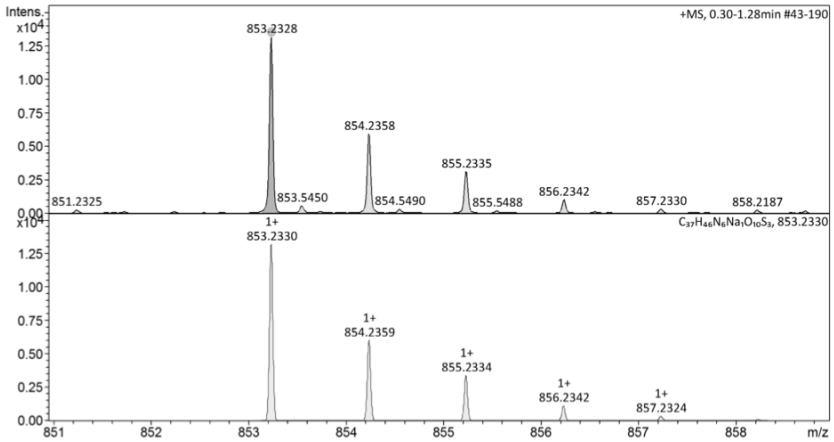
Analytical RP–HPLC of rhodamine-RGG-(Ac-Cys-NHMe) (8). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min.



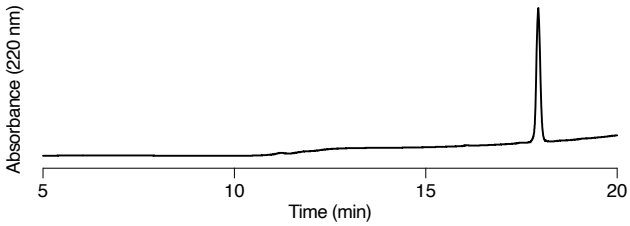
HR–MS (ESI) of rhodamine-RGG-(Ac-Cys-NHMe) (8). Obs. 987.3513 (top), calc. for $C_{43}H_{59}N_{10}O_{11}S_3$ $[M+H]^+$: 987.3521 (bottom).



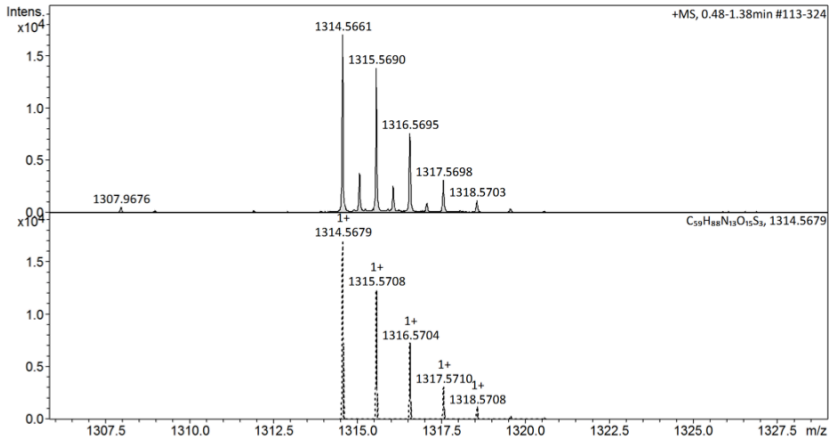
Analytical RP–HPLC of rhodamine-GG-(Ac-Cys-NHMe) (9). The product was purified by preparative RP–HPLC using a gradient of 5 to 60% solvent B over 20 min and 60 to 95% solvent B over 10 min.



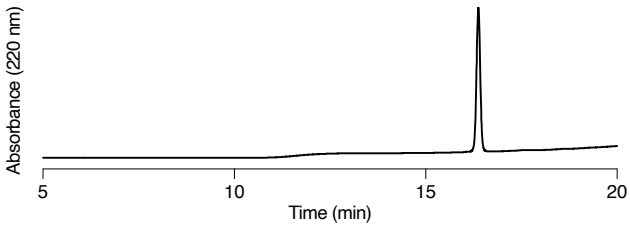
HR–MS (ESI) of rhodamine-GG-(Ac-Cys-NHMe) (9). Obs. 853.2328 (top), calc. for $C_{37}H_{46}N_6NaO_{10}S_3$ $[M+Na]^+$: 853.2330 (bottom).



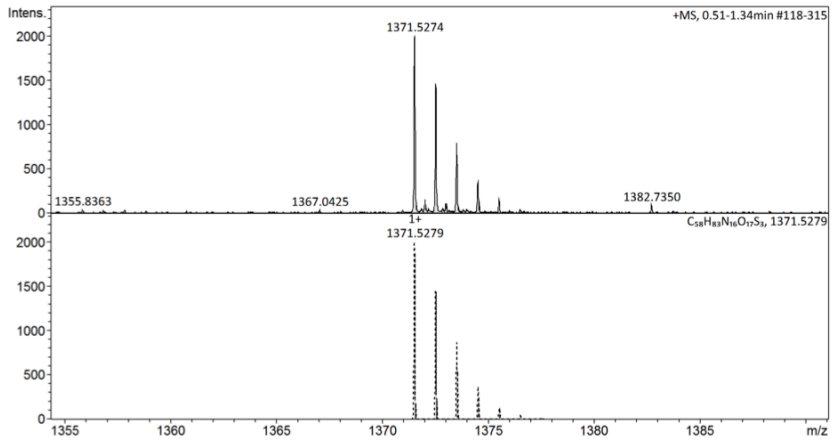
Analytical RP–HPLC of rhodamine-LRLTGG-(Ac-Cys-NHMe) (10). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min.



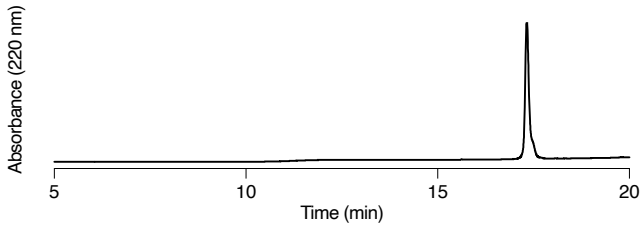
HR–MS (ESI) of rhodamine-LRLTGG-(Ac-Cys-NHMe) (10). Obs. 1314.5661 (top), calc. for $C_{59}H_{88}N_{13}O_{15}S_3$ $[M+H]^+$: 1314.5679 (bottom).



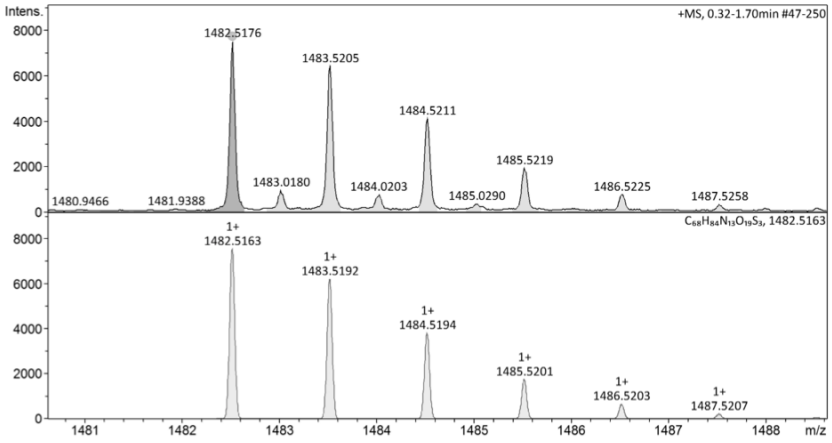
Analytical RP–HPLC of rhodamine-QQQRGG-(Ac-Cys-NHMe) (11). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min.



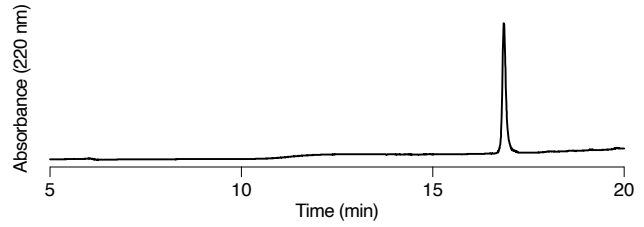
HR–MS (ESI) of rhodamine-QQQRGG-(Ac-Cys-NHMe) (11). Obs. 1371.5274 (top), calc. for $C_{58}H_{83}N_{16}O_{17}S_3$ $[M+H]^+$: 1371.5279 (bottom).



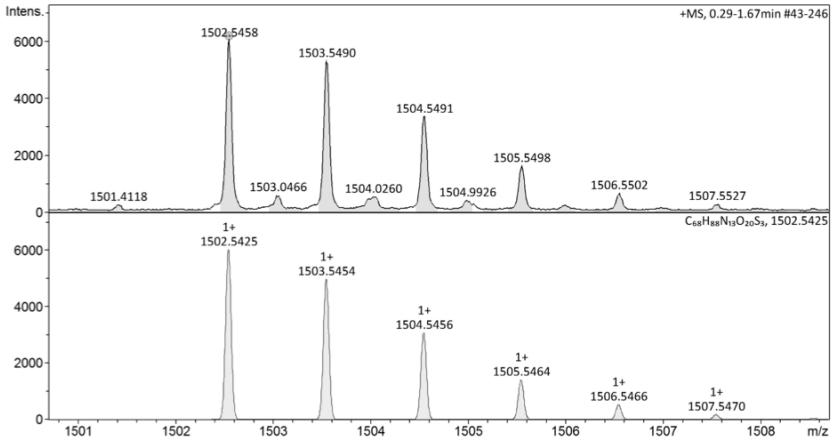
Analytical RP–HPLC of rhodamine-YQWSSGG-(Ac-Cys-NHMe) (14). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 70% solvent B over 30 min.



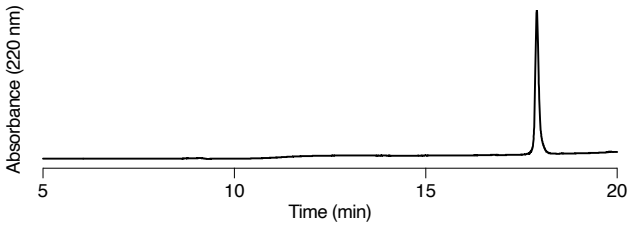
HR–MS (ESI) of rhodamine-YQWSSGG-(Ac-Cys-NHMe) (14). Obs. 1482.5176 (top), calc. for $C_{68}H_{84}N_{13}O_{19}S_3$ $[M+H]^+$: 1482.5163 (bottom).



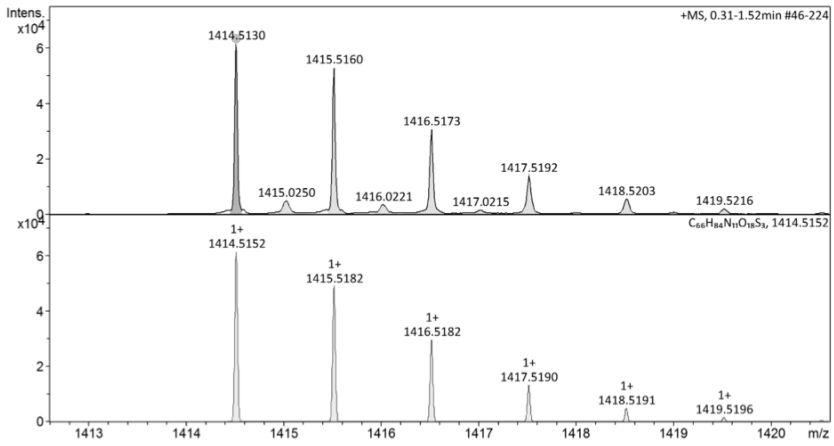
Analytical RP–HPLC of rhodamine-YEEHIGG-(Ac-Cys-NHMe) (15). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 70% solvent B over 30 min.



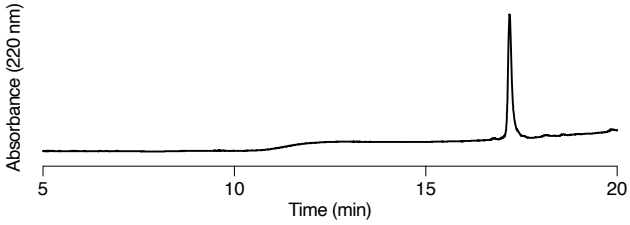
HR–MS (ESI) of rhodamine-YEEHIGG-(Ac-Cys-NHMe) (15). Obs. 1502.5458 (top), calc. for $C_{68}H_{88}N_{13}O_{20}S_3$ $[M+H]^+$: 1502.5425 (bottom).



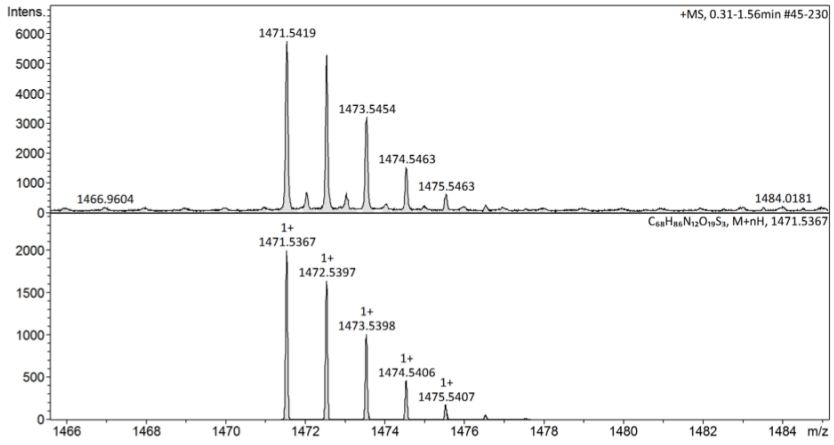
Analytical RP–HPLC of rhodamine-YSFVSGG-(Ac-Cys-NHMe) (16). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min.



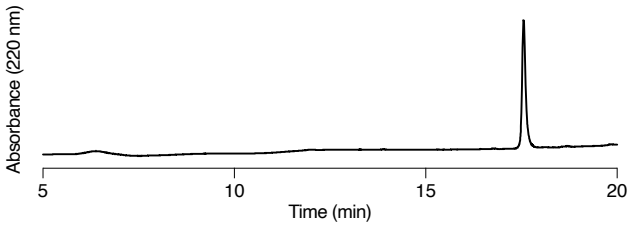
HR–MS (ESI) of rhodamine-YSFVSGG-(Ac-Cys-NHMe) (16). Obs. 1414.5130 (top), calc. for $C_{66}H_{84}N_{11}O_{18}S_3$ $[M+H]^+$: 1414.5152 (bottom).



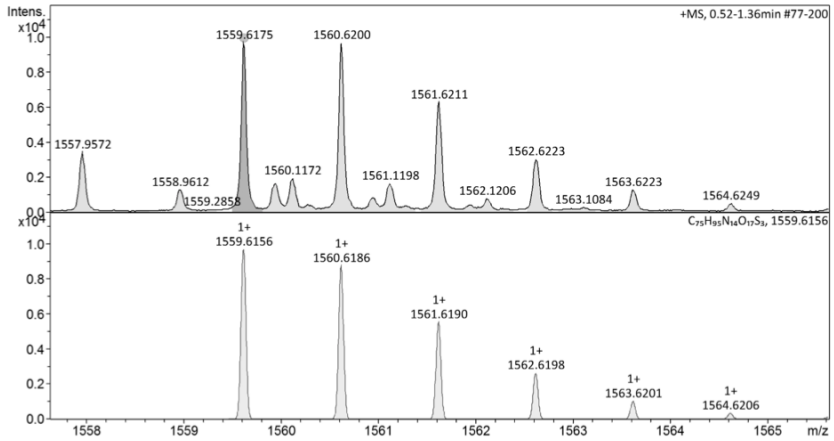
Analytical RP–HPLC of rhodamine-YGYVSGG-(Ac-Cys-NHMe) (17). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min.



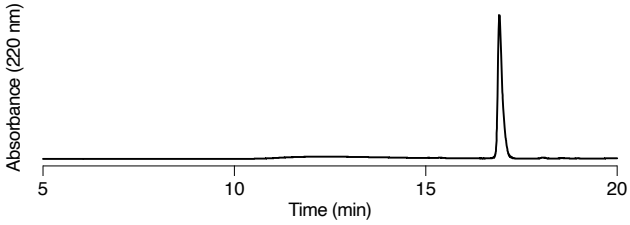
HR–MS (ESI) of rhodamine-YGYVSGG-(Ac-Cys-NHMe) (17). Obs. 1471.5419 (top), calc. for $C_{68}H_{87}N_{12}O_{19}S_3$ $[M+H]^+$: 1471.5367 (bottom).



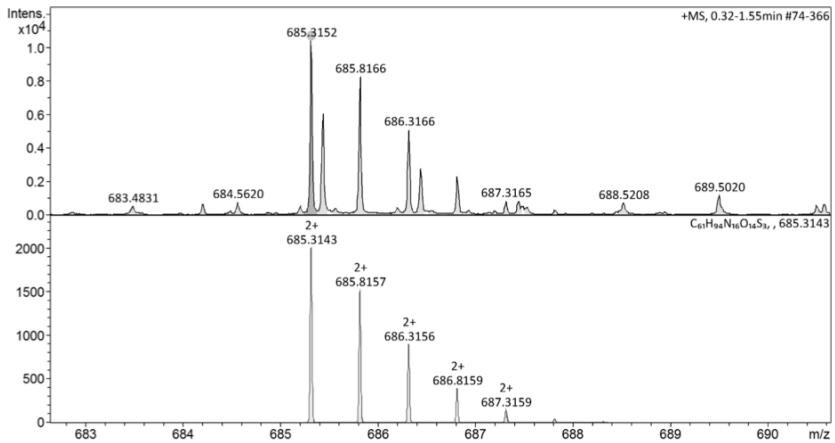
Analytical RP–HPLC of rhodamine-VYRFYGG-(Ac-Cys-NHMe) (18). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min.



HR–MS (ESI) of rhodamine-VYRFYGG-(Ac-Cys-NHMe) (18). Obs. 1559.6175 (top), calc. for $C_{75}H_{95}N_{14}O_{17}S_3$ $[M+H]^+$: 1559.6156 (bottom).

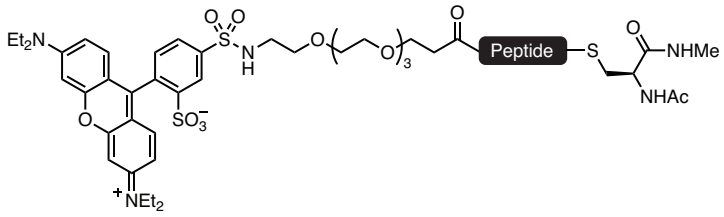


Analytical RP–HPLC of rhodamine-RLRLGG-(Ac-Cys-NHMe) (22).



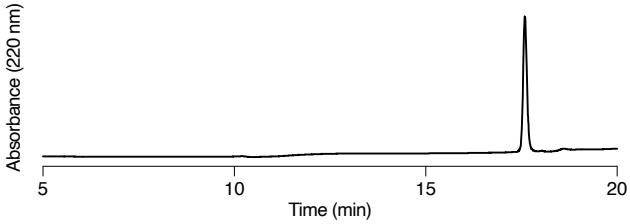
HR–MS (ESI) of rhodamine-RLRLGG-(Ac-Cys-NHMe) (22). Obs. 685.3152 (top), calc. for $C_{61}H_{94}N_{16}O_{14}S_3$ $[M+H]^+$: 685.3143 (bottom).

Rhodamine-PEG$_3$-peptide-(Ac-Cys-NHMe) thioester variants 12 and 13

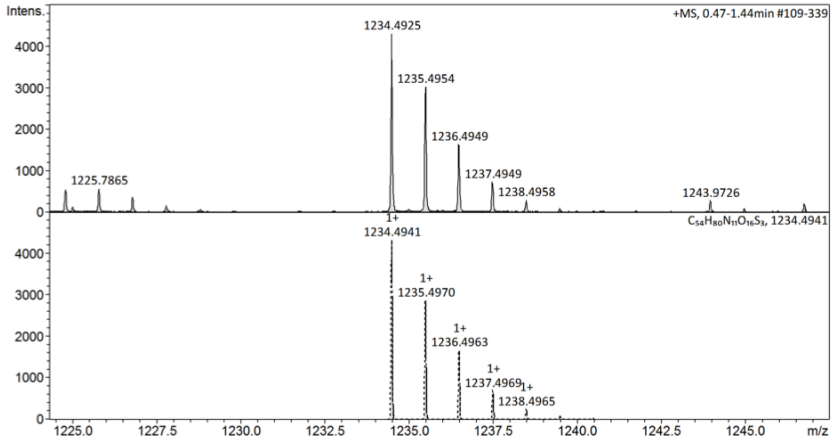


Thioester variant	peptide sequence
12	RGG
13	GG

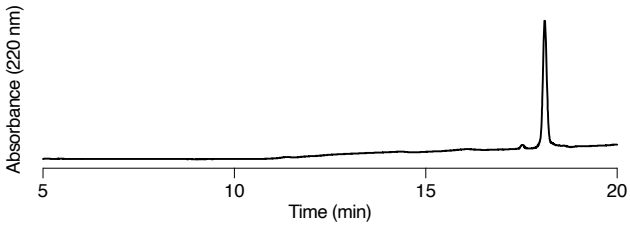
Rhodamine thioesters **12** and **13** were synthesized analogously to **5** with the respective peptide sequence in place of LRLRGG, and a PEG-linker was introduced by coupling Fmoc-NH-(PEG)$_3$-COOH (2.5 equiv) with HATU (2.45 equiv) and NMM (5 equiv) in DMF once to the resin. Following conversion to the thioester, the reaction mixtures were directly purified by preparative RP–HPLC. Analytical RP–HPLC and HR–MS are shown below, together with gradients used for purification by preparative RP–HPLC.



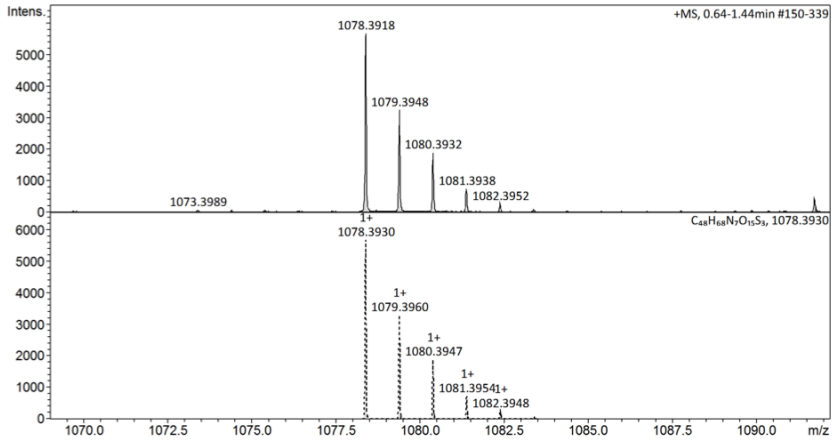
Analytical RP–HPLC of rhodamine-PEG$_3$-RGG-(Ac-Cys-NHMe) (12). The product was purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 75% solvent B over 20 min.



HR–MS (ESI) of rhodamine-PEG$_3$-RGG-(Ac-Cys-NHMe) (12). Obs. 1234.4925 (top), calc. for $C_{54}H_{80}N_{11}O_{16}S_3$ $[M+H]^+$: 1234.4941 (bottom).

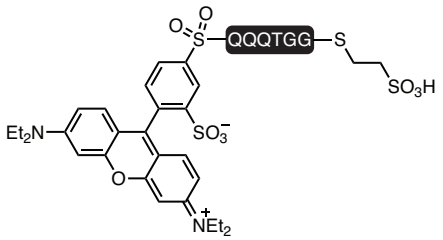


Analytical RP–HPLC of rhodamine-PEG$_3$-GG-(Ac-Cys-NHMe) (13). The product was purified by preparative RP–HPLC using a gradient of 5 to 60% solvent B over 20 min and 60 to 95% solvent B over 10 min.

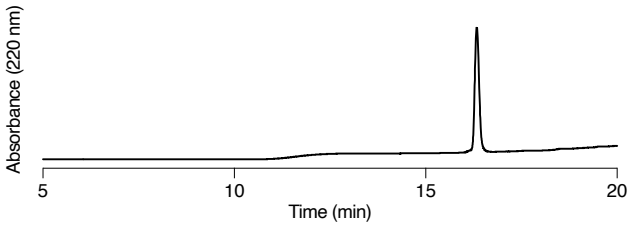


HR–MS (ESI) of rhodamine-PEG$_3$-GG-(Ac-Cys-NHMe) (13). Obs. 1078.3918 (top), calc. for $C_{48}H_{68}N_7O_{15}S_3$ $[M+H]^+$: 1078.3930 (bottom).

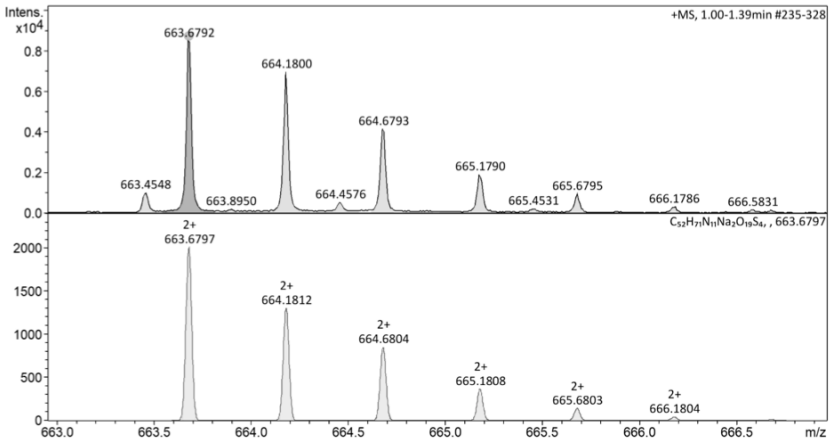
Rhodamine-QQQTGG-MesH (2b)



Rhodamine thioesters **2b** was synthesized analogously to **2a**, with the following deviation: Thioester formation was done by treating crude rhodamine-QQQTGG-NHNH$_2$ (1.7 mg, 1.29 μmol, 1 equiv) in oxidation buffer (200 μL, 0.2 M sodium phosphates pH 3.0, 6 M GndHCl) at –15 °C (ice/NaCl bath) with $NaNO_2$ (25.8 μL of a fresh 0.5 M solution in water, 12.9 μmol, 10 equiv). After 20 min, a solution of sodium 2-mercaptoethanesulfonate (MesNa) (10.6 mg, 64.5 μmol, 50 equiv) in thiolysis buffer (250 μL, 0.2 M sodium phosphates pH 7.0, 6 M GndHCl) was added to the mixture. The pH was adjusted to pH 6.9–7.0 using aq. 1 N NaOH (approximately 30 μL) and the reaction was incubated at rt for 1.5 h. The reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 75% solvent B over 20 min. Fractions containing pure product were pooled and lyophilized to obtain **2b** as a red solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).

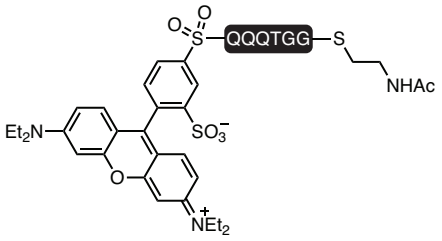


Analytical RP–HPLC of rhodamine-QQQTGG-MesH (2b).

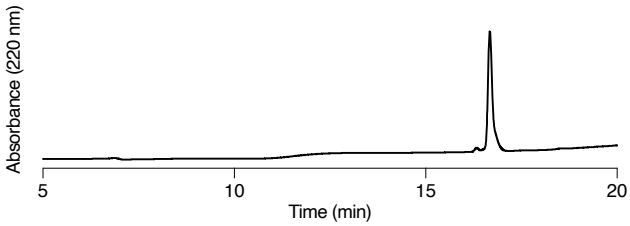


HR–MS (ESI) of rhodamine-QQQTGG-MesH (2b). Obs. 663.6792 (top), calc. for $C_{52}H_{71}N_{11}Na_2O_{19}S_4$ $[M+2Na]^{2+}$: 663.6797 (bottom).

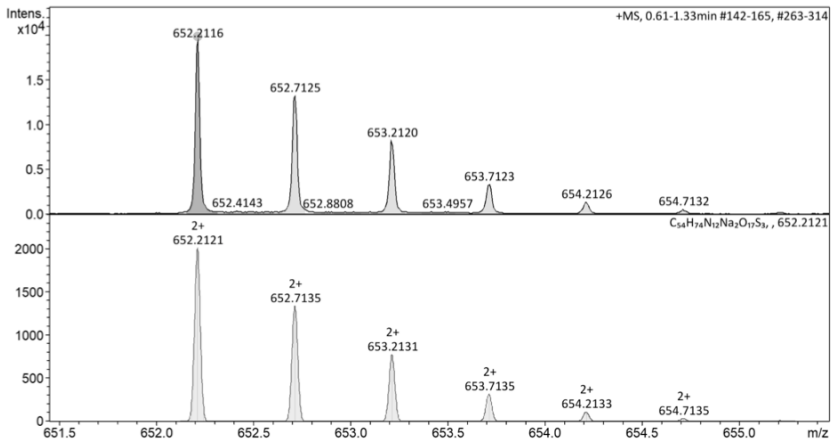
Rhodamine-QQQTGG-(*N*-acetylcysteamine) (2c)



Rhodamine thioesters **2c** was synthesized analogously to **2a**, with the following deviation: Thioester formation was done by treating crude rhodamine-QQQTGG-$NHNH_2$ (1.6 mg, 1.22 μmol, 1 equiv) in oxidation buffer (200 μL, 0.2 M sodium phosphates pH 3.0, 6 M GndHCl) at –15 °C (ice/NaCl bath) with $NaNO_2$ (24.4 μL of a fresh 0.5 M solution in water, 12.2 μmol, 10 equiv). After 20 min, a solution of *N*-acetylcysteamine (6.5 μL, 61.0 μmol, 50 equiv) in thiolysis buffer (250 μL, 0.2 M sodium phosphates pH 7.0, 6 M GndHCl) was added to the mixture. The pH was adjusted to pH 6.9–7.0 using aq. 1 N NaOH (approximately 30 μL) and the reaction was incubated at rt for 1.5 h. The reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 75% solvent B over 20 min. Fractions containing pure product were pooled and lyophilized to obtain **2c** as a red solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).



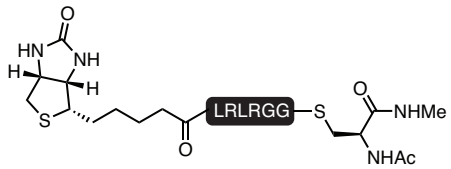
Analytical RP–HPLC of rhodamine-QQQTGG-(*N*-acetylcysteamine) (2c).



HR–MS (ESI) of rhodamine-QQQTGG-(*N*-acetylcysteamine) (2c). Obs. 652.2116 (top), calc. for $C_{54}H_{74}N_{12}Na_2O_{17}S_3$ $[M+2Na]^{2+}$: 652.2121 (bottom).

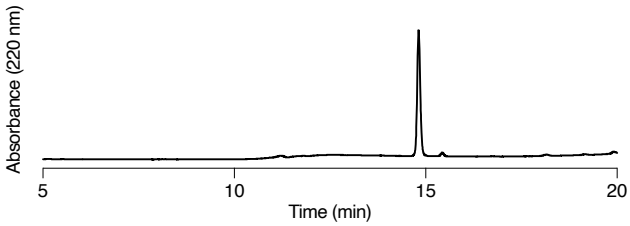
3.1.3. Biotin thioesters 24 and 25

Biotin-LRLRGG-(Ac-Cys-NHMe) (24)

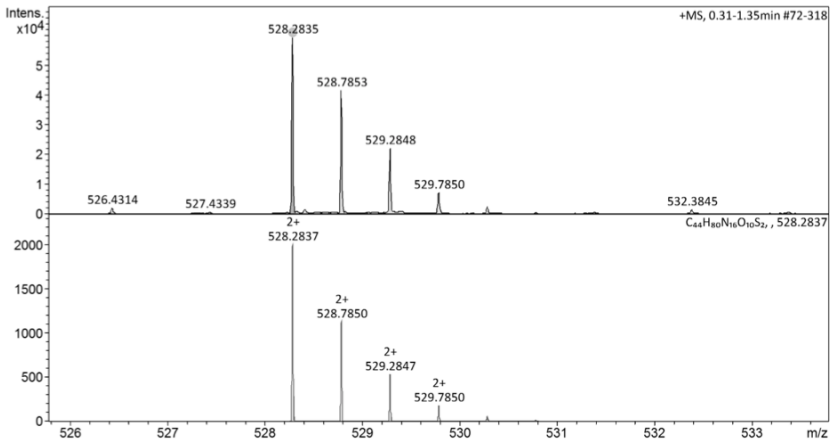


Biotin-LRLRGG-NHNH$_2$ (S9): Biotin-LRLRGG-NHNH$_2$ was synthesized analogously to rhodamine-LRLRGG-NHNH$_2$ (**S8**), with the following deviation: N-terminal biotin was coupled to the resin in place of sulforhodamine B according to the general coupling procedure for Fmoc amino acids.

Biotin-LRLRGG-(Ac-Cys-NHMe) (24): Conversion of **S9** to biotin thioester **24** was done analogously to **5**. Following conversion to the thioester, the reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 50% solvent B over 30 min and lyophilized to give **24** as a white solid. Analytical RP–HPLC and HR–MS are shown below.

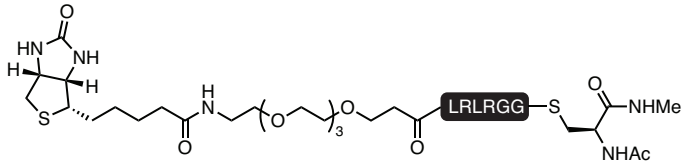


Analytical RP–HPLC of biotin-LRLRGG-(Ac-Cys-NHMe) (24).

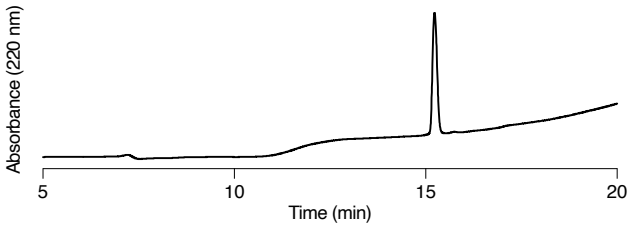


HR–MS (ESI) of biotin-LRLRGG-(Ac-Cys-NHMe) (24). Obs. 528.2835 (top), calc. for $C_{44}H_{80}N_{16}O_{10}S_2$ $[M+2H]^{2+}$: 528.2837 (bottom).

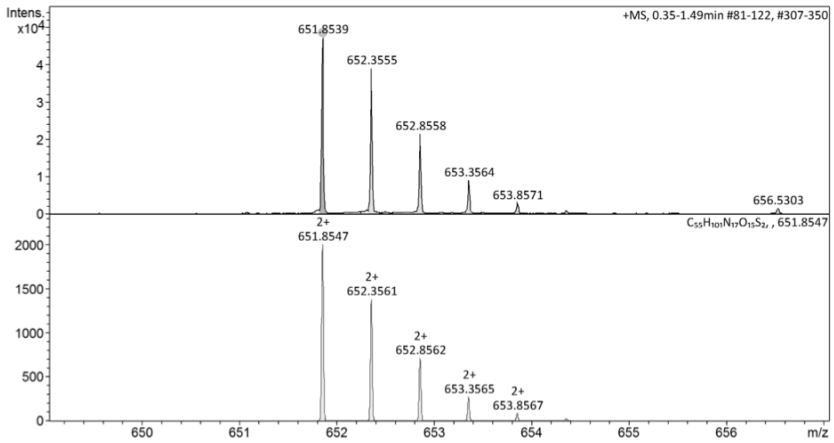
Biotin-PEG$_3$-LRLRGG-(Ac-Cys-NHMe) (25)



Biotin-PEG$_3$-LRLRGG-(Ac-Cys-NHMe) was synthesized analogously to **12** and **24**. Following conversion to the thioester, the reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 50% solvent B over 30 min. Analytical RP–HPLC and HR–MS are shown below.

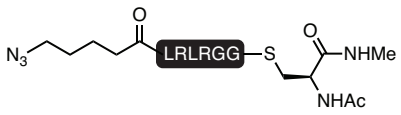


Analytical RP–HPLC of biotin-PEG$_3$-LRLRGG-(Ac-Cys-NHMe) (25).



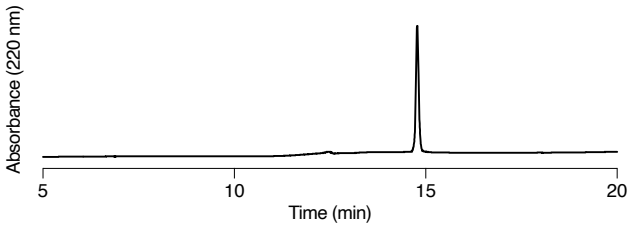
HR–MS (ESI) of biotin-PEG$_3$-LRLRGG-(Ac-Cys-NHMe) (25). Obs. 651.8539 (top), calc. for $C_{55}H_{101}N_{17}O_{15}S_2$ $[M+2H]^{2+}$: 651.8547 (bottom).

3.1.4. Azide-LRLRGG-(Ac-Cys-NHMe) (28)



H-LRLRGG-(Ac-Cys-NHMe) (26)

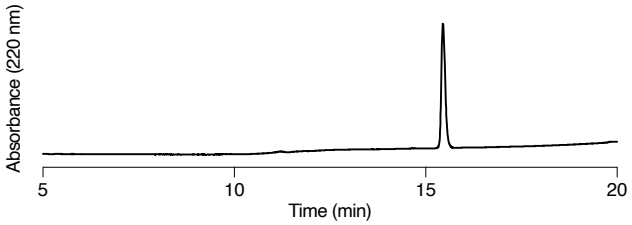
To a solution of H-LRLRGG-NHNH$_2$ (20.3 mg, 29.6 μmol, 1 equiv; prepared analogously to **S8** but omitting coupling with sulforhodamine B sulfonyl chloride) in oxidation buffer (2.30 mL, 0.2 M sodium phosphates pH 3.0, 6 M GndHCl) at –15 °C (ice/NaCl bath) was added $NaNO_2$ (593 μL of a fresh 0.5 M solution in water, 60 μmol, 10 equiv). After 20 min, a solution of Ac-Cys-NHMe (**1**) (156.5 mg, 889.3 μmol, 30 equiv) in thiolysis buffer (2.30 mL, 0.2 M sodium phosphates pH 7.0, 6 M GndHCl) was added to the reaction mixture. The pH was adjusted to pH 6.9–7.0 using aq. 5 N NaOH (approximately 80 μL) and the reaction was incubated at rt for 1 h 45 min. The reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 50% solvent B over 30 min. Fractions containing pure product were pooled and lyophilized to obtain **26** (approximately 10 mg, 12 μmol, 41% yield) as a white solid. The purity and identity were verified by analytical RP–HPLC (results shown below) and MALDI–MS.



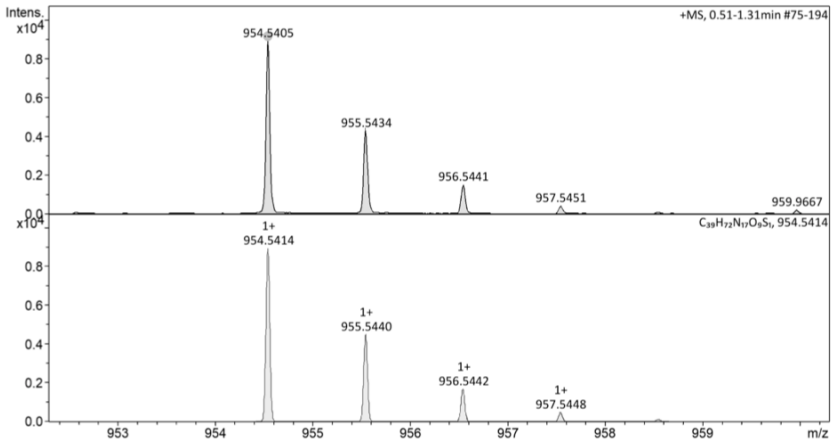
Analytical RP–HPLC of H-LRLRGG-(Ac-Cys-NHMe) (26).

5-azidopentanoyl-LRLRGG-(Ac-Cys-NHMe) (28)

To a solution of H-LRLRGG-(Ac-Cys-NHMe) (**26**) (10 mg, 9.5 μmol assuming a 2*TFA salt, 1 equiv) in 1:1 MeCN / 0.1 M potassium phosphates pH 6.5 (1.89 mL) was added a solution of **27** (5.68 mg, 23.6 μmol, 2.5 equiv) in DMSO (20 μL). The pH was readjusted to pH 6.5 with aq. 2 N NaOH (approximately 5 μL) and the mixture was agitated at 30 °C overnight. The reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 75% solvent B over 30 min. Fractions containing pure product were pooled and lyophilized to obtain **28** (6.3 mg, 6.6 μmol, 69% yield) as a white solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).



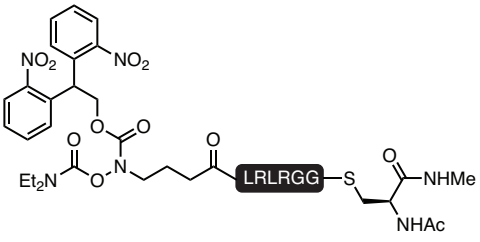
Analytical RP–HPLC of 5-azidopentanoyl-LRLRGG-(Ac-Cys-NHMe) (28).



HR–MS (ESI) of 5-azidopentanoyl-LRLRGG-(Ac-Cys-NHMe) (28). Obs. 954.5405 (top), calc. for $C_{39}H_{72}N_{17}O_9S$ $[M+H]^+$: 954.5414 (bottom).

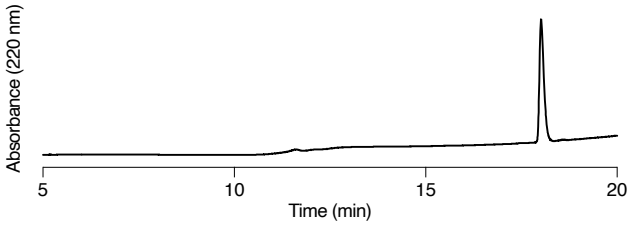
3.1.5. Hydroxylamine thioesters 31a and 31b

4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)-oxy)amino)butanoyl-LRLRGG-(Ac-Cys-NHMe) (31a)

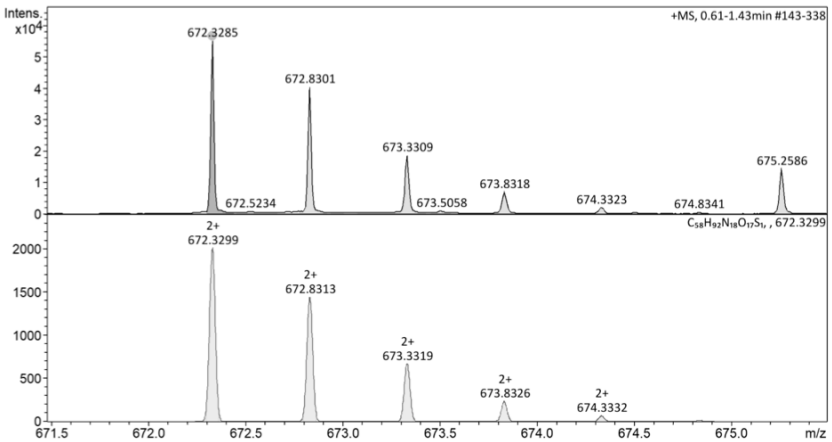


4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)-oxy)amino)butanoyl-LRLRGG-NHNH$_2$ (S10): S10 was synthesized analogously to rhodamine-LRLRGG-NHNH$_2$ (**S8**), with the following deviation: **S2** was coupled to the resin in place of sulforhodamine B. For the coupling, **S2** (36.0 mg, 67.5 μmol, 3.0 equiv) was preactivated with HATU (25.6 mg, 67.5 μmol, 3 equiv) and NMM (14.8 μL, 135 μmol, 6.0 equiv) in DMF (0.75 mL) for 3 min before adding the solution to preswollen H-LRLRGG-NHNH-2-chlorotrityl resin (50 mg resin, 0.45 mmol/g loading, 22.5 μmol). The mixture was agitated in the dark at rt for 2.5 h, before the resin was washed with DMF. Following SPPS, global deprotection and cleavage from resin was carried out as described in the general procedure. After lyophilization, the crude peptide hydrazide **S10** was obtained as a white solid (quant.) and stored in the dark at –20 °C.

4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)-oxy)amino)butanoyl-LRLRGG-Ac-Cys-NHMe (31a): Conversion of **S10** to thioester **31a** was done analogously to **5**. Following conversion to the thioester, the reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 75% solvent B over 30 min and lyophilized to give **31a** as a white solid, which was stored in the dark at –20 °C. Analytical RP–HPLC and HR–MS are shown below.

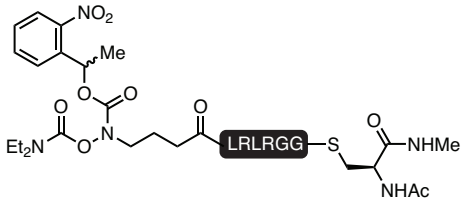


Analytical RP–HPLC of 4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)-oxy)amino)-butanoyl-LRLRGG-Ac-Cys-NHMe (31a).

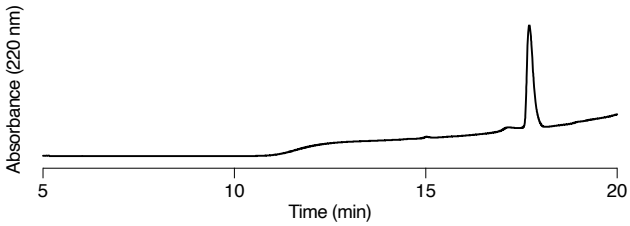


HR–MS (ESI) of 4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)-oxy)amino)butanoyl-LRLRGG-Ac-Cys-NHMe (31a). Obs. 672.3285 (top), calc. for $C_{58}H_{92}N_{18}O_{17}S$ $[M+2H]^{2+}$: 672.3299 (bottom).

