Development of auxiliary-mediated protein semisynthesis methods toward the study of histone SUMOylation

Caroline E. Weller

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Dr. Champak Chatterjee, Chair

Dr. Michael H. Gelb

Dr. Dustin J. Maly

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Development of auxiliary-mediated protein semisynthesis methods toward the study of histone SUMOylation

Caroline E. Weller

Chair of the Supervisory Committee:

Dr. Champak Chatterjee

Chemistry

Eukaryotic DNA is packaged into chromatin, which consists of a fundamental repeating unit, the nucleosome, and its associated proteins. Nucleosomes are made up of histone protein octamers, containing two each of histones H2A, H2B, H3, and H4, around which ~147 bp of DNA is wound. The tails of histones are disordered and extend beyond the DNA where they may be accessed by *reader*, *writer*, and *eraser* proteins. Post-translational modifications (PTMs) on histone tails are crucial to regulating chromatin structure and its interacting proteins, which collectively dictate the transcriptional state of the overlying DNA. The dysfunction of DNA-templated processes underlies a variety of human pathologies, including neurodegenerative and autoimmune diseases, and multiple human cancers. Importantly, these conditions are all associated with the misregulation of various histone PTMs. We focused on the poorly-understood effects of histone modification

with the small ubiquitin-like modifier (SUMO). The SUMOylation of histone H4 is conserved between yeast and humans and is implicated in transcriptional silencing and DNA repair. However, a mechanistic understanding of how H4 SUMOylation leads to these outcomes has remained elusive for well over a decade, and insight into its precise roles in gene regulation may lead to the identification of new therapeutic targets.

We were interested in the dynamic interplay, or *cross-talk*, between SUMOylated H4 (suH4) and other histone PTMs, potentially mediated by transcriptional repressive complexes. To generate suH4 for *in vitro* assays, we utilized protein semisynthesis. Toward this goal, we developed chemical methodology for facile, site-specific native SUMOylation of protein substrates, including suH4, based on the small molecule native chemical ligation auxiliary (2-aminooxy)ethanethiol. We further utilized this auxiliary to accomplish protein SUMOylation under non-denaturing conditions, which greatly expanded the scope of substrates accessible to this methodology. Over the course of developing non-denaturing ligation, we discovered that the thiol additive, 4-mercaptophenylacetic acid (MPAA), promoted *N-O* bond reductive cleavage and auxiliary removal without the low pH and harsh denaturants previously required. Extensive characterization of the reaction revealed that it is promoted by the spontaneous formation of MPAA thiyl radicals in solution. Finally, these methodologies enabled us to generate multi-milligram quantities of homogenous suH4 to test the hypothesis that SUMO recruits the transcriptional repressor histone deacetylase 1 (HDAC1) to chromatin. *In vitro* deacetylation assays with SUMOylated mononucleosomes (MN) demonstrated that a repressive complex of HDAC1 and the corepressor protein CoREST is preferentially recruited to SUMOylated MN, which enhances their deacetylation. Thus, we have identified a novel cross-talk relationship between histone

SUMOylation and acetylation, which represents one possible pathway by which histone SUMOylation can mediate transcriptional repression.

Acknowledgements

I would like to thank everyone that helped make this PhD possible. Champak of course welcomed me into his group, and provided abundant guidance and advice that I learned so much from. He challenged me to think about my research in new ways, which proved invaluable as I began to steer the course of my project and design meaningful experiments. His boundless enthusiasm and passion for science are infectious, and his support has helped get me through tough experimental stretches – when nothing seemed to work – and to succeed and grow as a scientist.

My labmates have been one of the best parts about graduate school. We have had many interesting discussions, related to science and a variety of other topics, and they always seem to look at things from unique points of view that I hadn't considered. They have helped me to progress my project, from making proteins or reagents, giving me advice, pushing me to do just one more experiment before I go home, or making a countdown calendar until my thesis due date. They helped make lab work fun, and made me feel like I was part of a team more than any job I've had before.

I also want to thank my parents and family for their endless support. Both of my parents have been through the process of graduate school, and even though a chemistry PhD is the "family business" they supported my choice in college when I initially wanted to be a journalist. After coming to my senses and beginning to study science, they helped me figure out that I should get my PhD, and were able to understand just what I was going through while I was working towards it. I will be forever grateful for all their help.

Finally, I need to thank my long-time boyfriend Chen. We've been together for 10 years, and for at least 5 of those years we were apart while I was in Seattle working toward my degree and he was on the East coast. It was difficult at times, but he was patient and stuck by me with constant love and support, and I'm so excited to continue our life together in California.

Table of Contents

List of Figures, Schemes, and Tables

Chapter 1

Chapter 2

Chapter 1

Introduction to chromatin, post-translational modification, and chemical protein modification

1.1 Epigenetics

Each eukaryotic cell contains the complete set of information necessary for maintaining basic cellular functions such as replication, metabolism, and signaling. Crucial to these processes is the synthesis of specific proteins at the appropriate times and in the correct quantities. These data are contained within a vast nucleoprotein complex known as chromatin. Packaging DNA into chromatin allows the approximately six billion base pairs (bp) of a diploid human cell – several linear meters – to fit within a ~ 6 μm-diameter nucleus. It has long been understood that the sequence of nucleotides in DNA encodes the amino acid chains that make up proteins.[1] However, early geneticists quickly found that the relationship between genotype and phenotype is flexible, and derives from a slew of complex intracellular processes.[2] Studies in the 1960s provided evidence that transcriptional activity is influenced by acetylation and methylation of the histone proteins of chromatin,[3] and in 1975, Holliday proposed that a chemical change to the DNA molecule itself, cytosine methylation, could be responsible for patterns of gene expression that persist after mitosis.[4] This early work laid the foundation for the field of epigenetics: the study of heritable changes in gene expression, caused by developmental processes or environmental signals, that are independent of the underlying DNA sequence. Many further decades of research have established an intimate connection between gene expression and small chemical changes in the DNA and proteins that make up chromatin.

