Chemical strategies for investigation of deubiquitinases

Samuel D. Whedon

A dissertation

submitted in partial fulfillment of the

requirements of the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Champak Chatterjee, Chair

Michael H. Gelb

Pradipsinh K. Rathod

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Chemical strategies for investigation of deubiquitinases

Samuel D. Whedon

Chair of the Supervisory Committee:

Dr. Champak Chatterjee

Chemistry

Regulation of protein structure and function by post-translational modification is a key mechanism in cellular homeostasis. Among known modifications the small protein ubiquitin is unique in the breadth of functions it directs. Regulation of protein ubiquitylation is the function of more than 600 ligases, and ~100 deubiquitinases. Dysregulated ubiquitylation is involved in infection, inflammation, neurodegeneration, metabolic syndromes and cancer. Therapeutic intervention in these conditions benefits from characterization of substrate-specific ligases and deubiquitinases. Substrate specificity is documented among deubiquitinases, but lags behind knowledge of ligases.

In the interest of characterizing specificity of cysteine protease deubiquitinases we have developed chemical methods for site-specific ubiquitylation and electrophile incorporation. We began by developing a ubiquitin-derived electrophile through installation of a C-terminal selenocysteine residue, and orthogonal oxidative conversion to dehydroalanine. Upon validation of the electrophile we expanded our substrate scope to a ubiquitylated peptide, with which we

captured the deubiquitinase USP15. In order to access ubiquitylated lysine residues in protein regions inaccessible by native chemical ligation we undertook the synthesis of two selenazolidine amino acids for amber suppression.

While conducting amber suppression selections we concurrently pursued semi-synthesis of p53, in order to introduce a C-terminal ubiquitin electrophile. Purification of the semi-synthetic protein proved materially intensive, prompting development of cleavable affinity handles for ligation product purification. A strategy for incorporation at glycine and incorporation at glutamine were developed. A glycine-derived biotin handle proved amenable to reduction by Zn, which furnishes a native glycine, and aromatic thiols, which formed novel site-specific conjugates at the previously modified glycine. Incorporation of a pendant flag tag at glutamine enabled immunoprecipitation followed by traceless tag removal with Zn.

Acknowledgements

I am deeply grateful to the community that has seen me through these past years of study. First and foremost, my family has been a constant inspiration. The kindness, warmth and generosity they show has been my sustenance. I hope to keep their example ever at the front of my mind, along with a reminder to call more often.

I thank my friends for their patience and remarkable steadfastness. It is a special caliber of person that meets you after midnight to pass the hours between time points. So too those who often told me simply that they were in town, as I could never commit to plans a week in advance. For the endless outreach (never badgering) to someone so lost in the moment, I am happily indebted.

My coworkers have been a daily joy. I have learned so much from all of you, and delighted in your company. The thoughtful collaboration, wise criticism, and resilient good humor have been essential. I count myself lucky to have shared so much of your company.

To my mentor, Champak, I owe more than I suspect I fully grasp. Your enthusiasm is infectious, and has carried me through a few bleak patches. Your relentlessness in pursuit of the truth, and challenge to weak arguments, has been instrumental in my development. I am tremendously thankful to have been given so much room to learn through exploration, and I plan to make good use of it.

Table of Contents

List of Figures, Schemes, and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 1: Introduction to ubiquitylation and protein semi-synthesis

1.1 Protein post-translational modifications

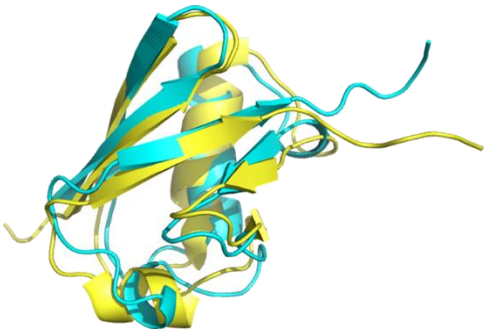
The twenty canonical amino acids exhibit a breadth of chemical properties, and their polymerization achieves a multitude of folds and functions. Non-covalent intermolecular interactions regulate protein function within the cell, but are not the sole regulatory mechanism. In the last half century post-translational modification (PTM) has been identified as a second mechanism by which protein activity is regulated.[1] Reversible chemical modification of amino acid side chains alters charge and steric properties and thereby protein structure, enzymatic activity, multimerization, and stability.

A few proteins associated with gene expression, such as transcription factors and histones, have tens of modifiable residues that enable coordination of diverse regulatory functions. Enzymatic regulation of the signals deposited at these hubs is critical to homeostasis. Some regulatory PTMs, like acetylation, are deposited extensively and nonspecifically, while others are tightly regulated.[2,3] Ubiquitin (Ub), a 76 amino acid protein, is a PTM subject to tight control by >600 conjugating enzymes and ~100 deconjugating enzymes.

A major function of ubiquitylation is marking proteins for degradation; pairing degradation with constitutive expression enables maintenance of a poised state, as in the case of the stress response directing transcription factor p53. Upon monoubiquitylation by MDM2 p53 undergoes first tetramer destabilization, then relocalization to the cytosol.[4] In the cytosol p53 is prevented from acting as a cell cycle arresting transcription factor and may be polyubiquitylated and degraded. Alternatively, p53 may be returned to the nucleus following deubiquitylation by one of several enzymes each of which responds to a different stressor.[5] The distinction between poly- and mono-ubiquitylation is the polymerization state of Ub, which enables signaling functions far beyond degradation.

1.2 Ubiquitin-like modifiers

The ubiquitin-like modifiers (Ubls) are short polypeptides the C-termini of which can act as donors in the acyl modification of amino acid side chains. Ubiquitylation occurs most often as a Lys ε amide isopeptide, and less frequently as an N-terminal α amide, Cys thioester, or Ser/Thr ester. In addition to a sharing the role of modifier, the Ubls are characterized by a common β grasp fold, and C-terminal glycine required for effective isopeptide conjugation and deconjugation (**Figure 1**). Ub is highly conserved across eukaryotes, and serves signaling functions in protein degradation, receptor endocytosis, DNA repair, activation and repression of transcription, and immune regulation among others. The repertoire of Ub functions is linked to the polymerization state of Ub, which arises from condensation of one Ub onto one of seven lysine residues (K6, K11, K27, K29, K33, K48, K63) or the N-terminus (M1) of another molecule of Ub (**Table 1**).



Ubiquitin	EDGRTLSDYNIQKESTLHLVLRLR**GG**	76
URM1	WELLGELDYQLQDQDSVLFISTLH**GG**	101
SUMO1	ADNHTPKELGMEEEDVIEVYQEQT**GG**	96
SUMO2	NETDTPAQLEMEDEDTIDVFQQQT**GG**	93
SUMO3	NETDTPAQLEMEDEDTIDVFQQQT**GG**	92
FAT10	EDGKMMADYGIRKGNLLFLACYCI**GG**	165
ISG15	EDQLPLGEYGLKPLSTVFMNLRLR**GG**	157
Nedd8	NDEKTAADYKILGGSVLHLVLALR**GG**	76

Figure 1.1. Homology of ubiquitin-like modifiers. (Left) Superimposition of Ubiquitin (cyan; pdb:1UBQ) and SUMO3 (yellow; pdb:1U4A) β grasp crystal structures. (Right) Sequence alignment of select human Ubls with C-terminal Gly-Gly motif bolded.[83,84]

The diversity of the Ub signal is not a general feature of the Ubls, which exhibit few polymeric linkage types and lower polymer abundance *in vivo*. The Ubls exhibit overlapping and inter-related activities, such that the interplay of these modifications imparts greater regulatory nuance to each signal. For example, Nedd8 shares 58% sequence identity with Ub and can be substituted for Ub in cells under proteasome inhibition; furthermore, the substitution of Nedd8

does not require expression of a separate set of enzymes, it arises from innate cross-reactivity in the conjugation machinery, which is mirrored in the deconjugation machinery (UCH-L3, UCH-L3, Otubain-1 and Ataxin3 exhibit deNeddylating activity).[6] Nedd8 also plays ubiquitin-adjacent roles, such as modification and activation of the Cullin E3 Ub ligases, and competitive modification of p53, which, unlike Ub, maintains the nuclear localization of p53, while recruiting Nub1 to induce p53 ubiquitylation.[7,8]

Table 1.1. Ubiquitin chain linkage functions.

Chain linkage	Functions	Conformation[85]
Mono	Receptor endocytosis, Chromatin structure[86]	n/a
Multi-mono	Receptor endocytosis, Lysosomal degradation	n/a
M1	NF-κB signaling[22]	open
K6	DNA damage response, Mitophagy[17,18]	compact
K11	Cell cycle related degradation[87]	compact
K27	DNA damage response[88], Protein aggregation[89]	open
K29	Protein aggregation[89]	open
K33	Golgi trafficking[90]	compact
K48	Proteasomal degradation, Replisome disassembly[91]	compact
K63	Endocytosis, DNA damage response[25], NF-κB signaling[22]	open
Branched	Degradation[92]	

The second best characterized of the Ubls is the Small Ubiquitin-like Modifier (SUMO), which was first identified as a nuclear pore targeting modification of Ran GTPase-activating protein 1.[9] Consistent with initial reports SUMOylation often directs proteins to the nucleus, and within the nucleus decorates a variety of transcription factors and chromatin associated proteins.[10,11] SUMOylation plays what appears to be an isoform dependent role in the cellular response to stress wherein SUMO1 conjugation predominates in the absence of stress, while SUMO2/3 prevail under conditions of hypoxia, heat shock, osmotic imbalance, and genotoxicity, and SUMO4 has only been observed under conditions of serum starvation.[12]

Cross-talk between SUMOylation and ubiquitylation takes a variety of forms including co-polymerization, antagonism, and recruitment. Where the relationship is antagonistic SUMO

serves as a non-degradative signal, blocking lysine residues where ubiquitylation is known to precede proteasomal degradation.[13] The activity of SUMO Targeted Ub Ligases (STUbLs) illustrates SUMO-dependent recruitment of ubiquitylation, and can give rise to mixed Ubl chains in which Ub caps poly-SUMO chains. The regulatory feedback enabled by this system is evident in STUbL recruitment to poly-SUMOylation signals at the site of DNA damage, which results in local ubiquitylation that both scaffolds the DNA damage response and targets SUMOylating enzymes for degradation thereby silencing the source of the STUbL recruitment signal.[14,15]

An emerging source of complexity in both the SUMO and Ub signals arises from PTM of the Ubls. Phosphorylation of SUMO1 at Ser2 has been detected in proteomic analysis of human and yeast SUMOylomes, though the precise function is unclear.[16] Ub undergoes lysine acetylation and phosphorylation at T14, S20, S57 and S65, of which S65 plays a key role in mitophagy by both stimulating the E3 ligase PARKIN and inhibiting the activity of the deubiquitylating enzyme (DUB) USP30. The localized antagonism between PARKIN and USP30 regulates deposition of K6 polyUb on mitochondrial outer membrane proteins in healthy mitochondria, while Ub phosphorylation by PINK signals mitochondrial instability and pushes the equilibrium in favor of polyUb chain elongation and mitophagy.[17,18] The PINK/PARKIN/USP30 system exemplifies the regulatory role of ubiquitin signaling, and the function of conjugating and deconjugating enzymes in dictating which signal is conveyed to the rest of the cell.

1.3 Ubiquitylation & Deubiquitylation

Conjugation of Ub is performed by the consecutive action of the E1, E2, and E3 enzymes (**Figure 2a**). The E1 Ub activating enzymes, two in humans, hydrolyze ATP to produce a short lived Ub-adenylate species, which is then transferred to a catalytic cysteine reside yielding a Ub-E1 thioester.[19] The E1 transfers Ub to one of ~40 E2 Ub conjugating enzymes by a transthioesterification reaction, again employing an activated cysteine residue.[20] The site specificity in ubiquitylation is conferred by the E3 enzyme, which binds the E2 and either positions

it adjacent to a substrate lysine residue (RING) or again transthioesterifies the activated Ub species and then conjugates it to a substrate lysine (HECT, RBR).[21]

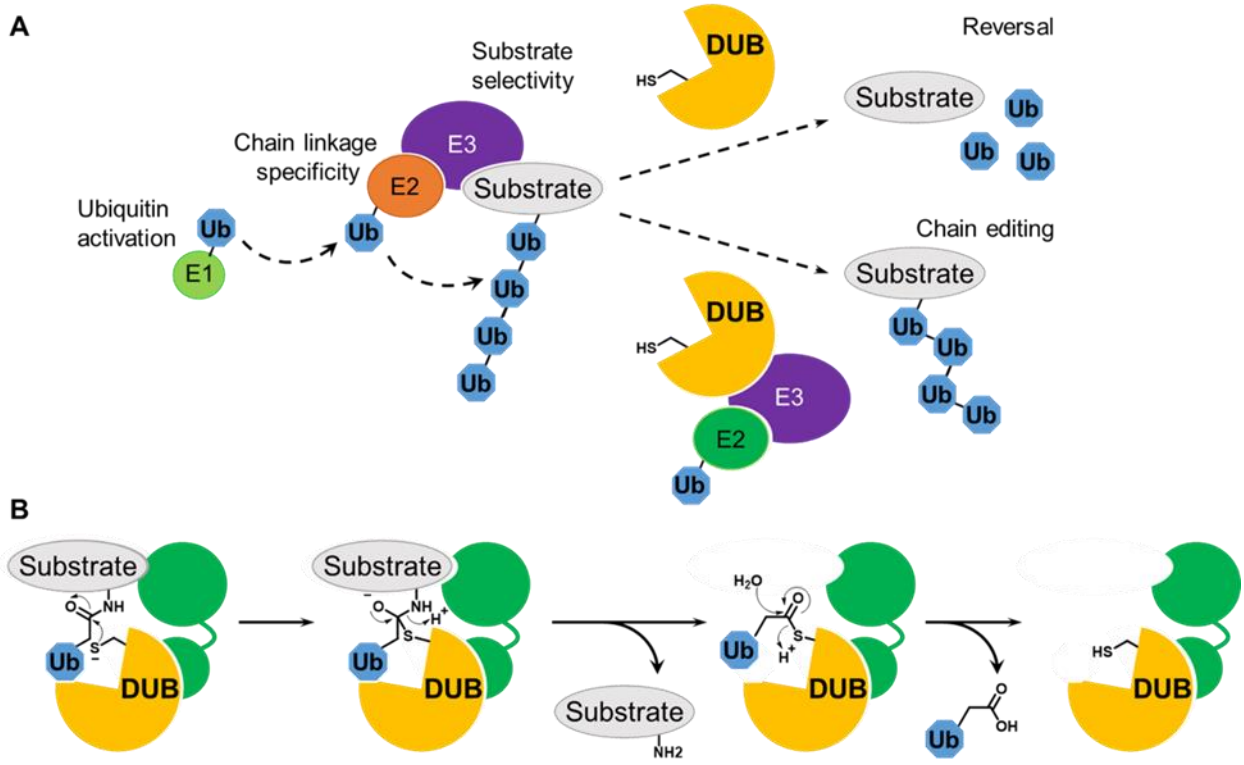


Figure 1.2. Ubiquitylation and Deubiquitylation. (**A**) An activated E1 Ub thioester undergoes transthiolation by an E2, which transfers the Ub to a growing polyUb chain on the substrate. The polyubiquitylated substrate can undergo deubiquitylation by a DUB, or Ub chain editing by a DUB-E3-E2 complex. (**B**) Mechanism of deubiquitylation by a cysteine protease DUB. The isopeptide undergoes nucleophilic attack by the catalytic thiolate, after which the tetrahedral intermediate collapses cleaving the isopeptide bond. The thioester formed following isopeptide cleavage is hydrolyzed, and the thiolate leaving group is protonated

Specificity in Ub chain assembly arises from the concerted action of E2s like Ube2S (K11), Ube2R1 and R2 (K48), Ube2G1 (K48) and Ube2K (K48), which preferentially assemble specific chain linkages, and E3s that position cognate E2s adjacent to a single Lys residue on the growing ubiquitin chain.[20] The activity of E2-E3 complexes can depend on multiple scaffolding proteins, as in the case of the Linear Ub Assembly Complex (LUBAC), in which HOIL1L and SHARPIN scaffold and enhance the activity of the E3 HOIP and an E2. As with USP30 and PARKIN, LUBAC

exists in close association with at least one cognate DUB, OTULIN; as LUBAC assembles an M1 linked Ub chain on the NF-κB essential modifier (NEMO) OTULIN binds to the N-terminus of HOIP and antagonizes ubiquitylation of NEMO.[22] Thus co-localization of E3 and DUB appears to be a significant mechanism by which substrate ubiquitylation is maintained in an equilibrium until an additional signal tips that equilibrium in favor of signal transduction. The co-association between MDM2 and USP7 regulates the ubiquitylation state of p53, and stabilizes MDM2 against its autoubiquitylating activity, which USP7 antagonizes.[23]

The deconjugation half of the ubiquitylation balance is the function of seven families of DUBs (**Figure 2b**). The Jab1/Mov34/MPN family (JAMMs) are metalloproteases that cleave Ub isopeptides with a water molecule activated by a His/Asp chelated Zn^{2+} ion. Six DUB families are papain-like cysteine proteases characterized by Cys, His, Asp/Asn catalytic triads and a Gln hydrogen bond donor that acts as an oxyanion-hole.[23] The cysteine protease families are dubbed Ub carboxy-terminal hydrolase (UCH), Ub specific protease (USP), ovarian tumor domain (OTU), Machado-Josephin disease (Josephin/MJD), motif interacting with Ub-containing novel DUB family (MINDY), and zinc finger with UFM1 specific peptidase domain (ZUFSP).[24,25] While the papain-like DUBs share a catalytic mechanism, they differ in the folds that position their catalytic residues. The UCHs generally processes short or unfolded Ub C-terminal amides due to an unstructured loop that limits access to the active site, though the proteasome associated UCH-37 can process polyUb due to a structural rearrangement that occurs on proteasome binding.[26,27] The USP family members exhibit different polyUb linkage specificities, and high variability in the insertions between the six conserved boxes of their catalytic domain, which are postulated to account for their substrate specificity.[28] The OTU family members exhibit significant, though different, linkage specificities, which appears associated with the numerous Ub-binding auxiliary domains common to the family (e.g. ZnF, UBA, UIM, Ubl).[29] The MJD, MINDY and ZUFSP families are less extensively characterized, though each exhibits polyUb chain specificity—K48 & K63,

K48, and K63 respectively. The ongoing characterization of DUB mechanism and specificity is dependent on biochemical probes to assay their activity, many of which are produced by protein semi-synthetic methods described in *sections 1.4* and *1.5*.

1.4 Chemical preparation of homogenously modified proteins

The precise function of a PTM can be difficult to parse as each is associated with multiple phenotypes and multi-protein complexes. A quantitative characterization of PTM contributions to structure and function evaluates the change in binding or *in vitro* enzyme kinetics relative to the wild type protein, which requires homogenously modified protein. Immunoprecipitation enables isolation of modified proteins but, even where a site-specific antibody exists, there is not yet a method for removing species bearing additional modifications without tremendous attrition in yield. Moreover, the majority of modified proteins lack corresponding specific antibodies. Proteomic analyses of ubiquitylated proteins have often identified multiple modified substrate lysine residues, but the functional significance of a modification site requires a homogenous sample. This issue has clinical significance, as demonstrated in the differing aggregation propensities of α synuclein species ubiquitylated or SUMOylated at Lys residues separated by as few as one amino acid, homogenous samples of which were accessible only by semi-synthesis.[30]

The simplest strategy for investigation of a modification is genetic replacement with a near mimic, as Gln for ε acetylated Lys, Asp for phosphorylated Ser, or Arg for constitutively deacetylated Lys. Where a more representative modification mimic is desired mutation of the modification site to Cys enables alkylation, directed disulfide formation, and thiol-ene functionalization.[31,32] Furthermore, elimination of Cys to dehydroalanine (Dha) enables nucleophilic substitution or cross-coupling at the site, at the cost of stereochemistry (**Figure 3**).[33–35]

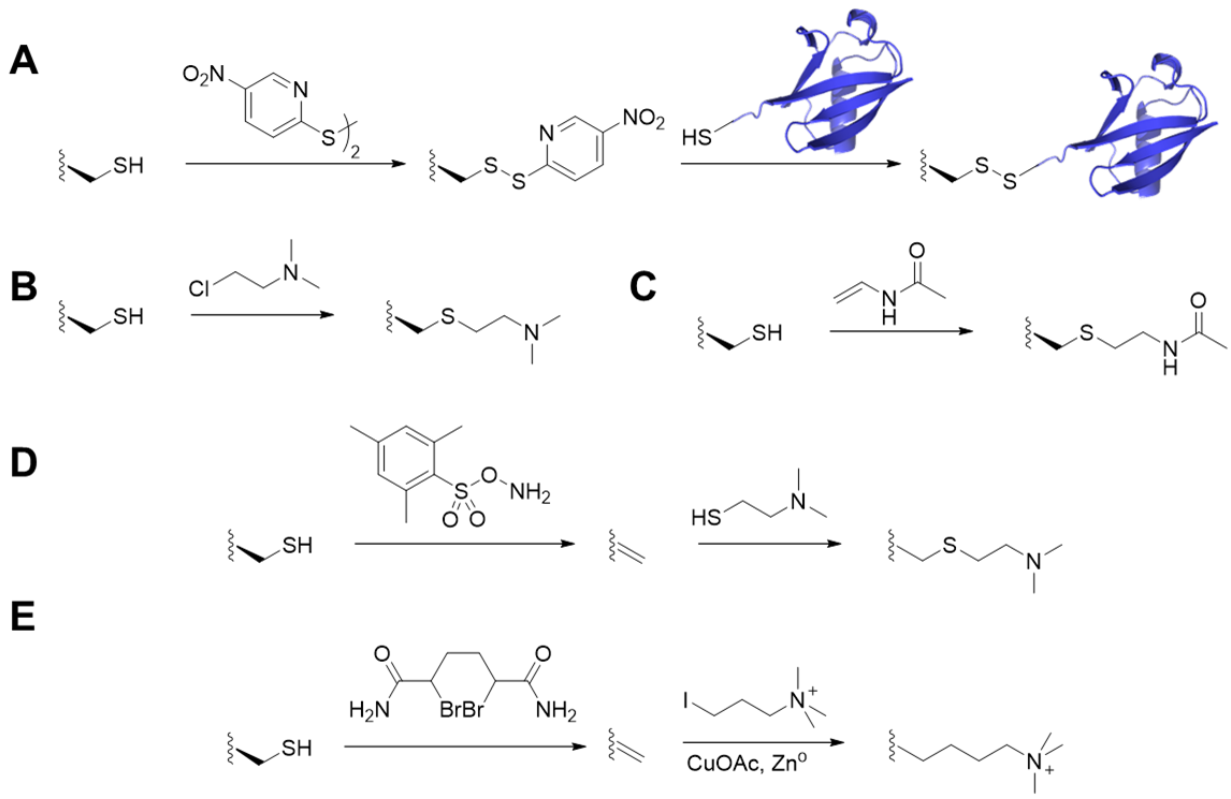


Figure 1.3. Chemical conversion of cysteine residues to modified lysine isosteres. (**A**) Cysteine activation by electrophilic DTNP disulfide, followed by disulfide exchange with a thiol functionalized protein (1UBQ). (**B**) Nucleophilic displacement of a primary alkyl chloride. (**C**) Thiol-ene radical addition to vinyl acetamide. (**D**) Elimination with O-(mesitylenesulfonyl)hydroxylamine, followed by Michael-type addition with N,N-dimethylthioethanolamine. (**E**) Elimination with α,α'-dibromoadipoyl(bis)amide, followed by cross coupling.

Amber suppression is the most attractive method for introduction and characterization of post-translational modifications, as it allows *in vivo* substitution of the desired modification at a single site. The technique utilizes a low abundance endogenous, amber, stop codon which can be matched to a non-canonical tRNA in cells transformed or transfected with orthogonal tRNA and tRNA synthetases.[36] The tRNA/synthetase pair is the product of directed evolution, which employs alternating positive selection for synthetase permissivity with negative selection for selectivity. Highly specific tRNA/synthetase pairs are essential for synthetic biology applications, however, recombinant protein expression may utilize promiscuous synthetases that incorporate multiple ncAAs so long as native amino acid incorporation remains low.[37–39] Amber suppression now enables incorporation of the native modifications phosphoserine, and phosphotyrosine, ε-

acetyl lysine, and a photoprotected monomethyl lysine precursor.[40–42] Installation of Ubls by this method is untenable, however lysine derivatives have been incorporated to enable site selective ubiquitylation by native chemical ligation (NCL).[43]

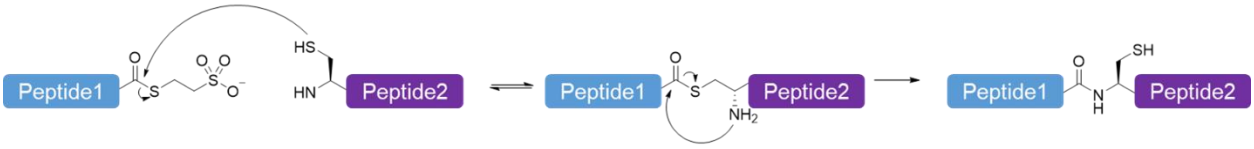


Figure 1.4. Native chemical ligation. Mechanism of a protein-peptide NCL.

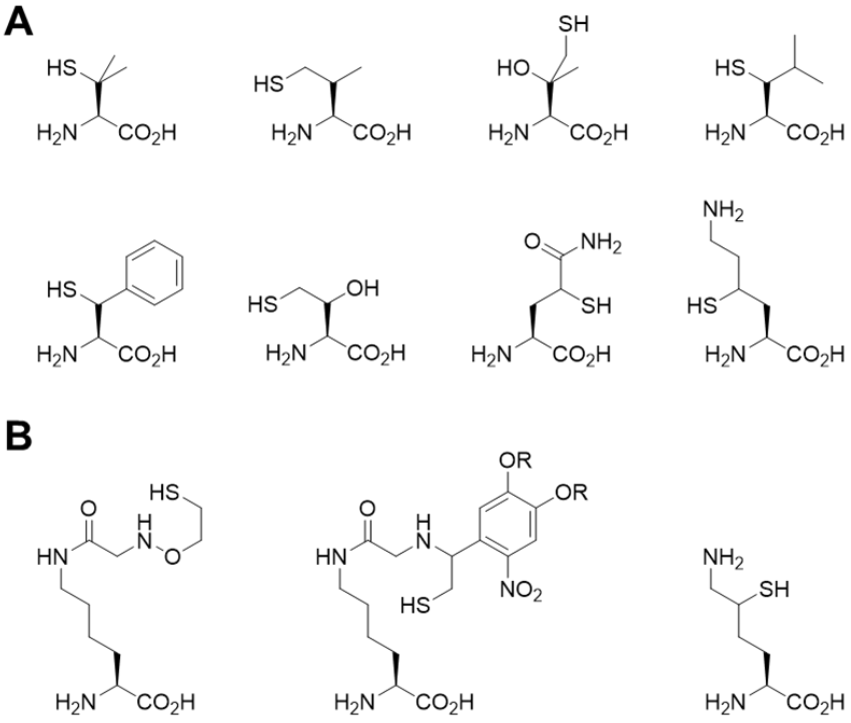


Figure 1.5. Non-canonical amino acids that enable NCL. (**A**) Synthetic thiol functionalized amino acids that can be desulfurized to afford native amino acids.[93] (**B**) Cleavable NCL handles for Ubl isopeptide formation.

First reported by Muir et. al., NCL links a C-terminal thioester and an N-terminal cysteine by equilibrium transthioesterification; the cysteinyl thioester undergoes an S-to-N acyl shift through a favorable 5-membered ring intermediate, and the stability of the resultant amide renders the reverse reaction energetically unfavorable (**Figure 4**).[44] The initial report utilized peptides

generated by solid phase peptide synthesis (SPPS), though recombinant proteins are used extensively in the study of PTMs.[45] The efficacy and generalizability of this reaction has driven research on thiol functionalization of native amino acids—reversible by radical desulfurization—cleavable thiol ligation auxiliaries, and thioesterification methods (**Figure 5a**).

Formation of the isopeptide linkage between Ub and a substrate lysine by NCL is readily achieved with a Ub α thioester, and a thiol functionalized lysine derivative (**Figure 5b**). Full length ubiquitylated proteins are thus accessible by either amber suppression of a lysine derivative, or a sequential ligation strategy employing orthogonally protected ligation handles incorporated on the solid phase. Both expressed protein ligation (EPL) strategies requires a protein thioester, which can be generated recombinantly by intein technology. Inteins are self-splicing protein elements that undergo an internal Cys N-to-S acyl shift, followed by transthioesterification onto a second Cys side chain, and excision of the intein (**Figure 6a**). The N- and C-terminal sequences, exteins, joined by a thioester then undergo an S-to-N acyl shift resulting in a native peptide linkage. Isolation of protein thioesters from intein fusions is achieved by removal of the C-extein sequence prior to expression, and addition of excess small molecule thiol to the purified protein-intein fusion (**Figure 6b**). Characterization and genomic analysis of inteins has led to discovery of naturally split inteins, which can be expressed separately and re-combined to splice their flanking extein sequences. Multiple C-terminal protein modifications have been incorporated by solid phase synthesis of a split intein fragment joined to the target C-terminal sequence, while application of this strategy to the N-terminus has been fewer.[46–48]

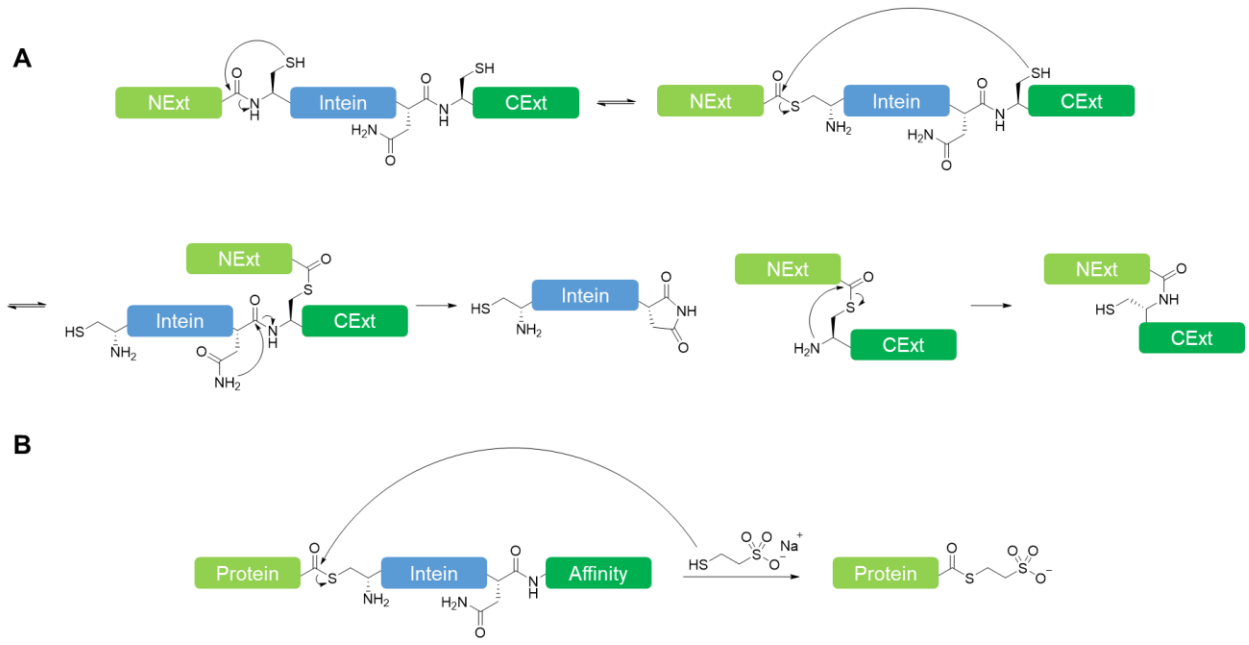


Figure 1.6. Intein splicing and protein thioester formation. (**A**) Mechanism of intein splicing. (**B**) Generation of an isolable protein thioester by transthioesterification of a protein-intein fusion in the presence of mercaptoethane sulfonate.

1.5 Chemical preparation of ubiquitin derivatives

Semi-synthetic ubiquitin-derived electrophiles generated by SPPS or direct amidation of Ub α thioesters have been instrumental in characterization of the deubiquitinases, though determinants of their substrate specificity remain vague. The earliest electrophiles produced were C-terminal thioester adducts of DTT isolated during attempted purification of an E1-Ub adduct, the hydrolysis of which was the first evidence of a Ub carboxy-terminal hydrolase (UCH).[49] The known substrate scope of UCH was expanded to amides and esters, which were produced by either incubation of the activated E1-Ub complex with amines or limited trypsination respectively.[50,51] These experiments further afforded the first insight that deubiquitinases could be substrate specific, when non-specifically ubiquitylated cytochrome C proved a poor substrate.

The first covalent capture Ub-derived electrophile to see extensive use was ubiquitin aldehyde (Ubal), the production of which by borohydride reduction confirmed the existence of an

acyl enzyme intermediate in UCH hydrolysis (**Figure 7a**).[52] The thiohemiacetal intermediate of Ubal in complex with yeast YUH1 afforded the first crystal structure of a DUB in action, lending credence to the postulated Cys-His-Asp/Asn catalytic triad and Gln oxyanion hole.[26,53] Confirmation of these catalytic features informed homology based annotation of DUB genes, which the next generation of reactive electrophiles served to validate.[54]

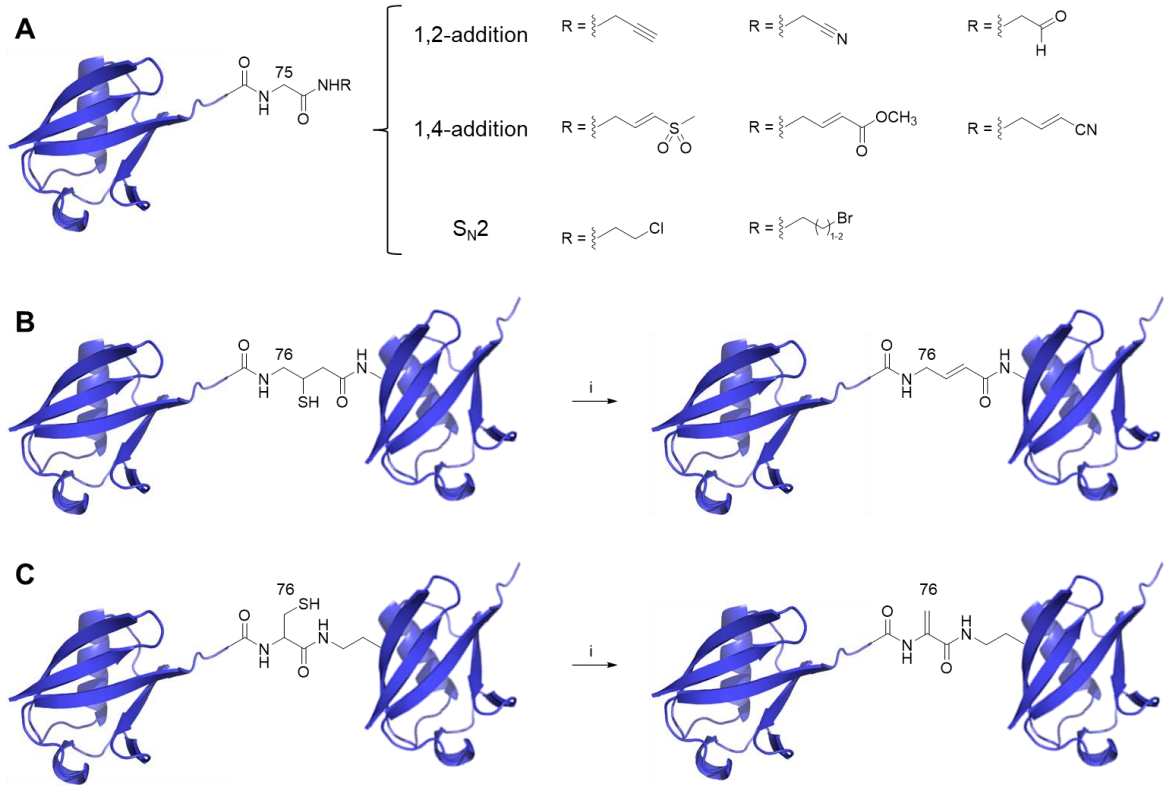


Figure 1.7. Ub- and diUb-derived electrophiles. (**A**) MonoUb electrophiles categorized by the type of covalent capture reaction with which they engage the catalytic cysteine; Gly75 annotated. (**B**) i) α,α'-dibromoadipoyl(bis)amide; elimination of ligation handle thiol to afford in-line DiUb electrophile; Gly76 isosteric position annotated. (**C**) i) α,α'-dibromoadipoyl(bis)amide; elimination of cysteine thiol to afford dehydroalanine diUb electrophile; Gly76 isosteric position annotated.

Amidation of a UbΔG α thioester with electrophilic allylamines and haloalkyl amines furnished a Ub bromopropylamine derivative that enabled analysis of the Crimean Congo Hemorrhagic Fever Virus DUB's activity toward ISG15.[55–57] The vinyl methyl ester and vinyl sulfone derivatives enabled capture and proteomic identification of three UCH and 20 USP DUBS

in EL4 cells, and have been extended to the identification of viral DUBs devoid of homology to others yet characterized.[58–60] Total chemical synthesis has recently become the strategy of choice for preparation of tagged and fluorescent ubiquitin derived electrophiles, and coupled with the discovery of the broadly reactive Ub propargylamine (UbPA) electrophile to enable characterization of numerous OTU family DUBs. Typically considered to be bioorthogonal, alkynes conjugated to the termini of Ubls exhibit strikingly selective labeling by active site thiolates, and forming the quaternary vinyl thioether exclusively.[61] Utilizing UbPA to confirm construct activity Tycho et al were able to evaluate linkage specificity of recombinant OTU catalytic domain constructs and identify which core-adjacent structural motifs impart polyUb linkage and chain length specificity, as in the case of OTUD2, which exhibits enhanced catalytic activity toward tetra- and tri-Ub and low activity toward diUb.[62]

The DUB-Ub interactions that mediate recognition of homotypic Ub chains have been further investigated with non-hydrolyzable polyUb chains and covalent capture diUb chains. The multiple distal Ub binding sites that confer selectivity to OTUs were elegantly assayed with triazole-linked non-hydrolyzable diUbs bearing C-terminal aminomethylcoumarin amides as fluorescent reporters.[63] In order to asses interactions between DUB, proximal Ub, and distal Ub Li et al synthesized a diUb electrophile by direct amidation of a Ub(1-75)-α-thioester with a ketal protected Michael acceptor, followed by nucleophilic substitution of the ketal deprotected acyl α bromide with a Ub K27C mutant.[64] Michael acceptor diUb probes have also been prepared by chemical synthesis of the Ub bearing a thiol ligation handle at the desired ubiquitylation site, which allows diUb formation by NCL, and electrophile installation by alkylative elimination of the thiol (**Figure 7b & c**).[65,66] The dehydroalanine (Dha) electrophile employed by Haj-Yahya et al in diUb probes was further demonstrated to have moderate orthogonality during activation and transthioesterification by E1, E2, E3 enzymes, which enabled its extension to *ex vivo* proteomic

characterization of ubiquitylation machinery and *in vivo* fluorescent visualization—the reactivity of related Ub dehydroalanine probes is discussed further in chapter 2.[67]

Ub-derived electrophiles have been observed to react with E1, E2 and E3 enzymes, which inspired the design of Ub-enzyme electrophilic probes. A pair of reports from Pao et al have made elegant use of a triazole-linked Ub activated vinyl sulfide, which undergoes Michael addition by E2s, but reverse Michael addition preferentially eliminates the aryl sulfinate, retaining the covalent link to the E2 and enabling a second Michael addition by cognate HECT and RING E3s.[68] This electrophile has enabled identification of an E3 ligase with Ub esterifying activity, though a bona fide substrate has yet to be identified.[69]

Given the development of a cascading electrophile, diubiquitin electrophiles, electrophilic Ub-enzyme complexes, and semi-synthetic ubiquitylated substrates, it comes as some surprise that a strategy for complexing DUB and ubiquitylated substrates has not been forthcoming. The most significant contribution to this question has been crystallization of the DUB module of the SAGA complex with a mononucleosome bearing non-hydrolyzably linked Ub-H2B.[70] In order to obtain a covalently linked DUB-substrate co-crystal structure a generalizable strategy for probe synthesis will require a method of site-specific conjugation through NCL, and orthogonal chemistry for electrophile incorporation, both of which are addressed in Chapter 2.

1.6 Purification of semi-synthetic proteins

Chemically ubiquitylated proteins like H2B and Ub are uniquely amenable to refolding. Semi-synthesis by NCL and EPL is generally more efficacious in the presence of denaturants, due to the reduction in steric hindrance at the site of transthioesterification. As split intein ligations become more widely employed it may be possible to express structured portions of a protein of interest and introduce modifications in refolded or unstructured domains, then stitch together the full length through strategic splits in unstructured regions.[46] Currently, however, chromatography

remains an essential step in the purification of semi-synthetic proteins and is often the limiting factor in overall yield.

The physical differences between ligation partners and ligation product may be sufficient to enable separation by conventional chromatographies such as size exclusion, ion exchange or ammonium sulfate fractionation. Where physical properties are insufficiently different protein purifications may employ a variety of affinity chromatographies that capitalize on highly specific, low K_D interactions from biological systems. Immunohistochemical purifications utilize antibody-antigen interactions with nanomolar K_D values, and are well suited to protein purification given the abundance of peptidic antigens (HA, FLAG, Myc, etc.) that can be genetically encoded with the protein to be expressed. Metal binding sequences (His_6, MAT) for immobilized metal affinity chromatography (IMAC) enable protein enrichment on agarose resin functionalized with nitroloacetic acid complexed to Ni^{2+}, or Co^{2+}.[71] The robustness of IMAC led Bang & Kent to incorporate a C-terminal His_6 tag during SPPS, enabling isolation of their desired ligation product twice in the course of the three segment synthesis of crambin by NCL.[72] The strong non-covalent interaction between biotin and streptavidin has driven research into endogenous biotinylation strategies including tagging with a sequence recognized by the *E. coli* biotin ligase BirA. Mutagenesis of the streptavidin pocket resulted in the peptidic Strep-tag/Streptactin purification system, which allows for enrichment and mild elution with desthiobiotin—whereas streptavidin requires detergent and heat for quantitative elution.[73] The aforementioned methods are of particular value because the peptide sequences required are small, and therefore less likely to affect protein fold or function; however, some recombinant proteins are poorly expressed ectopically, in which case whole protein domains (MBD, CBD, SUMO, GST, etc.) can enhance protein solubility in addition to affording an affinity handle for subsequent purification.

Alteration of protein structure and function by affinity tags must be assayed for on a protein by protein basis. Recapitulation of biologically relevant protein-protein interactions *in vitro*,

confirmation of enzyme activity, and/or spectroscopic evaluations like NMR or CD may be used as validation. Structural issues that arise may be addressed by shifting a tag from one terminus to the other, or introduction of a protease cleavage site within a flexible Ser/Gly linker between the tag and protein of interest. Tag removal by protease generally leaves some fragment of the recognition sequence at the modified terminus, and may necessitate a second purification step to remove both protease and uncleaved material. SUMO is particularly well addressed to this limitation as its Gly-Gly terminus acts as a flexible linker, and an additional affinity tag can be conjugate to the N-terminus of SUMO without impairing proteolysis by SUMO proteases.

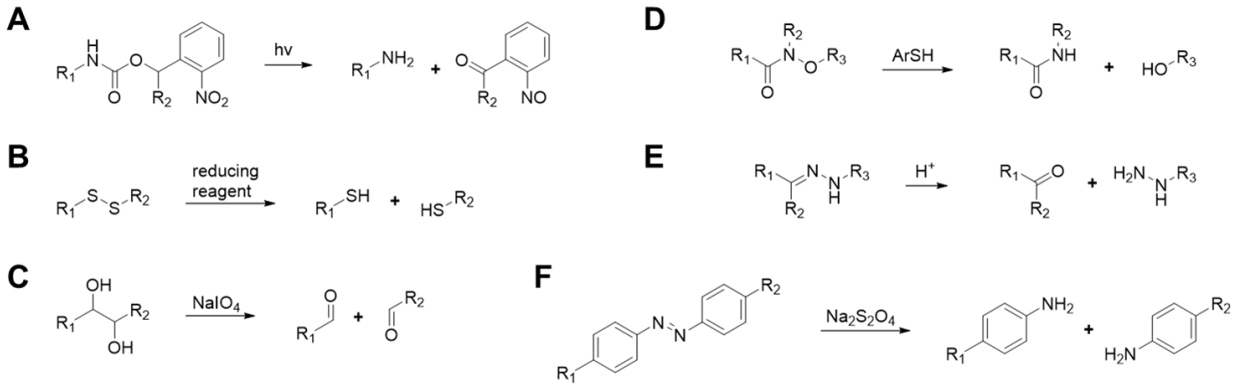


Figure 1.8. Selectively cleavable linkers employed in protein chemistry. (**A**) Photolysis of o-nitrobenzyl carbamate linker. (**B**) Reduction of disulfide linker. (**C**) Oxidative cleavage of vicinal diol. (**D**) Disulfide radical anion reduction of aminooxy amide. (**E**) Acid hydrolysis of hydrazone. (**F**) Dithionite reduction of aryl hydrazide.

Chemically cleavable linkers offer the advantage of leaving few or no atoms behind and may be cleaved under a variety of conditions (**Figure 8**). Arguably the mildest cleavage strategy, photocleavable *o*-nitrobenzyl carbamate derivatives of biotin have been introduced during SPPS and non-specifically as lysine reactive NHS-esters, of which the former afforded milligram quantities of full length *O*-GlcNAcylated Tau following cleavage from streptavidin resin.[74,75] Acid labile hydrazone linkers, and periodate labile vicinal diol linkers are both amenable to chemical cleavage, but yield non-native ketone, aldehyde or hydrazide functional groups on cleavage.[76] Reductive cleavage of disulfides enables traceless cysteine tagging, but is incompatible with the

reducing conditions employed in NCL; similarly, diazobenzenes undergo reductive cleavage with sodium dithionite, and have seen applications in both protein and polymer chemistry.[77,78] The diazobenzene linker is principally applicable to development of a traceless linker, undergoing self-immolation upon reduction.[79] Recent work by Weller et. al. divulged a non-denaturing cleavage of tertiary aminooxy amide N-O bonds by aromatic thiols, which motivates the work discussed in Chapter 3.[80]

1.7 Aim of this work

In this work I have developed novel chemical approaches to site specifically ubiquitylate proteins and incorporate electrophiles at the site of ubiquitylation in order to capture relevant DUBs. The synthesis of Ub-derived electrophiles and ncAAs for amber suppression are the subject of Chapter 2. In order to synthesize full length ubiquitylated targets I developed multiple strategies for reversibly appending affinity handles to peptides in order to facilitate isolation of products from EPL. The synthesis of aminooxy amide derivatives, their incorporation in peptides, and assessment of cleavage conditions are the subject of Chapter 3.