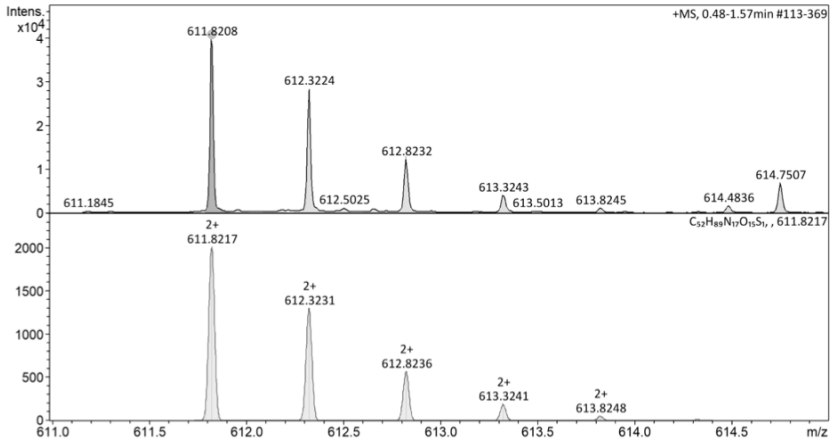
4-(((diethylcarbamoyl)oxy)((1-(2-nitrophenyl)ethoxy)carbonyl)amino)butanoyl-LRLRGG-(Ac-Cys-NHMe) (31b)



31b was synthesized analogously to **31a** by coupling **S6** in place of **S5**. Following conversion to the thioester, the reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min. Analytical RP–HPLC and HR–MS are shown below.

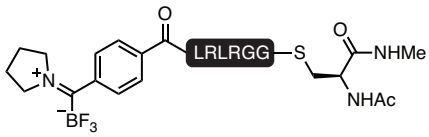


Analytical RP–HPLC of 4-(((diethylcarbamoyl)oxy)((1-(2-nitrophenyl)ethoxy)carbonyl)amino)-butanoyl-LRLRGG-(Ac-Cys-NHMe) (31b).

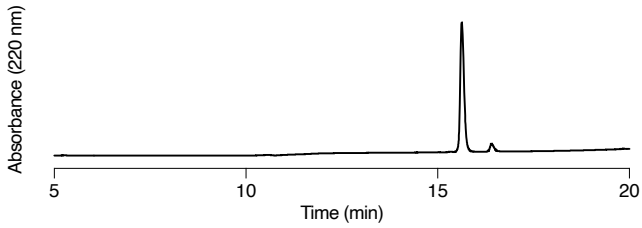


HR–MS (ESI) of 4-(((diethylcarbamoyl)oxy)((1-(2-nitrophenyl)ethoxy)carbonyl)amino)butanoyl-LRLRGG-(Ac-Cys-NHMe) (31b). Obs. 611.8208 (top), calc. for $C_{52}H_{89}N_{17}O_{15}S$ $[M+2H]^{2+}$: 611.8217 (bottom).

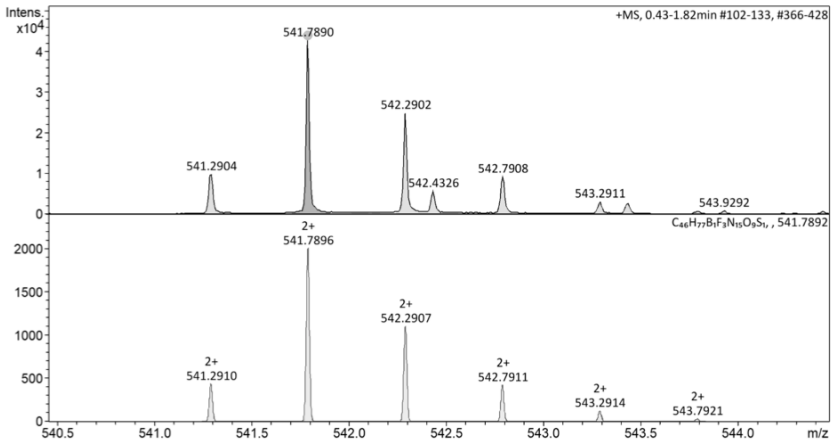
3.1.6. TIM thioester 39



H-LRLRGG-(Ac-Cys-NHMe) (**26**) (40.0 mg, 34.2 μmol assuming a 3*TFA salt, 1.0 equiv) and **38**[15] (31.1 mg, 68.4 μmol, 2.0 equiv) were dissolved in 9:1 DMF / *N*,*N*-diisopropylethylamine (DIPEA) (4 mL). The mixture was briefly vortexed and then agitated for 10 min. The reaction was quenched by the addition of 1:1 AcOH/H_2O (4 mL), diluted with aq. MeCN and directly purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 50% solvent B over 30 min. Fractions containing pure product were pooled and lyophilized to obtain **39** (16.8 mg, 15.7 μmol, 46% yield) as a white solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).



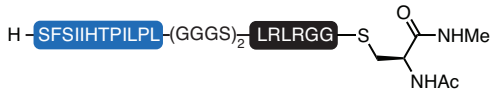
Analytical RP–HPLC of TIM thioester (39).



HR–MS (ESI) of TIM thioester (39). Obs. 541.7890 (top), calc. for $C_{46}H_{77}BF_3N_{15}O_9S$ $[M+2H]^{2+}$: 541.7896 (bottom).

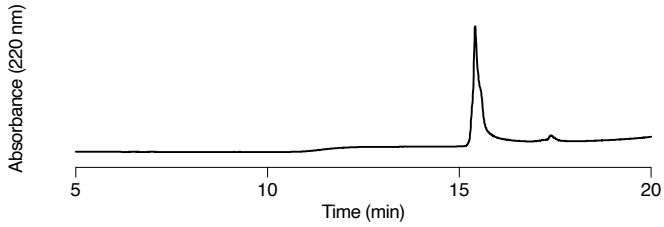
[15] Kindly provided by D. Schauenburg (Bode Group, ETH Zürich). Unpublished results.

3.1.7. SP94 peptide thioester 44



SP94-$(GGGS)_2$-LRLRGG-$NHNH_2$ (S11)

Onto Fmoc-Gly-NHNH-2-chlorotrityl resin (**S7**) (500 mg resin, 0.196 mmol/g loading, 0.1 mmol), peptide **S11** was elongated by automated SPPS (Multisyntech Syro I parallel synthesizer). The automated coupling cycle was analogous to the one described in the general procedure, with the difference that DIPEA was used instead of NMM. Following elongation of the sequence, the N-terminal Fmoc group was removed with 20% DMF (2 x 10 min), and the resin was washed with CH_2Cl_2 and dried under reduced pressure. Global deprotection and cleavage from resin was carried out as described in the general procedure. The crude product was dissolved in aq. 30% MeCN with 0.1% TFA and purified by preparative RP–HPLC on a Shiseido Capcell Pak C_{18} column (5 μm, 80 Å pore size, 50 x 250 mm) at 60 °C using a gradient of 10 to 60% solvent B over 35 min at a flow rate of 40 mL/min. Fractions containing mostly pure product were pooled, lyophilized and repurified by semi-preparative RP–HPLC on a Agilent Eclipse XDB C_8 column (7 μm, 21.2 x 250 mm) at 60 °C using a gradient of 10–70% solvent B over 35 min at a flow rate of 10 mL/min. Fractions containing pure product were pooled and lyophilized to obtain **S11** (25 mg, 10 μmol, 10% yield) as a white solid. The purity and identity were verified by analytical RP–HPLC on a Shiseido Capcell Pak C_{18} column (5 μm, 4.6 x 250 mm) (results shown below) and MALDI–MS.

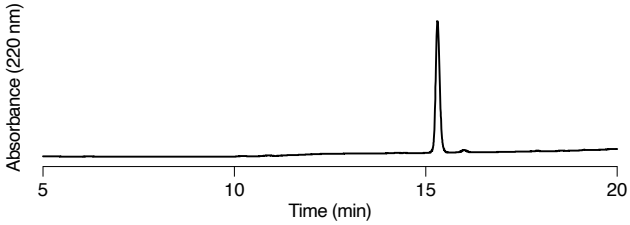


Analytical RP–HPLC of SP94-$(GGGS)_2$-LRLRGG-$NHNH_2$ (S11).

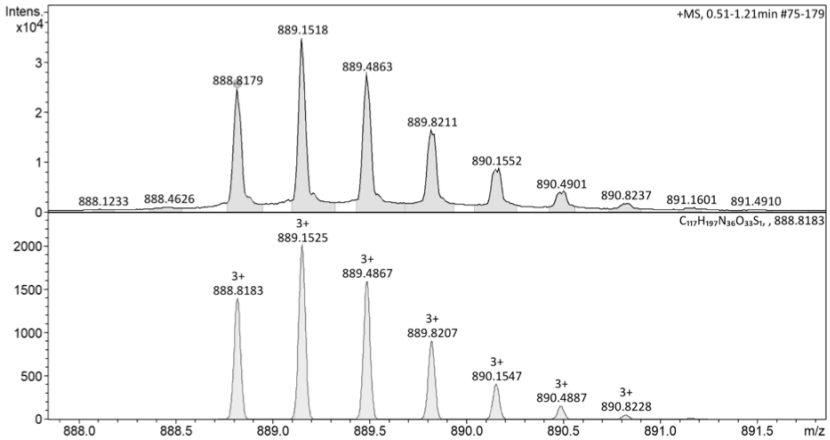
SP94-$(GGGS)_2$-LRLRGG-(Ac-Cys-NHMe) (44)

SP94-$(GGGS)_2$-LRLRGG-$NHNH_2$ (**S11**) (24.0 mg, 9.5 μmol, 1 equiv) was dissolved in oxidation buffer (800 μL, 0.2 M sodium phosphates pH 3.0, 6 M GndHCl) and cooled to –15 °C in an ice/NaCl bath. The solution was treated with $NaNO_2$ (190 μL of a fresh 0.5 M solution in water, 60 μmol, 10 equiv) at –15 °C for 20 min. Subsequently, a solution of Ac-Cys-NHMe (**1**) (58.6 mg, 333.2 μmol, 35 equiv) in thiolysis buffer (950 μL, 0.2 M sodium phosphates pH 7.0, 6 M GndHCl) was added to the reaction mixture. The pH was adjusted to pH 6.9–7.0 using aq. 5 N NaOH (approximately 35 μL) and the reaction was incubated at rt for 2 h. The reaction mixture was directly purified by preparative RP–HPLC on an Osaka-Soda Capcell Pack C_{18} column type MGIII

(5 μm, 80 Å pore size, 50 x 250 mm) using a gradient of 10 to 75% solvent B over 30 min. Fractions containing pure product were pooled and lyophilized to obtain **44** (15.0 mg, 5.6 μmol, 59% yield) as a white solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).

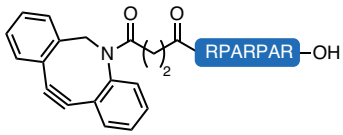


Analytical RP–HPLC of SP94-(GGGS)$_2$-LRLRGG-(Ac-Cys-NHMe) (44).



HR–MS (ESI) of SP94-(GGGS)$_2$-LRLRGG-(Ac-Cys-NHMe) (44). Obs. 888.8179 (top), calc. for $C_{117}H_{197}N_{36}O_{33}S$ $[M+3H]^{3+}$: 888.8183 (bottom).

3.2. DBCO-RPARPAR-OH (DBCO peptide 45)



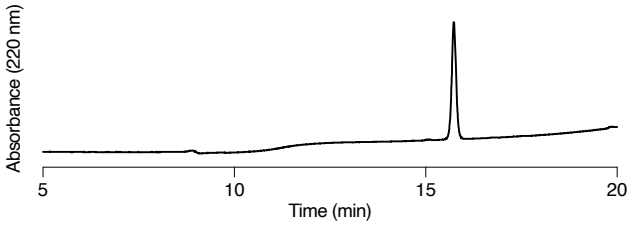
H-RPARPAR-OH (S12)

2-Chlorotrityl chloride resin (2 g resin, 1.6 mmol/g 2-chlorotrityl chloride loading) was swollen in CH_2Cl_2 for 30 min at rt. The solvent was drained off and to the resin was added a solution of Fmoc-Arg(Pbf)-OH (389 mg, 0.6 mmol, 0.3 mmol/g resin) and NMM (264 μL, 2.4 mmol, 4 equiv relative to the amino acid) in CH_2Cl_2 (5 mL) with a few drops of DMF. The mixture was agitated for 2 h. The resin was washed with CH_2Cl_2 and DMF and treated with 10% NMM / 5% MeOH in DMF (3 x 3 min with fresh solution each time). The resin was washed with DMF and CH_2Cl_2 and dried under reduced pressure. Onto a portion of the prepared resin (382 mg resin, 0.262 mmol/g loading, 0.1 mmol), peptide **S12** was elongated by automated SPPS (Multisyntech Syro I parallel synthesizer). The automated coupling cycle was analogous to the one described in the general

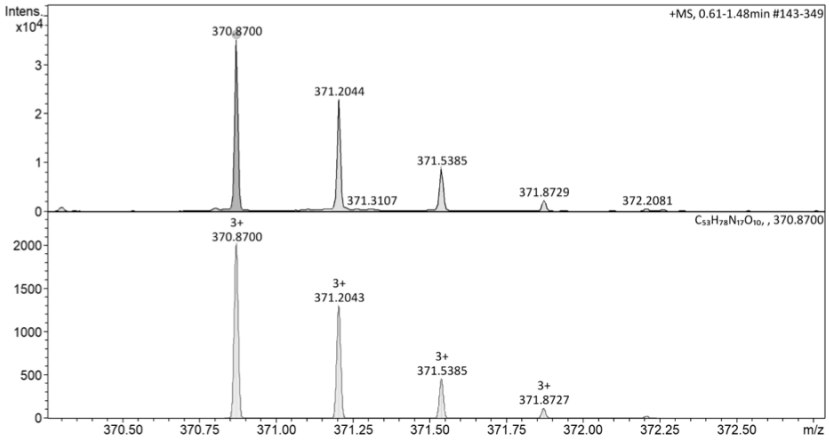
procedure, with the difference that DIPEA was used instead of NMM. Following elongation of the sequence, the N-terminal Fmoc group was removed with 20% DMF (2 x 10 min), and the resin was washed with CH_2Cl_2 and dried under reduced pressure. Global deprotection and cleavage from resin was carried out as described in the general procedure. After lyophilization, the crude peptide **S12** was obtained as a white solid (quant.), which was used without further purification.

DBCO-RPARPAR-OH (45)

To a solution of H-RPARPAR-OH (**S12**) (17.4 mg, 13.6 μmol assuming it was present as a 4*TFA salt, 1.0 equiv) in 1:1 MeCN / 0.1 M potassium phosphates pH 6.5 (2.72 mL) was added a solution of commercially available DBCO-NHS ester (13.7 mg, 34.0 μmol, 2.5 equiv) in DMSO (25 μL). The pH was readjusted to pH 6.5 with aq. 2 N NaOH (approximately 30 μL) and the mixture was agitated at 30 °C overnight. The reaction mixture was directly purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 60% solvent B over 30 min. Fractions containing pure product were pooled and lyophilized to obtain **45** (14.4 mg, 10.0 μmol, 73% yield) as a white solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).

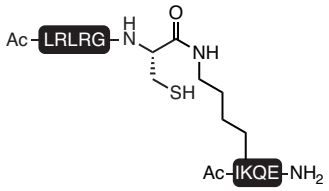


Analytical RP–HPLC of DBCO-RPARPAR-OH (45).

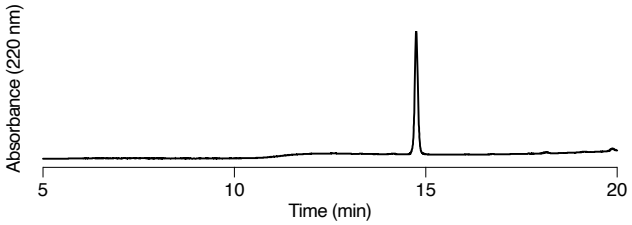


HR–MS (ESI) of DBCO-RPARPAR-OH (45). Obs. 370.8700 (top), calc. for $C_{53}H_{78}N_{17}O_{10}$ $[M+3H]^{3+}$: 370.8700 (bottom).

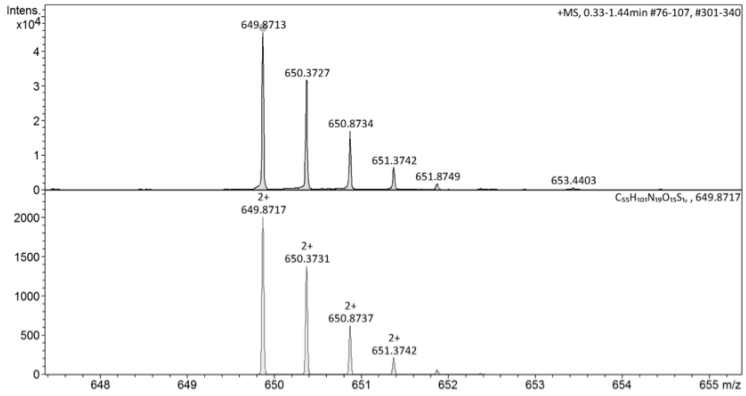
3.3. Ac-IK*iso*(Ac-LRLRGC)QE-NH_2 (isopeptide ligand 21)



According to the general SPPS procedure, Fmoc-Glu(OtBu)OH, Fmoc-Gln(Trt)-OH, Fmoc-Lys(Alloc)-OH and Fmoc-Ile-OH were coupled onto Fmoc-Rink-Amide resin (180 mg resin, 0.272 mmol/g loading, 50 μmol) and the peptide was N-terminally acetylated. Alloc deprotection on resin was done following the reported procedure.[16] Briefly, the resin was treated twice with $Pd(PPh_3)_4$ (11.5 mg, 10 μmol, 0.2 equiv) and $PhSiH_3$ (62.5 μL, 100 μmol, 10.0 equiv) in 1:1 CH_2Cl_2/DMF (1.5 mL) at rt for 45 min, with a CH_2Cl_2 wash (5x) in-between. The resin was washed with CH_2Cl_2 (3x), 0.5% w/v sodium diethyldithiocarbamate trihydrate in DMF (5x) and DMF (5x). Fmoc-Cys(Trt)-OH (117 mg, 200 μmol, 4.0 equiv) was coupled to the free lysine ε-amine with DIC (31.3 μL, 200 μmol, 4.0 equiv) and HOBt hydrate (27.0 mg, 200 μmol, 4.0 equiv) in DMF twice for 1.5 h. The isopeptide was then elongated, N-terminally acetylated and cleaved from resin according to the general procedure. The crude product was purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 50% solvent B over 30 min. Fractions containing pure product were pooled and lyophilized to obtain **21** (approximately 12.1 mg, 9.3 μmol, 19% yield) as a white solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).

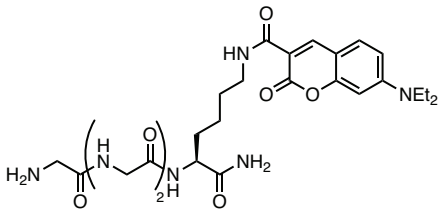


Analytical RP–HPLC of Ac-IK*iso*(Ac-LRLRGC)QE-NH_2 (21).

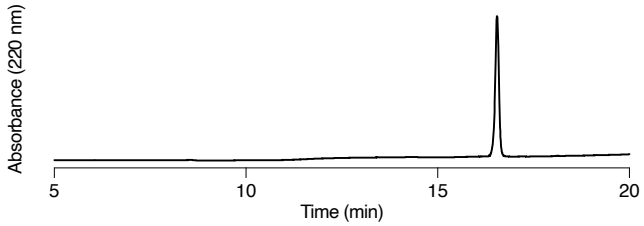


HR–MS (ESI) of Ac-IK*iso*(Ac-LRLRGC)QE-NH_2 (21). Obs. 649.8713 (top), calc. for $C_{55}H_{101}N_{19}O_{15}S$ $[M+2H]^{2+}$: 649.8717 (bottom).

3.4. H-GGGK(7-[diethylamino]-3-carboxycoumarin)-NH_2 (sortase-reactive probe 23)

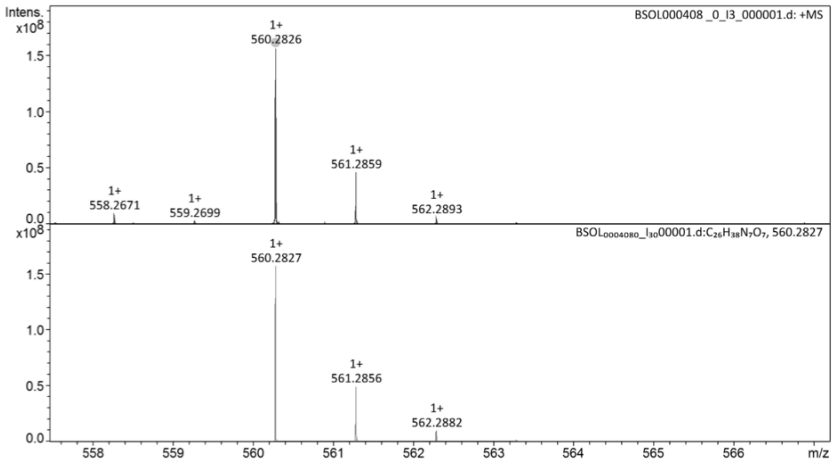


Peptide **23** was elongated on Fmoc-Rink-Amide resin (600 mg resin, 0.265 mmol/g loading, 0.159 mmol) by manual SPPS according to the general procedure. Fmoc-Lys(Alloc)-OH was coupled as first amino acid and the last Gly was coupled as Boc-Gly-OH. Alloc deprotection on resin was done following a reported procedure.[16] Briefly, the resin was treated twice with $Pd(PPh_3)_4$ (46 mg, 40 μmol, 0.25 equiv) and $PhSiH_3$ (250 μL, 2.03 mmol, 12.80 equiv) in 1:1 CH_2Cl_2/DMF (8 mL) at rt for 45 min, with a CH_2Cl_2 wash (5x) in-between. The resin was washed with CH_2Cl_2 (3x), 0.5% w/v sodium diethyldithiocarbamate trihydrate in DMF (5x) and DMF (5x). Onto the free lysine ε-amine was coupled 7-(diethylamino)coumarin-3-carboxylic acid (131 mg, 0.50 mmol, 3.15 equiv) with HATU (183 mg, 0.48 mmol, 3.00 equiv) and NMM (110 μL, 1.00 mmol, 6.30 equiv) in DMF (4 mL) for 2.5 h. The peptide was cleaved from resin according to the general procedure and purified by preparative RP–HPLC using a gradient of 5 to 15% solvent B over 5 min and 15 to 60% solvent B over 30 min. Fractions containing pure product were pooled and lyophilized to obtain **23** (approximately 64 mg, 0.114 μmol, 72% yield) as a yellow solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below).



Analytical RP–HPLC of H-GGGK(7-[diethylamino]-3-carboxycoumarin)-NH_2 (23).

[16] Zheng, J.-S.; Yu, M.; Qi, Y.-K.; Tang, S.; Shen, F.; Wang, Z.-P.; Xiao, L.; Zhang, L.; Tian, C.-L.; Liu, L. Expedient total synthesis of small to medium-sized membrane proteins via Fmoc chemistry. *J. Am. Chem. Soc.* **2014**, *136*, 3695–3704.

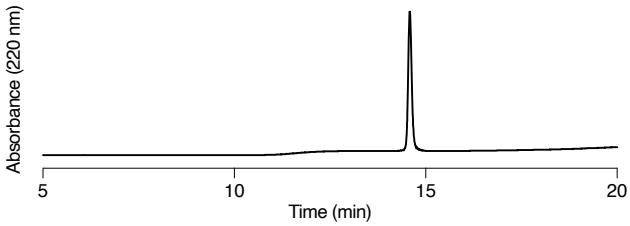


HR–MS (MALDI–FTICR–MS) of H-GGGK(7-[diethylamino]-3-carboxycoumarin)-NH_2 (23). Obs. 560.2826 (top), calc. for $C_{26}H_{38}N_7O_7$ $[M+H]^+$: 560.2827 (bottom).

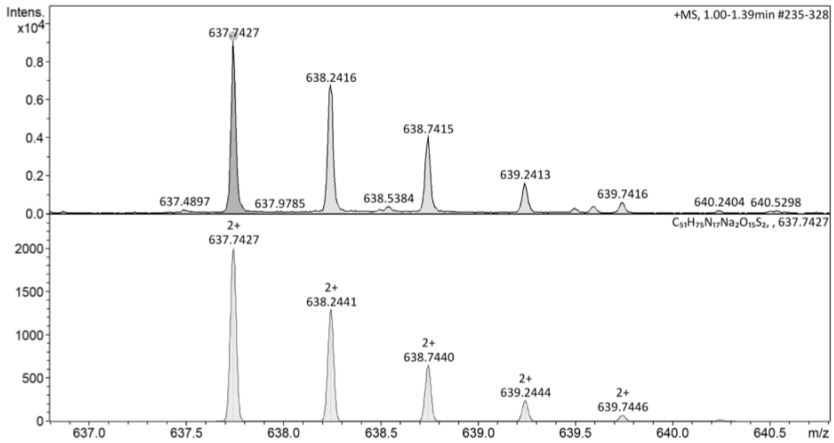
3.5. Disulfide tag library standards (peptides 48–50)

Peptide	Structure
48	Ac-VYC(Cam)C(Cam)GFGGGRG-NH_2
49	Ac-VAC(Cam)C(Cam)FDGGGRG-NH_2
50	Ac-VLC(Cam)C(Cam)NAGGGRG-NH_2

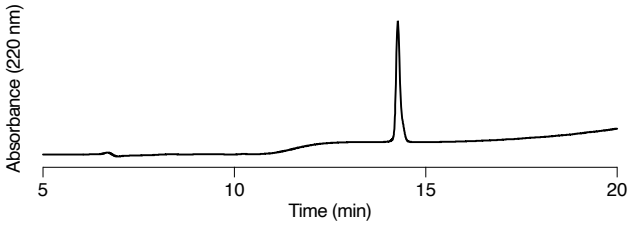
Peptides **48**–**50** were each elongated on Fmoc-Rink-Amide resin (50 mg resin, 0.35 mmol/g loading, 17.5 μmol) by automated SPPS (Multisyntech Syro I parallel synthesizer). The automated coupling cycle was analogous to the one described in the general procedure. Fmoc-Cys(Cam)-OH (250 μL of a 0.40 M stock solution in DMF, 5.7 equiv) was coupled manually to the resin at the respective positions by preactivating the amino acid with HATU (200 μL of a 0.45 M stock solution in DMF, 5.1 equiv) and 2,4,6-trimethylpyridine (100 μL of a 2 M stock solution in DMF, 11.4 equiv) for 3 min before addition of the mixture to the resin. The coupling was performed twice for 30 min at rt. Following elongation of the sequences, the N-terminal Fmoc group was removed and the N-terminus was capped as described in the general procedure. The peptides were cleaved from resin according to the general procedure and purified by preparative RP–HPLC using a gradient of 5 to 25% solvent B over 5 min and 25 to 60% solvent B over 30 min in the case of **48**, 5 to 50% solvent B over 10 min and 50 to 95% solvent B over 20 min in the case of **49**, and 5 to 25% solvent B over 5 min and 25 to 75% solvent B over 20 min in the case of **50**. Fractions containing pure product were pooled and lyophilized to obtain **48**–**50** (approximately 5 mg each, ~4 μmol, ~23% yield) as a white solid. The purity and identity were verified by analytical RP–HPLC and HR–MS (results shown below). MS/MS peptide sequencing results are shown in Figure 96.



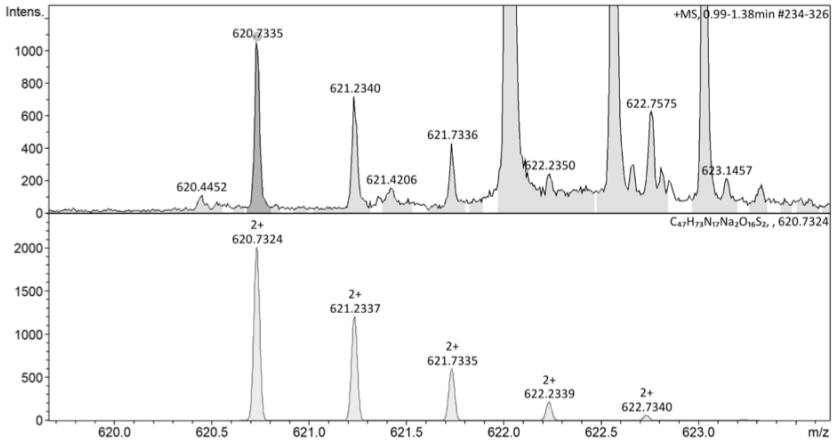
Analytical RP–HPLC of Ac-VYC(Cam)C(Cam)GFGGGRG-NH_2 (48).



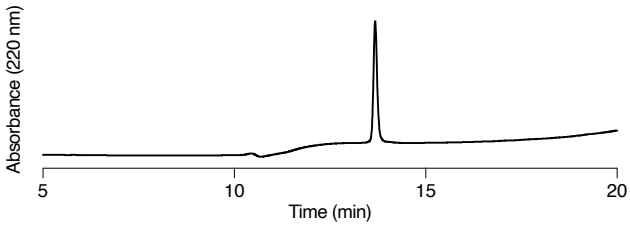
HR–MS (ESI) of Ac-VYC(Cam)C(Cam)GFGGGRG-NH_2 (48). Obs. 637.7427 (top), calc. for $C_{51}H_{75}N_{17}Na_2O_{15}S_2$ $[M+2Na]^{2+}$: 637.7427 (bottom).



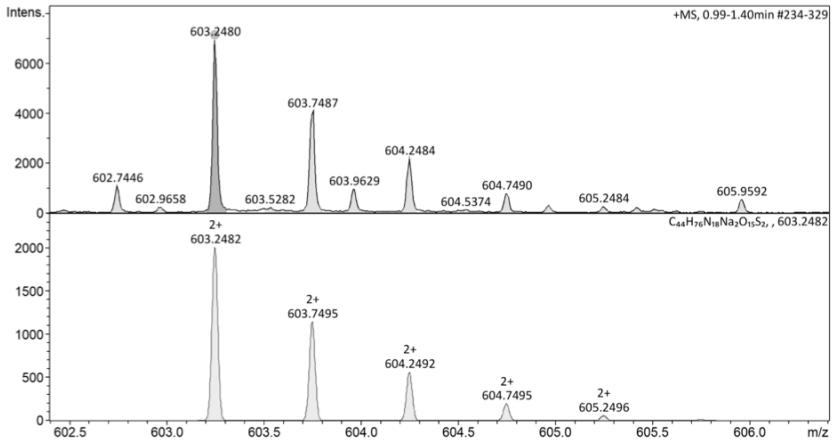
Analytical RP–HPLC of Ac-VAC(Cam)C(Cam)FDGGGRG-NH_2 (49).



HR–MS (ESI) of Ac-VAC(Cam)C(Cam)FDGGGRG-NH_2 (49). Obs. 620.7335 (top), calc. for $C_{47}H_{73}N_{17}Na_2O_{16}S_2$ $[M+2Na]^{2+}$: 620.7324 (bottom). The intense signals at 622.0289, 622.5661 and 623.0.21 m/z correspond to calibrant peaks.



Analytical RP–HPLC of Ac-VLC(Cam)C(Cam)NAGGGRG-NH_2 (50).



HR–MS (ESI) of Ac-VLC(Cam)C(Cam)NAGGGRG-NH_2 (50). Obs. 603.2480 (top), calc. for $C_{44}H_{76}N_{18}Na_2O_{15}S_2$ $[M+2Na]^{2+}$: 603.2482 (bottom).

4. Protein expression

4.1. Plasmids

All new genes were codon-optimized for *E. coli*, synthesized and subcloned into pET28 expression vectors (Merck Novagen). Variants and mutations were prepared using the QuickChange protocol (Agilent) or the Q5 protocol (New England BioLabs). All constructs were verified by DNA Sanger sequencing.

4.2. Ubc9 variants

Ubc9 constructs

A gene encoding for wild type human Ubc9 (UniProt P63279) was subcloned in frame with an N-terminal thrombin-cleavable His_6-tag GSSHHHHHHSSGLVPRGSHM (first Met processed during expression) using *Nde*I (5') and *Xho*I (3'). Further variants (R13A, K14R, N85A-C138A, S89A-C138A, T91A-C138A, C93A, C93A-C138A, D102A-C138A, D102R-C138A, E118A-C138A, E118R-C138A, E122A-C138A, E122R-C138A, D127A-C138A, C138A) were prepared using site-directed mutagenesis. A gene encoding for the X-ray construct of Ubc9 with the mutations K48A-K49A-E54A-C138A (K48A-K49A-E54A as reported in a previous X-ray study[17] and C138A to prevent disulfide formation with ligand **21** at this site) was subcloned with an N-terminal TEV-cleavable His_6-tag GSSHHHHHHSSGAENLYFQG (first Met processed during expression) using *Nco*I (5') and *Xho*I (3'). Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

Ubc9 variant expressions

Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmids and single colonies were used to inoculate overnight precultures in selective lysogeny broth (LB) Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.5 mM. Expressions were carried out for 4–5 h at 30 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C), resuspended in 20 mL lysis buffer per liter cell culture (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8.0, 350 mM NaCl, 20 mM imidazole) and stored at -80 °C until purification.

---

17 Hewitt, W. M.; Lountos, G. T.; Zlotkowski, K.; Dahlhauser, S. D.; Saunders, L. B.; Needle, D.; Tropea, J. E.; Zhan, C.; Wei, G.; Ma, B.; Nussinov, R.; Waugh, D. S.; Schneekloth, J. S. Insights into the allosteric inhibition of the SUMO E2 enzyme Ubc9. *Angew. Chem. Int. Ed.* **2016**, *55*, 5703–5707.

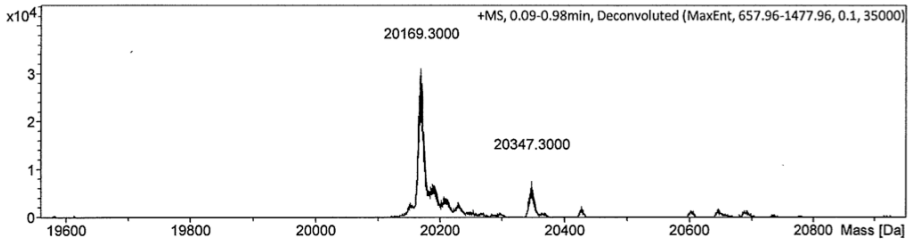
Ubc9 variants: Ubc9 variants were purified similarly to previous reports.[18] Briefly, cell suspensions were thawed at rt and placed on ice, supplemented with 1 mM final DTT, lysozyme (20 μg/mL) and DNase I (0.1 mg/mL), and nutated at 4 °C for 1 h. Cells were lysed by sonication and the suspensions were cleared by centrifugation (16,000 x g, 30 min, 4 °C) and filtration (0.2 μm membrane filter). Supernatants were subjected to gravity-flow Ni-NTA affinity purification using binding buffer (20 mM HEPES pH 8.0, 250 mM NaCl, 20 mM imidazole) and elution buffer (binding buffer containing 400 mM imidazole and 1 mM DTT), to obtain His_6-Ubc9 variants in >95% purity by SDS–PAGE. Variants were further polished by cation exchange chromatography (Mono S 5/50 GL). Samples were dialyzed against 20 mM potassium phosphates pH 7.0, 50 mM NaCl and 1 mM DTT at 4 °C and purified with buffer A (20 mM potassium phosphates pH 7.0, 1 mM DTT) and a gradient of buffer B (buffer A with 1 M NaCl). Purified samples were exchanged to reaction buffer (50 mM HEPES pH 7.6, 50 mM KCl, 1 mM reduced Ac-Cys-NHMe **1**) by dialysis or spin diafiltration (10 kDa MWCO) and concentrated to 150–250 μM by spin diafiltration (10 kDa MWCO). The samples were portioned into aliquots, flash-frozen in liquid N_2 and stored at –80 °C for months without noticeable loss of activity. Typical yields were found to be approximately 30 mg per liter cell culture for all variants. All Ubc9 variants were used with the His_6-tag attached, except the X-ray construct.

Ubc9-K48A-K49A-E54A-C138A (Ubc9-3A-C138A, X-ray construct)*:* Following lysis and Ni-NTA affinity purification of His_6-Ubc9-3A-C138A as described for the other Ubc9 variants, the sample was dialyzed against TEV cleavage buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.5 mM EDTA, 1 mM DTT) at 4 °C, diluted to approximately 100 μM (2 mg/mL) with TEV cleavage buffer, supplemented with His_6-TEV protease (1:40 weight ratio) and nutated overnight at 4 °C. The sample was subjected directly to reverse gravity-flow Ni-NTA purification (resin equilibrated in binding buffer) to remove TEV protease and uncleaved Ubc9. The flow-through containing Ubc9-3A-C138A was dialyzed against 50 mM HEPES pH 8.0, 100 mM NaCl and 1 mM DTT at 4 °C. The sample was supplemented with 1 mM TCEP and concentrated to approximately 100 μM by spin diafiltration (10 kDa MWCO). The sample was portioned into aliquots, flash-frozen in liquid N_2 and stored at –80 °C.

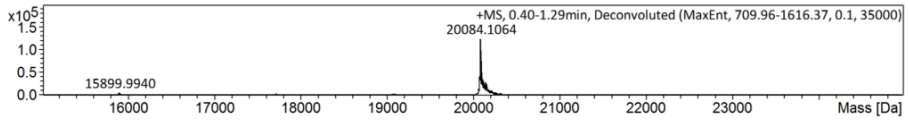
An aliquot of each variant was desalted to 0.1% formic acid using a desalting column and analyzed by ESI–MS (results shown below). Variants with an N-terminal His_6-tag were found to

[18] Yunus, A. A.; Lima, C. D. Purification of SUMO conjugating enzymes and kinetic analysis of substrate conjugation in *SUMO Protocols*; Ulrich, H. D., Ed.; Methods in Molecular Biology; Humana Press Inc.: New York, NY, 2009; Vol. 497, pp 167–186.

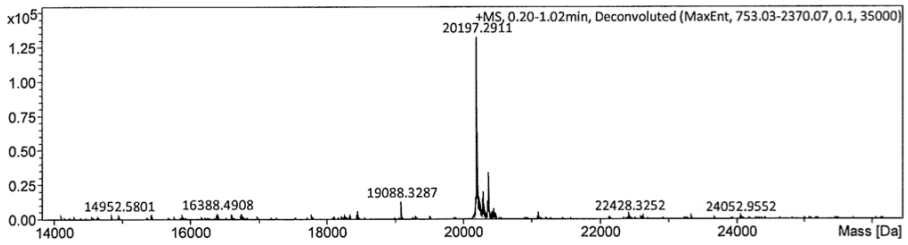
be gluconoylated to approximately 25% (+178 Da).[19,20] Met1 was found to be processed in all variants.



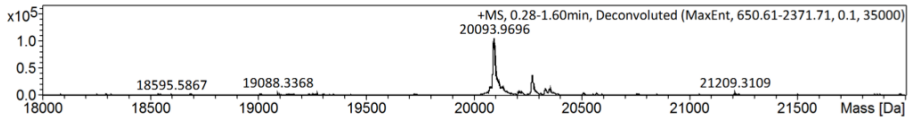
Deconvoluted ESI–MS of His_6-Ubc9 (wt). Calc. 20,170.2 Da, obs. 20,169.3 Da.



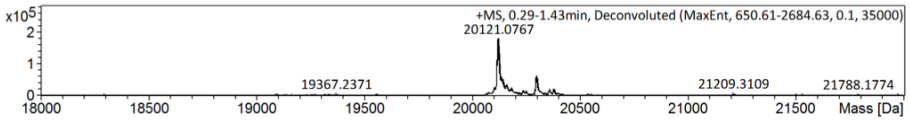
Deconvoluted ESI–MS of His_6-Ubc9-R13A. Calc. 20,085.0 Da, obs. 20,084.1 Da.



Deconvoluted ESI–MS of His_6-Ubc9-K14R. Calc. 20,198.2 Da, obs. 20,197.3 Da.



Deconvoluted ESI–MS of His_6-Ubc9-N85A-C138A. Calc. 20,095.1 Da, obs. 20,094.0 Da.

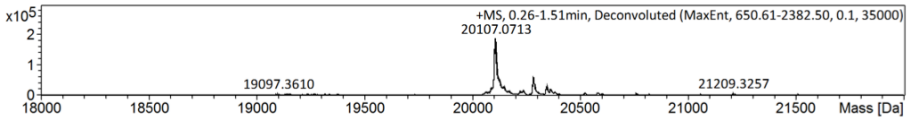


Deconvoluted ESI–MS of His_6-Ubc9-S89A-C138A. Calc. 20,122.1 Da, obs. 20,121.1 Da.

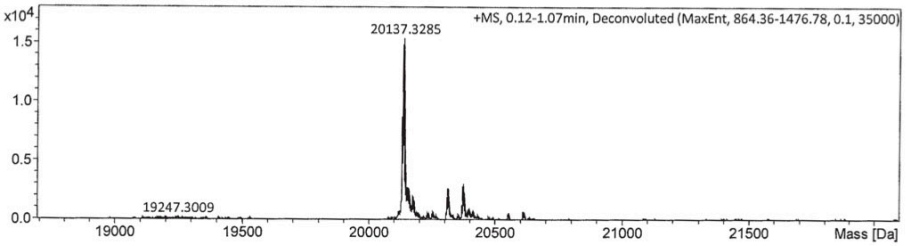
---

19 Aon, J. C.; Caimi, R. J.; Taylor, A. H.; Lu, Q.; Oluboyede, F.; Dally, J.; Kessler, M. D.; Kerrigan, J. J.; Lewis, T. S.; Wysocki, L. A.; Patel, P. S. Suppressing posttranslational gluconoylation of heterologous proteins by metabolic engineering of *Escherichia coli*. *Appl. Environ. Microbiol.* **2008**, *74*, 950-958.

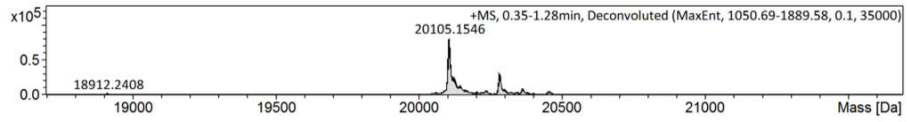
20 Geoghegan, K. F.; Dixon, H. B.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, A. S. Spontaneous alpha-*N*-6-phosphogluconoylation of a "His tag" in *Escherichia coli*: The cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **1999**, *267*, 169–184.



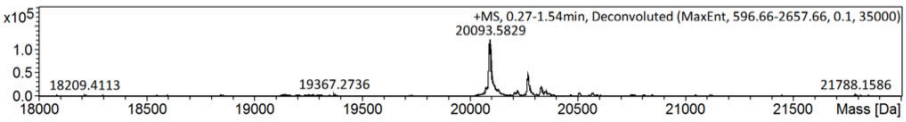
Deconvoluted ESI–MS of His_6-Ubc9-T91A-C138A. Calc. 20,108.1 Da, obs. 20,107.1 Da.



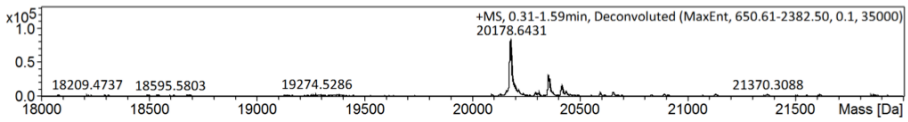
Deconvoluted ESI–MS of His_6-Ubc9-C93A. Calc. 20,138.1 Da, obs. 20,137.3 Da.



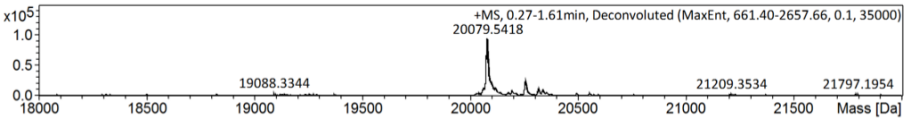
Deconvoluted ESI–MS of His_6-Ubc9-C93A-C138A. Calc. 20,106.0 Da, obs. 20,105.2 Da.



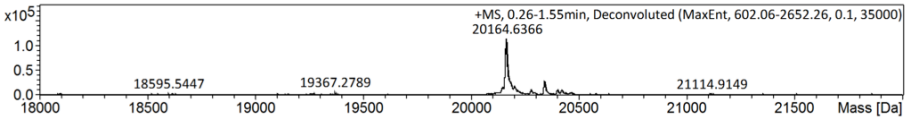
Deconvoluted ESI–MS of His_6-Ubc9-D102A-C138A. Calc. 20,094.1 Da, obs. 20093.6 Da.



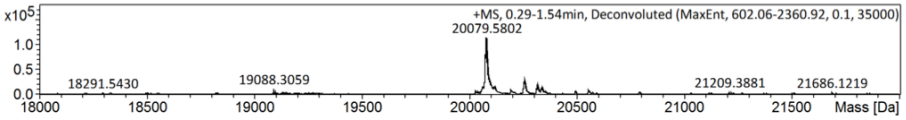
Deconvoluted ESI–MS of His_6-Ubc9-D102R-C138A. Calc. 20,179.2 Da, obs. 20,178.6 Da.



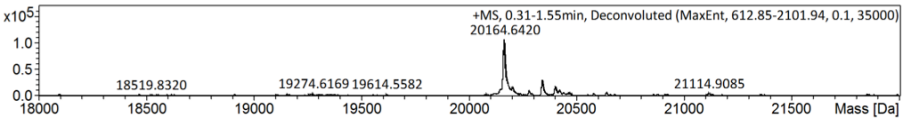
Deconvoluted ESI–MS of His_6-Ubc9-E118A-C138A. Calc. 20,080.1 Da, obs. 20,079.5 Da.



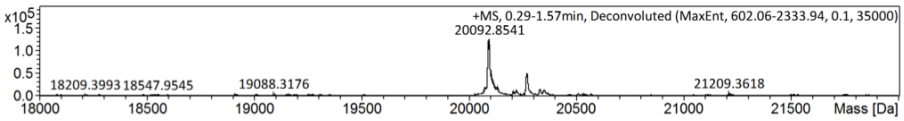
Deconvoluted ESI–MS of His_6-Ubc9-E118R-C138A. Calc. 20,165.2 Da, obs. 20,164.6 Da.