1.2 Chromatin structure and organization

The first order of chromatin organization is the formation of the nucleosome. It was initially visualized by electron microscopy as a 'beads on a string' structure under low salt conditions,[5,6] and established as the basic repeating unit of chromatin.[7,8] The nucleosome core particle (NCP) consists of an octameric protein

complex of two copies of each core histone, H2A, H2B, H3, and H4, around which 147 bp of DNA is wrapped ~1.65 superhelical turns (**Figure 1.1a**). There are typically 10-90 bp of linker DNA between each NCP.[9] The octamer itself is composed of an H3-H4 tetramer, bound by two H2A-H2B dimers and mediated by H2B-H4 interactions.[10] Each histone has a histone fold motif (3 α-helices connected by a pair of loops) that is responsible for the heterodimeric interactions in the octamer core.[11] DNA affinity for the octamer varies greatly and is sequence pattern dependent. Flexibility of the DNA is essential, and intrinsically stiff, homopolymeric stretches are strongly disfavored.[12] The most favorable sequences contain AT/TA and GC/CG dimers that occur every ~10 bp, but directly out-of-phase with each other. With this pattern, the AT/TA and GC/CG pairs face toward (major groove out) or away (major groove in) from the octamer core, respectively, for every helical turn of the DNA (every ~10 bp). This corresponds to the ability of those pairs to adopt a favorable bend, and, in the case of the AT/TA pairs, the ability to adopt a narrow minor groove for maximal salt bridge formation between the phosphates of the DNA backbone and surface residues of the histone octamer.[10,12]

In addition to the well-defined NCP structure, each histone contains a significant unstructured region at its N-terminus of approximately 20-40 amino acids, and H2A and H2B have substantial unstructured C-terminal regions as well. These histone 'tails' extend past the DNA surface and are solvent-exposed,[13] leaving them prone to enzymatic modification (discussed in section 1.3). The high density of basic amino acids in the tails gives them a positive charge under physiological conditions, and the tails engage in stabilizing intra- and inter-nucleosomal interactions, which are important for the formation of higher-order chromatin structures.[14] The H4 tail is particularly important, and stabilizes inter-nucleosomal interactions by binding an acidic patch at the surface of the H2A/H2B interface on an adjacent nucleosome.[10,15]

The next order of compaction above the nucleosome occurs when linear arrays of nucleosomes are folded into three-dimensional structures known as 30 nm fibers. These structures form from nucleosomal arrays *in vitro* with the addition of divalent cations, and arrange in a zig-zag pattern, which was observed in an X-ray crystal structure of a tetranucleosome.[16] In this structure, the nucleosomes 1 and 3, and 2 and 4, are in

contact and form a ‘two-start’ stack of nucleosomes that can form twists or coils in higher-order structures.[17] The short linker length of DNA in the tetranucleosome structure, however, is atypical in eukaryotes. A solenoid model of the 30 nm fiber has also been proposed,[18] but the zig-zag model is often favored because it is conducive to the stabilizing interaction between H2A and the H4 tail.[19] The tetranucleosome structure is informative, yet it omits the linker histone H1, which binds to the nucleosome at the entry and exit point of DNA, and interacts with ~20 bp of linker DNA.[20] H1 is crucial for stabilizing the compacted state of chromatin. Depletion of H1 causes aberrations in the formation of mitotic chromosomes, which exhibit the highest order of chromatin compaction.[21]

When not engaged in cell division, higher-order structures form when 30 nm fibers form loops and contacts defined by proteins such as CTCF and cohesin.[22] These topologically associating domains (TADs) are 800 kb on average, and appear to be conserved in both position and size. Interestingly, TAD boundaries are regions of high transcriptional activity, or euchromatin, and tend to be located on the interior of the nucleus.[22] Transcriptionally inactive chromatin, or heterochromatin, is enriched at the nuclear periphery.[23] In this way, chromatin appears to be organized into nuclear neighborhoods of similar transcriptional activity. The establishment and maintenance of euchromatin and heterochromatin is key to the regulation of transcription. In general, euchromatin is loosely compacted and accessible to the transcriptional machinery, while heterochromatin is compacted into higher-order structures. Both structures are heavily influenced by histone modification.