1.8 References

1. Walsh, C. T., Garneau-tsodikova, S. & Gatto, G. J. Protein Chemistry Protein Posttranslational Modifications : The Chemistry of Proteome Diversifications. *Angew. Chemie Int. Ed.* **44,** 7342–7372 (2005).

2. Sabari, B. R. *et al.* Intracellular Crotonyl-CoA Stimulates Transcription Article Intracellular Crotonyl-CoA Stimulates Transcription through p300-Catalyzed Histone Crotonylation. *Mol. Cell* **58,** 203–215 (2015).

3. Huang, J. *et al.* 2-Hydroxyisobutyrylation on histone H4K8 is regulated by glucose homeostasis in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.* **114,** (2017).

4. Li, M. *et al.* Mono- versus polyubiquitination: differential control of p53 fate by Mdm2. *Science (80-.).* **302,** 1972–5 (2003).

5. Yuan, J., Luo, K., Zhang, L., Cheville, J. C. & Lou, Z. USP10 regulates p53 localization and stability by deubiquitinating p53. *Cell* **140,** 384–96 (2010).

6. van der Veen, A. G. & Ploegh, H. L. Ubiquitin-like proteins. *Annu. Rev. Biochem.* **81,** 323–57 (2012).

7. Carter, S. & Vousden, K. H. p53-Ubl fusions as models of ubiquitination, sumoylation and neddylation of p53. *Cell Cycle* **7,** 2519–2528 (2008).

8. Liu, G. & Xirodimas, D. P. NUB1 promotes cytoplasmic localization of p53 through cooperation of the NEDD8 and ubiquitin pathways. *Oncogene* **29,** 2252–2261 (2010).

9. Mahajan, R., Delphin, C., Guan, T., Gerace, L. & Melchior, F. A Small Ubiquitin-Related Polypeptide Involved in Targeting RanGAP1 to Nuclear Pore Complex Protein RanBP2. *Cell* **88,** 97–107 (1997).

10. Stindt, M. H., Carter, S. a., Vigneron, A. M., Ryan, K. M. & Vousden, K. H. MDM2 promotes SUMO-2/3 modification of p53 to modulate transcriptional activity. *Cell Cycle* **10,** 3176–3188 (2011).

11. Shiio, Y. & Eisenman, R. N. Histone sumoylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 13225–30 (2003).

12. Flotho, A. & Melchior, F. Sumoylation: a regulatory protein modification in health and disease. *Annu. Rev. Biochem.* **82,** 357–85 (2013).

13. Desterro, J. M. ., Rodriguez, M. S. & Hay, R. T. SUMO-1 Modification of IκBα Inhibits NF-κB Activation. *Mol. Cell* **2,** 233–239 (1998).

14. Kumar, R., González-Prieto, R., Xiao, Z., Verlaan-de Vries, M. & Vertegaal, A. C. O. The STUbL RNF4 regulates protein group SUMOylation by targeting the SUMO conjugation machinery. *Nat. Commun.* **8,** 2–16 (2017).

15. Nie, M., Moser, B. A., Nakamura, T. M. & Boddy, M. N. SUMO-targeted ubiquitin ligase activity can either suppress or promote genome instability , depending on the nature of the DNA lesion. *PLoS Genet.* 1–25 (2017).

16. Matic, I., Macek, B., Hilger, M., Walther, T. C. & Mann, M. Phosphorylation of SUMO-1 Occurs in Vivo and Is Conserved through Evolution research articles. *J. Proteome Res.* **7,** 4050–4057 (2008).

17. Wauer, T., Simicek, M., Schubert, A. & Komander, D. Mechanism of phospho-ubiquitin-induced PARKIN activation. *Nature* **524,** 370–4 (2015).

18. Gersch, M. *et al.* Mechanism and regulation of the Lys6-selective deubiquitinase USP30. *Nat. Struct. Mol. Biol.* **24,** 920–930 (2017).

19. Spasser, L. & Brik, A. Chemistry and biology of the ubiquitin signal. *Angew. Chem. Int. Ed. Engl.* **51,** 6840–62 (2012).

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

21. Vittal, V., Stewart, M. D., Brzovic, P. S. & Klevit, X. R. E. Regulating the Regulators : Recent Revelations in the Control of E3 Ubiquitin Ligases. *J. Biol. Chem.* **290,** 21244–21251 (2015).

22. Iwai, K., Fujita, H. & Sasaki, Y. Linear ubiquitin chains: NF-κB signalling, cell death and beyond. *Nat. Rev. Mol. Cell Biol.* **15,** 503–508 (2014).

23. Komander, D., Clague, M. J. & Urbé, S. Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **10,** 550–63 (2009).

24. Arif, S. *et al.* MINDY-1 Is a Member of an Evolutionarily Conserved and Structurally Distinct New Family of Deubiquitinating Enzymes Short Article MINDY-1 Is a Member of an Evolutionarily Conserved and Structurally Distinct New Family of Deubiquitinating Enzymes. *Mol. Cell* **63,** 146–155 (2016).

25. Kwasna, D. *et al.* Distinct Deubiquitinase Class Important for Genome Article Discovery and Characterization of ZUFSP / ZUP1 , a Distinct Deubiquitinase Class Important for Genome Stability. *Mol. Cell* **70,** 150–164.e6 (2018).

26. Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D. & Hill, C. P. Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 A resolution. *EMBO J.* **16,** 3787–96 (1997).

27. Yao, T. *et al.* Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nat. Cell Biol.* **8,** 994–1002 (2006).

28. Ye, Y., Scheel, H., Hofmann, K. & Komander, D. Dissection of USP catalytic domains reveals five common insertion points. *Mol. Biosyst.* **5,** 1797–808 (2009).

29. Mevissen, T. E. T. & Komander, D. Mechanisms of Deubiquitinase Specificity and Regulation. *Annu. Rev. Biochem.* **86,** 159–192 (2017).

30. Meier, F. *et al.* Semisynthetic, site-specific ubiquitin modification of α-synuclein reveals differential effects on aggregation. *J. Am. Chem. Soc.* **134,** 5468–71 (2012).

31. Chatterjee, C., McGinty, R. K., Fierz, B. & Muir, T. W. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat. Chem. Biol.* **6,** 267–9 (2010).

32. Dhall, A. *et al.* Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long-range Internucleosomal Interactions. *J. Biol. Chem.* **289,** 33827–33837 (2014).

33. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2,** 1666–1676 (2011).

34. Wang, Z. A. & Liu, W. R. Proteins with Site-Specific Lysine Methylation. *Chem. A Eur. J.* **23,** 11732–11737 (2017).

35. Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science (80-.).* **354,** 1–5 (2016).

36. Wang, L., Brock, A. & Schultz, P. G. Adding L-3-(2-Naphthyl)alanine to the genetic code of E. coli. *J. Am. Chem. Soc.* **124,** 1836–7 (2002).

37. Stokes, A. L. *et al.* Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol. Biosyst.* **5,** 1032–1038 (2009).

38. Young, D. D. *et al.* An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity †,‡. *Biochemistry* **50,** 1894–1900 (2011).

39. Cooley, R. B., Karplus, P. A. & Mehl, R. A. Gleaning Unexpected Fruits from Hard-Won Synthetases : Probing Principles of Permissivity in Non-canonical Amino Acid – tRNA Synthetases. *ChemBioChem* **15,** 1810–1819 (2014).

40. Rogerson, D. T. *et al.* Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* **11,** 496–503 (2015).

41. Luo, X. *et al.* Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* **13,** 845–849 (2017).

42. Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* **6,** 50–69 (2015).

43. Virdee, S. *et al.* Traceless and Site-Specific Ubiquitination of Recombinant Proteins. *J. Am. Chem. Soc.* **133,** 10708–10711 (2011).

44. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science (80-.).* **266,** 775–778 (1994).

45. Merrifield, R. B. Solid Phase Peptide Synthesis. I . The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149–2154 (1963).

46. Stevens, A. J. *et al.* A promiscuous split intein with expanded protein engineering applications. *Proc. Natl. Acad. Sci.* **114,** 8538–8543 (2017).

47. Ludwig, C., Pfeiff, M., Linne, U. & Mootz, H. D. Ligation of a Synthetic Peptide to the N Terminus of a Recombinant Protein Using Semisynthetic Protein trans -Splicing. *Angew. Chemie Int. Ed.* **45,** 5218–5221 (2006).

48. Appleby-tagoe, J. H. *et al.* Highly Efficient and More General cis - and trans -Splicing Inteins through Sequential Directed Evolution. *J. Biol. Chem.* **286,** 34440–34447 (2011).

49. Rose, I. A. & Warms, J. V. B. An Enzyme with Ubiquitin Carboxy-Terminal Esterase Activity. *Biochemistry* **414,** 4234–4237 (1983).

50. Pickart, C. M. & Rose, I. A. Ubiquitin Carboxyl-terminal Hydrolase Acts on Ubiquitin Carboxyl-terminal Amides. *J. Biol. Chem.* **260,** 7903–7910 (1985).

51. Wilkinson, K. D., Cox, M. J., Mayer, A. N. & Frey, T. Synthesis and Characterization of Ubiquitin Ethyl Ester , a New Substrate for Ubiquitin Carboxyl-Terminal Hydrolase. *Biochemistry* **25,** 6644–6649 (1986).

52. Pickart, C. M. & Rose, A. Mechanism of Ubiquitin Carboxyl-terminal Hydrolase. *J. Biol. Chem.* **261,** 10210–10217 (1986).

53. Johnston, S. C., Riddle, S. M., Cohen, R. E. & Hill, C. P. Structural basis for the specificity of ubiquitin C-terminal hydrolases. *EMBO J.* **18,** 3877–87 (1999).

54. Nijman, S. M. B. *et al.* A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123,** 773–86 (2005).

55. Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9,** 1149–59 (2002).

56. Akutsu, M., Ye, Y., Virdee, S., Chin, J. W. & Komander, D. Molecular basis for ubiquitin and ISG15 cross-reactivity in viral ovarian tumor domains. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 2228–33 (2011).

57. James, T. W. *et al.* Structural basis for the removal of ubiquitin and interferon-stimulated gene 15 by a viral ovarian tumor domain-containing protease. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 2222–7 (2011).

58. Kattenhorn, L. M., Korbel, G. a, Kessler, B. M., Spooner, E. & Ploegh, H. L. A deubiquitinating enzyme encoded by HSV-1 belongs to a family of cysteine proteases that is conserved across the family Herpesviridae. *Mol. Cell* **19,** 547–57 (2005).

59. Misaghi, S. *et al.* Chlamydia trachomatis-derived deubiquitinating enzymes in mammalian cells during infection. *Mol. Microbiol.* **61,** 142–50 (2006).

60. Artavanis-Tsakonas, K. *et al.* Identification by functional proteomics of a deubiquitinating/deNeddylating enzyme in Plasmodium falciparum. *Mol. Microbiol.* **61,** 1187–95 (2006).

61. Ekkebus, R. *et al.* On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. *J. Am. Chem. Soc.* **135,** 2867–70 (2013).

62. Mevissen, T. E. T. *et al.* OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. *Cell* **154,** 169–84 (2013).

63. Flierman, D. *et al.* Non-hydrolyzable Diubiquitin Probes Reveal Linkage-Specific Reactivity of Deubiquitylating Enzymes Mediated by S2 Pockets Article Non-hydrolyzable Diubiquitin Probes Reveal Linkage-Specific Reactivity of Deubiquitylating Enzymes Mediated by S2 Pockets. *Cell Chem. Biol.* **23,** 472–482 (2016).

64. Li, G., Liang, Q., Gong, P., Tencer, A. H. & Zhuang, Z. Activity-based diubiquitin probes for elucidating the linkage specificity of deubiquitinating enzymes. *Chem. Commun. (Camb).* **50,** 216–8 (2014).

65. Haj-Yahya, N. *et al.* Dehydroalanine-based diubiquitin activity probes. *Org. Lett.* **16,** 540–3 (2014).

66. Mulder, M. P. C., El Oualid, F., ter Beek, J. & Ovaa, H. A native chemical ligation handle that enables the synthesis of advanced activity-based probes: Diubiquitin as a case study. *ChemBioChem* **15,** 946–949 (2014).

67. Mulder, M. P. C. *et al.* A cascading activity-based probe sequentially targets E1–E2–E3 ubiquitin enzymes. *Nat. Chem. Biol.* **12,** 523–530 (2016).

68. Pao, K. *et al.* Probes of ubiquitin E3 ligases enable systamtic dissection of parkin activation.

Nat. Chem. Biol. **12,** 324–331 (2016).

69. Pao, K. *et al.* Activity-based E3 ligase profiling uncovers and E3 ligase with esterification activity. *Nature* **556,** 381–385 (2018).

70. Morgan, M. T. *et al.* Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science (80-.).* **351,** 1–5 (2016).

71. Hochuli, E., Dobeli, H. & Schacher, A. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr.* **411,** 177–184 (1987).

72. Bang, D. & Kent, S. B. H. His 6 tag-assisted chemical protein synthesis. *Proc. Natl. Acad. Sci.* **102,** 5014–5019 (2005).

73. Skerra, A. & Schmidt, T. O. M. Use of the Strep-Tag and Streptavidin for Detection and Purification of Recombinant Proteins. *Methods Enzymol.* **326,** 2000 (2000).

74. Olejnik, J., Sonar, S., Krzymanska-olejnik, E. & Rothschild, K. J. Photocleavable biotin derivatives : A versatile approach for the isolation of biomolecules. *Proc. Natl. Acad. Sci.* **92,** 7590–7594 (1995).

75. Schwagerus, S., Reimann, O., Despres, C., Smet-nocca, C. & Hackenberger, C. P. R. Semi-synthesis of a tag-free O -GlcNAcylated tau protein by sequential chemoselective ligation §. *J. Pept. Sci.* **22,** 327–333 (2016).

76. Böhme, D. & Beck-sickinger, A. G. Drug delivery and release systems for targeted tumor therapy. *J. Pept. Sci.* **21,** 186–200 (2015).

77. Yang, Y. & Verhelst, S. H. L. Cleavable trifunctional biotin reagents for protein labelling, capture and release. *Chem. Commun.* **49,** 5366–5368 (2013).

78. Wong, A. D. & Gillies, E. R. Multiresponsive Azobenzene End-Cap for Self-Immolative Polymers. *ACS Macro Lett.* **3,** 1191–1195 (2014).

79. Bogyo, M., Verhelst, S. H. L. & Fonovic, M. A Mild Chemically Cleavable Linker System for Functional Proteomic Applications. *Angew. Chemie Int. Ed.* **46,** 1284–1286 (2007).

80. Weller, C. E. *et al.* Aromatic thiol-mediated cleavage of N–O bonds enables chemical ubiquitylation of folded proteins. *Nat. Commun.* **7,** 12979 (2016).

81. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res.* **21,** 564–578 (2011).

82. Blanc, R. S. & Richard, S. Arginine Methylation : The Coming of Age. *Mol. Cell* **65,** (2016).

83. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **194,** 531–44 (1987).

84. Ding, H. *et al.* Solution structure of human SUMO-3 C47S and its binding surface for Ubc9. *Biochemistry* **44,** 2790–9 (2005).

85. Akutsu, M., Dikic, I. & Bremm, A. Ubiquitin chain diversity at a glance. *J. Cell Sci.* **129,** 875–880 (2016).

86. Fierz, B. *et al.* Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat. Chem. Biol.* **7,** 113–9 (2011).

87. Matsumoto, M. L. *et al.* K11-linked polyubiquitination in cell cycle control revealed by a K11 linkage-specific antibody. *Mol. Cell* **39,** 477–484 (2010).

88. Gatti, M. *et al.* to Signal DNA Damage Article RNF168 Promotes Noncanonical K27 Ubiquitination to Signal DNA Damage. *CellReports* **10,** 226–238 (2015).

89. Nucifora Jr, F. C. *et al.* Ubiqutination via K27 and K29 chains signals aggregation and neuronal protection of LRRK2 by WSB1. *Nat. Commun.* **7,** 1–11 (2016).

90. Yuan, W. *et al.* K33-Linked Polyubiquitination of Coronin 7 by Cul3-KLHL20 Ubiquitin E3 Ligase Regulates Protein Trafficking. *Mol. Cell* **54,** 586–600 (2014).

91. Priego, S. *et al.* Polyubiquitylation drives replisome disassembly at the termination of DNA replication. *Science (80-.).* **326,** 477–481 (2014).

92. Meyer, H.-J. & Rape, M. Enhanced Protein Degradation by Branched Ubiquitin Chains. *Cell* **157,** 910–921 (2014).

93. Wong, C. T. T., Tung, C. L. & Li, X. Synthetic cysteine surrogates used in native chemical ligation. *Mol. Biosyst.* **9,** 826–833 (2013).

Chapter 2: Development of bioorthogonal chemistry for ubiquitin ligation and electrophile incorporation

2.1 Introduction

Ubiquitin-derived electrophilic probes have been instrumental in characterizing both conjugating and deconjugating enzymes (DUBs) in the ubiquitin system. Existing strategies for electrophile incorporation are, however, incompatible with application to most known ubiquitylated proteins. The C-terminus of ubiquitin (Ub) can be activated as a thioester for direct amidation with primary amines bearing β or δ alkyl halides, Michael acceptors, and alkynes;[1–3] Alternatively, C-terminal thioesters can be more selectively functionalized through native chemical ligation (NCL) with an electrophile precursor. Historically, limited trypsination in the presence of excess diglycine analogue can drive exchange of the analogue for the two C-terminal glycine residues of Ub.[4] Finally, solid phase synthesis of Ub by the Ovaa and Brik groups has enabled incorporation of terminal covalent capture and fluorogenic features, as well as internal modifications for assembling diUbs. Each of these strategies yields an electrophile either isosteric to, or one carbon removed from the isopeptide carbonyl carbon with which the active site cysteine reacts.

Each of the described electrophilic probes removes or displaces the isopeptide amide, which participates in hydrogen bonding interactions with some DUBs.[5] The Michael acceptor probes stand out as the most similar to the native Ub-substrate linkage as the enone preserves the sp2 hybridization of the electrophile and a proximal sp2 hybrid oxygen for hydrogen bonding. This class of probes, until recently, comprised exclusively in-line Michael acceptors of the vinyl methyl ester, vinyl sulfone, and acrylonitrile types, leaving unanswered the question of whether an electrophile preserving the native isopeptide would alter reactivity toward DUBs.

Dehydroalanine (Dha) and dehydrobutyrene are naturally occurring post-translational modifications that arise from β elimination of phosphoserine and phosphothreonine respectively.

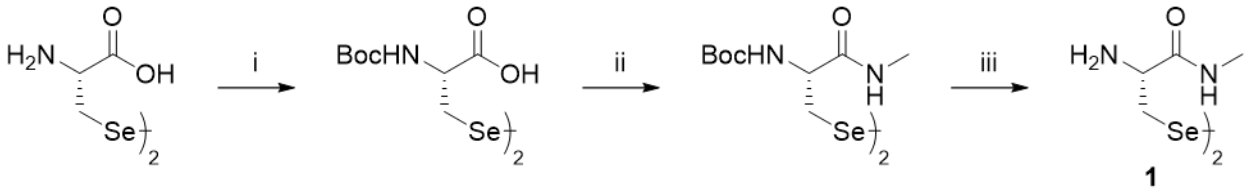
The α,β unsaturated enones that result are key intermediates in the biosynthesis of cyclic lanthipeptides, so named for the characteristic lanthionine linkage that results from 1,4 addition of a cysteine thiol into the enone.[6] The evidenced reactivity of Dha toward thiol nucleophiles suggests potential application as an irreversible covalent trap for cysteine protease DUBs. The biosynthesis of Dha *in situ* suggests that mild, selective chemistries can be employed in generation of the electrophile, and that the selection of an appropriate electrophile precursor can afford multiple modes of reactivity.

Chemical synthesis of Dha has been achieved by β elimination of aminated and alkylated cysteine residues, as well as oxidation of phenylselenocysteine.[7,8] Cysteine as a Dha precursor facilitates installation at the G76 position of Ub by NCL chemistry similar to the direct amidation employed in preparation of the monoubiquitin derived electrophiles. Furthermore, the application of sequential NCLs to semi-synthesis of post-translationally modified proteins suggests that condensation of Cys as onto Lys as an ε isopeptide could be used to semi-synthesize ubiquitylated proteins, and then apply the Dha derivatized proteins as probes for deubiquitinases. The generalizability of this approach is limited by the natural occurrence of cysteine residues, many of which play structurally significant roles in disulfide bridges and zinc fingers. Selenocysteine, by contrast, naturally occurs in far fewer proteins and exhibits reactivity distinct from that of cysteine; thus, application of Sec as both ligation handle and Dha precursor renders a large portion of the proteome accessible to semi-synthetic preparation of covalent capture probes for deubiquitinases. The extension of the covalent capture probes to full length proteins may enable delineation of substrate specificity among the deubiquitinases, characterization of which is largely incomplete.

2.2 Results and Discussion

2.2.1 Synthesis and reactivity of Ub Gly76Dha N-methylamide

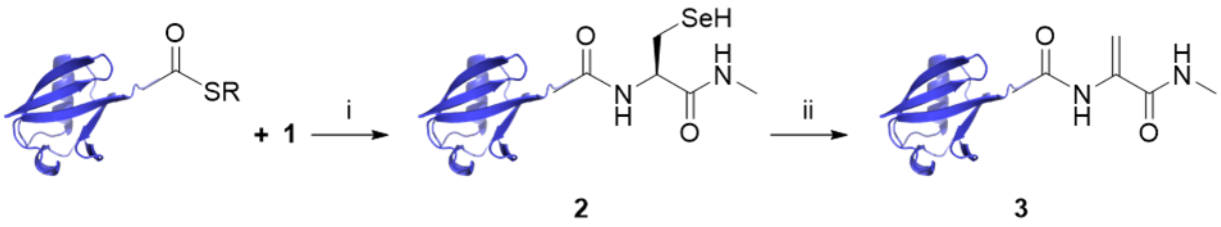
The viability of this chemical strategy depends on the criteria of (1) Dha reactivity toward active site thiolates, (2) Sec ligation efficacy, and (3) selective the conversion of Sec to Dha. This first criteria might be directly addressed through mutation of Ub to G76C, and alkylative elimination to Dha; however, the predominance at the C-terminus of the deprotonated carboxylate under physiological conditions employed in deubiquitinase assays can be reasonably assumed to inhibit thiolate 1,4 addition by decreasing stability of the enolate intermediate. This property has been neatly evidenced in the case of fumarate ester electrophilic probes that react with cysteine proteases through either ester hydrolysis or 1,4 addition, and have been observed cease 1,4 addition following hydrolysis.[9] In order to address experiments to all three criteria the terminal glycine of Ub was to be replaced with selenocysteine N-methylamide (**1**), enabling optimization of selenol selective chemistry, and subsequent characterization of enamide reactivity (**Scheme 2.1**).



Scheme 2.1. Synthesis of Selenocystine N-methyl amide. (i) di-tert-butyl dicarbonate, TEA, water, 90%; (ii) methylamine, EDCI, HOBt, TEA, DMF, 68%; (iii) 1:1 TFA:DCM, quant.

Synthesis of Ub Gly76Sec N-methylamide (**2**) was achieved by NCL of Ub(1-75) α-thioester and **1** (**Scheme 2.2**). **2** was isolated as both symmetric diselenide and reduced selenol, but *in situ* reduction of the ligation reaction with thiol reducing reagents was insufficient for complete reduction. Reduction with water soluble phosphine tris-carboxylethylphosphine (TCEP) either *in situ* or following reverse phase high performance liquid chromatography (RP-HPLC)

purification was rapid, but a competing deselenization of Sec to Ala was unavoidable. While the free amino acid amide is the least sterically hindered substrate profiled herein, the ligation was, consistent with prior reports of selenol NCL, rapid and efficient.



Scheme 2.2. Synthesis of Ub Gly76Sec N-methylamide and Ub Gly76Dha N-methylamide. (i) (1) **1**, mercaptophenylacetic acid, Na_2HPO_4, pH 7.5; (2) dithiothreitol, pH 8. (ii) $NaIO_4$, sodium acetate, pH 5.5; (PDB 1UBQ).[10]

Orthogonal conversion of Sec to Dha capitalizes on marked differences in the observed pKa, nucleophilicity, and redox potential of Sec and Cys.[11–14] Oxidative elimination of selenocysteine had been observed during preparation of selenoprotein samples for proteomic characterization by LC-MSMS.[15] Oxidation of diselenides and selenylsulfides is reported to precede cleavage of the bond between the heavier chalcogens, which facilitates manipulation of the protein diselenide that forms rapidly under oxidative conditions.[16,17] The nucleophilicity of Sec has enabled its manipulation in the presence of Cys by pH regulated single alkylation and subsequent oxidative elimination, wherein sulfinic and sulfenic acid formation was skirted by disulfide protection of thiol side chains in the polypeptide. Direct oxidation of the **2** diselenide is parsimonious in both steps and reagent, and furnished the desired product (**3**) in reasonable yield (**Scheme 2.2**).

The electrophilicity of **3** was established first by *in vitro* assay with purified DUBs USP2 and UCH-L3, and the reaction with UCH-L3 confirmed by LC-MS (**Figure 2.1**). A FLAG tagged Ub Gly76Dha (**4**) was employed for visualization of reactions in *E. coli* lysate, where specificity

for the catalytic cysteine of UCH-L3 was affirmed by comparison to the catalytically inactive C95A mutant (**Figure 2.2a**). Activity assays against a broader panel of deubiquitinases (conducted by Dr. Ambar Rana, University of Madison Wisconsin) confirmed reactivity toward UCH family DUBs and the OTU Cezanne (**Figure 2.2b**). Labeling intensity varied across DUBs surveyed, which is consistent with all previously reported Ub-derived electrophiles.[18]

Human HeLa cell lysates were probed (conducted by Dr. Caroline Weller, Chatterjee Lab), and higher molecular weight adducts were observed. Inactivation of lysate DUBs with iodoacetamide and N-ethyl maleimide dramatically decreased high molecular weight bands (**Figure 2.3a**). A competition between **4** and commercial Ub vinyl methyl ester (UbVME) reduced the α-FLAG signal in a ratio-dependent fashion (**Figure 2.3b**). Concurrent work by the Mulder et al indicated that UbDha carboxylate can cycle through the E1-E2-E3 cascade in HeLa cells and react with deubiquitinases, though E1s and E2s were the most robustly labelled in this proteomic analysis.[19]

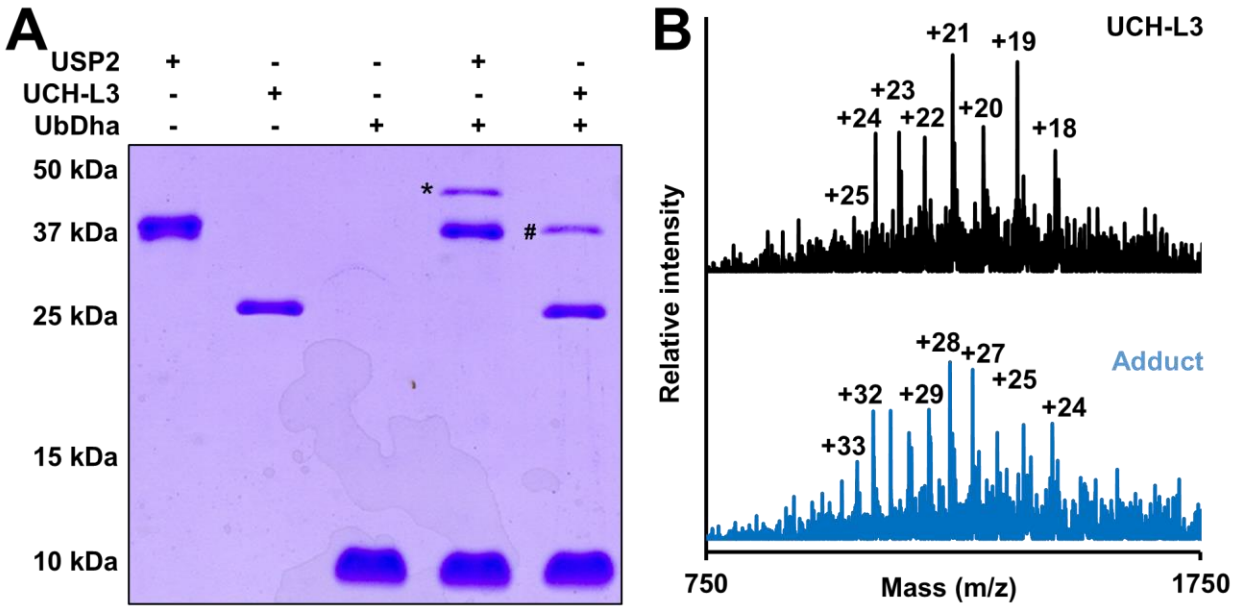


Figure 2.1. Covalent capture of UCH-L3 and USP2 with 3. (**A**) 15% SDS-PAGE resolved covalent adducts of **3** and USP2 (*), and **3** and UCH-L3 (#) visualized with coomassie brilliant blue. (**B**) ESI-MS of UCH-L3 (top) and UCH-L3-**3** adduct (bottom). Calculated for UCH-L3 $[M+H]^+$ 26183 Da, observed 26194.1 ± 4.4 Da. Calculated for UCH-L3-**3** adduct $[M+H]^+$ 34773 Da, observed 34774.0 ± 4.9 Da.

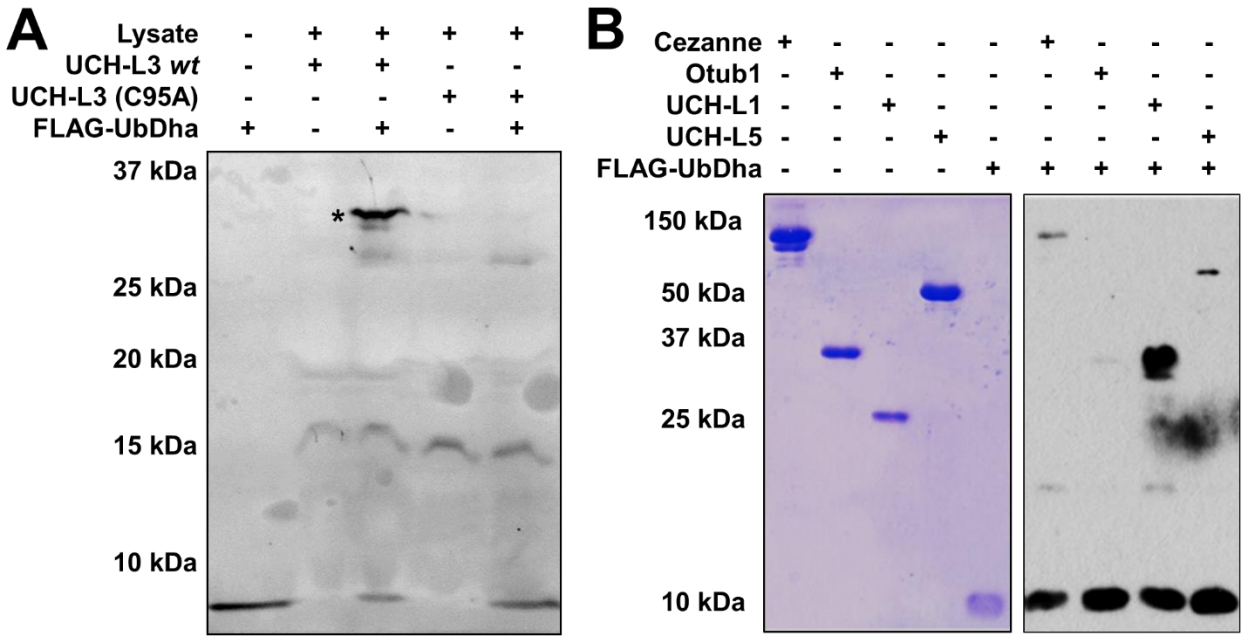


Figure 2.2. In vitro and ex vivo covalent capture with 4. (**A**) Western Blot of E. coli lysates containing overexpressed wild-type UCH-L3 or UCH-L3(C95A) incubated with **4**. The covalent UCH-L3-**4** adduct (*) was detected with an anti-FLAG antibody (F3165, Sigma-Aldrich). (**B**) Western blot of Cezanne (left), Otub1, UCH-L1, and UCH-L5 (right) incubated with **4** for 12 hours at 37 ˚C. The DUB−5 adducts were detected with an anti-Ub antibody (P4D1, Santa Cruz Biotech).

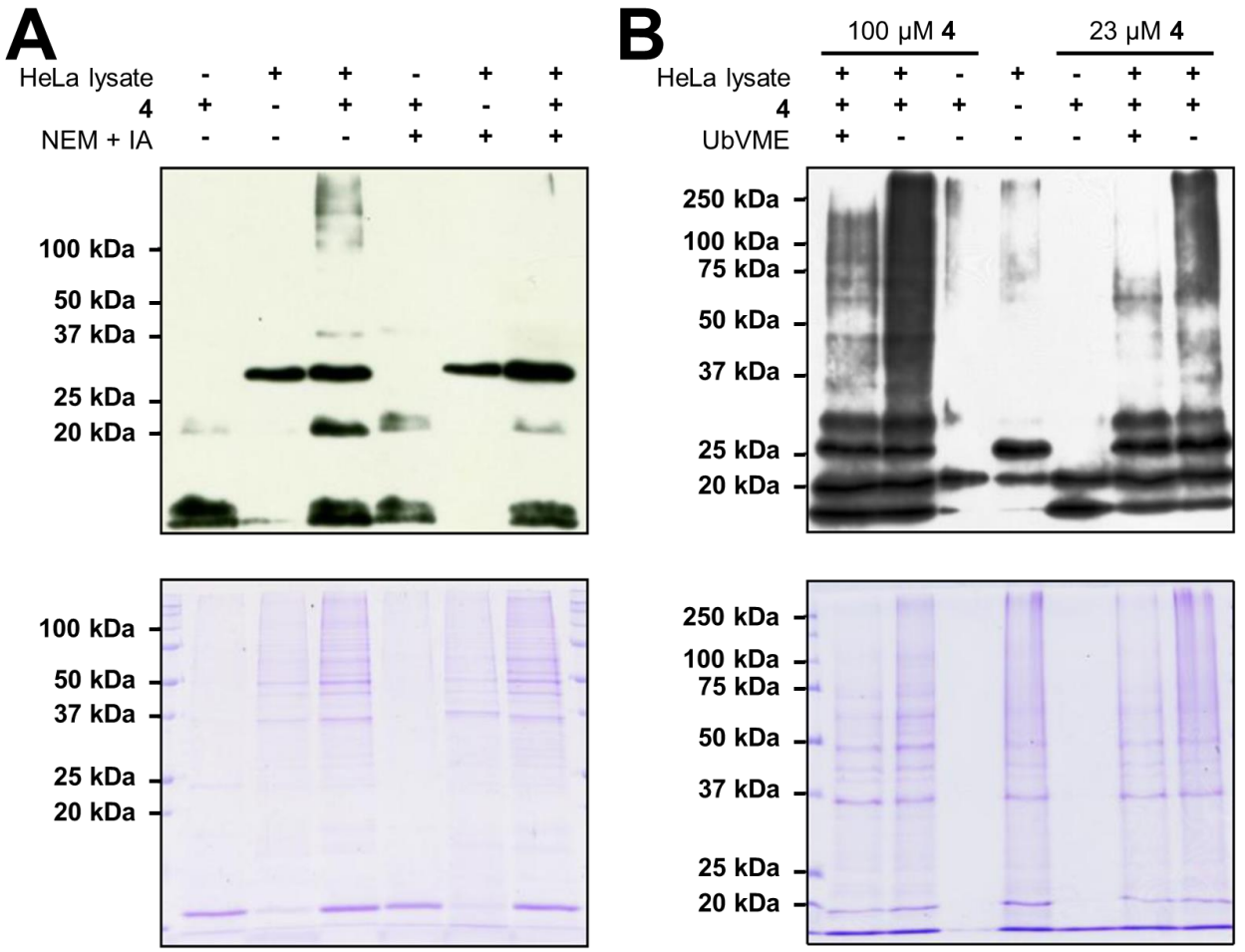


Figure 2.3. HeLa proteome labeling. (**A**) Western blot of HeLa proteome labeling by **4** (100 μM, 1.5 h, 37 ˚C) in the presence or absence of the DUB inhibitors N-Ethylmaleimide (NEM; 10 mM) and 2-Iodoacetamide (IA; 20 mM) using an anti-FLAG antibody (F2555, Sigma-Aldrich). A coomassie-stained loading control is shown at bottom. (**B**) Western blot of HeLa proteome labeling by **4** (left 100 μM; right 23 μM; 1.5 h, 37 ˚C) in the presence or absence of the DUB labeling probe Ub-VME (23 μM) using an anti-FLAG antibody (ab1162, Abcam). A coomassie-stained loading control is shown at bottom

2.2.2 Synthesis of SUMO3 Gly92Sec N-methylamide

Direct oxidation of **2** affords the desired Dha product, and demonstrates oxidative orthogonality to methionine (M1), but does not afford insight into cysteine orthogonality. Extension of the oxidative chemistry to the small ubiquitin-like modifier isoform 3 (SUMO3) broadens the scope to a substrate with a single Cys (C47) and three Met residues (M43, 54 and 77), and would enable characterization of more regulatory proteases. Synthesis of SUMO3 G92U N-methylamide (**5**) is facile, and proceeds in good yield, although the propensity for symmetric diselenide formation is markedly lower; while diselenides are more stable than selenylsulfides due to more

extensive orbital overlap, the intramolecular selenylsulfide of SUMO3Sec forms faster than the diselenide both in the presence and absence of denaturant (**Figure 2.4**).

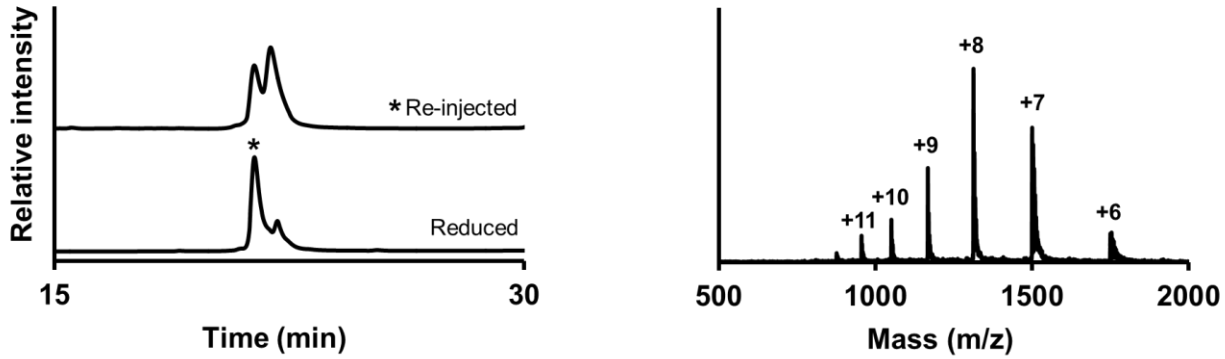


Figure 2.4. Rapid oxidation of 5 from selenol to selenylsulfide. Reduction of purified **5** with 1 M DTT, pH 8 proceeds nearly to completion (*); however, immediate repurification of the reduced fraction shows rapid formation of a second peak. The two peaks are indistinguishable by ESI-MS. Calculated for **5** $[M+H]^+$ 10500.6 Da, observed 10500.8 ± 1.5 Da.

Rapid oxidation to a stable selenylsulfide renders direct oxidation of the selenol ineffectual as a route to Dha, a finding that is consistent with both the small molecule and selenoprotein literature. Selenoproteins with peroxidase activity employ C-X-X-U motifs that enable selenocysteine to reduce peroxides, followed by cysteine reduction of the resultant selenenic acid, a feature that prevents terminal oxidative inactivation of the catalytic pair by over oxidation.[20] Were this the case in the SUMO3Sec oxidation it would do nothing to prevent further oxidation of the resultant selenylsulfide, which class is known to undergo further oxidation to selenosulfonates rather than elimination .[17] Furthermore, in the event that selenol oxidation proceeds to the seleninic acid prior to thiol reduction, then the result is thioseleninate formation, which isomerizes to the selenosulfinate and can react with an equivalent of hydroxyl radical to afford the selenenic acid and thioxyl radical.[17] Oxidative approaches predictably failed to yield Dha products without concomitant oxidation of other residues in **5**.

Where controlled selenol oxidation is untenable, the differences in the pKa's and nucleophilicities of Sec and Cys proffer an alternate route to Dha. Selective selenol alkylation is

attainable at low pH, though the local environment contributes significantly to the selenol pKa.[21,22] Selective selenol alkylation followed by oxidation was pursued, and the dehydroalanine product detected, but ultimately abandoned due to the chromatographic inseparability of significant byproducts.[23]

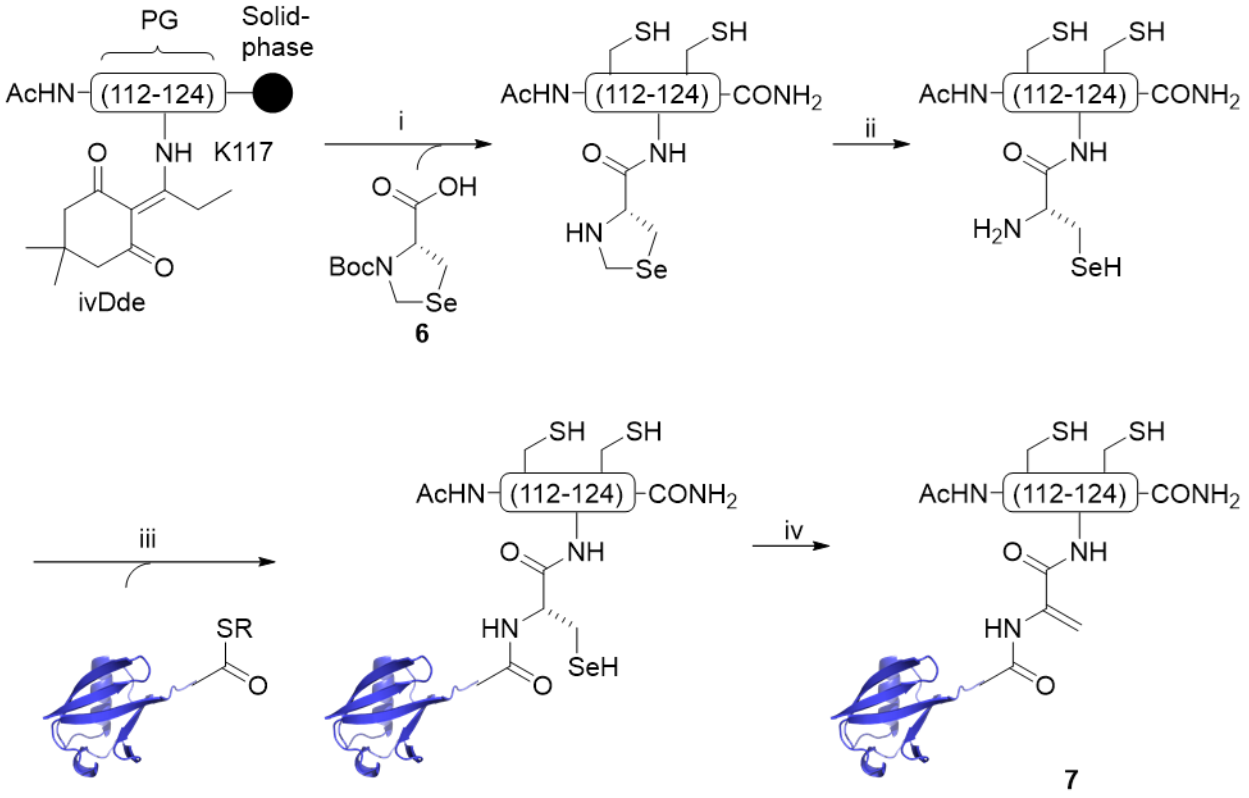
2.2.3 Synthesis and reactivity of TRIM25(112-124) K117(UbDha)

Chemical control of Dha installation enables preparation of electrophile-bearing ubiquitylated proteins and peptides, with which the specificity of deubiquitinases can be surveyed. An optimal test case satisfies a few criteria: (1) the protein is known to be monoubiquitylated at a specific lysine; (2) the deubiquitinase that targets the specific lysine is known; (3) the lysine is proximal to one or more cysteines. The site-specific monoubiquitylation criterion (1) serves to exclude substrates for which the affinity of the cognate deubiquitinase depends on binding multiple Ubs in a polyUb chain. Polyubiquitylated proteins are synthetically tractable, though adding numerous additional steps to the synthesis.[24–26] The known deubiquitinase criterion (2) simplifies validation of probe reactivity *in vitro*. The proximal cysteine criterion (3) enables demonstration of cysteine orthogonality, and opens access to semi-synthesis of a full length ubiquitylated protein probe.

The E3 ligase TRIM25 conforms to the established criteria with identified monoubiquitylation sites at K117, K264, K320 and K416, of which K117 is flanked by five cysteine residues in the 20 surrounding amino acids; furthermore, multiply ubiquitylated TRIM25 is completely deubiquitylated by USP15.[27,28] Moreover, the interaction between USP15 and deubiquitylated TRIM25 has been partially characterized by co-immunoprecipitation of tagged truncant constructs, and the mediating domains narrowed to the USP15 C-terminal His-Box (residues 535-952) and the TRIM25 N-terminal B-box (residues 91-200).[28] Thus, installation of a Ub Gly76Dha probe at K117 of TRIM25 enables characterization of cysteine orthogonality, and conservation of some sequence elements important to recognition by the deubiquitinase.