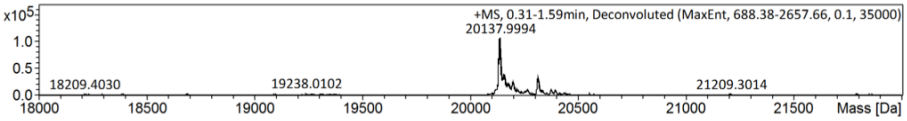
Deconvoluted ESI–MS of His_6-Ubc9-E122A-C138A. Calc. 20,080.1 Da, obs. 20,079.6 Da.



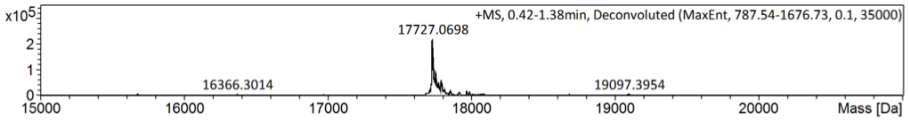
Deconvoluted ESI–MS of His_6-Ubc9-E122R-C138A. Calc. 20,165.2 Da, obs. 20,164.6 Da.



Deconvoluted ESI–MS of His_6-Ubc9-D127A-C138A. Calc. 20,094.1 Da, obs. 20,092.9 Da.



Deconvoluted ESI–MS of His_6-Ubc9-C138A. Calc. 20,138.1 Da, obs. 20,138.0 Da.



Deconvoluted ESI–MS of His_6-3A-C138A. Calc. 17,728.4 Da, obs. 17,727.1 Da.

4.3. GFP variants

GFP constructs

LACE: A gene encoding for GFP[21] was subcloned with an N-terminal TEV-cleavable His_6-tag GSSHHHHHHSSGAENLYFQG (first Met processed during expression) using *Nco*I (5') and *Xho*I (3'). Test constructs of GFP were prepared with C-terminal tags GSGHMGLL**K**SEDKV (GFP-RanGAP1, including residues 519–529 of RanGAP1, UniProt P46060) and GSGPRDGL**K**KERLL (GFP-IκBα, including residues 16–26 of IκBα, UniProt P25963). The LACE tag GSGPRKVI**K**MESEE (including residues 485–495 of PML, UniProt P29590) was fused to the C-terminus of GFP (GFP-$LACE_C$). The internal LACE tag GSGPRKVI**K**MESEEGS was inserted into the GFP sequence after residue D173 (GFP-$LACE_I$). GFP-$LACE_N$ carried an N-terminal LACE tag inserted after the TEV cleavage site of the His_6-tag (GSSHHHHHHSSGAENLYFQGPRKVI**K**MESEEGSG). GFP-$LACE_{I,C}^{2M}$ contained an internal LACE tag as well as the C-terminal LACE tag[2M] GSGPRAVI**K**QESEE. For GFP variants with minimal LACE tags, the residues GSGI**K**QE and GSGI**K**QEGS were fused to the C-terminus (GFP-$LACE_C^{minimal}$) or inserted internally after residue D173 (GFP-$LACE_I^{minimal}$), respectively. Further variants and point mutants were prepared by site-directed mutagenesis (Acceptor lysine to arginine tag mutations, alanine scan of GFP-$LACE_C$ tag).

Disulfide tag constructs: Constructs encoding for GFP[21] with N-terminal CysTag variants (MXXCCXXGSG) and a C-terminal His_6-tag (SLEHHHHHH) were prepared in pET28b. The tested CysTag sequences are shown in Figure 98.

Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

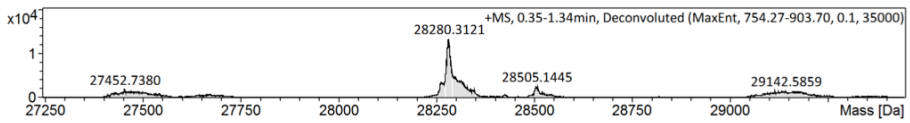
---

[21] Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.

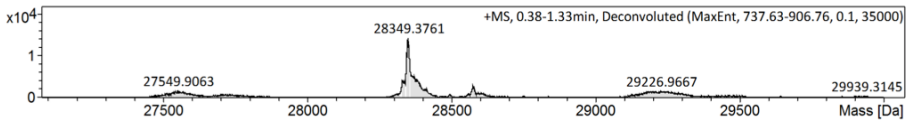
GFP variant expressions

LACE substrates: Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmids and single colonies were used to inoculate overnight precultures in selective LB Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 0.5 mM, Expressions were carried out overnight at 18 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C), resuspended in 20 mL lysis buffer per liter cell culture (50 mM HEPES pH 8.0, 350 mM NaCl, 20 mM imidazole) and stored at –80 °C until purification. Purification of GFP variants and His_6-tag cleavage using TEV was performed as described for Ubc9-3A-C138A. Variants were further purified by anion exchange chromatography (Mono Q 5/50 GL). The samples were dialyzed against 25 mM Tris-HCl pH 8.5 at 4 °C and purified with buffer A (25 mM Tris-HCl pH 8.5) and a gradient of buffer B (buffer A with 1 M NaCl). Purified samples were exchanged to reaction buffer, concentrated to 100–200 μM by spin diafiltration (10 kDa MWCO) and stored at 4 °C until use. For screening purposes (alanine scanning), the His_6-tag was kept and Ni-NTA purified samples (>95% purity by SDS–PAGE) were exchanged to reaction buffer by dialysis or spin diafiltration (10 kDa MWCO) and assayed directly. Typical yields were found to be approximately 150 mg per liter cell culture for all variants. An aliquot of each variant was desalted to 0.1% formic acid using a desalting column and analyzed by ESI–MS (results shown below). Variants with an N-terminal His_6-tag were found to be gluconoylated to approximately 25% (+178 Da).[19,20] Met1 was found to be processed in all variants.

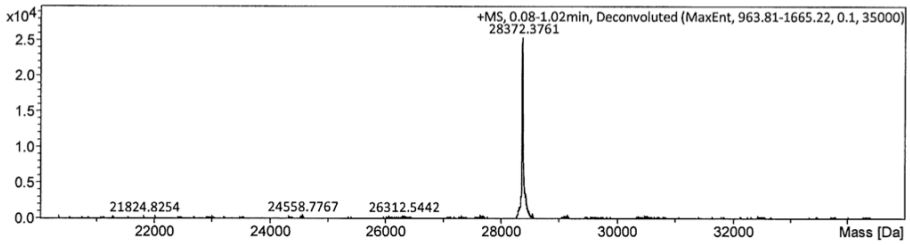
CysTag variants: CysTag variants were expressed, harvested and lysed as described for the GFP LACE substrates (also without any addition of thiol reducing agents during the purification). The cleared lysates from 50 mL cell culture expressions were subjected to gravity-flow Ni-NTA purification (1 mL equilibrated slurry) using binding buffer (20 mM HEPES pH 8.0, 250 mM NaCl, 20 mM imidazole) and elution buffer (binding buffer containing 400 mM imidazole and 1 mM DTT). The eluted CysTag variants (approximately 1 mL per sample, 150–200 μM concentration) were dialyzed against reaction buffer (50 mM HEPES pH 7.6, 50 mM KCl) and stored at 4 °C until use. The dialyzed samples were analyzed by non-reducing SDS–PAGE according to the general procedure. Results of the non-reducing SDS–PAGE analyses are shown in Figure 98b.



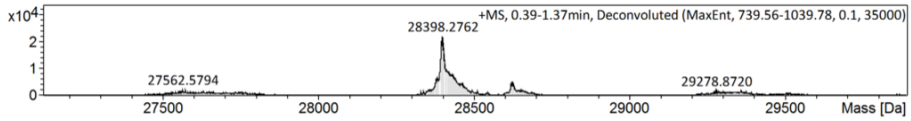
Deconvoluted ESI–MS of GFP-RanGAP1. Calc. 28,282.9 Da, obs. 28,280.3 Da.



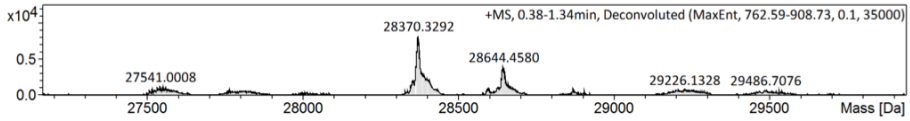
Deconvoluted ESI–MS of GFP-IκBα. Calc. 28,351.0 Da, obs. 28,349.4 Da.



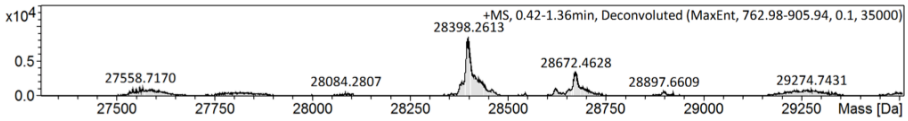
Deconvoluted ESI–MS of GFP-LACE$_{C}$ (PML). Calc. 28,372.0 Da, obs. 28,372.4 Da.



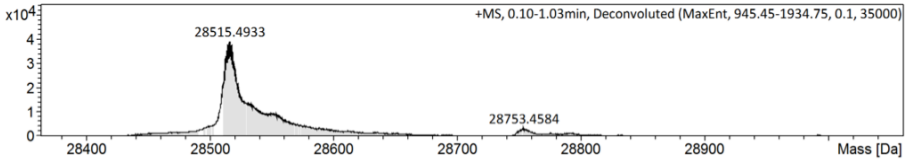
Deconvoluted ESI–MS of GFP-LACE$_{C}^{K6R}$. Calc. 28,400.1 Da, obs. 28,398.3 Da.



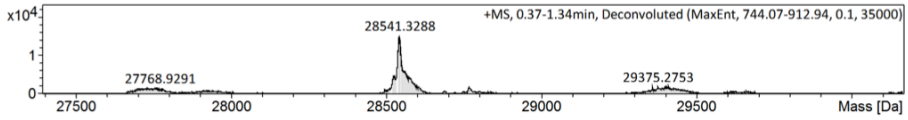
Deconvoluted ESI–MS of GFP-LACE$_{N}$. Calc. 28,372.0 Da, obs. 28,370.3 Da.



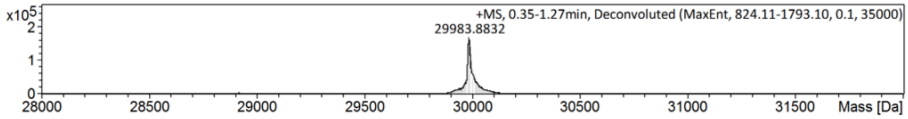
Deconvoluted ESI–MS of GFP-LACE$_{N}^{K6R}$. Calc. 28,400.1 Da, obs. 28,398.3 Da.



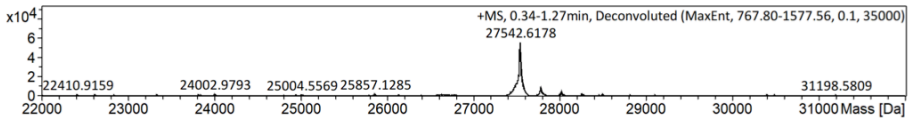
Deconvoluted ESI–MS of GFP-LACE$_{I}$. Calc. 28,516.2 Da, obs. 28,515.5 Da.



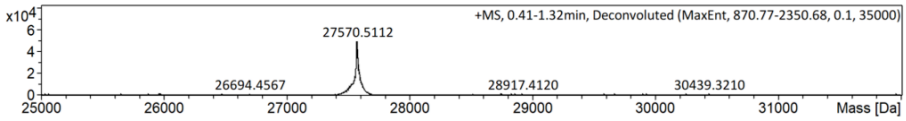
Deconvoluted ESI–MS of GFP-LACE$_{I}^{K6R}$. Calc. 28,544.2 Da, obs. 28,541.3 Da.



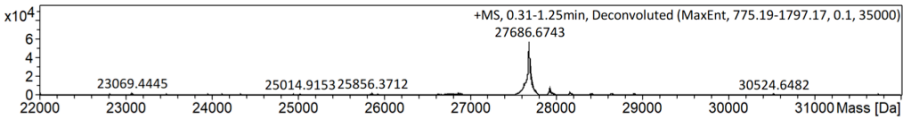
Deconvoluted ESI–MS of GFP-LACE$_{I,C}^{2M}$. Calc. 29,984.8 Da, obs. 29,983.9 Da.



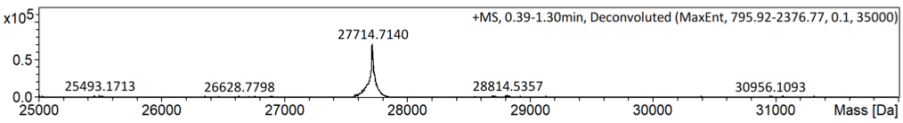
Deconvoluted ESI–MS of GFP-LACE$_{C}^{minimal}$. Calc. 27,543.2 Da, obs. 27,542.6 Da.



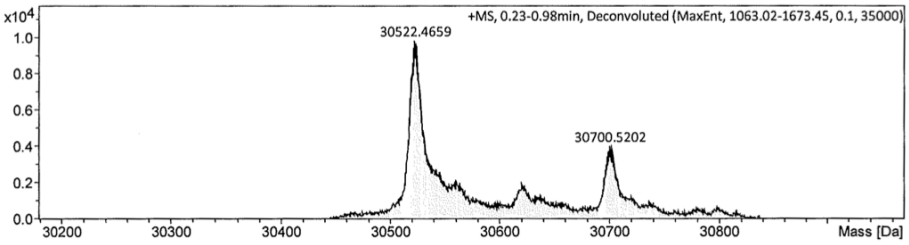
Deconvoluted ESI–MS of GFP-LACE$_C^{minimal\text{-}KtoR}$. Calc. 27,571.1 Da, obs. 27,570.5 Da.



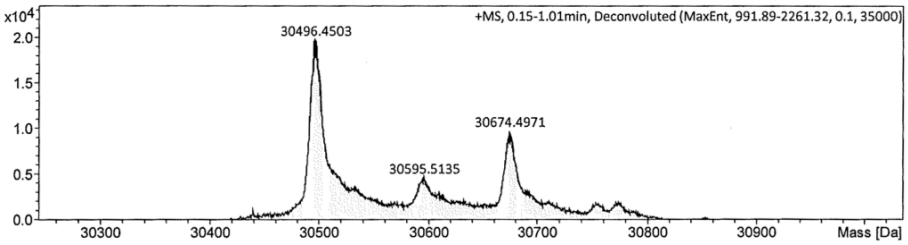
Deconvoluted ESI–MS of GFP-LACE$_I^{minimal}$. Calc. 27,687.2 Da, obs. 27,686.7 Da.



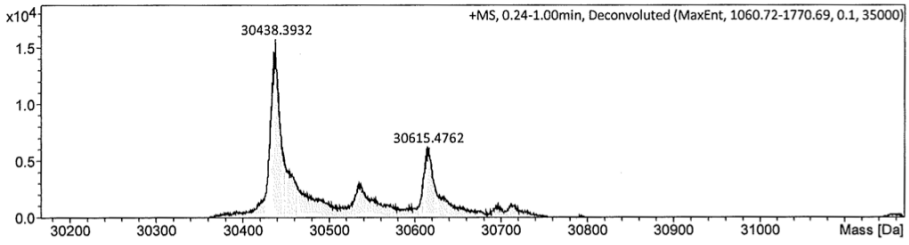
Deconvoluted ESI–MS of GFP-LACE$_I^{minimal\text{-}KtoR}$. Calc. 27,715.2 Da, obs. 27,714.7 Da.



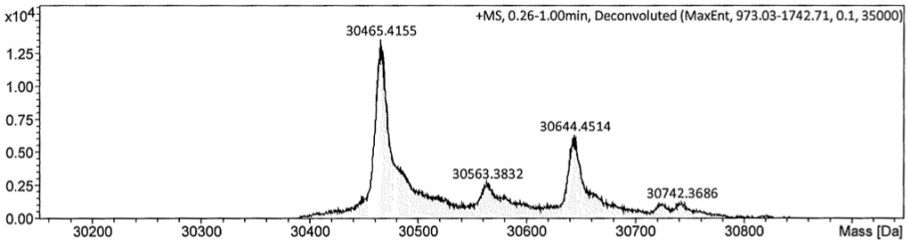
Deconvoluted ESI–MS of His$_6$-GFP-LACE$_C$. Calc. 30,523.2 Da, obs. 30,522.5 Da.



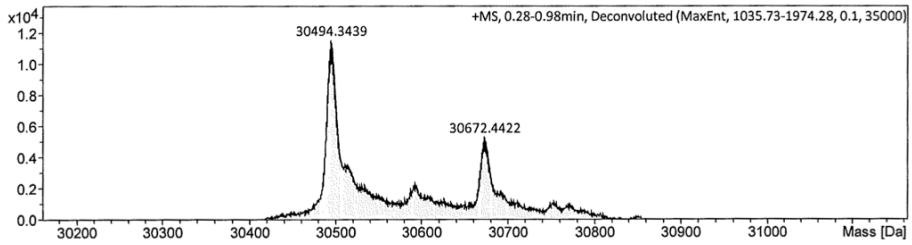
Deconvoluted ESI–MS of His$_6$-GFP-LACE$_C^{P1A}$. Calc. 30,497.2 Da, obs. 30,496.5 Da.



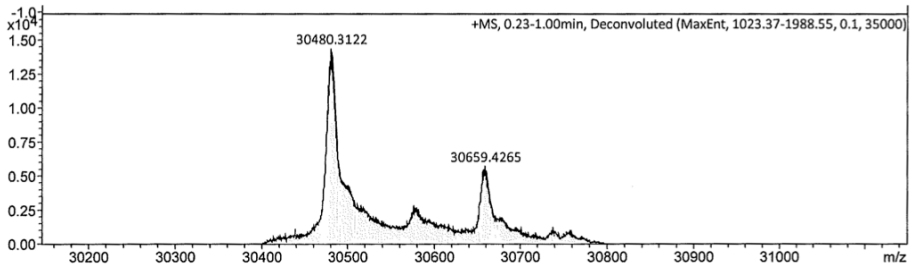
Deconvoluted ESI–MS of His$_6$-GFP-LACE$_C^{R2A}$. Calc. 30,438.1 Da, obs. 30,438.4 Da.



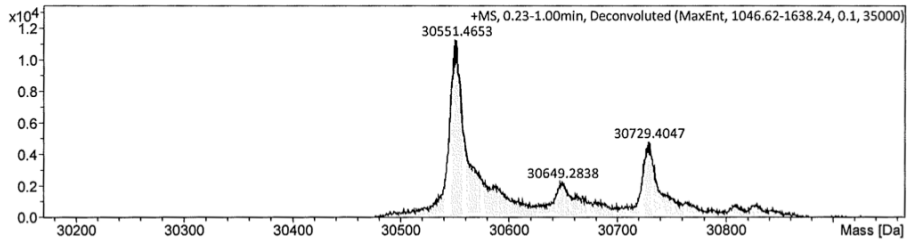
Deconvoluted ESI–MS of His$_6$-GFP-LACE$_C^{K3A}$. Calc. 30,466.1 Da, obs. 30,465.4 Da.



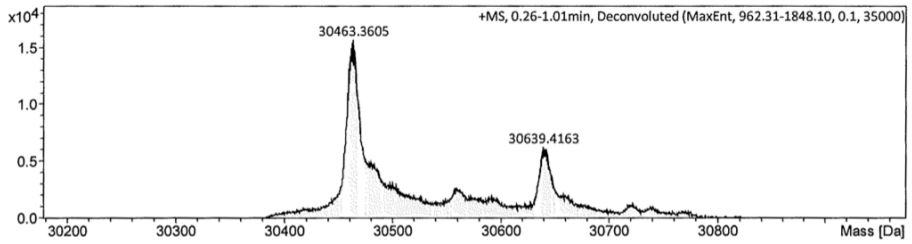
Deconvoluted ESI–MS of His_6-GFP-$LACE_C^{V4A}$. Calc. 30,495.2 Da, obs. 30,494.3 Da.



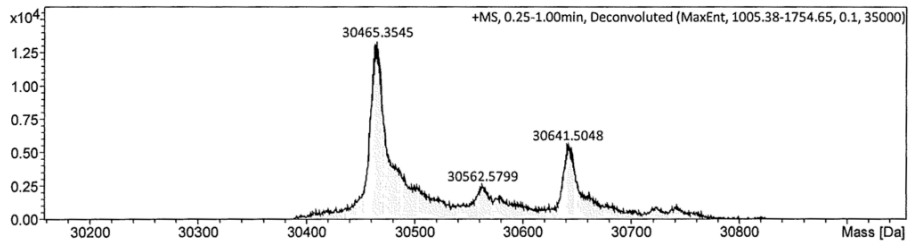
Deconvoluted ESI–MS of His_6-GFP-$LACE_C^{I5A}$. Calc. 30,481.2 Da, obs. 30,480.3 Da.



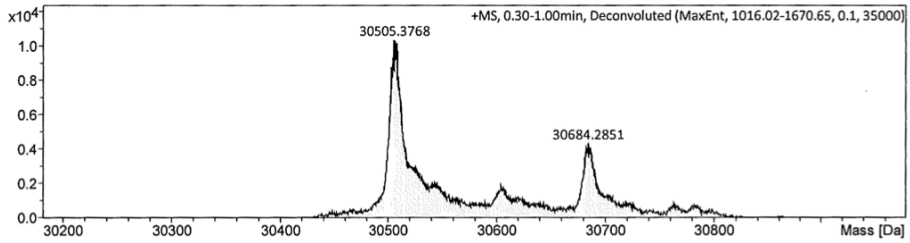
Deconvoluted ESI–MS of His_6-GFP-$LACE_C^{K6R}$. Calc. 30,551.3 Da, obs. 30,551.5Da.



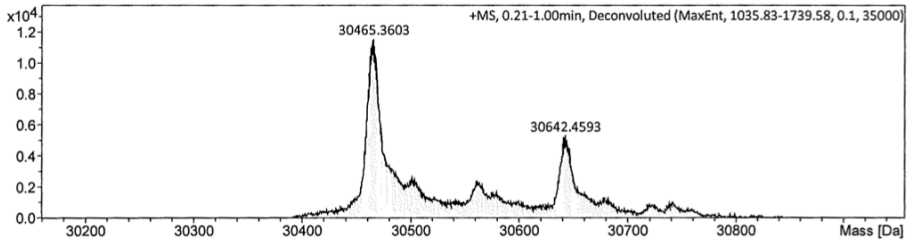
Deconvoluted ESI–MS of His_6-GFP-$LACE_C^{M7A}$. Calc. 30,463.1 Da, obs. 30,463.4Da.



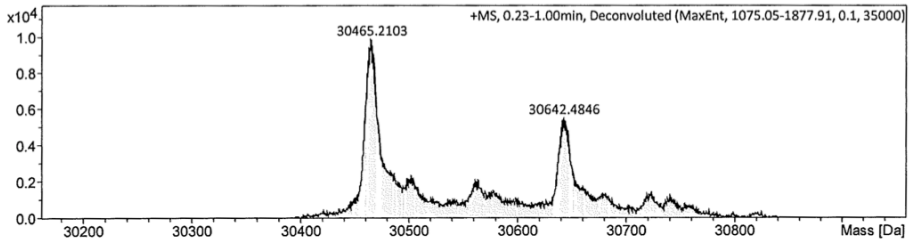
Deconvoluted ESI–MS of His_6-GFP-$LACE_C^{E8A}$. Calc. 30,465.2 Da, obs. 30,465.4 Da.



Deconvoluted ESI–MS of His_6-GFP-$LACE_C^{S9A}$. Calc. 30,507.2 Da, obs. 30,505.4Da.



Deconvoluted ESI–MS of His_6-GFP-$LACE_C^{E10A}$. Calc. 30,465.2 Da, obs. 30,465.4Da.



Deconvoluted ESI–MS of His_6-GFP-$LACE_C^{E11A}$. Calc. 30,465.2 Da, obs. 30,465.2 Da.

4.4. Titin IG27 variants

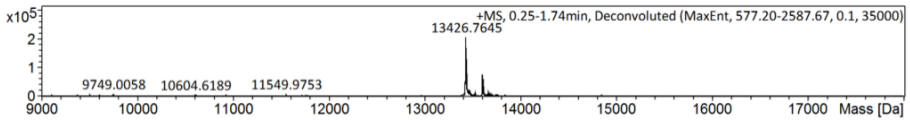
Titin constructs

A gene encoding for the human immunoglobulin 27 domain of titin (UniProt Q8WZ42) with the mutations C47S and C63S was subcloned with an N-terminal TEV-cleavable His_6-tag GSSHHHHHHSSGAENLYFQG (first Met processed during expression) using *Nco*I (5') and *Xho*I (3'). The LACE tag^{2M} GSGPRAVI**K**QESEE was fused to the C-terminus of titin (titin-$LACE_C^{2M}$). The acceptor tag mutant K6R (GSGPRAVI**R**QESEE) was prepared by site-directed mutagenesis. Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

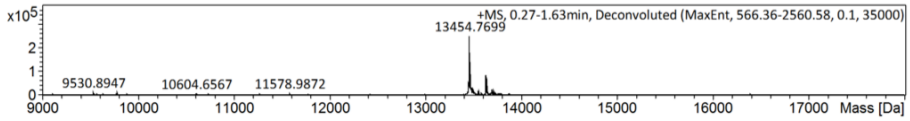
Titin variant expressions

Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmids and single colonies were used to inoculate overnight precultures in selective LB Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 0.5 mM. Expressions were carried out overnight at 18 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C), resuspended in 20 mL lysis buffer per liter cell culture (50 mM HEPES pH 8.0, 350 mM NaCl, 20 mM imidazole) and stored at –80 °C until purification. Purification of titin variants and His_6-tag cleavage using

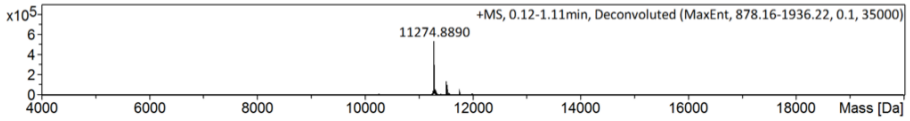
TEV was performed as described for Ubc9-3A-C138A. Titin variants were obtained in >95% purity by SDS–PAGE. Samples were exchanged to reaction buffer, concentrated to approximately 100 μM by spin diafiltration (3 kDa MWCO) and stored at 4 °C until use. Typical yields were found to be approximately 30 mg per liter cell culture for all variants. An aliquot of each variant was desalted to 0.1% formic acid using a desalting column and analyzed by ESI–MS (results shown below). Variants with an N-terminal His_6-tag were found to be gluconoylated to approximately 25% (+178 Da).[19,20] Met1 was found to be processed in all variants.



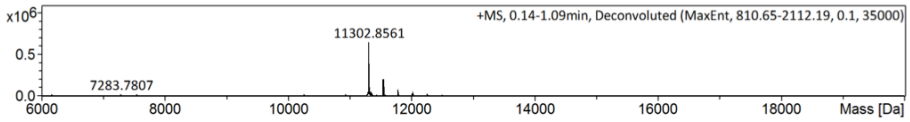
Deconvoluted ESI–MS of His_6-titin-$LACE_C^{2M}$. Calc. 13,426.9 Da, obs. 13,426.8 Da.



Deconvoluted ESI–MS of His_6-titin-$LACE_C^{2M\text{-}K6R}$. Calc. 13,455.0 Da, obs. 13,454.8 Da.



Deconvoluted ESI–MS of titin-$LACE_C^{2M}$. Calc. 11,275.8 Da, obs. 11,274.9 Da.



Deconvoluted ESI–MS of titin-$LACE_C^{2M\text{-}K6R}$. Calc. 11,303.8 Da, obs. 11,302.9 Da.

4.5. T4L variants

T4L constructs

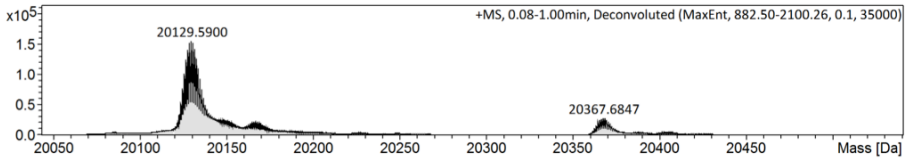
A pET28 plasmid harboring T4L with mutations C54T and C97A was a gift from Christopher J. White.[22] The LACE tag GSGPRKVI**K**MESEE was fused to the C-terminus of T4L (T4L-$LACE_C$). An acceptor tag mutant T4L-$LACE_C^{2M\text{-}K6R}$ was prepared by site-specific mutagenesis. Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

T4L variant expressions

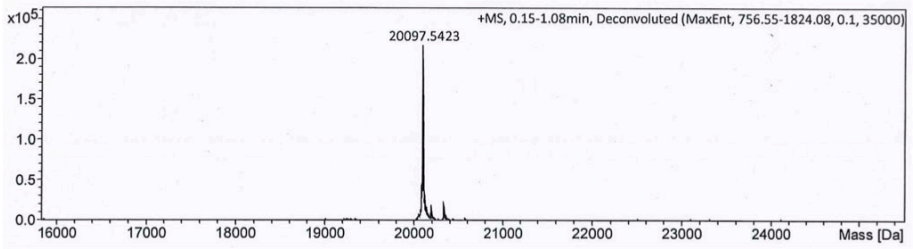
Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmids and single colonies were used to inoculate overnight precultures in selective LB Miller medium.

[22] White, C. J.; Bode, J. W. PEGylation and dimerization of expressed proteins under near equimolar conditions with potassium 2-pyridyl acyltrifluoroborates. *ACS Cent. Sci.* **2018**, *4*, 197–206.

Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 1 mM. Expressions were carried out for 2 h at 37 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C), resuspended in in 20 mL lysis buffer per liter cell culture (25 mM Tris, 25 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS), 0.2 mM EDTA, pH 7.6) and stored at –20 °C until purification. T4L variants were purified similarly to previous reports.[23,24] Briefly, cell suspensions were thawed at rt and placed on ice, supplemented with 1 mM final DTT and Benzonase (1 μL per 10 mL suspension) and incubated on ice for 1 h to induce self-lysis. Cells were further lysed by sonication, and the suspension was cleared by centrifugation (16,000 x g, 30 min, 4 °C) and filtration (0.2 μm membrane filter). Cleared lysates were purified by cation exchange chromatography (Mono S 5/50 GL) with buffer A (25 mM Tris, 25 mM MOPS, 0.2 mM EDTA, pH 7.6) and a gradient of buffer B (buffer A with 1 M NaCl). Fractions containing the T4L variant were pooled and passed through a 50 kDa MWCO filter via spin diafiltration to remove higher molecular weight impurities. Purified T4L variants were exchanged to reaction buffer and concentrated to a final concentration of 100–200 μM by spin diafiltration (10 kDa MWCO), and stored at 4 °C until use. Typical yields were found to be approximately 20 mg per liter cell culture. An aliquot of each variant was desalted to 0.1% formic acid using a desalting column and analyzed by ESI–MS (results shown below). Met1 was found to be processed in all variants.



Deconvoluted ESI–MS of T4L-LACE$_C$. Calc. 20,130.1 Da, obs. 20,129.6 Da.



Deconvoluted ESI–MS of T4L-LACE$_C^{2M\text{-}K6R}$. Calc. 20,098.0 Da, obs. 20,097.5 Da.

---

23 Sauer, U. H.; Dao-pin, S.; Matthews, B. W. Tolerance of T4 lysozyme to proline substitutions within the long interdomain alpha-helix illustrates the adaptability of proteins to potentially destabilizing lesions. *J. Biol. Chem.* **1992**, *267*, 2393–2399.

24 Columbus, L.; Kálai, T.; Jekö, J.; Hideg, K.; Hubbell, W. L. Molecular motion of spin labeled side chains in alpha-helices: analysis by variation of side chain structure. *Biochemistry* **2001**, *40*, 3828–3846.

4.6. SpyCatcher variants and SpyTagMBP

SpyCatcher and SpyTagMBP constructs

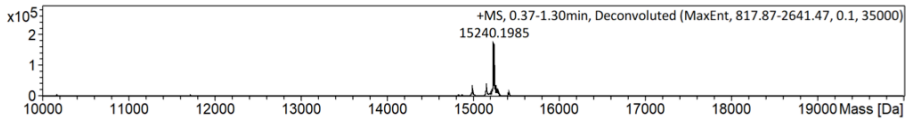
His_6-tagged SpyCatcher in pDEST14 (Addgene plasmid #35044) and SpyTagMBP in pET28a (Addgene plasmid #35050) were gifts from Mark Howarth. The competent variant (^{33}IKSE^{36} instead of ^{33}LSSE^{36}, residue numbers relative to Met1 of the Addgene construct) and the acceptor mutant (^{33}IRSE^{36}) were prepared by site-directed mutagenesis. Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

SpyCatcher variant and SpyTagMBP expressions

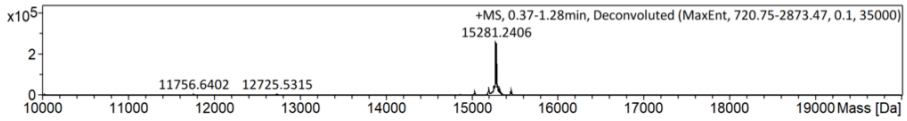
Chemically competent BL21 (DE3) cells, or BL21 (DE3) pLysS cells for expression of SpyCatcher variants, were heat-shock transformed with the plasmids and single colonies were used to inoculate overnight precultures in selective LB Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 0.5 mM. Expressions were carried out for 4–5 h at 30 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C), resuspended in 20 mL lysis buffer per liter cell culture (50 mM Tris pH 7.6, 150 mM NaCl, 20 mM imidazole) and stored at –80 °C until purification. SpyCatcher variants and SpyTagMBP were purified similarly to previous reports.[25] Lysis of cell suspensions was performed as described for Ubc9 variants, and the lysates were additionally supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF). The cleared supernatants were subjected to gravity-flow Ni-NTA affinity purification using binding buffer (50 mM Tris pH 7.6, 150 mM NaCl, 20 mM imidazole) and elution buffer (binding buffer containing 400 mM imidazole), to obtain variants in >95% purity by SDS–PAGE. SpyCatcher variants were exchanged to PBS and stored at a final concentration of 300–400 μM at 4 °C until use. SpyTagMBP was exchanged to reaction buffer, concentrated to a final concentration of 250 μM by spin diafiltration (10 kDa MWCO), and stored at 4 °C until use. Yields were found to be approximately 40 mg and 100 mg per liter cell culture for SpyCatcher variants and SpyTagMBP, respectively. An aliquot of each variant was desalted to 0.1% formic acid using a desalting column and analyzed by ESI–MS (results shown below). Met1 was found to be processed in all variants.

---

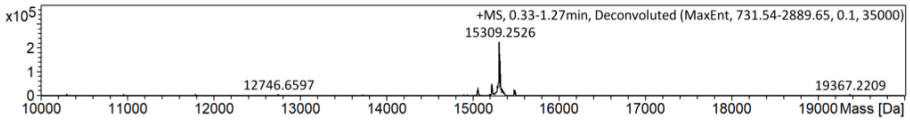
[25] Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E690–E697.



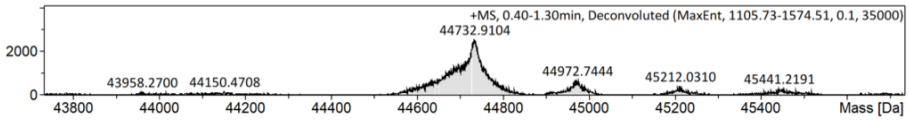
Deconvoluted ESI–MS of SpyCatcher (LSSE, wt). Calc. 15,240.6 Da, obs. 15,240.2 Da.



Deconvoluted ESI–MS of SpyCatcher (IKSE). Calc. 15,281.7 Da, obs. 15,281.2 Da.



Deconvoluted ESI–MS of SpyCatcher (IRSE). Calc. 15,309.7 Da, obs. 15,309.3 Da.



Deconvoluted ESI–MS of SpyCatcher (SpyTagMBP). Calc. 44,733.6 Da, obs. 44,732.9 Da.

4.7. Trastuzumab Fab

Trastuzumab Fab construct

A bicistronic gene encoding for the light and heavy chain of anti-HER2 (trastuzumab) Fab, each with a PelB leader sequence MKYLLPTAAAGLLLLAAQPAMA (cleaved during periplasmic expression) and separated by a non-coding internal ribosomal binding site (taataatggtaccgccattcggccgcagaaataattttgtttaactttaagaaggagatatacg), was subcloned using the restriction enzymes *Xba*I (5') and *Nde*I (3'). The light chain contained a C-terminal LACE tag^{2M} (GGSGRGSGPRAVI**K**QESEE), and the heavy chain contained a C-terminal sortase-reactive sequence with a His_6-tag (GGSGRSLPETGGHHHHHHV). Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

Trastuzumab Fab expression

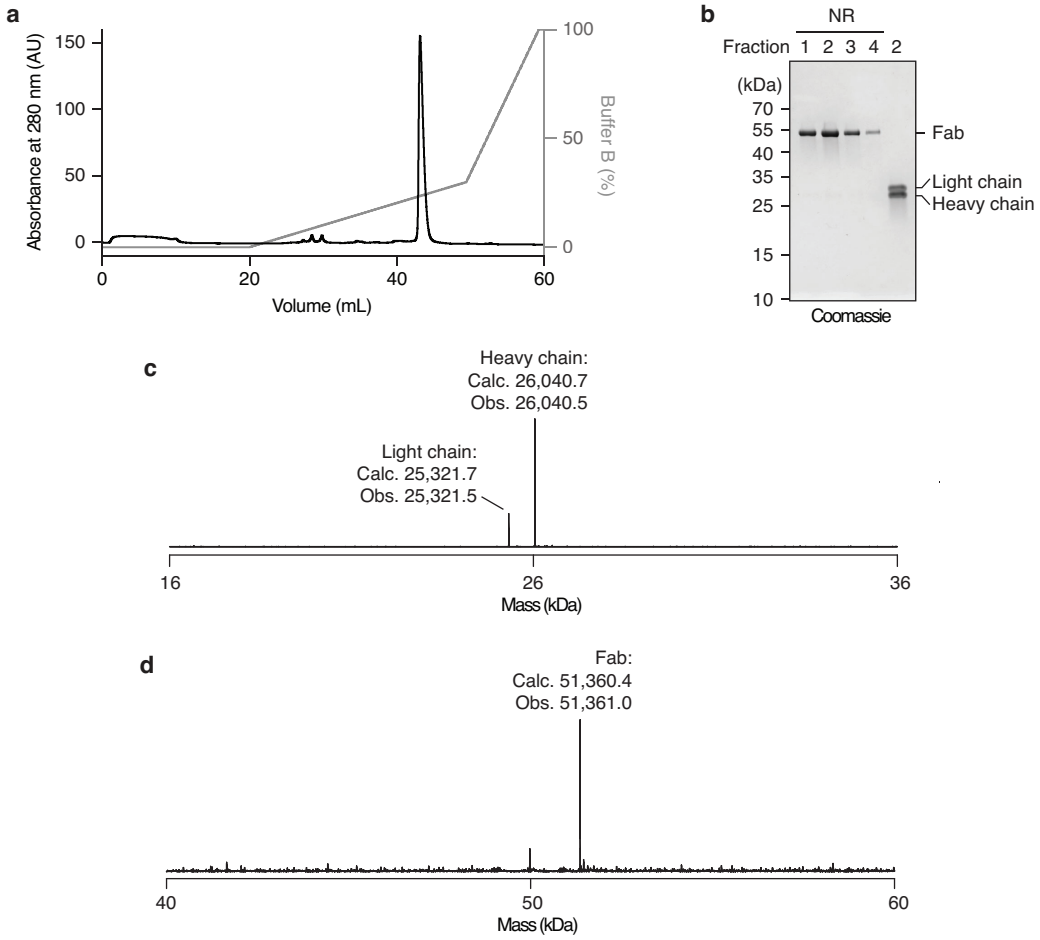
Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmid and a single colony was used to inoculate an overnight preculture in selective LB Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 0.5 mM. Expressions were carried out for 4–5 h at 30 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C) and the protein was immediately purified after harvesting. Fab was purified from the periplasm by osmotic shock similarly to previous reports.[26,27] Briefly, freshly harvested cells were resuspended in 20 mM Tris-HCl pH 8.0, 2 mM EDTA and 40% w/v sucrose (one tenth of the cell culture volume), agitated at 25 °C at 100 rpm for 10 min and pelleted by centrifugation (16,000 x g, 20 min, 4 °C). The supernatant was decanted and the pellet was quickly resuspended in ice-cold osmotic shock buffer (one tenth of the cell culture volume, 10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$) by vortexing and pipetting. The suspension was incubated on ice for 5 min before centrifugation (16,000 x g, 20 min, 4 °C). The supernatant containing the Fab fragment was supplemented with 1 M HEPES pH 8.0 to reach a final concentration of 10 mM HEPES, filtered (0.2 μm membrane filter) and subjected to gravity-flow Ni-NTA affinity purification as described in the section for Ubc9 variants. The eluted sample was exchanged to 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.5 by spin diafiltration (30 kDa MWCO) and further purified by cation exchange chromatography (Mono S 5/50 GL) with buffer A (50 mM MES pH 6.5) and a gradient of buffer B (buffer A with 1

---

[26] Nossal, N. G.; Heppel, L. A. The release of enzymes by osmotic shock from *Escherichia coli* in exponential phase. *J. Biol. Chem.* **1966**, *241*, 3055–3062.

[27] Hsu, C.-C.; Thomas, O. R. T.; Overton, T. W. Periplasmic expression in and release of Fab fragments from *Escherichia coli* using stress minimization. *J. Chem. Technol. Biotechnol.* **2016**, *91*, 815–822.

M NaCl). The protein eluted as a single species and was obtained as the disulfide-bonded heterodimer as shown by reducing and non-reducing SDS–PAGE and ESI–MS (results shown below). The PelB leader sequence was found to be processed. Fractions containing the Fab fragment were pooled, exchanged to reaction buffer, concentrated to 80 μM by spin diafiltration (30 kDa MWCO), and stored at 4 °C until use. The yield was approximately 50 μg per liter cell culture.



Fab purification and characterization. (**a**) Cation exchange chromatogram of Ni-NTA-purified Fab sample. (**c**) SDS–PAGE analysis of four peak fractions from cation exchange purification under reducing and non-reducing (NR) conditions. (**c**,**d**) Deconvoluted ESI–MS after reduction of the interchain disulfide bond (**c**) and of the intact Fab molecule (**d**).

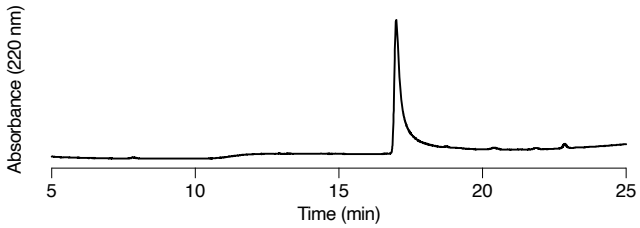
4.8. SUMO2 variants

SUMO2 constructs

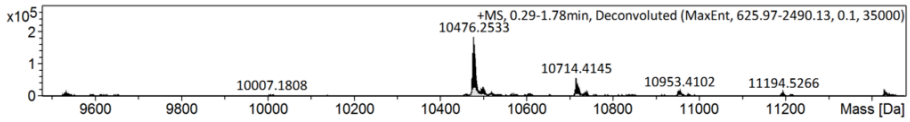
Wild type human SUMO2 (UniProt P61956) in pET11a was a gift from Frauke Melchior (Addgene plasmid #53143). Further variants (K11R or V10I) were prepared by site-directed mutagenesis. Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

SUMO2 variant expressions

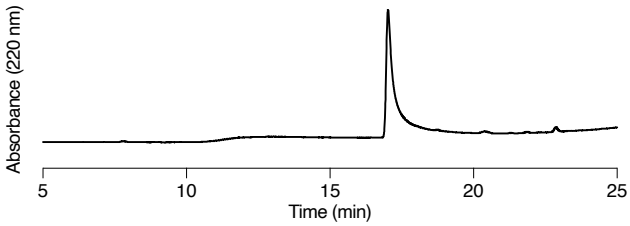
Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmids and single colonies were used to inoculate overnight precultures in selective LB Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 0.5 mM. Expressions were carried out for 4 h at 37 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C), resuspended in 20 mL lysis buffer per liter cell culture (50 mM Tris pH 8.0, 50 mM NaCl) and stored at –80 °C until purification. Lysis of cell suspensions was performed as described for Ubc9 variants. The cleared supernatants were dialyzed against Milli-Q water with 1% AcOH, 30% MeCN and purified by preparative RP–HPLC using a gradient of 20 to 50% solvent B over 30 min. Fractions containing pure SUMO2 variant were pooled and lyophilized. Yields were found to be approximately 20 mg per liter cell culture. The purified variants were characterized by analytical RP–HPLC and ESI–MS (results shown below). Met1 was found to be processed in all variants.



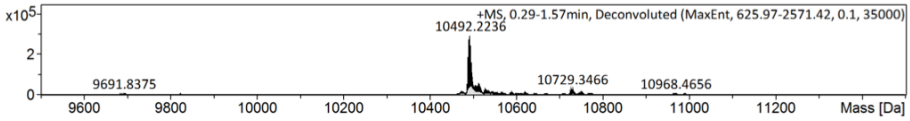
Analytical RP–HPLC of SUMO2 (wt). 10 to 95% solvent B over 20 min.



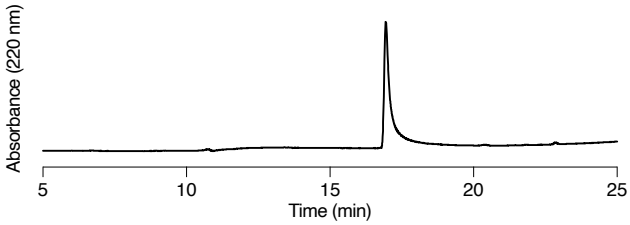
Deconvoluted ESI–MS of SUMO2 (wt). Calc. 10,477.2 Da, obs. 10,476.3 Da.



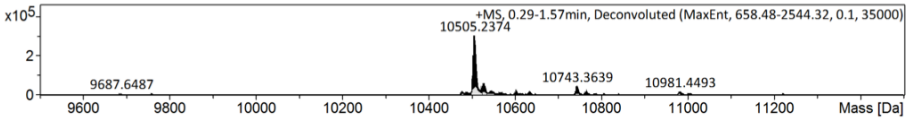
Analytical RP–HPLC of SUMO2-V10I (K11*). 10 to 95% solvent B over 20 min.