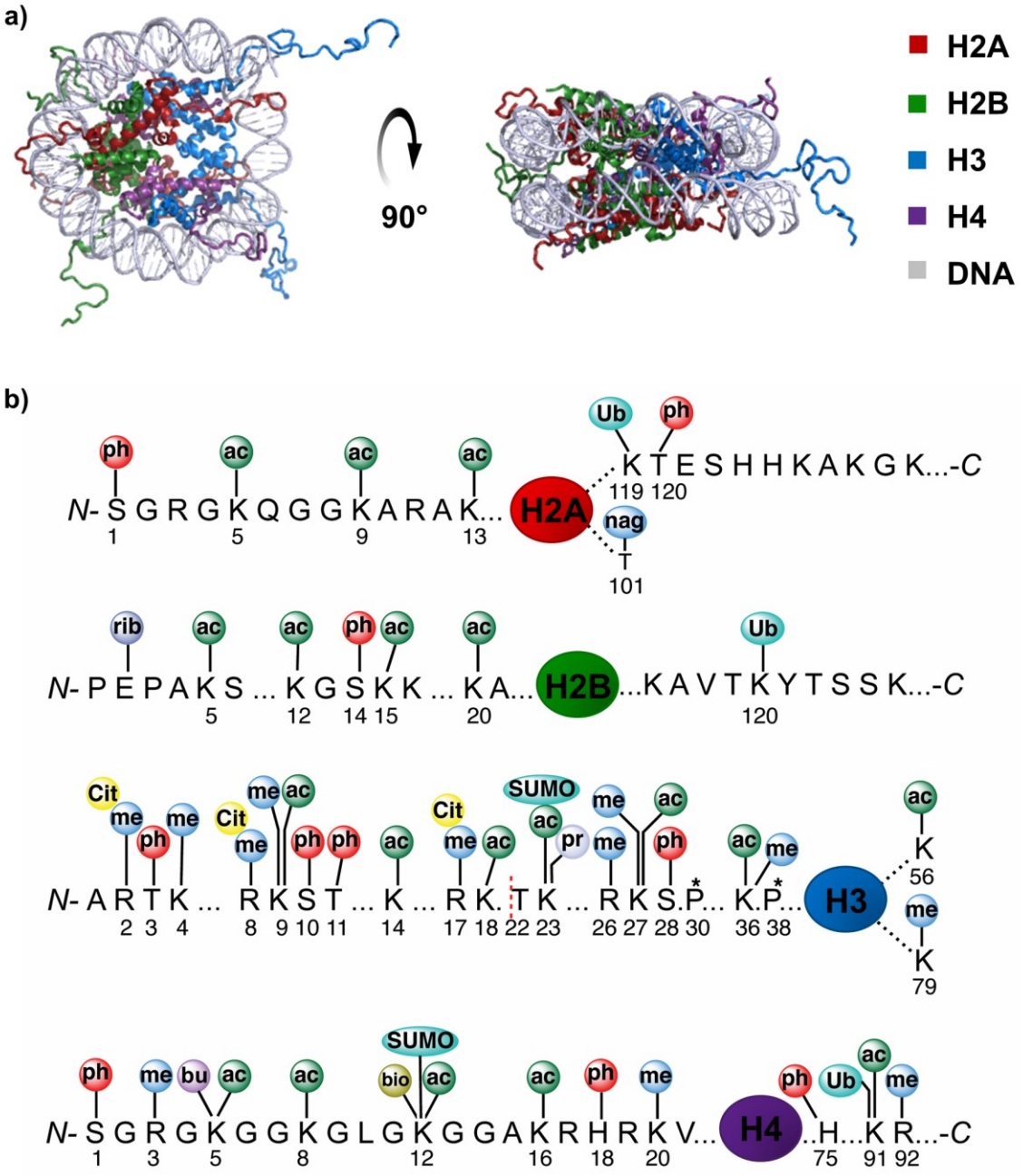


Figure 1.1. Mononucleosome structure. a) Structure of the mononucleosome at 1.9 Å resolution (PDB code 1KX5), with individual histones and 147 bp double-stranded DNA. The N-terminal tails extend beyond the DNA for post-translational modification. Red, H2A; Green, H2B; Blue, H3; Purple, H4; Gray, DNA. b) Schematic of histone tail modifications. Chemical groups: ac, acetyl; bio, biotinyl; bu, butyryl; cit, ciyrullyl; me, methyl; nag, *N*-acetylglucosaminyl; ph, phosphoryl; pr, propionyl; rib, ADP-ribosyl; SUMO, SUMOyl; ub, ubiquitinyl. Figure adapted from ref. 50.

1.3 Post-translational modification of histones

Chemical modification of chromatin can occur on its DNA or protein constituents and can dramatically affect its structure and function – a key epigenetic route of transcriptional control. The primary modification on DNA is methylation, which occurs at CpG dinucleotide sequences in the form of 5-methylcytosine, and has a repressive effect on gene expression.[24] In contrast, the solvent-exposed N-terminal tails of histones are decorated with a wide variety of chemical groups known as post-translational modifications (PTMs) that can have disparate effects on transcription (**Figure 1.1b**).

There are ~24,000 protein-coding genes in the human genome, and further proteome variety is generated by alternative pre-mRNA splicing, which occurs in >90% of multi-exonic genes.[25] By far the greatest source of functional protein diversity comes from PTMs, and estimates of chemically unique species in the human proteome exceed 1,000,000.[26] In addition, the reversibility of PTMs (excepting arginine methylation) represents a mechanism for the rapid cellular signaling and stimuli response that is necessary for homeostasis. For example, phosphorylation and dephosphorylation signal cascades are well characterized intracellular pathways that transform extracellular information into enzymatic activity or changes in gene expression.

Advances in mass spectrometry techniques have enabled the identification of over 700 isoforms of the core histones in HeLa cells,[27] with various combinations of PTMs that include acetyl, methyl, and phosphoryl groups, and even modification by the ~10 kDa protein ubiquitin.[28] Histone PTMs can favor either repressed or active transcription. Early recognition of this effect was described in the mid-1960s by Allfrey and coworkers, who showed that global histone acetylation with acetic anhydride destabilized higher-order chromatin structures and promoted transcription.[3] Importantly, acetylation was later revealed to be non-random. This is exemplified by X-chromosome inactivation, in which one of two X-chromosomes in female mammalian cells is hypoacetylated and transcriptionally silent.[29] It was not until 1996 that the first enzymes responsible for depositing and removing acetylation were identified.[30] Histone acetyltransferases (HATs)

and histone deacetylases (HDACs) are now understood to be part of a broad group of histone-modifying enzymes known as *writers* and *erasers*.