Introduction of a Sec ligation handle as a lysinyl isopeptide requires preparation of an SPPS compatible Sec. Protection of Sec as an N-boc selenaproline (**6**) enables installation as a lysinyl isopeptide and simplifies initial peptide purification by preventing intra- and intermolecular diselenide and selenylsulfide formation. Furthermore, an N-boc selenazolidine lysinyl isopeptide is a good candidate for amber suppression, as the thioether isostere has already been successfully incorporated by this strategy.[29] Incorporation of a probe in full length TRIM25 is simplified by the amber suppression strategy, which, though reducing the protein yield per liter of cells grown, avoids the issues of yield and fold preservation entailed in sequential ligation strategies. Thus the synthesis of **6** from L-selenocysteine in two steps with a 34% isolated yield (**Figure 2.5.1**)—though a synthesis reported during this investigation achieved 90% yield.[30]

The TRIM25 sequence spanning K112 to A124 incorporates C119 and C122, as well as M123 to survey the swath of sulfur containing functional groups. The peptide is synthesized on Rink amide resin by standard Fmoc chemistry, with K117 bearing ε-ivDde protection for orthogonal deprotection and functionalization with **6**, and N-terminal is acetylation to mimic the electronics of the full length protein (**Scheme 2.4**). Peptide cleavage and RP-HPLC purification furnishes the desired product in 13% yield based on resin loading. Selenazolidine deprotection proceeds by nucleophilic attack on the bridging methylene of the selenoaminal, and is readily accomplished with a nucleophile like methoxyamine the oxime product of which is sufficiently electron rich to minimize any reverse reaction. Nonetheless, the solubility of the TRIM25 peptide, and propensity of the unmasked selenol to oxidize necessitated reaction in a degassed solution of 5 M guanidine hydrochloride, 1.6 M methoxyamine hydrochloride, pH 5, which afforded the desired product in 80% yield after one hour at 25˚C. The solubility and oxidation susceptibility of the peptide substrate presage the challenge of the ligation with Ub(1-75) α-thioester, which requires 6 M guanidine hydrochloride and 5% DMF as a co-solvent to prevent peptide precipitation over the course of the ligation (**Scheme 2.3**).



Scheme 2.3. Synthesis of 7. (i) (1) 5% H2N−NH2 in DMF; (2) **6**, DIC/Oxyma, 13%. (ii) (1) 5 M Gn-HCl, 1.6 M MeONH2, pH 5, 80%; (iii) Ub(1-75) α-thioester, 5% DMF, 6 M Gn-HCl, 100 mM MPAA, 100 mM NaPi, 1 mM EDTA, pH 7.5, 43%. (iii) 28 mM α,α′-dibromoadipoyl(bis)amide, pH 3.4, 0 °C, 12 h, 27% (1UBQ).

Purified TRIM25(112-124)K117(UbSec) was not reliably oxidized to TRIM25(112-124)K117(UbDha) (**7**) without two to six overoxidations (**Figure 2.5.3**). Selective alkylation with iodoacetamide, cysteine disulfide protection, and oxidation with peroxide was initially promising, affording a 34% isolated yield. However, sensitivity to small variations in pH and temperature hindered reproducibility and scalability. The selective selenol alkylation step was reasonably well controlled, prompting investigation of the bis-alkylating reagent α,α'-dibromoadipoyl(bis)amide deployed by Chalker et al in the conversion of Cys residues to Dha.[7] We were gratified to observe that low pH was sufficient for selective conversion of Sec to Dha with a modest 27% yield. Subsequent efforts have demonstrated the feasibility of alkylative elimination at a more mild pH

in different protein contexts, which is consistent with the observed orbital-interaction-dependent properties of selenol and selenide small molecules in conventional organic contexts.[21,22,31]

Evaluation of **7** by circular dichroism confirmed the integrity of the β grasp fold. Incubation of **7** with USP15 resulted in labeling detectable by α-Ub western blot (**Figure 2.5**). Consistent with what has been shown by Mulder et al, the probe appears to undergo competing cleavage of the isopeptide as observed by SDS-PAGE.[19] These results validated the steps for site specific ubiquitylation at a Sec modified Lys, and electrophile formation; however, proteolysis of the isopeptide by DUBs inspired synthesis of a second selenol designed to mitigate proteolysis.

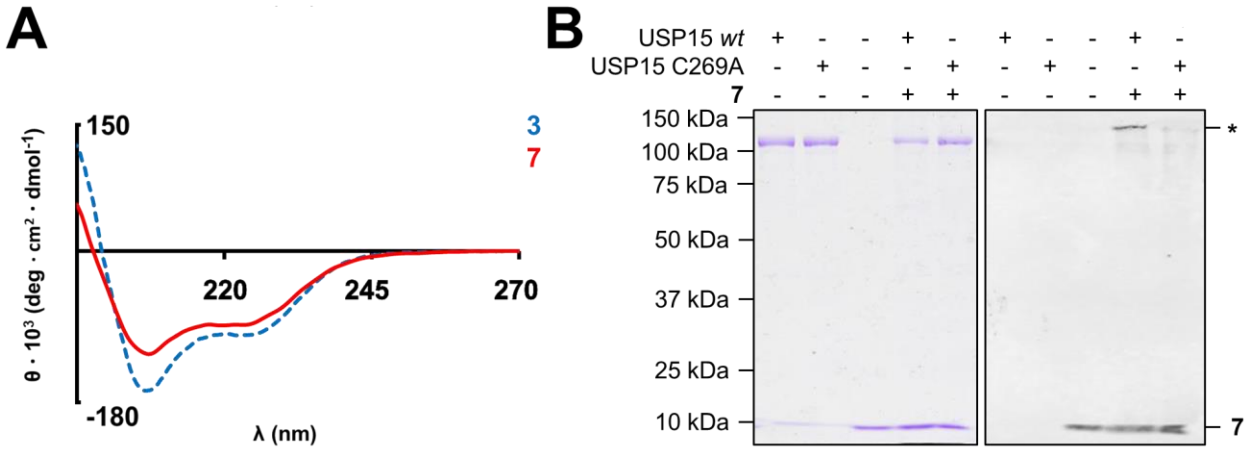
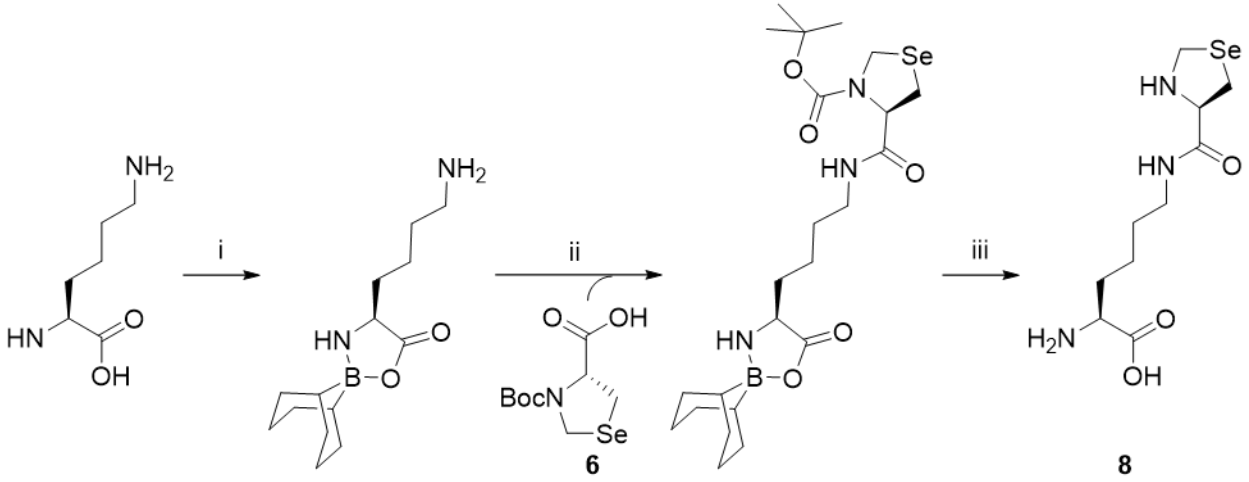


Figure 2.5. Structure and reactivity of 7. (**A**) Comparison of the CD spectra of **7** and **3**. (**B**) 10% SDS-PAGE (left) and Western blot (right) of **7** incubated with full-length USP15; (*) **7**-USP15 wt adduct (P4D1, Santa Cruz Biotech).

2.2.4 Synthesis and Incorporation of N$^{\varepsilon}$-L-Selenaprolyl-L-Lysine

Probe incorporation at lysine residues distal from the N- and C-termini, in structured regions, and in proteins not amenable to refolding is feasible by amber suppression of a non-canonical lysine isopeptide. The lysyl isopeptide of **6** is synthesized by condensation onto L-lysinato-borabicyclononane, followed by global deprotection with trifluoroacetic acid to afford N$^{\varepsilon}$-L-Selenaprolyl-L-Lysine (**8**) (**Scheme 2.4**) (prepared by Nicholas Senger, Chatterjee Lab). The amber suppression of Nε-L-Thiaprolyl-L-Lysine has been reported, and enabled site-specific modification with Ubls following methoxylamine deprotection of the thiazolidine.[29] **8** was screened against a >80 promiscuous tRNA/aminoacyl synthetase pairs in *E. coli* (In collaboration with Dr. Ryan Mehl, Oregon State University).[32–34] Incorporation was observed with neither the N$^{\varepsilon}$-L-Thiaprolyl-L-Lysine synthetase, nor any other.



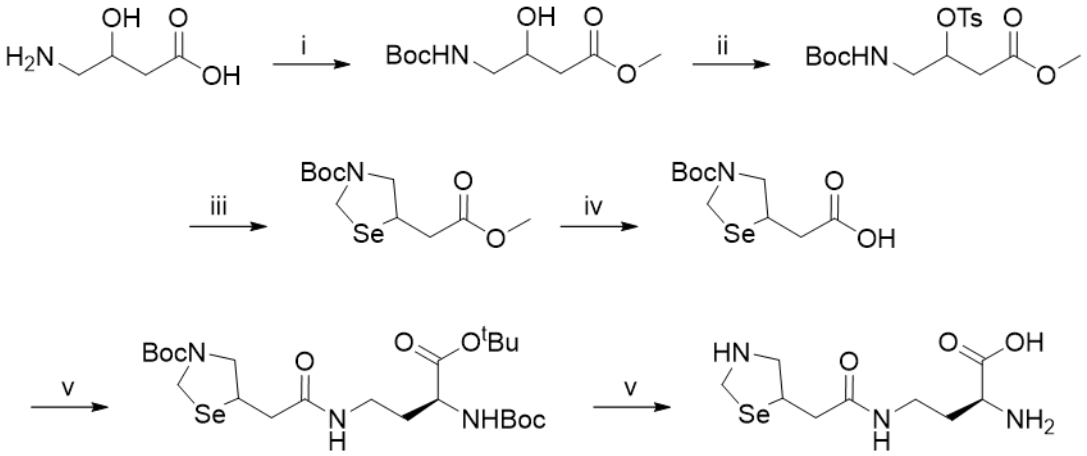
Scheme 2.4. Synthesis of 8. (i) 9-BBN, DIEA, MeOH, reflux. (ii) DIC, TEA, ACN, 71%. (iii) 95:5 TFA:water, 80%.

Initial incorporation tests lacked a preliminary evaluation of ncAA cytoxicity and growth inhibition. A counter screen for toxicity attributable to the trifluoroacetate counter ion found no significant inhibition of growth. Tolerance of linear selenides is evidenced by substitution of selenomethionine for methionine in defined media, as well as non-canonical Se alkylated

selenocysteine-derived photocrosslinkers.[35,36] Nonetheless, selenazolidine derivatives of selenocysteine have been identified as potentially cytotoxic in bacterial and mammalian cells.[37] Selenaproline and Sec toxicity in *E. coli* is attributed to impaired L-Cystine uptake, though a precise mechanism is lacking.[38]

2.2.5 Synthesis and Incorporation of (R)-4-[2-[5-selenazolidinyl]acetyl]amino-2-aminobutanoic acid

(R)-4-[2-[5-selenazolidinyl]acetyl]amino-2-aminobutanoic acid (**9**) cannot be metabolized to a Cys isostere, reducing the likelihood of interference with Cys pathways.[22] Furthermore, **9** moves the isopeptide that we had observed to proteolyze, and selenol β elimination installs the electrophilic methinyl carbon isosteric to the carbonyl carbon of the native isopeptide. The viability of this electrophile has already been demonstrated in a diubiquitin that successfully captured linkage specific DUBs.[39] The thiol isostere of the ligation handle is also validated in NCL, though Mulder et al introduced the thiol during chemical synthesis of Ub on the solid phase. Rather than revisit incorporation on the solid phase, **9** was synthesized for amber suppression (**Scheme 2.2.6**).



Scheme 2.5. Synthesis of 9. i) (1) $SOCl_2$, MeOH. (2) $Boc_2(O)$, TEA, ACN, 75% 2 steps; ii) TsCl, pyridine, 76%; iii) (1) Se^0, $NaBH_4$, 3:1 DMF:H_2O; (2) 1:1 TFA:DCM; (3) $NaBH_4$, 37% H_2CO, MeOH; (4) $Boc_2(O)$, TEA, MeOH, 37% 4 steps; iv) Me_3SnOH, $C_2H_4Cl_2$, 75 ˚C, 50 hrs, quant; v) HATU, DIEA, N^{α}-Boc-L-DAB-OtBu, DMF, 49%; vi) 90:5:5 TFA:TIS:H_2O, 79%.

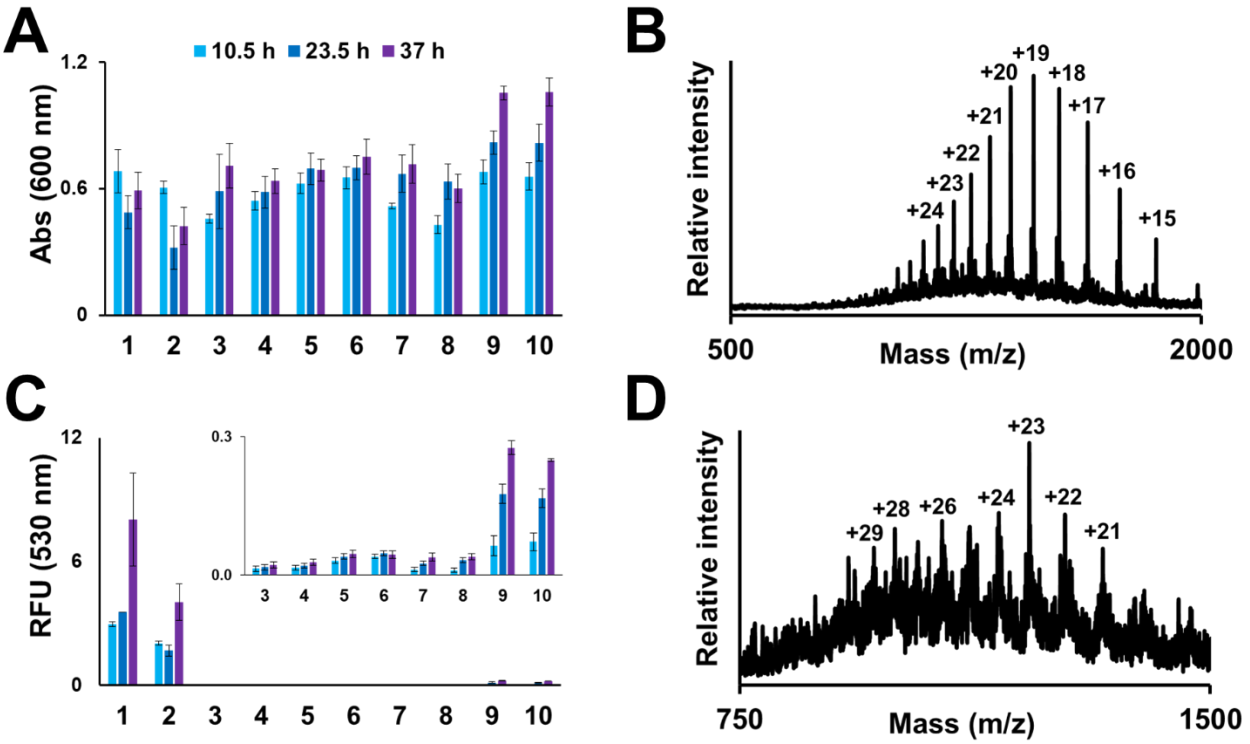


Figure 2.6. Testing incorporation of 9. (**A**) Evaluation of BL21ai E. coli (1-4) and Dh10b (5-10) growth in the presence or absence of **9**: (1) sfGFP wt; (2) sfGFP wt + **9**; (3) sfGFP TAG150; (4) sfGFP TAG150 + **9**; (5) C6; (6) C6 + **9**; (7) C7; (8) C7 + **9**; (9) D6; (10) D6 + **9**. (**B**) ESI-MS of sfGFP purified from wt growth; calculated for sfGFP $[M+H]^+$ 27827.3 Da, observed 27830.8 ± 3.4 Da. (**C**) Evaluation of sfGFP expression by fluorescence of BL21ai cultures (485 nm excitation, 530 nm emission). (**D**) ESI-MS of sfGFP TAG150 purified from D6 growth; calculated for sfGFP N150**9** $[M+H]^+$ 27989.41 Da, observed 27876.5 ± 22.8 Da.

Synthesis of **9** from 4-amino-3-hydroxybutyric acid proceeds in a meager 8% yield over 10 steps. The first three steps follow the synthetic scheme of Mulder et al, and are reasonably robust (57%). The four steps in which the selenazolidine is formed afford only 37% and represents a reasonable target for optimization stepwise optimization. Elimination of the tosylate, or subsequently installed chalcogen was a reported risk from the synthesis of the thiol ligation handle. We thus opted to use disodium diselenide (Na_2Se_2) produced by *in situ* reduction of selenium in the nucleophilic substitution at the β carbon.[40,41] The boc protection of the cyclized selenazolidine is likely the limiting step in this series, as with the boc protection of **6**. Bases tested in this step included DBU, K_2CO_3, DIEA and TEA, of which the first two were ineffective, and the last was the most easily removed organic base, which critically reduced elimination during concentration of the crude.

Following amidation and global deprotection, purified **9** was tested for cytotoxicity. At the 1 mM concentration of ncAA employed in our incorporation screens no *E. coli* grow in the presence of **9**, however, at 0.5 mM cultures grow at rate only slightly slower than wild type. Selections conducted at 0.5 and 0.25 mM recapitulated colony growth, but gave weak candidates for incorporation. The three most promising cell lines were tested, and again exhibited healthy growth; however, the major protein products were truncants. The expressed sfGFP wt was purified and compared to the amber codon sfGFP (N150*) cell lines with orthogonal synthetase/tRNA pairs (C6, C7, D6), and LCMS confirmed background non-specific amino acid incorporation at N150 (**Figure 2.6**). However, the survival and near wildtype growth rates were sufficient evidence to merit a selection trial, the first round of which (conducted by Dr. Ryan Mehl, Oregon State University) has been successful.

2.3 Conclusion and outlook

The interactions by which DUBs recognize specific ubiquitylated proteins remain unknown except in a few cases.[42] It is increasingly apparent that many DUBs employ multiple domains to recognize structural features of polyUb chains, and it is hypothesized that similar interactions guide substrate specific USP DUBs.[26,43,44] While Ub and diUb derived electrophiles have enabled identification of many DUBs, and investigation of the mechanisms by which they bind Ub, there has been no forthcoming general method for incorporation of electrophiles at a Ub-substrate linkage. Here we have reported the use of Sec as a handle for NCL, and an orthogonal handle for electrophile incorporation. β elimination of the Sec selenol in the presence of two Cys residues adjacent to the primary monoubiquitylation site of TRIM25 enabled covalent capture of the DUB known to act on that site.[28]

In order to generalize this strategy to ubiquitylation sites that are onerous semisynthetic targets we prepared two ncAAs incorporating selenazolidine precursors to 1,2-aminoselenol ligation handles. In the process of screening these ncAAs we identified conditions under which one is tolerated, which can be used in a rescreen of the first ncAA tested. If the first ncAA proves cytotoxic at lower concentrations we will have learned that the selenaproline moiety is uniquely challenging to *E. coli*. Meanwhile, the success of **9** in even a single round of selections holds forth hope of a promiscuous synthetase.

2.4 Experimental Procedures

2.4.1 General Methods

Rink-amide resin (0.46 mmol/g substitution) was purchased from Chem-Impex (Wood Dale, IL). Fmoc-L-amino acids were purchased from either AnaSpec (Fremont, CA) or AGTC Bioproducts (Wilmington, MA). Other chemical reagents were purchased from either Fisher Scientific (Pittsburgh, PA) or Sigma-Aldrich Chemical Company (St. Louis, MO). DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA). Gene sequencing was performed by and Genewiz (South Plainfeld, NJ). Plasmid mini-prep, PCR purification and gel extraction kits were purchased from Qiagen (Valencia, CA). Restriction enzymes were purchased from either Fermentas (Thermo Fisher Scientific, Philadelphia, PA) or New England BioLabs (Ipswitch, MA). Reversed-phase HPLC (RP-HPLC) was performed on a Varian (Palo Alto, CA) ProStar HPLC with either a Grace-Vydac (Deerfield, IL) analytical C18 column (5 micron, 150 x 4.6 mm) at a flow rate of 1 ml/min, or a Grace-Vydac preparative C18 column (10 micron, 250 x 22 mm) at a flow rate of 9 ml/min; RP-HPLC was performed using 0.1% trifluoroacetic acid (TFA) in water (A) and 90% acetonitrile, 0.1% TFA in water (B) as the mobile phases. Solid phase peptide synthesis was performed on a Liberty Blue Automated Microwave Peptide Synthesizer (CEM Corporation, Matthews, NC). Mass spectrometric analysis was conducted on a Bruker (Billerica, MA) Esquire, or a Thermo Scientific (Waltham, MA) LTQ Obitrap ESI-MS. Analytical RP-HPLC-mass spectrometry (LC-ESI-MS) was performed on a Hewlett-Packard (Palo Alto, CA) 1100-series LC linked to the Bruker Esquire ESI-MS with an Agilent (Santa Clara, CA) Zorbax C18 column (3.5 micron, 100 x 2.1 mm) using 5% acetonitrile, 1% acetic acid in water (C) and acetonitrile, 1% acetic acid (D) as the mobile phases. NMR spectra were recorded on Bruker Avance AV-300, AV-301, or AV-500 instruments. Circular dichroism measurements were performed on a JASCO (Easton, MD) J-720 spectropolarimeter. Fluorescence was measured with a Biotek synergy H1 hybrid multimode microplate reader (Winooski, Vermont).

2.4.2 Synthesis of (2R,2'R)-3,3'-diselanediylbis(2-amino-N-methylpropanamide)

*Nα,Nα'-di-Boc-L-selenocystine (**1a**)*

To a suspension of L-selenocystine (250 mg, 0.748 mmol) in water were added triethylamine (313 µL, 2.244 mmol) and di-tert-butyldicarbonate (500 mg, 2.244 mmol). The reaction mixture was stirred at room temperature for 2 h. After completion of the reaction, the solvent was evaporated under reduced pressure. The residue was dissolved in ethyl acetate and washed with 10% HCl. The organic layer was dried over anhydrous sodium sulfate, filtered, and evaporated under reduced pressure to obtain the crude product which was purified using column chromatography (10% MeOH/CH_2Cl_2) to obtain **1a** (361 mg, 90%).

^{1}H NMR (500 MHz, $CDCl_3$) δ = 6.22-6.92 (br, 2H), 5.46-5.73 (br, 2H), 4.56-4.87 (br, 2H), 3.50 (d, J = 10.18 Hz, 4H), 1.50 (s, 18H). ^{13}CNMR (125 MHz, CD_3OD) δ 171.75, 155.10, 78.07, 52.91, 26.21, 26.03. HRMS (ESI):m/z Calcd for $C_{16}H_{29}N_2O_8Se_2$ $[M+H]^+$ 537.0254, found 537.8787 $[M+H]^+$.

*tert-butyl (2R,2'R)-3,3'-diselanediylbis(1-(methylamino)-1-oxopropane-3,2-diyl)dicarbamate (**1b**)*

To a solution of **1a** (200 mg, 0.373 mmol) in anhydrous DMF were added methyl amine hydrochloride (75 mg, 1.199 mmol), HOBt (176 mg, 1.305 mmol), and triethylamine (182 µL, 1.305 mmol). The reaction mixture was stirred at 0 °C for 30 min, followed by addition of EDCI (203 mg, 1.305 mmol) at 0 °C. The reaction mixture was stirred at room temperature for 12 h, followed by evaporation of the solvent under reduced pressure. The residue was then dissolved in ethyl acetate and washed with water. The organic solvent was dried over anhydrous sodium sulfate, filtered, and evaporated under reduced pressure to obtain the crude product which was purified using column chromatography (50% EtOAc/hexanes) (142 mg, 68%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.89 (s, 2H), 5.64 (s, 1H), 5.62 (s, 1H), 4.80 - 4.89 (m, 2H), 3.20 (d, J = 6.59 Hz, 4H), 2.84 (s, 3H), 2.83 (s, 3H), 1.48 (s, 18H). ^{13}CNMR (125 MHz, $CDCl_3$) δ 171.08,

156.03, 80.24, 55.64, 36.96, 28.37, 25.92. HRMS (ESI): m/z Calcd for $C_{18}H_{35}N_4O_6Se_2$ $[M+H]^+$ 563.0887, found 563.0874 $[M+H]^+$.

*(2R,2'R)-3,3'-diselanediylbis(2-amino-N-methylpropanamide) (**1**)*

1b was stirred in (100 mg, 0.178 mmol) in 1:1 of CH_2Cl_2 and TFA at room temperature for 2 hours. Solvent was evaporated under reduced pressure to obtain the trifluoroacetate salt of compound **1** in quantitative yield.

^{1}H NMR (500 MHz, D2O) δ 4.25 (d, J = 6.27 Hz, 2H), 3.35-3.49 (m, 4H), 2.79 (s, 6H). ^{13}CNMR (125 MHz, D2O) δ 168.20, 53.50, 28.00, 26.40. HRMS (ESI): m/z Calcd for $C_8H_{19}N_4O_2Se_2$ $[M+H]^+$ 362.9838, found 362.9827 $[M+H]^+$.

2.4.3 Synthesis of (R)-3-(tert-butoxycarbonyl)-1,3-selenazolidine-4-carboxylic acid

*(R)-1,3-selenazolidine-4-carboxylic acid (**6a**)*

L-selenocystine (1 g, 2.99 mmol) was suspended in 38.4 ml of degassed 0.05 M sodium hydroxide containing 11.45 mL ethanol. Sodium borohydride (340mg, 8.89mmol) was added slowly over 10 minutes, and the reaction stirred until colorless (20 minutes). The mixture was then cooled in an ice-bath, and excess sodium borohydride quenched with addition of acetic acid. Formaldehyde (90 mmol, 8 mL 37%) was then added, and the reaction allowed to proceed for 2 hours under nitrogen atmosphere. Ethanol (150 mL) was then added, and the reaction cooled for several hours in an ice bath. A small amount of red precipitate was filtered off, and then the solvent removed under reduced pressure. The residue was taken up in 20mL of ethanol, and cooled to give a cream colored solid, which vacuum filtered and used without further purification (942mg, 87%).

*(R)-3-(tert-butoxycarbonyl)-1,3-selenazolidine-4-carboxylic acid (**6**)*

6a (1 g, 5.56 mmol) was taken up in 15 mL of H_2O. The solution was cooled and diisopropylethylamine (16.48 mmol) was added, followed by di-tert-butyldicarbonate (16.68 mmol)

in 2 mL of methanol. The reaction was allowed to proceed under nitrogen atmosphere for 16 hours, after which the solvent was removed under high vacuum. The residue was taken up in ethyl acetate (100 mL), washed with 1M HCl (18 mL), H_2O (20 mL), then dried and evaporated in vacuo. The crude material was purified by column chromatography (1% MeOH/DCM) to give **3** (654mg, 39%; 34% over two steps).

^{1}H NMR (500 MHz, $CDCl_3$) δ 9.85 (bs, 1H, -COOH), δ 5.14-5.36 (m, 1 H, C(O)CαHN), δ 4.88 (bd, 1H, $SeCH_2N$), δ 4.50 (bd, J = 7.35, 1H, $SeCH_2N$), δ 3.38-3.40 (dd, J = 3.42, $SeCH_2C\alpha$), δ 3.32 (m, 1 H, $SeCH_2C\alpha$), 1.48 (s, 9 H, O^{t}Bu). ^{13}CNMR (500 MHz, $CDCl_3$) δ 176.02, 175.49 (COOH), 153.94, 153.59 (NC(O)O), 81.85 (O*C*$(CH_3)_3$), 62.74 (Cα), 39.92, 38.69 ($SeCH_2N$), 28.25 (OC(*C*$H_3)_3$), 25.19, 24.33 (Se*C*$H_2C\alpha$). HRMS (ESI) m/z Calcd for $C_9H_{16}NO_4Se$ $[M+H]^+$ 282.0244, found 181.9724 $[M\text{-}Boc]^+$, 225.9627 $[M\text{-}tBu]^+$, 282.0258 $[M+H]^+$.

2.4.4 Synthesis of α,α'-di-bromoadipoyl(bis)amide

As previously described.[7]

2.4.5 Synthesis of N$^{\varepsilon}$-L-Selenaprolyl-L-lysine

*L-lysinato-bicyclononylboron (**8a**)*

To 180mL of methanol was added 6.48g of L-lysine monohydrochloride (35.6mmol), followed by 4.6g (6.2mL, 36.0mmol) of N,N-diisopropylethylamine under argon atmosphere. The mixture was stirred for 10 minutes, then 5g 9-borabicyclononane dimer (20.5mmol) was added, and the reaction refluxed for 6 hours under a continuous stream of argon. The solvent was then removed under vacuum, and the residue co-evaporated twice with methylene chloride to give 13.5g of an off-white powder which was used without further purification.

*N_ε-((N-Boc-selenazolidine)-2-carboxamido)-L-lysinato-bicyclononylboron (**8b**)*

500mg of **6** (1.79mmol) was taken up in 20mL of acetonitrile and stirred until fully dissolved. Diisopropylcarbodiimide (280uL, 1.79mmol) was added followed by an equimolar quantity of triethylamine. After a 15 minute equilibration period, 576mg of **8a** was added and the reaction stirred under argon atmosphere for 12 hours. The solvent was then removed in vacuo, and the residue purified directly by flash chromatography to give 392mg of ***8b*** (71%).

*N_ε-L-Selenaprolyl-L-lysine (**8**)*

800mg of **8b** was taken up in 20mL of 95% TFA. The mixture was stirred for 24 hours at room temperature, then the solvent removed and the residue taken up in water. The organic material was extracted extensively with methylene chloride, then the aqueous phase separated and evaporated under high vacuum. The residue so obtained was dissolved in the minimum amount of 0.1% TFA solution, and passed through a short plug of C-18 reverse phase silica gel. The eluent was then evaporated under high vacuum to give 513mg of **8** as the trifluoroacetate salt (80.6%).

^{1}H NMR (301 MHz, D_2O): δ 4.56 (d, 1H), δ 4.55 (t, 1H), δ 4.45 (d, 1H), δ 4.15 (t, 1H), δ 3.85 (s, 3H), δ 3.56 (dd, 1H), δ 3.28 (t, 2H), δ 3.28 (dd, 1H), δ 1.96 (m, 2H), δ 1.59 (m, 2H), 1.45 (m, 2H). ESI-MS m/z Calcd for $C_{10}H_{19}N_3O_3Se$ $[M+H]^+$ 310.0670, found 310.1 $[M+H]^+$.

2.4.6 Synthesis of (R)-4-[2-[5-selenazolidinyl]acetyl]amino-2-aminobutanoic acid

*Methyl 4-((tert-butoxycarbonyl)amino)-3-hydroxybutanoate (**9a**)*

4-amino-3-hydroxybutanoic acid (2.5 g, 21 mmol, Sigma Aldrich) was suspended in methanol dried over 4 Å molecular sieves (75 mL, Fisher Scientific) and brought to 0 ˚C in an ice bath. Thionyl chloride (2.4 mL, 32.76 mmol, Sigma Aldrich) was added dropwise under argon, after which the ice bath was removed and the reaction was allowed to come to room temperature

overnight. After 12 h the reaction was concentrated *in vacuo*, co-evaporated 5x with DCM, and used without further purification. The crude was solubilized in dry acetonitrile (50 mL, Fischer Scientific) and dry TEA was added (11.72 mL, 84 mmol). Di-*tert*-butyl dicarbonate (6.88 g, 31.5 mmol, Sigma Aldrich) was taken up in dry acetonitrile (20 mL, Fischer Scientific), and transferred under argon to the flask containing crude from the previous step. After 12 h the reaction was concentrated *in vacuo*. The resulting residue was taken up in ethyl acetate (100 mL, bulk solvent) and washed thrice (50 mL) with 1 M $KHSO_4$, then thrice with saturated $NaHCO_3$, dried with Na_2SO_4, and concentrated *in vacuo*. The crude was column purified on silica gel (2:1 hexanes:ethyl acetate) to furnish the desired product as a yellow oil (3.6979 g, 75%) (SW-VII-35).

^{1}H NMR (300 MHz, $CDCl_3$): δ 5.08 (bs, 1H, OH), 4.12 (bs, 1H, βCH), 3.72 (s, 3H, OCH_3), 3.58 (m, 1H, NH), 3.11-3.33 (m, 2H, γCH_2), 2.52 (m, 2H, αCH_2), 1.45 (s, 1H, $C(CH_3)_3$). ESI-MS m/z Calcd for $C_{10}H_{19}NO_5$ $[M+H]^+$ 233.26, found 256.3 $[M+Na]^+$ (SW-VII-34).

*Methyl 4-((tert-butoxycarbonyl)amino)-3-(tosyloxy)butanoate (**9b**)*

9a (2.0 g, 8.58 mmol) was solubilized in dry pyridine (12 mL) with stirring and chilled on ice. 4-toluensufonyl chloride (4.9 g, 25.72 mmol, Sigma Aldrich) was added slowly with stirring. The reaction vessel was then purged with argon, and the ice bath removed. After 12 h stirring at room temperature the pyridine was stripped under vacuum, and the crude was co-evaporated thrice with toluene, then thrice with dichloromethane. The residue was taken up in dichloromethane, washed twice with 1 M $KHSO_4$, then twice with saturated NaHCO3. The organic layer was dried with Na_2SO_4, concentrated, and purified on silica gel (3:1 hexane:ethyl acetate to 1:1 hexane:ethyl acetate) to furnish the desired product as an off-white crystalline solid (2.5 g, 76%) (SW-VII-40).

^{1}H NMR (300 MHz, CDCl3): δ 7.78-7.80 (d, 2H, PhCH), 7.33-7.36 (d, 2H, PhCH), 4.92-4.95 (t, 1H, βCH), 4.81 (bs, 1H, NH), 3.59 (s, 3H, OCH_3), 3.41-3.43 (t, 2H, γCH_2), 2.63-2.66 (d, 2H, αCH_2),

2.46 (s, 3H, $PhCH_3$), 1.41 (s, 9H, $C(CH_3)_3$). ESI-MS m/z Calcd for $C_{17}H_{25}NO_7$ $[M+H]^+$ 387.45, found 287.9 $[M\text{-}Boc]^+$ (SW-VII-40).

*Methyl 3-(tert-butoxycarbonyl)-5-selenazolidinylacetoate (**9c**)*

Amorphous selenium powder (305 mg g, 3.87 mmol) was suspended in H_2O (3 mL) on ice, with stirring. To this was added $NaBH_4$ (293 mg, 7.74 mmol) and a further 3 mL H_2O, resulting in bubbling and clarification of the suspension to a colorless solution (*Note: foul vapors*). After 20 minutes a further 3.87 mmol of amorphous selenium was added, and the reaction was gently heated in a 50 ˚C oil bath for a further 10 minutes, which caused the solution to turn burnt red. The solution was removed from heat and a solution of **9b** (3 g, 7.74 mmol) in DMF (18 mL) was added, then the reaction was stirred for 2 hr. The reaction was concentrated *in vacuo*, taken up in EtOAc, and extracted with ice to remove residual DMF, then concentrated *in vacuo*. Boc deprotection of the crude was achieved by stirring with 55:45 DCM:TFA (10 mL) for 30 minutes on ice, then concentrated *in vacuo*. Deprotection was repeated twice more.

This deprotected crude was taken up in MeOH (120 mL) and chilled on ice. $NaBH_4$ (1.464 g, 38.7 mmol) was added portion wise and the vessel was purged with argon, then reaction was stirred for 20’ on ice. A 37% formaldehyde solution (20 mL, 116 mmol) was added, and the reaction was stirred for 2 hours under argon, then concentrated *in vacuo*. The crude solid was washed extensively with EtOAc, and the EtOAc was then concentrated *in vacuo* (1.0 g crude).

Crude (1.0 g) was taken up in acetonitrile (10 mL) and chilled on ice with stirring. Di-*tert*-butyl dicarbonate (3.3 g) was added, followed by TEA (6 mL) added dropwise on ice. The reaction was stirred for 2 hours, then concentrated *in vacuo*. The crude was taken up in DCM (30 ml) and extracted thrice with 0.5 M Acetic acid (3 ml). The organic layer was dried with Na_2SO_4, and concentrated *in vacuo.* Purification was by column chromatography over 0 to 10% EtOAc in hexanes (0.88 g, 2.85 mmol, 37%). (SW-X-60)

1H NMR (300 MHz, VT 288K, $CDCl_3$): δ 4.87 (dd, 1H, *J* = 100.71, 7.72 Hz, NCHSe), 4.49 (dd, 1H, *J* = 40.17, 7.74 Hz, NCHSe), 3.95 (m, 1.5H, SeCHCC, NCHC), 3.79 (m, 1H, NCHC), 3.69 (s, 3H, OCH_3), 3.58 (m, 0.5H, NCHC), 2.90 (m, 1H, CCHC(O)), 2.65 (dd, 1H, *J* = 16.91, 8.62 Hz, CCHC(O)), 1.47 (s, 9H, $C(CH_3)_3$) (SW-IX-14). HRMS (ESI) m/z Calcd for $C_{11}H_{20}NO_4Se$ $[M+H]^+$ 310.06, found 253.9167 $[M-^tBu]^+$, 210.0000 $[M-Boc]^+$, 309.6667 $[M+H]^+$.

*3-(tert-butoxycarbonyl)-5-selenazolidinylacetic acid (**9d**)*

9c (30 mg, 0.1 mmol) was dissolved in 1,2-dichloroethane (3 ml), to which was added trimethyltin hydroxide (54.2 mg, 0.3 mmol). The reaction was heated to 75 ˚C for 50 hrs, then washed with 0.1 M $KHSO_4$, dried with Na_2SO_4, and concentrated *in vacuo*. The crude was purified over a silica plug column in 5% MeOH/DCM to remove residual trimethyltin species (28.7 mg, 0.98 mmol, 98%) (SW-IX-20).

1H NMR (500 MHz, $CDCl_3$): δ 4.96 (bd, 1H, *J* = 91.58 Hz, NCHSe), 4.53 (dd, 1H, *J* = 25.41, 7.77 Hz, NCHSe), 3.95 (m, 1H, βCH), 3.84 (bm, 1.5H, γCH), 3.65 (bm, 0.5H, γCH), 2.86 (bm, 1H, αCH), 2.71 (dd, 1H, *J* = 17.21, 8.47 Hz, αCH), 1.48 (s, 9H, $C(CH_3)_3$).

*1,1-dimethylethyl (R)-4-[2-[3-(tert-butoxycarbonyl)-5-selenazolidinyl]acetyl]amino-2-[tert-butoxycarbonyl]aminobutanoate (**9e**)*

9d (80 mg, 0.27 mmol) is dissolved in dry DMF (500 µL) with stirring and chilled on ice under dry inert gas. To this solution is added first DIEA (56 µL, 0.32 mmol), then HATU (121.7 mg, 0.32 mmol). After 5 minutes of mixing N_α-Boc-diaminobutyric acid tert-butyl ester (99.3 mg, 1.2 mmol) is added, and the reaction is removed from ice and allowed to stir for 18.5 hrs. DMF is removed by partitioning the reaction between ice and ethyl acetate. The organic layer is dried with Na_2SO_4, concentrated *in vacuo* and column purified on silica gel from 0 to 2% MeOH/DCM (74 mg, 49%). *Note:* Run at 2.16 mmol scale the yield drops to 39%. (SW-IX-66/67)

^{1}H NMR (500 MHz, CDCl3): δ 6.48 (bs,1H, NHC(O)), 5.42 (m, 1H, *J* = 32.29, 7.86 Hz, SeCHCC), 4.81 (m, 1H), 4.54 (bs, 1H), 4.21 (bs, 1H), 4.03 (m, 1H), 3.74 (bm, 2H), 3.02 (bs, 1H), 2.69 (m, 1H), 2.51 (m, 1H), 2.06 (s, 1H), 1.5 (m, 2H), 1.47 (s, 9H, $C(CH_3)_3$), 1.46 (s, 9H, $C(CH_3)_3$), 1.44 (s, 9H, $C(CH_3)_3$). ESI-MS m/z Calcd for $C_{23}H_{41}N_3O_7Se$ $[M+H]^+$ 551.56, found 451.198 $[M\text{-}Boc]^+$, 551.110 $[M+H]^+$, 573.155 $[M+Na]^+$.

*(R)-4-[2-[5-selenazolidinyl]acetyl]amino-2-aminobutanoic acid (**9**)*

9e (400 mg, 0.73 mmol) was treated with chilled 90:5:5 TFA:TIS:H_2O (8 ml) for 30' with stirring, then precipitated with ice cold diethyl ether. The precipitate was purified by preparative RP-HPLC (235 mg, 0.58 mmol, 79%). (SW-X-98)

^{1}H NMR (500 MHz, D_2O): δ 4.43-4.49 (m, 2H, NCH_2Se), 4.22 (m, 1H, SeCHCC), 4.02 (m, 1H, NCHC), 3.89 (m, 1H, NCHC), 3.35-3.47 (m, 3H, γCH_2, αCH), 3.01-3.07 (m, 1H, CHC(O)N), 2.72-2.77 (m, 1H, CHC(O)N), 2.06-2.23 (m, 2H, βCH_2). ^{13}CNMR (500 MHz, D_2O): δ 175.00, 173.29, 173.24, 171.46, 54.07, 54.03, 53.69, 50.47, 39.79, 39.76, 38.89, 38.23, 38.20, 35.07, 35.01, 34.68, 34.01, 29.30 (diastereomers). ESI-MS m/z Calcd for $C_9H_{18}N_3O_3Se$ $[M+H]^+$ 296.05, found 590.90 $[2M+H]^+$, 296.04 $[M+H]^+$.

2.4.7 Molecular cloning

Cloning FLAG-Ub(1-75)-MxeGyrA-CBD and SUMO3(1-91)-MxeGyrA-CBD

The plasmid pTXB1-ubiquitin(1-75)-MxeGyrA-CBD, containing the human ubiquitin gene, *ub(1-76)*, was used to generate the plasmid pTXB1-FLAG-ubiquitin(1-75)-MxeGyrA-CBD, which adds an N-terminal FLAG tag (MDYKDDDDKA).[45] The plasmid pTXB1-SUMO3(1-92), containing the human SUMO3 gene *Smt3(1-92)*, was used to generate the plasmid pTXB1-SUMO3(1-91), which lacks the C-terminal Gly of SUMO3.[46] The following primers were used to generate a FLAG-ubiquitin(1-75) amplicon for restriction cloning, and prepare pTXB1-SUMO3(1-91) by site-directed mutagenesis (QuikChange kit, Agilent Technologies, Santa Clara, CA):

Primer	DNA Sequence (5'- to -3')
FLAG-Ub(1-75)-FP	CTTTAGAAGGAGATATACATATGGATTACAAGGATGACGACGATAAAGCGATGCAGATCTTCGTGAAGACTCTG
FLAG-Ub(1-75)-RP	GGTGGTTGCTCTTCCGCAACCTCTGAGACGGAGTACCAGGTGCAGGGT
SUMO3(1-91)-FP	ATCGACGTGTTCCAGCAGCAGACGGGATGCATCACGGGAGATGCACTAGTTGCC
SUMO3(1-91)-RP	GGCAACTAGTGCATCTCCCGTGATGCATCCCGTCTGCTGCTGGAACACGTCGAT

The FLAG-ubiquitin(1-75) amplicon was digested with *NdeI* and *SapI*, treated with Calf Alkaline Intestinal Phosphatase, purified by agarose gel, and ligated into a similarly digested pTXB1 vector with T4 DNA ligase (New England Biolabs, Ipswitch, MA). The ligation mix was used to transform ultracompetent DH5 alpha *E. coli*. Gene sequences were confirmed by sequencing with the T7 forward primer (Genewiz).

2.4.8 Protein overexpression and purification

Overexpression of protein-MxeGyrA-CBD fusion and purification of protein-MESNa α-thioester

Overexpression was achieved in BL21 *E. coli*, which were grown at 37 ˚C to an OD_{600} of 0.4, then cooled to 16 ˚C and induced over 9 h with 0.3 mM IPTG. Cells were harvested by centrifugation at 6,000xg for 15 min at 4 ˚C. The cell pellet was resuspended in 22 mL lysis buffer (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 Mm KH_2PO_4, pH 7.4) and lysed by sonication, and centrifuged for a further 30 min at 20,000xg at 4 ˚C. The lysate supernatant was passed through a 0.45 µm filter then applied to a 10 mL chitin res pre-equilibrated with lysis buffer and bound over night at 4 ˚C. The column was washed with lysis buffer (20 Column Volumes, CV), and 35 mL thiolysis buffer was applied to the column (100 mM sodium 2-mercaptoethanesulfonate (MESNa), 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 Mm KH_2PO_4, pH 7.4). The first elution was collected after 48 hr at 4 ˚C, and a second 35 mL of thiolysis buffer was applied for a further 48

hr. The eluted protein was purified by C18 Preparative RP-HPLC (30-70% B, 60 min) to yield 9.2 mg FLAG-ubiqiuitin(1-75)-MESNa thioester was obtained from 3 L.

Overexpression of UCH-L3 WT and C95A mutant

UCH-L3 wt and C95A were expressed as previously described.[45]

Expression and purification of His_6-MBP-USP15

USP15 was expressed recombinantly in BL21(DE3) *E. coli*, as an N-terminal 6xHis-MBP (maltose binding protein) fusion. Cells were grown at 37ºC with shaking at 180 rpm until OD_{600} reached 0.4. Protein expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.2mM). Cells were then grown overnight at 16ºC and then harvested by centrifugation for 30 minutes at 4000 x g, 4ºC. The cell pellet was resuspended in lysis buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1mM TCEP, and 0.25 mM PMSF) and then lysed by sonication. The lysate was clarified by centrifugation for 30 min at 75000 x g, 4ºC. USP15 was then purified using two chromatographic steps: nickel affinity chromatography followed by anion exchange after the MBP fusion was removed using TEV protease. Enzyme activity was confirmed with the fluorogenic probe Ub-AMC, as previously described.[47]

2.4.9 Expressed protein ligation of 1 and Ub(1-75) or SUMO3(2-92) α-thioester

Protein α-thioester (0.5 mM) was reacted with **1** (1 mM) in ligation buffer (6 M Guanidine HCl (Gn), 200 mM NaPi, 200 mM mercaptophenylacetic acid (MPAA), pH 7.5) at 25 ˚C. After 12 hrs the reaction was brought to pH 3 and extracted thrice with diethyl ether to remove residual MPAA. The reaction was then brought to 1 M in dithiothreitol (DTT), and adjusted to pH 8 for a further 12 h at 25 ˚C. The reaction was purified by C18 analytical RP-HPLC (30-55% B, 30 min) (Ub 76%; SUMO3 84%).