Deconvoluted ESI–MS of SUMO2-V10I (K11*). Calc. 10,491.2 Da, obs. 10,492.2 Da.



Analytical RP–HPLC of SUMO2-K11R. 10 to 95% solvent B over 20 min.



Deconvoluted ESI–MS of SUMO2-K11R. Calc. 10,505.2 Da, obs. 10,505.2 Da.

4.9. α-Synuclein variants

α-Synuclein constructs

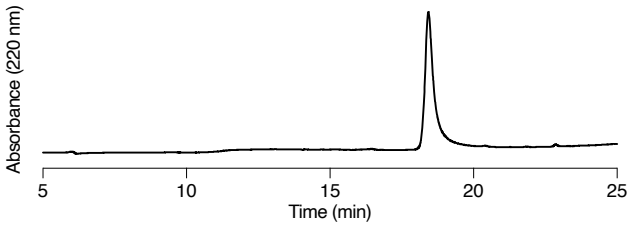
A gene encoding for wild type α-synuclein (UniProt P37840) was subcloned using *Nco*I (5') and *Xho*I (3'). Further variants (K96R-K102R, V95I-D98E-K102R, K96R-G101I) were prepared by site-directed mutagenesis. Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

α-Synuclein variant expressions

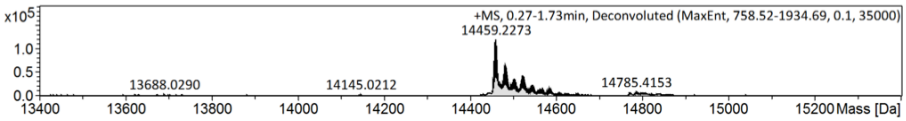
Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmids and single colonies were used to inoculate overnight precultures in selective LB Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 0.5 mM. Expressions were carried out for 4 h at 37 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C) and the protein was immediately purified after harvesting. α-Synuclein variants were purified from the periplasm by osmotic shock, similarly to previous reports.[28] Briefly, freshly harvested cells were resuspended in 30 mM Tris-HCl pH 7.2, 2 mM EDTA and 40% w/v sucrose (one tenth of the cell culture volume), agitated at 25 °C at 100 rpm for 10 min and pelleted by centrifugation (16,000 x g, 20 min, 4 °C). The supernatant was decanted and the pellet was quickly resuspended in ice-cold 7 mM $MgCl_2$ (one twentieth of the cell culture volume) by vortexing and pipetting. The suspension was incubated on ice for 5 min before centrifugation (16,000 x g, 20 min, 4 °C). The supernatant containing the α-synuclein variant was dialyzed into 20 mM bis-(2-hydroxyethyl)-amino-tris(hydroxymethyl)methane (Bis-Tris) pH 6.5 and purified by HiTrap Capto Q using buffer A (20 mM Bis-Tris pH 6.5) and a gradient of buffer B (buffer A with 1 M NaCl). The fraction containing α-synuclein was collected and purified by preparative RP–HPLC using a gradient of 5 to 60% solvent B over 30 min. The eluted sample was directly dialyzed into reaction buffer, concentrated to a final concentration of 100–150 μM by spin diafiltration (10 kDa MWCO) and stored at 4 °C until use. The purified variants were characterized by analytical RP–HPLC and ESI–MS (results shown below).

---

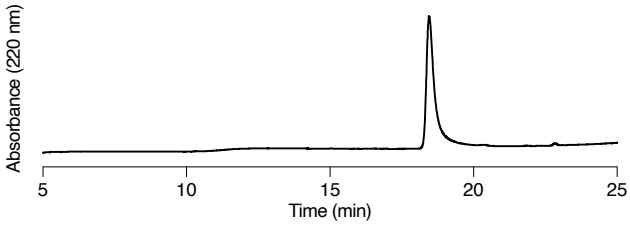
[28] Huang, C.; Ren, G.; Zhou, H.; Wang, C. C. A new method for purification of recombinant human alpha-synuclein in *Escherichia coli*. *Protein. Expr. Purif.* **2005**, *42*, 173–177.



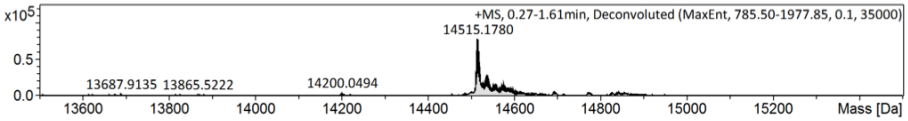
Analytical RP–HPLC of α-synuclein (wt). 10 to 95% solvent B over 20 min.



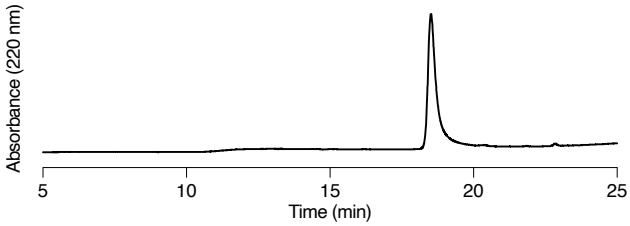
Deconvoluted ESI–MS of α-synuclein (wt). Calc. 14,459.2 Da, obs. 14,459.2 Da.



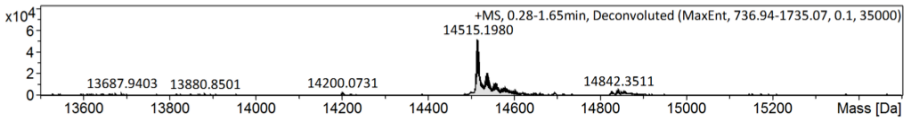
Analytical RP–HPLC of α-synuclein-K96R-K102R. 10 to 95% solvent B over 20 min.



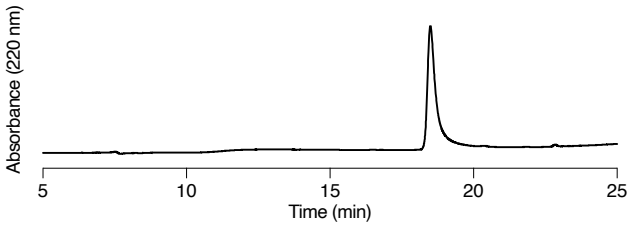
Deconvoluted ESI–MS of α-synuclein- K96R-K102R. Calc. 14,515.3 Da, obs. 14,515.2 Da.



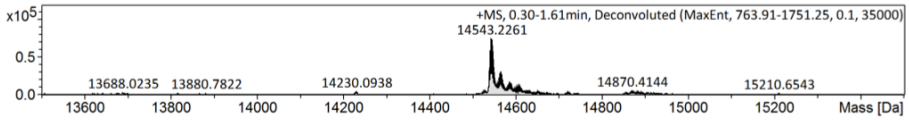
Analytical RP–HPLC of α-synuclein- V95I-D98E-K102R (K96*). 10 to 95% solvent B over 20 min.



Deconvoluted ESI–MS of α-synuclein-V95I-D98E-K102R (K96*). Calc. 14,515.2 Da, obs. 14,515.2 Da.



Analytical RP–HPLC of α-synuclein-K96R-G101I (K102*). 10 to 95% solvent B over 20 min.



Deconvoluted ESI–MS of α-synuclein-K96R-G101I (K102*). Calc. 14,543.3 Da, obs. 14,543.2 Da.

4.10. TNFα-LACE$_N^{3M}$

TNFα-LACE construct

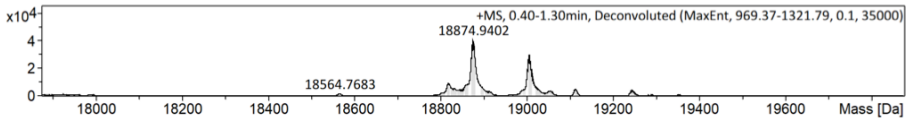
A gene encoding for wild type TNFα (UniProt P01375) with an N-terminal LACE tag^{3M} (MGPRAVI**K**QESAEGGSG) (TNFα-LACE$_N^{3M}$) was subcloned using *Nco*I (5') and *Xho*I (3'). Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

TNFα-LACE$_N^{3M}$ expression

Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmid and a single colony was used to inoculate an overnight preculture in selective LB Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 1 mM. Expressions were carried out at 30 °C for 4 h. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C), resuspended in 50 mL lysis buffer per liter cell culture (20 mM Tris-HCl pH 8.0, 2 mM EDTA) and stored at –80 °C until purification. Purification of the TNFα variant was performed similarly to previous reports.[29] Briefly, cell suspensions were thawed at rt and placed on ice, supplemented with DNase I (0.1 mg/mL), PMSF (1 mM) and protease inhibitor cocktail (Roche), and nutated at 4 °C for 1 h. Cells were lysed by sonication and the suspensions were cleared by centrifugation (16,000 x g, 30 min, 4 °C). To the supernatant was added $(NH_4)_2SO_4$ to a final concentration of 35%, and the sample was stirred at 4 °C for 1 h. The suspension was centrifuged (16,000 x g, 30 min, 4 °C), and to the supernatant was added again $(NH_4)_2SO_4$ to a final concentration of 65%. The suspension was stirred at 4 °C overnight. The suspension was centrifuged (16,000 x g, 30 min, 4 °C) and the

[29] Corti, A.; Ghezzi, P.; Curnis, F. Production and characterization of recombinant human and murine TNF. In *Tumor Necrosis Factor, Methods and Protocols*; Corti, A., Ghezzi, P., Ed.; Methods in Molecular Medicine; Humana Press Inc.: Totowa, NJ, 2004; Vol. 98, pp 9–22.

obtained pellet was resuspended in 15 mL H_2O per liter cell culture by pipetting. To the resuspended pellet was slowly added an equal volume of ice cold 2x TA buffer (200 mM Tris-HCl pH 8.0, 2M $(NH_4)_2SO_4$, 10% methanol), the sample was placed on ice for 10 min and then centrifuged (16,000 x g, 30 min, 4 °C). The obtained supernatant was filtered (0.45 μm) and purified at 4 °C on a HiPrep Phenyl HP HIC column (GE Healthcare) using a gradient of TA buffer (200 mM Tris-HCl pH 8.0, 2M $(NH_4)_2SO_4$, 10% methanol) and TB buffer (100 mM Tris-HCl pH 8.0, 70% ethylene glycol, 5% methanol). Fractions containing the protein as judged by SDS–PAGE were pooled and dialyzed four times against 24 volumes of 20 mM Tris-HCl pH 8.0, 2 mM EDTA at 4 °C. The dialyzed sample was purified by anion exchange chromatography (Mono Q 5/50 GL) using a gradient of buffer A (20 mM Tris-HCl pH 8.5) and buffer B (buffer A with 1 M NaCl). Fractions containing the protein as judged by SDS–PAGE were pooled, concentrated by spin diafiltration (10 kDa MWCO) and purified by size-exclusion chromatography (Superdex 75) equilibrated in reaction buffer. Fractions containing the protein as judged by SDS–PAGE were concentrated by spin diafiltration (10 kDa MWCO) and stored at 4 °C until use. Yields were found to be approximately 2 mg per liter cell culture. An aliquot was desalted to 0.1% formic acid using a desalting column and analyzed by ESI–MS (results shown below). TNFα-$LACE_N^{3M}$ was obtained as a mixture with and without Met1 processing. The intramolecular disulfide bond was formed as judged by ESI–MS.



Deconvoluted ESI–MS of TNFα-$LACE_N^{3M}$. Without Met1 (processed): Calc. 18,875.4 Da, obs. 18,874.9Da. With Met1: Calc. 19,006.6 Da, obs. 19,005.9 Da.

4.11. Recombinant protein thioesters

Constructs for recombinant protein thioesters

Genes encoding for mature ubiquitin with an additional Gly at the N-terminus for cloning purposes, and ISG15 (UniProt P05161) with the mutation C78A, were each prepared as a C-terminal MXe GyrA-His_6 intein fusion (*Mycobacterium xenopi* DNA gyrase subunit A)[30] and were subcloned using *Nco*I (5') and *Xho*I (3'). Genes encoding the affibody clone ZTNF-α-185[31] with a C-terminal flexible extension (GGGGSGGGGSGPQAIAGQGGGGSGGGGS**LRLRGG**) containing an MMP1

30 Chatterjee, C.; McGinty, R. K.; Pellois, J.-P.; Muir, T. W. Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chem. Int. Ed.* **2007**, *119*, 2872–2876.

31 Jonsson, A.; Wållberg, H.; Herne, N.; Ståhl, S.; Frejd, F. Y. Generation of tumour-necrosis-factor-α-specific affibody molecules capable of blocking receptor binding *in vitro*. *Biotechnol. Appl. Bioc.* **2009**, *54*, 93–103.

cleavage site (underlined)[32] and a ubiquitin-derived C-terminal hexapeptide (bold), and the affibody clone ZHER2:342[33] with a C-terminal ubiquitin-derived hexapeptide (LRLRGG) were each also prepared as Mxe GyrA-His_6 intein fusions, analogously to the ubiquitin and ISG15 constructs. Full amino acid sequences of the constructs are given in Table 5 of the Appendix.

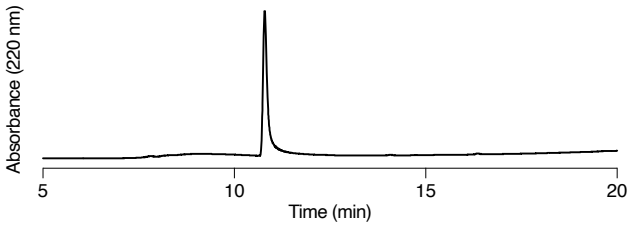
Expression of ubiquitin-Mes thioester 41

Chemically competent BL21 (DE3) cells were heat-shock transformed with the plasmid and a single colony was used to inoculate an overnight preculture in selective LB Miller medium. Following 1:100 dilution with fresh selective LB Miller medium, cultures were grown in baffled shake flasks at 37 °C until an OD_{600} of approximately 0.6 was reached. Protein expression was induced by addition of IPTG at a final concentration of 0.5 mM. Expressions were carried out overnight at 25 °C. Cells were collected by centrifugation (4,500 x g, 15 min, 4 °C), resuspended in 20 mL lysis buffer per liter cell culture (50 mM Tris-HCl pH 7.4, 200 mM NaCl) and stored at –80 °C until purification. Ubiquitin-GyrA-His_6 was purified and converted to Ubiquitin-Mes thioester similarly to a previous report.[30] Lysis was performed as described for Ubc9 variants. The cleared lysate was subjected to gravity-flow Ni-NTA affinity purification at 4 °C using binding buffer (50 mM Tris-HCl pH 7.4, 200 mM NaCl) and elution buffer (50 mM Tris-HCl pH 7.2, 200 mM NaCl, 300 mM imidazole). The sample was diluted to 100 μM concentration with 50 mM Tris-HCl pH 7.4, 200 mM NaCl, 1 mM EDTA and dialyzed against the same buffer at 4 °C. The sample was supplemented with 0.1 M final MesNa and 1 mM final TCEP to initiate intein thiolysis. The sample was agitated at rt for 3 days and fresh MesNa was added after 1 and 2 days (30 mM final concentration each). About 30% conversion to the desired thioester was achieved as judged by analytical RP–HPLC and SDS–PAGE. The sample was dialyzed against 50 mM Tris-HCl pH 7.4, 200 mM NaCl at 4 °C, and subjected to reverse gravity-flow Ni-NTA purification to remove uncleaved starting material and spliced GyrA-His_6. The flow-through containing Ub-Mes was dialyzed against Milli-Q water with 1% AcOH, lyophilized and purified by preparative RP–HPLC using a gradient of 20 to 50% solvent B over 30 min. Fractions containing pure ubiquitin-Mes were pooled and lyophilized. Yields were found to be approx. 5 mg per liter cell culture. The purified product was characterized by analytical RP–HPLC and ESI–MS (results shown below). Met1 was found to be processed.

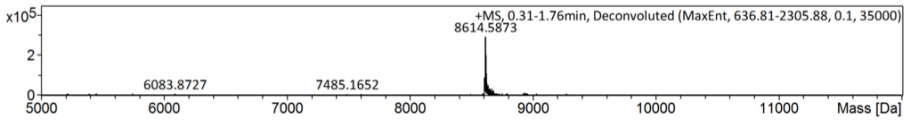
---

32 Sandersjöö, L.; Jonsson, A.; Löfblom, J. A new prodrug form of affibody molecules (pro-affibody) is selectively activated by cancer-associated proteases. *Cell Mol. Life Sci.* **2015**, *72*, 1405–1415.

33 Orlova, A.; Magnusson, M.; Eriksson, T. L. J.; Nilsson, M.; Larsson, B.; Höidén-Guthenberg, I.; Widström, C.; Carlsson, J.; Tolmachev, V.; Ståhl, S.; Nilsson, F. Y. Tumor imaging using a picomolar affinity HER2 binding affibody molecule. *Cancer Res.* **2006**, *66*, 4339–4348.



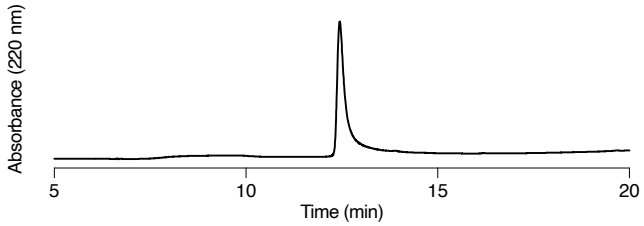
Analytical RP–HPLC of ubiquitin-MesH thioester 41. 40 to 95% solvent B over 14 min.



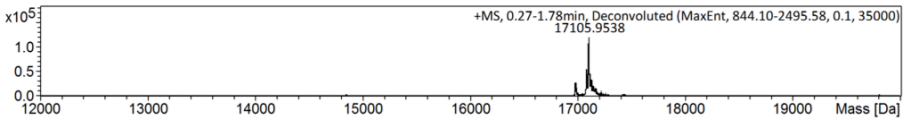
Deconvoluted ESI–MS of ubiquitin-MesH thioester 41. Calc. 8614.7 Da, obs. 8614.6 Da.

Expression of ISG15-Mes thioester 42

Expression, lysis and conversion of ISG15-GyrA-His_6 to the thioester was performed as described for **41** (Ub). After thiolysis and reverse gravity-flow Ni-NTA purification, ISG15-Mes **42** was found to be pure by SDS–PAGE (>95%). ISG15-Mes was insoluble in buffer following purification by preparative RP–HPLC and lyophilization, likely due to disruption of the tertiary structure. After reverse Ni-NTA purification, the sample was instead directly exchanged to reaction buffer and concentrated to approximately 450 μM. The sample was portioned into aliquots, flash-frozen in liquid N_2 and stored at –20 °C until use. Yields were found to be approx. 15 mg per liter cell culture. The purified product was characterized by analytical RP–HPLC and ESI–MS (results shown below). Met1 was found to be processed.



Analytical RP–HPLC of ISG15-MesH thioester 42. 40 to 95% solvent B over 14 min.

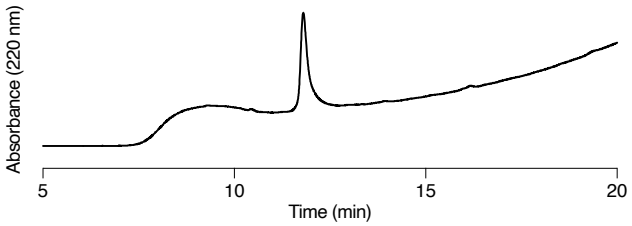


Deconvoluted ESI–MS of ISG15-MesH thioester 42. Calc. 17,106.7 Da, obs. 17,106.0 Da.

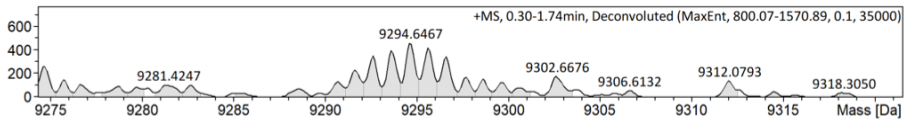
Expression of affibody thioester Z_{TNF}-Mes 43

Expression, lysis and conversion of Z_{TNF}-GyrA-His_6 to the thioester was performed as described for **42** (ISG15), with the difference that, following thiolysis, the sample was dialyzed against 20 mM HEPES pH 7.2, 200 mM NaCl at 4 °C. The pH of the dialyzed sample was adjusted to pH 7.4 and the sample was subjected to reverse Ni-NTA purification as described for **42** (ISG15). After thiolysis and reverse Ni-NTA purification, Z_{TNF}-Mes **43** was found to be pure by SDS–PAGE

(>95%). The sample was concentrated to approximately 400 μM, portioned into aliquots, flash-frozen in liquid N_2 and stored at –80 °C until use. Yields were found to be approx. 20 mg per liter cell culture. The purified product was characterized by analytical RP–HPLC and ESI–MS (results shown below). Met1 was found to be processed.



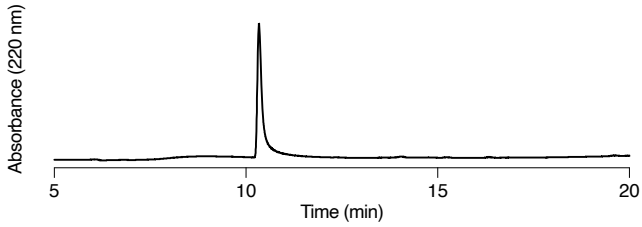
Analytical RP–HPLC of Z_{TNF}-MesH thioester 43. 40 to 95% solvent B over 14 min.



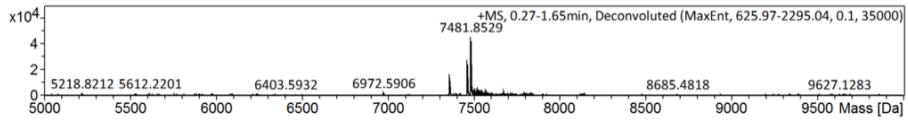
Deconvoluted ESI–MS of Z_{TNF}-MesH thioester 43. Calc. 9295.1 Da, obs. 9294.6 Da.

Expression of affibody thioester Z_{HER2}-Mes 46

Expression, lysis and conversion of Z_{HER2}-GyrA-His_6 to the thioester was performed as described for **42** (ISG15), with the difference that, following thiolysis, the sample was dialyzed against 20 mM potassium phosphates pH 7.2, 200 mM NaCl at 4 °C. The pH of the dialyzed sample was adjusted to pH 7.4 and the sample was subjected to reverse Ni-NTA purification as described for **42** (ISG15). After thiolysis and reverse Ni-NTA purification, Z_{HER2}-Mes **44** was found to be pure by SDS–PAGE (>95%). The sample was concentrated to approximately 400 μM, portioned into aliquots, flash-frozen in liquid N_2 and stored at –80 °C until use. Yields were found to be approx. 20 mg per liter cell culture. The purified product was characterized by analytical RP–HPLC and ESI–MS (results shown below). Met1 was found to be processed.



Analytical RP–HPLC of Z_{HER2}-MesH thioester 46. 40 to 95% solvent B over 14 min.



Deconvoluted ESI–MS of Z_{HER2}-MesH thioester 46. Calc. 7482.3 Da, obs. 7481.9 Da.

5. Experimental part for Chapter 2: Lysine acylation using conjugating enzymes (LACE)

5.1. General LACE method

Unless otherwise mentioned, reactions were assembled on ice in polypropylene microcentrifuge tubes. Final assay volumes were 30–100 μL. The exact volumes used are indicated for preparative reactions. Reactions were incubated in a water bath at the indicated temperature, the standard temperature for LACE being 30 °C. Unless otherwise mentioned, reactions were carried out in 50 mM HEPES pH 7.6, 50 mM KCl (reaction buffer). Indicated thiol buffer additives were added from a 50 mM stock solution in water. For assembly of the reactions, stock solutions of lyophilized synthetic peptide thioesters were prepared at 1 mM concentration in reaction buffer as determined by dry weight, or, if applicable, by rhodamine fluorescence (ε_{567} 93,000 $M^{-1}cm^{-1}$). Thioester stock solutions were stored for up to one month at –20 °C.

5.2. Canonical SUMOylation

SUMOylation of RanGAP1 with E1, Ubc9 and ATP (or without ATP for the control reaction) was carried out using a commercial SUMOylation kit according to the instructions by the manufacturer (Enzo Life Sciences). A synthetic rhodamine-labeled full length SUMO3[34] (2.5 μM final concentration) was used in place of the provided SUMO from the kit. Reactions were quenched and analyzed by reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 25c.

5.3. Identification of a peptide thioester as acyl donor for Ubc9

Leaving group screening by RanGAP1 labeling with thioesters 2a–2c

Commercially available RanGAP1 (1 μM final concentration) (residues 418–587 of RanGAP1 with an N-terminal GST tag, Enzo Life Sciences) was incubated with 10 μM His_6-Ubc9 and 100 μM thioesters **2a**–**2c** in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) for 1 h at 37 °C. Reactions were quenched and analyzed by reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 27b.

---

[34] Kindly provided by Dr. S. Shimura (Bode Group, ETH Zürich). See also: (a) Wucherpfennig, T. G.; Pattabiraman, V. R.; Limberg, F. R. P.; Ruiz-Rodríguez, J.; Bode, J. W. Traceless preparation of C-terminal α-ketoacids for chemical protein synthesis by α-ketoacid–hydroxylamine ligation: Synthesis of SUMO2/3. *Angew. Chem. Int. Ed.* **2014**, *53*, 12248–12252. (b) Wucherpfennig, T. G. Chemical protein synthesis by α-ketoacid-hydroxylamine ligation. Doctoral dissertation, ETH Zürich, Switzerland, 2016.

pH screening by RanGAP1 labeling with thioester 2a

Commercially available RanGAP1 (1 μM final concentration) (Enzo Life Sciences) was incubated with 0 or 10 μM His_6-Ubc9 and 100 μM thioesters **2a** in buffer with 1 mM Ac-Cys-NHMe (**1**) for 1 h at 37 °C. His_6-Ubc9 and thioester **2a** were added from stock solutions in standard reaction buffer, and the pH was adjusted using a 10x reaction buffer (500 mM HEPES, 500 mM KCl, 50 mM $MgCl_2$) with either pH 6.8, 7.6 or 8.2. Reactions were quenched and analyzed by reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 27c.

Thioester sequence optimization by RanGAP1 labeling with peptide thioesters 2a–13

Commercially available RanGAP1 (2 μM final concentration) (Enzo Life Sciences) was incubated with 12 μM His_6-Ubc9 and 50 μM thioester **2a**–**13** in reaction buffer wit 1 mM Ac-Cys-NHMe (**1**) for 1 h at 30 °C. Reactions were quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 28b.

MS/MS analysis of RanGAP1 conjugate with thioester 5

The Coomassie-stained SDS–PAGE band corresponding to RanGAP1–**5** was excised, cut in small pieces and washed with 100 mM NH_4HCO_3 / 50% MeCN (2 x 100 μL) and MeCN (1 x 50 μL). Trypsin (1 ng/μL final concentration) in 10 mM Tris pH 8.2, 2 mM $CaCl_2$ (50 μL) was added to the gel. The sample was microwaved at 60 °C for 34 min. The supernatant was removed and the gel extracted with 0.1% TFA / 50% MeCN (150 μL) by ultrasonication for 15 min. The solution was dried, dissolved in 0.1% formic acid (20 μL) and analyzed by LC–MS/MS on a nanoAcquity UPLC coupled to a Q-Exactive mass spectrometer (Thermo Fisher). Data was analyzed using the PEAKS engine.[35] by searching against the expected sequence with the variable modification oxidation (M) and GlyGly (K). Results of the MS/MS analysis are shown in Figure 29.

5.4. Identification of a LACE tag

LACE tag screening by labeling GFP-RanGAP1, -IκBα and -PML with thioester 5

GFP-RanGAP1, GFP- IκBα or GFP-PML (GFP-$LACE_C$) (15 μM) was incubated with 0 or 60 μM His_6-Ubc9 or His_6-Ubc9-C93A and 150 μM thioester **5** in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) for 1 h at 30 °C. Reactions were quenched and analyzed by Coomassie-stained reducing

[35] Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G.A.; Ma, B. PEAKS DB: *De novo* sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **2012**, *11*, M111.010587.

SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 30c,d.

LACE tag screening by labeling GFP-PML, -PML/Elk1 and -Elk1 with thioester 5

GFP-PML (GFP-$LACE_C$), GFP-PML/Elk1, GFP-Elk1 or commercially available RanGAP1 (1 μM) was incubated with 0 or 50 μM His_6-Ubc9 and 75 μM thioester **5** in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) for 1 h at 30 °C. Reactions were quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 31c.

Specificity determination of LACE tag by alanine scanning

Variants of GFP-$LACE_C$ with single point mutations (P1A, R2A, K3A, V4A, I5A, K6R, M7A, E8A, S9A, E10A or E11A) (15 μM) were incubated with 0 or 60 μM His_6-Ubc9 and 150 μM thioester **5** in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) for 1 h at 30 °C. Reactions were quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 34d.

5.5. Identification of useful reaction conditions

Ubc9 stability determination by preincubation experiment

His_6-Ubc9-K14R (60 μM) was incubated in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) for 12 h at 30 °C. To the sample was added fresh GFP-$LACE_I$ and thioester **5** to a final concentration of 15 and 150 μM, respectively, and the reactions were incubated at 30 °C. SDS–PAGE samples were taken from the reaction mixture after 0, 0.5, 1, 2, 3 and 8 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the quantification are shown in Figure 32a.

Ubc9 denaturation curve with GndHCl

Freshly thawed aliquots of His_6-Ubc9 (45 μL, 220 μM) were diluted to a final concentration of 20 μM with 455 μL of cold buffer (0.2 M sodium phosphates pH 7.0, 10 mM TCEP) containing either 0, 1.0, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0 or 6.0 M GndHCl. Samples were incubated at 4 °C for 4 h and centrifuged (10,000 x g, 10 min, 4 °C). The supernatants were transferred to fresh tubes, and the protein concentration was determined by A_{280}. Fluorescence was measured at rt using a spectrofluorometer equipped with an 814 Photomultiplier Detection System (both Photon Technology International) in 400 μL cuvettes (10 x 2 mm) at an excitation wavelength of 280 nm and the emission was scanned between 300 and 370 nm at a speed of 1 nm/s. Tryptophan fluorescence was evaluated at 330 nm where the largest intensity change was observed. The signal was normalized by the protein concentration. Correction for background signal was

performed with analogous samples in which His_6-Ubc9 was substituted with reaction buffer. The results are shown in Figure 32b.

Ubc9 refolding and activity test

Refolding: A freshly thawed aliquot of His_6-Ubc9 (50 μL, 220 μM) was diluted to a final concentration of 22 μM with 450 μL of cold unfolding buffer (0.2 M sodium phosphates pH 7.0, 6.0 M GndHCl, 10 mM TCEP) to achieve a final concentration of 5.4 M GndHCl. The sample was incubated at 4 °C for 5 h. The sample was dialyzed against fresh refolding buffer (50 mM HEPES pH 7.6, 50 mM KCl, 5 mM $MgCl_2$ and 10 mM TCEP) for 2 h, overnight and again for 2 h. No obvious precipitation was observed. The sample was exchanged to reaction buffer with 1 mM Ac-Cys-NHMe (**1**) by spin diafiltration (10 kDa MWCO). The protein concentration was determined by A_{280}, based on which a 35% recovery of remaining protein was observed compared to the initial amount.

Activity test: Commercially available RanGAP1 (1 μM final concentration) was incubated with a fresh aliquot of His_6-Ubc9 or the refolded His_6-Ubc9 (10 μM) and thioester **2a** (100 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 37 °C. SDS–PAGE samples were taken from the reaction and quenched after 2, 4 and 8 h according to the general procedure, for comparison of the activity at three time points. The time point samples were analyzed by reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the quantification are shown in Figure 32c.

Reaction concentration optimization

GFP-$LACE_C$ (15 μM or 50 μM) was incubated with His_6-Ubc9-K14R (0, 7.5, 15, 30, 50 or 60 μM) and thioester **5** (150 or 250 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 0, 1, 3, 8 and 24 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence. Results of the SDS–PAGE analyses are shown in Figure 33.

5.6. Tag position screening in the GFP scaffold, dual labeling, and minimal LACE tags

Labeling of GFP-$LACE_C$, -$LACE_I$ and -$LACE_N$ with thioester 5

GFP-$LACE_C$, GFP-$LACE_I$ or GFP-$LACE_N$ or the respective K6R acceptor tag mutants (15 μM) was incubated with His_6-Ubc9 (60 μM) and thioester **5** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 0, 0.5, 1, 2, 4, 6 and 8 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 34b,c and Figure 35.

Dual-labeling of GFP-LACE$_{I,C}^{2M}$ with thioester 5

GFP-LACE$_{I,C}^{2M}$ (15 μM) was incubated with His$_6$-Ubc9 (0 or 60 μM) and thioester **5** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 0, 0.5, 1, 2, 4, 6 and 8 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 36b,c.

Labeling of GFP variants with a minimal internal or C-terminal LACE tag with thioester 5

GFP variants with a minimal LACE tag (IKQE) positioned internally (GFP-LACE$_I^{minimal}$) or at the C-terminus (GFP-LACE$_C^{minimal}$), or the corresponding acceptor tag mutants (IRQE) (15 μM) was incubated with His$_6$-Ubc9 (0 or 60 μM) and thioester **5** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 0, 1, 3, 8 and 27 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 37.

5.7. Labeling of titin IG27 and T4L with thioester 5

Titin-LACE$_C^{2M}$ labeling with thioester 5

Titin-LACE$_C^{2M}$ or the corresponding acceptor tag mutant K6R (15 μM) was incubated with His$_6$-Ubc9 (60 μM) and thioester **5** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 1, 3, 8 and 24 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 38a.

T4L-LACE$_C$ labeling with thioester 5

T4L-LACE$_C$ (15 μM) was incubated with His$_6$-Ubc9 (0 or 60 μM) and thioester **5** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. Samples (20 μL) were taken from the reaction mixtures after 4, 8 and 24 h, treated with AET-HCl (1 mM final concentration) for 5 min and mixed with an equal volume of Ni-NTA agarose resin (1:1 slurry in reaction buffer). The suspensions were agitated for 5 min at rt and then passed through a fritted spin column (Bio-Rad) to remove resin-bound His$_6$-Ubc9 from the solution. The eluted samples were then quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 38b.

5.8. Mechanistic studies

5.8.1. Self-labeling of Ubc9

MS/MS analysis of self-labeled Ubc9

Sample preparation: His_6-Ubc9 (60 μM) was incubated with thioester **5** (150 μM) in reaction buffer (final volume 213 μL) with 1 mM Ac-Cys-NHMe (**1**) at 30 °C for 4.5 h. The sample was diluted to 500 μL with reaction buffer and excess thioester was removed using desalting column. A final sample of 500 μL volume with approximately 0.4 μg/μL protein as determined by A_{280} was obtained.

MS/MS analysis: The sample (30 μL) was treated with 0.9 μL of 100 mM TCEP and 1.4 μL of 500 mM iodoacetamide, and incubated at 700 rpm and 60 °C in the dark for 30 min. Trypsin (5 μL of a 100ng/μL stock solution in 10 mM HCl) was added to the sample and microwaved at 60 °C for 30 min. The sample was dried, dissolved in 0.1% formic acid (20 μL) and analyzed by LC–MS/MS on a nanoAcquity UPLC coupled to a Q-Exactive mass spectrometer (Thermo Fisher). Data was analyzed using the PEAKS engine[35] by searching against the expected sequence with the fixed modification Cam (C) and the variable modification oxidation (M) and GlyGly (KC). Results of the MS/MS analysis are shown in Figure 39 and Figure 101 of the Appendix.

Self-labeling of Ubc9-C138A and Ubc8-K14R

His_6-Ubc9 (wt), His_6-Ubc9-C138A or His_6-Ubc9-K14R (25 μM) was incubated with thioester **5** (50 μM) and GFP-$LACE_C$ (2 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C for 3 h. Reactions were quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 40b,c.

5.8.2. Influence of buffer additives

Buffer additive screen

Reactions were assembled by pre-dispensing stock solutions of the indicated additives (stock solutions were prepared in water, max. 10% v/v of the final reaction volume) and the volumes were equalized with reaction buffer. Dilution of the pre-dispensed samples with an equal volume of premixed concentrated solution (freshly prepared on ice) of GFP-$LACE_C$, His_6-Ubc9 and thioester **5** in reaction buffer to reach final concentrations of 15 μM GFP-$LACE_C$, 60 μM His_6-Ubc9 and 150 μM thioester **5** marked the start of the reactions. The samples were incubated at 25 °C for 1 h. The reactions were quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 41.

MPAA catalysis

GFP-LACE$_C$ or the respective K6R acceptor tag mutant (15 μM) was incubated with His$_6$-Ubc9 (60 μM) and thioester **5** (150 μM) in reaction buffer with 0, 1 or 10 mM MPAA at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 0, 1 and 3 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 43a.

5.8.3. Reaction profile studies

Thioester formation assay

His$_6$-Ubc9-C138A (60 μM) was incubated with thioester **5** (Ub-derived) or thioester **2a** (SUMO2/3-derived) (150 μM) in reaction buffer at 30 °C. After 1, 15 and 45 min, 40 μL of the reaction mixture was desalted to 50 mM citrate buffer pH 5.5, 50 mM NaCl, 5% glycerol to remove unreacted peptide thioester using pre-equilibrated Zeba Spin columns (7K MWCO, 0.5 mL, Thermo Fisher Scientific). The desalted samples were kept on ice and collectively treated with an equal volume of 2x non-reducing SDS–PAGE buffer at the end of the assay. Samples were not boiled and directly analyzed by non-reducing SDS–PAGE and in-gel fluorescence. Quantification was done according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 44a.

Preincubation experiment

His$_6$-Ubc9-C138A (60 μM) was incubated in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) in the presence or absence of thioester **5** (150 μM) at 30 °C. The reactions were placed on ice after 1 h of preincubation. To the samples were added the remaining reaction components – GFP-LACE$_C$ (15 μM final concentration) and thioester (150 μM final concentration) – or just GFP-LACE$_C$ for the sample that was already preincubated with thioester. Samples were placed at rt to initiate the GFP-LACE$_C$ labeling reactions. SDS–PAGE samples were taken from the reaction mixtures after 0, 5, 15, 30 and 45 min and quenched according to the general procedure. The time point samples were analyzed by reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 44b.

Single discharge reaction

His$_6$-Ubc9-C138A (60 μM), and thioester **5** (ubiquitin-derived) or thioester **2a** (SUMO2/3-derived) (150 μM) were incubated in reaction buffer at 30 °C for 1 h. The sample (100 μL) was desalted to remove unreacted peptide thioester by passing the sample two times through Zeba Spin columns pre-equilibrated with reaction buffer (7K MWCO, 0.5 mL, Thermo Fisher Scientific). The amount of Ubc9~peptide thioester intermediate was quantified by UV-Vis based on the absorbance of rhodamine (A_{567}). The concentrations of Ubc9~peptide thioester intermediate were equalized by diluting the more concentrated sample with a solution of 60 μM His$_6$-Ubc9-C138A in reaction

buffer. The equalized samples of pre-formed Ubc9~peptide thioester were treated at rt with an excess of GFP-LACE$_C$ (50 μM final concentration). SDS–PAGE samples were taken from the reaction mixtures after 0, 5, 15 and 45 min and quenched by the addition of an equal volume of 2x non-reducing SDS–PAGE buffer. The time point samples were kept on ice and collectively analyzed by non-reducing SDS–PAGE at the end of the assay. Results of the SDS–PAGE analysis are shown in Figure 44c.

5.8.4. Reactivity comparison of ubiquitin and SUMO2/3 sequence with reported peptides

GFP-LACE$_C$ (15 μM) was incubated with His$_6$-Ubc9 (0 or 60 μM) and thioesters **5**, **2a**, and **14**–**18** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C for 45 min. Reactions were quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 45.

5.9. Structural analysis by protein X-ray crystallography

5.9.1. Disulfide-bonded complex Ubc9-3A-C138A–21

Ubc9-K48A-K49A-E54A-C138A (Ubc9-3A-C138A) (2 mL, 107.8 μM, 216 nmol, 1 equiv) was thawed on ice. To remove TCEP, the buffer was exchanged to 50 mM HEPES pH 8.0, 100 mM NaCl by passage through a desalting column. The protein solution was added to solid isopeptide ligand **21** (10 mg, 7.7 μmol, 35 equiv) and incubated at 30 °C for 48 h. The reaction was exchanged to 20 mM potassium phosphates pH 7.0 by passage through a desalting column, and the desired conjugate was separated from unreacted Ubc9-3A-C138A by cation exchange chromatography (Mono S 5/50 GL) with buffer A (20 mM potassium phosphates pH 7.0) and a gradient of buffer B (buffer A with 1 M NaCl) (95% conversion to the conjugate based on peak area). Pure fractions were pooled, exchanged to 25 mM Tris pH 7.2, 150 mM NaCl and concentrated to a final concentration of 15.7 mg/mL (886 μM) by spin diafiltration (10 kDa MWCO). The sample was flash-frozen in liquid N_2 and stored at –80 °C until crystallization. Purification of the complex by cation exchange chromatography and ESI–MS are shown in Figure 49.

5.9.2. Crystallization of Ubc9-3A-C138A–21 and structure determination

The disulfide-bonded complex Ubc9-3A-C138A-**21** was crystallized at 20 °C using the sitting drop vapor diffusion method. Diffraction-quality crystals appeared within two days after mixing 0.1 μL protein solution (15.7 mg/mL, in 25 mM Tris pH 7.2, 150 mM NaCl) with 0.1 μL reservoir solution (0.1 M trisodium citrate pH 5.0, 30% w/v PEG8000). Crystals were cryoprotected by soaking in 1 μL of well solution supplemented with 25% v/v PEG400 and flash-frozen in liquid N_2. X-ray diffraction data was collected at the Swiss Light Source (PSI, Viligen, Switzerland) beamline PXIII equipped with the PILATUS 2M-F detector system (Dectris, Switzerland) at a wavelength of 1.0 Å,

while the crystal was kept at 100 K. The XDS software package was used for data processing.[36] Space group assignment was re-evaluated after analysis of the observed translational non-crystallographic symmetry. Structure determination was performed by molecular replacement using Phaser[37] and a previously published structure of human Ubc9 (PDB entry 5f6e)[17] as search model (4 molecules per asymmetric unit in $P2_1$). The structural model was built and refined iteratively using Coot,[38] Refmac5 (CCP4 suite),[39] Phenix,[40] and Buster. Protein structure visualization and analysis was performed using the PyMOL Molecular Graphics System (Version 1.7.4.5 Schrödinger LLC) and the USCF Chimera package.[41] Data collection and refinement statistics are shown in Table 1.

5.10. Structure validation

5.10.1. Reactivity comparison of Ub- and inverse-Ub-derived sequences

Ubc9~peptide thioester formation

His_6-Ubc9-C138A or His_6-Ubc9-C138A-C93A (60 μM) was incubated with thioester **5** (Ub-derived) or thioester **22** (inverse Ub-derived) (150 μM) in reaction buffer at 30 °C. After 45 min, 30 μL of the reaction mixture was desalted to 50 mM citrate buffer pH 5.5, 50 mM NaCl, 5% glycerol to remove unreacted peptide thioester using pre-equilibrated Zeba Spin columns (7K MWCO, 0.5 mL, Thermo Fisher Scientific). The desalted samples were treated with an equal volume of 2x non-reducing SDS–PAGE buffer. Samples were not boiled and directly analyzed by non-reducing SDS–PAGE and in-gel fluorescence. Quantification was done according to the general procedure.

---

36 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 133–144.

37 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Cryst.* **2007**, *40*, 658–674.

38 Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst.* **2004**, *D60*, 2126–2132.

39 Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst.* **2011**, *D67*, 355–367.

40 Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Cryst.* **2019**, *D75*, 861–877.

41 Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

Full GFP-LACE$_C$ labeling reaction

GFP-LACE$_C$ (15 μM) was incubated with either His$_6$-Ubc9-C138A or His$_6$-Ubc9-C138A-C93A (60 μM) and with either thioester **5** (Ub-derived) or thioester **22** (inverse Ub-derived) (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C for 1 h. Reactions were quenched and analyzed by reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure.

Results of the SDS–PAGE analyses of Ubc9~peptide thioester formation and of full GFP-LACE$_C$ labeling are shown in Figure 53c.

5.10.2. Ubc9 mutagenesis analysis

Ubc9~peptide thioester formation

His$_6$-Ubc9-C138A, or His$_6$-Ubc9-C138A bearing the mutation N85A, S89A, T91A, D102A, D102R, E118A, E118R, E122A, E122R or D127A (60 μM) was incubated with thioester **5** (Ub-derived) or thioester **2a** (SUMO2/3-derived) (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. After 45 min, 30 μL of the reaction mixture was desalted to 50 mM citrate buffer pH 5.5, 50 mM NaCl, 5% glycerol to remove unreacted peptide thioester using pre-equilibrated Zeba Spin columns (7K MWCO, 0.5 mL, Thermo Fisher Scientific). The desalted samples were treated with an equal volume of 2x non-reducing SDS–PAGE buffer. Samples were not boiled and directly analyzed by non-reducing SDS–PAGE and in-gel fluorescence. Quantification was done according to the general procedure.

Full GFP-LACE$_C$ labeling reaction

GFP-LACE$_C$ (15 μM) was incubated with His$_6$-Ubc9-C138A or His$_6$-Ubc9-C138A bearing the mutations N85A, S89A, T91A, D102A, D102R, E118A, E118R, E122A, E122R or D127A (60 μM) in the presence of thioester **5** (Ub-derived) or thioester **2a** (SUMO2/3-derived) (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe **1** at 30 °C. Reactions with thioester **5** were quenched after 1.5 h and reactions with thioester **2a** after 3.0 h. Reactions were quenched and analyzed by reducing SDS–PAGE and in-gel fluorescence and quantified according to the general procedure.

Results of the quantification of Ubc9~peptide thioester formation and of full GFP-LACE$_C$ labeling are shown in Figure 55.

6. Experimental part for Chapter 3: Protein functionalization and conjugation by LACE

6.1. Compatibility with other chemoenzymatic methods

6.1.1. Dual-labeling with SpyTag/SpyCatcher and LACE

Reactivity test of SpyCatcher variants with SpyTagMBP

SpyCatcher (wt) or SpyCatcher with a minimal LACE tag (IKSE) was incubated with 1.0, 1.2 or 2.0 equiv of SpyTagMBP in reaction buffer at rt. A control reaction omitting SpyCatcher was performed. SDS–PAGE samples were taken from the reaction mixture after 1 and 15 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 56b.