A major focus of ongoing research concerns the cause-and-effect relationship between histone PTMs and gene expression outcomes. For a long time, it was thought that the primary means of transcriptional activation by acetylation was global histone charge neutralization. Indeed, the high density of positively-charged lysine and arginine residues in the histone N-terminal tails form stabilizing interactions with the DNA sugar-phosphate backbone, mainly in linker DNA regions. Acetylation of lysine neutralizes its positive charge and reduces this interaction, resulting in a less compact chromatin structure that allows access of transcription factors and machinery to the DNA.[31] However, research performed by Shogren-Knaack and coworkers showed that acetylation of H4K16 is sufficient for chromatin decompaction, by disrupting its interaction with the H2A/H2B acidic patch.[32] In this way, even minor changes to chromatin can have large effects on gene expression. Similarly, H3K56 is located on nucleosomes at the DNA entry/exit point. Acetylation at this position loosens the grip of the nucleosome on DNA and increases its accessibility, and is strongly correlated with active transcription.[33] These cases of histone PTMs that lead to a direct change in chromatin structure and function are examples of *cis* effects.

Alternatively, *trans* effects of histone PTMs are those that indirectly lead to changes in chromatin structure and function by recruiting effector proteins and enzymes. This requires specific *readers* with unique structural domains responsible for identifying the various histone PTMs. The first reader domain, the acetyllysine-binding bromodomain of the HAT p300/CBP-associated factor (PCAF), was documented in 1999.[34] This acetylation feed-forward mechanism leads to a high local concentration of acetyllysine, which favors transcription. Another example of a trans effect is the recognition of di- or tri-methylated H3K4 at the 5’ end of active genes by the chromodomain helicase DNA binding protein 1 (CHD1), which has ATP-dependent nucleosome remodeling activity and promotes nucleosome disassembly and transcription through chromatin.[35] Finally, some histone PTMs can exert trans effects by preventing the association of readers rather than recruiting them. For instance, H3T6 phosphorylation by protein kinase C beta I (PKCβI)

prevents repressive H3K4 demethylation by lysine-specific demethylase 1 (LSD1) during androgen receptor-dependent gene activation.[36]

Since discovery of the PCAF bromodomain many more binding modules have been identified. They include the plant homeodomain (PHD) and plextrin homology (PH) domains which also recognize acetyllysine; the chromodomain, Tudor, malignant brain tumor (MBT), and PHD domains, to name a few that recognize methyllysine; and the 14-3-3 proteins that recognize phosphoserine.[37] Interestingly, the PHD domain, which is primarily methyllysine binding, binds H3K14Ac in the first of two tandem PHD finger domains in double PHD fingers 3 (DPF3). These tandem PHD domains are packed together to form a single functional unit that recognizes PTMs along the H3 tail.[38] Combinations of reader domains within a single protein or complex are common, and allow preferential binding to specific combinations of PTMs. The discovery of trans effects and multivalent recognition of histone PTMs led to the proposal of a 'histone code hypothesis,' which argues that particular patterns of histone PTMs direct specific downstream biological outcomes.[39] While genomic and biochemical studies have provided ample evidence for *cross-talk* between highly-correlated modifications, this hypothesis is a great source of debate in the field. Another hypothesis suggests that histone marks represent a signaling network in which chromatin integrates diverse messages into a small number of biological outcomes with variable 'intensity' (e.g. degree of transcriptional activation).[40] Therefore, much work remains to be done to understand the complex interplay between histone marks and effectors that leads to changes in gene expression.

1.4 Histone post-translational modification and transcriptional regulation

Transcription in eukaryotes typically begins with transcription factor (TF) binding upstream of the core promoter region and transcription start site (TSS). This leads to recruitment of adapter complexes such as Spt-Ada-Gcn5 acetyltransferase (SAGA) and Mediator, which results in binding of general TFs to the promoter, and finally formation of the pre-initiation complex (PIC).[41] Transcription ultimately depends on the ability of RNA polymerase II (Pol II) to access DNA. Not surprisingly, nucleosomes impede both formation

of the PIC and elongation.[42,43] Nucleosomes also regulate access of TFs to DNA in regulatory elements such as promoters and enhancers, although promoters often achieve nucleosome-depleted regions with homopolymeric dA or dT tracts.[12] Pol II can only traverse a nucleosome if at least one H2A/H2B dimer is removed.[44] Hence it relies on chromatin remodeling complexes, such as Swi/Snf, chromatin structure remodeling (RSC), or facilitates chromatin transcription (FACT), to evict part or all of individual nucleosomes as transcription proceeds. Nucleosomes are then redeposited behind Pol II with the help of chaperones such as Nap1.[41]

Histone PTMs are crucial to transcription and to maintaining a euchromatic or heterochromatic state, and genome-wide ChIP-seq and RNA-seq studies have identified marks associated with active or repressed transcription. In general, H3/H4 acetylation, H3K4me2/3, H3K36me2/3, and H2BK120ub appear to be activating marks, and H3K9me3, H3K27me3, H2AK119ub, and H4K12 SUMOylation (acylation with the small ubiquitin-like modifier) are repressive.[41,45] These trends do have exceptions. For instance, ubiquitylated H2B is highly-activating within gene bodies, but inhibits Pol II recruitment when located in promoter regions.[46] Cross-talk between PTMs is especially important in the establishment of active or repressed transcription. In one example, H3S10 phosphorylation by RSK2 occurs on nucleosomes of a small subset of inactive immediate early genes in response to epidermal growth factor (EGF). This stimulates Gcn5 towards H3K14 acetylation and leads to selective, transient derepression.[47,48] However, this effect is not seen at all phosphorylated promoters, indicating that there are combinatorial and context-dependent effects at play. Further research into transcriptional changes arising from histone PTM trans effects is therefore needed to unravel these complex interactions. Isolating the fundamental components of these systems for experimentation *in vitro* greatly facilitates the characterization of cross-talk events. In my thesis I will describe the development of chemical tools for native, site-specific histone SUMOylation, and this mark's cross-talk with acetylation.