*2.4.10 Oxidative conversion of **2** to **3***

2 (0.35 mM) is sonicated to give a heterogeneous suspension in 100 mM NaOAc, pH 5.5. To this suspension is added a solution of $NaIO_4$ in 100 mM NaOAc, pH 5.5, such that the final concentration of $NaIO_4$ is 0.35 mM. The suspension in allowed to react for 2 hr at 25 ˚C protected from light, after which time crystalline Gn is added to a concentration of 6 M. Following denaturation the reaction was allowed to proceed for a further 48 hr. The reaction was purified by RP-HPLC (C18 analytical, 30-55% B, 45 min, 37%).

2.4.11 Iodoacetamide alkylation of SUMO3 Gly93Sec N-methylamide

SUMO3 Gly93Sec (1.1 mg scale, 84 µM S3Sec) is dissolved in buffer (6 M Gn, 100 mM NaOAc, pH 4) and allowed to stand 4 hours under ambient atmosphere at ambient temperature, then cooled to on ice. The samples are reduced with chilled TCEP (1.6 mM) for 2 minutes, then treated with chilled iodoacetamide (17.7 mM) for 60 minutes on ice. The reaction is purified by RP-HPLC (C18 Analytical, 30-40% B, 45 min).

2.4.12 Solid-phase peptide synthesis

Synthesis of AcHN-TRIM25(112-124)- Rink-amide resin

The peptide AcHN-KEAAVK(U)TCLVCMA-CONH2 was synthesized by microwave-assisted SPPS on a 0.1 mmol scale utilizing standard 9-fluorenylmethoxycarbonyl (Fmoc)-based Nα-deprotection chemistry. From Rink-amide resin (0.22 g, 0.46 mmol/g) each amino acid was coupled in 5.25 molar excess based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 20% piperidine in DMF for 65 s at 90 °C. Coupling reactions were undertaken for 2 min at 90 °C with a mixture of Fmoc-amino acid (0.53 mmol), O-(Benzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HBTU, 0.51 mmol) and N,N-Diisopropylethylamine (DIEA, 1.1 mmol) in DMF. The N_ε of K117 was orthogonally protected with the 1-(4,4-dimethyl-2,6-dioxocyclohexylidine)-3-methylbutyl (ivDde) protecting group. The

peptide was acetylated at the N_α position reaction with acetic anhydride (2 mmol) and DIEA (2 mmol) in DMF twice, each time for 60 minutes.

Installation of K117-iso-U

Deprotection of K117 Nε was accomplished by reacting resin bound peptide with a solution of 5% hydrazine in DMF overnight. **6** (0.15 mmol) was coupled to K117 Nε with *N,N′*-Diisopropylcarbodiimide (DIC, 0.15 mmol) and Ethyl cyano(hydroxyimino)acetate (Oxyma, 0.15 mmol) overnight. Coupling completion was confirmed by test cleavage, HPLC purification and subsequent ESI-MS. The remaining resin-bound peptide was cleaved and deprotected with Reagent K (TFA: thioanisole: H_2O: phenol: 1,2-ethanedithiol 82.5:5:5:5:2.5 v/v) for 1 hr at room temperature. Peptide was precipitated with ice cold diethyl ether, then washed twice more. Dried peptide was taken up in 6 M Gn, 100 mM NaOAc (pH 4), 50 mM tris(2-carboxyethyl)phosphine (TCEP), and purified by C18 preparative RP-HPLC (0-73% B, 60 min), yielding 13% based on resin loading. Deprotection of the selenazolidine was accomplished by dissolving the peptide (2 mM) in 5 M Gn, 1.6 M Methoxylamine HCl, pH 5 and reacting for 4 h. Immediately prior to purification samples were chilled on ice and reduced with TCEP (10 mM) for 30 sec. and purified by RP-HPLC (C18 preparative, 0% B, 15 min; 0-100% B, 30 min, >80%).

2.4.13 Expressed protein ligation of TRIM25(112-124) K117(Sec) and Ub(1-75) α-thioester

TRIM25(112-124) K117(Sec) (2.5 mM, 1.2 mg) was ligated with Ub(1-75)-MESNa α-thioester (0.5 mM, 2.2 mg) in 500 µL of degassed ligation buffer (6 M Gn, 100 mM MPAA, 100 mM NaPi, 1 mM EDTA, 5% DMF, pH 7.5) at 25 ˚C. After 24 h the reaction was acidified with phosphoric acid to pH 3 and washed five times with diethyl ether to remove excess MPAA. The reaction was then chilled on ice and reduced for 30 sec with TCEP (25 mM). The reduced reaction mix was purified by RP-HPLC (C18 Analytical, 20-60% B, 30 min gradient) (1.1 mg, 42.6%).

2.4.14 Synthesis of 7 from TRIM25(112-124) K117(UbSec) and α,α'-dibromoadipoyl(bis)amide

TRIM25(112-124) K117(UbSec) (0.3 mM, 3 mg/ml) is dissolved in denaturing buffer (8 M Urea, 100 mM NaOAc, pH 3.4), chilled on ice, and reduced with TCEP (15 mM) for 10 min. A solution of α,α'-dibromoadipoyl(bis)amide (280 mM) in DMF added to 10% of the total reaction volume, and kept on ice for a further 12 hr. The reaction is purified by RP-HPLC (C18 analytical, 30-50% B, 30 min, 27%).

2.4.15 General protease labeling

The probe **3** (1.8 nmol) is solubilized in a volume of 8 M Urea that is 5% of the final reaction volume. The protease (0.252 nmol) is activated with 1 mM TCEP for 10 min at 25 ˚C, then added to the solubilized probe. Reactions are placed in a 37 ˚C water bath for 12 hrs, and stopped by the addition of 1% TFA to a concentration of 0.2% TFA. Laemmli sample buffer is added and samples are boiled for 5 minutes with 100 mM TCEP, then analyzed by SDS-PAGE and western blotting.

2.4.16 USP15 labeling

The probe **7** (1 mg) is solubilized in 8 M Urea to make a 2 mM stock. It is further diluted in reaction buffer (50 mM HEPES, 100 mM NaCl, pH 7.5) to make 200 μM working stock. USP15 (1.2 μM) is incubated in the reaction buffer for 10 min at 25˚C, followed by addition of the probe (70 μM), and the reaction is run for 12 h at 37 ˚C. Laemmli sample buffer is added and samples are boiled for 5 minutes with 10 mM DTT, then analyzed by SDS-PAGE and western blotting.

2.4.17 HeLa whole-cell proteome labeling competition assays

*Competition of **4** with N-Ethylmaleimide and 2-Iodoacetamide*

HeLa cells were cultured under standard conditions and collected by trypsinization followed by centrifugation. The pellet was washed twice with PBS, then cells were resuspended at $3x10^6$ cells per 100 µL in lysis buffer consisting of 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM Methionine, 1 mM PMSF, and 0.5% (v/v) IGEPAL CA-630. A DUB-inhibited lysate was prepared by lysing cells in an identical lysate buffer supplemented with 20 mM 2-Iodoacetamide (IA) and 10 mM N-Ethylmaleimide (NEM). Cells were incubated on ice for 20 min then lysed by brief sonication. The lysate was clarified by centrifugation in a 4 ˚C microcentrifuge at 13,500 rpm for 20 min. The total protein concentration in lysates was determined by a Bradford assay.

DUB-labeling assays were performed in 100 mM Na2HPO4, 150 mM NaCl, pH 7.5. To 100 µg of clarified lysate, **4** was added to a final concentration of 100 µM from a stock solution of **4** dissolved in 50:50 (v/v) H_2O:ACN (≤2% total reaction volume). Reactions were incubated at 37 °C for 90 minutes and then quenched with four volumes of 8 M Gn-HCl. Samples were purified by Analytical RP-HPLC (Grace-Vydac, C4, 5 micron, 150 x 4.6 mm) over a 15 minute 10-100% B gradient in order to reduce signal from unreacted 5. The entire collected eluate was lyophilized, resolublized in 40 □L of 8 M Urea and 100 mM DTT, and mixed with 8 µL of 6x Laemmli gel-loading dye. The mixture was heated for 10 min at 60 ˚C and samples were resolved by 10% SDS-PAGE, run at 200 V for 40 min. Protein was transferred to PVDF membrane (Bio-Rad #162-0177) at 4 °C for 12 h at 35 V in Towbin blot buffer with SDS, then immunoblotted with rabbit anti-FLAG (1: 4,000 dilution, Sigma #F2555). The membrane was incubated with secondary HRP-conjugated anti-rabbit (1: 40,000 dilution, Jackson Immunoresearch Laboratories #111-035-003). The blot was visualized by enhanced chemiluminescence with SuperSignal West Dura Extended Duration Substrate (Thermo Scientific #34075) and Kodak BioMax Light autoradiography film (Carestream Health #819-4540).

*Competition of **4** with Ub-VME*

HeLa cells were cultured under standard conditions and collected by trypsinization followed by centrifugation. The pellet was washed twice with PBS, then cells were resuspended at $3x10^6$ cells per 100 µL in lysis buffer consisting of 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM Methionine, 1 mM PMSF, and 0.5% (v/v) IGEPAL CA-630. Cells were incubated on ice for 20 min then lysed by brief sonication. The lysate was clarified by centrifugation in a 4 ˚C microcentrifuge at 13,500 rpm for 20 min. The total protein concentration in lysates was determined by a Bradford assay. Competition assays were performed in 100 mM Na_2HPO_4, 150 mM NaCl, pH 7.5. A stock solution of buffered Ub-VME (Boston Biochem, U-203) was added to 5 dissolved in 50:50 H2O:ACN (≤2% total reaction volume) such that the final assay concentrations were 23 µM in UbVME and either 23 µM or 100 µM in **4**. To the combination of probes in reaction buffer were added 100 µg total lysate protein. Assays were incubated at 37 °C for 90 min and quenched with four volumes of 8 M Gn-HCl. Samples were purified by Analytical RP-HPLC (Grace-Vydac, C4, 5 micron, 150 x 4.6 mm) over a 15 minute 10-100% B gradient in order to reduce signal from unreacted **4**. The collected eluate was lyophilized and then dissolved in 40 µL of 8 M Urea, 100 mM DTT. The resulting solution was mixed with 8 µL of 6x Laemmli gel-loading dye, and heated for 10 min at 60 ˚C. Samples were resolved by 10% SDS-PAGE run at 200 V for 40 min. Proteins were transferred to PVDF membrane (Bio-Rad #162-0177) at 4 °C for 12 h at 35 V in Towbin blot buffer with SDS, then immunoblotted with rabbit anti-FLAG (1: 4,000 dilution, Abcam #ab1162). The membrane was incubated with secondary HRP-conjugated anti-rabbit (1: 40,000 dilution, Jackson Immunoresearch Laboratories #111-035-003). The blot was visualized by enhanced chemiluminescence with SuperSignal West Dura Extended Duration Substrate (Thermo Scientific #34075) and Kodak BioMax Light autoradiography film (Carestream Health#819-4540)

2.4.18 BL21ai expression of sfGFP and sfGFP TAG150

BL21ai *E.coli* were transformed with pDule-tet2.0 and either pBAD-sfGFP-wt (wt) or pBAD-sfGFP-TAG150 (TAG150), and Dh10b were provided transformed with either C6, C7 or D6 mutant tRNA synthetase encoding pBK plasmid and pyrrolysine tRNA coding pALS-sfGFP-TAG150; cells and plasmids were a kind gift of the Mehl Lab (Oregon State University). Cultures of wt, TAG150, C6, C7, and D6 were started in non-inducing media containing tetracycline (100 µg ml^{-1}) and either ampicillin (25 µg ml^{-1}; wt & TAG150) or Kanamycin (50 µg ml^{-1}; C6, C7 & D6).[48,49] Non-inducing cultures were grown overnight and each was used to inoculate two cultures of auto-induction media (non-inducing media supplemented to 0.5% w/v glycerol, 0.05% w/v arabinose, 0.02% w/v lactose) containing the same antibiotics. Cultures were shaken for 75 min. at 37 ˚C, then one of each culture was supplemented with **9** (0.5 mM), while the second culture was supplemented with an equal volume (10 µl) sterile H_2O to serve as a negative control for non-specific incorporation. Culture growth was monitored by absorbance at 600 nm, and sfGFP expression was monitored by fluorescence (485 nm excitation, 530 nm emission) on a synergy H1 hybrid multimode microplate reader. Cultures were grown to 54 hours, then pelleted by centrifugation (5 min at 3 k rpm). Cell pellets were taken up in lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 5 mM imidazole, pH 8) and lysed by three rounds of freeze thawing with LN2, followed by a single pulse of sonication. Cell lysates were clarified by centrifugation (10 min at 17,000 rcf, 4˚C) and syringe filtered (0.45 µm). Each cell lysate was bound to 0.5 mL NiNTA agarose for 1 hr, then washed with 10 column volumes (CV) lysis buffer, 20 CV 20 mM imidazole in lysis buffer, 5 CV 50 mM imidazole in lysis buffer, and eluted in 5 single CV fractions of 250 mM imidazole in lysis buffer. Elution fractions were dialyzed twice against water (1 L at 4 ˚C; 3.5 kDa MWCO), lyophilized, and evaluated by LC-MS.

2.5 Product characterization and supplemental data

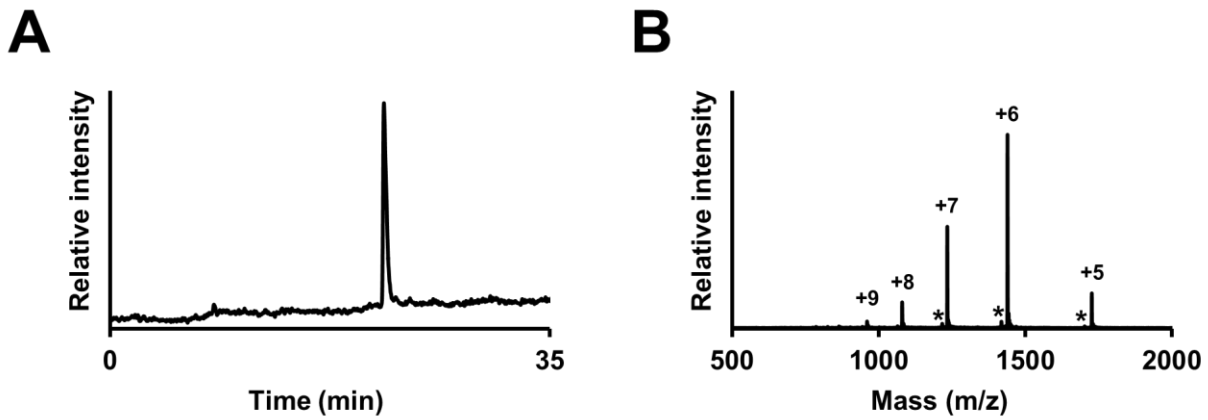


Figure 2.5.1. Purification of Ub(1-75) α-MESNa thioester (**A**) C18 Analytical RP-HPLC chromatogram of purified Ub(1-75) α-MESNa thioester, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified Ub(1-75)- α-MESNa thioester; observed mass 8633.3 ± 1.8 Da (cacl'd mass 8632.0 Da); asterisks (*) indicate hydrolyzed Ub(1-75); observed mass 8508.5 ± 1.9 Da (calc'd mass 8507.7 Da); all reported masses are isotopically averaged.

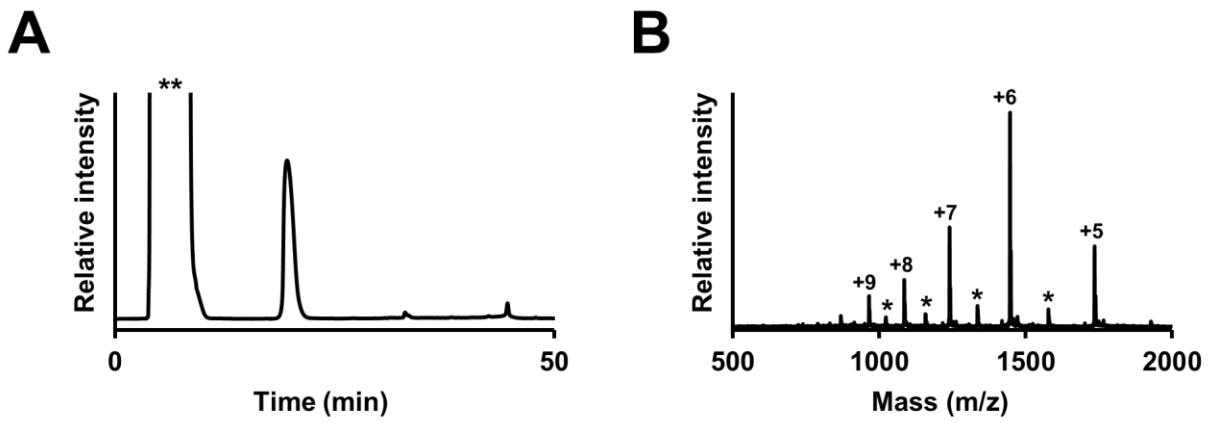


Figure 2.5.2 Purification of Ub Gly76Sec N-methylamide (**2**). (**A**) C18 Analytical RP-HPLC chromatogram of purified **2**, 30 min. 0-73% B gradient; (**) injection peak. (**B**) ESI-MS of purified **2**; observed mass 8673.7 ± 2.1 Da (cacl'd mass 8673.7 Da); asterisks (*) indicate dimer; observed mass 17345.7 ± 3.6 Da (calc'd mass 17341.5 Da); all reported masses are isotopically averaged.

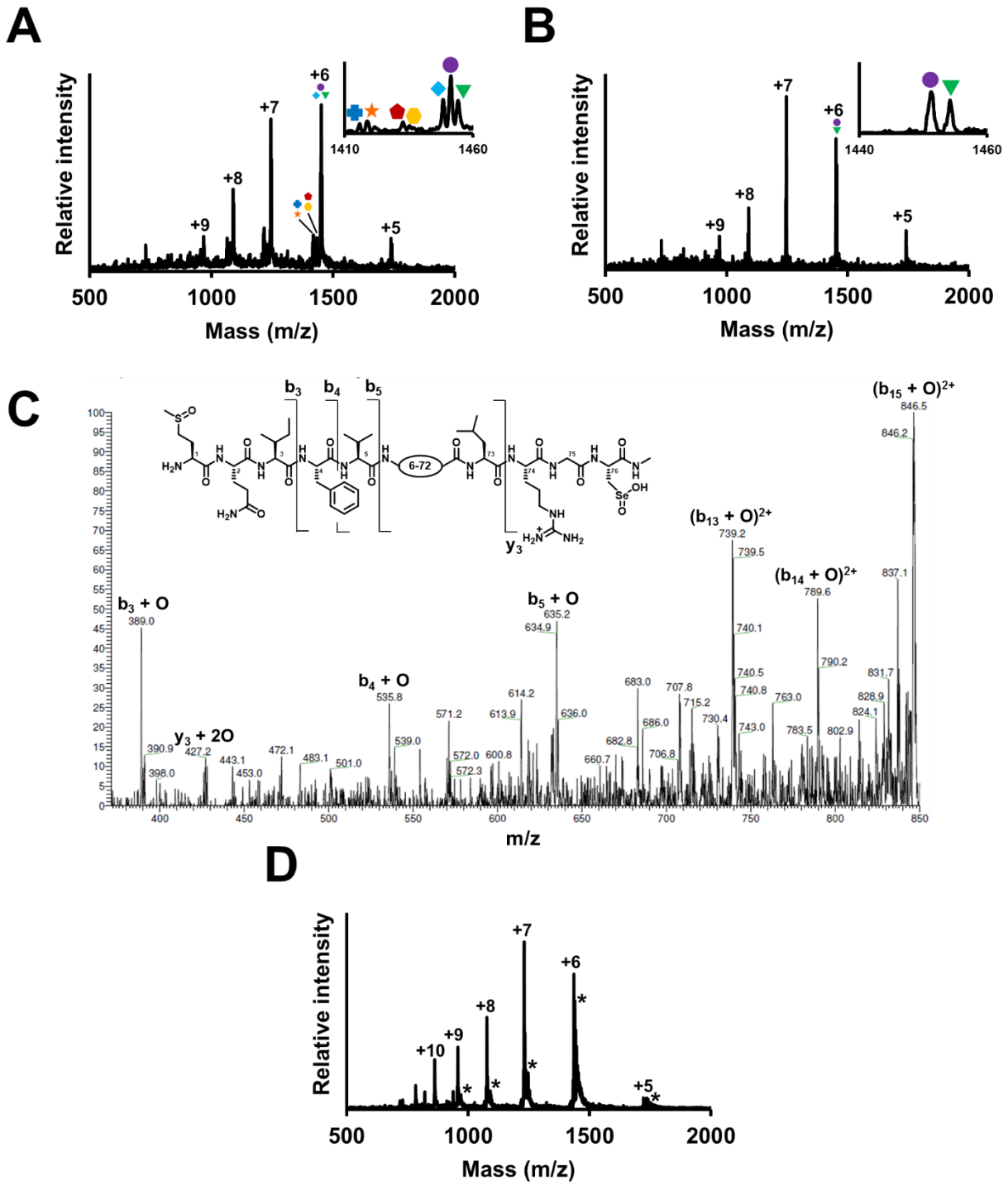


Figure 2.5.3 Products resulting from oxidation of Ub-SeCys-N-methylamide. (a) ESI-MS of oxidation products: Ub-SeCys-N-methylamide [O_3] (green triangle); observed mass 8720.9 ± 1.5 Da (calc'd 8719.8 Da); Ub-SeCys-N-methylamide [O_2] (purple circle); observed mass 8703.3 ± 1.5 Da (calc'd 8703.8 Da); Ub-SeCys-N-methylamide [O_1] (light blue diamond); observed mass 8685.7 ± 2.0 Da (calc'd 8687.8 Da); Ub-Dha-N-methylamide [O_1] (yellow hexagon); observed mass 8607.0 ± 1.9 Da (calc'd 8606.8 Da); Ub-Dha-N-methylamide (red pentagon); observed mass 8591.0 ± 1.6 Da (calc'd 8590.8 Da); hydrolyzed Ub(1-75) (blue cross); hydrolyzed Ub(1-75) [O_1] (orange star); observed mass 8525.4 ± 2.2 Da (calc'd 8523.7 Da); observed mass 8508.8 ± 1.5 Da (calc'd 8507.8 Da); all reported masses are isotopically averaged. (b) ESI-MS of oxidation products: Ub-SeCys-N-methylamide [O_3] (green triangle); observed 8719.1 mass ± 6.1 Da (calc'd 8719.8 Da); Ub-SeCys-N-methylamide [O_2] (purple circle); observed mass 8703.4 ± 2.9 Da (calc'd 8703.8 Da); all reported masses are isotopically averaged. (c) ESI-MS/MS of Ub-SeCys-N-methylamide [O_3] +7 charge state (1246.0 Da); all reported masses are monoisotopic. (d) Ub-Dha-N-methylamide [O_1] (yellow hexagon); observed mass 8609.6 ± 2.7 Da (calc'd Da); asterisks (*) indicates sodium adducts; all reported masses are isotopically averaged.

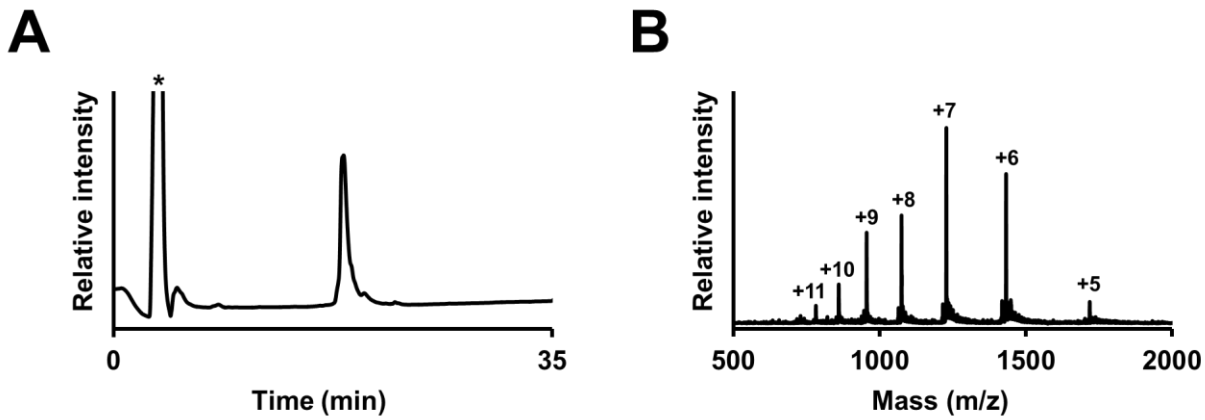


Figure 2.5.4 Purification of Ub Gly76Dha N-methylamide (**2**). (**A**) C18 Analytical RP-HPLC chromatogram of purified **3**, 30 min. 0-73% B gradient; (*) injection peak. (**B**) ESI-MS of purified **3**; observed mass 8593.1 ± 2.2 Da (cacl'd mass 8590.77 Da); all reported masses are isotopically average.

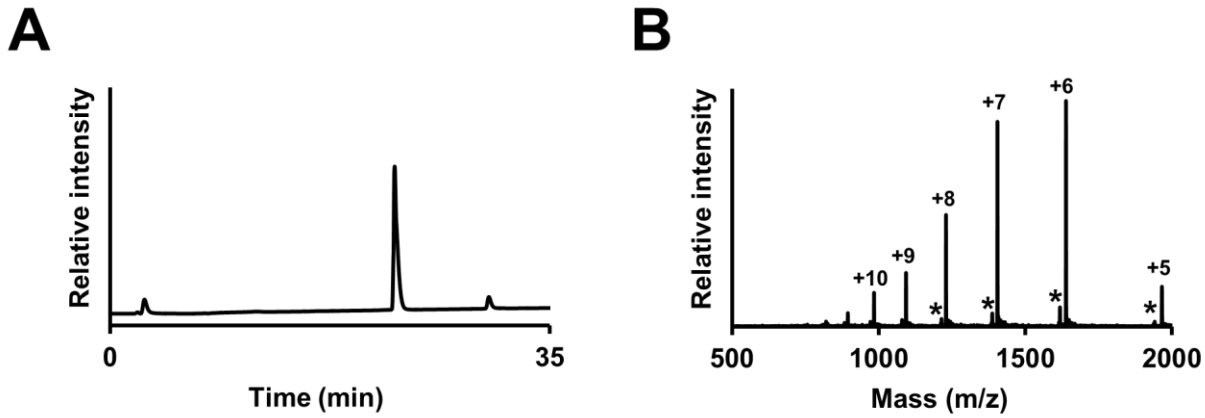


Figure 2.5.5 Purification of FLAG-Ub(1-75) α-MESNa thioester. (**A**) C18 Analytical RP-HPLC chromatogram of purified FLAG-Ub(1-75)-α-MESNa thioester, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified FLAG-Ub(1-75)-α-MESNa thioester; observed mass 9830.9 ± 2.5 Da (calc'd mass 9830.3 Da); asterisks indicate hydrolysis product FLAG-ubiquitn(1-75); observed mass 9706.6 ± 2.0 Da (calc'd mass 9705.0 Da); all reported masses are isotopically averaged.

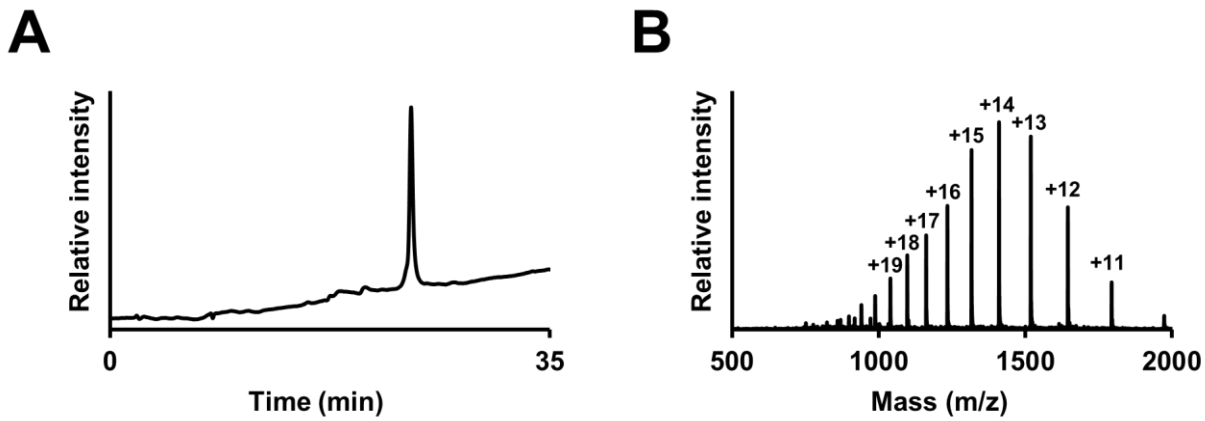


Figure 2.5.6 Purification of FLAG-ubiquitin Gly76Sec N-methylamide. (**A**) C18 Analytical RP-HPLC chromatogram of purified FLAG-ubiquitin Gly76Sec, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified FLAG-ubiquitin Gly76Sec thioester; observed mass 19738.0 ± 3.5 Da (calc'd mass 19736.1 Da); all reported masses are isotopically averaged.

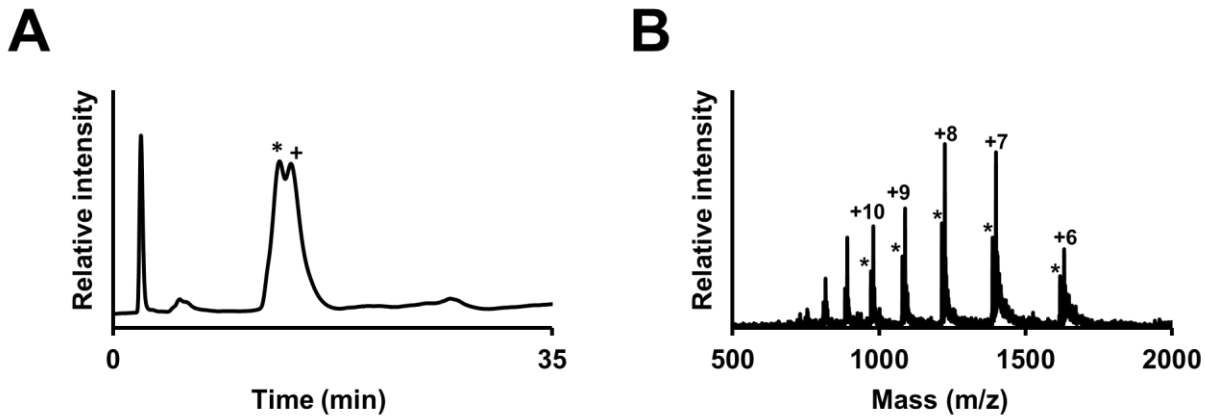


Figure 2.5.7 Purification of FLAG-ubiquitin Gly76Dha N-methylamide. (**A**) C18 Analytical RP-HPLC chromatogram of purified FLAG-ubiquitin Gly76Dha N-methylamide, 30 min. 0-73% B gradient; (*) hydrolyzed FLAG-Ub(1-75); (+) product. (**B**) ESI-MS of purified FLAG-ubiquitin Gly76Dha N-methylamide; observed mass 9789.9 ± 1.9 Da (calc'd mass 9788.1 Da); (*)hydrolyzed FLAG-Ub(1-75); observed mass 9708.0 ± 2.8 Da (calc'd mass 9705.0 Da); all reported masses are isotopically averaged.

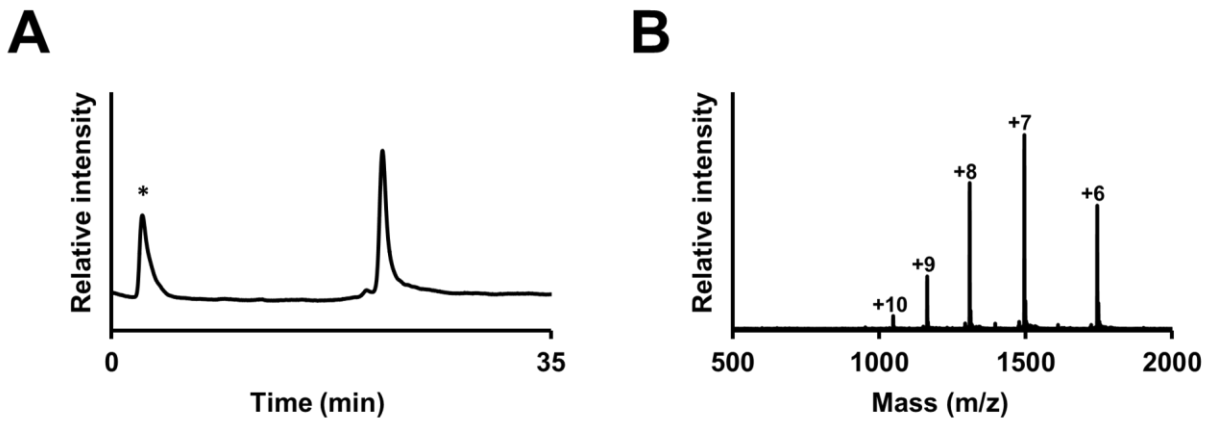


Figure 2.5.8 Purification of SUMO3(2-92)-MESNa α-thioester. (**A**) C18 Analytical RP-HPLC chromatogram of purified SUMO3(2-92)-MESNa α-thioester, 30 min. 0-73% B gradient; (*) injection peak. (**B**) ESI-MS of purified SUMO3(2-92)-MESNa α-thioester; observed mass 10462.5 ± 2.0 Da (calc'd mass 10461.7 Da); all reported masses are isotopically averaged.

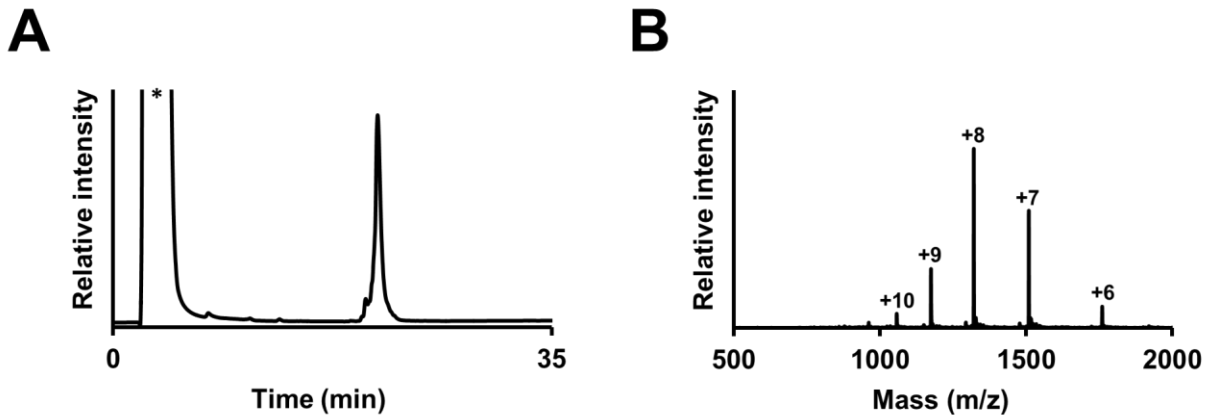


Figure 2.5.9 Purification of SUMO3 Gly93Sec(acetamide) N-methylamide. (**A**) C18 Analytical RP-HPLC chromatogram of purified SUMO3 Gly93Sec(acetamide) N-methylamide, 30 min. 0-73% B gradient; (*) injection peak. (**B**) ESI-MS of purified SUMO3 Gly93Sec(acetamide) N-methylamide; observed mass 10559.0 ± 2.3 Da (calc'd mass 10557.6 Da); all reported masses are isotopically averaged.

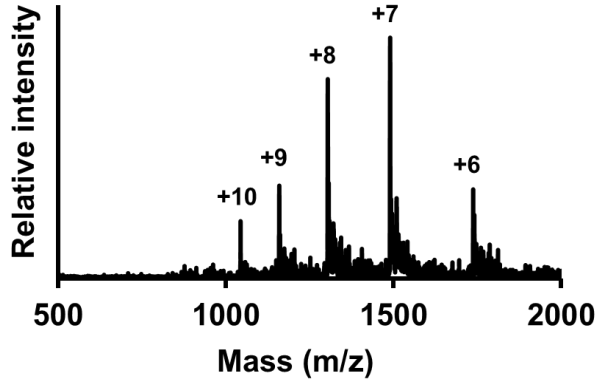


Figure 2.5.10 ESI-MS of SUMO3 Gly93Dha N-methylamide produced during oxidation of SUMO3 Gly93Sec(acetamide) N-methylamide. Observed mass 10423.6 ± 2.8 Da (calc'd mass 10419.6 Da); all reported masses are isotopically averaged.

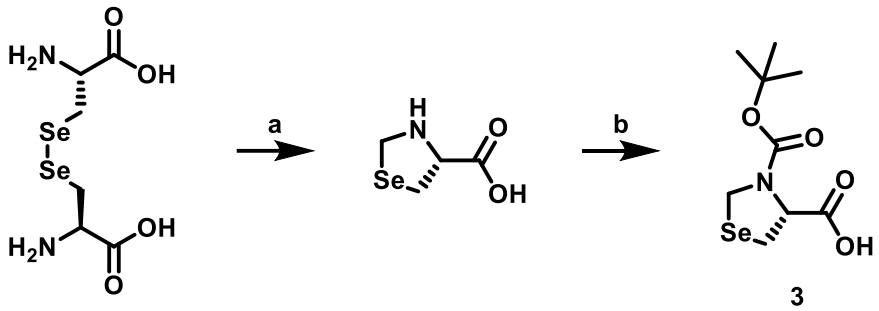


Figure 2.5.11. Synthesis of (R)-3-(tert-butoxycarbonyl)-1,3-selenazolidine-4-carboxylic acid **6**: (a) 1) $NaBH_4$, NaOH, H_2O/EtOH, RT; 2) AcOH, $CH_2(O)$, on ice, 87%. (b) Boc_2O, DIEA, H_2O/MeOH, on ice, 39%.[30]

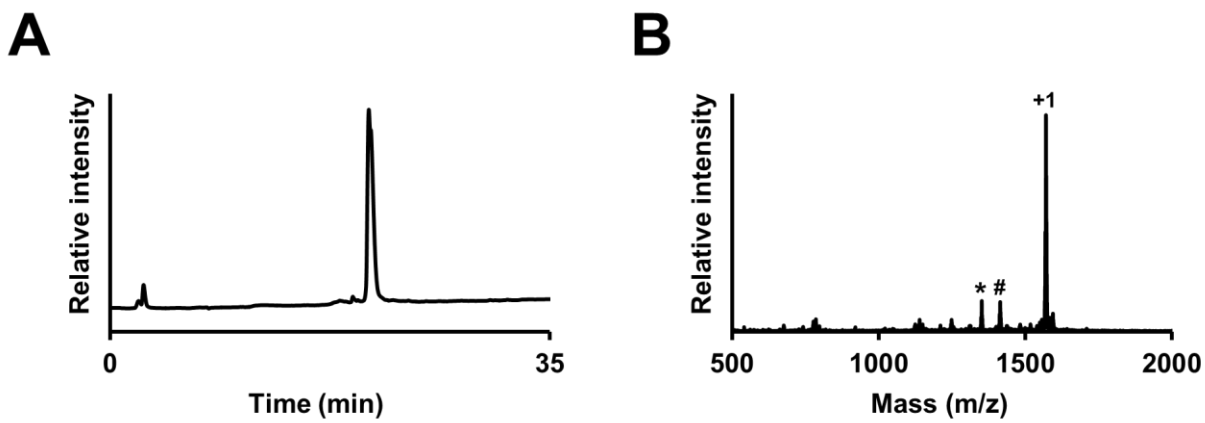


Figure 2.5.12. Purification of TRIM25(112-124) K117(Selenaproline). (a) C18 Analytical RP-HPLC chromatogram of purified ubiquitin TRIM25(112-124) K117(Selenaproline), 30 min. 0-73% B gradient; (b) ESI-MS of purified TRIM25(112-124) K117(Selenaproline); observed mass 1569.8 Da (calc'd mass 1569.8 Da); b12 ion (asterisk); observed mass 1350.8 Da (calc'd mass 1351.5 Da); Singly oxidized y13 ion (hashtag); observed mass 1413.7 Da (calc'd mass 1414.6 Da); all reported masses are isotopically averaged.

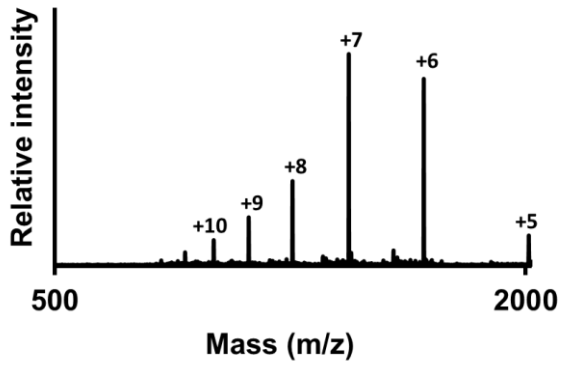


Figure 2.5.13. ESI-MS of purified TRIM25(112-124) K117(UbSec); observed mass 10047.8 ± 2.4 Da (calc'd mass 10047.5 Da); all reported masses are isotopically averaged.

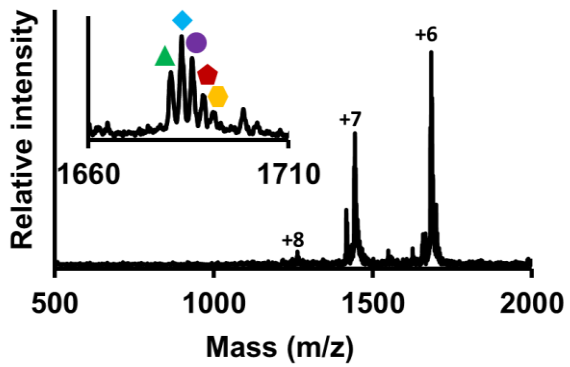


Figure 2.5.14. ESI-MS of TRIM25(112-124) K117(UbSec) overoxidation. TRIM25(112-124) K117(UbSec) [O_2] (green triangle); observed mass 10079.4 ± 1.6 Da (calc'd 10078.6 Da); TRIM25(112-124) K117(UbSec) [O_3] (blue diamond); observed mass 10094.5 ± 2.1 Da (calc'd 10094.6 Da); TRIM25(112-124) K117(UbSec) [O_4] (purple circle); observed mass 10110.6 ± 1.3 Da (calc'd 10110.2 Da); TRIM25(112-124) K117(UbSec) [O_5] (red pentagon); observed mass 10126.8 Da ± 0.8 (calc'd 10126.6 Da); TRIM25(112-124) K117(UbSec) [O_6] (yellow hexagon); observed mass 10142.1 ± 1.2 Da (calc'd 10142.6 Da); all reported masses are isotopically averaged; all reported masses are isotopically averaged.