Attempt at labeling SpyCatcher directly with thioester 5

SpyCatcher (wt), the SpyCatcher variant with a minimal LACE tag (IKSE) or the corresponding acceptor mutant (IRSE) (15 μM) was incubated with thioester **5** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. SDS–PAGE samples were taken from the reaction mixture after 1, 3, 8 and 27 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 57a.

One-pot dual-labeling of SpyCatcher with SpyTagMBP and thioester 5

SpyCatcher (wt), the SpyCatcher variant with a minimal LACE tag (IKSE) or the corresponding acceptor mutant (IRSE) (30 μM) was incubated in the presence of SpyTagMBP (1.05 equiv) in reaction buffer at rt for 1 h. The samples were then directly diluted with an equal volume of a premixed 2x solution (freshly prepared on ice) of His_6-Ubc9, thioester **5** and Ac-Cys-NHMe (**1**) in reaction buffer to arrive at a final concentration of 15 μM SpyCatcher/SpyTagMBP, 60 μM His_6-Ubc9, 150 μM thioester **5** and 1 mM Ac-Cys-NHMe (**1**). The reactions were incubated at 30 °C and SDS–PAGE samples were taken from the reaction mixture after 0, 1, 3, 6, 8 and 20 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 58b.

6.1.2. Dual-labeling of trastuzumab Fab with sortase and LACE

One-pot dual-labeling of trastuzumab Fab with thioester 5 and sortase-reactive probe 23

Trastuzumab Fab (3 nmol, 0.15 mg, final concentration was 15 μM) was combined with His_6-Ubc9 (60 μM), thioester **5** (150 μM) and sortase-reactive probe **23** (2 mM) in reaction buffer (200 μL).

For reaction assembly, a stock solution of **23** was prepared at 4.5 mM concentration in reaction buffer. The sample was incubated at 30 °C for 6 h before the reaction was placed at 4 °C, $SrtA^{7M}$ (2 μM final concentration) was added, and the reaction was incubated at 4 °C for another 6 h. SDS–PAGE samples were taken from the reaction mixture after 0, 3, 6, 9 and 12 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained SDS–PAGE and in-gel fluorescence according to the general procedure. The reaction mixture was separated by size-exclusion chromatography (Superdex 75) equilibrated with PBS. The peak corresponding to the Fab molecule was collected. Clean and near quantitatively dual-modified trastuzumab Fab–**5**–**23** was obtained, with approximately 30% recovery based on the initial amount as judged by coumarin and sulforhodamine absorbance (38,740 and 93,000 $M^{-1}cm^{-1}$, respectively). Results of the SDS–PAGE analyses are shown in Figure 59c.

Mass spectrometry analysis

Prior to ESI–MS analysis, a sample was treated with DTT (50 mM final concentration) for 1 h at rt to reduce the interchain disulfide bond, desalted using C4 ZipTips (Millipore, USA) and analyzed in MeOH:isopropanol:0.2% formic acid (30:20:50) on a Synapt G2-Si mass spectrometer (Waters, UK). Unreduced samples of intact Fab were desalted and analyzed using the same procedure. The recorded spectra were deconvoluted by applying the maximum entropy algorithm MaxEnt1 (MaxLynx) with an output resolution of 0.5 Da/channel and a Uniform Gaussian Damage Model at the half height of 0.7 Da. The ESI–MS results are shown in Figure 60.

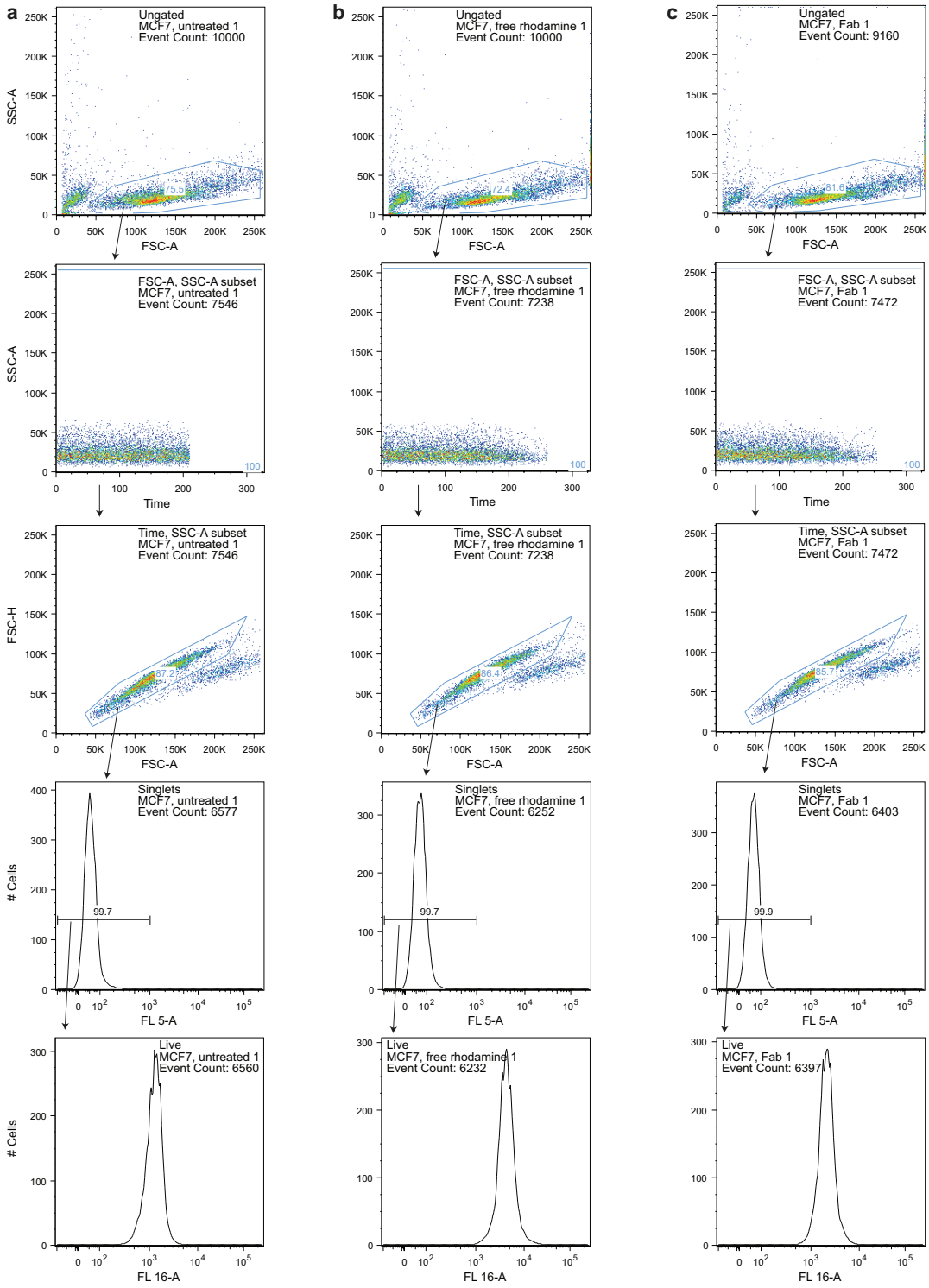
Cell culture

SKBR3 and SKOV3 cells were cultured in McCoy's 5A medium (Sigma M8403) containing penicillin with streptomycin (Sigma P4333). MCF7 cells were cultured in Iscove's modified Dulbecco's medium containing 50 μg/ml gentamycin (Sigma G1397), 2 mM L-glutamine (Sigma G7513) and 1 mM alanyl glutamine (Sigma G8541). All media contained 10% fetal bovine serum (FBS). Cells were seeded at $1x10^6$ cells per T75 flask and allowed to grow for 3–4 days at 37 °C in 5% CO_2 before being passaged. Cells were removed from the flask by incubation with trypsin (Gibco 25200-056) and counted. Trypan Blue staining was used to exclude dead cells.

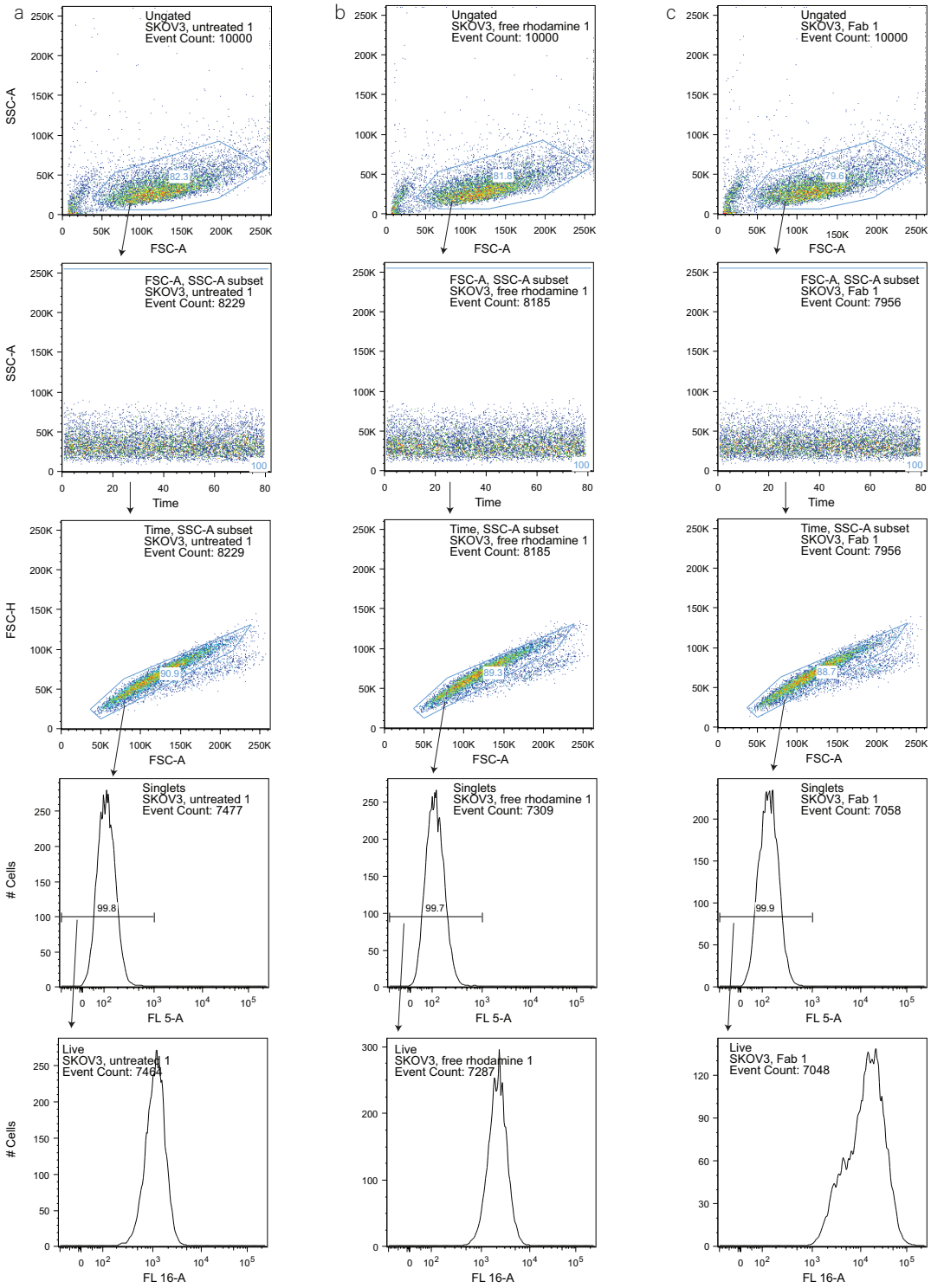
Flow cytometry

Cells were seeded at a density of 50,000 cells per well (MCF7) or 100,000 cells per well (SKBR3 and SKOV3) in a 24-well plate in 500 μL of culture medium and allowed to recover at 37 °C and 5% CO_2 for 24 hours. The culture medium was removed and cells were washed once with 500 μL of fresh medium without FBS. To each well was added 355 μL of PBS (untreated), 70 nM sulforhodamine B (free rhodamine), or 70 nM trastuzumab Fab–**5**–**23** (each in 355 μL PBS). Cells were incubated for 2 h in 5% CO_2 at 37 °C before washing with ice-cold PBS (500 μL). Cells were trypsinized by treating each well with 100 μL of 0.05% trypsin-EDTA (Thermo Fisher Scientific,

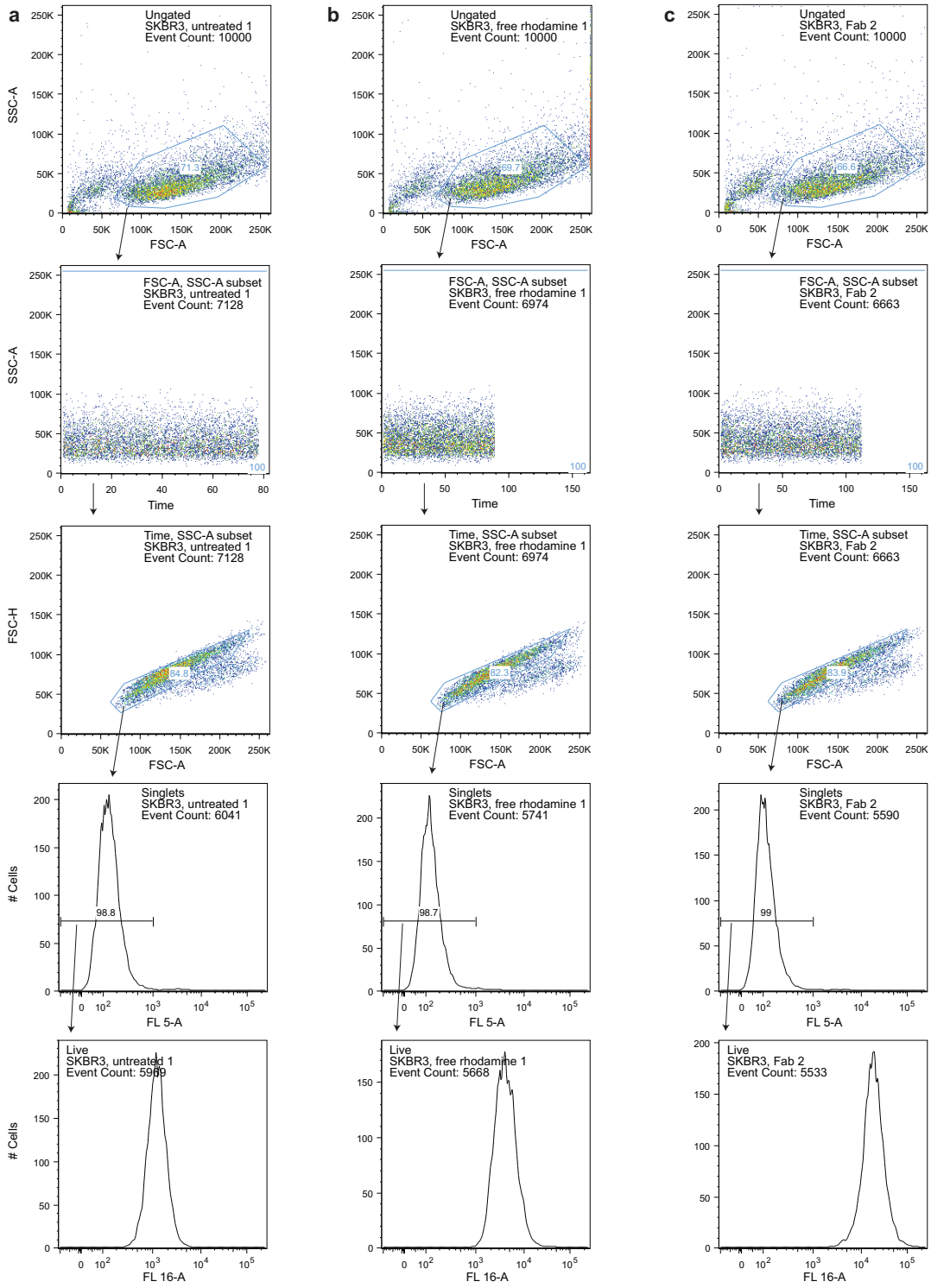
USA) for 5 minutes at 37 °C. Cells were collected in cold culture medium (500 µL with 10% FBS) and washed twice with ice cold PBS (5 mL each time). All centrifugation was carried out at 500 rpm for 5 min at rt. The cells were resuspended in flow cytometry buffer (500 µL PBS with 3% FBS). Samples were filtered and supplemented with 1 µL SYTOX Red (Thermo Fisher Scientific) to exclude dead cells. Samples were analyzed on an LSRFortessa flow cytometer (BD Bioscences). Experiments were carried out in duplicates. Flow cytometry data was analyzed using FlowJo (Version 9.7.6). The flow cytometry results are shown in Figure 61. Gating strategy, number of events and percentage of populations are shown below for representative samples.



Flow cytometry results for MCF7. Gating strategy, number of events and percentage of populations are shown for representative samples of untreated cells (**a**), or cells treated with free sulforhodamine B (**b**) or trastuzumab–**5–23** (**c**). FL 5-A was used to detect SYTOX™ Red (dead cells) and FL 16-A to detect rhodamine.



Flow cytometry results for SKOV3. Gating strategy, number of events and percentage of populations are shown for representative samples of untreated cells (**a**), or cells treated with free sulforhodamine B (**b**) or trastuzumab–**5**–**23** (**c**). FL 5-A was used to detect SYTOX™ Red (dead cells) and FL 16-A to detect rhodamine.



Flow cytometry results for SKBR3. Gating strategy, number of events and percentage of populations are shown for representative samples of untreated cells (**a**), or cells treated with free sulforhodamine B (**b**) or trastuzumab–**5–23** (**c**). FL 5-A was used to detect SYTOX™ Red (dead cells) and FL 16-A to detect rhodamine.

6.2. Transfer of small molecules and bioorthogonal handles

6.2.1. Labeling of titin-LACE$_C^{2M}$ with biotin thioesters 24 or 25

Titin-LACE$_C^{2M}$ or the corresponding acceptor tag mutant K6R (15 μM) was incubated with His$_6$-Ubc9 (60 μM) and thioester **24** or **25** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 0, 0.5, 1, 2, 4, 6 and 8 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE and western blot using streptavidin and quantified according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 64.

6.2.2. Labeling of GFP-LACE$_C$ with DBCO thioester 30

GFP-LACE$_C$ (15 μM) was incubated with His$_6$-Ubc9 (60 μM) and thioester **30**[42] (150 μM) in reaction buffer at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 0, 1, 3, 6 and 8 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 66.

6.2.3. Labeling of GFP-LACE$_C$ with hydroxylamine thioesters 31a or 31b

GFP-LACE$_C$ (15 μM) was incubated with His$_6$-Ubc9 (60 μM) and thioester **31a or 31b** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C in the dark. SDS–PAGE samples were taken from the reaction mixtures after 0, 4 and 8 h and quenched according to the general procedure. The time point samples were analyzed by Coomassie-stained reducing SDS–PAGE according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 69b.

6.2.4. Two-step labeling of GFP-LACE$_I$ with hydroxylamine thioesters 31a or 31b and PEG-KAT 32

GFP-LACE$_I$ (15 μM) was incubated with His$_6$-Ubc9 (60 μM) and thioester **31a or 31b** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) (200 μL) at 30 °C in the dark. SDS–PAGE samples were taken from the reaction mixtures after 0, 4 and 8 h and quenched according to the general procedure. The reaction was with diluted with reaction buffer containing 100 mM AET-HCl (200 μL) and incubated at rt for 5 min. Excess thioester was removed and the buffer exchanged to 50 mM Gly-HCl pH 3.6, 50 mM KF, 5% DMF by passing the sample through a desalting column. The sample was concentrated by spin diafiltration (10 kDa MWCO) to 200 μL. To the sample was

[42] Kindly provided by Dr. R. Boehringer (Bode Group, ETH Zürich).

added PEG-KAT **32**[43] (3 μL from a 10 mM stock solution in DMF, 10 equiv assuming a GFP concentration of 15 μM). The sample was placed into a glass vial and irradiated under a handheld UV lamp (365 nm) for 1 h at rt, and incubated at rt for 15 h. SDS–PAGE time points were taken from the reaction mixture after 0, 1 and 15 h, mixed with 2x reducing SDS–PAGE sample buffer and neutralized with NaOH (approximately 0.2 μL of 1 N NaOH) until the contained bromophenol blue indicator turned from yellow to a blue color prior to boiling. The time point samples were analyzed by Coomassie-stained SDS–PAGE according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 70.

6.2.5. Two-step labeling of GFP-LACE$_C$ with TIM thioester 39 and rhodamine hydroxylamine 40

GFP-LACE$_C$ (15 μM) was incubated with His$_6$-Ubc9 (60 μM) and thioester **39** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) (200 μL) at 30 °C. SDS–PAGE samples were taken from the reaction mixtures after 0, 2, 4 and 8 h and quenched according to the general procedure. Excess thioester was removed and the buffer exchanged to 50 mM Gly-HCl pH 3.6, 50 mM KF, 5% DMF by passing the sample through a desalting column. The sample was concentrated by spin diafiltration (10 kDa MWCO) to 300 μL. Half of this sample (150 μL) was treated with rhodamine hydroxylamine **40**[44] (5 μL from a 5 mM stock solution in DMF, approximately 15 equiv assuming a GFP concentration of 10 μM). The sample was briefly vortexed and incubated at rt for 16 h. SDS–PAGE samples were taken from the reaction mixture after 0, 1, 3 and 16 h, mixed with 2x reducing SDS–PAGE sample buffer and neutralized with NaOH (approximately 0.2 μL of 1 N NaOH) until the contained bromophenol blue indicator turned from yellow to a blue color prior to boiling. The time point samples were analyzed by Coomassie-stained SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 73.

6.2.6. IL-13–IL-4 conjugation by LACE and KAT ligation

Labeling of IL-13-LACE^{2M} with TIM thioester 39

For IL-13-LACE^{2M} labeling,[45] His$_6$-Ubc9 stocks were exchanged to reaction buffer without thiols using a desalting column. IL-13-LACE^{2M} (420 μg, 31 nmol, 280 μL from a 110 μM stock solution

---

43 Kindly provided by Dr. C. White (Bode Group, ETH Zürich). See footnote 22 for a reference.

44 Kindly provided by D. Schauenburg (Bode Group, ETH Zürich). See also: Schauenburg, D.; Divandari, M.; Neumann, K.; Spiegel, C. A.; Hackett, T.; Dzeng, Y.; Spencer, N. D.; Bode, J. W. Synthesis of polymers containing potassium acyltrifluoroborates (KATs) and post-polymerization ligation and conjugation. *Angew. Chem. Int. Ed.* **2020**, *59*, 14656–14663.

45 Kindly provided by M. Ninomiya. Unpublished results.

in water) (final concentration 15 μM) was incubated with His_6-Ubc9 (60 μM final concentration, from the prepared stock without thiols) and thioester **39** (150 μM) in reaction buffer (2053 μL) at 30 °C for 2 h. The reaction mixture was purified by preparative RP–HPLC on a Shiseido proteonavi column (5 μm, 300 Å pore size, 10 mm I.D. x 250 mm) with a gradient of 30–85% solvent B over 30 min at a flow rate of 5 mL/min. IL-13-$LACE^{2M}$–**39** coeluted with the starting material IL-13-$LACE^{2M}$, and MALDI–MS analysis showed approximately 25% conversion to the conjugate. The sample was used directly for KAT ligation with IL-4-HA-PG.

KAT ligation between IL-13-$LACE^{2M}$–39 and IL-4-HA-PG

The RP–HPLC fraction containing the product (approximately 5 mL) was diluted with 50 mM Gly-HCl, 50 mM KF, pH 3.6 (15 mL) and concentrated by spin diafiltration (3 kDa MWCO) to 250 μL. This sample was added to the lyophilized powder of IL-4-HA-PG^{45} (115 μg, 8.3 nmol, approximately 1–2 equiv relative to IL-13-$LACE^{2M}$–**39**) and irradiated for 20 min with handheld UV lamp (365 nm), and incubated at rt for 2 days. Samples (15 μL) were taken from the ligation mixture at various time points and analyzed by analytical RP–HPLC to check the reaction progress. The reaction mixture was purified by analytical RP–HPLC on a Shiseido proteonavi column (5 μm, 300 Å pore size, 4.6 mm I.D. x 250 mm) with a gradient of 20 to 95% solvent B over 14 min at flow rate 1 mL/min. MALDI–MS confirmed formation of the dimer (calc. 27.82 kDa, obs. 27.84 kDa). The purified product was analyzed by Coomassie-stained reducing SDS–PAGE according to the general procedure. Results of the RP–HPLC and SDS–PAGE analyses are shown in Figure 75b,c.

pSTAT assay

The pSTAT assays were performed by C. Egholm (Boyman Group, UZH). The results are shown in Figure 76.

6.3. Transfer of Ubls

6.3.1. GFP-$LACE_I$ labeling with full-length ubiquitin thioester 41

GFP-$LACE_I$ or the corresponding acceptor tag mutant K6R (15 μM) was incubated with His_6-Ubc9 (wt) or His_6-Ubc9-C93A (0 or 60 μM) and thioester **41** (150 μM; for reaction assembly, lyophilized thioester **41** was first dissolved in DMSO followed by dilution to a stock concentration of 300 μM with reaction buffer, less than 2.5% v/v final DMSO in the stock solution) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C for 24 h. Samples (20 μL) were taken from the reaction mixtures, treated with AET-HCl (1 mM final concentration) for 5 min at rt and mixed with an equal volume of Ni-NTA agarose resin (1:1 slurry in reaction buffer). The suspensions were agitated for 5 min and then passed through a fritted spin column (Bio-Rad) to remove resin-bound His_6-Ubc9 from the solution. The eluted samples were then quenched and analyzed by

Coomassie-stained reducing SDS–PAGE and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 81b.

6.3.2. SUMO2 and α-synuclein ubiquitination and ISG15ylation

For reaction assembly of ubiquitination reactions, lyophilized ubiquitin thioester **41** was first dissolved in a small amount of DMSO followed by reaction buffer to obtain a final stock solution with a concentration of 300 μM (less than 2.5% v/v final DMSO in the stock solution).

SUMO2 labeling with full length ubiquitin thioester 41

SUMO2 (wt), the consensus motif variant (IKTE) or the acceptor mutant (VRTE) (15 μM) was incubated with His_6-Ubc9-R13A (60 μM) and thioester **41** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C for 24 h. Samples (20 μL) were taken from the reaction mixtures, treated with AET-HCl (1 mM final concentration) for 5 min at rt and mixed with an equal volume of Ni-NTA agarose resin (1:1 slurry in reaction buffer). The suspensions were agitated for 5 min and then passed through a fritted spin column (Bio-Rad) to remove resin-bound His_6-Ubc9 from the solution. The eluted samples were then quenched and analyzed by Coomassie-stained reducing SDS–PAGE and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 82b.

α-Synuclein labeling with full length ubiquitin thioester 41 or ISG15 thioester 42

α-Synuclein (wt) or one of the variants as described in Figure 82 (15 μM) was incubated with His_6-Ubc9 (60 μM) and thioester **41** or **42** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) at 30 °C for 24 h. Samples (20 μL) were taken from the reaction mixtures, treated with AET-HCl (1 mM final concentration) for 5 min at rt and mixed with an equal volume of Ni-NTA agarose resin (1:1 slurry in reaction buffer). The suspensions were agitated for 5 min and then passed through a fritted spin column (Bio-Rad) to remove resin-bound His_6-Ubc9 from the solution. The eluted samples were then quenched and analyzed by Coomassie-stained reducing SDS–PAGE and quantified according to the general procedure. Results of the SDS–PAGE analysis are shown in Figure 82c and Figure 83.

6.4. TNFα-affibody conjugate

Labeling of TNFα-$LACE_N^{3M}$ with full length Z_{TNF} affibody thioester 43

TNFα-$LACE_N^{3M}$ (15 μM) was incubated with His_6-Ubc9 (0 or 60 μM) and thioester **43** (150 μM) in reaction buffer with 1 mM Ac-Cys-NHMe (**1**) (300 μL) at 30 °C. The reaction mixtures were purified by size-exclusion chromatography (Superdex 75) equilibrated in reaction buffer after 4 h (reaction in the absence of Ubc9) and 6 h (reaction in the presence of Ubc9). SDS–PAGE samples were taken from the SEC fractions, quenched and analyzed by Coomassie-stained reducing SDS–

PAGE according to the general procedure. Results of the SEC and SDS–PAGE analyses are shown in Figure 85a,b.

MS/MS analysis of TNFα-LACE$_N^{3M}$–43

Fraction 1 (see Figure 85a) (1 mL) containing the TNFα-LACE$_N^{3M}$–**43** conjugate was concentrated by spin diafiltration (10 kDa MWCO) to 0.5 mL. The sample contained 0.03 μg/μL of protein as determined by A_{280}. From this sample 50 μL was taken, diluted with 50 mM NH_4HCO_3 pH 8.0 (50 μL), and treated with 5 μL of 0.5 M TCEP (25 mM final concentration) at 45 °C for 30 min. To the sample was added 5 μL of 0.5 M iodoacetamide (25 mM final concentration) and the sample was incubated at rt for 30 min in the dark. The sample was treated with trypsin (0.2 μg, 2 μL of a freshly prepared 0.1 μg/μL stock in 50 mM NH_4HCO_3 pH 8.0) at 37 °C overnight in the dark. The digested sample was desalted using C_{18} ZipTips (100 μL bed size, Thermo Fisher) according to the instructions of the manufacturer and eluted in 70% aqueous MeCN with 0.1% TFA (30 μL). The sample was dried (Savant DNA Speed Vac 110), dissolved in 5% aqueous MeCN with 0.1% formic acid (10 μL) and measured on a timsTOF pro (Bruker). Data was analyzed using the Mascot server[46] (Matrix Science) by searching against the SwissProt database[47] and a custom database containing His_6-Ubc9, TNFα-LACE$_N^{3M}$ and the sequence of Z_{TNF} **43** with the fixed modification Cam (C) and the variable modifications oxidation (M), deamidation (NQ) and GlyGly(K). Results of the MS/MS analysis are shown in Figure 85c.

6.5. Post-assembly functionalization of AaLS-13-LACE$_C^{3M}$

For labeling of AaLS-13-LACE$_C^{3M}$, His_6-Ubc9 stocks and affibody thioester stocks (Z_{HER2}-Mes **46**) were thawed on ice and exchanged to cage buffer (50 mM sodium phosphates pH 7.6, 200 mM NaCl, 5 mM EDTA) using desalting columns prior to the LACE. Exchanged stocks were concentrated by spin diafiltration if required for reaction assembly (10 kDa and 3 kDa MWCO, respectively).

---

46 Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–3567.

47 Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **2000**, *28*, 45–48.

6.5.1. Functionalization of AaLS-13-LACE$_C^{3M}$ with small molecule thioesters

Labeling of AaLS-13-LACE$_C^{3M}$ with rhodamine thioester 5

Pre-assembled AaLS-13-LACE$_C^{3M}$ cages[48] (15 μM) were incubated with His$_6$-Ubc9 (30 μM) and thioester **5** (120 μM) in cage buffer (1 mL) at 30 °C for 1 h. Excess thioester was removed by passing the sample through a desalting column. The reaction mixture was purified at rt by size-exclusion chromatography (Superose 6 Increase, NGC Chromatography System, Bio-Rad) equilibrated in cage buffer at pH 8.0 to isolate the cage fraction. SDS–PAGE samples were taken from the reaction mixture after 0 and 1 h and of the SEC cage fraction, and quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. The cage fraction was analyzed by TEM.[49] Results of the SEC, SDS–PAGE and TEM analyses are shown in Figure 86b,c.

Labeling of AaLS-13-LACE$_C^{3M}$ with biotin thioester 24

Pre-assembled AaLS-13-LACE$_C^{3M}$ cages[48] (15 μM) were incubated with His$_6$-Ubc9 (30 μM) and thioester **24** (120 μM) in cage buffer (1 mL) at 30 °C for 2 h (low labeling), or with His$_6$-Ubc9 (60 μM) and thioester **24** (150 μM) in cage buffer (1 mL) at 30 °C for 8 h (high labeling). Excess thioester was removed by passing the samples through a desalting column at the end of the reaction. The reaction mixtures were purified at rt by size-exclusion chromatography (Superose 6 Increase, NGC Chromatography System, Bio-Rad) equilibrated in cage buffer at pH 8.0 to isolate the cage fractions. SDS–PAGE samples were taken from the reaction mixture after 0, 2 and 8 h and of the SEC cage fractions, and quenched and analyzed by Coomassie-stained reducing SDS–PAGE according to the general procedure. The cage fractions were analyzed by TEM[49] and ESI–MS. Results of the SEC, SDS–PAGE, ESI–MS and TEM analyses are shown in Figure 86b,d,e.

---

48 Kindly provided by M. Levasseur (Hilvert Group, ETH Zurich). Unpublished results. See also: (a) Tytgat, H. L. P.; Lin, C.; Levasseur, M. D.; Tomek, M. B.; Rutschmann, C.; Mock, J.; Liebscher, N.; Terasaka, N.; Azuma, Y.; Wetter, M.; Bachmann, M. F.; Hilvert, D.; Aebi, M.; Keys, T. G. Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. *Nat. Commun.* **2019**, *10*, 5403. (b) Sasaki, E.; Böhringer, D.; Waterbeemd, M. van de; Leibundgut, M.; Zschoche, R.; Heck, A. J. R.; Ban, N.; Hilvert, D. Structure and assembly of scalable porous protein cages. *Nat. Commun.* **2017**, *8*, 14663.

49 Azuma, Y.; Hilvert, D. Enzyme encapsulation in an engineered lumazine synthase protein cage. In *Protein Scaffolds: Design, Synthesis, and Applications*; Udit, A. K., Ed.; Methods in Molecular Biology; Humana Press: New York, NY, 2018; Vol. 1798, pp 39–55.

6.5.2. Functionalization of AaLS-13-LACE$_{C}^{3M}$ with peptides and affibody domains

Labeling of AaLS-13-LACE$_{C}^{3M}$ with peptide thioester 44

Pre-assembled AaLS-13-LACE$_{C}^{3M}$ cages[48] (15 μM) were incubated with His$_6$-Ubc9 (60 μM) and thioester **44** (150 μM) in cage buffer (5 mL) at 30 °C for 6 h. The reaction mixture was purified at rt by size-exclusion chromatography (Superose 6 Increase, NGC Chromatography System, Bio-Rad) equilibrated in cage buffer at pH 8.0 to isolate the cage fraction. SDS–PAGE samples were taken from the reaction mixture after 0 and 6 h and of the SEC cage fraction, and quenched and analyzed by Coomassie-stained reducing SDS–PAGE according to the general procedure. The cage fractions were analyzed by TEM[49] and ESI–MS. Results of the SDS–PAGE, ESI–MS and TEM analyses are shown in Figure 87c,d.

Attempt at two-step modification of AaLS-13-LACE$_{C}^{3M}$ with azide thioester 28 and DBCO peptide 45

LACE reaction with azide thioester 28: Pre-assembled AaLS-13-LACE$_{C}^{3M}$ cages[48] (15 μM) were incubated with His$_6$-Ubc9 (60 μM) and thioester **28** (150 μM) in cage buffer (1 mL) at 30 °C for 8 h. The reaction mixture was purified at rt by size-exclusion chromatography (Superose 6 Increase, NGC Chromatography System, Bio-Rad) equilibrated in cage buffer at pH 8.0 to isolate the cage fraction. SDS–PAGE samples were taken from the reaction mixture after 0, 4 and 8 h and of the SEC cage fraction, and quenched and analyzed by Coomassie-stained reducing SDS–PAGE according to the general procedure.

Click reaction with DBCO peptide 45: The azide-functionalized SEC cage fraction was concentrated by spin diafiltration (22.5 μM, 230 μL). To a part of this sample (90 μL) was added DBCO peptide **45** (5 equiv, 6.75 μL of a 1.5 mM stock solution in cage buffer) and the sample was incubated at rt for 12 h. An SDS–PAGE sample was taken at the end of the reaction and quenched and analyzed by Coomassie-stained reducing SDS–PAGE according to the general procedure. Analysis of the reaction mixture by size-exclusion chromatography (Superose 6 Increase, NGC Chromatography System, Bio-Rad) equilibrated in cage buffer at pH 8.0 at rt did not afford any cage fraction.

Results of the SEC and SDS–PAGE analyses are shown in Figure 88c,d.

Labeling of AaLS-13-LACE$_{C}^{3M}$ with full length ZHER2 affibody thioester 46

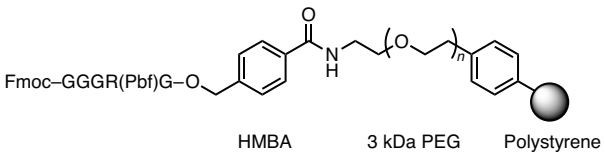
Pre-assembled AaLS-13-LACE$_{C}^{3M}$ cages[48] (15 μM) were incubated with His$_6$-Ubc9 (60 μM) and thioester **46** (150 μM) in cage buffer (4.8 mL) at 30 °C for 3 h (low labeling) or 8 h (high labeling). The reaction mixtures were purified at rt by size-exclusion chromatography (HiPrep 16/60 Sephacryl S-400 HR, NGC Chromatography System, Bio-Rad) equilibrated in cage buffer at pH 8.0 to isolate the cage fractions. SDS–PAGE samples were taken from the reaction mixtures after 0, 3 and 8 h and of the SEC cage fractions, and quenched and analyzed by Coomassie-

stained reducing SDS–PAGE according to the general procedure. The cage fractions were analyzed by TEM[49] and ESI–MS. Results of the SEC, SDSPAGE, ESI–MS and TEM analyses are shown in Figure 89b,c,d.

7. Experimental part for Chapter 4: Library screening workflow towards a redox-switchable disulfide tag

7.1. One-bead one-compound peptide library synthesis

7.1.1. Preparation of Fmoc-GGGR(Pbf)G-HMBA-tentagel resin (S13)



Fmoc-G-HMBA-tentagel resin (S14)

Tentagel MB HMBA resin (1 g resin, 0.4–0.5 mmol/g HMBA loading, 0.45 mmol, 1.0 equiv; Rapp Polymers GmBH, Tübingen, Germany; Cat. Nr. MB160140, 3 kDa PEG linker, 140–170 μm particle size, 450,000–500,000 beads/g) was washed with CH_2Cl_2 and DMF. The resin was swollen in DMF for 30 min and the solvent was drained. Separately, to a stirred solution of Fmoc-Gly-OH (1.338 g, 4.5 mmol, 10.0 equiv) in dry CH_2Cl_2 (30 mL) and DMF (2 mL) was added DIC (352 μL, 2.25 mmol, 5.0 equiv). After stirring at 0 °C for 20 min, the mixture was concentrated under reduced pressure. The residue was dissolved in DMF (10 mL), and the solution was added to the swollen resin. To the mixture was added DMAP (5.5 mg, 0.05 mmol, 0.1 equiv) and the suspension was agitated at rt for 1 h. The resin was washed with DMF and CH_2Cl_2 and dried under reduced pressure to give **S14**. A resin substitution of 0.455 mmol/g (Fmoc-Gly) was obtained as determined by Fmoc quantification.

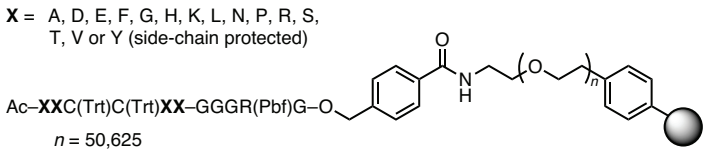
Fmoc-R(Pbf)G-HMBA-tentagel resin (S15)

S14 (1 g resin, 0.455 mmol/g loading, 0.455 mmol) was swollen in DMF, and Fmoc deprotection was performed according to the general procedure. The resin was then resuspended in DMF (10 mL). Separately, to a solution of Fmoc-Arg(Pbf)-OH (65 mg, 0.100 mmol, 0.100 mmol/g resin) and HCTU (39 mg, 0.095 mmol, 0.095 mmol/g resin) in a minimal amount of DMF was added NMM (22 μL, 0.200 mmol, 0.200 mmol/g resin). The solution was incubated at rt for 3 min, and the preactivated amino acid solution was added dropwise to the agitated resin suspension in order to achieve an even distribution of Fmoc-Arg coupling product across the resin batch. The suspension was agitated at rt for 1.5 h before the resin was washed and capped according to the general procedure. The resin was washed with DMF and CH_2Cl_2 and dried under reduced pressure to give **S15**. A resin substitution of 0.068 mmol/g (Fmoc-Arg) was obtained as determined by Fmoc quantification. We aimed for a relatively low resin substitution to reduce the chance for intermolecular interactions between the solid-supported peptides during library screening.

Fmoc-GGGR(Pbf)G-HMBA-tentagel resin (S13)

Onto Fmoc-R(Pbf)G-HMBA-tentagel (**S15**) (1 g resin, 0.068 mmol/g loading, 68 μmol), peptide **S13** was elongated by manual SPPS similarly to the general procedure using Fmoc-Gly-OH (121 mg, 408 μmol, 6.0 equiv), HCTU (166 mg, 401 mmol, 5.9 equiv) and NMM (90 μL, 816 μmol, 12.0 equiv) in DMF (10 mL). Couplings and Fmoc deprotection were performed twice, and capping was omitted.

7.1.2. Split and mix peptide library synthesis: Ac-XXC(Trt)C(Trt)XX-GGGR(Pbf)G-HMBA-tentagel resin



Library design and general synthetic strategy

The prepared resin **S13** was used to synthesis the peptide library by split-and-mix synthesis. Briefly, variable positions X were incorporated by distributing the pooled resin batch into 15 aliquots, and each aliquot was elongated by one of 15 Fmoc amino acids bearing suitable acid-labile protecting groups as described in the general procedure (A, D, E, F, G, H, K, L, N, P, R, S, T, V, Y). To enable peptide sequencing by MS/MS, Ile, which is isobaric to Leu, and Gln, which is nearly isobaric to Lys, were omitted. Additionally, Cys, Met and Trp which are prone to oxidation were omitted. Following elongation by one variable position X, the resin batch was pooled and mixed before proceeding to the next residue. Fmoc-Cys(Trt)-OH was coupled to the entire resin batch at the fixed positions, and the entire library was N-terminally acetylated.

Coupling of variable position X

The dry resin batch was evenly distributed into 15 peptide synthesis vessels (approximately 67 mg each, approximately 5 μmol scale each). The resin was swollen in DMF for 30 min, the N-terminal Fmoc group was removed with 20% piperidine in DMF (2 x 10 min), and the resin was washed. To each resin aliquot was coupled one of the 15 Fmoc amino acids (240 μL of a 0.50 M stock solution in DMF) using HATU (240 μL of a 0.49 M stock solution in DMF) and NMM (120 μL of a 2 M stock solution in DMF). Preactivation for 3 min and addition of the coupling solution was done manually. Fmoc deprotection, wash steps and agitation of the resin during reactions was performed in an automated fashion using a Multisyntech Syro I parallel synthesizer. The couplings were performed twice for 30 min. Following elongation by one variable position X, the resin was pooled, washed with DMF and CH_2Cl_2 and dried under reduced pressure.

Coupling of Fmoc-Cys(Trt)-OH at the fixed positions

The pooled resin (1 g resin, 0.068 mmol scale) was swollen in DMF for 30 min and Fmoc deprotection was performed according to the general procedure. Fmoc-Cys(Trt)-OH (8 mL of a 0.17 M stock solution in DMF, 1.360 mmol, 20.0 equiv) DIC (207 μL, 1.326 mmol, 19.5 equiv) and HOBt monohydrate (203 mg, 1.326 mmol, 19.5 equiv) were mixed and preactivated for 15 min before addition to the resin. Couplings were performed twice at rt for 1.5 h.

N-terminal acetylation of the library

Final Fmoc deprotection and capping was performed on the pooled resin batch according to the general procedure. The prepared library with the side chain protecting groups still in place was washed with DMF and CH_2Cl_2, dried under reduced pressure and stored under N_2 at –20 °C.

7.2. Removal of side chain protecting groups from bead-supported library

The acid-labile side chain protecting groups were removed from a portion of the library prior to use in subsequent analyses and library screening experiments. To a portion of the library (25 mg resin) in a fritted syringe was added a cocktail of 95% TFA, 2.5% TIPS and 2.5% water (1 mL). The suspension was agitated at rt for 2 h. The resin was extensively washed with DMF and CH_2Cl_2, dried under reduced pressure and stored under N_2 at –20 °C until use.

7.3. Cysteine alkylation, peptide cleavage and elution from single beads for library quality control by MALDI–MS

Complete reduction and alkylation of cysteine residues

Reduction and alkylation of cysteine side chains on resin for analysis by MALDI–MS was performed similarly to a previous report.[50] Briefly, to side chain-deprotected library on resin (1.7 mg) in a fritted syringe in 0.4 M $NaHCO_3$ with 30% MeCN (0.5 mL) was added TCEP (28 μL of a 0.9 M neutralized stock,[51] 50 mM final concentration). The suspension was agitated at rt under N_2 for 1.5 h. To the mixture was added iodoacetamide (200 μL of a 0.5 M stock solution in aq. 30% MeCN, 0.143 mM final concentration), and the suspension was agitated at rt under N_2 in the dark for 1 h. The resin was washed with aq. 30% MeCN and dried under reduced pressure.

---

50 Juskowiak, G. L.; McGee, C. J.; Greaves, J.; Vranken, D. L. V. Synthesis, Screening, and Sequencing of Cysteine-Rich One-Bead One-Compound Peptide Libraries. *J Comb. Chem.* **2008**, *10* (5), 726–731.

51 A fresh neutralized stock solution of TCEP (0.9 M, 100 μL, approximately pH 7–8) was prepared by dissolving TCEP hydrochloride (25.8 mg) in 5 N NaOH (57 μL) and 1 M K_2HPO_4 (43 μL) immediately before use.

Manual isolation of single beads

A portion of the alkylated library was placed onto a glass plate and suspended in a drop of aq. 50% MeCN. Under an inverted microscope (Nikon Eclipse TS100), eight beads were picked by taking them up individually with a micropipette set to 3 μL volume and dispensing the solution containing the bead into separate polypropylene microcentrifuge tubes (0.5 mL size).