1.5 Chemical histone modification: small molecules

The past few decades have seen a rapid expanse in the variety of known histone PTMs. To date, over 150 unique positions have been identified, mainly in the unstructured N-terminal tail regions.[49] A drawback to cell-based assays is the presence of PTMs other than the mark of interest, that cannot be completely controlled and may influence results in unforeseen or unrecognized ways. Studies *in vitro* with chemically-defined substrates avoid these complications, and research into the functional outcomes of different PTM combinations has relied heavily on solid-phase peptide synthesis (SPPS). Short, synthetic histone N-terminal tail peptides have been used to determine the binding preference, and substrate specificity of readers, writers, and erasers in the context of various PTM combinations.[50] However, peptide substrates may not accurately reflect *in vivo* processes. Early work by Allfrey and coworkers found that HDACs purified from calf thymus could not deacetylate a short H4(15-21) tail peptide acetylated at Lys16.[51] However, later studies showed that a H4(1-37) peptide[52] or a H4(14-21) peptide acetylated at its N- and C-termini[53] were in fact substrates for the same HDAC. In another example, surface plasmon resonance was used to determine the binding constant between immobilized Sir3, a heterochromatin spreading factor in yeast, and H4 tail peptides acetylated in various combinations at Lys 5, 8, 12, and 16.[54] Binding *in vitro* showed an additive effect, and decreased upon each successive acetyl mark regardless of position. In contrast, *in vivo* studies showed a much greater reduction in silencing due to a Lys to Gln mutation at Lys16 than at any of the other positions. It is therefore useful to incorporate substrates into a physiologically relevant system, such as synthetic chromatin, to accurately study the interplay between cis and trans effects of histone PTMs. This requires the synthesis of full-length modified histones.

There are many different approaches to generating full-length histones with defined PTMs. The simplest method is genetic amino acid substitution to mimic modification. For instance, Gln resembles acetyllysine, Arg resembles constitutively unacetylated Lys, and Glu resembles phosphorylated Ser. These mimics have been used widely in yeast to screen for functionally significant sites of modification. They are also amenable to high-yielding purification from *E. coli*, which is advantageous for biochemical or biophysical techniques,

including crystallography.[55] However, these mimics cannot undergo native modification. Cysteine-based methods can generate better mimics that are labile to enzymatic removal. For these techniques histones are purified from *E. coli* with a Lys to Cys mutation at the desired site of modification. Because histones contain no other Cys residues, with the exception of C110 in H3 that can be mutated to Ala without consequence,[56] site selective reactivity is achieved. Shokat and coworkers successfully alkylated cysteines of histone H3 with mono- di- or tri-methylated electrophilic ethylamines, and demonstrated similar antibody recognition and substrate binding as the natural substrates (**Figure 1.2a**).[57] Alternatively, the thiol-ene 'click' reaction has been used to alkylate Cys with *N*-vinylacetamide to mimic acetyllysine. In this reaction, a thermal- or UV light-inducible radical initiator generates a Cys thiyl radical that adds across the alkene to yield the anti-Markovnikov thioether product (**Figure 1.2b**).[58] The authors showed that an H4K16Ac mimic in nucleosomal arrays inhibited compaction to an identical degree as native H4K16, and that deacetylation by SIRT2 was only mildly impaired. They also demonstrated identical antibody recognition. However, although replacement of the γ-methylene with a sulfide results in only ~0.28 Å increased length,[57] antibody recognition is highly substrate dependent. As described in Chapter 4, the failure of commercial H3K14Ac-specific antibodies to recognize the γ-sulfide mimic necessitated synthesis of the native substrate for deacetylation assays on nucleosomal substrates.

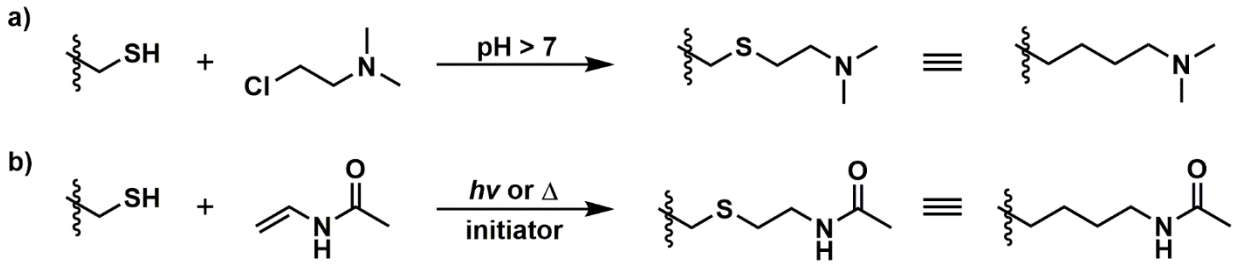


Figure 1.2. Cysteine-based analogs of modified lysine. a) Cys reaction with an electrophilic ethylamine, such as (2-chloroethyl)dimethylamine, yields a methyllysine mimic. b) Thiol-ene 'click' reaction between Cys and *N*-vinylacetamide yields an acetyllysine mimic.

Amber suppression is a powerful, alternative method that exploits ribosomal incorporation of mimic or native versions of modified amino acids into proteins, usually in *E. coli*. During translation, the 'empty' amber stop codon is recognized by a tRNA that is co-expressed with its cognate aminoacyl tRNA synthetase. The synthetase has been mutated by directed evolution to accept an unnatural amino acid, such as N^{ε}-acetyl-L-lysine,[59] or N^{ε}-*tert*-Butoxycarbonyl (Boc)-N^{ε}-methyl-L-lysine[60] or photo-caged N^{ε}-methyl-L-lysine,[61] which must bear a bulky group at the ε-amine to ensure discrimination from unmodified Lys by the synthetase. More flexibility comes from the incorporation of L-*Se*-phenylselenocysteine, which can be oxidized by H_2O_2 to the Michael-acceptor dehydroalanine and reacted with *N*-acetylated or *N*-methylated 2-aminoethanethiol.[62] This yields the γ-sulfide Lys as discussed above, but may result in undesireable side-effects such as Met oxidation. While the methods discussed in this section are quite useful, it is important to note that the accuracy of studies with PTM mimics must be determined on a case-by-case basis. Further, incorporation of more complex modifications or of multiple different marks into the same histone by these techniques is not trivial. In these cases, the ability to tracelessly join chemically modified peptide fragments is especially useful.