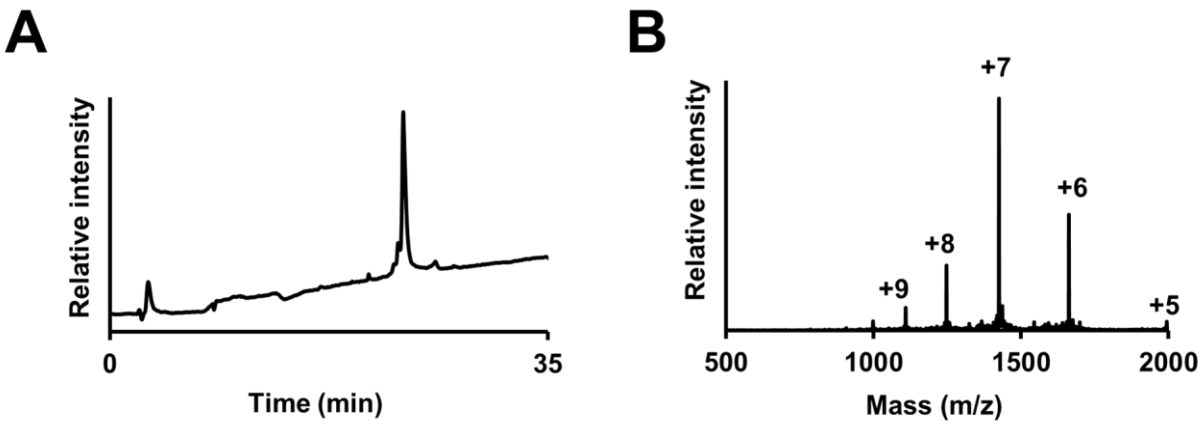
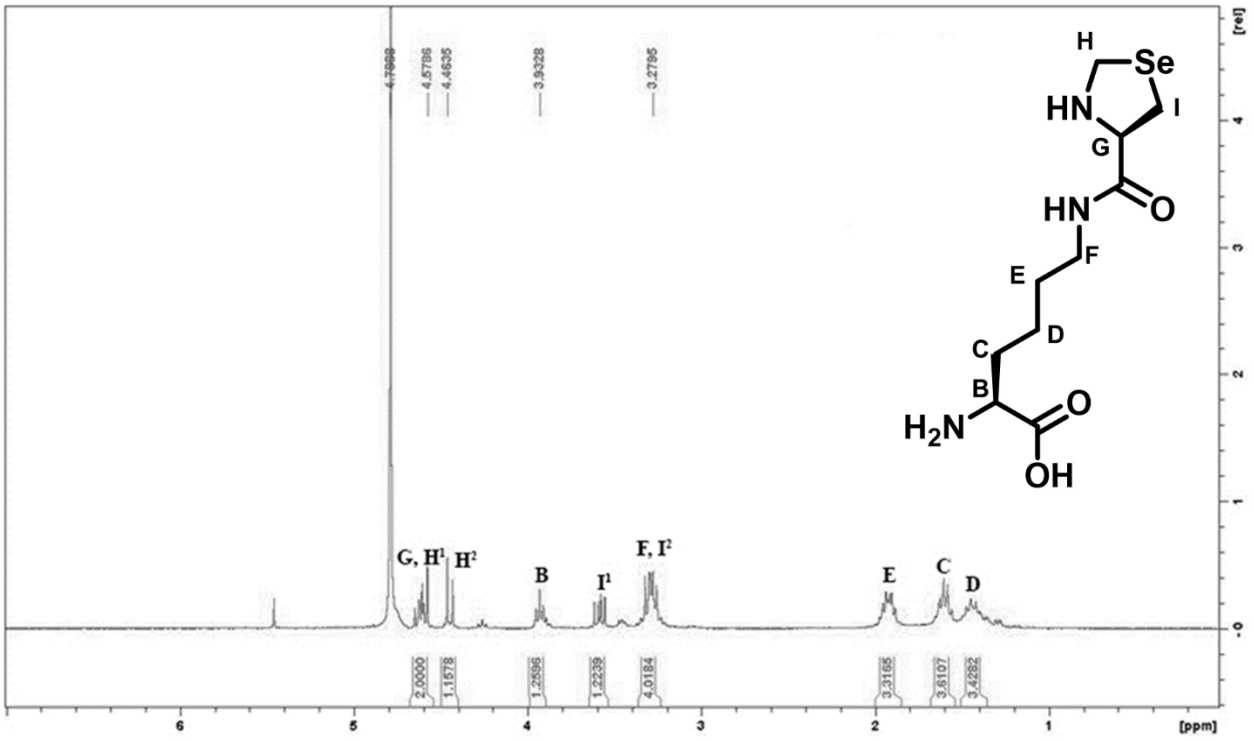
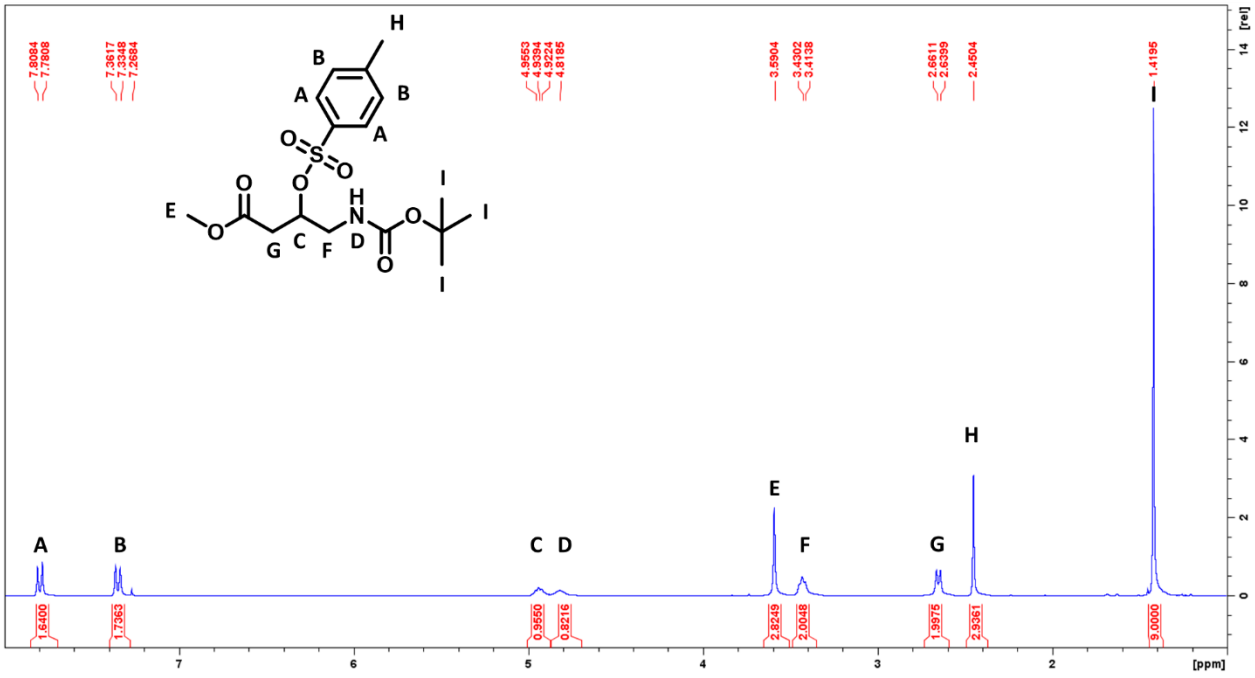


Figure 2.5.15. Purification of TRIM25(112-124) K117(UbDha). (a) C18 Analytical RP-HPLC chromatogram of purified ubiquitin TRIM25(112-124) K117(UbDha), 30 min. 0-73% B gradient; (b) ESI-MS of purified TRIM25(112-124) K117(UbDha); observed mass 9968.0 ± 1.4 Da (calc'd mass 9968.5 Da); all reported masses are isotopically averaged; all reported masses are isotopically averaged.

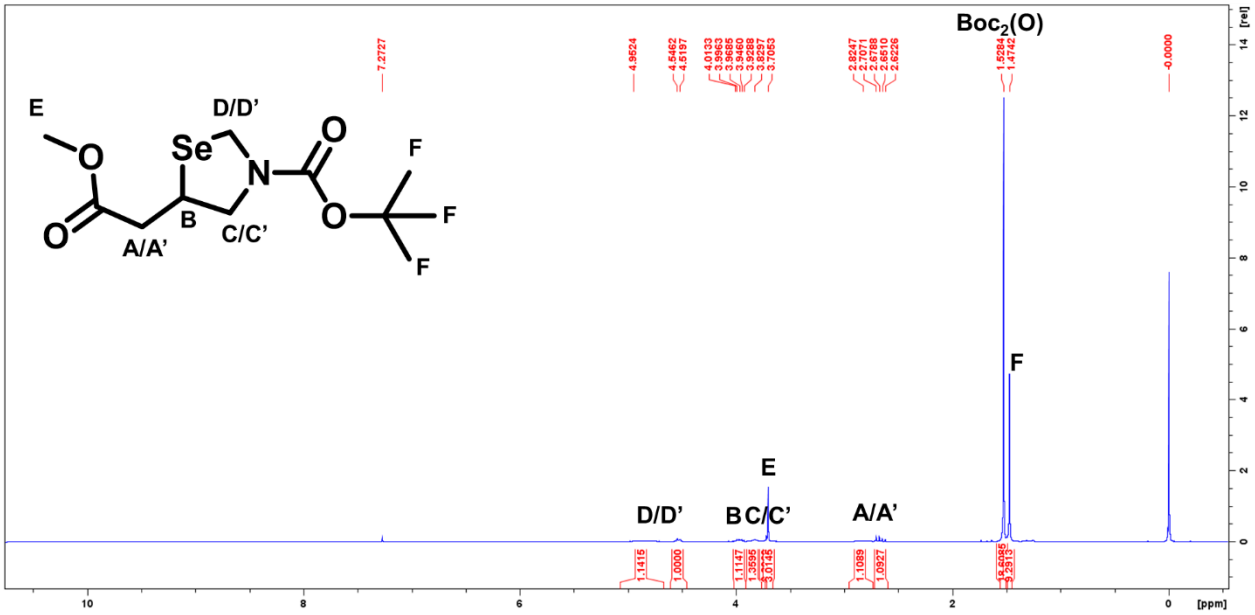


2.5.16. ^{1}H NMR of 8

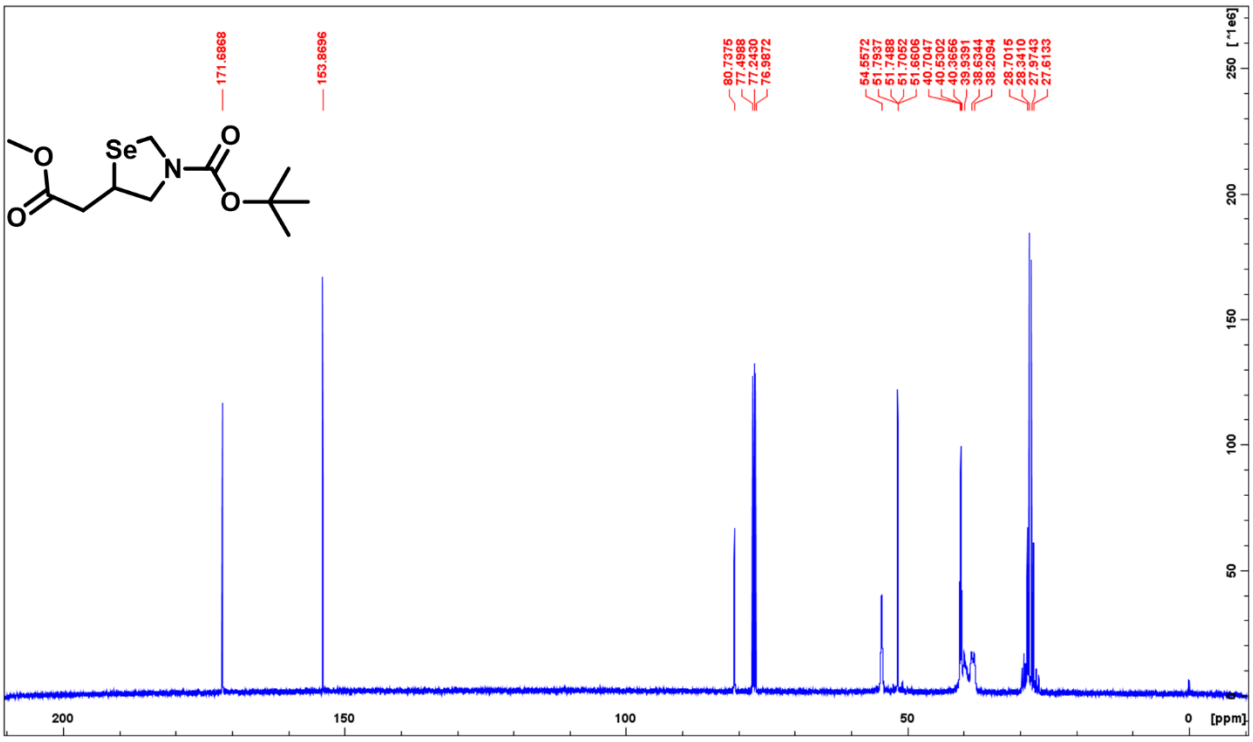
2.5.17. ^{1}H NMR of 9a.



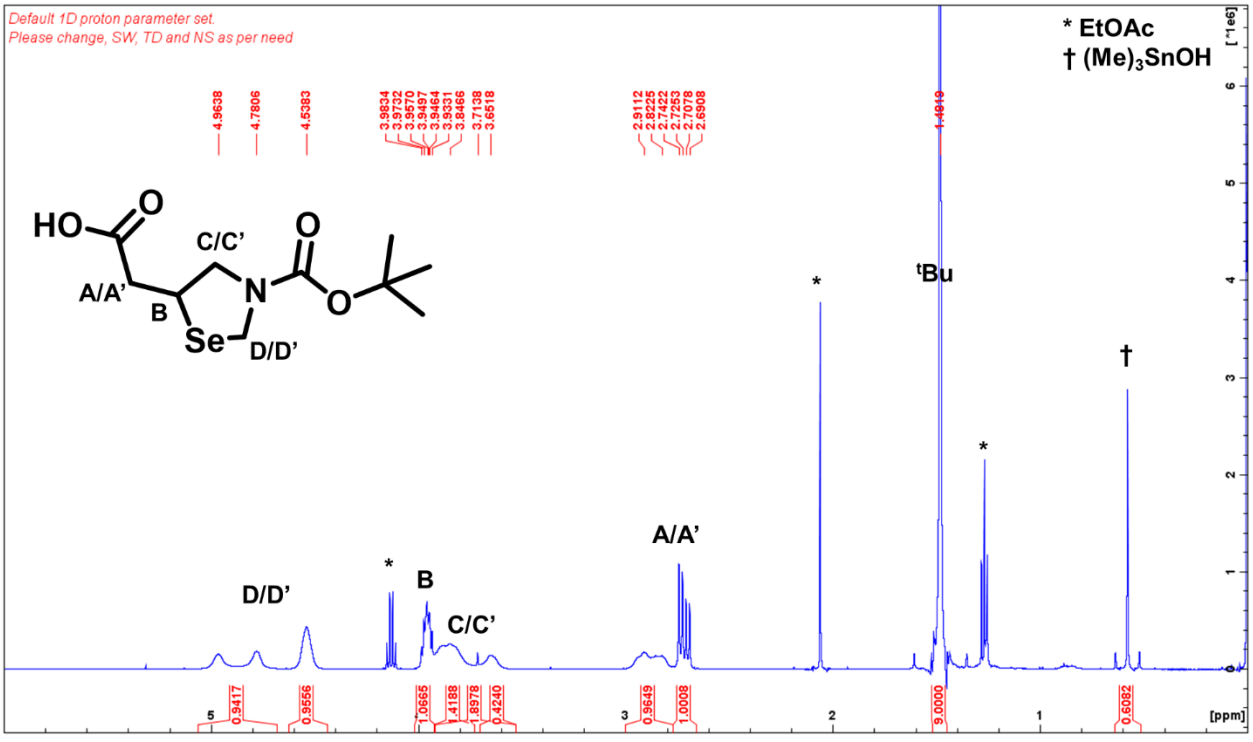
2.5.18. ^{1}H NMR of 9b.



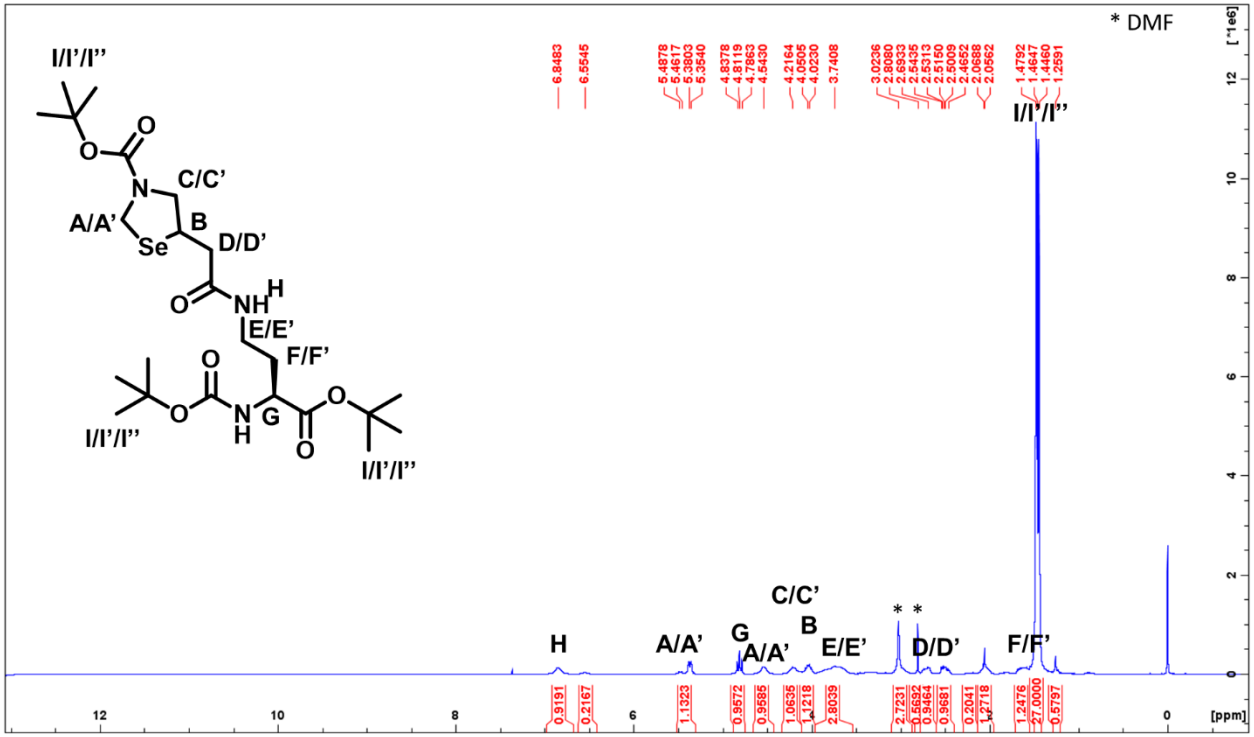
2.5.19. ^{1}H NMR of 9c.



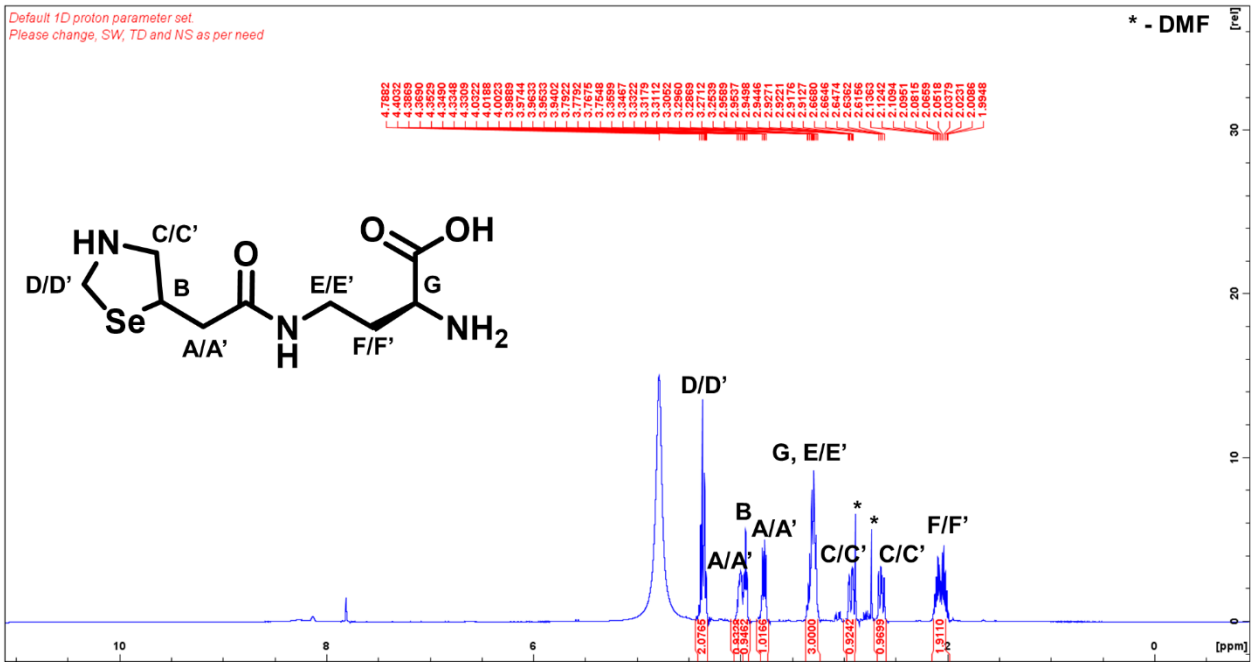
2.5.20. ^{13}C NMR of 9c.



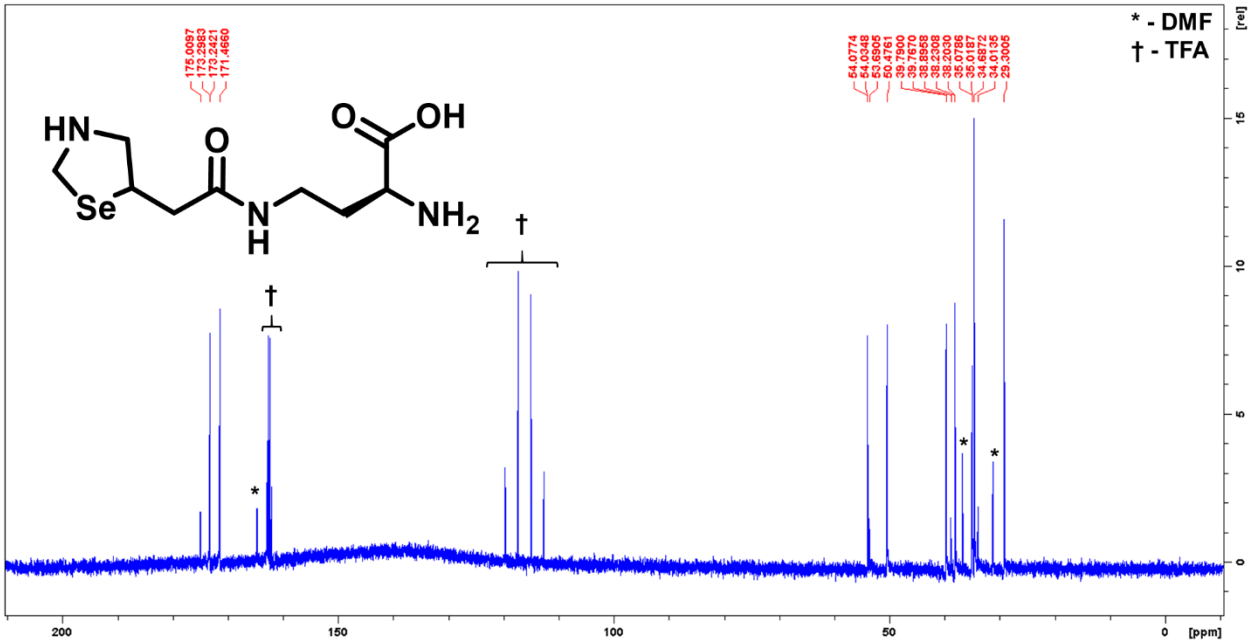
2.5.2. ^{1}H NMR of 9d.



2.5.22. ^{1}H NMR of 9e.



2.5.23. ^{1}H NMR of 9.



2.5.24. ^{13}C NMR of 9.

2.6 References

Portions of this chapter have been published as:

Whedon, S. D. *et al.* Selenocysteine as a Latent Bioorthogonal Electrophilic Probe for Deubiquitylating Enzymes. *J. Am. Chem. Soc.* **138**, 13774-7 (2016).

1. Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9,** 1149–59 (2002).

2. Hemelaar, J. *et al.* Specific and Covalent Targeting of Conjugating and Deconjugating Enzymes of Ubiquitin-Like Proteins. *Mol. Cell. Biol.* **24,** 84–95 (2004).

3. Ekkebus, R. *et al.* On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. *J. Am. Chem. Soc.* **135,** 2867–70 (2013).

4. Wilkinson, K. D., Cox, M. J., Mayer, A. N. & Frey, T. Synthesis and Characterization of Ubiquitin Ethyl Ester , a New Substrate for Ubiquitin Carboxyl-Terminal Hydrolase. *Biochemistry* **25,** 6644–6649 (1986).

5. Shrestha, R. K. *et al.* Insights into the Mechanism of Deubiquitination by JAMM Deubiquitinases from Cocrystal Strucures of the Enzyme with the Substrate and Product. *Biochemistry* **53,** 3199–3217 (2014).

6. Zhu, Y. & van der Donk, W. a. Convergent Synthesis of Peptide Conjugates Using Dehydroalanines for Chemoselective Ligations. *Org. Lett.* **3,** 1189–1192 (2001).

7. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2,** 1666–1676 (2011).

8. Okeley, N. M., Zhu, Y. & van Der Donk, W. a. Facile chemoselective synthesis of dehydroalanine-containing peptides. *Org. Lett.* **2,** 3603–6 (2000).

9. Zaro, B. W., Whitby, L. R., Lum, K. M. & Cravatt, B. F. Metabolically Labile Fumarate Esters Impart Kinetic Selectivity to Irreversible Inhibitors. *J. Am. Chem. Soc.* **138,** 2–5 (2016).

10. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **194,** 531–44 (1987).

11. Malins, L. R., Mitchell, N. J. & Payne, R. J. Peptide ligation chemistry at selenol amino acids. *J. Pept. Sci.* **20,** 64–77 (2013).

12. Besse, D., Siedler, F., Diercks, T., Kessler, H. & Moroder, L. The Redox Potential of Selenocystine in Unconstrained Cyclic Peptides. *Angew. Chemie Int. Ed. English* **36,** 883–885 (1997).

13. Hondal, R. J., Nilsson, B. L. & Raines, R. T. Selenocysteine in Native Chemical Ligation and Expressed Protein Ligation. *J. Am. Chem. Soc.* **123,** 5140–5141 (2001).

14. Metanis, N., Beld, J. & Hilvert, D. *The chemistry of selenocysteine*. *Organic Selenium and Tellurium Compounds* (John Wiley & Sons, Ltd., 2012). doi:10.1002/9780470682531.pat0582

15. Ma, S., Caprioli, R. M., Hill, K. E. & Burk, R. F. Loss of selenium from selenoproteins: conversion of selenocysteine to dehydroalanine in vitro. *J. Am. Soc. Mass Spectrom.* **14,** 593–600 (2003).

16. Kloc, K., Mlochowski, J. & Syper, L. Oxidation of Organic Diselenides with Hydrogen Peroxide to Alkane and Areneseleninic Acids and Selenium-Containing Heterocycles. *Liebigs Ann. der Chemie* 811–813 (1989).

17. Abdo, M. & Knapp, S. Mechanism of a Redox Coupling of Seleninic Acid with Thiol. *J. Org. Chem.* **77,** 3433–3438 (2012).

18. Love, K. R., Pandya, R. K., Spooner, E. & Ploegh, H. L. Ubiquitin C-Terminal Electrophiles Are Activity-Based Probes for Identification and Mechanistic Study of Ubiquitin Conjugating Machinery. *ACS Chem. Biol.* **4,** 275–287 (2009).

19. Mulder, M. P. C. *et al.* A cascading activity-based probe sequentially targets E1–E2–E3 ubiquitin enzymes. *Nat. Chem. Biol.* **12,** 523–530 (2016).

20. Liu, J. & Rozovsky, S. Contribution of Selenocysteine to the Peroxidase Activity of Selenoprotein S. *Biochemistry* **52,** 5514–5516 (2013).

21. Mobli, M., Morgenstern, D., King, G. F., Alewood, P. F. & Muttenthaler, M. Site-Specific pKa Determination of Selenocysteine Residues in Selenovasopressin by Using 77 Se NMR Spectroscopy **. *Angew. Chemie Int. Ed. English* **50,** 11952–11955 (2011).

22. Liu, J., Chen, Q. & Rozovsky, S. Utilizing Selenocysteine for Expressed Protein Ligation and Bioconjugations. *J. Am. Chem. Soc.* **139,** 3430–3437 (2017).

23. Seebeck, F. P. & Szostak, J. W. Artificial lantipeptides from in vitro translations w. *Chem. Commun.* **47,** 6141–6143 (2011).

24. Valkevich, E. M. *et al.* Forging isopeptide bonds using thiol-ene chemistry: site-specific coupling of ubiquitin molecules for studying the activity of isopeptidases. *J. Am. Chem. Soc.* **134,** 6916–9 (2012).

25. Trang, V. H. *et al.* Nonenzymatic polymerization of ubiquitin: single-step synthesis and isolation of discrete ubiquitin oligomers. *Angew. Chem. Int. Ed. Engl.* **51,** 13085–8 (2012).

26. Flierman, D. *et al.* Non-hydrolyzable Diubiquitin Probes Reveal Linkage-Specific Reactivity of Deubiquitylating Enzymes Mediated by S2 Pockets Article Non-hydrolyzable Diubiquitin Probes Reveal Linkage-Specific Reactivity of Deubiquitylating Enzymes Mediated by S2 Pockets. *Cell Chem. Biol.* **23,** 472–482 (2016).

27. Inn, K.-S. *et al.* Linear Ubiquitin Assembly Complex Negatively Regulates RIG-I- and TRIM25-Mediated Type I Interferon Induction. *Mol. Cell* **41,** 354–365 (2011).

28. Pauli, E.-K. *et al.* The ubiquitin-specific protease USP15 promotes RIG-I-mediated antiviral signaling by deubiquitylating TRIM25. *Sci. Signal.* **7,** 1–11 (2014).

29. Nguyen, D. P., Elliott, T., Holt, M., Muir, T. W. & Chin, J. W. Genetically Encoded 1,2-Aminothiols Facilitate Rapid and Site-Specific Protein Labeling via a Bio-orthogonal

Cyanobenzothiazole Condensation. *J. Am. Chem. Soc.* **133,** 11418–11421 (2011).

30. Reddy, P. S., Dery, S. & Metanis, N. Chemical Synthesis of Proteins with Non-Strategically Placed Cysteines Using Selenazolidine and Selective Deselenization. *Angew. Chemie Int. Ed.* **55,** 992–995 (2015).

31. Mukherjee, A. J., Zade, S. S., Singh, H. B. & Sunoj, R. B. Organoselenium Chemistry : Role of Intramolecular Interactions. *Chem. Rev.* **110,** 4357–4416 (2010).

32. Stokes, A. L. *et al.* Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol. Biosyst.* **5,** 1032–1038 (2009).

33. Young, D. D. *et al.* An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity †,‡. *Biochemistry* **50,** 1894–1900 (2011).

34. Cooley, R. B., Karplus, P. A. & Mehl, R. A. Gleaning Unexpected Fruits from Hard-Won Synthetases : Probing Principles of Permissivity in Non-canonical Amino Acid – tRNA Synthetases. *ChemBioChem* **15,** 1810–1819 (2014).

35. Lin, S. *et al.* Genetically Encoded Cleavable Protein Photo-Cross-Linker. *J. Am. Chem. Soc.* **136,** 11860–11863 (2014).

36. He, D. *et al.* Quantitative and Comparative Profiling of Protease Substrates through a Genetically Encoded Multifunctional Photocrosslinker. *Angew. Chemie Int. Ed.* **56,** 14521–14525 (2017).

37. Short, M. D., Xie, Y., Li, L., Cassidy, P. B. & Roberts, J. C. Characteristics of Selenazolidine Prodrugs of Selenocysteine : Toxicity and Glutathione Peroxidase Induction in V79 Cells. *J. Med. Chem.* **46,** 3308–3313 (2003).

38. Deutch, C. E., Spahija, I. & Wagner, C. E. Susceptibility of Escherichia coli to the toxic L-proline analogue L-selenaproline is dependent on two L-cystine transport systems. *J. Appl. Microbiol.* **117,** 1487–1499 (2014).

39. Mulder, M. P. C., El Oualid, F., ter Beek, J. & Ovaa, H. A native chemical ligation handle that enables the synthesis of advanced activity-based probes: Diubiquitin as a case study. *ChemBioChem* **15,** 946–949 (2014).

40. Back, T. G. & Moussa, Z. Diselenides and Allyl Selenides as Glutathione Peroxidase Mimetics . Remarkable Activity of Cyclic Seleninates Produced in Situ by the Oxidation of Allyl ω -Hydroxyalkyl Selenides. *J. Am. Chem. Soc.* **125,** 13455–13460 (2003).

41. Scianowski, J., Rafinski, Z., Szuniewicz, A. & Wojtczak, A. New chiral selenium electrophiles derived from functionalized terpenes. *Tetrahedron* **65,** 10162–10174 (2009).

42. Morgan, M. T. *et al.* Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science (80-.).* **351,** 1–5 (2016).

43. Arif, S. *et al.* MINDY-1 Is a Member of an Evolutionarily Conserved and Structurally Distinct New Family of Deubiquitinating Enzymes Short Article MINDY-1 Is a Member of an Evolutionarily Conserved and Structurally Distinct New Family of Deubiquitinating

Enzymes. *Mol. Cell* **63,** 146–155 (2016).

44. Mevissen, T. E. T. *et al.* OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. *Cell* **154,** 169–84 (2013).

45. Weller, C. E., Huang, W. & Chatterjee, C. Facile synthesis of native and protease-resistant ubiquitylated peptides. *ChemBioChem* **15,** 1263–1267 (2014).

46. Dhall, A. *et al.* Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long-range Internucleosomal Interactions. *J. Biol. Chem.* **289,** 33827–33837 (2014).

47. Van der Veen, A. G. *et al.* Role of the ubiquitin-like protein Urm1 as a noncanonical lysine-directed protein modifier. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 1763–1770 (2011).

48. Hammill, J. T., Miyake-stoner, S., Hazen, J. L., Jackson, J. C. & Mehl, R. A. Preparation of site-specifically labeled fluorinated proteins for 19 F-NMR structural characterization. *Nat. Protoc.* **2,** 2601–2607 (2007).

49. Peeler, J. C. & Mehl, R. A. Chapter 8 Site-Specific Incorporation of Unnatural Amino Acids as Probes for Protein Conformational Changes. *Methods Mol. Biol.* **794,** 125–134 (2012).

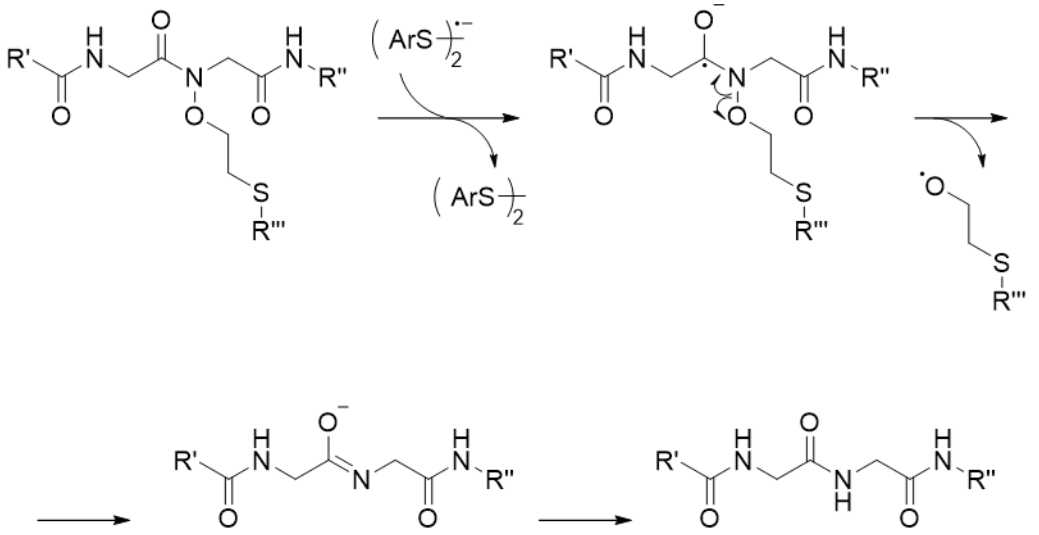
Chapter 3: Traceless affinity purification handles for the purification of semi-synthetic proteins

3.1 Introduction

Characterization of protein post-translational modifications' (PTMs) contributions to protein-protein interaction, allostery, and enzyme activity relies on access to homogenously modified protein.[1] Native chemical ligation (NCL) facilitates the production of chemically defined protein by overcoming the 50-100 amino acid limitation of solid phase peptide synthesis (SPPS).[2,3] Ease of synthesis increases where terminal modifications are studied, as either the N-terminal Cys or α-thioester component can be generated recombinantly for use in expressed protein ligation (EPL). Preservation of protein structure is a strength of EPL, while products of NCL are frequently isolated by denaturing reverse-phase high performance liquid chromatography (RP-HPLC). In ligation reactions of either type the separation strategy employed is a major contributor to overall yield, and as a consequence numerous techniques have been developed.[4]

Non-denaturing separation strategies for isolation of recombinant proteins include immobilized metal affinity chromatography (IMAC), immunoprecipitation, size exclusion, and ion exchange. Any one of these strategies may be applied to purification of EPL reactions by either genetic incorporation of the peptidic tag at either terminus of the recombinant fragment, or the synthetic fragment. The addition of a peptide or small molecule tag (e.g. biotin, chlorohexyl, O^6-benzylguanine) can interfere with structural and functional characterization of the semisynthetic protein, necessitating tag removal. Tag removal can be enzymatic or chemical, though traceless removal is most attainable by chemical cleavage given the diversity of functional groups that can be introduced on the synthetic fragment.

Photodeprotection of *o*-nitrobenzyl carbamates is among the mildest chemical means of tag removal, though photo-oxidation may also occur.[5] Reducing conditions are well tolerated by proteins without structurally important disulfide bonds, and enable cleavage of disulfide and aryl hydrazide-linked tags.[6] Weller et al recently reported a non-denaturing reduction of aminooxy amide N-O bonds by aromatic thiols, which proceeds at near neutral pH under aerobic conditons.[7] Single electron reduction of the carbonyl followed by N-O homolysis to the alkoxyl radical and imidate is the proposed mechanism, and is supported by the calculated LUMO of a model dipeptide (**Scheme 3.1**). We reasoned that this single electron reduction could be generalized to a broader array of O-alkyl functionalized aminooxy amides, enabling traceless removal of affinity handles under mild conditions.



Scheme 3.1. Proposed mechanism of N-O bond cleavage by disulfide radical anion. Adapted from Weller, et al.[8]

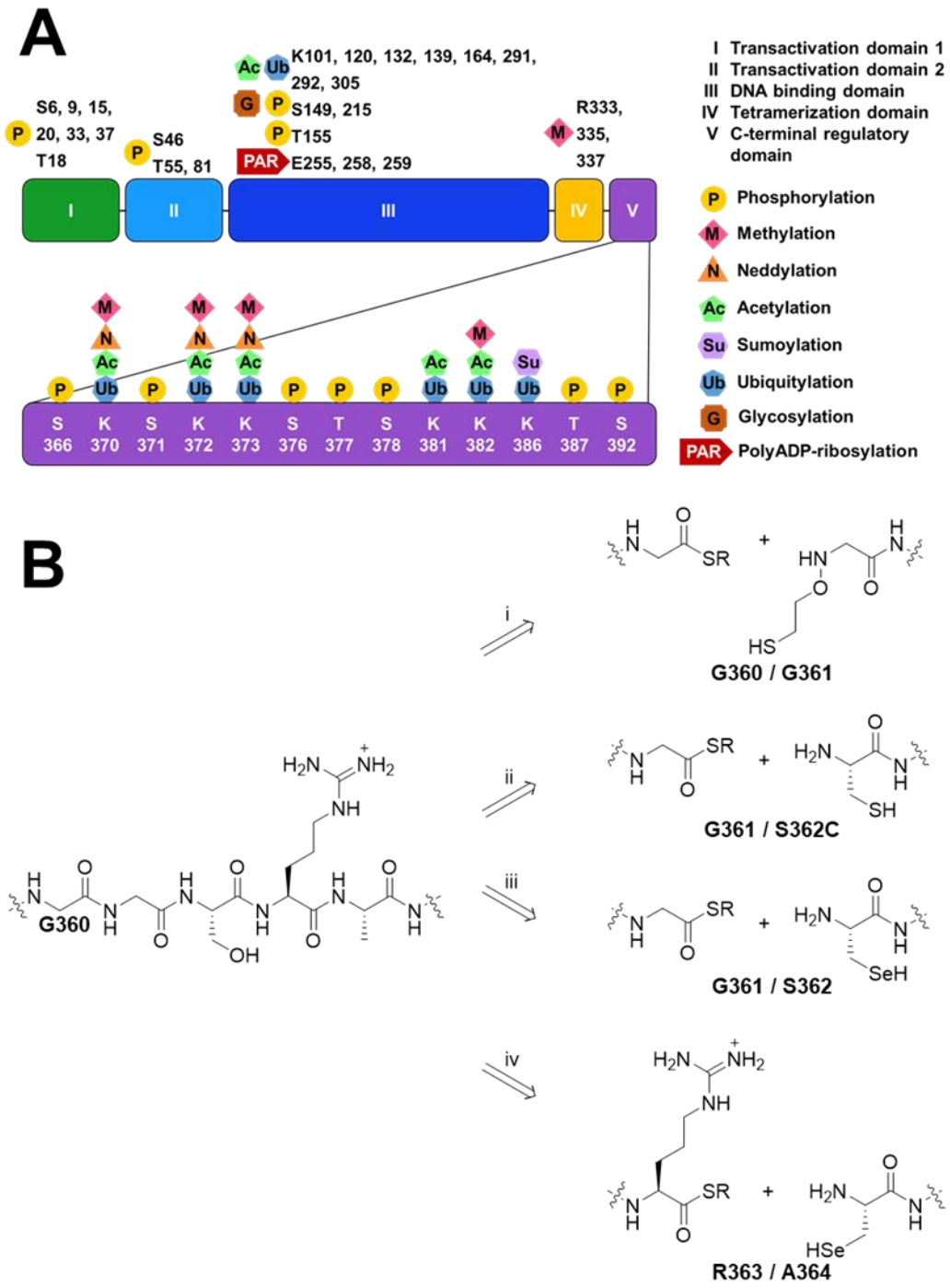


Figure 3.1. Routes to homogenously modified p53. (A) Post-translational modifications of p53, with emphasis on CTD PTMs accessible by semisynthesis.[9] (B) Ligations applicable to semi-synthesis of C-terminally modified p53: (i) auxiliary-mediated ligation;[10] (ii) S362C mutant NCL; (iii) Sec ligation followed by oxidative conversion to Ser;[11] (iv) Sec ligation followed by selective deselenization.[12]

The intended application of a traceless affinity handle is in isolation of semi-synthetic p53. The transcription factor p53 is a nexus in the cellular stress response, integrating stress-specific threats to genome stability into a homeostatic response. Stress-specific information can be relayed through altered protein-protein interactions, location and residence time. In stressed and unstressed cells post-translational modifications (PTMs) encode some information about the environmental conditions on p53. The C-terminal regulatory domain (CTD) of p53 harbors 13 modifiable residues in a 30 amino acid sequence, which endows this sequence with tremendous combinatorial diversity, redundancy, or both (**Figure 3.1**). Furthermore, the CTD is unstructured and thus amenable to EPL, which we sought to employ for the introduction of single PTMs and subsequent interrogation of their functional consequences.

3.2 Results and discussion

3.2.1 p53 ligation proof of principle

Native, modified p53 is the end goal of a semi-synthetic strategy, however, viability of semi-synthetic p53 needed to be established first. The introduction of an S376C mutation affords access to modification of three C-terminal Lys residues subject to modification with Ubls, methylation and acetylation. Furthermore, a p53(1-375)-intein fusion will thiolyze to a stable C-terminal Gln α-thioester. The peptide was produced by standard Fmoc SPPS (prepared by Patrick Shelton, Chatterjee Lab), and the thioester was expressed as an MxeGyrA-CBD fusion (prepared by Nicholas Senger, Chatterjee lab). Ligation of peptide and thioester was followed by gel, and observed to reach an endpoint within hours (**Figure 3.2.1a**). Attempts at purification by RP-HPLC could not resolve the ligation product and hydrolyzed thioester, though it did allow for validation of the ligation product by ESI-MS (**Figure 3.2.1b**).

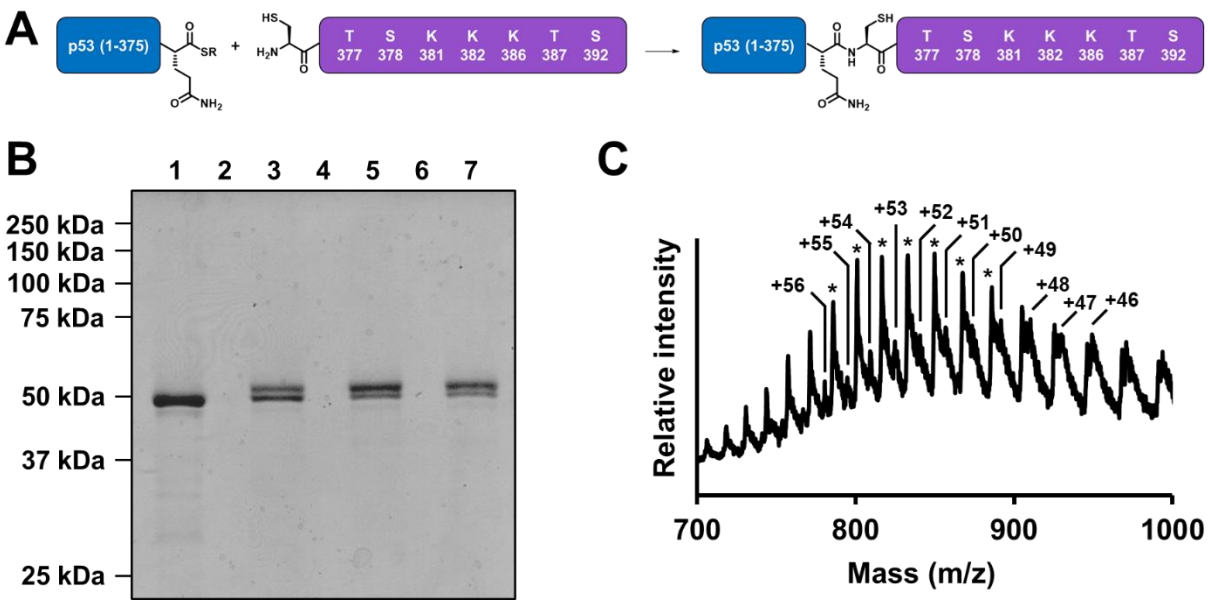
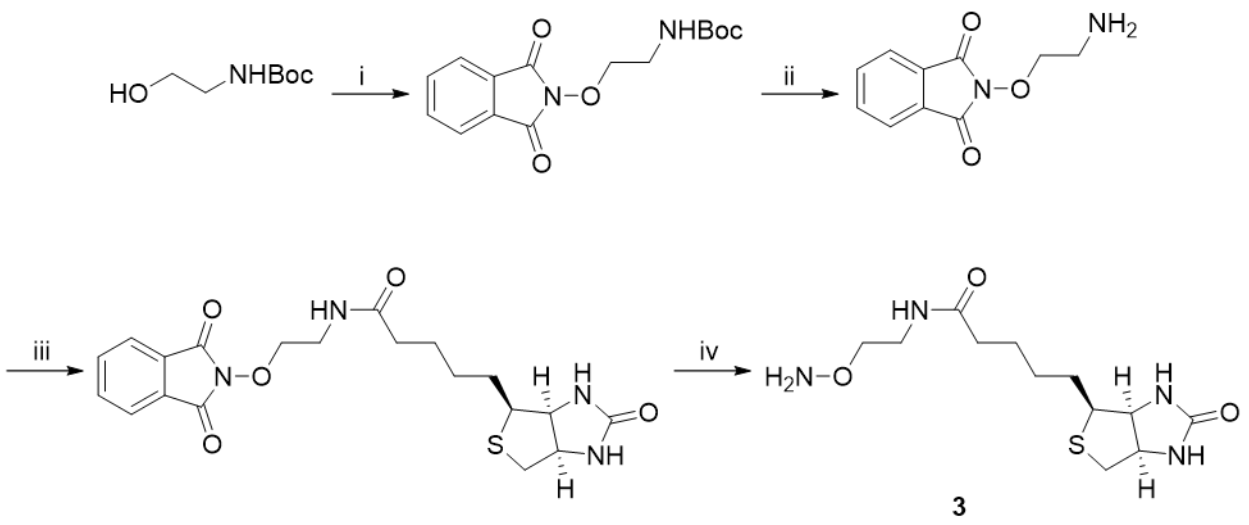


Figure 3.2. Semi-synthesis of p53. (**A**) Schematic illustration of p53 semi-synthesis from p53(1-375) α-thioester and p53(376-393)S376C peptide with accessible C-terminal modification sites highlighted. (**B**) Ligation visualization by 10% SDS-PAGE: (1) 0 hours; (2) empty; (3) 15 min; (4) empty; (5) 24 h; (6) empty; (7) 48 h. (**C**) ESI-MS of co-eluting proteins: (labeled) ligation product found 43647.5 Da ± 15.1 Da (calc'd 43669.2 Da); (*) hydrolysis product found 41606.2 ± 7.2 Da (calc'd 41606.8 Da).