Peptide cleavage using ammonia vapor and elution

Peptide cleavage by aminolysis of the HMBA ester by ammonia vapor to afford C-terminal amides, and elution of the peptides from the beads was performed similarly to previous reports.[52,53] Briefly, the open tubes containing single beads were placed in a desiccator and dried under high vacuum (8 h). Into the desiccator was added an open container containing aq. 25% NH_4OH (100 mL) and the lid was closed tightly using clamps.[54] After 18 h at rt, the container with aq. NH_4OH was removed, and vacuum was applied for another 1 h. To each tube was then added elution solution (10 μL, 3:4:3 AcOH/MeCN/H_2O), the samples were briefly centrifuged and incubated at rt for 18 h.

Optimization of MALDI sample preparation and MALDI–MS analysis

Given the limited peptide sample amount that can be obtained from single bead elution for analysis of a library member, we tested various spotting procedures and matrices and their influence on the signal-to-noise ratio in MALDI mass spectra. From the tested spotting procedures, the protocol described in Entry 4 afforded the best spectra quality, which is consistent with previous reports.[52]

---

52 Martínez-Ceron, M. C.; Giudicessi, S. L.; Marani, M. M.; Albericio, F.; Cascone, O.; Erra-Balsells, R.; Camperi, S. A. Sample preparation for sequencing hits from one-bead–one-peptide combinatorial libraries by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* **2010**, *400*, 295–297.

53 Giudicessi, S. L.; Gurevich-Messina, J. M.; Martínez-Ceron, M. C.; Erra-Balsells, R.; Albericio, F.; Cascone, O.; Camperi, S. A. Friendly strategy to prepare encoded one bead–one compound cyclic peptide library. *ACS Comb. Sci.* **2013**, *15*, 525–529.

54 Note of caution: Significant pressure build-up was observed. Care should be taken to secure the chamber lid.

Optimization of MALDI spotting procedure.

Entry	Matrix	Spotting method	Result
1	Matrix 1: 10 mg/mL CHCA in aq. 90% acetone Matrix 2: 10 mg/mL CHCA in aq. 50% MeCN / 0.1% TFA	0.7 μL matrix 1 (air-dried), overlaid with 0.7 μL of a 1:1 mixture of sample and matrix 2	High background, almost no peptide signal
2	4 mg/mL CHCA in aq. 50% MeCN / 0.1% TFA	0.7 μL matrix (air-dried), overlaid with 0.7 μL sample	Medium background, medium peptide signal
3	4 mg/mL CHCA in aq. 50% MeCN / 0.1% TFA	0.7 μL matrix (air-dried), overlaid with 0.7 μL sample (air-dried), followed again by 0.7 μL matrix	Medium background, medium peptide signal
4	4 mg/mL CHCA in aq. 50% MeCN / 0.1% TFA	0.7 μL sample (air-dried), overlaid with 0.7 μL matrix	Low background, strong peptide signal
5	10 mg/mL CHCA in aq. 50% MeCN / 0.1% TFA	0.7 μL sample (air-dried), overlaid with 0.7 μL matrix	Medium background, medium peptide signal

According to the optimized spotting procedure, eluted peptide solution (0.7 μL) was spotted onto a 96-spot MALDI target plate and the sample was allowed to air-dry. Onto the dried sample was spotted matrix solution (0.7 μL of 4 mg/mL CHCA in aq. 50% MeCN with 0.1% TFA), and the sample was again allowed to air-dry. Mass spectra were acquired by MALDI–MS (Bruker Microflex) using a reflective positive measuring mode. The MALDI–MS results are shown in Figure 93.

7.4. Library screening assay

7.4.1. Complete assay

Complete reduction

Side-chain deprotected library on resin (75 mg resin) was placed in a fritted syringe (5 mL syringe size) with buffer 1 (2.9 mL, 0.1 M sodium phosphates pH 7.4, aq. 30% MeCN) while agitating the suspension with a magnetic stir bar at 250 rpm.[55] After 15 min of swelling, TCEP (175 μL of a 0.9 M neutralized stock solution,[51] 50 mM final concentration) was added and the suspension was agitated at rt for 1.5 h under N_2.

Partial oxidation, rhodamine labeling and capping

To remove TCEP, the suspension was diluted repeatedly to 5 mL with buffer 1 followed by draining to 1 mL (ten times). The suspension (1.9 mL buffer 1) was agitated at rt for 15 min under

[55] MeCN was added as cosolvent to adjust the density of the buffer to allow the resin to stay submerged during the library screen. Stirring was set to a sufficient speed to keep the resin suspended in the buffer.

air. To the suspension was added sulforhodamine chloroacetamide **46** (100 μL of a 5 mM stock solution in MeCN, 250 μM final concentration). The suspension was agitated at rt for 5 min under N_2 before iodoacetamide (500 μL of a 0.5 M stock solution in MeCN, 0.1 M final concentration) was added. The suspension was agitated at rt for 30 min under N_2 in the dark, and excess reagents were removed by repeated dilution with buffer 2 (degassed buffer 1 by argon bubbling and sonication for 30 min; ten times as above).

Partial reduction, fluorescein labeling

To the suspension (2.36 mL buffer 2) was added TCEP (140 μL of a 9 mM neutralized stock solution in buffer 2, 0.5 mM final concentration). The suspension was agitated at rt for 15 min under N_2, and 5(6)-(chloroacetamido)fluorescein **47** (500 μL of a 5 mM stock solution in MeCN, 0.83 mM final concentration, 2 equiv relative to TCEP) was added. The suspension was agitated at rt for 5 min under N_2 and excess reagents were removed by repeated dilution with buffer 2 (ten times as above).

Complete reduction and alkylation

To the suspension (2.36 mL buffer 2) was added TCEP (140 μL of a 0.9 M neutralized stock solution,[51] 50 mM final concentration) and the suspension was agitated at rt for 1 h under N_2. Iodoacetamide (500 μL of a 0.5 M stock solution in MeCN, 83 mM final concentration) was added and the suspension was agitated at rt for 30 min under N_2 in the dark. The resin was washed with aq. 30% MeCN, MeCN, DMF and CH_2Cl_2. The resin was dried under reduced pressure and stored at –20 °C until sorting.

7.4.2. Control assays

No TCEP control

Side-chain deprotected resin (8 mg resin) was placed in a 5 mL fritted syringe. The assay was performed as described for the complete assay using the same amounts and volumes of reagents (Section 7.4.1), with the difference that addition of 0.5 mM TCEP during partial reduction before fluorescein labeling was omitted. At the end of the assay, the resin was completely reduced and alkylated as described for the complete assay.

No rhodamine control

Side-chain deprotected resin (15 mg resin) was placed in a 5 mL fritted syringe. The assay was performed as described for the complete assay using the same amounts and volumes of reagents (Section 7.4.1), with the difference that addition of sulforhodamine chloroacetamide **46** was omitted after partial oxidation for 15 min under air. Instead, the suspension was agitated for 5 min under N_2 before iodoacetamide was added. The rest of the assay, including partial reduction,

fluorescein labeling and complete reduction and alkylation at the end of the assay was performed as described for the complete assay.

No fluorescein control

Side-chain deprotected resin (8 mg resin) was placed in a 5 mL fritted syringe. Complete reduction, partial oxidation, rhodamine labeling and capping was performed as described for the complete assay using the same amounts and volumes of reagents (Section 7.4.1). After rhodamine labeling, the resin was completely reduced and capped as described in the last step of the complete assay.

7.5. Library sorting by fluorescence-activated bead sorting

The solid-supported library (20–100 mg resin) was resuspended in 10 mM NMM/AcOH volatile buffer pH 7.2–7.4 and 10% EtOH (40 mL) and agitated for 1 h to equilibrate the resin. The suspension was vortexed and passed through a 355 μm sieve. The flow through containing dispersed resin was sorted using a fluorescence-activated sorter (BioSorter, Union Biometrica) equipped with a 50 mL sample cup with an overhead stirrer and a FOCA 1000 flow cell (200–700 μm particle size sorting) using Milli-Q water as sheath fluid. Single beads were selected based on the axial length (first gate, time-of-flight, approximately 80–90% of the total events), and single beads were sorted using green peak height (excitation at 488 nm) versus red peak height (excitation at 561 nm) with 90% red compensation to isolate library hits (second gate, 2.1% of single bead events across all sorts). Hit beads were dispensed individually into 96-well semi-skirted PCR plates in approximately 75 μL sheath fluid per well. A representative example of a library sorting result is shown in Figure 95b. The number of events and sorting statistics are given in Table 2.

7.6. Peptide sequencing of library hits by tandem MALDI–MS/MS (TOF–TOF)

Sample preparation

Sorted beads in 96-well semi-skirted PCR plates were inspected under a handheld UV lamp (365 nm). Under the UV light, individual fluorescein-labeled beads were visible by eye and wells that contained more than one bead were excluded from further analysis (typically less than 10% of the wells). Using a 96-well liquid handler, MeCN was added to a final volume of 150 μL, the plates were briefly centrifuged (4000 x g) and the supernatant (120 μL) was removed. The wash step with MeCN was repeated two times. The plates containing washed beads in MeCN (30 μL each) were placed in a desiccator and dried under reduced pressure for 24 h. Resin cleavage by ammonia vapor and elution of peptides from individual beads was performed directly in the 96-well plates as described in Section 7.3 of the Experimental Part (“Peptide cleavage using ammonia vapor and elution”).

Tandem MALDI–MS/MS (TOF–TOF)

Eluted peptide solution (0.7 μL) was spotted onto a 384-spot MALDI target plate and the sample was allowed to air-dry. Onto the dried sample was spotted matrix solution (0.7 μL of 4 mg/mL CHCA in aq. 50% MeCN with 0.1% TFA), and the sample was again allowed to air-dry. Samples were analyzed by tandem MALDI–MS/MS (TOF–TOF) (Bruker Ultraflex) in batch mode by automated identification of intense parent ions and selection for fragmentation (typically 3–5 MS/MS spectra per spot).

Semi-automated workflow for analysis of peptide sequencing data

Mascot search: MS/MS data was analyzed using the Mascot server[46] (Matrix Science) by searching against a custom database containing all possible peptide sequences of the library (XXCCXXGGGRG, where X is either A, D, E, F, G, H, K, L, N, P, R, S, T, V or Y) with the fixed modifications acetyl (N-terminal) and amidated (C-terminal), and the variable modifications Cam (C), Cys -> Dap (C), Cys -> Dha (C) and fluorescein (C).

Semi-automated data curation and analysis: Search results were exported from the Mascot server as text file and curated using a custom Python script (see Section 6 of the Appendix for the Python script and a short output example). Briefly, for automated data curation, we took advantage of the fact that the eluted peptide sample from a single bead was often obtained as a mixture with variable modifications on the cysteine residues, especially Cam and its side products Dha (base-promoted elimination during ammonia cleavage from resin) and Dap (addition of ammonia to Dha), and fluorescein (incomplete fluorophore labeling during library screening). Since the differently modified species were fragmented and sequenced separately, this redundancy allowed us to independently verify the sequencing results in these cases. To this end, sequence results of all parent ions from a particular MALDI target spot, representing one sample of a library hit, were collected and compared. Based on this group, the sequence of a particular sample was deemed to be positively identified if the two peptide sequence results with the highest Mascot peptide score were identical (if more than one parent ion was sequenced), and if the highest Mascot peptide score was 25 or greater. Else, the results were rejected. Additionally, the mass differences between the parent ions were compared to expected values based on the various possible combinations of variable modifications. Expected differences (m/z): 512.07119, 495.04469, 478.01819, 421.06201, 404.03551, 330.05283, 182.01836, 164.99186, 147.96536, 108.03568, 91.00918, 73.98268, 56.95618, 34.053, 17.0265; Search range (m/z): ±0.6. If parent ions were found that were unrelated (mass differences other than the expected values), the result was flagged for manual inspection to determine whether the sample may have contained multiple beads. If found so, the results were also rejected. A complete list of the library hit sequences is given in Section 5 of the Appendix.

7.7. CysTag-GFP variant reactivity tests

7.7.1. Competition experiment with RNase A

CysTag-GFP variants (25 μM, a control reaction without GFP variant was performed) and RNase A (25 μM, from a fresh 100 μM stock solution in reaction buffer; Boehringer Mannheim GmbH) were incubated with TCEP (25 μM final concentration, 1 equiv, added from a fresh 0.9 M neutralized stock solution[51]) for 15 min at rt in reaction buffer (50 μL total volume). The samples were placed on ice, cysteine-reactive rhodamine probe **48** (100 μM, 4 equiv, from a 5 mM stock solution in DMF) was added and the samples were briefly vortexed. After incubating the samples at 0 °C for 2 h, the reactions were quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 99a.

7.7.2. Competition experiment with chicken egg lysozyme

CysTag-GFP variants (25 μM, a control reaction without GFP variant was performed) and chicken egg lysozyme (25 μM, from a fresh 100 μM stock solution in reaction buffer; PanReac AppliChem) were incubated in the presence or absence of TCEP (25 μM final concentration, 1 equiv, added from a fresh 0.9 M neutralized stock solution[51]) for 15 min at rt in reaction buffer (50 μL total volume). The samples were placed on ice, cysteine-reactive rhodamine probe **48** (100 μM, 4 equiv, from a 5 mM stock solution in DMF) was added and the samples were briefly vortexed. After incubating the samples at 0 °C for 2 h, the reactions were quenched and analyzed by Coomassie-stained reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 99c.

7.7.3. Reduction and labeling of CysTag-GFP variants

CysTag-GFP variants (25 μM) were incubated in the presence TCEP (0, 0.5, 1 or 2 equiv, 0–50 μM final concentration, added from a fresh 0.9 M neutralized stock solution[51]) for 15 min at rt in reaction buffer (50 μL total volume). The samples were placed on ice, cysteine-reactive rhodamine probe **48** (100 μM, 4 equiv, from a 5 mM stock solution in DMF) was added and the samples were briefly vortexed. After incubating the samples at 0 °C for 2 h, the reactions were quenched and analyzed by reducing SDS–PAGE and in-gel fluorescence according to the general procedure. Results of the SDS–PAGE analyses are shown in Figure 99d.

APPENDIX

1. Diffusion-limited protein–protein conjugation

Table 3. Data, calculations and references for estimation of the diffusion-limited protein–protein conjugation rate in H_2O as a function of molecular weight.

Protein	MW (kDa)	r [a] (Å)	δ [b] (°)	F [b] (10^{-3})	D_T [c] (10^{-7} cm^2 s^{-1})	D_R [d] (10^{-7} s^{-1})	k [e] (M^{-1} s^{-1})
EGF	6.6	12.2	11.9	2.67	11.72	12.12	2.29E+06
Cytochrome *c*	13.4	15.5	9.4	1.66	11.63	5.97	1.28E+06
RNase	13.8	15.6	9.3	1.63	10.49	5.80	1.17E+06
RNase	13.8	15.6	9.3	1.63	11.10	5.80	1.21E+06
RNase	13.8	15.6	9.3	1.63	10.32	5.80	1.16E+06
Lysozyme	14.3	15.8	9.2	1.59	11.19	5.59	1.18E+06
α-Lactalbumin	14.2	15.8	9.2	1.60	10.58	5.63	1.15E+06
Lactalbumin	14.2	15.8	9.2	1.60	9.97	5.63	1.11E+06
Trypsin	15.1	16.1	9.0	1.53	10.93	5.30	1.12E+06
Myoglobin	16.9	16.7	8.7	1.42	10.32	4.73	9.87E+05
Myoglobin	16.9	16.7	8.7	1.42	11.28	4.73	1.04E+06
α-Chymotrypsin	21.6	18.2	8.0	1.20	10.23	3.70	8.04E+05
Chymotrypsinogen	21.6	18.2	8.0	1.20	9.53	3.70	7.72E+05
Pepsin	35.0	21.4	6.8	0.87	9.01	2.29	5.05E+05
Ovalbumin	43.5	23.0	6.3	0.75	6.82	1.84	3.60E+05
Ovalbumin	43.5	23.0	6.3	0.75	7.17	1.84	3.70E+05
Ovalbumin	43.5	23.0	6.3	0.75	7.26	1.84	3.73E+05
BSA	66.5	26.5	5.5	0.57	6.30	1.20	2.43E+05
BSA	66.5	26.5	5.5	0.57	5.60	1.20	2.27E+05
BSA	66.5	26.5	5.5	0.57	6.03	1.20	2.37E+05
Fibrinogen	339.7	45.9	3.1	0.19	2.97	0.24	4.19E+04
Fibrinogen	339.7	45.9	3.1	0.19	2.01	0.24	3.37E+04

[a] Minimal radius of a sphere that can contain the protein, estimated according to Erickson *et al.*[56] [b] Polar angle of reactive surface (δ) and reactive surface fraction (F), calculated for a reactive area of 5 Å^2 and a sphere with radius r. [c] Experimentally determined translational diffusion coefficient as compiled in Table 2 of Nauman *et al.*[57] Data was rescaled to 20 °C as described by Ryabov *et al.*[58] using $D_T^* \eta(T)/T$ = constant, where T is the temperature in Kelvin and $\eta(T) = 1.7753 - 0.0565 * (T - 273) + 1.0751 \times 10^{-3} * (T - 273)^2 - 9.2222 \times 10^{-6} * (T - 273)^3$. [d] Rotational diffusion coefficient, calculated for the above proteins using $D_R = (2 * \tau_C)^{-1}$. Rotational correlation time τ_C was estimated using the linear relationship $\tau_C = 0.625 * MW$ (kDa),

56 Erickson, H. P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol. Proced. Online* **2009**, *11*, 32–51.

57 Nauman, J. V.; Campbell, P. G.; Lanni, F.; Anderson, J. L. Diffusion of insulin-like growth factor-I and ribonuclease through fibrin gels. *Biophys. J.* **2007**, *92*, 4444–4450.

58 Ryabov, Y. E.; Geraghty, C.; Varshney, A.; Fushman, D. An efficient computational method for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation. *J. Am. Chem. Soc.* **2006**, *128*, 15432–15444.

based on experimentally determined τ_C as a function of MW as compiled in Figure 1 of Su *et al.*[59] and Table 1 of Ryabov *et al.*[58] [e] The diffusion-limited rate constant (k) with the orientation constraints of δ and F was calculated according to previous reports[60,61,62] using $k = 4\pi * N_A * 2r * 2D_T * F * \xi * \tan(\delta/2)$, where $\xi = [(1 + D_R * (2r)^2 / D_T) / 2]^{1/2}$ and N_A is Avogadro's number.

59 Su, X.-C.; Jergic, S.; Ozawa, K.; Burns, N. D.; Dixon, N. E.; Otting, G. Measurement of dissociation constants of high-molecular weight protein–protein complexes by transferred ^{15}N-relaxation. J. *Biomol. NMR* **2007**, *38*, 65–72.

60 Berg, O. G. Orientation constraints in diffusion-limited macromolecular association. The role of surface diffusion as a rate-enhancing mechanism. *Biophys. J.* **1985**, *47*, 1–14.

61 Northrup, S. H.; Erickson, H. P. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3338–3342.

62 Schreiber, G.; Haran, G.; Zhou, H.-X. Fundamental aspects of protein-protein association kinetics. *Chem. Rev.* **2009**, *109*, 839–860.

Table 4. Data, calculations and references for estimation of the diffusion-limited protein–protein conjugation rate in *E. coli* cytoplasm as a function of molecular weight.

Protein	MW (kDa)	r [a] (Å)	δ [a] (°)	F [a] (10^{-3})	D_T [b] (10^{-7} cm^2 s^{-1})	D_R [a] (10^{-7} s^{-1})	k [a] (M^{-1} s^{-1})
eYFP	26.7	19.5	7.4	1.04	0.71	3.00	1.62E+05
PtsH-YFP	35.8	21.5	6.7	0.86	0.38	2.23	9.25E+04
Crr-YFP	45.0	23.3	6.2	0.74	0.20	1.78	5.58E+04
CFP-CheW-YFP	71.5	27.2	5.3	0.54	0.15	1.12	3.26E+04
CFP-CheR-YFP	86.2	28.9	5.0	0.48	0.17	0.93	2.96E+04
DnaK-YFP	95.8	30.0	4.8	0.44	0.07	0.84	1.70E+04
HtpG-YFP	198.0	38.3	3.8	0.27	0.17	0.40	1.45E+04
CFP-CheA-YFP	249.6	41.4	3.5	0.23	0.04	0.32	6.16E+03

[a] See footnote in Table 3. [b] Experimentally determined translational diffusion coefficient D_T at 20 °C by Kumar *et al.*[63]

[63] Kumar, M.; Mommer, M. S.; Sourjik, V. Mobility of cytoplasmic, membrane, and DNA-binding proteins in *Escherichia coli*. *Biophys. J.* **2010**, *98*, 552–559.

2. Sequence alignment of E2 conjugating enzymes

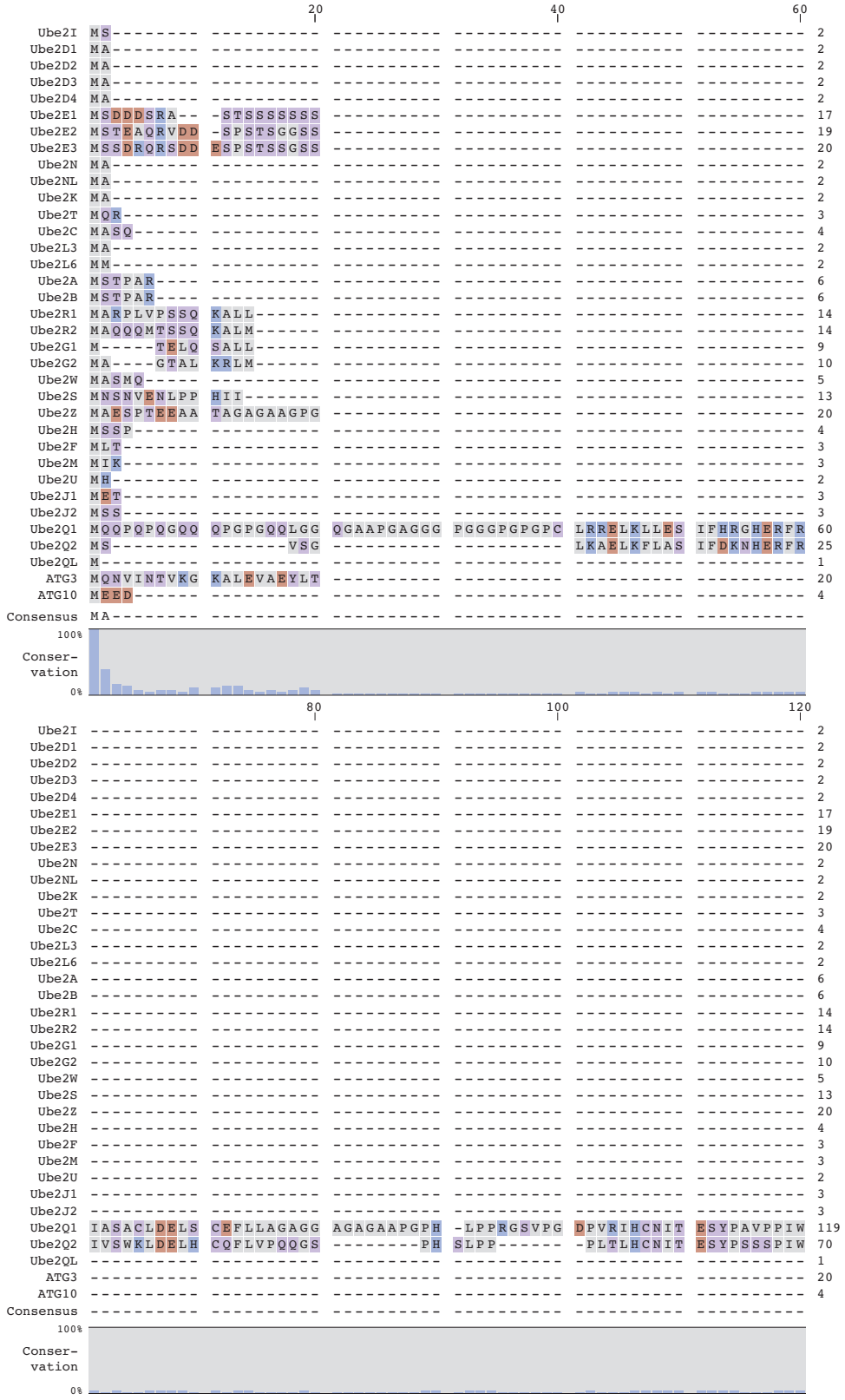


Figure 100. Sequence alignment of E2 conjugating enzymes. Sequence alignment was generated using CLC Genomics Workbench 12.0.3. **Figure continues to next page.**

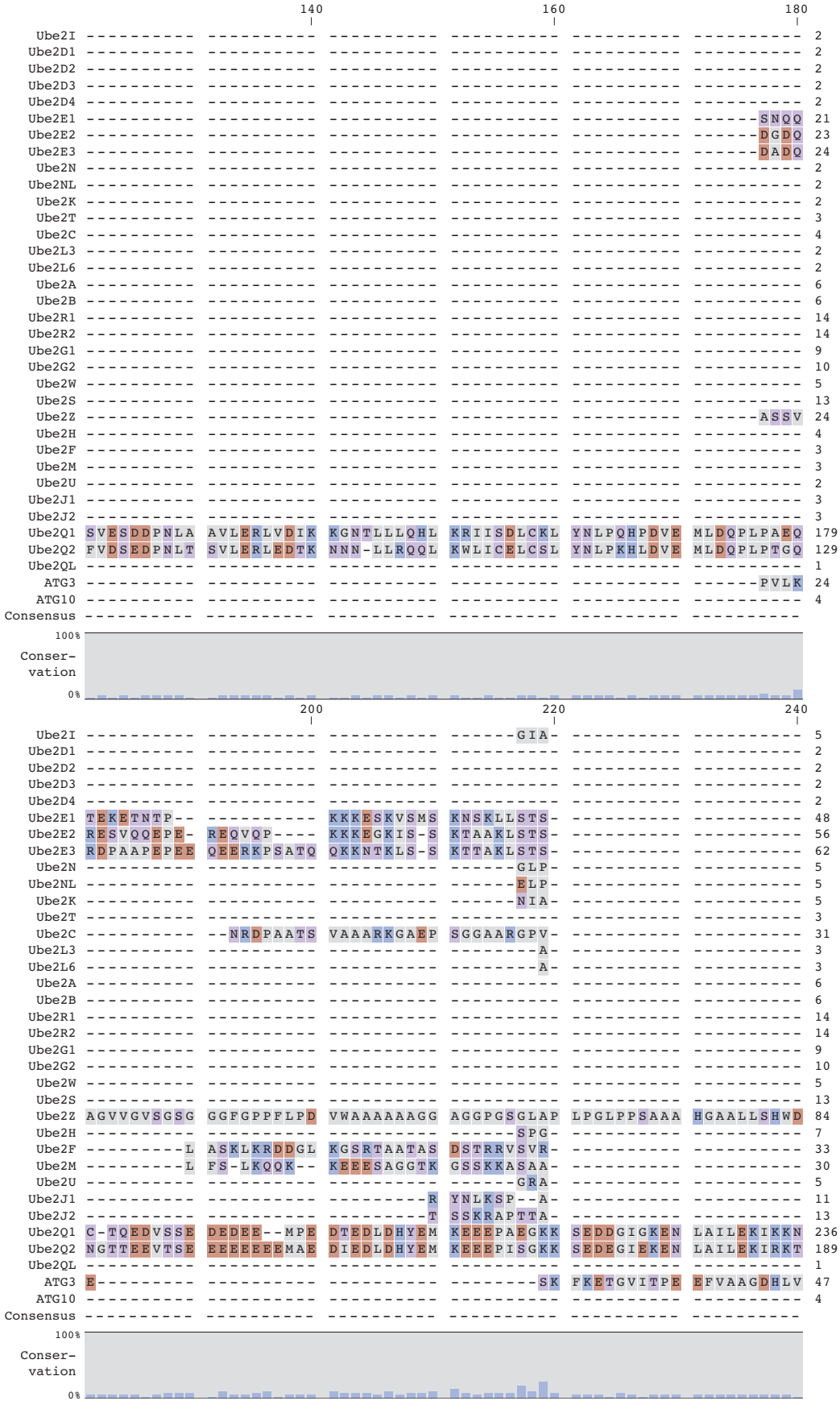


Figure 100. Continued.

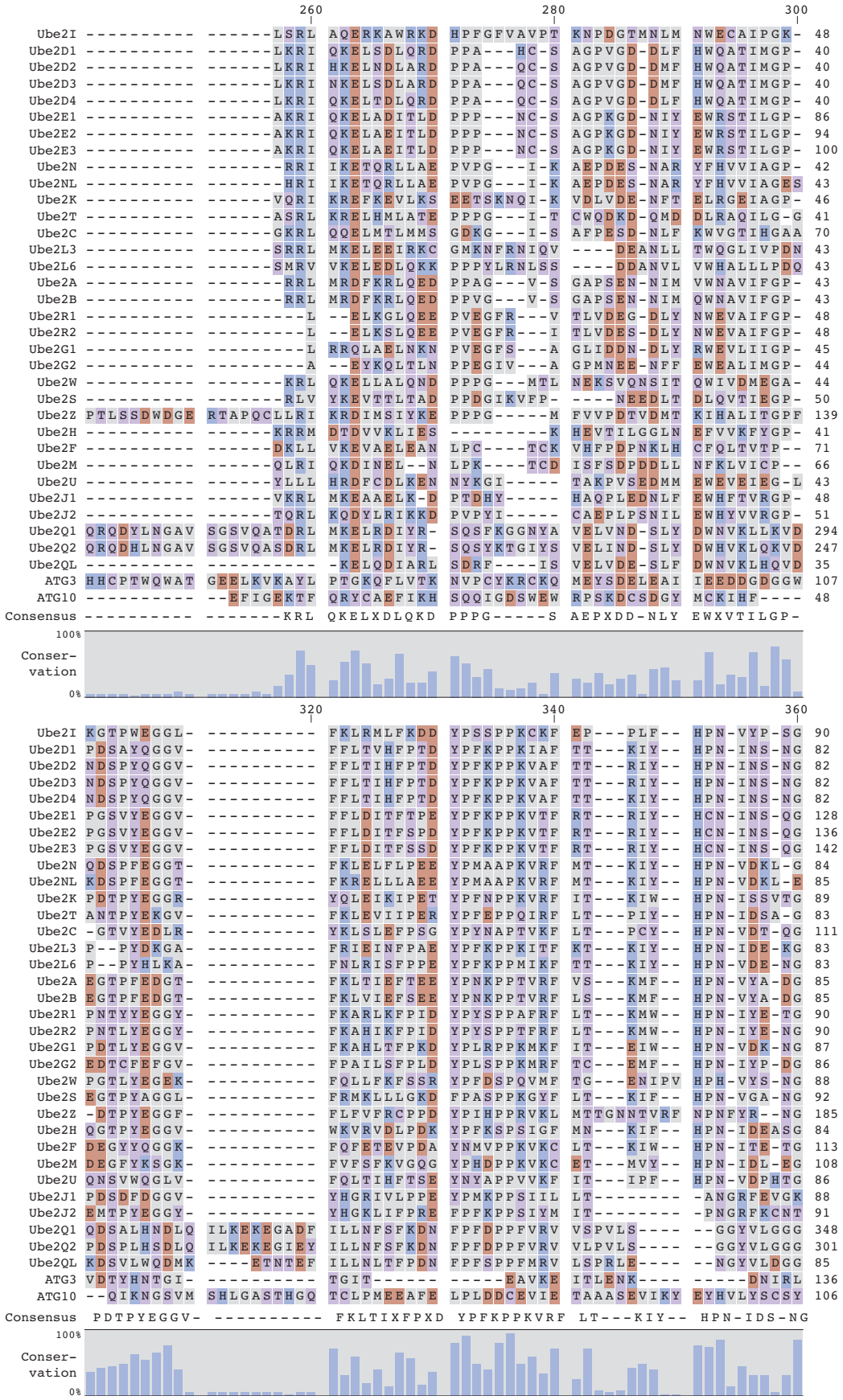


Figure 100. Continued.

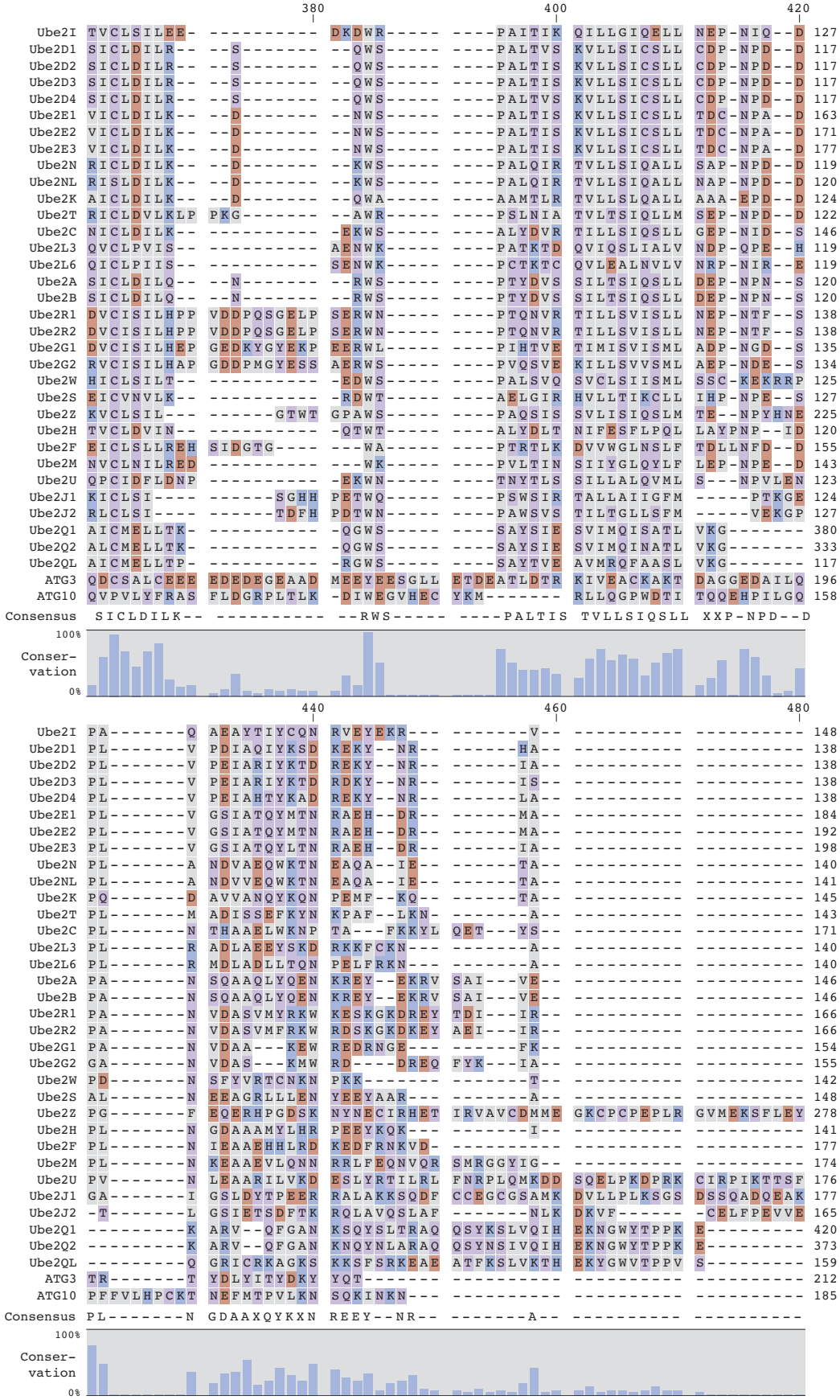


Figure 100. Continued.

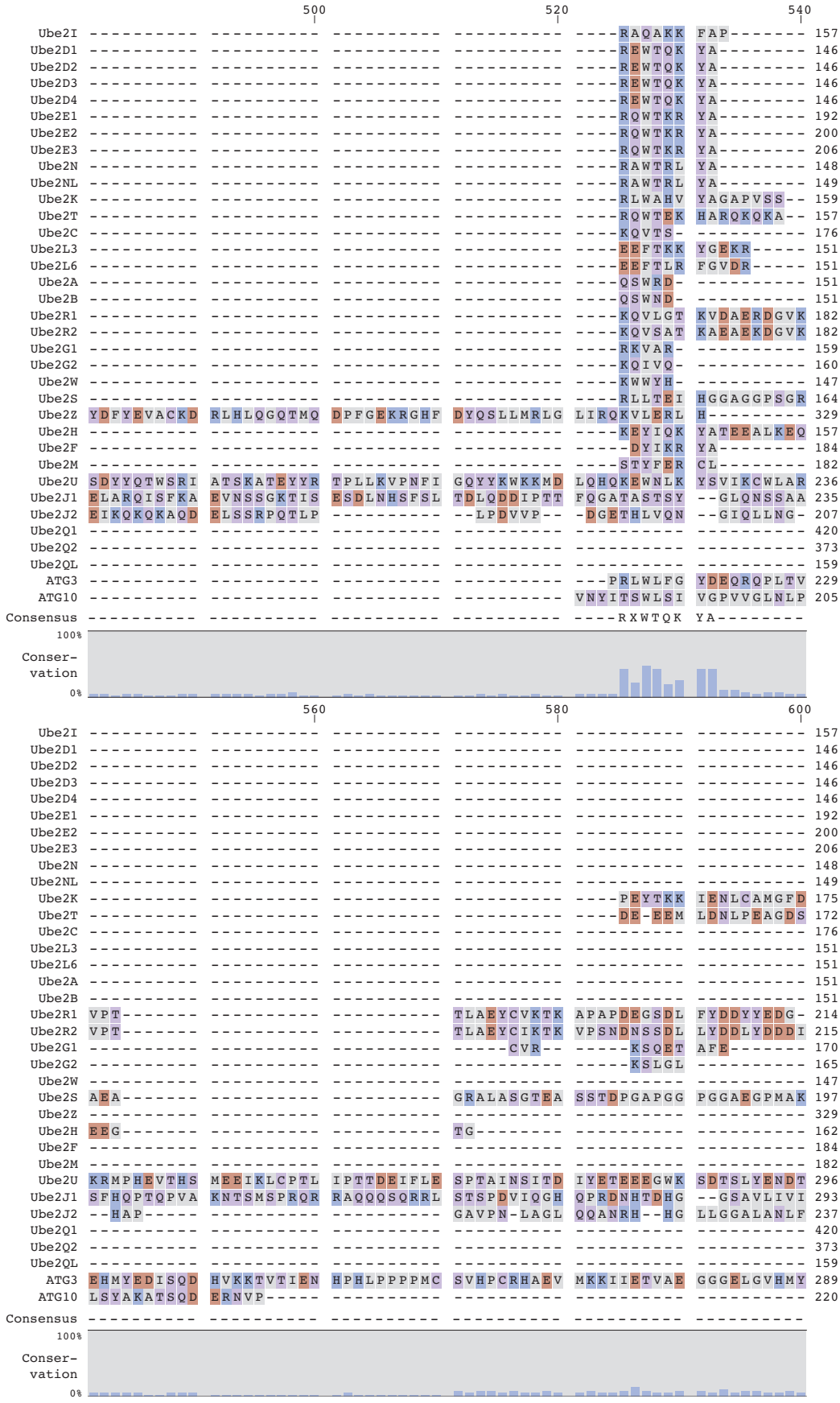


Figure 100. Continued.

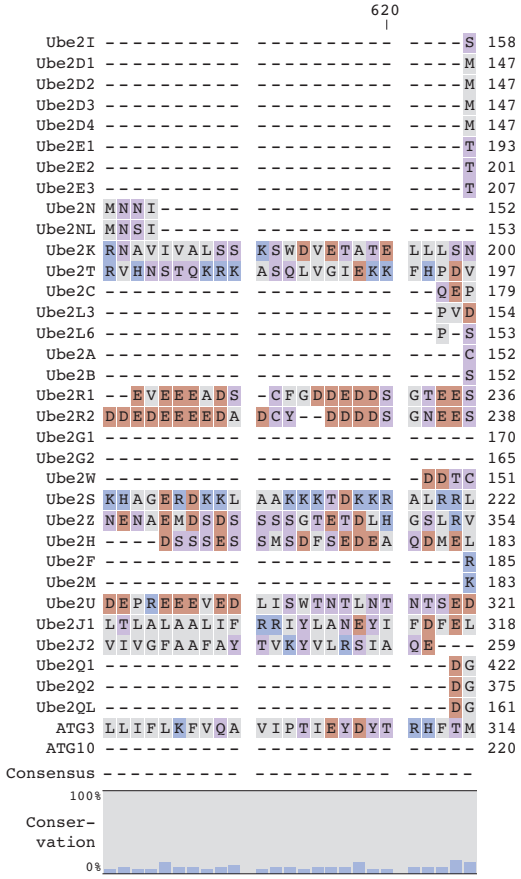


Figure 100. Continued.

3. Tandem mass spectrometry analysis of self-labeled Ubc9

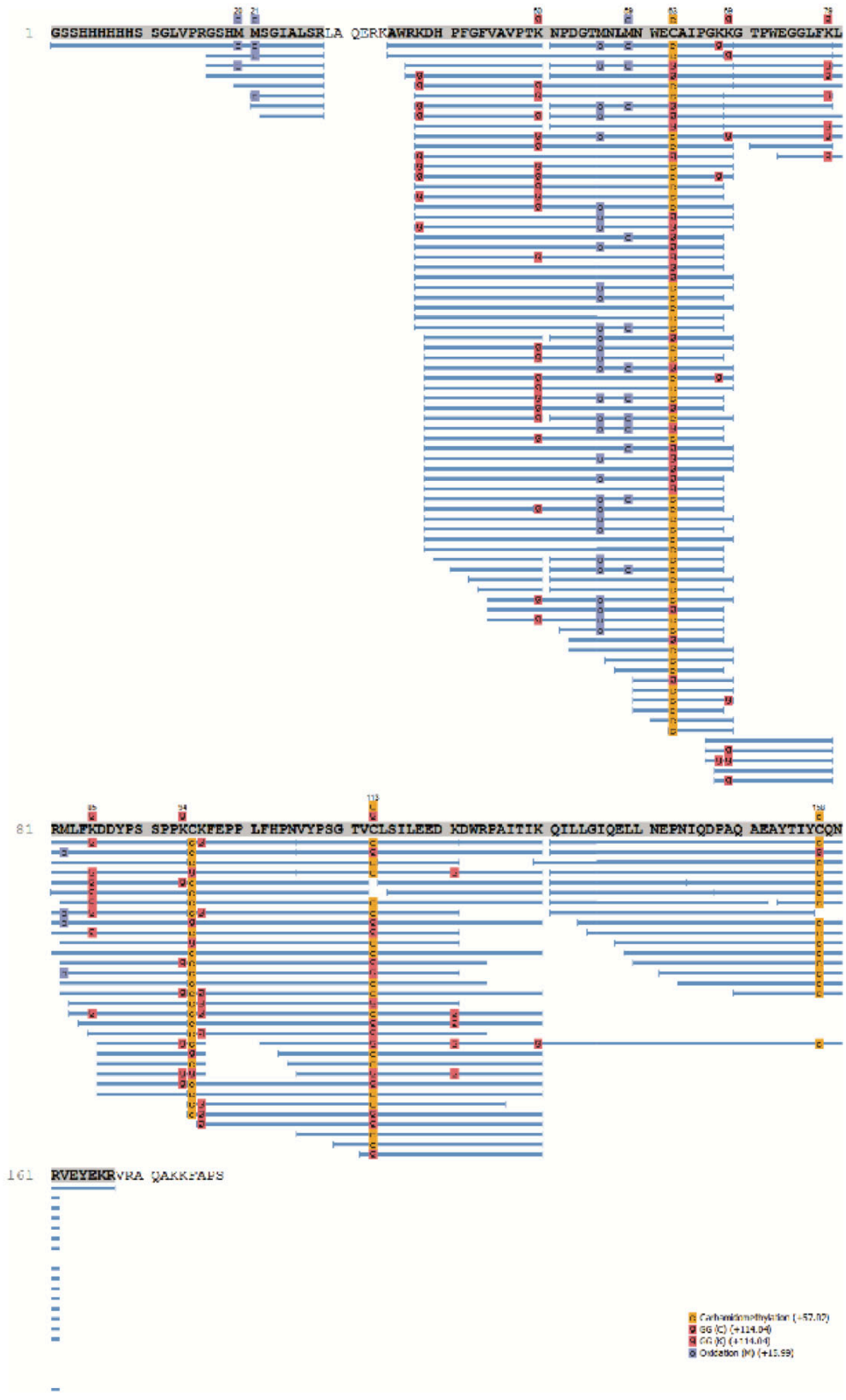


Figure 101. MS/MS analysis of Ubc9–**5** self-labeling.

4. Protein sequences

Table 5. Protein sequences.