1.6 Chemical histone modification: native chemical ligation

The direct, total chemical synthesis of proteins by SPPS is limited to proteins of 50-100 residues due to attrition with each amino acid coupling step.[63] Native chemical ligation (NCL) circumvents this problem and was first reported by Kent and coworkers in 1994 (**Figure 1.3**).[64] This reaction forms a native peptide bond between two unprotected peptide fragments. The only requirements are a C-terminal α-thioester on one fragment and an N-terminal Cys residue on the other. First, the N-terminal Cys performs transthioesterification to form a peptide-peptide thioester. Next, a thermodynamically favorable *S*-to-*N* acyl shift, via a 5-membered ring intermediate, forms the native amide bond between the two fragments. The polypeptide fragments for ligation can be generated by SPPS, and incorporate a wide variety of chemical PTMs on the solid phase. Histone PTMs are clustered at the N- and C-termini, so modified histones can readily be synthesized from a short synthetic peptide and a much longer recombinant fragment in a

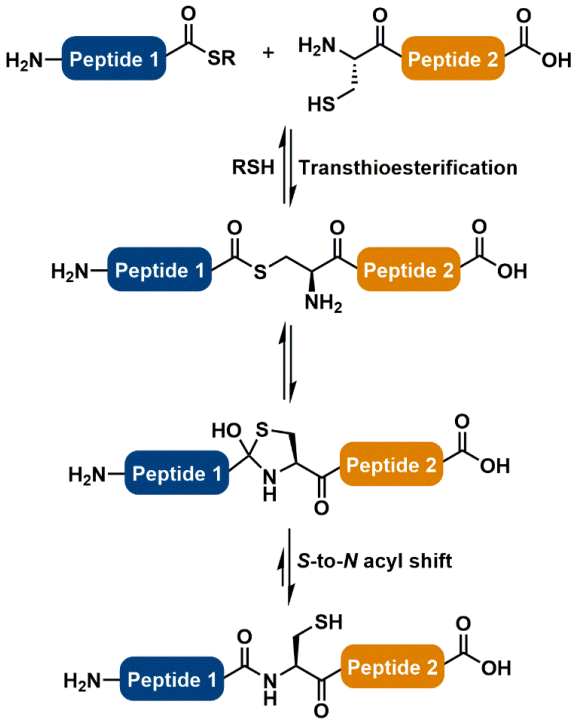


Figure 1.3. Native chemical ligation. Peptides containing a C-terminal α-thioester react with peptides containing a N-terminal Cys to form a thioester, followed by spontaneous rearrangement to the native amide linkage.

technique termed expressed protein ligation (EPL).[65] The approach to generating an N-terminal Cys or C-terminal α-thioester fragment is quite different depending on its origin.

Recombinant proteins with C-terminal α-thioesters can be effectively purified from *E. coli* as intein fusions. Inteins are single-turnover protein splicing enzymes. They naturally occur between extein fragments in polypeptide chains, and catalyze extein joining followed by self-removal. The first step of intein catalysis involves formation of a thioester at the C-terminus of the first extein and N-terminus of the intein. Protein chemists have taken advantage of this natural process by fusing their protein of interest to the N-terminus of an intein bearing an affinity tag. The addition of external thiol releases the protein of interest as a C-terminal α-thioester (**Figure 1.4**).[63] The intein historically used in EPL is the *M. xenopi* GyrA intein, which can take days to complete catalysis, although the recent discovery and characterization of 'ultrafast' inteins has reduced that time to a matter of hours.[66] On the other hand, many approaches exist for synthesizing

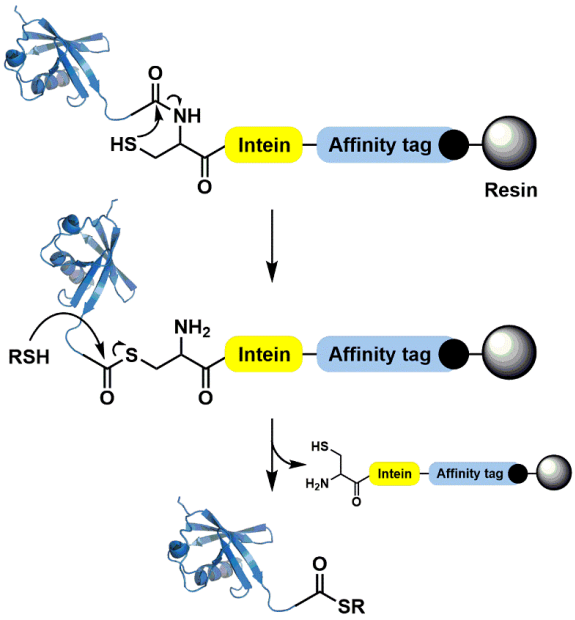


Figure 1.4. Purification of protein C-terminal α-thioester from intein fusion. Protein of interest, shown here as Ub (PDB code 1UBQ), is expressed and purified from *E. coli* as a C-terminal intein fusion. The intein catalyzes rearrangement of its N-terminal Cys to a thioester with the C-terminus of Ub. Addition of small molecule thiol releases Ub from the intein as a C-terminal α-thioester.