3.2.2 Synthesis and peptide incorporation of N-(2-aminooxyethyl)-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide

The ligation auxiliary employed by Weller et al was incorporated at Gly residues because the Ubls under study terminate in GlyGly, and for synthetic ease. Auxiliary incorporation by Sn2 on bromoacetylated peptides does not yield peptide diastereomers, which advantage we intended to employ in our affinity handle synthesis. The affinity moiety selected was biotin, as numerous synthetic biotinylation reagents are commercially available, and enrichment on an avidin species is generally robust. In order to minimize synthetic steps and preserve similarity to the reported auxiliary, we determined to prepare N-(2-aminooxyethyl)-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide (**3**) (**Scheme 3.2**).



Scheme 3.2. Synthesis of 3. (i) N-hydroxyphthalimide, DIAD, $P(Ph)_3$, THF, 78%; (ii) 4N HCl, dioxane, quant.; (iii) Biotin, DIC, Oxyma, DMF, 30%; (iv) 1:1 DCM:MeOH, 5 eq 50% hydrazine in H_2O, 48%.

Synthesis of **3** afforded 11% overall yield. The amidation of biotin proved difficult with benzotriazole-based coupling reagents, as HOBt was observed to partially deprotect the phthalimide and form a stable ester. Hydrazine deprotection of the phthalimide was facile, however the hydrazine phthalimide byproduct and **3** were poorly resolved by RP-HPLC; a solution to this issue is described in the preparation of **5**.

A peptide model substrate derived from the terminus of eGFP was selected for N-O bond cleavage tests. Incorporation at Gly233 revealed that double alkylation of **3**, effectively crosslinking two nascent peptides, was the major product. Movement of **3** to Gly 229 reduced product loss to double alkylation and the peptide (**4**) was isolated in 0.5% yield (**Figure 3.3a**). Cleavage of the N-O bond was attempted with the aromatic thiol 4-mercaptophenylacetic acid (MPAA) and activated Zn, both of which caused disappearance of the starting material, but furnished unexpected product masses (**Figure 3.3b**). Tandem MS was used to identify the source of the mass defects, and confirmed that the mass changes occurred in the N-terminal CysGly (**Figure 3.3c**). Among the masses identified was one retaining the ethoxy fragment of the linker, which suggests a transamidation of the biotinamide of **3** during conjugation to the solid phase.

The combination of unusual cleavage products and potential for rearrangement prompted a redesign of the affinity handle.

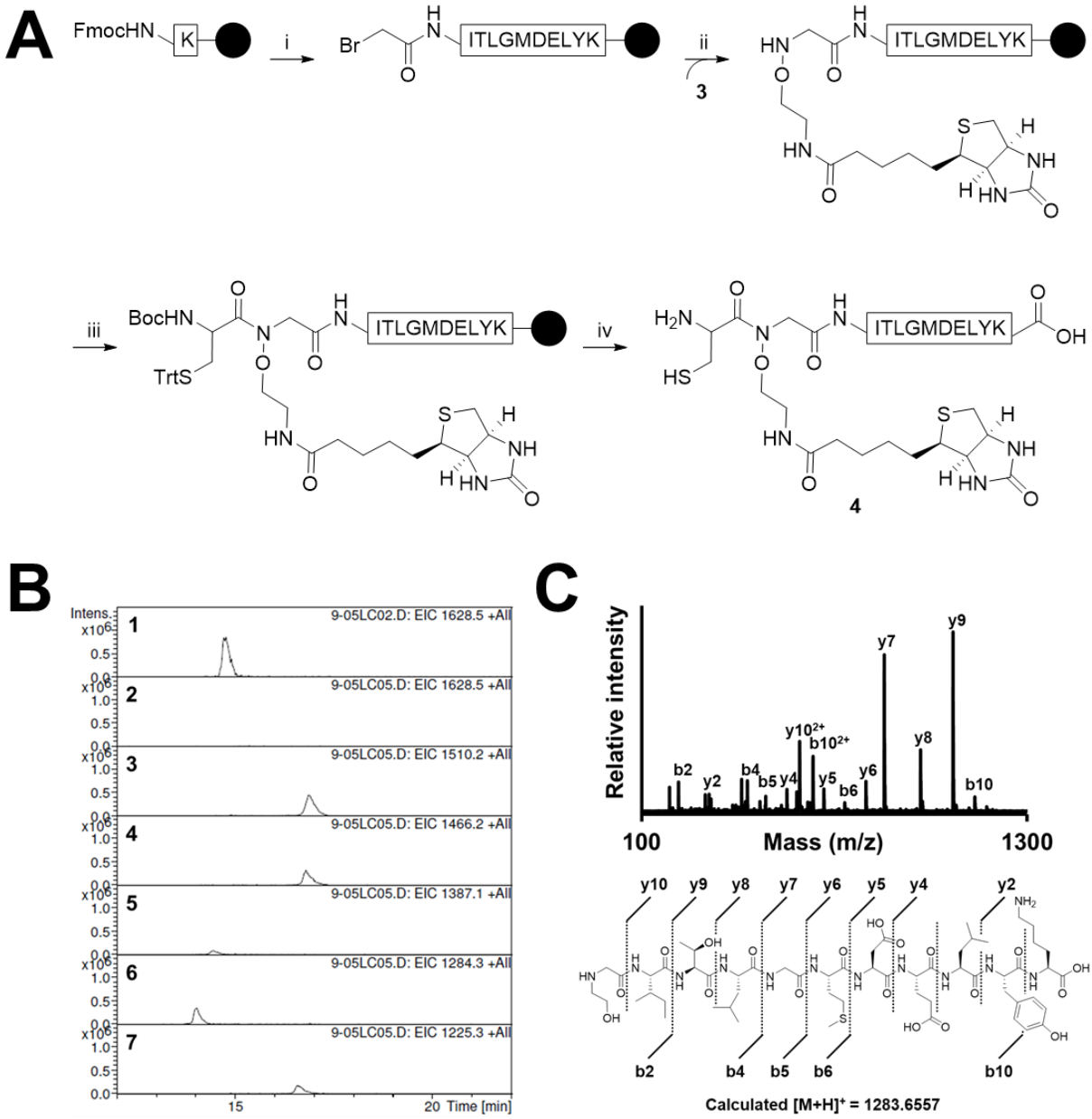
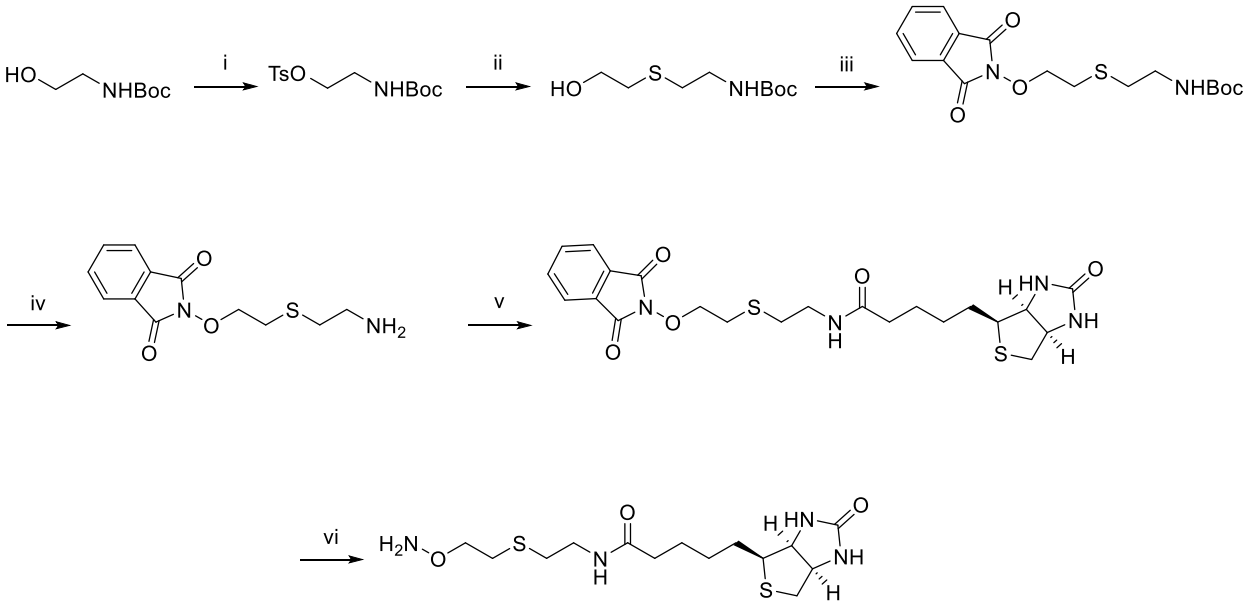


Figure 3.3. Synthesis and N-O bond cleavage of 4. (**A**) (i) standard deprotection: 20% piperidine/DMF; standard coupling: Fmoc-AA, DIC, Oxyma; bromoacetylation: bromoacetic acid, DIC; (ii) 10 eq **3**, 1:1 DMSO:DCM; (iii) Boc-Cys(Trt)-OH, HCTU, DIEA, DMF; (iv) Reagent K: 82.5/5/5/2.5 TFA/phenol/water/thioanisole/1,2-ethanedithiol (3% yield). (**B**) Extracted ion chromatograms of Zn N-O bond cleavage of **4**: (1) **4** - 0 hrs; (2) **4** – 36 hrs; (3-7) new peptide species – 36 hrs. (**C**) Tandem MS characterization of the ion in EIC 6 (m/z = 1284.3).

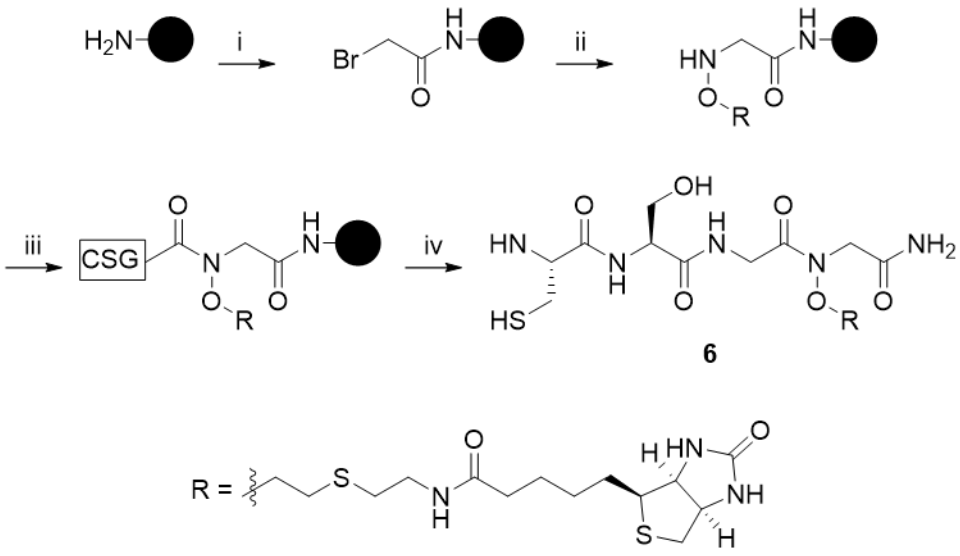
3.2.3 Synthesis and peptide incorporation of N-[2-aminooxyethylthioethanyl]-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide

The initial report of the N-O bond cleavage claimed that the pendant thiol was irrelevant, as the reaction was observed to proceed with the trityl thioether and following thiol alkylation with *N*-(2-chloroethyl)-*N,N*-dimethyl ammonium chloride. In light of this result, and to combat potential rearrangement the synthesis of N-[2-aminooxyethylthioethanyl]-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide (**5**) was undertaken (**Scheme 3.3**). The synthesis of **5** was slightly lower in overall yield (8%), however, two steps were markedly improved: (1) amidation of biotin with the biotin N-hydroxysuccinimidyl ester was higher yielding;[13] (2) phthalimide deprotection with methylamine rendered **5** and the N-methyl phthalimide byproduct readily separable.[14]



Scheme 3.3. Synthesis of 5. (i) TsCl, TEA, DCM, 88%; (ii) 2-mercaptoethanol, DBU, benzene, 82%;[15] (iii) N-hydroxyphthalimide, $P(Ph)_3$, DIAD, 40%; (iv) 4 N HCl/dioxane, quant; (v) Biotin-NHS, DIEA, DMF, 44%;[13] (vi) methylamine, 1:1 DMSO:100 mM Na_2HPO_4, pH 7, 64%.[14]

Incorporation of **5** into simpler model peptides allowed rapid assessment of N-O bond cleavage conditions. The sequence CSGG was selected for ease of synthesis, and reconstitution of the GlyGly junction at which the chemistry had first been observed. Resin bromoacetylation, the initial Sn2 and first coupling were as reported by Shelton et al in the related synthesis of mercaptoethoxy glycinamide resin.[16] The CSGG peptide (**6**) was obtained in 5% isolated yield (**Scheme 3.4**).



Scheme 3.4. Synthesis of 6. (i) bromoacetic acid, DIC; (ii) **4**, DIEA, DMSO; (iii) first coupling: 10 eq Fmoc-Gly, 9.8 eq HATU, 10 eq DIEA, DMF, overnight; standard deprotection: 20% piperidine/DMF; second coupling: 4 eq Fmoc-Ser(OtBu)-OH, 4 eq DIC, 4 eq Oxyma, DMF, overnight; standard deprotection; final coupling: 8 eq Boc-Cys(Trt)-OH, 7.9 eq HATU, 8 eq DIEA, DMF, 1 hr; (iv) Reagent K: 82.5/5/5/2.5 TFA/phenol/water/thioanisole/1,2-ethanedithiol (5% yield).

Ligation of **6** to SUMO3(C47S) α-thioester was executed, and the biotin handle enabled enrichment on streptactin, and monomeric avidin resins, though the former to a far greater extent. Enrichment on streptactin was impervious to competitive elution with desthiobiotin or HABA, however it did allow a test of on-resin handle cleavage with MPAA. The on-resin reaction successfully cleaved the biotin tag, however, LCMS indicated a mass increase equivalent to MPAA (**Figure 3.4**). Thiol adducts had been infrequently observed in the characterization of

MPAA-mediated N-O bond cleavage, though these adducts had been short-lived and reversible (Dr. Caroline Weller, personal communication). The desired N-O bond reduction product was attainable through Zn reduction, however, this was incompatible with the elution resistant ligation product.

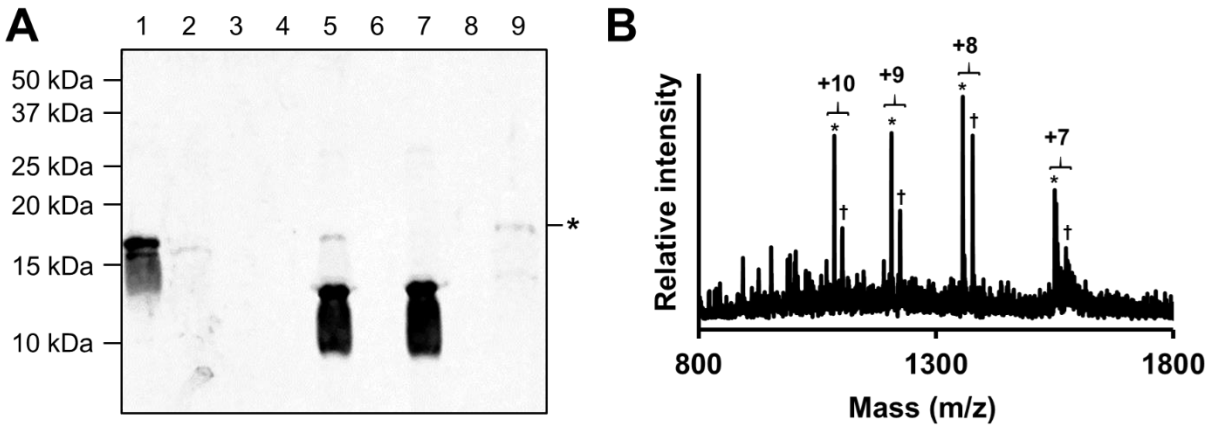


Figure 3.4. On-resin N-O bond cleavage. (**A**) SDS-PAGE (15%) of ligation, enrichment and N-O bond cleavage: (1) SUMO3(C47S) α-thioester ligation to **6**; (2) flow through; (3) final column wash; (4) empty; (5) loaded resin; (6) empty; (7) resin following on-resin cleavage; (8) empty; (9) on-resin cleavage soluble fraction. (**B**) ESI-MS of on-resin cleavage soluble fraction: (*) SUMO3(C47S)CSGG-$MPAA_1$ found 10848.9 ± 2.0 Da (Calc'd 10848.11 Da); (†) SUMO3(C47S)CSGG-$MPAA_2$ found 11014.5 ± 2.8 Da (Calc'd 11014.3 Da).

Incubation of **6** with MPAA under the conditions reported (200 mM ArSH, 50 mM Na_2HPO_4, pH 7.3, under air) also resulted in the desired N-O bond cleavage, and apparent covalent conjugation to MPAA. Sample reduction with TCEP failed to rectify the mass difference, and tandem MS narrowed the site of the mass defect to the C-terminal Gly residue. While the N-O bond cleavage is believed to proceed through a radical mechanism, the possibility of a direct nucleophilic attack by an equivalent of aromatic thiol has also been proposed. To this end the less nucleophilic aromatic thiols 3-mercaptobenzoic acid and thiosalicylic acid were also tested, and in both cases adduct formation was observed (**Figure 3.5**). A low intensity z1 ion supports formation of a C_α-S bond, though further spectroscopic evidence of this connectivity is required.

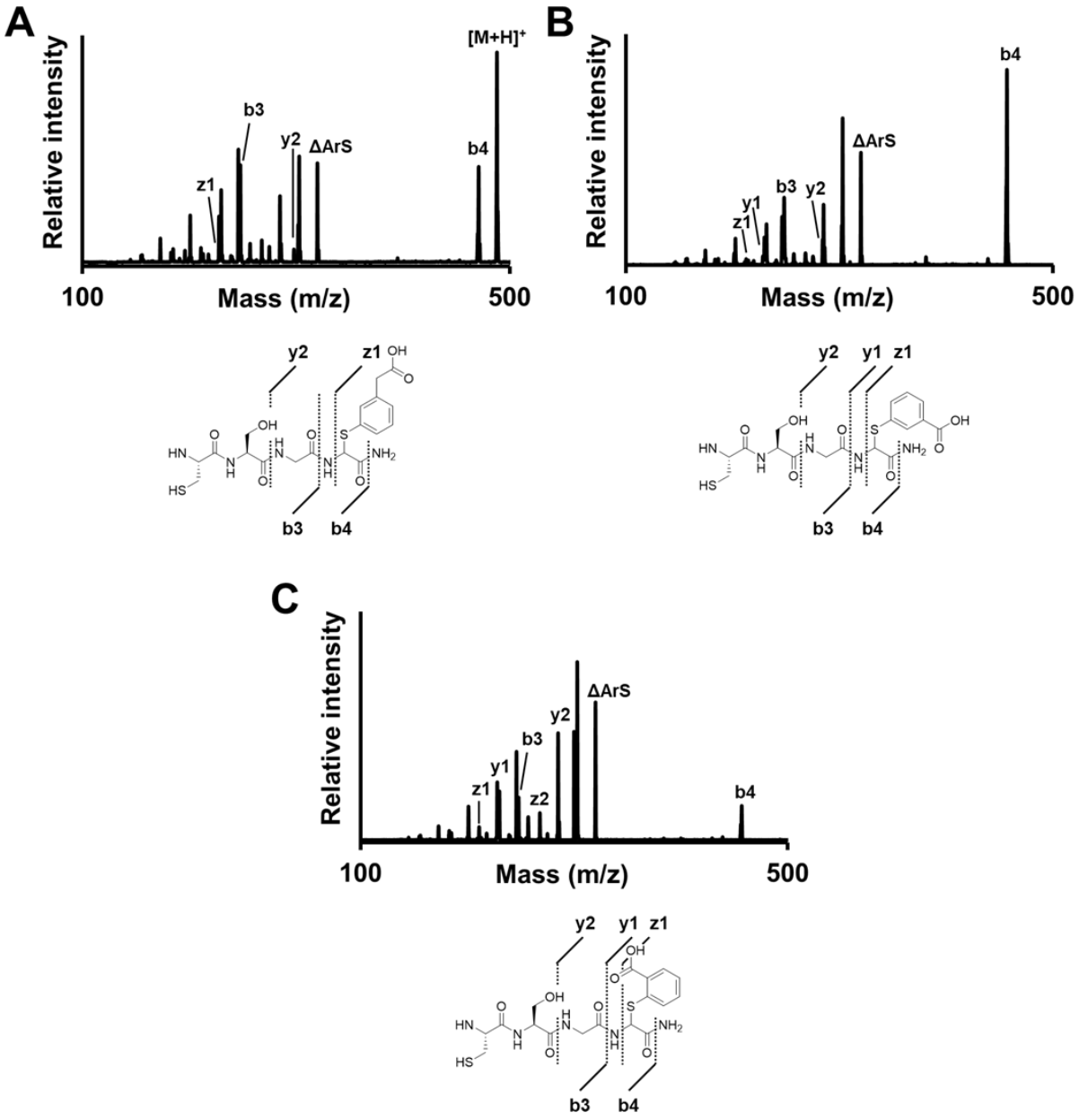


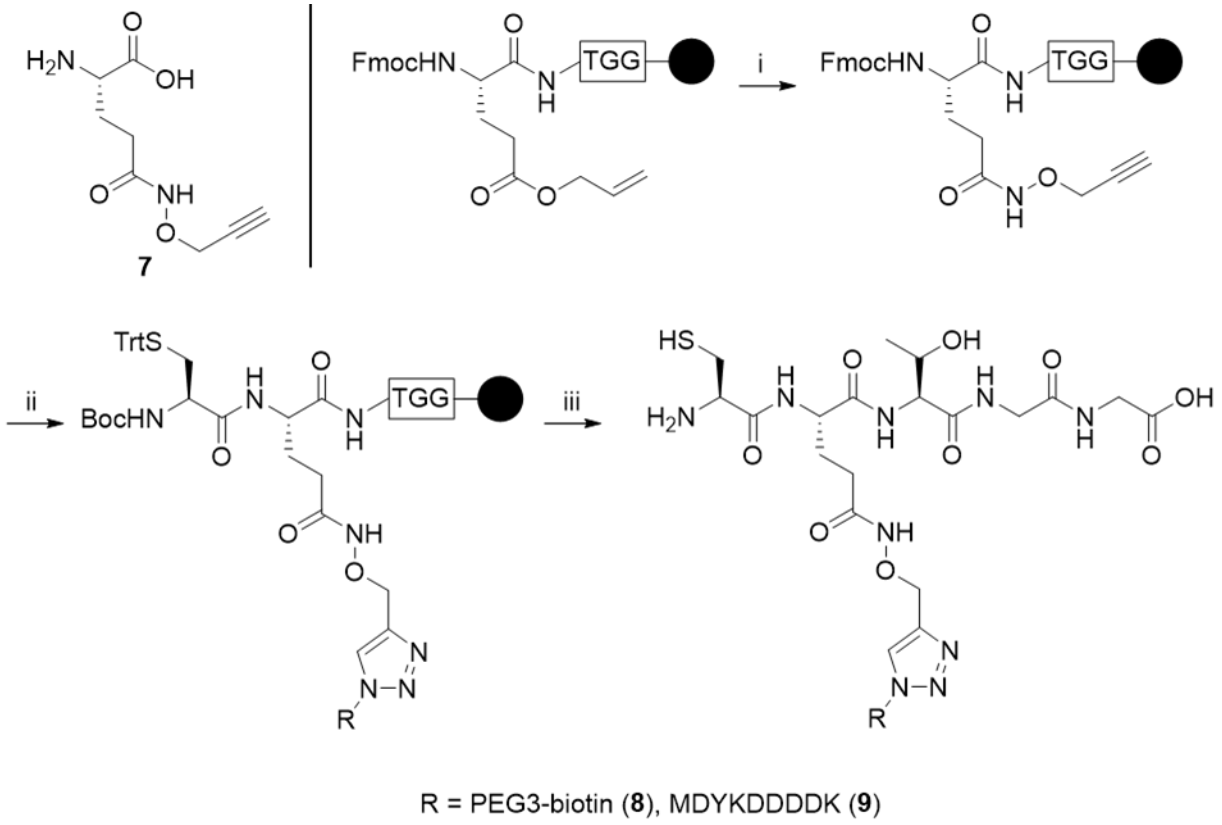
Figure 3.5. Tandem MS of adducts arising from the reaction of 6 and aromatic thiols. (**A**) Fragmentation of m/z 488.1 from reaction of **6** and MPAA: $[M+H]^+$ found 488.1 Da (calc'd 487.1 Da); b4 found 471.0 Da (calc'd 471.1 Da); ΔArS found 320.0 Da (calc'd 320.1 Da); y2 found 297.9 Da (calc'd 297.1 Da); b3 found 247.8 Da (calc'd 248.1 Da); z1 found 227.8 Da (calc'd 225.0 Da). (**B**) Fragmentation of m/z 474.3 from reaction of **6** and 3-mercaptobenzoic acid: b4 found 457.1 Da (calc'd 456.9 Da); ΔArS found 320.1 Da (calc'd 320.1 Da); y2 found 284.3 Da (calc'd 283.1 Da); z2 found 268.2 Da (calc'd 268.1 Da); b3 found 248.2 Da (calc'd 247.1 Da); y1 found 228.3 Da (calc'd 226.0 Da); z1 found 212.6 Da (calc'd 211.0 Da). (**C**) Fragmentation of m/z 474.3 from reaction of **6** and thiosalicylic acid: b4 found 456.9 Da (calc'd 457.1 Da); ΔArS found 320.0 Da (calc'd 320.1 Da); y2 found 285.0 Da (calc'd 283.1 Da); z2 found 267.9 Da (calc'd 268.1 Da); b3 found 248.0 Da (calc'd 248.1 Da); y1 found 228.0 Da (calc'd 226.0 Da); z1 found 211.0 Da (calc'd 211.0 Da).

Mild removal of the biotin moiety proved unattainable without thiol adduct formation, however, enrichment had been demonstrated, and Zn reduction had been demonstrated. Thus, incorporation of **5** at the G389 position in the p53 C-terminus was attempted on the solid phase, and proved successful, affording a 5% overall yield. Ligation of this peptide to p53(1-375)-MESNa α-thioester has been demonstrated by SDS-PAGE. Purification of the ligation product on streptavidin resin has not yet been demonstrated.

3.2.4 Incorporation and functionalization of Glutamine N_{δ}-(O-2-propyne)

The formation of a persistent covalent thiol adduct on the previously aminooxy functionalized Gly inspired the design of a third affinity handle. Incorporation of the aminooxy amide at a Gln sidechain removes the C_{α}, and thus blocks one avenue of peptide-thiol bond formation. Furthermore, reduction of a related Asn sidechain hydroxamic acid is reported with both Zn and aromatic thiols.[17] Thus **7** was contrived, wherein biotin is replaced with an alkyne to enable installation of multiple affinity handles. The incorporation of **7** on the solid phase was achieved by orthogonal protection of Glu as an O_{δ}-allyl ester, which on deprotection was condensed with O-2-propynyl hydroxylamine (**Scheme 3.5**). The peptide sequence selected corresponds to the five C-terminal residues of SUMO3 (QQTGG), with a Q89C mutation to enable the native chemical ligation. Functionalization by copper azide-alkyne click (CuAAC) was conducted on the solid phase, which mitigates off target oxidation typical to the copper catalyzed reaction (**Scheme 3.5**).[18]

CuAAC was attempted with biotin-PEG3-azide, azidoacetyl-FLAG peptides, and azidoacetyl-His6 peptides. Deprotected peptide azides proved untenable in the organic solvents required to maintain resin swelling, and rapidly precipitated on addition of Copper salts. A biotin conjugated CQ(biotin)TGG (**9**) was obtained in 33% yield, while a fully protected azidoacetyl-FLAG was employed for synthesis of triazole linked CQ(FLAG)TGG peptide conjugate (**8**) in a scant 6% overall yield.



Scheme 3.5. Structure of 7 and incorporation in 8 and 9. (i) (1) 20 mM $Pd(PPh_3)_4$, 0.4 M phenylsilane, DCM, 1 hr;[17] (2) HATU, DIEA, O-2-propynyl hydroxylamine; (ii) (1) Boc-Cys(Trt)-OH, HCTU, DIEA, 1 hr; (2) 2eq azide, 10% piperidine/DMSO, 2 eq Cu(I)Cl, 2 eq ascorbic acid, 12 hr; (iii) reagent K: 85/5/5/5/2.5 TFA/phenol/water/thioanisole/1,2-ethanedithiol, 33% (**8**), 6% (**9**).

Peptides **8** and **9** were subjected to N-O bond cleavage tests with MPAA and 3-mercaptobenzoic acid, and were found unreactive toward either thiol. Exploration of oxidant additives and Fe(III)Cl, both of which were previously reported to enhance yields, yielded no reactivity.[7] Structurally related aminooxy amides have been cleaved by hv light, though the substituent effect of the O-R group is significant, and no bond reduction was observed.[19,20] Thus chastened, Zn reduction was conducted and found to proceed quantitatively in the presence and absence of chaotropes, and across a surprising pH range (**Figure 3.6**). To confirm that this reactivity is generalizable to proteins peptides **8** and **9** were ligated to a SUMO3(2-88)C47S α-thioester. The ligation of **8**, followed by alkylation of C89 to mimic the native Q, affords SUMO3(2-93)C47S/Q89C(acetamide)/Q90(PEG3-biotin) (**10**), which is isolable by enrichment on

streptavidin resin, and elution with 90% DMF. The ligation of **9** and subsequent alkylation affords SUMO3(2-93)C47S/Q89C(acetamide)/Q90(FLAG) (**11**), which is isolable by α-FLAG immunoprecipitation or RP-HPLC. Zn reduction of **11** required 6 M guandine and low pH, and afforded a 56% yield after 24 hours at 37 ˚C (**Figure 3.7**).

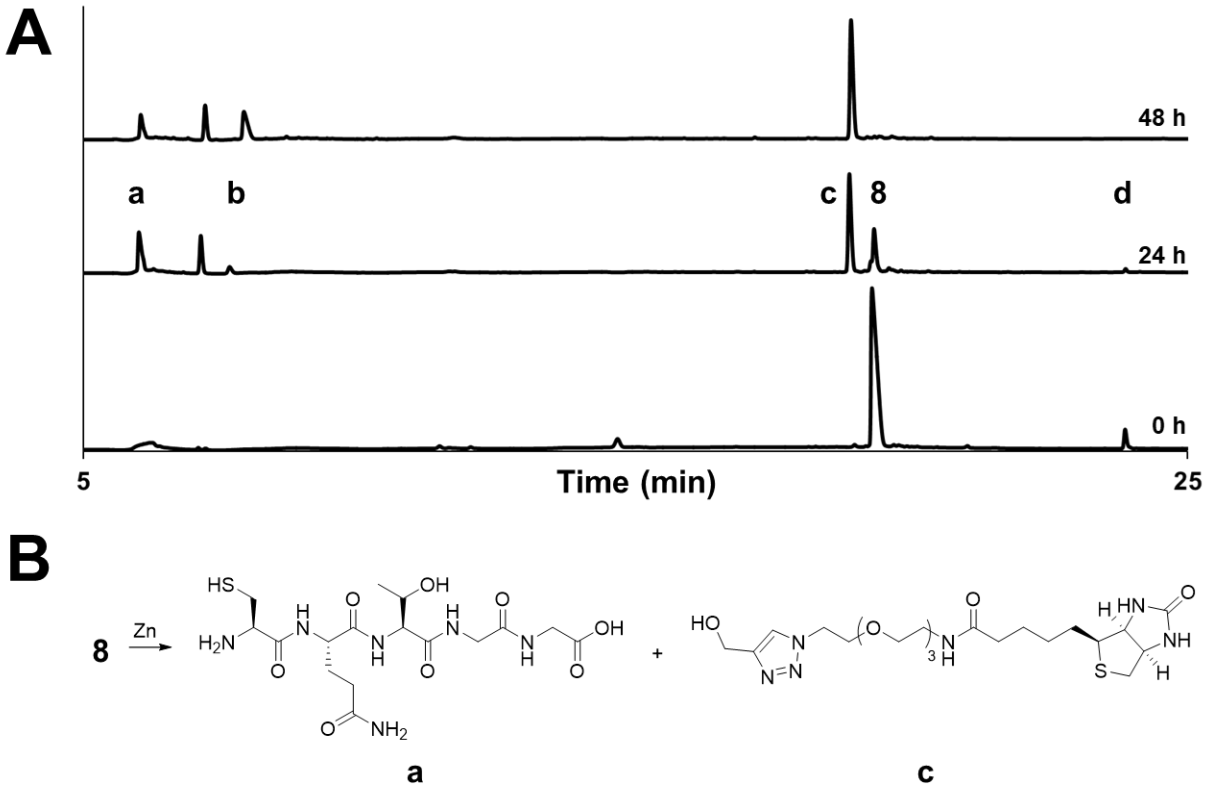


Figure 3.6. Non-denaturing cleavage of 8 N-O bond. (**A**) RP-HPLC chromatograms collected at 0, 24 and 48 hrs (C18 Poroshell, 0-35% B, 30'): (a) CQTGG peptide; (b) CQTGG peptide disulfide; (c) Biotin-PEG3-triazole; (d) **8** disulfide. **8** (100 ug, 0.5 mM) was treated with Zn (1.4 mg) in citric acid / sodium phosphate buffer at pH 6, and nutated at 25 ˚C. (**B**) Depiction of N-O bond cleavage products.

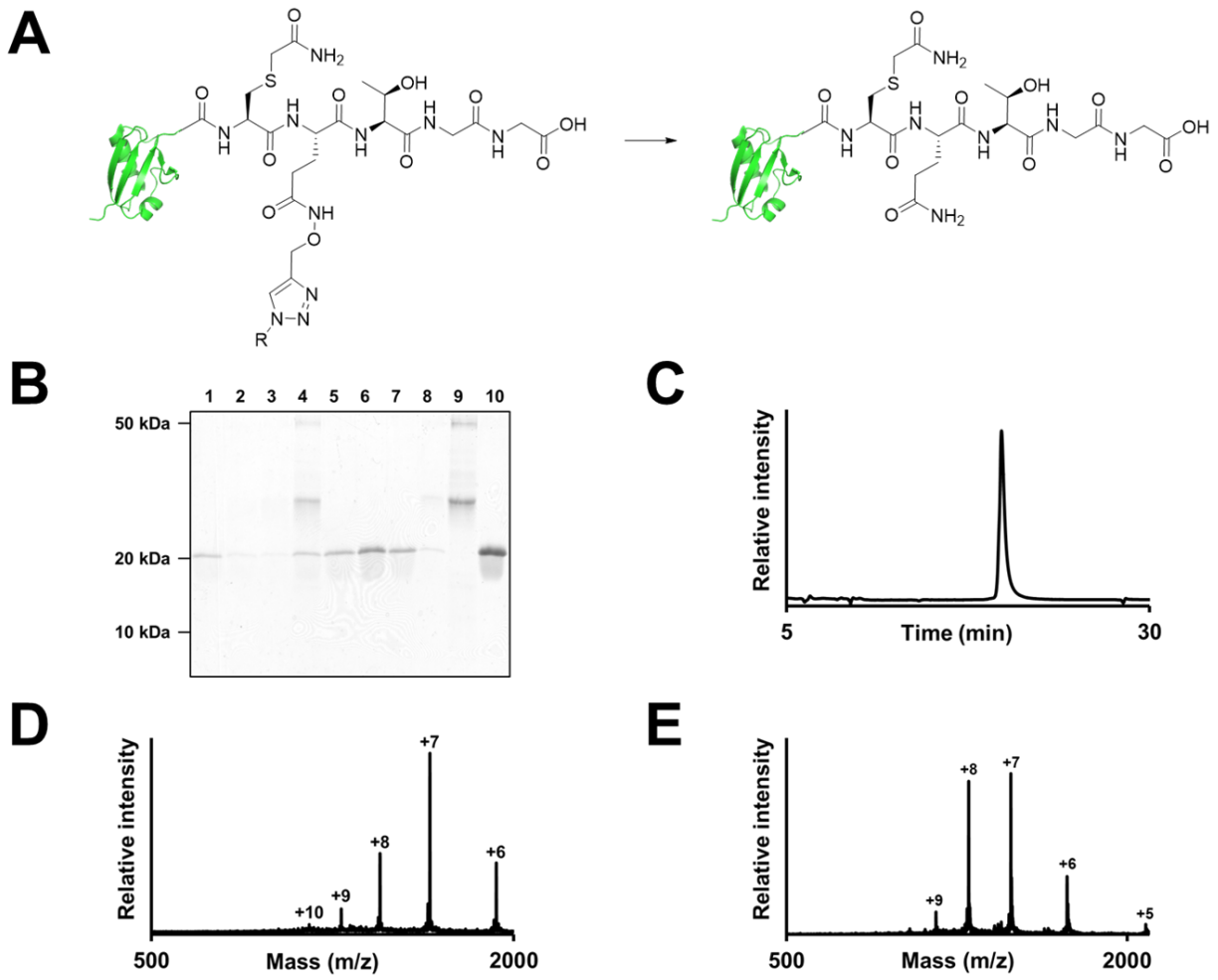


Figure 3.7. N-O bond cleavage of 11. (**A**) Schematic depiction of N-O bond cleavage: **11** (1 mg/ml) in 6 M guanidine, pH 3, Zn (20 mg/100 µg protein), 37 C, 24 h, 56% (pdb 1U4A).[22] (**B**) α-FLAG purification of **11**; (1) input; (2) flow through; (3) final wash; (4) loaded resin; (5) elution 1; (6) elution 2; (7) elution 3; (8) elution 4; (9) eluted resin; (10) concentrated elution. (**C**) RP-HPLC chromatogram of N-O bond cleavage product. (**D**) ESI-MS of **11** found 11560.7 ± 1.9 Da (calc'd 11560.7 Da). (**E**) ESI-MS of N-O bond cleavage product found 10410.0 ± 2.1 Da (calc'd 10409.5 Da).

3.3 Conclusions and outlook

Methods for affinity handle incorporation and traceless removal are highly useful in NCL and EPL reactions. Traceless chemical removal is practically limit to Lys N_{ε}-carbamates, which we have addressed through the development of strategies for Gly and Gln. Development of the cleavage chemistry has shed new light on the mechanism of aminooxy amide N-O bond reduction by aromatic thiols. The data reported here suggest that formation of a C_{α} radical may be critical

to N-O bond. Each of reported purification handles has proven reducible by Zn, though effective reduction of protein conjugates requires denaturing conditions and low pH. Thus these approaches present a potential solution to semi-syntheses in which a peptide component lacks Lys residues for reversible modification.

The unanticipated reactions observed reveal multiple potential future lines of investigation. Further characterization of the reduction by aromatic thiol could afford a more satisfying explanation for the adducts observed. The selective chemistry detected may offer a method for site-specific protein labeling. Incorporation of a glycine Weinreb amide on the solid phase is far more synthetically facile, and may prove useful as the literature on selective photocatalytic protein conjugation grows.[21]

3.4 Experimental Procedures

3.4.1 General Methods

Rink-amide resin (0.46 mmol/g substitution) was purchased from Chem-Impex (Wood Dale, IL). Fmoc-L-amino acids were purchased from either AnaSpec (Fremont, CA) or AGTC Bioproducts (Wilmington, MA). Other chemical reagents were purchased from either Fisher Scientific (Pittsburgh, PA) or Sigma-Aldrich Chemical Company (St. Louis, MO). DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA). Gene sequencing was performed by and Genewiz (South Plainfeld, NJ). Plasmid mini-prep, PCR purification and gel extraction kits were purchased from Qiagen (Valencia, CA). Restriction enzymes were purchased from either Fermentas (Thermo Fisher Scientific, Philadelphia, PA) or New England BioLabs (Ipswitch, MA). Reversed-phase HPLC (RP-HPLC) was performed on a Varian (Palo Alto, CA) ProStar HPLC with either a Grace-Vydac (Deerfield, IL) analytical C18 column (5 micron, 150 x 4.6 mm) at a flow rate of 1 ml/min, or a Grace-Vydac preparative C18 column (10 micron, 250 x 22 mm) at a flow rate of 9 ml/min; RP-HPLC was performed using 0.1% trifluoroacetic acid (TFA) in water (A) and 90% acetonitrile, 0.1% TFA in water (B) as the mobile phases. Solid phase peptide synthesis was performed on a Liberty Blue Automated Microwave Peptide Synthesizer (CEM Corporation, Matthews, NC). Mass spectrometric analysis was conducted on a Bruker (Billerica, MA) Esquire, or a Thermo Scientific (Waltham, MA) LTQ Obitrap ESI-MS. Analytical RP-HPLC-mass spectrometry (LC-ESI-MS) was performed on a Hewlett-Packard (Palo Alto, CA) 1100-series LC linked to the Bruker Esquire ESI-MS with an Agilent (Santa Clara, CA) Zorbax C18 column (3.5 micron, 100 x 2.1 mm) using 5% acetonitrile, 1% acetic acid in water (C) and acetonitrile, 1% acetic acid (D) as the mobile phases. NMR spectra were recorded on Bruker Avance AV-300, AV-301, or AV-500 instruments. Circular dichroism measurements were performed on a JASCO (Easton, MD) J-720 spectropolarimeter. Fluorescence was measured with a Biotek synergy H1 hybrid multimode microplate reader (Winooski, Vermont).

3.4.2 Synthesis of N-(2-aminooxyethyl)-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide

*R2-(1,3-Dioxo-1,3-dihydro-isoindol-2-yloxy)-ethyl-1-carbamic acid t-butyl ester (**3a**)*

N-hydroxyphthalimide (2.81 g, 17.24 mmol, Sigma Aldrich) was combined with triphenylphosphine (3.39 g, 12.93 mmol, Fluka) and t-butyl 2-hydroxyethylcarbamate (1.39 g, 8.62 mmol) in tetrahydrofuran (28.7 mL, anhydrous, Fisher Scientific), stirred for 10' at room temperature, then brought to 0 ˚C in an ice bath. Diisopropylazidodicarboxylate (DIAD, 1.86 mL, 9.48 mmol) was added dropwise over 1 h while on ice, after which the ice bath was removed, and the reaction allowed to stir over night. The reaction was concentrated under vacuum, taken up in ethyl acetate, washed 3x with 1 volume of saturated $NaHCO_3$, then 3x with 1 volume of H_2O, then 3x with 1 volume of brine, and finally the organics were dried over Na_2SO_4 and concentrated. The reaction mixture was dry loaded on silica gel (100 g silica, 14x1 inch) and run from 20% ethyl acetate in hexanes to 30% ethyl acetate in hexanes (2.06 g, 78% yield, SW-VII-28).[1]

^{1}H NMR (500 MHz, $CDCl_3$): δ 7.83 (m, 4H, Ar), 5.73 (bs, 1H, NH), 4.29 (t, 2H, CH_2O), 3.47 (q, 2H, CH_2N), 1.48 (s, 9H, $C(CH_3)_3$). ^{13}CNMR (500 MHz, $CDCl_3$): δ 164.01, 155.98, 134.73, 128.83, 123.71, 79.50, 77.57, 38.84, 28.40. ESI-MS m/z Calcd for $C_{15}H_{18}N_2O_6$ $[M+H]+$ 306.32, 207.1 $[M-Boc]^+$, 329.1 $[M+Na]^+$ (SW-VII-28; VIII-11).

*R2-(1,3-Dioxo-1,3-dihydro-isoindol-2-yloxy)-ethyl-1-amine (**3b**)*

3a (2.06 g, 4.0 mmol) is taken up in 4N HCl/dioxane (10 mL) under argon. The reaction is stirred for 3 hours, then concentrated under vacuum and used without further purification (SW-VII-33, 46).[1]

^{1}H NMR (300 MHz, $CDCl_3$): δ 7.87-7.89 (m, 2H, Ar), 7.74-7.75 (m, 2H, Ar), 5.64 (bs, 2H, NH_2), 3.96-3.99 (q, 2H, CH_2), 3.82-3.85 (t, 2H, CH_2). ESI-MS m/z Calcd for $C_{10}H_{10}N_2O_3$ $[M+H]^+$ 207.07, found 207.00 $[M+H]^+$ (SW-VII-48).

*N-((1,3-Dioxo-1,3-dihydro-isoindol-2-yloxy)-ethyl)-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide (**3c**)*

Biotin (1.28 g, 5.26 mmol) and Oxyma (1.49 g, 10.52 mmol) were combined in a round bottom flask with stir bar and dried under vacuum. DMF (70 ml) is added on ice with stirring, followed by diisopropylcarbodiimide (1.63 ml, 10.52 mmol), and the mixture is stirred until homogenous. **3b** (1.08 g, 5.26 mmol) is added as a solid, and the reaction is removed from ice and stirred a further 2 hrs. The reaction is quenched by addition of water acidified with TFA (1 ml water: 1 µl TFA), then vacuum concentrated. The reaction is purified by silica gel chromatography over 3 to 6% MeOH/DCM (0.68 g, 1.57 mmol, 30% yield) (SW-VIII-19/30).

^{1}H NMR (500 MHz, $CDCl_3$): δ 9.68 (bs, 1H, NH), 7.83-7.85 (m, 2H, Ar), 7.72-7.74 (m, 2H, Ar), 6.65 (bs, 1H, NH), 5.74 (bs, 1H, NH), 4.50 (m, 1H, CH), 4.31 (m, 1H, CH), 4.09 (m, 2H, OCH_2), 3.97-3.98 (m, 2H, NCH_2), 2.88-3.16 (m, 2H, SCH_2), 2.74-2.76 (m, 1H, SCH), 2.18 (m, 2H, $C(O)CH_2$), 1.25-1.69 (m, 6H, $CH_2CH_2CH_2$). HRMS (ESI) m/z Calcd for $C_{20}H_{24}N_4O_4S$ $[M+H]^+$ 433.1546, found 433.1550 $[M+H]^+$ (SW-VIII-10).