Protein	Amino acid sequence	Comments
Ubc9 variants		
His_6-Ubc9	**GSSHHHHHHSSGLVPRGSHM**MSGIALSRLAQERKAWRKDHPFGFV AVPTKNPDGTMNLMNWECAIPGKKGTPWEGGLFKLRMLFKDDYPS SPPKCKFEPPLFHPNVYPSGTVCLSILEEDKDWRPAITIKQILLG IQELLNEPNIQDPAQAEAYTIYCQNRVEYEKRVRAQAKKFAPS	Thrombin-cleavable His_6-tag (bold)
His_6-Ubc9-K48A-K49A-E54A-C138A	**GSSHHHHHHSSGAENLYFQ**GSGIALSRLAQERKAWRKDHPFGFVA VPTKNPDGTMNLMNWECAIPGAAGTPWAGGLFKLRMLFKDDYPSS PPKCKFEPPLFHPNVYPSGTVCLSILEEDKDWRPAITIKQILLGI QELLNEPNIQDPAQAEAYTIYAQNRVEYEKRVRAQAKKFAPS	X-ray construct; TEV-cleavable His_6-tag (bold)
LACE substrates		
His_6-GFP-$LACE_C$	**GSSHHHHHHSSGAENLYFQ**GMRKGEELFTGVVPILVELDGDVNGH KFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC FARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNG IKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQS VLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK*GSG*PRKVIKMES EE	LACE tag (underlined), TEV-cleavable His_6-tag (bold), additional residues (italic)
His_6-GFP-$LACE_I$	**GSSHHHHHHSSGAENLYFQ**GMRKGEELFTGVVPILVELDGDVNGH KFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC FARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNG IKANFKIRHNVED*GSG*PRKVIKMESEE*GS*GSVQLADHYQQNTPIG DGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMD ELYK	LACE tag (underlined), TEV-cleavable His_6-tag (bold), additional residues (italic)
His_6-GFP-$LACE_N$	**GSSHHHHHHSSGAENLYFQ**GPRKVIKMESEE*GSG*MRKGEELFTGV VPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVP WPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFK DDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFN SHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDG PVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDEL YK	LACE tag (underlined), TEV-cleavable His_6-tag (bold), additional residues (italic)
His_6-GFP-$LACE_{I,C}^{2M}$	**GSSHHHHHHSSGAENLYFQ**GMRKGEELFTGVVPILVELDGDVNGH KFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC FARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNG IKANFKIRHNVED*GSG*PRKVIKMESEE*GS*GSVQLADHYQQNTPIG DGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMD ELYK*GSG*PRAVIKQESEE	LACE tags (internal tag, C-terminal tag^{2M}) (underlined), TEV-cleavable His_6-tag (bold), additional residues (italic)
His_6-GFP-$LACE_C^{minimal}$	**GSSHHHHHHSSGAENLYFQ**GMRKGEELFTGVVPILVELDGDVNGH KFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC FARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNG IKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQS VLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK*GSG*IKQE	Minimal LACE tag (underlined), TEV-cleavable His_6-tag (bold), additional residues (italic)
His_6-GFP-$LACE_I^{minimal}$	**GSSHHHHHHSSGAENLYFQ**GMRKGEELFTGVVPILVELDGDVNGH KFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC FARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNG IKANFKIRHNVED*GSG*IKQE*GS*GSVQLADHYQQNTPIGDGPVLLP DNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK	Minimal LACE tag (underlined), TEV-cleavable His_6-tag (bold), additional residues (italic)

His_6-Titin-$LACE_C^{2M}$	**GSSHHHHHHSSGAENLYFQ**GLIEVEKPLYGVEVFVGETAHFEIEL SEPDVHGQWKLKGQPLTASPDSEIIEDGKKHILILHNSQLGMTGE VSFQAANAKSAANLKVKEL*GSG*PRAVIKQESEE	LACE tag^{2M} (underlined), TEV-cleavable His_6-tag (bold), additional residues (italic)
T4L-$LACE_C$	MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSE LDKAIGRNTNGVITKDEAEKLFNQDVDAAVRGILRNAKLKPVYDS LDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAK SRWYNQTPNRAKRVITTFRTGTWDAYKNL*GSG*PRKVIKMESEE	LACE tag (underlined), additional residues (italic)
SpyCatcher with minimal LACE tag	**SYYHHHHHHDYDIPTTENLYFQ**GAMVDTLSG***IK***SEQGQSGDMTIE EDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKD FYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGD AHI	LACE tag (underlined), mutations (bold italic), TEV-cleavable His_6-tag (bold)
Trastuzumab Fab light chain with $LACE\ tag^{2M}$	**MKYLLPTAAAGLLLLAAQPAMA**DIQMTQSPSSLSASVGDRVTITC RASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSRSG TDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAP SVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGN SQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSS PVTKSFNRGEC*GGSGRGSG*PRAVIKQESEE	LACE tag^{2M} (underlined), PelB leader sequence (bold, cleaved during expression), additional residues (italic)
Trastuzumab Fab heavy chain with sortase tag	**MKYLLPTAAAGLLLLAAQPAMA**EVQLVESGGGLVQPGGSLRLSCA ASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVKGRF TISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQG TLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVT VSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYIC NVNHKPSNTKVDKKVEPKSCDKT*GGSGRS*LPETGGHHHHHHV	Sortase-His_6-tag (underlined), PelB leader sequence (bold, cleaved during expression), additional residues (italic)
SUMO2 with minimal LACE tag (K11)	ADEKPKEG***I***KTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKA YCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQT GG	LACE tag (underlined), mutations (bold italic)
α-Syn with minimal LACE tag (K96)	MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTK EGVVHGVATVAEKTKEQVTNVGGAVVTGVTAVAQKTVEGAGSIAA ATGF***I***KK***E***QLG***R***NEEGAPQEGILEDMPVDPDNEAYEMPSEEGYQD YEPEA	LACE tag (underlined), mutations (bold italic)
α-Syn with minimal LACE tag (K96)	MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTK EGVVHGVATVAEKTKEQVTNVGGAVVTGVTAVAQKTVEGAGSIAA ATGFV***R***KDQL***I***KNEEGAPQEGILEDMPVDPDNEAYEMPSEEGYQD YEPEA	LACE tag (underlined), mutations (bold italic)
TNFα-$LACE_N^{3M}$	*MG*PRAVIKQESAE*GGSG*VRSSSRTPSDKPVAHVVANPQAEGQLQW LNRRANALLANGVELRDNQLVVPSEGLYLIYSQVLFKGQGCPSTH VLLTHTISRIAVSYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPI YLGGVFQLEKGDRLSAEINRPDYLDFAESGQVYFGIIAL	LACE tag^{3M} (underlined), additional residues (italic)
AaLS-13-$LACE_C^{3M}$	MEIYEGKLTAEGLRFGIVASRFNHALVGRLVEGAIDCIVRHGGRE EDITLVCVPGSWEIPVAAGELARKEDIDAVIAIGVLIEGAEPHFD YIASEVSKGLANLSLELRKPISFGDITDDELEEAIECAGTEHGNK GWEAALSAIEMANLFKSLRLE*GSGSGA*PRAVIKQESAE**HHHHHH**	LACE tag^{3M} (underlined, His_6-tag (bold), additional residues (italic)

Protein thioesters

Ubiquitin-GyrA-His_6	*G*QIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIF AGKQLEDGRTLSDYNIQKESTLHLVLRLRGG**CITGDALVALPEGE SVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVY TVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAV IQRSAFSVDCAGFARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDA QAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGF VSHALEHHHHHH**	GyrA-His_6 (bold), additional residue (italic)
ISG15-C78A-GyrA-His_6	GWDLTVKMLAGNEFQVSLSSSMSVSELKAQITQKIGVHAFQQRLA VHPSGVALQDRVPLASQGLGPGSTVLLVVDKADEPLSILVRNNKG RSSTYEVRLTQTVAHLKQQVSGLEGVQDDLFWLTFEGKPLEDQLP LGEYGLKPLSTVFMNLRLRGG**CITGDALVALPEGESVRIADIVPG ARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRV TGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDC AGFARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTDG RFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSHALEHHHH HH**	GyrA-His_6 (bold)
Z_{TNF}-GyrA-His_6	GVDNKFNKELGWAIGEIGTLPNLNHQQFRAFILSLWDDPSQSANL LAEAKKLNDAQAPK*GGGGSGGGGSG*PQAIAGQ*GGGGSGGGGS**LRL RGG**CITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHG NPVLADRLFHSGEHPVYTVRTVEGLRVTGTANHPLLCLVDVAGVP TLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVG VPGLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPV YSLRVDTADHAFITNGFVSHALEHHHHHH	MMP1 cleavage site (underlined), Ub-derived C-terminus (bold italic underlined), GyrA-His_6 (bold), additional residues (italic)
Z_{HER2}-GyrA-His_6	VDNKFNKEMRNAYWEIALLPNLNNQQKRAFIRSLYDDPSQSANLL AEAKKLNDAQAPK**LRLRGG**CITGDALVALPEGESVRIADIVPGAR PNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTG TANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAG FARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTDGRF YYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSHALEHHHHHH	Ub-derived C-terminus (bold italic underlined), GyrA-His_6 (bold)

Disulfide tag

CysTag-GFP-His_6	*M*XXCCXX*GSG*RKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGD ATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQH DFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELK GIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNV EDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRD HMVLLEFVTAAGITHGMDELYK*SLE***HHHHHH**	Disulfide tag (underlined), His_6-tag (bold), additional residues (italic)

5. Sequences of library hits

Sequence	Charged (R,H,K,D,E)	Net Charge	Polar (S,T,N)	Apolar (A,L,F,Y,V)	Special (G,P)
ARCCDEGGGRG	3	-1	0	1	0
GSCCEHGGGRG	2	0	1	0	1
EHCCAGGGGRG	2	0	0	1	1
SDCCHFGGGRG	2	0	1	1	0
ERCCHEGGGRG	4	0	0	0	0
YKCCEDGGGRG	3	-1	0	1	0
EPCCALGGGRG	1	-1	0	2	1
DGCCLEGGGRG	2	-2	0	1	1
EHCCLGGGGRG	2	0	0	1	1
ENCCPAGGGRG	1	-1	1	1	1
STCCVEGGGRG	1	-1	2	1	0
GHCCEGGGGRG	2	0	0	0	2
GGCCSEGGGRG	1	-1	1	0	2
KECCKEGGGRG	4	0	0	0	0
SGCCDGGGGRG	1	-1	1	0	2
HPCCAEGGGRG	2	0	0	1	1
EECCKKGGGRG	4	0	0	0	0
EECCKFGGGRG	3	-1	0	1	0
SVCCGEGGGRG	1	-1	1	1	1
KDCCAPGGGRG	2	0	0	1	1
HDCCKPGGGRG	3	1	0	0	1
RDCCHDGGGRG	4	0	0	0	0
HECCTLGGGRG	2	0	1	1	0
ETCCKEGGGRG	3	-1	1	0	0
EHCCSDGGGRG	3	-1	1	0	0
PECCPDGGGRG	2	-2	0	0	2
SGCCDGGGGRG	1	-1	1	0	2
ERCCVEGGGRG	3	-1	0	1	0
NACCHDGGGRG	2	0	1	1	0
SGCCNEGGGRG	1	-1	2	0	1
ERCCDHGGGRG	4	0	0	0	0
ESCCERGGGRG	3	-1	1	0	0
EGCCDRGGGRG	3	-1	0	0	1
EHCCKEGGGRG	4	0	0	0	0
TYCCEHGGGRG	2	0	1	1	0
SFCCLHGGGRG	1	1	1	2	0
RHCCEDGGGRG	4	0	0	0	0
RECCDGGGGRG	3	-1	0	0	1
SGCCLEGGGRG	1	-1	1	1	1
RECCHEGGGRG	4	0	0	0	0
EHCCSAGGGRG	2	0	1	1	0
HACCPDGGGRG	2	0	0	1	1
HHCCNDGGGRG	3	1	1	0	0
RECCDGGGGRG	3	-1	0	0	1
DNCCERGGGRG	3	-1	1	0	0
RSCCEDGGGRG	3	-1	1	0	0
ETCCLHGGGRG	2	0	1	1	0
ETCCDRGGGRG	3	-1	1	0	0
EGCCLHGGGRG	2	0	0	1	1

SGCCLEGGGRG	1	-1	1	1	1
ENCCGDGGGRG	2	-2	1	0	1
DECCARGGGRG	3	-1	0	1	0
EACCHHGGGRG	3	1	0	1	0
ETCCTGGGGRG	1	-1	2	0	1
GHCCVEGGGRG	2	0	0	1	1
AACCLEGGGRG	1	-1	0	3	0
TVCCNEGGGRG	1	-1	2	1	0
EVCCHLGGGRG	2	0	0	2	0
HTCCEVGGGRG	2	0	1	1	0
TACCNDGGGRG	1	-1	2	1	0
NTCCPEGGGRG	1	-1	2	0	1
HSCCDGGGGRG	2	0	1	0	1
EHCCGSGGGRG	2	0	1	0	1
EACCHHGGGRG	3	1	0	1	0
DNCCERGGGRG	3	-1	1	0	0
DTCCDRGGGRG	3	-1	1	0	0
NECCNNGGGRG	1	-1	3	0	0
NHCCTEGGGRG	2	0	2	0	0
ERCCEYGGGRG	3	-1	0	1	0
HECCGAGGGRG	2	0	0	1	1
DNCCLAGGGRG	1	-1	1	2	0
DGCCKDGGGRG	3	-1	0	0	1
DNCCHDGGGRG	3	-1	1	0	0
NNCCHDGGGRG	2	0	2	0	0
GACCHEGGGRG	2	0	0	1	1
ETCCPAGGGRG	1	-1	1	1	1
EGCCVVGGGRG	1	-1	0	2	1
LNCCPDGGGRG	1	-1	1	1	1
HECCLDGGGRG	3	-1	0	1	0
DNCCGHGGGRG	2	0	1	0	1
VHCCHDGGGRG	3	1	0	1	0
KDCCHPGGGRG	3	1	0	0	1
KECCYPGGGRG	2	0	0	1	1
RDCCPTGGGRG	2	0	1	0	1
EGCCVAGGGRG	1	-1	0	2	1
NSCCHDGGGRG	2	0	2	0	0
EHCCSGGGGRG	2	0	1	0	1
KACCEDGGGRG	3	-1	0	1	0
VRCCDEGGGRG	3	-1	0	1	0
ENCCDRGGGRG	3	-1	1	0	0
ENCCVVGGGRG	1	-1	1	2	0
EHCCREGGGRG	4	0	0	0	0
GACCHEGGGRG	2	0	0	1	1
THCCNEGGGRG	2	0	2	0	0
GECCTHGGGRG	2	0	1	0	1
HACCAEGGGRG	2	0	0	2	0
NACCAEGGGRG	1	-1	1	2	0
AECCAAGGGRG	1	-1	0	3	0
ERCCSEGGGRG	3	-1	1	0	0
PACCNEGGGRG	1	-1	1	1	1
HTCCSEGGGRG	2	0	2	0	0

EGCCAVGGGRG	1	-1	0	2	1
ARCCEEGGGRG	3	-1	0	1	0
TTCCNDGGGRG	1	-1	3	0	0
DRCCHDGGGRG	4	0	0	0	0
TECCVVGGGRG	1	-1	1	2	0
SHCCDNGGGRG	2	0	2	0	0
NHCCDGGGGRG	2	0	1	0	1
EYCCKDGGGRG	3	-1	0	1	0
NTCCYEGGGRG	1	-1	2	1	0
ETCCHDGGGRG	3	-1	1	0	0
DACCSKGGGRG	2	0	1	1	0
KLCCEDGGGRG	3	-1	0	1	0
ADCCNAGGGRG	1	-1	0	2	1
KECCVPGGGRG	2	0	0	1	1
GPCCAEGGGRG	1	-1	1	1	1
HACCNEGGGRG	2	0	1	1	0
SNCCNDGGGRG	1	-1	3	0	0
NFCCPDGGGRG	1	-1	1	1	1
DACCKPGGGRG	2	0	0	1	1
EACCERGGGRG	3	-1	0	1	0
HHCCDGGGGRG	3	1	0	0	1
HNCCLEGGGRG	2	0	1	1	0
FACCLAGGGRG	0	0	0	4	0
KECCVPGGGRG	2	0	0	1	1
GTCCGEGGGRG	1	-1	1	0	2
DTCCRDGGGRG	3	-1	1	0	0
TNCCSDGGGRG	1	-1	3	0	0
EGCCAAGGGRG	1	-1	0	2	1
NECCPVGGGRG	1	-1	1	1	1
EECCVRGGGRG	3	-1	0	1	0
EECCRHGGGRG	4	0	0	0	0
PRCCDEGGGRG	3	-1	1	0	0
HSCCAEGGGRG	2	0	1	1	0
GHCCTEGGGRG	2	0	1	0	1
GECCPVGGGRG	1	-1	0	1	2
VECCAHGGGRG	2	0	0	2	0
NLCCPDGGGRG	1	-1	1	1	1
HDCCKHGGGRG	4	2	0	0	0
DNCCLTGGGRG	1	-1	2	1	0
GGCCAEGGGRG	1	-1	0	1	2
VLCCLFGGGRG	0	0	0	4	0
EHCCLAGGGRG	2	0	0	2	0
ESCCHGGGGRG	2	0	1	0	1
PTCCEGGGGRG	1	-1	2	0	1
DRCCAEGGGRG	3	-1	0	1	0
TVCCDGGGGRG	1	-1	1	1	1
DLCCKPGGGRG	2	0	0	1	1
HLCCDGGGGRG	2	0	0	1	1
NDCCRDGGGRG	3	-1	1	0	0
HPCCPDGGGRG	2	0	0	0	2
EKCCPTGGGRG	2	0	1	0	1
DDCCRRGGGRG	4	0	0	0	0

ELCCSKGGGRG	2	0	1	1	0
SDCCKPGGGRG	2	0	1	0	1
NECCRDGGGRG	3	-1	1	0	0
EKCCSPGGGRG	2	0	1	0	1
EKCCERGGGRG	4	0	0	0	0
SDCCKDGGGRG	3	-1	1	0	0
DKCCTPGGGRG	2	0	1	0	1
EKCCPSGGGRG	2	0	1	0	1
VDCCGKGGGRG	2	0	0	1	1
GNCCADGGGRG	1	-1	1	1	1
KTCCKDGGGRG	3	-1	1	0	0
GSCCAFGGGRG	0	0	1	2	1
PNCCPDGGGRG	1	-1	1	0	2
GLCCSDGGGRG	1	-1	1	1	1
KRCCVRGGGRG	3	3	0	1	0
KNCCETGGGRG	2	0	2	0	0
RGCCETGGGRG	2	0	1	0	1
HKCCEHGGGRG	4	2	0	0	0
EKCCGDGGGRG	3	-1	0	0	1
AECCRPGGGRG	2	0	0	1	1
SDCCRPGGGRG	2	0	1	0	1
RDCCGDGGGRG	3	-1	0	0	1
TGCCTDGGGRG	1	-1	2	0	1
KHCCEDGGGRG	4	0	0	0	0
PDCCHRGGGRG	3	1	0	0	1
YKCCSFGGGRG	1	1	1	2	0
DACCKTGGGRG	2	0	1	1	0
VGCCPDGGGRG	1	-1	0	1	2
GACCNDGGGRG	1	-1	1	1	1
EHCCKSGGGRG	3	1	1	0	0
RECCPTGGGRG	2	0	1	0	1
ADCCKDGGGRG	3	-1	0	1	0
SKCCDTGGGRG	2	0	2	0	0
FNCCFHGGGRG	1	1	1	2	0
ARCCEDGGGRG	3	-1	0	1	0
STCCSDGGGRG	1	-1	3	0	0
TDCCHKGGGRG	3	1	1	0	0
DSCCHRGGGRG	3	1	1	0	0
LRCCDDGGGRG	3	-1	0	1	0
YRCCEVGGGRG	2	0	0	2	0
RDCCNPGGGRG	2	0	1	0	1
RACCEDGGGRG	3	-1	0	1	0
RDCCKEGGGRG	4	0	0	0	0
PKCCDTGGGRG	2	0	1	0	1
VKCCTEGGGRG	2	0	1	1	0
NKCCEDGGGRG	3	-1	1	0	0
SACCPDGGGRG	1	-1	1	1	1
ENCCKDGGGRG	3	-1	1	0	0
DLCCRPGGGRG	2	0	0	1	1
TLCCPDGGGRG	1	-1	1	1	1
DKCCADGGGRG	3	-1	0	1	0
YDCCKPGGGRG	2	0	0	1	1

RHCCEHGGGRG	4	2	0	0	0
HHCCVPGGGRG	2	2	0	1	1
KHCCEVGGGRG	3	1	0	1	0
YACCSHGGGRG	1	1	1	2	0
ESCCAAGGGRG	1	-1	1	2	0
DKCCNSGGGRG	2	0	2	0	0
HNCCSDGGGRG	2	0	1	0	1
NDCCHKGGGRG	3	1	1	0	0
KDCCGFGGGRG	2	0	0	1	1
RNCCTEGGGRG	2	0	2	0	0
FLCCTRGGGRG	1	1	1	2	0
GNCCSDGGGRG	1	-1	2	0	1
KSCCTEGGGRG	2	0	2	0	0
KVCCESGGGRG	2	0	1	1	0
VNCCADGGGRG	1	-1	1	2	0
DKCCVTGGGRG	2	0	1	1	0
KECCSDGGGRG	3	-1	1	0	0
KDCCHHGGGRG	4	2	0	0	0
DKCCADGGGRG	3	-1	0	1	0
KECCDKGGGRG	4	0	0	0	0
KFCCEFGGGRG	2	0	0	2	0
DYCCKDGGGRG	3	-1	0	1	0
ERCCTDGGGRG	3	-1	1	0	0
LTCCPDGGGRG	1	-1	1	1	1
NHCCDPGGGRG	2	0	1	0	1
KECCNKGGGRG	3	1	1	0	0
HACCHHGGGRG	3	3	0	1	0
KECCSLGGGRG	2	0	1	1	0
SNCCSDGGGRG	1	-1	3	0	0
ADCCYRGGGRG	2	0	0	2	0
KECCPHGGGRG	3	1	0	0	1
RDCCEKGGGRG	4	0	0	0	0
KDCCVTGGGRG	2	0	1	1	0
RECCGDGGGRG	3	-1	0	0	1
VDCCRPGGGRG	2	0	0	1	1
DKCCHDGGGRG	4	0	0	0	0
ADCCRHGGGRG	3	1	0	1	0
SKCCEGGGGRG	2	0	1	0	1
DKCCHDGGGRG	4	0	0	0	0
EYCCPKGGGRG	2	0	0	1	1
YRCCVPGGGRG	1	1	0	2	1
DRCCLPGGGRG	2	0	0	1	1
TDCCRPGGGRG	2	0	1	0	1
RDCCHVGGGRG	3	1	0	1	0
KDCCKEGGGRG	4	0	0	0	0
GACCSDGGGRG	1	-1	1	1	1
LNCCPPGGGRG	0	0	1	1	2
HDCCPKGGGRG	3	1	0	0	1
HKCCEPGGGRG	3	1	0	0	1
HDCCSRGGGRG	3	1	1	0	0
RDCCTPGGGRG	2	0	1	0	1
SDCCRDGGGRG	3	-1	1	0	0

HDCCNKGGGRG	3	1	1	0	0
DKCCPAGGGRG	2	0	0	1	1
EKCCSPGGGRG	2	0	1	0	1
DNCCRSGGGRG	3	-1	1	0	0
KECCERGGGRG	4	0	0	0	0
NPCCPKGGGRG	1	1	1	0	2
YECCHKGGGRG	3	1	0	1	0
KDCCHVGGGRG	3	1	0	1	0
DHCCKPGGGRG	3	1	0	0	1
EHCCKDGGGRG	4	0	0	0	0
NECCKTGGGRG	2	0	2	0	0
NDCCSRGGGRG	2	0	2	0	0
RECCTTGGGRG	2	0	2	0	0
RECCNSGGGRG	0	0	3	0	1
YDCCKSGGGRG	2	0	1	1	0
EKCCTPGGGRG	2	0	1	0	1
DGCCKSGGGRG	2	0	1	0	1
SECCPKGGGRG	2	0	1	0	1
LNCCDKGGGRG	2	0	1	1	0
VDCCKTGGGRG	2	0	1	1	0
KECCREGGGRG	4	0	0	0	0
KPCCEDGGGRG	3	-1	0	0	1
KECCSRGGGRG	3	1	1	0	0
HECCKAGGGRG	3	1	0	1	0
PVCCHSGGGRG	1	1	1	1	1
HKCCGDGGGRG	3	1	0	0	1
DKCCSVGGGRG	2	0	1	1	0
RGCCSEGGGRG	2	0	1	0	1
ESCCVRGGGRG	2	0	1	1	0
HPCCEKGGGRG	3	1	0	0	1
TSCCPSGGGRG	0	0	3	0	1
ANCCERGGGRG	2	0	1	1	0
EGCCRPGGGRG	2	0	0	0	2
PKCCHDGGGRG	3	1	0	0	1
RGCCSEGGGRG	2	0	1	0	1
KPCCEAGGGRG	2	0	0	1	1
GPCCKEGGGRG	2	0	0	0	2
HSCCREGGGRG	3	1	1	0	0
ANCCDKGGGRG	2	0	1	1	0
KACCEFGGGRG	2	0	0	2	0
HDCCTRGGGRG	3	1	1	0	0
PDCCKSGGGRG	2	0	1	0	1
HRCCTEGGGRG	3	1	1	0	0
RACCDTGGGRG	2	0	1	1	0
KVCCTEGGGRG	2	0	1	1	0
DTCCLKGGGRG	2	0	1	1	0
HRCCEHGGGRG	4	2	0	0	0
KTCCVEGGGRG	2	0	1	1	0
TKCCPEGGGRG	2	0	1	0	1
VSCCVPGGGRG	0	0	1	2	1
DFCCKHGGGRG	3	1	0	1	0
SPCCVPGGGRG	0	0	1	1	2

RSCCDGGGGRG	2	0	1	0	1
DKCCADGGGRG	3	-1	0	1	0
PKCCEVGGGRG	2	0	0	1	1
PECCKAGGGRG	2	0	0	1	1
KDCCLKGGGRG	3	1	0	1	0
DLCCYKGGGRG	2	0	0	2	0
KFCCETGGGRG	2	0	1	1	0
NKCCDHGGGRG	3	1	1	0	0
KPCCGEGGGRG	2	0	0	0	2
APCCKDGGGRG	2	0	0	1	1
DACCARGGGRG	2	0	0	2	0
RECCHGGGGRG	3	1	0	0	1
KTCCESGGGRG	2	0	2	0	0
HDCCKVGGGRG	3	1	0	1	0
RLCCVEGGGRG	2	0	0	2	0
DACCGKGGGRG	2	0	0	1	1
STCCDRGGGRG	2	0	2	0	0
HVCCNPGGGRG	1	1	1	1	1
NKCCVDGGGRG	0	0	2	2	0
PACCREGGGRG	2	0	0	1	1
VACCATGGGRG	0	0	1	3	0
ERCCHAGGGRG	3	1	0	1	0
GHCCEKGGGRG	3	1	0	0	1
DLCCKVGGGRG	2	0	0	2	0
KACCEDGGGRG	3	-1	0	1	0
NKCCGEGGGRG	2	0	1	0	1
GGCCSPGGGRG	0	0	1	0	3
RNCCFAGGGRG	1	1	1	2	0
VRCCSSGGGRG	1	1	2	1	0
RLCCRSGGGRG	2	2	1	1	0
FKCCESGGGRG	2	0	1	1	0
ASCCDKGGGRG	2	0	1	1	0
EGCCRDGGGRG	3	-1	0	0	1
SNCCGSGGGRG	0	0	3	0	1
ERCCPTGGGRG	2	0	1	0	1
AECCTRGGGRG	2	0	1	1	0
DSCCAKGGGRG	2	0	1	1	0
EGCCFYGGGRG	1	-1	0	2	1
RNCCEGGGGRG	2	0	1	0	1
PECCLKGGGRG	2	0	0	1	1
SFCCGSGGGRG	0	0	2	1	1
AKCCAPGGGRG	1	1	0	2	1
KTCCDVGGGRG	2	0	1	1	0
ERCCPKGGGRG	3	1	0	0	1
PDCCKVGGGRG	2	0	0	1	1
ENCCRGGGGRG	2	0	1	0	1
EKCCDTGGGRG	2	0	2	0	0
KLCCSEGGGRG	2	0	1	1	0
RECCSAGGGRG	2	0	1	1	0
ESCCAKGGGRG	2	0	1	1	0
NECCRAGGGRG	2	0	1	1	0
RECCRPGGGRG	3	1	0	0	1

TRCCDTGGGRG	2	0	2	0	0
RECCSGGGGRG	2	0	1	0	1
RDCCPRGGGRG	3	1	0	0	1
DACCKTGGGRG	2	0	1	1	0
NDCCRGGGGRG	2	0	1	0	1
EFCCPKGGGRG	2	0	0	1	1
AVCCKEGGGRG	2	0	0	2	0
HKCCEAGGGRG	3	1	0	1	0
PTCCKDGGGRG	2	0	1	0	1
PECCRTGGGRG	2	0	1	0	1
VDCCKAGGGRG	2	0	0	2	0
ARCCDTGGGRG	2	0	1	1	0
TDCCLKGGGRG	2	0	1	1	0
DRCCDRGGGRG	4	0	0	0	0
DRCCGPGGGRG	2	0	0	0	2
AECCRAGGGRG	2	0	0	2	0
STCCERGGGRG	2	0	2	0	0
ERCCHGGGGRG	3	1	0	0	1
FECCKHGGGRG	3	1	0	1	0
DVCCKAGGGRG	2	0	0	2	0
VNCCNHGGGRG	1	1	2	1	0
PDCCRDGGGRG	3	-1	0	0	1
DVCCTRGGGRG	2	0	1	1	0
EACCLKGGGRG	2	0	0	2	0
EGCCKAGGGRG	2	0	0	1	1
EKCCAVGGGRG	2	0	0	2	0
KLCCEPGGGRG	2	0	0	1	1
HFCCEKGGGRG	3	1	0	1	0
NHCCKTGGGRG	2	2	2	0	0
APCCPKGGGRG	1	1	0	1	2
ATCCDHGGGRG	2	0	1	1	0
NNCCDPGGGRG	1	-1	2	0	1
KTCCHEGGGRG	3	1	1	0	0
KPCCEPGGGRG	2	0	0	0	2
DTCCLKGGGRG	2	0	1	1	0
FSCCFYGGGRG	0	0	1	3	0
TKCCEVGGGRG	2	0	1	1	0
GNCCSRGGGRG	1	1	2	0	1
RNCCEHGGGRG	3	1	1	0	0
LDCCLKGGGRG	2	0	0	2	0
VKCCEAGGGRG	2	0	0	2	0
ADCCKVGGGRG	2	0	0	2	0
EFCCKPGGGRG	2	0	0	1	1
NKCCDHGGGRG	3	1	1	0	0
DKCCPVGGGRG	2	0	0	1	1
KFCCESGGGRG	2	0	1	1	0
NVCCSKGGGRG	1	1	2	1	0
YPCCYLGGGRG	0	0	0	3	1
NPCCTTGGGRG	0	0	3	0	1
RNCCEDGGGRG	3	-1	1	0	0

6. Python script for analysis of peptide sequencing data

6.1. Script

```
#Empty lists and variables, constants
startFiles = 0
startSequences = 0
spotList = []
fileList = []
spotListSingle = []
indicesList = []
sequenceLines=[]
massDifferencesExp =
[512.07119,495.04469,478.01819,421.06201,404.03551,330.05283,182.01836,164.99186,147.96536
,108.03568,91.00918,73.98268,56.95618,34.053,17.0265]
precursorMassAccuracy = 0.6
pepScoreCutOff = 25
mProt = 1.00728
mCam = 57.02146
positive = ["R","H","K"]
negative = ["D","E"]
charged = ["R","H","K","D","E"]
polar = ["S","T","N"]
apolar = ["A","L","F","Y","V"]
special = ["G","P"]

#Open and create files
file1 = open("Mascot_output.txt", "r")
file2 = open("Curated_results.txt", "w")

#Write title line
file2.write("Rules for automated identification of sequences:"+'\n')
file2.write('\t'+">Most confident sequence and second-most confident sequence from a
different raw file (if present) have identical sequence"+'\n')
file2.write('\t'+">Most confident sequence has a peptide score >= 25"+'\n')
file2.write('\t'+'\n')
file2.write("Spot"+'\t'+"Raw File"+'\t'+"Sequence Name"+'\t'+"Mass Check
MH+"+'\t'+"Sequence"+'\t'+"Protein Score"+'\t'+"Modifications"+'\t'+"Peptide
Rank"+'\t'+"Peptide Score"+'\t'+"Experimental MZ"+'\t'+"Experimental MR"+'\t'+"Calculated
MR"+'\t'+"Delta"+'\t'+"Charged (R,H,K,D,E)"+'\t'+"Net Charge"+'\t'+"Polar
(S,T,N)"+'\t'+"Apolar (A,L,F,Y,V)"+'\t'+"Special (G,P)"+'\n'+'\t'+'\n')

#Determine MALDI plate spot in each file
for line in file1:
    currentline = line.split(",")
    if currentline[0] == '_DISTILLER_MDRO_VERSION':
        startFiles = startFiles + 1
    if startFiles == 1 and currentline[0] != '_DISTILLER_MDRO_VERSION':
        if len(currentline) >= 2:
            positionSpot = (currentline[1].find("MSMS"))
            if positionSpot != -1:
                spotString = ""
                if currentline[1][positionSpot-5] == "\\":
                    for i in range (1,4):
                        spotString = spotString + currentline[1][positionSpot-5+i]
                else:
                    for i in range (1,3):
```

```
                        spotString = spotString + currentline[1][positionSpot-4+i]
                spotList.append(spotString)
#Determine RawFile Names in each file
            positionFile = (currentline[0].find("RAWFILE"))
            if positionFile != -1:
                fileString = ""
                for i in range (positionFile+8,len(currentline[0])):
                    if currentline[0][i] != "]":
                        fileString = fileString + currentline[0][i]
                fileList.append(fileString)
#Create workable list of all lines with Sequencing details
    if currentline[0] == 'prot_hit_num':
        startSequences = startSequences + 1
    if startSequences == 1 and currentline[0] != 'prot_hit_num':
        sequenceLines.append(currentline)

#Iterate through MALDI plate spots to find unique spots
for i in range (len(spotList)):
    if spotList[i] not in spotListSingle:
        spotListSingle.append(spotList[i])

#Iterate through MALDI plate spots to find all RawFiles that belong to that spot
for i in range (len(spotListSingle)):
    indices = [j for j, x in enumerate(spotList) if x == spotListSingle[i]] #gives all
indices where this spot occurs
    indicesList.append(indices)

#Go through MALDI plate spots, search sequence lines whether they belong to a file
associated with this spot, print the line to file if peptide rank is 1 or 2
for i in range (len(spotListSingle)):
    counter = 0 #to find out if there was no match. In case of no match, write an empty
new line into file
    peptideScoreBest = 0 #to determine most confident sequencing results from all files
that belong to this spot
    peptideRankBest = 0
    mostConfidentSeq = ""
    rawFileBest = ""
    precursorList = [] #add all precursors selected per spot to this list. Then compare
deltaMW to identify spots that contained multiple beads
    deltaPrecursor = [] #list of all delta between precursors to compare to expected
values for precursor masses stemming from a single peptide
    precursorOffCount = 0 #for every precursor which does not match the expected mass
difference pattern, counter is increased by one
    for j in range (len(indicesList[i])):
        for k in range (len(sequenceLines)):
            positionFileSeq = (sequenceLines[k][28].find("file"))
            fileStringSeq = ""
            for l in range (positionFileSeq+6,len(sequenceLines[k][28])-4):
                    fileStringSeq = fileStringSeq + sequenceLines[k][28][l]
            if fileStringSeq == fileList[indicesList[i][j]]:
                if sequenceLines[k][13] not in precursorList:
                    precursorList.append(sequenceLines[k][13])
                if sequenceLines[k][10] in ["1","2"]:
                    counter = counter + 1
                    if float(sequenceLines[k][19]) > float(peptideScoreBest):
                        peptideScoreBest = float(sequenceLines[k][19])
                        mostConfidentSeq = sequenceLines[k][22]
                        peptideRankBest = float(sequenceLines[k][10])
```

```
                    rawFileBest = fileStringSeq
                if sequenceLines[k][24] =='"2 Carbamidomethyl (C)"':
                    mCalcH = mProt + float(sequenceLines[k][16])
                else:
                    mCalcH = ""
                #determine peptide properties
                overallCharge = 0
                chargedAA = 0
                polarAA = 0
                apolarAA = 0
                specialAA = 0
                for m in (0,1,4,5):
                    if sequenceLines[k][22][m] in positive:
                        overallCharge = overallCharge + 1
                    if sequenceLines[k][22][m] in negative:
                        overallCharge = overallCharge -1
                    if sequenceLines[k][22][m] in charged:
                        chargedAA = chargedAA + 1
                    if sequenceLines[k][22][m] in polar:
                        polarAA = polarAA + 1
                    if sequenceLines[k][22][m] in apolar:
                        apolarAA = apolarAA + 1
                    if sequenceLines[k][22][m] in special:
                        specialAA = specialAA + 1

file2.write(str(spotListSingle[i])+'\t'+str(fileStringSeq)+'\t'+str(sequenceLines[k][2])+'
\t'+str(mCalcH)+'\t'+str(sequenceLines[k][22])+'\t'+str(sequenceLines[k][3])+'\t'+str(sequ
enceLines[k][24])+'\t'+str(sequenceLines[k][10])+'\t'+str(sequenceLines[k][19])+'\t'+str(s
equenceLines[k][13])+'\t'+str(sequenceLines[k][14])+'\t'+str(sequenceLines[k][16])+'\t'+st
r(sequenceLines[k][17])+'\t'+str(chargedAA)+'\t'+str(overallCharge)+'\t'+str(polarAA)+'\t'
+str(apolarAA)+'\t'+str(specialAA)+'\n')
    if peptideScoreBest != 0: #print the most confident sequencing result belonging to
this spot
        file2.write("most
confident:"+'\t'+str(rawFileBest)+'\t'+""+'\t'+""+'\t'+str(mostConfidentSeq)+'\t'+""+'\t'+
""+'\t'+str(peptideRankBest)+'\t'+str(peptideScoreBest)+'\n')
    #determining the raw file which provides the second most confident sequencing result
    peptideScoreSecondBest = 0
    peptideRankSecondBest = 0
    secondMostConfidentSeq = ""
    rawFileSecondBest = ""
    for j in range (len(indicesList[i])):
        for k in range (len(sequenceLines)):
            positionFileSeq = (sequenceLines[k][28].find("file"))
            fileStringSeq = ""
            for l in range (positionFileSeq+6,len(sequenceLines[k][28])-4):
                    fileStringSeq = fileStringSeq + sequenceLines[k][28][l]
            if fileStringSeq == fileList[indicesList[i][j]]:
                if sequenceLines[k][10] in ["1","2"]:
                    if float(sequenceLines[k][19]) > float(peptideScoreSecondBest):
                        if fileStringSeq != rawFileBest:
                            peptideScoreSecondBest = float(sequenceLines[k][19])
                            secondMostConfidentSeq = sequenceLines[k][22]
                            peptideRankSecondBest = float(sequenceLines[k][10])
                            rawFileSecondBest = fileStringSeq
    if peptideScoreSecondBest != 0: #print the most confident sequencing result belonging
to this spot
```

```
        file2.write("second most
confident:"+'\t'+str(rawFileSecondBest)+'\t'+""+'\t'+""+'\t'+str(secondMostConfidentSeq)+'
\t'+""+'\t'+""+'\t'+str(peptideRankSecondBest)+'\t'+str(peptideScoreSecondBest)+'\n')
    #finish up this spot by adding a new line. Also add extra new line if no files have
been associated with this spot (counter == 0)
    if mostConfidentSeq == secondMostConfidentSeq and mostConfidentSeq != "" and
peptideScoreBest >= pepScoreCutOff:
        file2.write(""+'\t'+""+'\t'+""+'\t'+""+'\t'+"IDENTIFIED"+'\n')
    elif mostConfidentSeq == secondMostConfidentSeq and mostConfidentSeq != "" and
peptideScoreBest < pepScoreCutOff:
        file2.write(""+'\t'+""+'\t'+""+'\t'+""+'\t'+"NOT IDENTIFIED (score < 25)"+'\n')
    elif secondMostConfidentSeq == "" and mostConfidentSeq != "" and peptideScoreBest >=
pepScoreCutOff:
        file2.write(""+'\t'+""+'\t'+""+'\t'+""+'\t'+"IDENTIFIED"+'\n')
    elif secondMostConfidentSeq == "" and mostConfidentSeq != "" and peptideScoreBest <
pepScoreCutOff:
        file2.write(""+'\t'+""+'\t'+""+'\t'+""+'\t'+"NOT IDENTIFIED (score < 25)"+'\n')
    elif mostConfidentSeq != secondMostConfidentSeq and mostConfidentSeq != "":
        file2.write(""+'\t'+""+'\t'+""+'\t'+""+'\t'+"NOT IDENTIFIED (different
sequences)"+'\n')

#Determine if there are multiple beads present based on deltaMW from precursors collected
in the list precursorList
    for q in range (len(precursorList)-1):
        for v in range (q+1,len(precursorList)):
            delta = abs(float(precursorList[q])-float(precursorList[v]))
            deltaPrecursor.append(delta)
    for w in range(len(deltaPrecursor)):
        foundMatch = 0
        for y in range(len(massDifferencesExp)):
            if (massDifferencesExp[y]-precursorMassAccuracy) <= deltaPrecursor[w] <=
(massDifferencesExp[y]+precursorMassAccuracy):
                foundMatch = foundMatch + 1
        if foundMatch == 1:
            precursorOffCount = precursorOffCount
        else:
            precursorOffCount = precursorOffCount + 1
    if precursorOffCount == 0:
        file2.write(""+'\t'+""+'\t'+""+'\t'+""+'\t'+"single bead"+'\n')
    elif precursorOffCount != 0:
        file2.write(""+'\t'+""+'\t'+""+'\t'+""+'\t'+"evtl. multiple beads"+'\n')
    if counter == 0:
        file2.write(str(spotListSingle[i])+'\n')
    file2.write('\t'+'\n')
    counter = 0 #to find out if there was no match. In case of no match, write an empty
new line into file
    for j in range (len(indicesList[i])): #write out all the lower ranks too below
        for k in range (len(sequenceLines)):
            positionFileSeq = (sequenceLines[k][28].find("file"))
            fileStringSeq = ""
            for l in range (positionFileSeq+6,len(sequenceLines[k][28])-4):
                    fileStringSeq = fileStringSeq + sequenceLines[k][28][l]
            if fileStringSeq == fileList[indicesList[i][j]]:
                counter = counter + 1
                if sequenceLines[k][24] =='"2 Carbamidomethyl (C)"':
                    mCalcH = mProt + float(sequenceLines[k][16])
                else:
                    mCalcH = ""
```

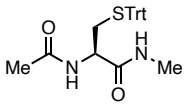
```
                #Determine peptide properties
                overallCharge = 0
                chargedAA = 0
                polarAA = 0
                apolarAA = 0
                specialAA = 0
                for m in (0,1,4,5):
                    if sequenceLines[k][22][m] in positive:
                        overallCharge = overallCharge + 1
                    if sequenceLines[k][22][m] in negative:
                        overallCharge = overallCharge -1
                    if sequenceLines[k][22][m] in charged:
                        chargedAA = chargedAA + 1
                    if sequenceLines[k][22][m] in polar:
                        polarAA = polarAA + 1
                    if sequenceLines[k][22][m] in apolar:
                        apolarAA = apolarAA + 1
                    if sequenceLines[k][22][m] in special:
                        specialAA = specialAA + 1
                if sequenceLines[k][10] not in ["1","2"]:

file2.write(str(spotListSingle[i])+'\t'+str(fileStringSeq)+'\t'+str(sequenceLines[k][2])+'
\t'+str(mCalcH)+'\t'+str(sequenceLines[k][22])+'\t'+str(sequenceLines[k][3])+'\t'+str(sequ
enceLines[k][24])+'\t'+str(sequenceLines[k][10])+'\t'+str(sequenceLines[k][19])+'\t'+str(s
equenceLines[k][13])+'\t'+str(sequenceLines[k][14])+'\t'+str(sequenceLines[k][16])+'\t'+st
r(sequenceLines[k][17])+'\t'+str(chargedAA)+'\t'+str(overallCharge)+'\t'+str(polarAA)+'\t'
+str(apolarAA)+'\t'+str(specialAA)+'\n')
    if counter == 0:
        file2.write(str(spotListSingle[i])+'\n')

    file2.write('\t'+'\n')

#Close files
file1.close()
file2.close()
```