peptides as C-terminal α-thioesters, which reflects a need for methods that are compatible with various chemical groups, or that accommodate different side chains at the C-terminal position. Most rely on 9-fluorenylmethoxycarbonyl (Fmoc) based SPPS. Due to the basic conditions required for peptide synthesis, thioesters are prepared through a modified resin linker followed by chemical manipulation,[67,68] or through protected thiols poised for nucleophilic attack of the C-terminal peptidyl amide, termed crypto-thioesters (**Figure 1.5**).[69–71] One such crypto-thioester method is based on a ligation auxiliary molecule that I utilize in Chapters 2 and 3 for native histone ubiquitylation and SUMOylation. This method enables one-pot thioester formation and NCL, as well as single-step formation of cyclic peptides.[72] Another particularly important method was developed by the Liu research group, in which peptides are purified as C-terminal hydrazides. These peptides are compatible with convergent, multi-step ligation strategies, and are only converted to reactive thioesters upon diazotization reaction with sodium nitrite, followed by azide displacement with a thiol and ligation *in situ*.[73]

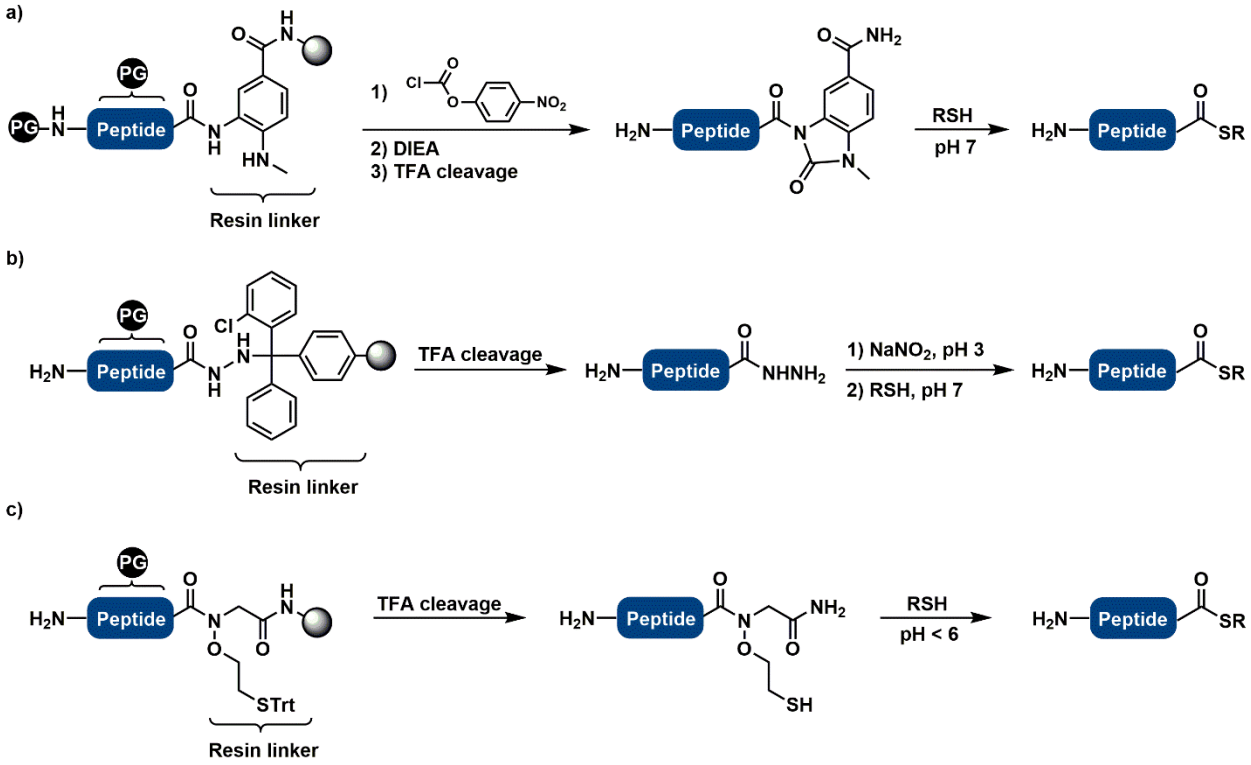


Figure 1.5. Synthesis of peptidyl C-terminal α-thioesters by Fmoc-SPPS. a) Peptides are synthesized from a 4-(methylamino)-3-aminobenzoic acid derived linker. On-resin treatment with 4-nitrophenyl chloroformate activates the C-terminus as an *N*-acylurea. Cleavage and global deprotection and subsequent treatment with thiol under basic conditions yields the thioester. b) Peptides are synthesized from 2-chlorotrityl hydrazine resin. Cleavage and global deprotection results in the peptidyl C-terminal hydrazine. Oxidation with $NaNO_2$ and subsequent treatment with thiol under basic conditions yields the thioester. c) Peptides are synthesized from a *N*-mercaptoethoxyglycinamide linker. Cleavage and global deprotection and subsequent treatment with thiol under acidic conditions yields the thioester. DIEA, diisopropylethylamine; PG, acid-labile protecting group; TFA, trifluoroacetic acid; Trt, trityl.

Peptides containing a N-terminal Cys are readily accessible by Fmoc-SPPS in most cases, and compatibility with multiple ligation steps is achieved by protecting the thiol as a thiazolidine (Thz). The N-terminal Cys must be uncapped with methoxylamine prior to ligation.[74] Preparation of recombinant proteins containing N-terminal Cys is more complex. This is mainly due to incomplete processing of Met in *E. coli* to reveal the N-terminal Cys.[75] To avoid this problem protein fusions are employed, with a protease cleavage site inserted immediately before Cys in the sequence. An affinity tag can be added N-terminal to the cleavage site to facilitate purification, and removed during incubation with the appropriate protease. Common proteases for this purpose are the tobacco etch virus (TEV) protease, with cleavage site

ENLYFQ\S(C),[76] or SUMO proteases such as Ulp1 or SENP2.[77] N-terminal SUMO fusions may also increase the solubility and yield of the recombinant protein.