*N-(2-aminooxyethyl)-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide (**3**)*

3c (0.68 g, 1.57 mmol) was taken up in a minimal volume of 1:1 DCM:MeOH. To this solution was added 50% hydrazine hydrate (493 µl, 7.86 mmol), and the reaction was stirred for 70 min, then concentrated *in vacuo*. The precipitate was vacuum filtered and washed extensively with DCM. The crude was solubilized in a minimal volume of water acidified with TFA (1000:1 water:TFA), and purified by C18 preparative RP-HPLC over a (C18 preparative, 0-20% B, 90 min, 0.33 g, 0.76 mmol, 48% yield).

^{1}H NMR (500 MHz, DMSO-d_6): δ 10.70 (bs, 1H, NH), 8.14 (t, 1H, *J* = 5.72, NH), 6.47 (bs, 1H, NH), 6.43 (s, 1H, NH), 4.33 (m, 1H), 4.14 (m, 1H), 3.94 (t, 2H, *J* = 5.39, OCH_2), 3.31 (q, 2H, *J* =

5.49, NCH_2), 3.10 (m, 1H), 2.82 (dd, 1H, *J* = 12.46, 5.07), 2.56-2.59 (d, 1H), 2.10 (t, 2H, *J* = 7.42), 1.28-1.65 (m, 6H, CH_2 CH_2CH_2). ^{13}CNMR (500 MHz, $CDCl_3$): δ 173.54, 163.35, 73.65, 62.10, 59.10, 56.39, 55.32, 37.98, 37.01, 35.44, 28.62, 25.63. HRMS (ESI) m/z Calcd for $C_{12}H_{22}N_4O_3S$ $[M+H]^+$ 303.1491, found 303.0833 $[M+H]^+$ (SW-VIII-36/37).

3.4.3 Synthesis of N-[2-aminooxyethylthioethanyl]-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide

*N-(tert-butoxycarbonyl)-(O-p-toluenesulfonyl)-2-aminoethanol (**5a**)*

As previously reported.[2] Briefly, t-butyl 2-hydroxyethylcarbamate (5.0 g, 31 mmol) and triethylamine (9.7 ml, 69 mmol) are combined in DCM (45 ml) on ice and stirred. In a separate flask *p*-toluenesulfonyl chloride (6.5 g, 34.5 mmol) is dissolved in DCM (45 ml) and the two solutions are combined on ice with stirring. The reaction is removed from ice and stirred for 48 hours at ambient temperature. After 48 hours precipitate is removed by vacuum filtration, and the solids are washed with DCM. The organic layer is extracted thrice with 10% citric acid (3 x 25 ml), then thrice with saturated NaCl (3 x 25 ml), then dried with Na_2SO_4, and concentrated *in vacuo* (8.26 g, 88%).

^{1}H NMR (300 MHz, CDCl3): δ 7.79 (d, 2H, *J* = 8.29, Ar), 7.35 (d, 2H, *J* = 7.99, Ar), 4.93 (bs, 1H, NH), 4.06 (t, 2H, *J* = 4.99, CH_2O), 3.38 (q, 2H, *J* = 5.16, CH_2N) (SW-IX-22).

*2-[S-[N-(tert-butoxycarbonyl)-2-aminoethyl]]mercaptoethanol (**5b**)*

As previously reported.[2] Briefly, β-mercaptoethanol (39 mmol), 1,8-diazabicyclo[5.4.0]undec-7-ene (5.8 ml, 39 mmol), and **5a** (9.73 g, 31 mmol) were combined in benzene (150 mL) and stirred overnight at ambient temperature. The solution was extracted once with water (50 ml), twice with 1 N HCl (2 x 50 ml), and once with saturated NaCl (50 ml). The organic layer was dried over Na_2SO_4, and concentrated *in vacuo* (4.93 g, 22 mmol, 82%).

^{1}H NMR (300 MHz, CDCl3): δ 5.02 (bs, 1H, OH), 3.75 (t, 2H, *J* = 5.56, CH_2O), 3.33 (q, 2H, *J* = 6.44, CH_2N), 2.74 (t, 2H, *J* = 5.95, SCH_2CO), 2.66 (d, 2H, *J* = 6.52, SCH_2CN) (SW-IX-22).

*(1,3-Dioxo-1,3-dihydro-isoindol-2-yloxy)-2-[S-[N-(tert-butoxycarbonyl)-2-aminoethyl]]thioethane (**5c**)*

5b (4.93 g, 22 mmol), triphenylphosphine (8.78 g, 33 mmol), and N-hydroxyphthalimide (7.28 g, 44.64 mmol) are combined in anhydrous THF (100 ml) and chilled on ice. Diisopropyl azidocarboxylate (4.82 mL, 25 mmol) is added dropwise over 60 minutes with stirring. After 12 hours the organics are stripped *in vacuo* and taken up in DCM, then washed twice with water (2 x 50 ml), thrice with saturated $NaHCO_3$ (3 x 50 ml), and twice with saturated NaCl (2 x 50 ml), then dried over Na_2SO_4 and concentrated *in vacuo.* The crude is dry loaded on silica gel and column purified from 0 to 8% acetonitrile in toluene (3.21 g, 8.8 mmol) (SW-IX-29). *Note: Diisopropyl hydrazinecarboxylate runs just above the product (product R_f 0.45 in 1:1 hexanes:ethyl acetate).*

^{1}H NMR (300 MHz, $CDCl_3$): δ 7.85 (m, 2H, *J* = 5.52, 3.06, Ar), 7.76 (m, 2H, *J* = 5.54, 3.15, Ar), 4.96 (bs, 1H, NH), 4.35 (t, 2H, *J* = 6.87, CH_2O), 3.34 (q, 2H, *J* = 6.18, CH_2N), 2.93 (t, 2H, *J* = 6.87, SCH_2CO), 2.75 (t, 2H, *J* = 6.46, SCH_2CN) (SW-IX-22).

*1-(1,3-Dioxo-1,3-dihydro-isoindol-2-yloxy)-2-(aminoethanyl)thioethane (**5d**)*

5c (3.21 g, 8.8 mmol) is taken up in 4N HCl/dioxane (20 mL) under argon and stirred for 1 hr, then concentrated *in vacuo*, washed with DCM, dried and used without further purification (SW-IX-31).

^{1}H NMR (300 MHz, $CDCl_3$): δ 8.30 (bs, 2H, NH_2), 7.70-7.81 (m, 4H, Ar), 4.39 (t, 2H, *J* = 6.52, CH_2O), 4.35 (t, 2H, *J* = 6.87, CH_2O), 3.35 (bs, 2H, CH_2N), 3.10 (t, 2H, *J* = 5.89, SCH_2CN), 2.98 (t, 2H, *J* = 6.16, SCH_2CO).

*2,5-dioxopyrrolidin-1-yl 5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanoate (**5e**)*

As previously reported.[3] Briefly, biotin (1.36 g, 5.6 mmol) and N-hydroxysuccinimide (0.64 g, 5.6 mmol) are combined in a round bottom flask and dried under vacuum. DMF (41 ml) is added, and the mixture is heated to 70 ˚C with stirring so that the solids dissolve. Dicyclohexyl carbodiimide (1.49 g, 7.2 mmol) is added to the heated solution, and the reaction is removed from heat and allowed to stir overnight at ambient temperature. The solids are removed by vacuum filtration, and the organics are concentrated *in vacuo*. The crude is triturated with diethyl ether, then washed with diethyl ether and used without further purification (2.07 g) (SW-IX-30).

*N-[1,3-Dioxo-1,3-dihydro-isoindol-2-yloxyethylthioethanyl]-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide (**5f**)*

5d (1.04 g, 3.4 mmol) and **5e** (1.49 g, 4.4 mmol) are combined in a round bottom flask and dried under vacuum. Under argon, dry DMF (80 ml) is added with stirring, followed by dry DIEA (4.65 ml, 26.4 mmol) added dropwise. The reaction is stirred overnight, then concentrated *in vacuo*, and purified by silica gel column chromatography from 0 to 8% MeOH/DCM (0.75 g, 44% yield) (SW-IX-39).

^{1}H NMR (300 MHz, $CDCl_3$): δ 7.76-7.84 (m, 4H, Ar), 6.57 (bs, 1H, NH), 5.89 (bs, 1H, NH), 5.17 (bs, 1H, NH), 4.51-4.53 (m, 1H), 4.37 (t, 2H, *J* = 6.39, CH_2O), 4.33-4.35 (m, 1H), 3.48 (q, 2H, *J* = 6.07, CH_2N), 3.15-3.19 (m, 1H), 2.93 (t, 2H, *J* = 3.19, SCH_2CO), 2.91 (m, 1H), 2.80 (t, 2H, *J* = 6.22, SCH_2CN), 2.27-2.29 (m, 1H), 1.54-1.78 (m, 8H).

*N-[2-aminooxyethylthioethanyl]-5-[(3aS,4S,6aR)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanamide (**5**)*

5f (121 mg, 0.25 mmol) is taken up in a 1:1 mixture of DMSO (5 ml) and 0.1 M Na_2HPO_4, pH 7 (5 ml), and methylamine hydrochloride (60 mg, 0.89 mmol) is added. The reaction is mixed for 44 hrs at 25 ˚C, then the soluble fraction is purified by RP-HPLC (C18 Preparative, 0-73% B – 60 min, 74 mg, 0.15 mmol, 64% yield) (SW-IX-53).

^{1}H NMR (300 MHz, D_2O): δ 4.50 (dd, 1H, *J* = 7.90, 4.78), 4.32 (dd, 1H, *J* = 7.90, 4.45), 4.15 (t, 2H, *J* = 6.10, CH_2O), 3.30 (t, 2H, *J* = 6.45, CH_2N), 3.21 (m, 1H), 2.88 (dd, 1H, *J* = 13.07, 4.93), 2.78 (t, 2H, *J* = 6.32, SCH_2CO), 2.63-2.68 (m, 3H), 2.17 (t, 2H, *J* = 7.12, $CH_2C(O)$), 1.28-1.64 (m, 6H). HRMS (ESI) m/z Calcd for $C_{14}H_{27}N_4O_3S_2$ $[M+H]^+$ 363.1525, found 363.1539 $[M+H]^+$ (SW-IX-35/43).

3.4.4 Synthesis of azidoacetic acid

As previously reported.[4] Briefly, sodium azide (6.95 g, 50 mmol) is dissolved in milliQ water (30 ml) and chilled on ice with stirring. To this solution is added bromoacetic acid (7.15 g, 100 mmol) slowly over 10 min. The reaction is removed from ice and allowed to stir overnight, then the pH is adjusted to 1 with hydrochloric acid and extracted five times with diethyl ether (5 x 10 ml). The ether layers are combined and dried over Na_2SO_4, then concentrated and used without further purification (3.0 g, 29.7 mmol, 59%).

^{1}H NMR (300 MHz, $CDCl_3$): δ 11.69 (s, 1H, CO_2H), 3.99 (s, 2H, CH_2).

3.4.5 Molecular cloning

Cloning pTXB1-p53(1-375)-MxeGyrA-CBD, and pTXB1-SUMO3(1-88)C47S-AvaDnaE-AAFN-His6

The plasmid pTXB1-SUMO3(C47S)-AvaDnaE-AAFN-His6, containing the human SUMO3 gene *Smt3(1-93)* with a C47S mutation,[5] was used to generate the plasmid pTXB1-SUMO3(1-88)C47S, which lacks the 5 C-terminal amino acids of SUMO3. The plasmid containing human p53-(1-393) was a gift from Cheryl Arrowsmith (Addgene plasmid # 24859), and was used to clone pTXB1-p53(1-373)-MxeGyrA-CBD (performed by Nicholas Senger, Chatterjee lab). The plasmid pTXB1-p53(1-373)-MxeGyrA-CBD was used to generate pTXB1-p53(1-375)-MxeGyrA-CBD.

Primer	**DNA Sequence (5'- to -3')**
SUMO3(C47S)Δ5 Forward Primer	TGCCTGAGCTATGATACCGAAGTGCTGACCG
SUMO3(C47S)Δ5 Reverse Primer	CTGGAACACGTCGATGGTGTCCTCGTCCTCC
p53(1-373) Forward Primer	GGGAATTCCATATGGAGGAGCCGCAGTCAGATCCTAGC
p53(1-373) Reverse Primer	GGTGGTTGCTCTTCCGCACTTTTTGGACTTCAGGTGGCTGGAGTG
p53 667-688 sequencing	CACCATGAGCGCTGCTCAGATA
p53(1-375) Forward Primer	GCC ACC TGA AGT CCA AAA AGG GTC AAT GCA TCA CGG GAG ATG C
p53(1-375) Reverse Primer	GCA TCT CCC GTG ATG CAT TGA CCC TTT TTG GAC TTC AGG TGG C

The pTXB1-SUMO3(1-88)C47S Q5 mutagenesis product was transformed into chemically competent DH5α *E. coli*. A plasmid stock was obtained by miniprep following the growth of a single transformant; the plasmid was sequenced and then used to transform BL21(DE3) *E. coli* for overexpression. The PCR product p53(1-373) was isolated by gel extraction, then subject to restriction digest with *NdeI* and *SapI*. Identically digested pTXB1 was treated with calf intestinal alkaline phosphatase, then gel purified. Digested p53(1-373) and pTXB1 were combined and treated with T4 DNA ligase for 16 hours, then transformed into chemically competent DH5α *E.*

coli. A plasmid stock was obtained by miniprep. The quickchange product of p53(1-375) site directed mutagenesis was treated with *DpnI* for 1 hr at 37 ˚C, then 20 min at 80 ˚C, and used to transform chemically competent DH5α *E. coli*. A plasmid stock was obtained by miniprep. All sequences were verified by sequencing with T7 forward primer, and both p53 sequences were further verified with a T7 reverse primer, as well as the p53 sequencing primer described above (Genewiz).

3.4.6 Protein overexpression and purification

Overexpression of SUMO3(1-88)C47S-AvaDnaE-His6 fusion and purification of SUMO3(2-88)C47S-MESNa α-thioester

Four liters of BL21 *E. coli* transformed with pTXB1-SUMO3(1-88)C47S-AvaDnaE-His6 were grown (25 g LB/L, 100 mg Ampicillin, 10 mL starter culture used to inoculate) at 37 ˚C to an OD_{600} of 0.4, then chilled on ice and moved to a 16 ˚C shaker and grown to an OD_{600} of 0.6. Induction was accomplished with 300 µL 1 M IPTG / 1 L culture, and cells were grown for a further 16 h. Cells were pelleted by centrifugation (6 k rpm, 4 ˚C, 15') to give a pellet mass of 29.45 g. Lysis buffer (90 ml, 50 mM NaPi, 300 mM NaCl, pH 8; 2 ml/1 g pellet) was used to resuspend the cell pellet, and cells were lysed by sonication (3x3', on ice). Insoluble material was pelleted by centrifugation (16 k rpm, 4 ˚C, 30'), after which the supernatant was filtered (0.45 µm) and divided between two 50 mL columns each loaded with 5 mL NiNTA resin pre-equilibrated with lysis buffer. After binding for 60' the lysate supernatant was drained and the columns were each washed with 20 column volumes (CV) of lysis buffer containing 5 mM imidazole, then 20 CV 20 mM imidazole in lysis buffer, 5 CV 50 mM imidazole in lysis buffer. The desired SUMO3(2-88)C472-AvaDnaE-His6 product was eluted in 5x2 CV 250 mM imidazole in lysis buffer fractions (M1 is processed in *E. coli*). Evaluation of fractions by SDS-PAGE (15%) revealed that the 50 mM wash and all 250 mM elutions could be pooled (150 mL total). This pool was dialyzed twice against 4 L of buffer containing 100 mM NaPi, 150 mM NaCl, 1 mM EDTA, 1 mM MESNa, pH 7.2 (2 h / dialysis, 4˚C).

The dialyzed pool was brought to 100 mM in MESNa, (Vf = 160 mL), and allowed to react for 12 h at 30 ˚C. The reaction was lyophilized to dryness and purified by RP-HPLC (C18 Preparative, 30-70% B, 60' gradient).

Overexpression of p53(1-375)-MxeGyrA-CBD and purification of p53(1-375)-MESNa α-thioester

One liter of BL21 *E. coli* transformed with pTXB1-p53(1-375)-MxeGyrA-CBD was grown (25 g LB/L, 100 mg Ampicillin, 10 mL starter culture used to inoculate) at 37 ˚C to an OD_{600} of 0.4. Induction was accomplished with 300 µL 1 M IPTG / 1 L culture, and cells were grown for a further 3 h. Cells were pelleted by centrifugation (6 k rpm, 4 ˚C, 15') to give a pellet mass of 4.04 g. Lysis buffer (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 Mm KH_2PO_4, pH 7.4; 2 ml/1 g pellet) was used to resuspend the cell pellet, and cells were lysed by sonication (3x3', on ice). Insoluble material, including p53(1-375)-MxeGyrA-CBD, was pelleted by centrifugation (16 k rpm, 4 ˚C, 30'), and the supernatant was discarded. The pellet was washed with 2 M urea (20 mL), and sonicated (3x3', on ice), then pelleted by centrifugation (16 k rpm, 4 ˚C, 15') and the supernatant was discarded. The washed pellet was taken up in refolding buffer (6 M urea, 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 Mm KH_2PO_4, pH 7.4; 3.3 ml/1 g pellet), and then clarified by centrifugation (16 k rpm, 4 ˚C, 15'). The supernatant was filtered (0.45 µM), and dialyzed against 4 M urea, 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 Mm KH_2PO_4, pH 7.4 (2 hr, 4 ˚C, 1 L/10 mL dialysate), then against 2 M urea, 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 Mm KH_2PO_4, pH 7.4 (2 hr, 4 ˚C, 1 L/10 mL dialysate). The dialysate was bound to 5 mL chitin resin pre-equilibrated with dialysis buffer and bound over night at 4 ˚C. The column was washed with dialysis buffer (20 Column Volumes, CV), and 35 mL thiolysis buffer was applied to the column (100 mM sodium 2-mercaptoethanesulfonate (MESNa), 2 M urea, 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 Mm KH_2PO_4, pH 7.4). The first elution was collected after 48 hr at 4 ˚C, and a second 35 mL of thiolysis buffer was applied for a further 48 hr. The chitin column eluent was frozen and lyophilized, then purified by preparative RP-HPLC (30-70% B, 60 min).

3.4.7 Solid-phase peptide synthesis

General Method

Peptides were synthesized by microwave-assisted SPPS on a 0.1 mmol scale utilizing standard 9-fluorenylmethoxycarbonyl (Fmoc)-based N_{α}-deprotection chemistry. Peptides were prepared from Rink-amide resin (0.53 mmol/g), Fmoc-Lys Wang resin (0.6 mmol/g), Fmoc-Asp Wang resin (0.31 mmol/g), Fmoc-Gly Wang resin (0.32 mmol/g), and 3-chlorotrityl resin (1.6 mmol/g). Each amino acid was coupled in 5.25 molar excess based on resin loading. Deprotection of the Fmoc group was performed one of two ways: (1) treating resin with 20% piperidine in DMF for 65 s at 90 °C; (2) treating resin with 5% piperazine, 0.1 M HOBt in DMF for 10 min at ambient temperature. Coupling reactions were undertaken in one of two ways: (1) for 2 min at 90 °C with a mixture of Fmoc-amino acid (0.53 mmol), *N,N′*-Diisopropylcarbodiimide (DIC, 0.51 mmol), and Ethyl cyano(hydroxyimino)acetate (Oxyma, 0.51 mmol) in DMF; for 10 min at 50 °C with a mixture of Fmoc-amino acid (0.53 mmol), *N,N′*-Diisopropylcarbodiimide (DIC, 0.51 mmol), and Ethyl cyano(hydroxyimino)acetate (Oxyma, 0.51 mmol) in DMF.

Synthesis of p53(376-393)S376C

p53(376-393)S376C was synthesized from Fmoc-Asp Wang resin (320 mg resin, 0.1 mmol) by the general method (1) described above, with double coupling of H380. The crude peptide was purified by RP-HPLC (C18 preparative, 0-50% B, 60 min; performed by Priscilla Kwong, Chatterjee lab).

*Synthesis of CG*ITLGMDELYK*

CG*ITLGMDELYK (* indicates **3** incorporation) was synthesized from Fmoc-Lys Wang resin (170 mg resin, 0.1 mmol) by the general method (1) described above for amino acids ITLGMDELY. Incorporation of **3** was achieved by acylation with bromoacetic acid (0.112 g, 0.8 mmol) and DIC (124 µl, 0.8 mmol) mixed by bubbling with N_2 for 30 min, followed by washing with DCM, both steps were then repeated a further time. Substitution of the terminal alkyl bromide was accomplished with **3** (315 mg, 1.0 mmol) and DIEA (400 µl, 2.3 mmol) in 4 mL 1:1 DMSO:DCM for 48 hrs under argon. The resin was washed with DMF, then DCM to remove **3**, and Boc-Cys(Trt)-OH (371 mg, 0.8 mmol) was coupled with HCTU (327 mg, 0.79 mmol) and DIEA (279 µl, 1.6 mmol) in DMF. Cleavage of the peptide with Reagent K (82.5:5:5:5:2.5 TFA:water:phenol:thioanisole:1,2-ethanedithiol; 1 ml/100 mg resin) afforded 115 mg crude, which was purified by RP-HPLC (C18 preparative, 22-50% B, 60 min; 0.8 mg, 0.49 µmol, 0.5%). Calculated for $C_{70}H_{114}N_{16}O_{22}S_3$ *m/z* $[M+H]^+$ 1627.8 Da, observed 1628.5 Da.

*Synthesis of CSGG**

CSGG* (* indicates **5** incorporation) employed Rink amide resin (188 mg, 0.1 mmol) was swelled in 1:1 DCM:DMF for 30', then drained and deprotected with 20% piperidine/DMF at ambient temperature (5 min x 5). Resin was drained, washed, and bromoacetylated with bromoacetic acid (112 mg, 0.8 mmol) and DIC (124 µl, 0.8 mmol) twice. Resin was washed, and substitution of the terminal alkyl bromide was accomplished with **5** (80 mg, 0.22 mmol) in DMSO (500 µl) with DIEA (30 µl, 0.17 mmol) for 48 hrs. Resin was washed, then incubated with Fmoc-Gly-OH (297 mg, 1 mmol), HATU (372 mg, 0.98 mmol) and DIEA (174 µl, 1 mmol) overnight. Resin was washed, dried, and deprotected as described above. Fmoc-Ser(OtBu)-OH (154 mg, 0.4 mmol) was coupled with DIC (63 µl, 0.4 mmol) and Oxyma (57 mg, 0.4 mmol) for 12 hrs. Resin was washed and dried, then Boc-Cys(Trt)-OH (371 mg, 0.8 mmol) was coupled with HATU (300 mg, 0.79 mmol) and DIEA (278 µl, 1.6 mmol) for 1 hr. Resin was washed, dried and cleaved with reagent

K. The ether precipitated crude was purified by RP-HPLC (C18 semi-preparative, 0-73% B, 30'; 3 mg, 5%). Calculated for $C_{24}H_{41}N_8O_8S_3$ *m/z* $[M+H]^+$ 666.2 Da, observed 666.5 Da.

*Synthesis of p53(376-393)S376C/K381ac/K382ac/K386ac/G389**

p53(376-393)S376C/K381ac/K382ac/K386ac/G389* (*indicates **5** incorporation) was prepared from Fmoc-Asp Wang resin (0.5 meq/g; 200 mg, 0.1 mmol). All deprotections were conducted with 5% piperazine, 0.1 M HOBt at ambient temperature due to the C-terminal Asp. The first three amino acids (Ser392, Asp391, Pro390) were each coupled employing 0.5 mmol Fmoc-AA in 0.1 M HCTU, 0.1 M HOBt (5 ml) (performed by Dr. Paul Lawrence, Chatterjee Lab). Following deprotection of Pro390, the resin was bromacetylated with bromoacetic acid (112 mg, 0.8 mmol) and DIC (126 µl, 0.8 mmol) twice (30'/acylation). Substitution of the terminal alkyl bromide with **5** (326 mg, 0.9 mmol) in DMSO (900 µl) with DIEA (17 µl, 0.1 mmol) occurred over 41 hrs at ambient temperature. Following draining and washing Fmoc-Glu(OtBu)-OH (425 mg, 1 mmol) was coupled overnight with HATU (372 mg, 0.98 mmol) and DIEA (174 µl, 1 mmol) in DMF. The remainder of the sequence (376-387) was performed on a Liberty Blue peptide synthesizer employing general conditions (2) described above. Peptide was cleaved with Reagent K, and ether precipitated, and purified by RP-HPLC (C18 Preparative, 10-50% B, 60 min; 12.1 mg, 5%). Calculated for $C_{107}H_{171}N_{29}O_{35}S_4$ *m/z* $[M+H]^+$ 2552.94 Da, observed 2553.65 ± 0.51 Da.

*Synthesis of CQ*TGG*

Fmoc-Gly Wang resin (313 mg, 0.1 mmol) was deprotected, and the first two amino acids coupled on a Liberty Blue peptide synthesizer employing general conditions (1) described above. Following Fmoc deprotection of Thr91, Fmoc-Glu(OAll)-OH (163.7 mg, 0.4 mmol) with DIC (62 µl, 0.4 mmol) and Oxyma (57 mg, 0.4 mmol) overnight. The resin was washed, and the allyl ester deprotected with $Pd(PPh_3)_4$ (0.0462 g, 40 nmol) and phenylsilane (50 µl, 1.2 µmol) in DCM (3 ml) for 2 hrs. The resin was drained and washed extensively, then the side chain was activated with

HATU (304 mg, 0.8 mmol) and DIEA (200 µl, 1.1 mmol) in DMF (2 ml) for one minute, after which O-2-propynyl hydroxylamine (40 µl, 0.6 mmol) was added in DMF (2 ml) and mixed by bubbling with nitrogen (40 min). Fmoc deprotection was as in general procedure (1), and Boc-Cys(Trt)-OH (371 mg, 0.8 mmol) was coupled with HCTU (325 mg, 0.79 mmol) and DIEA (278 µl, 1.6 mmol) for 1 hr.

Synthesis of fully protected Azidoacetyl-DYKDDDDK

2-chlorotrityl resin (100 mg, 0.15 mmol) was swelled in DCM for 30 min, then drained. Fmoc-Lys(boc)-OH (187 mg, 0.4 mmol) was solubilized in DCM (8 ml) with collidine (300 µl, 2.25 mmol) and added to the 2-chlorotrityl resin. The resin is shaken overnight at ambient temperature, then washed thrice with DCM. The resin is capped with 18/2/1 DCM/MeOH/DIEA (9.5 ml) for 60 min, then washed thrice with DCM. The remaining canonical amino acids are coupled using the Liberty Blue peptide synthesizer with coupling and deprotection conditions for peptides without an aminooxy amide. Azidoacetic acid is coupled manually with azidoacetic acid (30 µl, 0.4 mmol), DIC (62 µl, 0.4 mmol), and Oxyma (57 mg, 0.4 mmol) for 2 hrs. Cleavage of the protected peptide is accomplished nutating for 60 min with 1:5 1,1,1,3,3,3-hexalfuoroisopropanol:DCM (10 ml/1 g resin), filtering away resin, and concentrating *in vacuo*; alternatively 1:2:7 acetic acid:2,2,2-trifluoroethanol:DCM (10 ml/1 g resin) may be used, in which case the combined organics are extracted twice with water, then concentrated *in vacuo*. The cleavage mixture is filtered away from resin, concentrated *in vacuo*, and used without further purification (132 mg, 0.083 mmol crude).

3.4.8 Expressed protein ligation

Ligation of p53(1-375)-MESNa α-thioester and p53(376-393)S376C

Full length p53 S376C was synthesized by ligation of p53(1-375)-MESNa and p53(376-393)S376C peptide (0.05 mM thioester, 2 mM peptide, 8 M Urea, 1 M HEPES, 1 mM EDTA, 50 mM MPAA, pH 7.5, 25 ˚C, under nitrogen, dark). Timepoints were taken (3 µL rxn, 12 µL 2x dye, 2 µL βME) and checked by SDS PAGE (10% gel). After 12 h the reaction was treated for 5' with 10 mM TCEP and HPLC purified (C18 Analytical, 30-70% B, 30' gradient). Product peaks (17-19' elution time) were checked by ESI-MS on an LTQ Orbitrap (Thermo), and all peaks were found to contain both hydrolysis product and ligation product.

*Ligation of SUMO3(2-93)C47S-MESNa α-thioester and CSGG**

SUMO3(2-93)C47S-MESNa α-thioester (2 mg, 0.19 µmol) and CSGG* (0.32 mg, 0.48 µmol) are dissolved in ligation buffer (6 M Gn, 10 mM TCEP, 1 mM EDTA, 100 mM Na_2HPO_4, pH 7.5) and nutated for 12 hrs under argon at 25 ˚C. The reaction was reduced with fresh TCEP (10 mM) and purified by RP-HPLC (C18 analytical, 30-40% B, 45'; 0.3 mg, 14%).

*Ligation of SUMO3(2-88)C47S-MESNa α-thioester and CQ*TGG*

SUMO3(2-88)C47S-MESNa α-thioester (1 mM) was reacted with CQ*TGG peptide (**8** or **9**, 2 mM) in ligation buffer (6 M Gn HCl, 150 mM NaCl, 100 mM Na_2HPO_4, 50 mM mercaptophenylacetic acid, 10 mM EDTA, 10 mM TCEP, pH 7.5) for 12 hr at 25 ˚C. After 12 hrs the reaction was brought to pH 3 and extracted thrice with diethyl ether to remove residual MPAA. The reaction is brought to pH 7.5, 50 mM TCEP with 1.5 M Tris, pH 8.8, then reduced for 15 min. The ligation product is alkylated by bringing to 50 mM iodoacetamide for 30 min, then quenched with DTT (250 mM) for 30 min. The products may be purified affinity chromatography or RP-HPLC (**10**: 800 µL streptavidin resin; 2.6 mg at 0.28 µmol scale, 85%; **11**: C18 analytical, 30-45% B, 30'; 1.0 mg at 0.4 µmol scale, 23%)

3.4.9 N-O bond cleavage

N-O bond cleavage with aromatic thiols

Peptide or protein is dissolved (0.1 mM) in a solution of 200 mM aromatic thiol, 100 mM Na_2HPO_4, pH 7.3. The reaction vessel is selected such that the ratio of headspace to reaction volume is 15:1. Reactions are nutated for 24 to 48 hours under ambient atmosphere at 25 ˚C. The completed reaction is brought to pH 2-3 with phosphoric acid, and extracted thrice (3 x 2 rxn volumes) with diethyl ether. The acidified reaction is neutralized and reduced with TCEP (50 mM), then purified by RP-HPLC.

N-O bond cleavage with activated Zn

Peptide or protein is dissolved (1 mg/ml; **11**: 0.8 mg) in degassed 6 M guanidine (pH 3). Activated Zn (20 mg/ 0.1 mg protein or peptide) is added, and the reaction is placed under inert atmosphere. The mixture is sonicated to suspend the Zn, then the reaction is transferred to a nutator in a 37 ˚C incubator, and mixed for 24 hrs. The Zn is pelleted by centrifugation (1 min, 13.3 k xg), the supernatant is removed, and the Zn is washed with one volume of 6 M guanidine, sonicated, pelleted, and the supernatant again removed. The combined supernatants are purified by RP-HPLC when volume is less than 300 µl, or dialyzed against water, then lyophilized and purified by RP-HPLC (**11**: C18 Analytical, 20-55% B, 30'; 0.4 mg, 56%).

3.4.10 Copper catalyzed Azide-Alkyne Cylcoaddition

Resin-bound Fmoc-CQ(N-(O-2-propyne))TGG peptide (50 µmol) is swelled with a minimal amound of dry DCM. A solution of Cu(I)Cl (25 mg, 0.25 mmol) and ascorbic acid (44 mg, 0.28 mmol) in N_2 sparged 20% piperidine/DMF (1 ml) is applied to the resin, and the slurry is nutated for 5 min under argon. The azide (0.1 mmol) is taken up in N_2 sparged 20% piperidine/DMF (500 µl) and transferred to the resin container, then the azide vial is washed with N_2 sparged 20% piperidine/DMF (500 µl), which is further transferred to the resin container. The reaction is nutated

for 24 hrs under argon. The resin is drained and washed extensively with 20% piperidine/DMF supplemented with 20 mM ascorbic acid, then DMF, then DCM, then MeOH, the vacuum dried. Resin is cleaved with reagent K and purified by RP-HPLC (**8**: C18 semi-preparative, 5-40% B, 45'; 15.8 mg, 33%; **9**: C18 semi-preparative, 0-20% B, 45'; 2.8 mg, reduced scale to 33 μmol resin, 5%).

3.5 Supplemental characterization

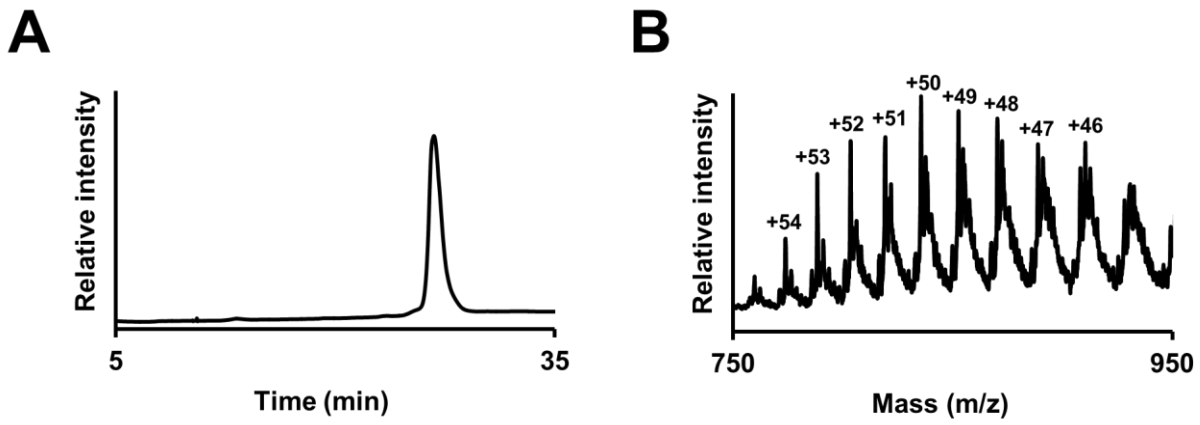


Figure 3.5.1. p53(1-375)-MESNa α-thioester. (**A**) C18 Analytical RP-HPLC chromatogram of purified p53(1-375)-MESNa α-thioester, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified p53(1-375)-MESNa α-thioester; observed mass 41737.63 ± 4.55 Da (calc'd 41731.03 Da); all reported masses are isotopically averaged.

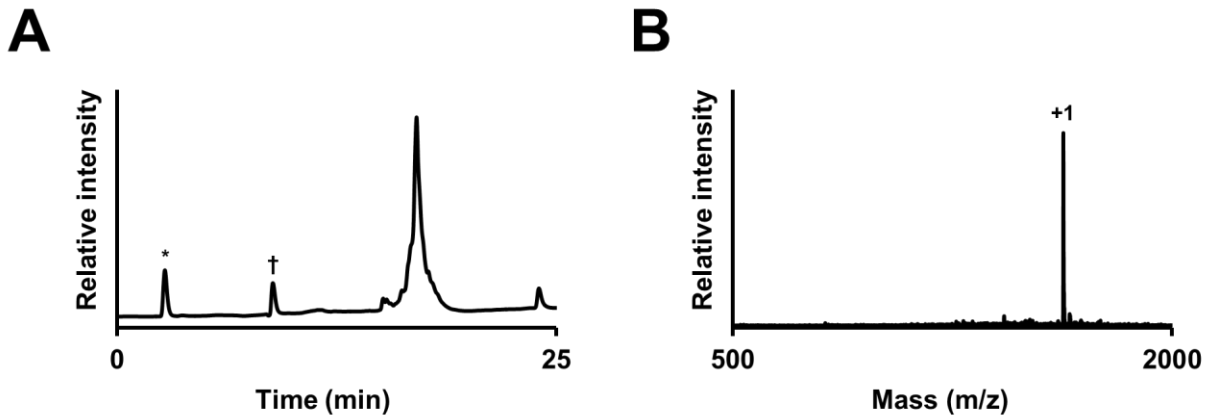


Figure 3.5.2. Purification of **4**. (**A**) C18 Analytical RP-HPLC chromatogram of purified **4**, 20 min. 0-73% B gradient; (*) injection peack; (†) TCEP. (**B**) ESI-MS of purified **4**; observed mass 1628.5 Da (calc'd mass 1627.7 Da); all reported masses are isotopically averaged.

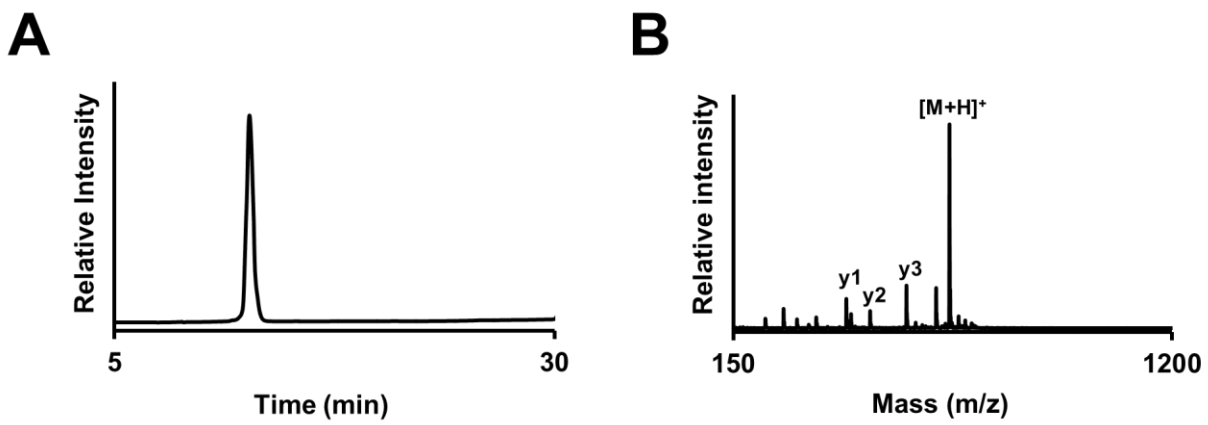


Figure 3.5.3. Purification of **6**. (**A**) C18 Analytical RP-HPLC chromatogram of purified **6**, 30 min. 0-73% B gradient; (*) injection peack; (†) TCEP. (**B**) ESI-MS of purified **6**; observed mass 667.5 Da (calc'd mass 666.8 Da); all reported masses are isotopically averaged.

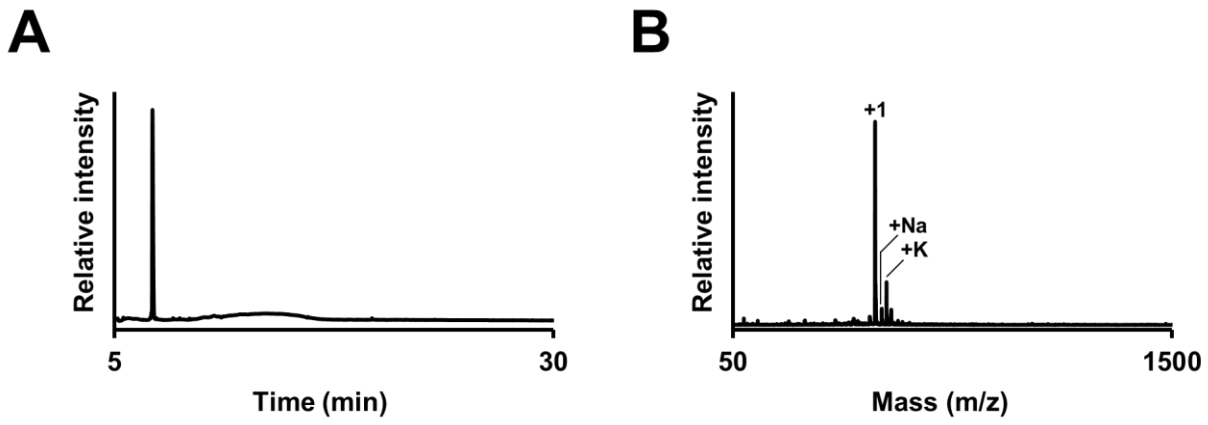


Figure 3.5.4. Purification of CQ(alkyne)TGG peptide. (**A**) C18 Analytical RP-HPLC chromatogram of purified CQ(alkyne)TGG, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified CQ(alkyne)TGG; observed mass 518.57 Da (calc'd mass 518.18 Da); all reported masses are isotopically averaged.

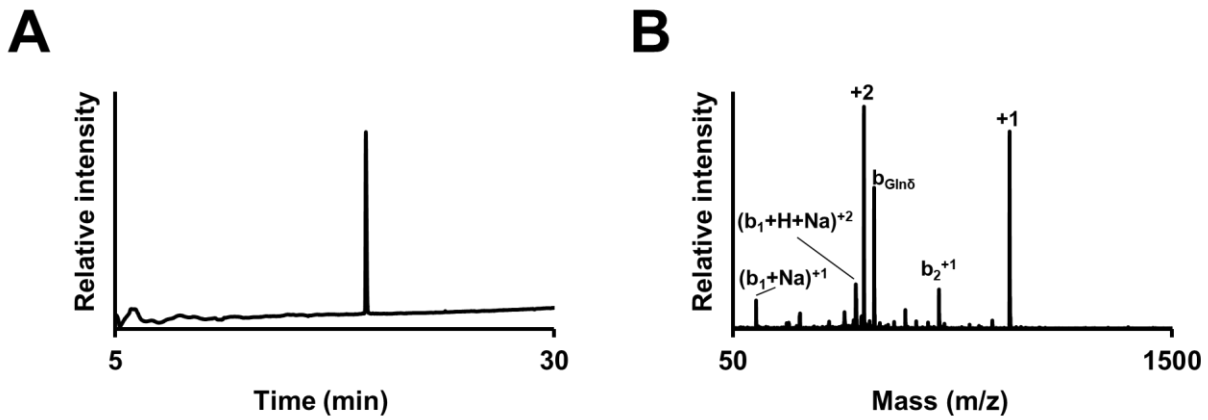


Figure 3.5.5. Purification of CQ(biotin)TGG peptide. (**A**) C18 Analytical RP-HPLC chromatogram of purified CQ(biotin)TGG, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified CQ(biotin)TGG; observed mass 963.18 ± 0.56 Da (calc'd mass 963.09 Da); all reported masses are isotopically averaged.

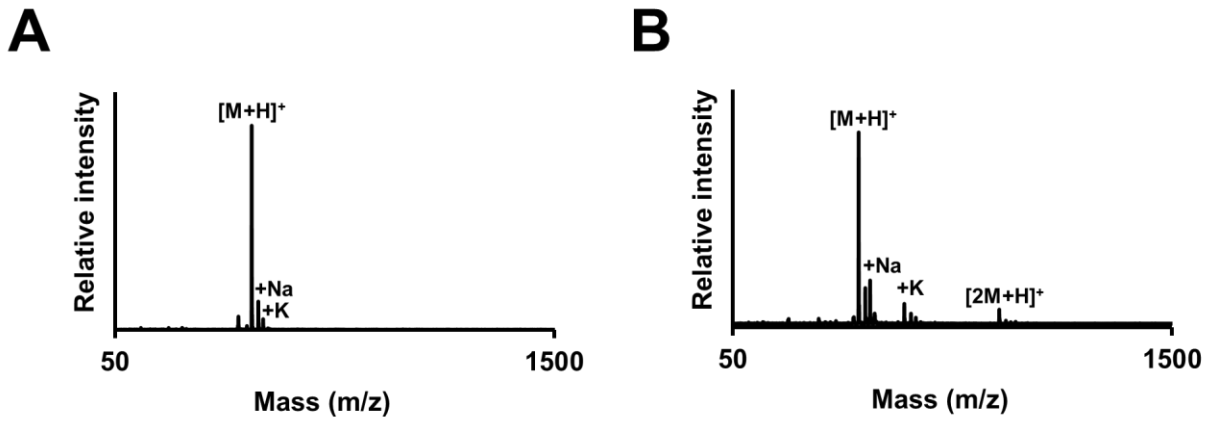


Figure 3.5.6. Products of **8** N-O bond cleavage with activated Zn. (**A**) ESI-MS of purified Biotin-PEG3-triazole; observed mass 500.5 Da (calc'd 500.6 Da). (**B**) ESI-MS of purified CQTGG; observed mass 464.5 Da (calc'd mass 464.5 Da); all reported masses are isotopically averaged.

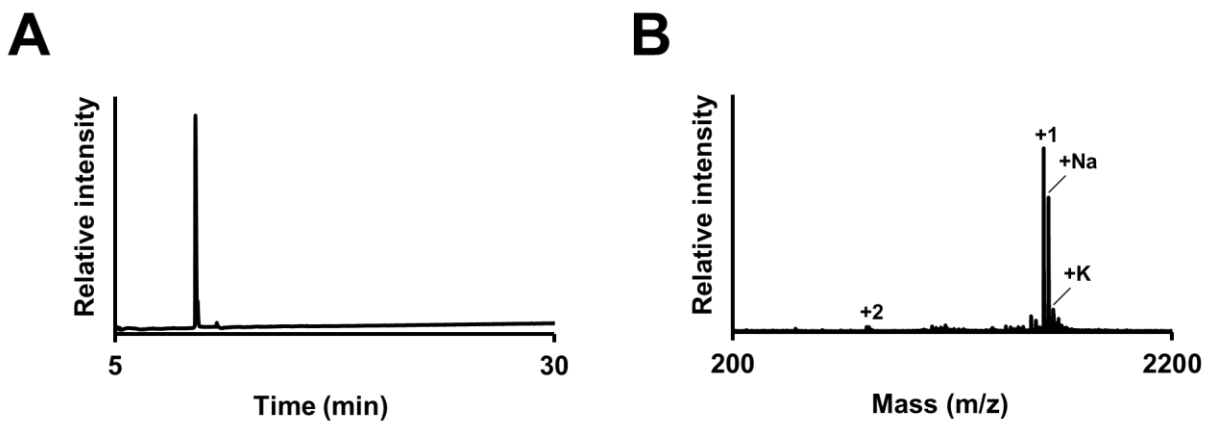


Figure 3.5.7. Purification of CQ(flag)TGG peptide. (**A**) C18 Analytical RP-HPLC chromatogram of purified CQ(flag)TGG, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified CQ(flag)TGG; observed mass 1614.65 ± 0.60 Da (calc'd mass 1614.57 Da); all reported masses are isotopically averaged.