6.2. Short output example

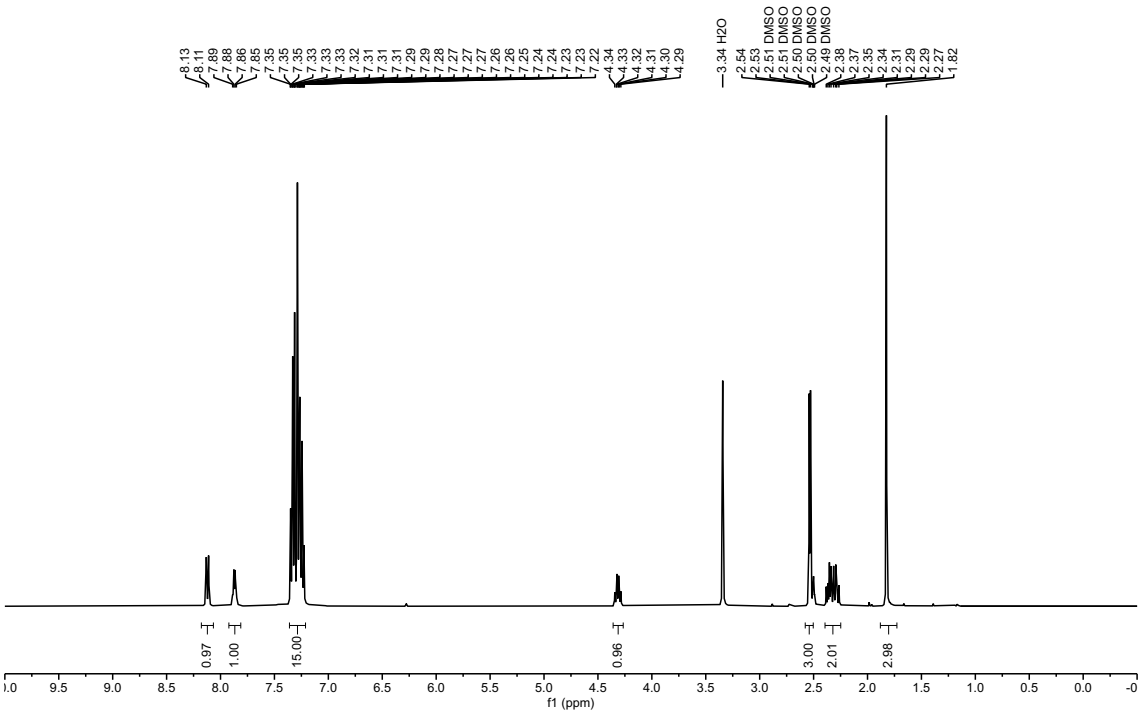
Rules for automated identification of sequences:																			
	>Most confident sequence and second-most confident sequence from a different raw file (if present) have identical sequence																		
	>Most confident sequence has a peptide score >= 25																		
Spot	Raw File	Sequence Name	Mass Check MH+	Sequence	Protein Score	Modifications	Peptide Rank	Peptide Score	Experimental MZ	Experimental MR	Calculated MR	Delta	Charged (R,H,K,D,E)	Net Charge	Polar (S,T,N)	Apolar (A,L,F,Y,V)	Special (G,P)
D10	0	SEQUENCE_46900 mH:1374.54672 mNa:1396.52866 mK:1412.5026		YYCCHFGGGRG	45	Carbamidomethyl (C); Cys->Dap (C)	1	8.19	1300.352	1299.3451	1299.5567	-0.2116	1	1	0	3	0
D10	0	SEQUENCE_46942 mH:1374.54672 mNa:1396.52866 mK:1412.5026		YYCCFHGGGRG	45	Carbamidomethyl (C); Cys->Dap (C)	1	8.19	1300.352	1299.3451	1299.5567	-0.2116	1	1	0	3	0
D10	1	SEQUENCE_46935 mH:1317.58277 mNa:1339.56471 mK:1355.53865	1317.58268	YYCCKVGGGRG	56	2 Carbamidomethyl (C)	1	56.49	1318.399	1317.3921	1316.5754	0.8167	1	1	0	3	0
D10	1	SEQUENCE_46816 mH:1317.55762 mNa:1339.53956 mK:1355.5135	1317.55748	YYCCRAGGGRG	54	2 Carbamidomethyl (C)	2	54.02	1318.399	1317.3921	1316.5502	0.8419	1	1	0	3	0
D10	4	SEQUENCE_47009 mH:1416.54606 mNa:1438.528 mK:1454.50194	1416.54598	YYCCYYGGGRG	28	2 Carbamidomethyl (C)	1	27.85	1416.368	1415.3605	1415.5387	-0.1782	0	0	0	4	0
most confident:	1			YYCCKVGGGRG			1	56.49					1	1	0	3	0
second most confident:	4			YYCCYYGGGRG			1	27.85									
				NOT IDENTIFIED (different sequences)													
				evtl. multiple beads		deltaMW of Precursors: ['18.0470', '116.0154', '97.9684']											
D11	5	SEQUENCE_22654 mH:1210.48411 mNa:1232.46605 mK:1248.43999		HPCCPDGGGRG	121	Carbamidomethyl (C); Cys->Dha (C)	1	46.22	1119.387	1118.3798	1118.4676	-0.0878	2	0	0	0	2
D11	6	SEQUENCE_22654 mH:1210.48411 mNa:1232.46605 mK:1248.43999		HPCCPDGGGRG	121	Carbamidomethyl (C); Cys->Dap (C)	1	66.52	1136.419	1135.4118	1135.4942	-0.0823	2	0	0	0	2
D11	6	SEQUENCE_20854 mH:1227.47428 mNa:1249.45622 mK:1265.43016		HNCCPDGGGRG	113	Carbamidomethyl (C); Cys->Dha (C)	1	66.52	1136.419	1135.4118	1135.4578	-0.0459	2	0	1	0	1
D11	7	SEQUENCE_22654 mH:1210.48411 mNa:1232.46605 mK:1248.43999	1210.48408	HPCCPDGGGRG	121	2 Carbamidomethyl (C)	1	52.39	1210.408	1209.4004	1209.4768	-0.0764	2	0	0	0	2
D11	7	SEQUENCE_22556 mH:1210.48411 mNa:1232.46605 mK:1248.43999	1210.48408	HPCCDPGGGRG	101	2 Carbamidomethyl (C)	2	51.56	1210.408	1209.4004	1209.4768	-0.0764	2	0	0	0	2
most confident:	6			HPCCPDGGGRG			1	66.52					2	0	0	0	2
second most confident:	7			HPCCPDGGGRG			1	52.39									
				IDENTIFIED													
				single bead													
D24	62	SEQUENCE_15476 mH:1205.51508 mNa:1227.49702 mK:1243.47096		EKCCSPGGGRG	146	Carbamidomethyl (C); Cys->Dap (C)	1	68.71	1131.481	1130.4732	1130.5251	-0.0519	2	0	1	0	1
D24	62	SEQUENCE_15462 mH:1205.51508 mNa:1227.49702 mK:1243.47096		EKCCPSGGGRG	114	Carbamidomethyl (C); Cys->Dap (C)	1	68.71	1131.481	1130.4732	1130.5251	-0.0519	2	0	1	0	1
D24	64	SEQUENCE_15476 mH:1205.51508 mNa:1227.49702 mK:1243.47096	1205.51498	EKCCSPGGGRG	146	2 Carbamidomethyl (C)	2	49.23	1205.475	1204.4672	1204.5077	-0.0405	2	0	1	0	1
D24	64	SEQUENCE_15462 mH:1205.51508 mNa:1227.49702 mK:1243.47096	1205.51498	EKCCPSGGGRG	114	2 Carbamidomethyl (C)	1	58.76	1205.475	1204.4672	1204.5077	-0.0405	2	0	1	0	1
most confident:	62			EKCCSPGGGRG			1	68.71					2	0	1	0	1
second most confident:	64			EKCCPSGGGRG			1	58.76									
				NOT IDENTIFIED (different sequences)													
				single bead													

7. NMR spectra

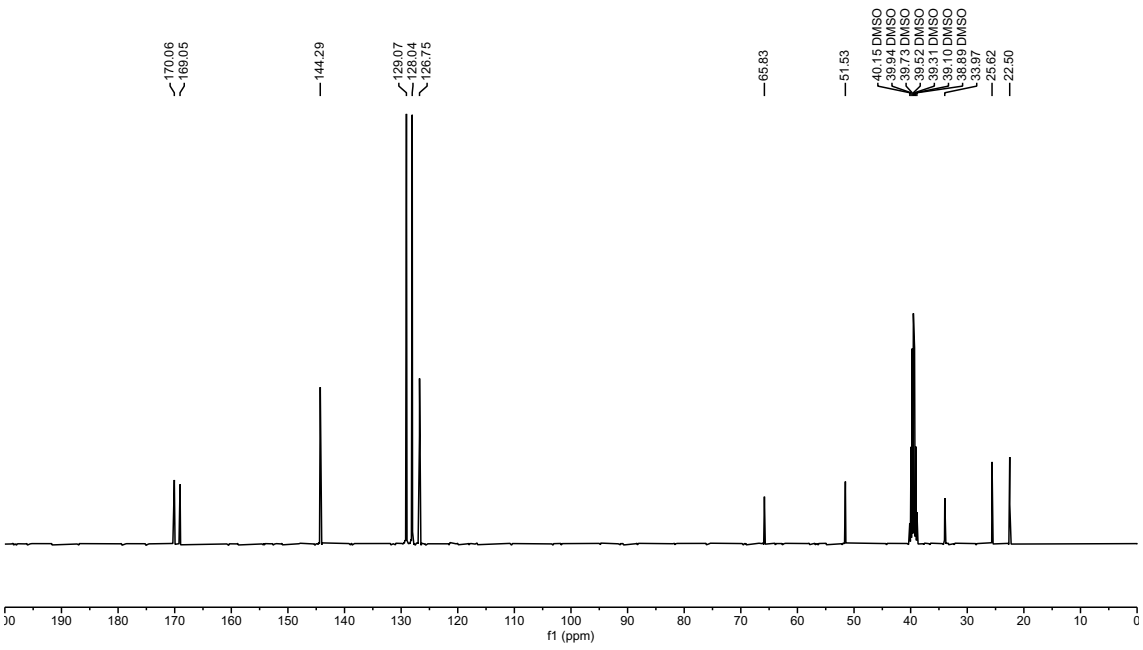
Ac-Cys(Trt)-NHMe (S1)



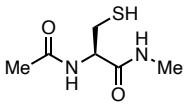
^{1}H NMR (400 MHz, DMSO-d_6)



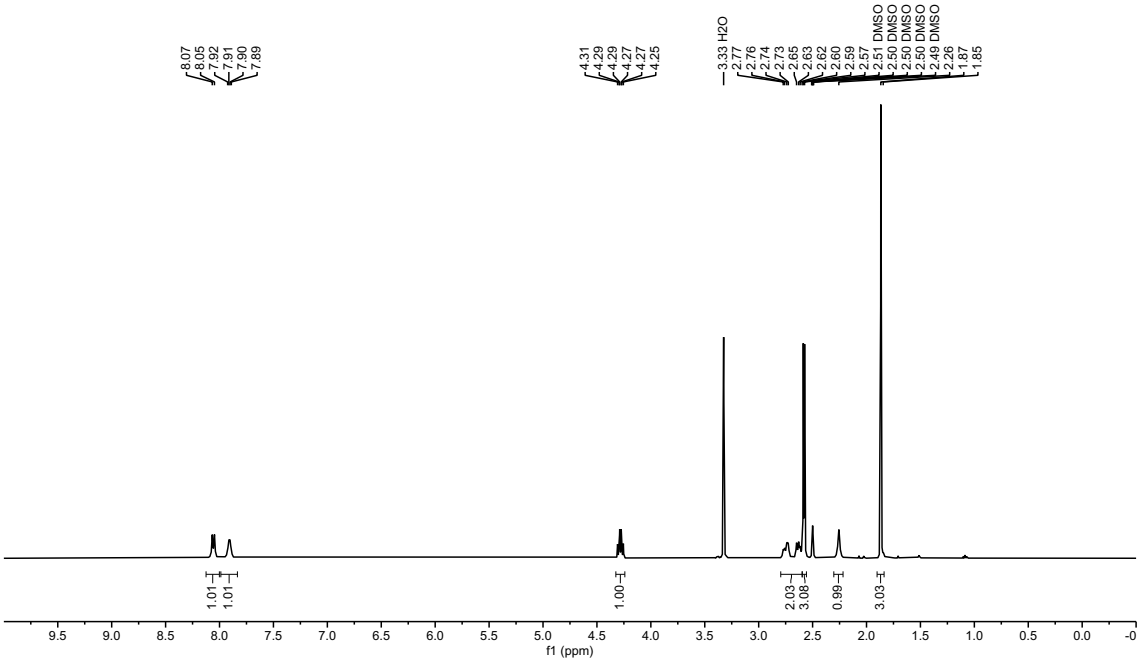
^{13}C NMR (101 MHz, DMSO-d_6)



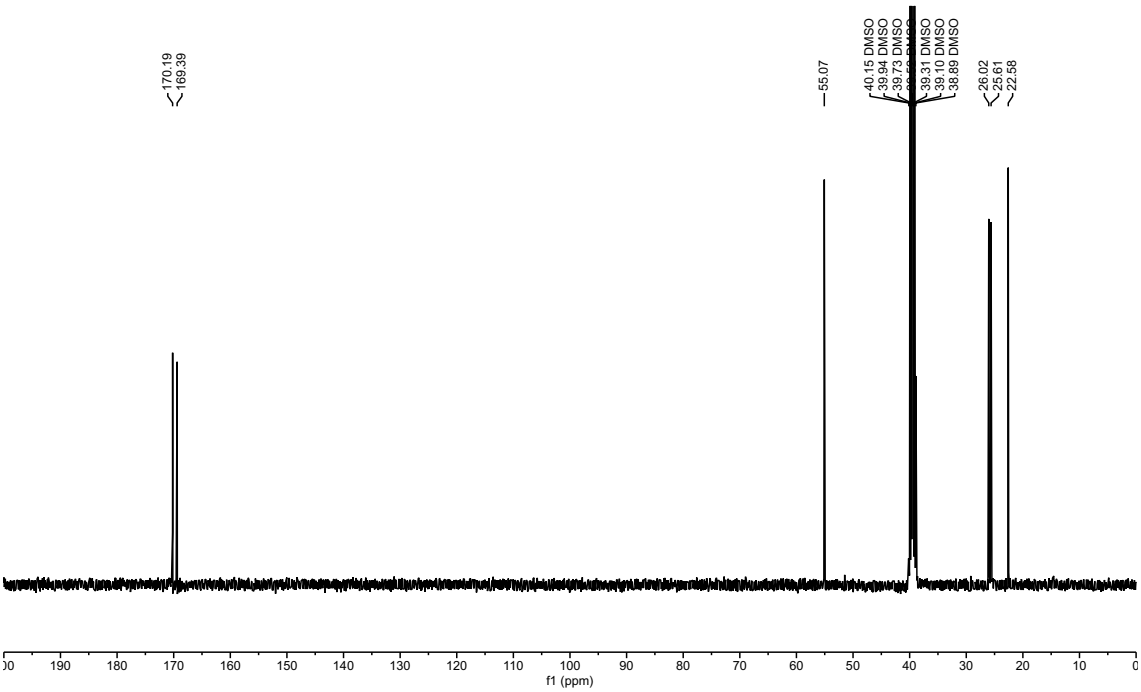
Ac-Cys-NHMe (1)



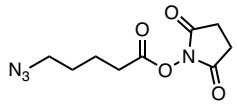
^{1}H NMR (400 MHz, DMSO-d_6)



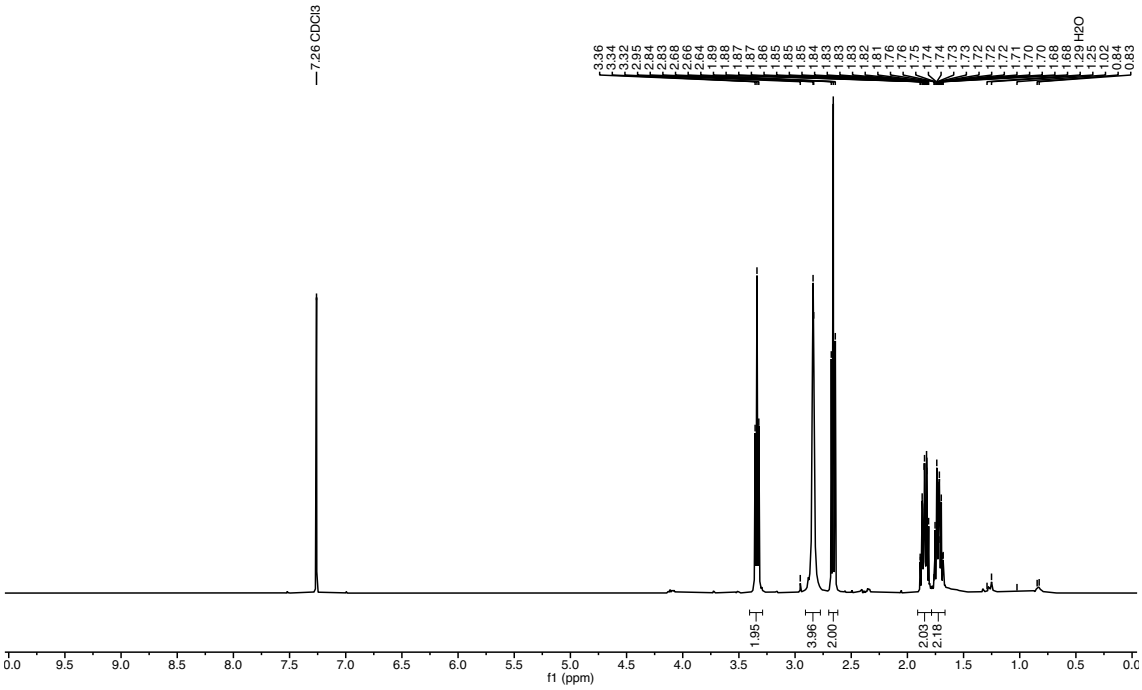
^{13}C NMR (101 MHz, DMSO-d_6)



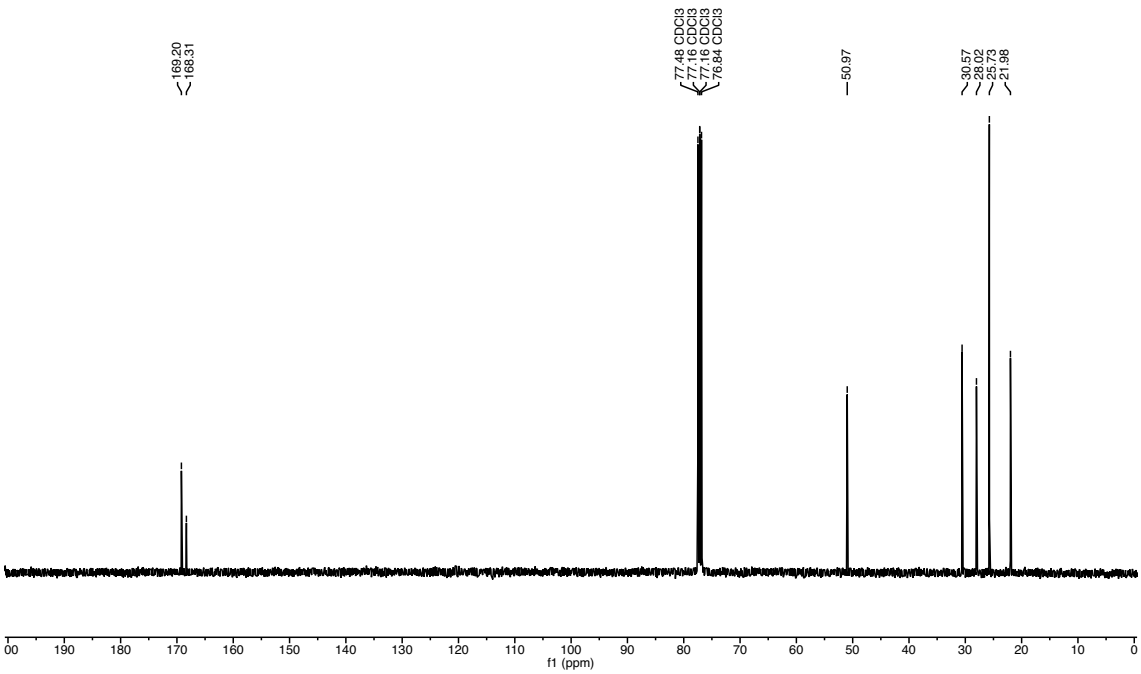
2,5-dioxopyrrolidin-1-yl 5-azidopentanoate (27)



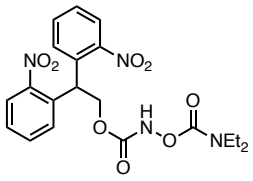
^{1}H NMR (400 MHz, $CDCl_3$)



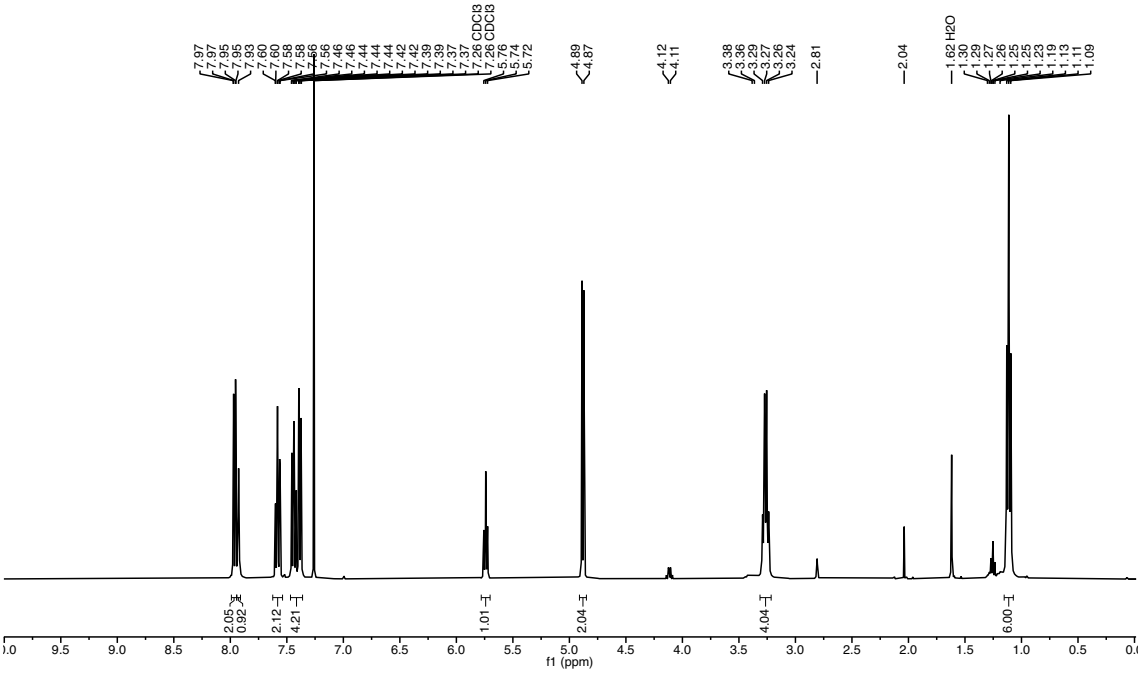
^{13}C NMR (101 MHz, $CDCl_3$)



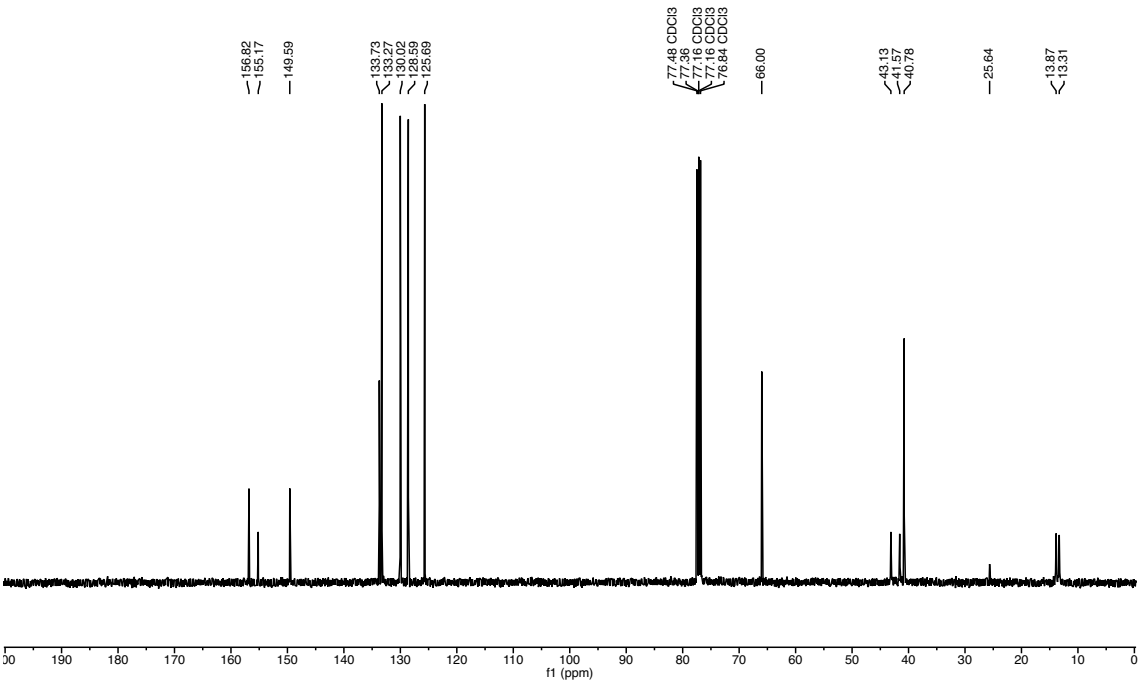
2,2-bis(2-nitrophenyl)ethyl ((diethylcarbamoyl)oxy)carbamate (S4)



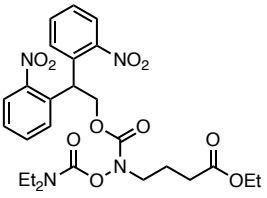
^{1}H NMR (400 MHz, $CDCl_3$)



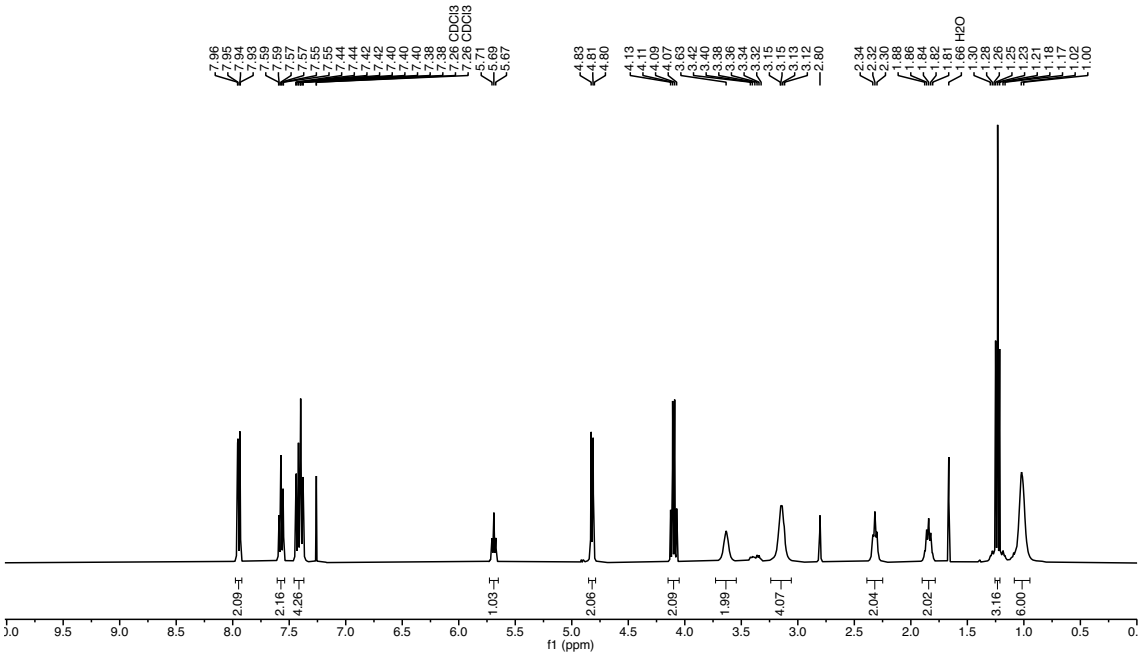
^{13}C NMR (101 MHz, $CDCl_3$)



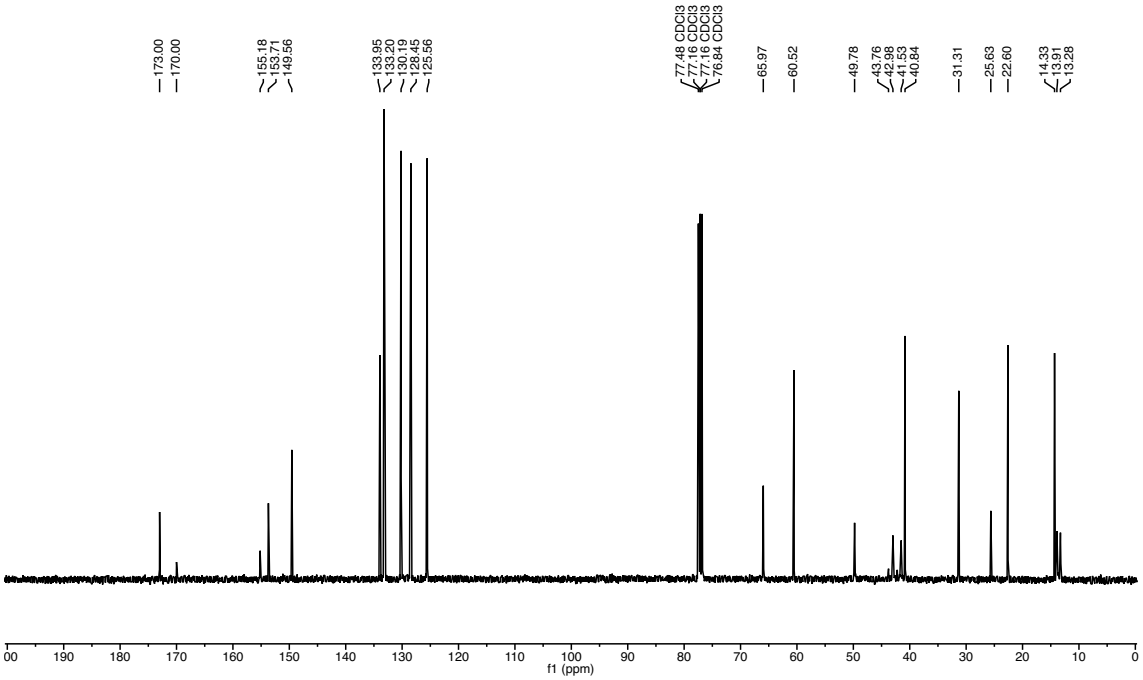
ethyl 4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)oxy)amino)butanoate (S5)



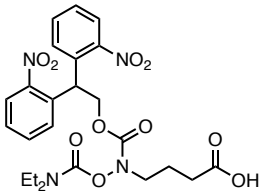
^{1}H NMR (400 MHz, $CDCl_3$)



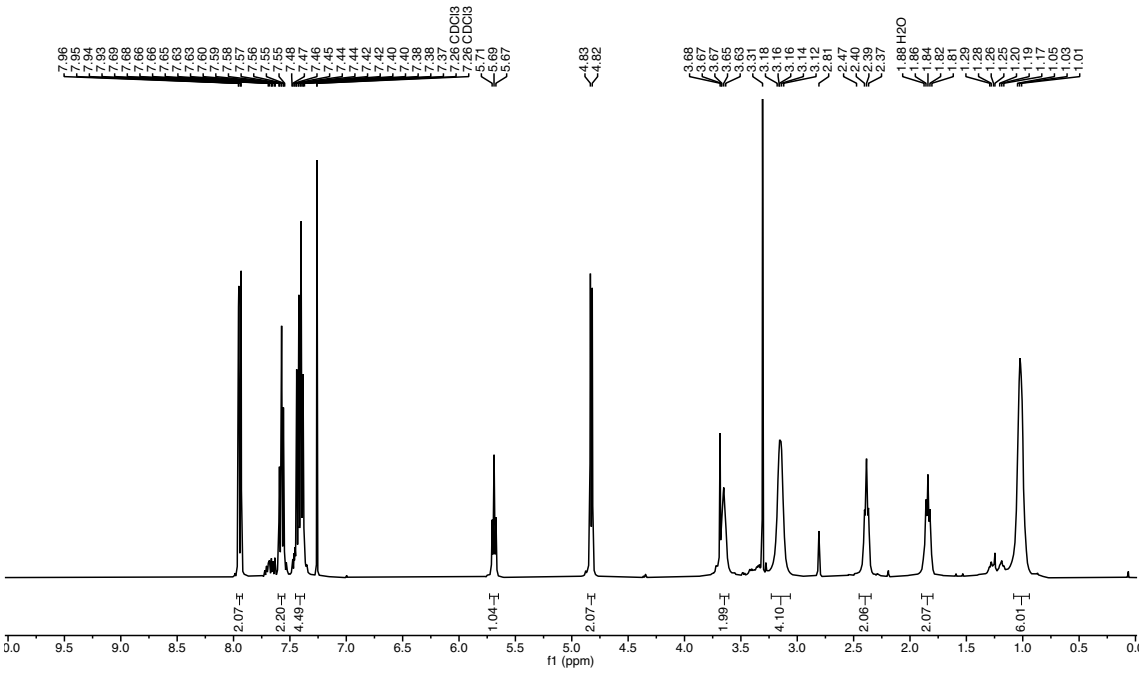
^{13}C NMR (101 MHz, $CDCl_3$)



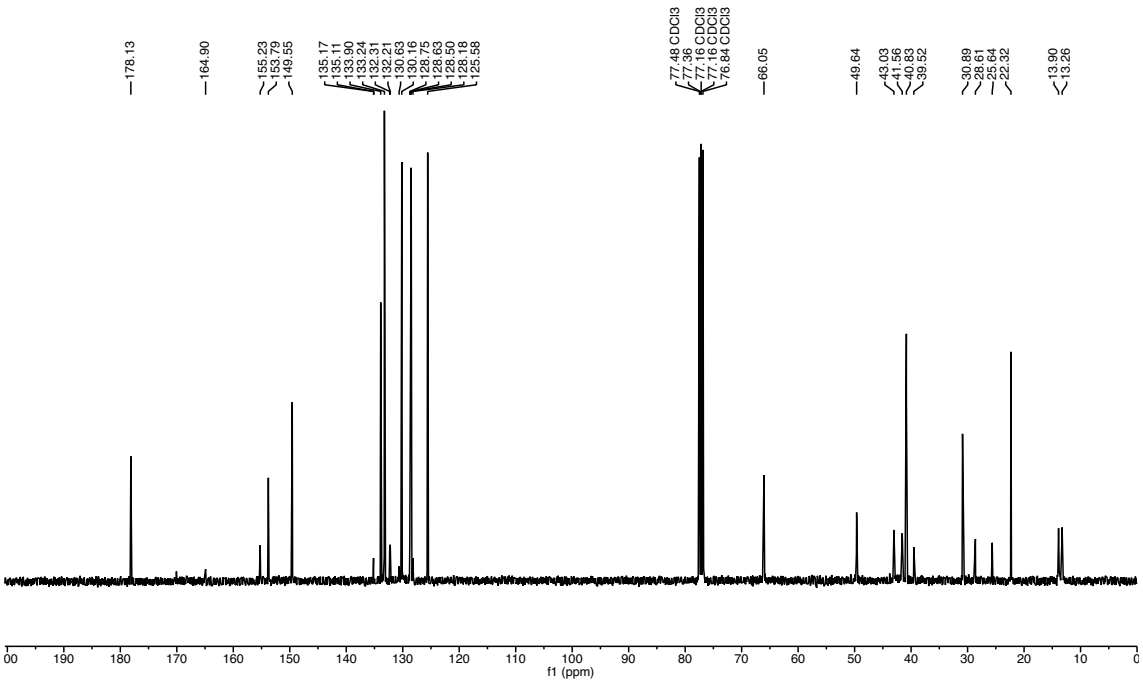
4-(((2,2-bis(2-nitrophenyl)ethoxy)carbonyl)((diethylcarbamoyl)-oxy)amino)-butanoic acid (S2)



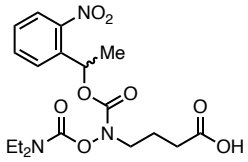
^{1}H NMR (400 MHz, $CDCl_3$)



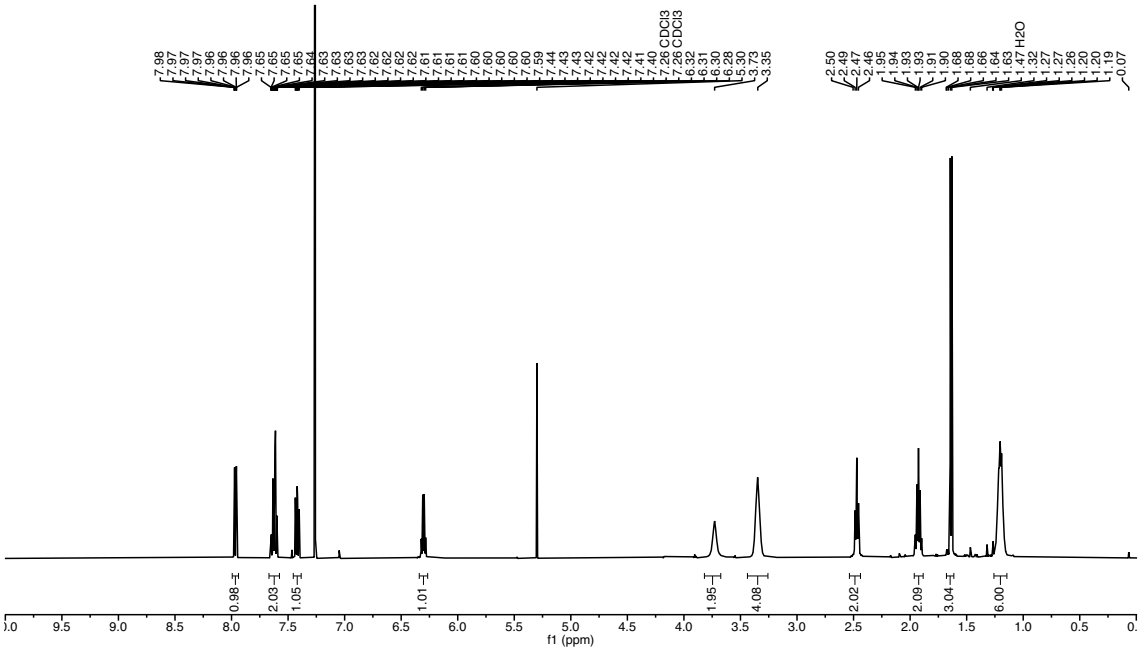
^{13}C NMR (101 MHz, $CDCl_3$)



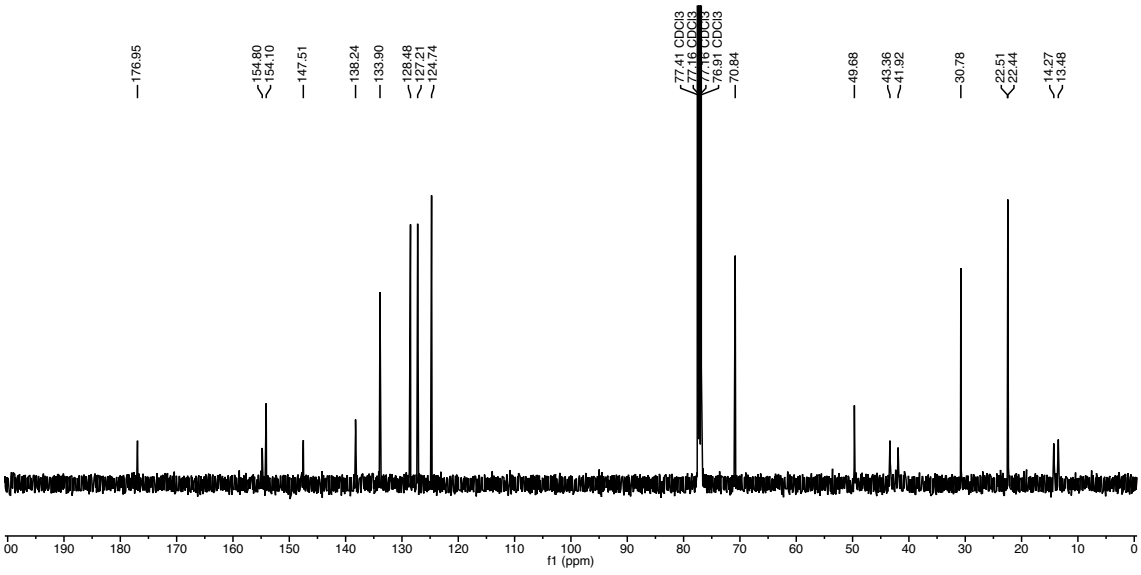
Synthesis of 4-(((diethylcarbamoyl)oxy)((1-(2-nitrophenyl)ethoxy)-carbonyl)amino)-butanoic acid (S6)



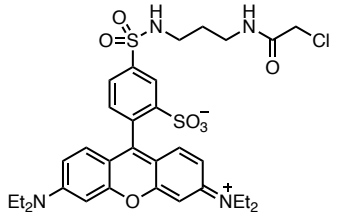
^{1}H NMR (500 MHz, $CDCl_3$)



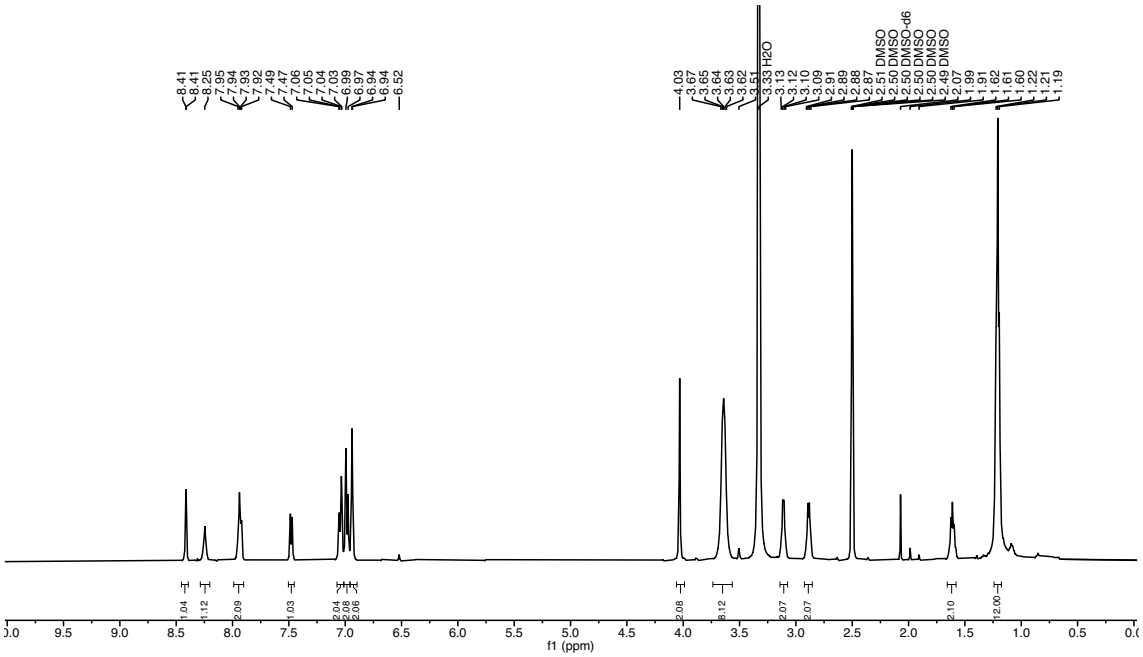
^{13}C NMR (126 MHz, $CDCl_3$)



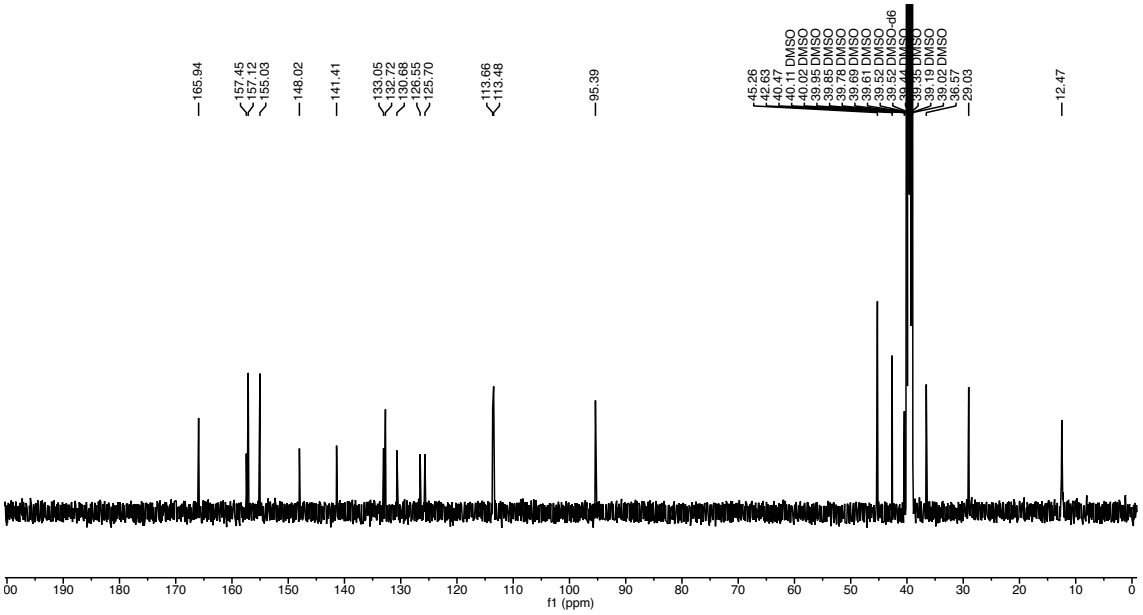
5-(*N*-(3-(2-chloroacetamido)propyl)sulfamoyl)-2-(6-(diethylamino)-3-(diethyliminio)-3*H*-xanthen-9-yl)benzenesulfonate (sulforhodamine chloroacetamide 46)



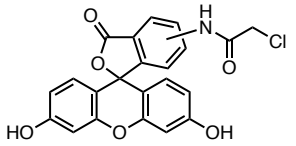
^{1}H NMR (500 MHz, DMSO-d_6)



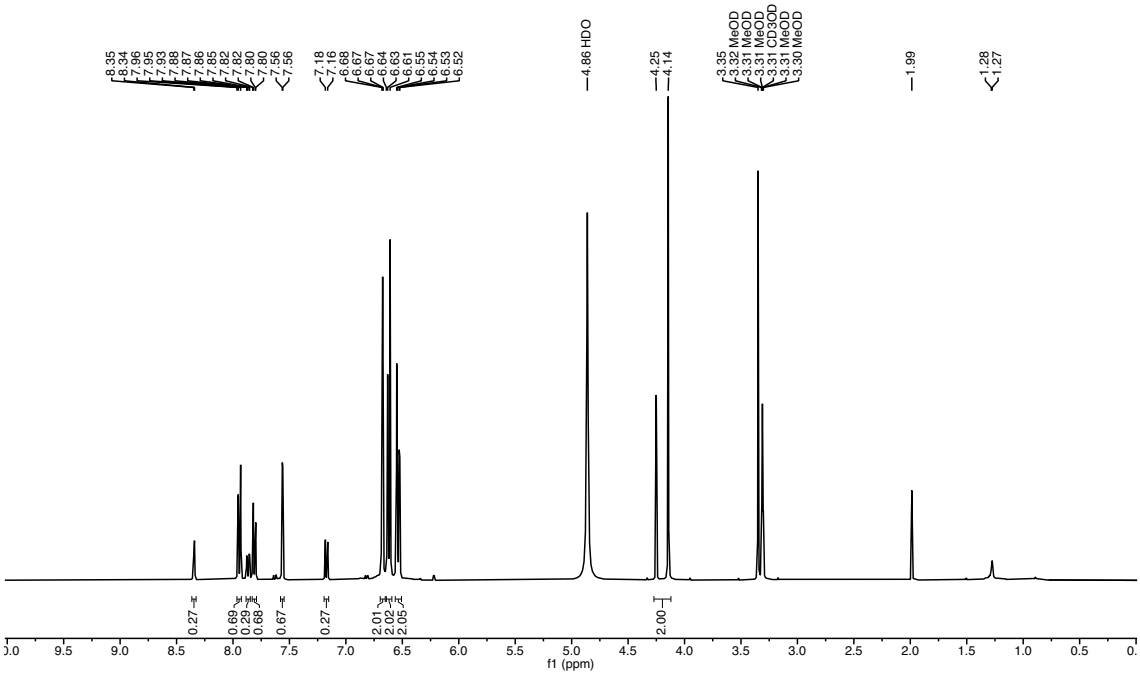
^{13}C NMR (126 MHz, DMSO-d_6)



2-chloro-*N*-(3',6'-dihydroxy-3-oxo-3*H*-spiro[isobenzofuran-1,9'-xanthen]-5(6)-yl)acetamide
(5(6)-(chloroacetamido)fluorescein 47, mixed isomers)



^{1}H NMR (400 MHz, MeOD-d_4)



^{13}C NMR (101 MHz, MeOD-d_4)