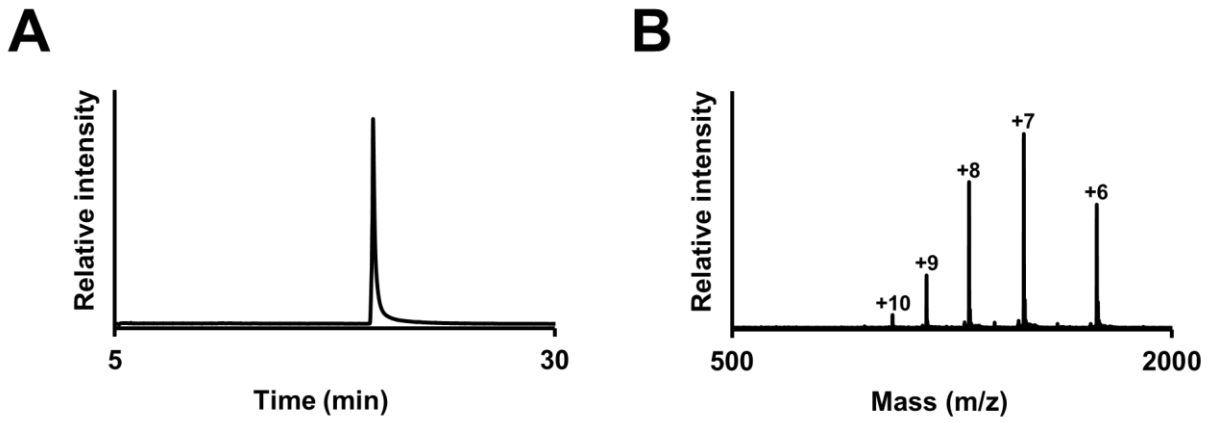


Figure 3.5.8. Purification of SUMO3 C47S Δ5 MES thioester. (**A**) C18 Analytical RP-HPLC chromatogram of purified SUMO3 C47S Δ5 thioester, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified SUMO3 C47S Δ5 thioester; observed mass 10462.46 ± 2.22 Da (calc'd mass 10461.7 Da); asterisks (*) indicate; observed mass Da (calc'd mass Da); all reported masses are isotopically averaged.

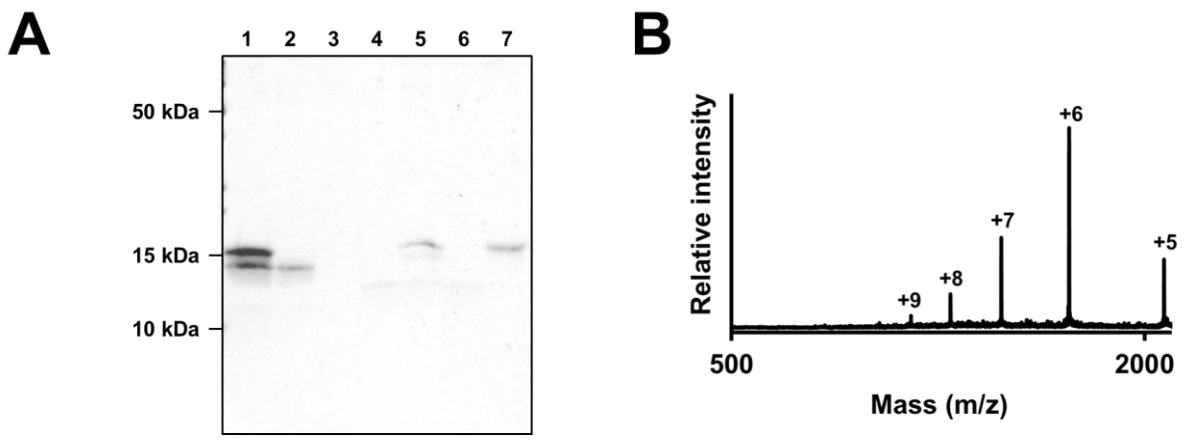


Figure 3.5.9. Isolation of SUMO3 C47S Q88C Q89(biotin) and N-O bond reduction. (**A**) enrichment of SUMO3 C47S Q88C Q89(biotin) on streptavidin resin: (1) input; (2) flow through; (3) final wash; (4) empty resin; (5) bound resin; (6) eluted resin; (7) elution. (**B**) LC-MS detection of SUMO3 C47S Q88C arising from Zn N-O bond cleavage of SUMO3 C47S Q88C Q89(biotin). Observed mass 10353.71 ± 2.44 Da (calc'd mass 10352.59 Da).

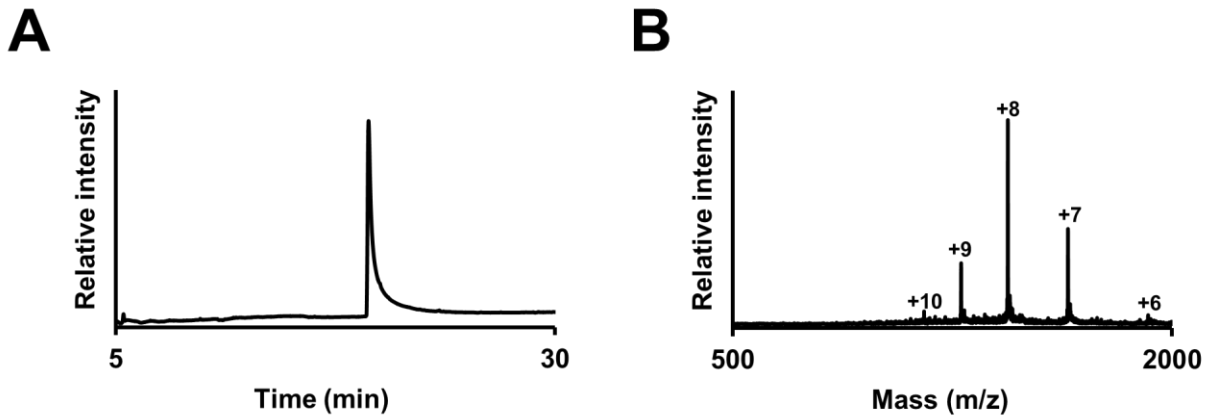


Figure 3.5.10. Purification of SUMO3 C47S Q88C Q89(flag). (**A**) C18 Analytical RP-HPLC chromatogram of purified SUMO3 C47S Q88C Q89(flag), 30 min. 0-73% B gradient. (**B**) ESI-MS of purified SUMO3 C47S Q88C Q89(flag); observed mass 11504.79 ± 2.96 Da (calc'd mass 11503.68 Da).

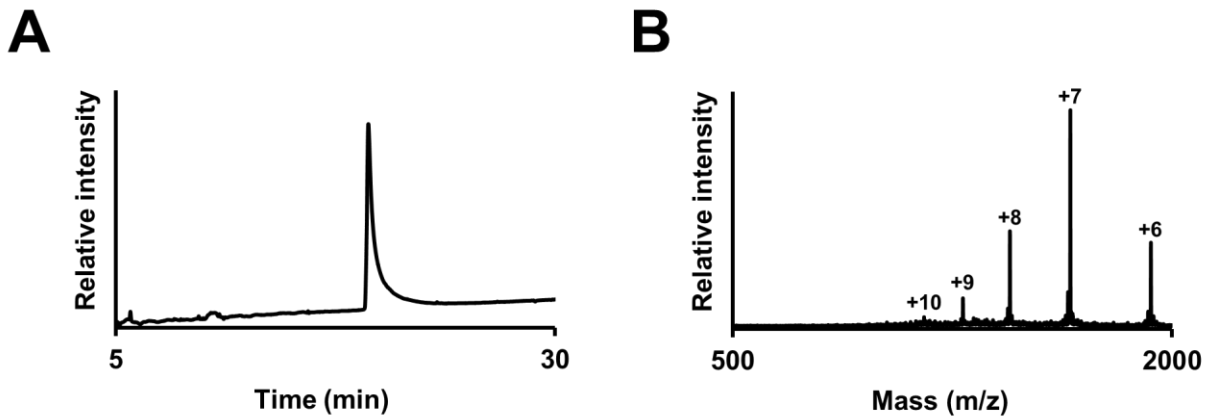


Figure 3.5.11. Purification of SUMO3 C47S Q88C(acetamide) Q89(flag). (**A**) C18 Analytical RP-HPLC chromatogram of purified SUMO3 C47S Q88C(acetamide) Q89(flag), 30 min. 0-73% B gradient. (**B**) ESI-MS of purified SUMO3 C47S Q88C(acetamide) Q89(flag); observed mass 11560.72 ± 1.92 Da (calc'd mass 11560.74 Da).

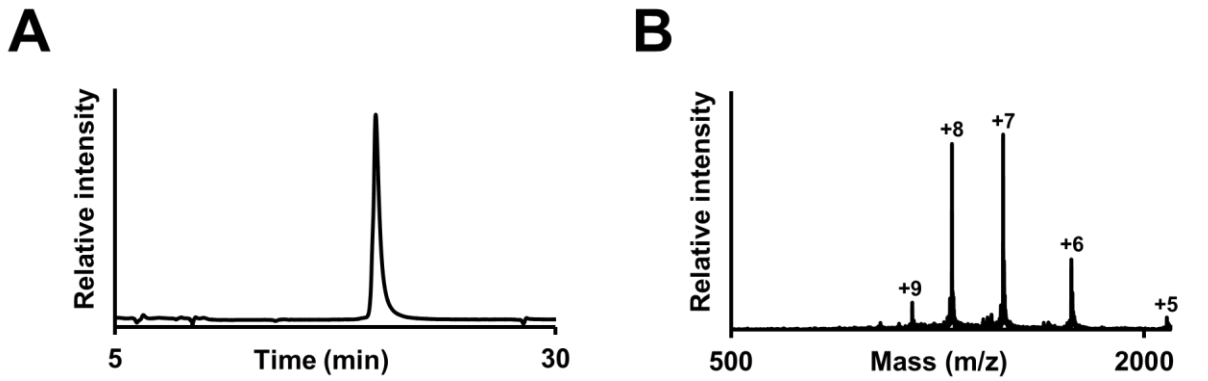


Figure 3.5.12. Purification of SUMO3 C47S Q88C(acetamide). (**A**) C18 Analytical RP-HPLC chromatogram of purified SUMO3 Q88C, 30 min. 0-73% B gradient. (**B**) ESI-MS of purified SUMO3 C47S Q88C(acetamide); observed mass 10353.71 ± 2.44 Da (calc'd mass 10352.59 Da).

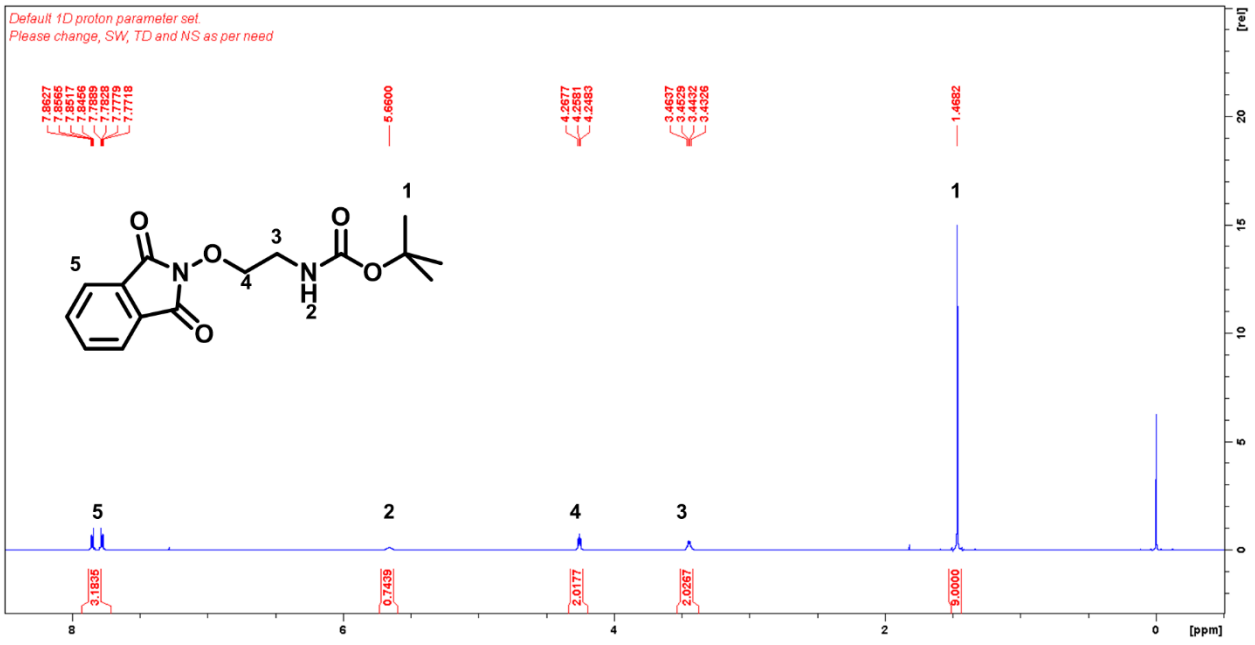


Figure 3.5.13. ^{1}H NMR of 3a.

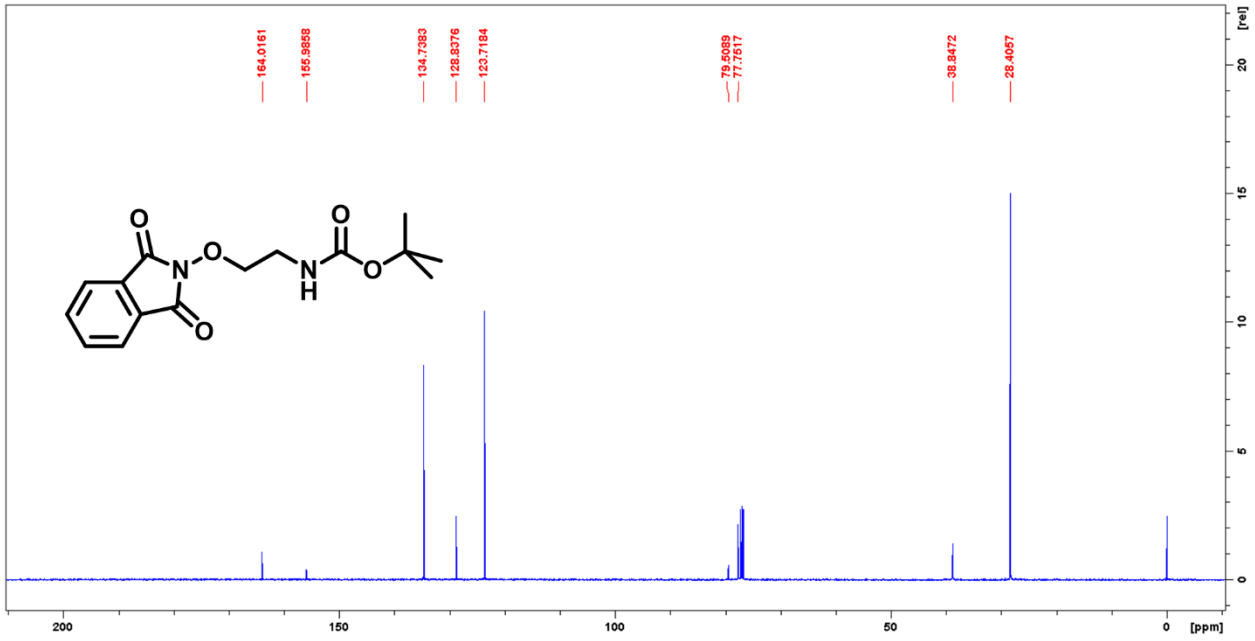


Figure 3.5.14. ^{13}C NMR of 3a.

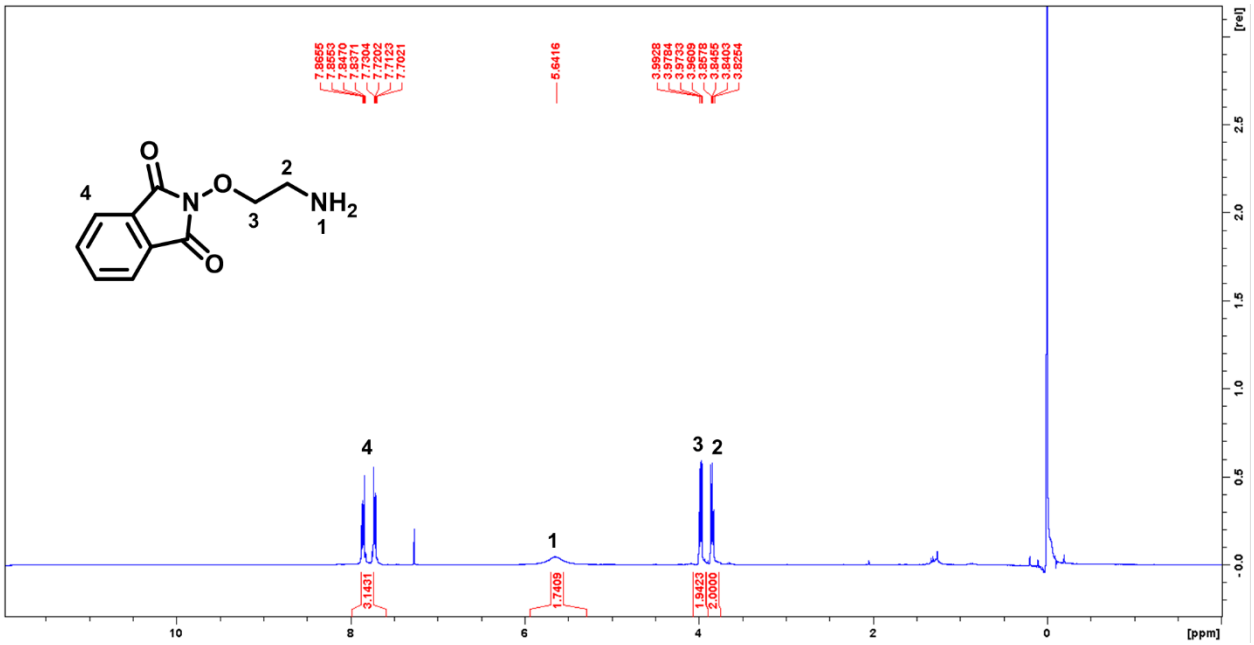


Figure 3.5.15. ^{1}H NMR of 3b.

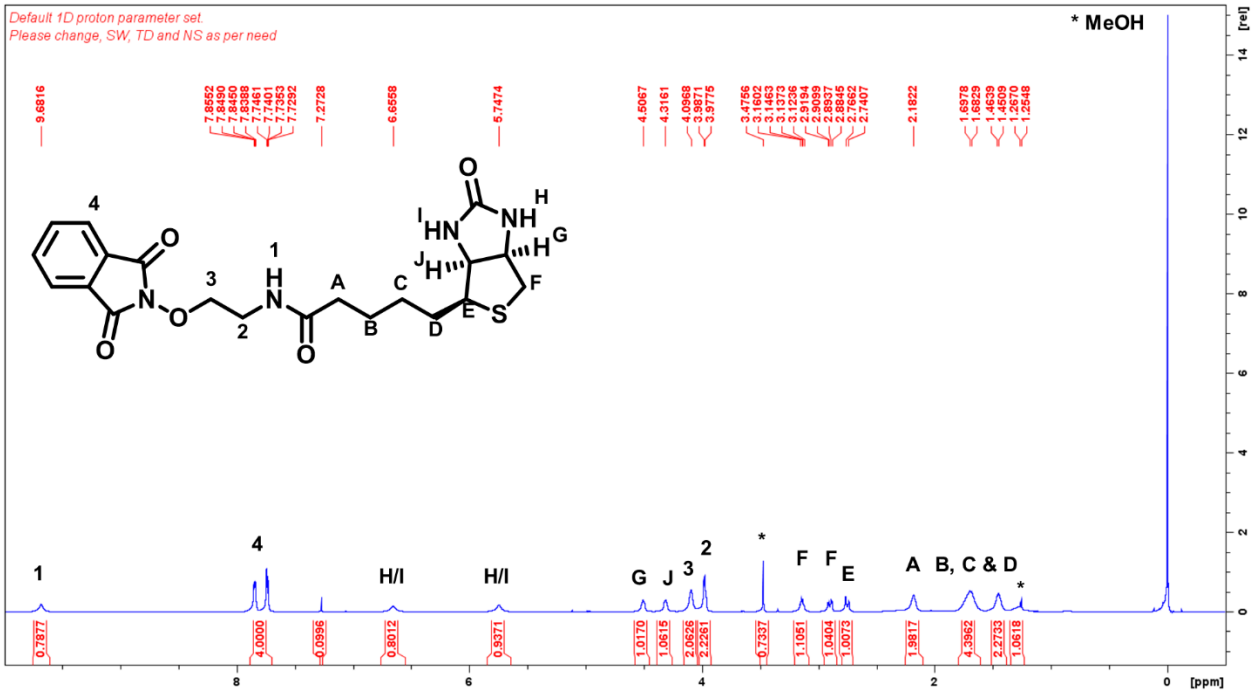


Figure 3.5.16. ^{1}H NMR of 3c.

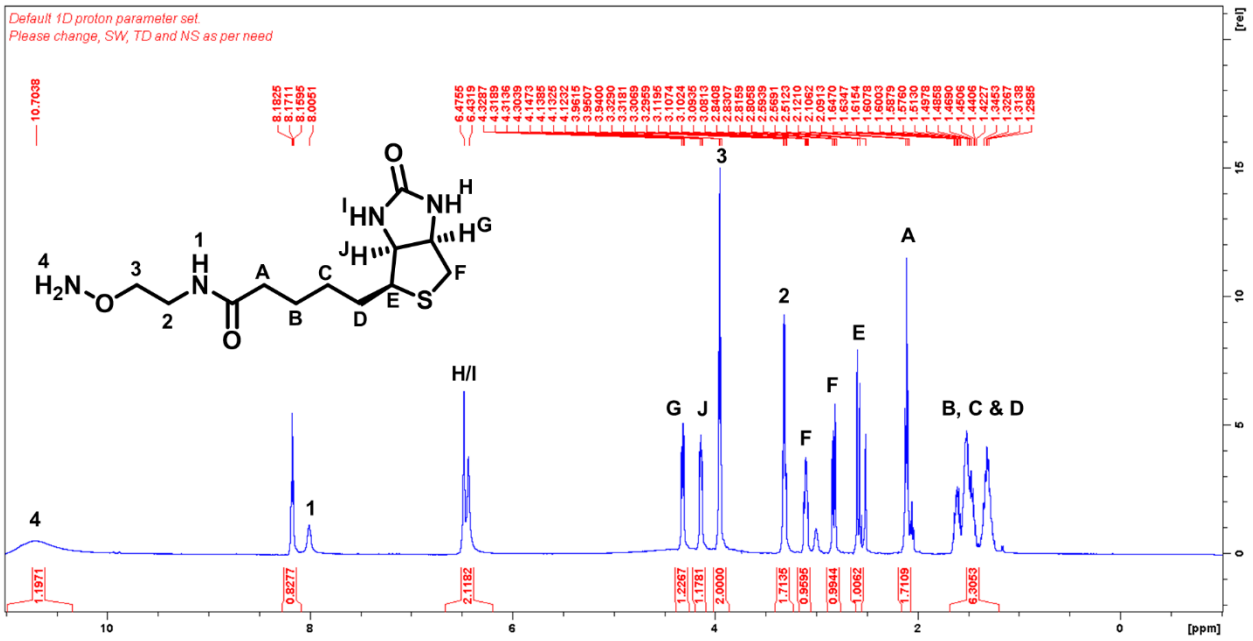


Figure 3.5.17. ^{1}H NMR of 3.

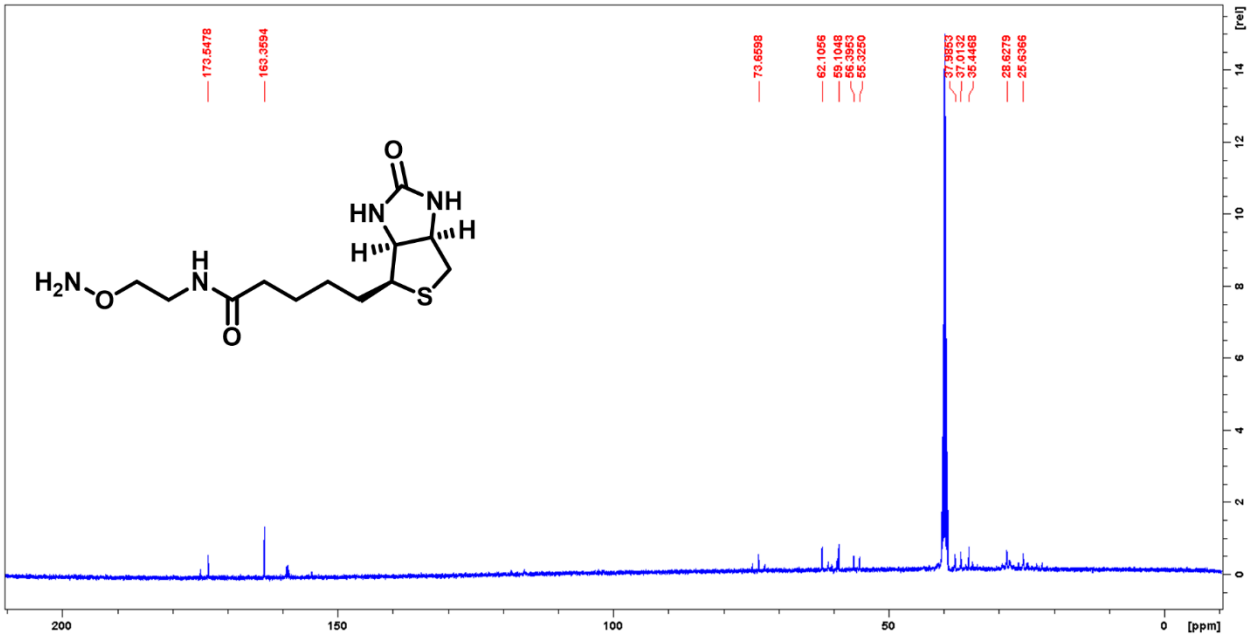


Figure 3.5.18. ^{13}C NMR of 3.

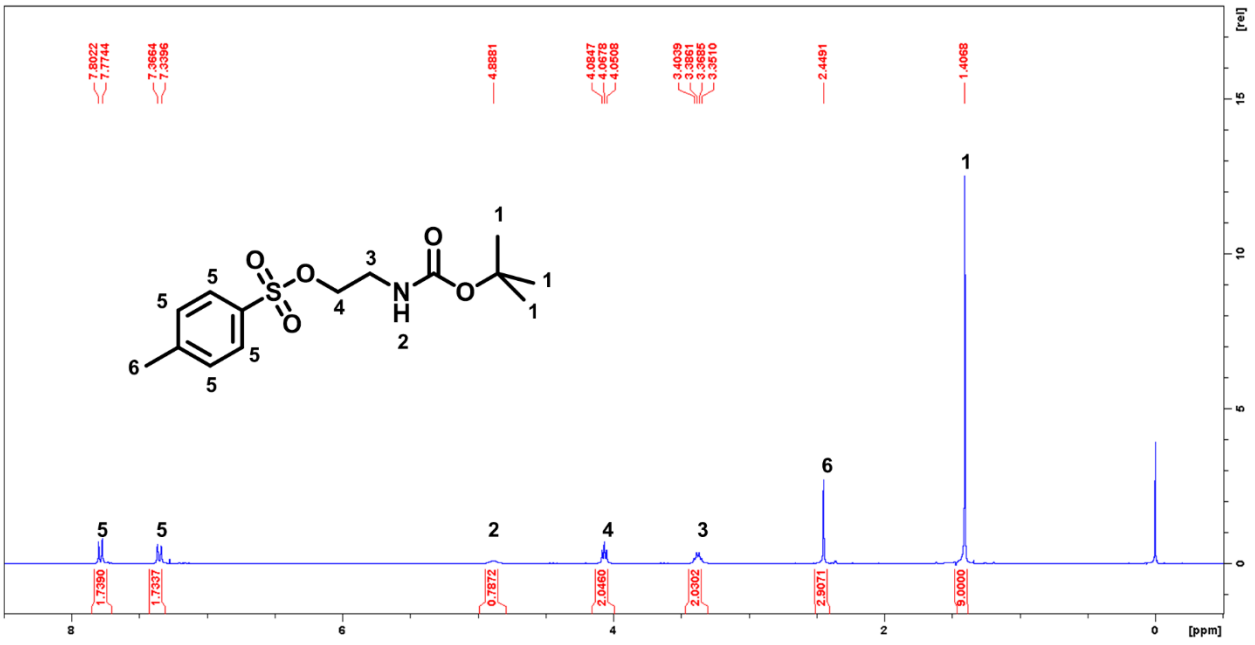


Figure 3.5.19. ^{1}H NMR of 5a.

Figure 3.5.20. ^{1}H NMR of 5b.

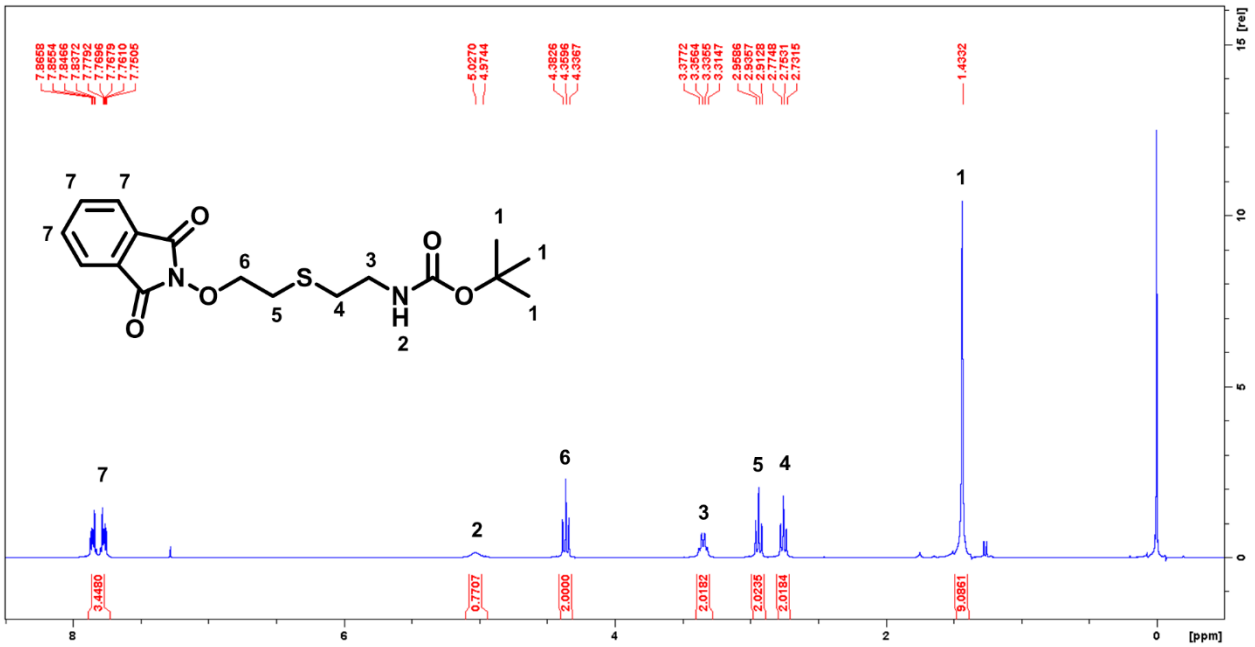


Figure 3.5.21. ^{1}H NMR of 5c.

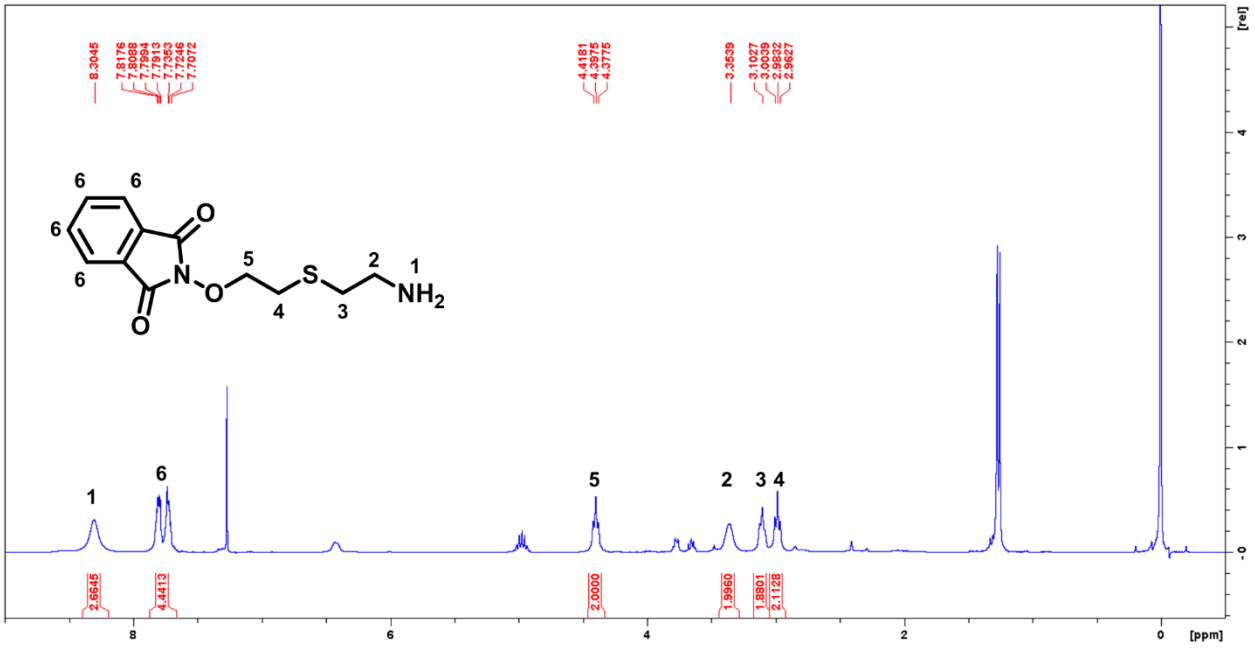


Figure 3.5.22. ^{1}H NMR of 5d.

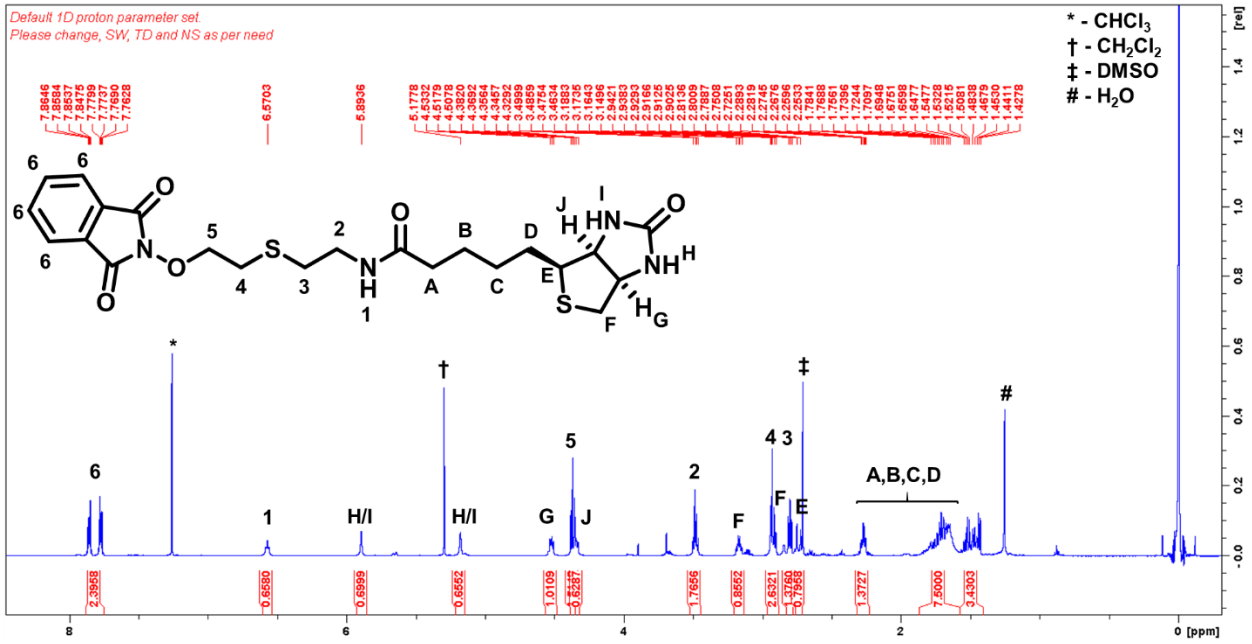


Figure 3.5.23. ^{1}H NMR of 5f.

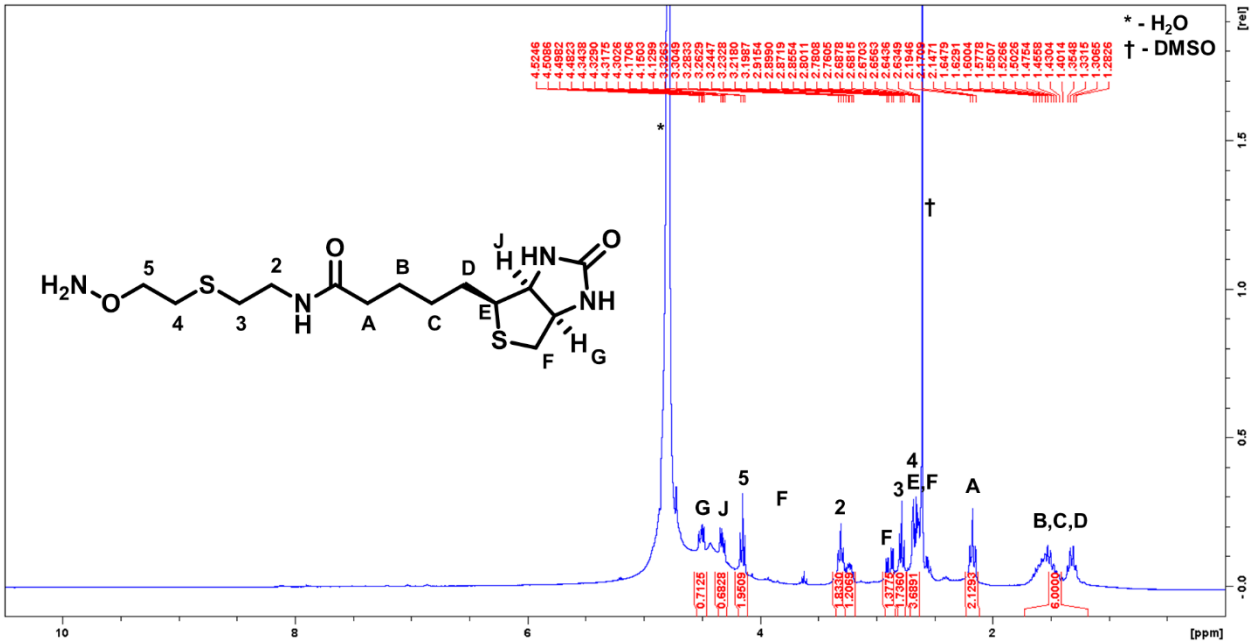


Figure 3.5.24. ^{1}H NMR of 5.

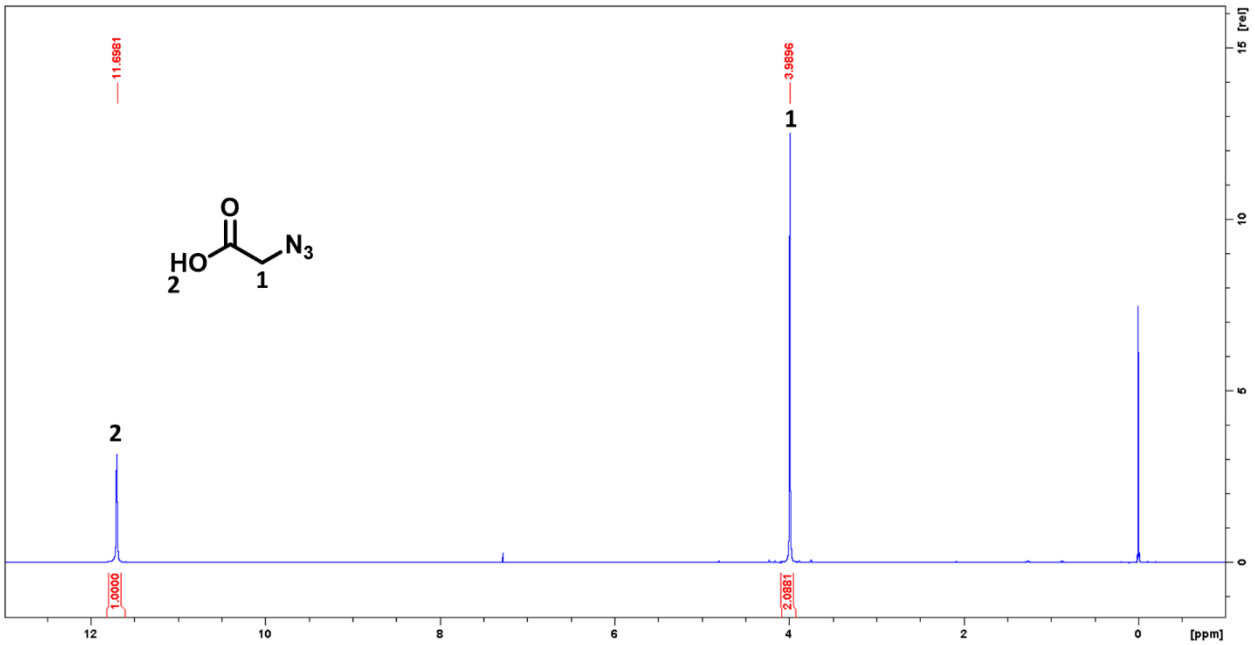


Figure 3.5.25. ^{1}H NMR of azidoacetic acid.

3.6 References

1. Walsh, C. T., Garneau-tsodikova, S. & Gatto, G. J. Protein Chemistry Protein Posttranslational Modifications : The Chemistry of Proteome Diversifications. *Angew. Chemie Int. Ed.* **44,** 7342–7372 (2005).

2. Merrifield, R. B. Solid Phase Peptide Synthesis. I . The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149–2154 (1963).

3. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science (80-.).* **266,** 775–778 (1994).

4. Bang, D. & Kent, S. B. H. His 6 tag-assisted chemical protein synthesis. *Proc. Natl. Acad. Sci.* **102,** 5014–5019 (2005).

5. Schwagerus, S., Reimann, O., Despres, C., Smet-nocca, C. & Hackenberger, C. P. R. Semi-synthesis of a tag-free O -GlcNAcylated tau protein by sequential chemoselective ligation §. *J. Pept. Sci.* **22,** 327–333 (2016).

6. Bogyo, M., Verhelst, S. H. L. & Fonovic, M. A Mild Chemically Cleavable Linker System for Functional Proteomic Applications. *Angew. Chemie Int. Ed.* **46,** 1284–1286 (2007).

7. Weller, C. E. *et al.* Aromatic thiol-mediated cleavage of N–O bonds enables chemical ubiquitylation of folded proteins. *Nat. Commun.* **7,** 12979 (2016).

8. Weller, C. E. *et al.* Aromatic thiol-mediated cleavage of N-O bonds enables chemical ubiquitylation of folded proteins. *Nat. Commun.* **7,** 1–10 (2016).

9. Meek, D. W. & Anderson, C. W. Posttranslational modification of p53: cooperative integrators of function. *Cold Spring Harb. Perspect. Biol.* **1,** a000950 (2009).

10. Weller, C. E., Huang, W. & Chatterjee, C. Facile synthesis of native and protease-resistant ubiquitylated peptides. *ChemBioChem* **15,** 1263–1267 (2014).

11. Malins, L. R., Mitchell, N. J., Mcgowan, S. & Payne, R. J. Angewandte Oxidative Deselenization of Selenocysteine : Applications for Programmed Ligation at Serine. *Angew. Chemie Int. Ed.* **54,** 12716–12721 (2015).

12. Metanis, N., Keinan, E. & Dawson, P. E. Traceless Ligation of Cysteine Peptides Using Selective. *Angew. Chemie Int. Ed.* **49,** 7049–7053 (2010).

13. Susumu, K. *et al.* Enhancing the Stability and Biological Functionalities of Quantum Dots via Compact Multifunctional Ligands. *J. Am. Chem. Soc.* **129,** 13987–13996 (2007).

14. Singudas, R., Adusumalli, S. R., Joshi, P. N. & Rai, V. A phthalimidation protocol that follows protein defined parameters. *Chem. Commun.* **51,** 473–476 (2014).

15. Canne, L. E., Winston, R. L. & Kent, S. B. H. Synthesis of a Versatile Purification Handle for Use with Boc Chemistry Solid Phase Peptide Synthesis. *Tetrahedron Lett.* **38,** 3361–3364 (1997).

16. Shelton, P. M. M., Weller, C. E. & Chatterjee, C. A Facile N - Mercaptoethoxyglycinamide (MEGA) Linker Approach to Peptide Thioesteri fi cation and Cyclization. *J. Am. Chem. Soc.* **139,** 1–4 (2017).

17. Dunkelmann, D. L. *et al.* Amide-forming chemical ligation via O -acyl hydroxamic acids.

Proc. Natl. Acad. Sci. 1–6 (2018). doi:10.1073/pnas.1718356115

18. Li, S. *et al.* Extent of the Oxidative Side Reactions to Peptides and Proteins During the CuAAC Reaction. *Bioconjug. Chem.* **27,** 2315–2322 (2016).

19. Cecere, G., Ko, C. M., Alleva, J. L. & Macmillan, D. W. C. Enantioselective Direct α - Amination of Aldehydes via a Photoredox Mechanism: A Strategy for Asymmetric Amine Fragment Coupling. *J. Am. Chem. Soc.* **135,** 10–13 (2013).

20. Davies, J., Svejstrup, T. D., Reina, D. F., Sheikh, N. S. & Leonori, D. Visible-Light-Mediated Synthesis of Amidyl Radicals: Transition- Metal-Free Hydroamination and N - Arylation Reactions. *J. Am. Chem. Soc.* **138,** 8092–8095 (2016).

21. Bloom, S. *et al.* Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **10,** 205–211 (2017).

22. Ding, H. *et al.* Solution structure of human SUMO-3 C47S and its binding surface for Ubc9. *Biochemistry* **44,** 2790–9 (2005).