Once the ligation between N- and C-terminal fragments is complete, a Cys residue is left at the site of ligation, which could make identifying a ligation site problematic due to the rare occurrence of Cys in proteins. However, Ser or Ala may sometimes be mutated to Cys without a major effect on structure or function.[78] Alternatively, free-radical or Raney-Nickel mediated desulfurization of Cys yields an Ala residue, which extends ligation to sites containing Ala, which is very common in proteins.[79,80] Desulfurization occurs at all Cys residues, which is a significant disadvantage unless the protein substrate contains no Cys. Fortunately, histones lack Cys completely (except H3, discussed above), so they are well suited to Ala-to-Cys mutations to accommodate NCL, followed by free radical desulfurization back to the native Ala.[55] Histones synthesized by this method are completely capable of being assembled into mononucleosomes or nucleosomal arrays, and have been used in a wide range of *in vitro* studies.[81]

1.7 Chemical histone modification: ubiquitin-like proteins

Ubiquitin (Ub) is a small, 76-amino acid protein that is highly conserved in eukaryotes.[82] Polyubiquitylation of substrate proteins, in which Ub is ligated via its C-terminus to the Lys ε-amine or N-terminal amine of a prior Ub to form chains, can result in targeted degradation by the proteasome.[83] In contrast, monoubiquitylation is primarily involved in non-proteolytic functions, and was originally discovered as a modification of histone H2A.[84] Ub is involved in myriad cellular pathways, including proteolysis, gene regulation, receptor endocytosis, and DNA double-stranded break repair.[85] Protein ubiquitylation *in vivo* is accomplished by a series of enzymatic reactions involving the E1, E2, and E3 Ub ligases (**Figure 1.6**). In humans, there are 2 E1, ~40 E2, and > 600 E3 enzymes.[86] Ub is expressed as an inactive poly-Ub precursor, which must be processed by deubiquitylating enzymes (DUBs) into its mature form.[87] First, the E1 Ub-activating enzyme (UBA1 or UBA6 in humans) charges the C-terminal Gly of Ub with ATP, which forms a Ub-adenylate and is accompanied by pyrophosphate release. Then, the active-site Cys forms a

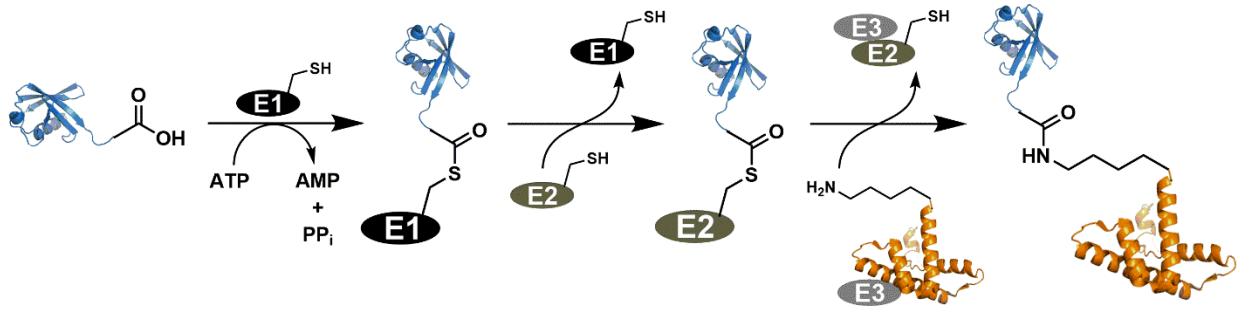


Figure 1.6. The protein ubiquitylation cascade. a) Ubiquitin is activated as a C-terminal thioester by an E1 activating enzyme, then transferred to an E2 conjugating enzyme. The E3 ligase catalyzes transfer of Ub to its substrate and imparts substrate specificity. Ub PDB code, 1UBQ. Substrate PDB code, 1KX5.

thioester with the C-terminus of Ub, and a transthioesterification reaction results in a Ub C-terminal thioester with an E2 Ub-conjugating enzyme. Finally, an E3 Ub-ligating enzyme catalyzes condensation of the Ub thioester with a Lys ε-amine of a target protein either through a covalent thioester intermediate (HECT-type), or by mediating the transfer of Ub directly from E2 to substrate (RING-type).[88] There is no known consensus sequence for ubiquitylation. In addition, the E3 ligase responsible for enzymatic ubiquitylation has not been identified for most substrates, rendering site-specific enzymatic ubiquitylation inaccessible.[89] There also exists a group of Ub-like proteins, including SUMO-1-3 and Nedd8, which share a similar sequence, three-dimensional structure, and enzymatic cascade for substrate conjugation with Ub.[90] These proteins and intracellular pathways are even less well-characterized than those of Ub, and warrant further study. Thus, semisynthetic access to chemically-defined substrates modified by Ub or Ub-like proteins is crucial to enabling *in vitro* study of these processes.

The synthesis and semisynthesis of Ub-modified proteins is a unique challenge. Because Ub and its substrate do not form a linear amino acid chain (except when the N-terminal amine is modified), direct recombinant expression or SPPS combined with NCL is not possible. Chemical strategies have been developed to overcome this issue, and are often demonstrated by synthesizing di-Ub and comparing the biological activity of the its various isomeric linkages (at Ub Lys 6, 11, 27, 29, 33, 48, and 63).[88] Ub is quite amenable to chemical manipulation. The first total chemical synthesis of Ub by Fmoc-SPPS was performed

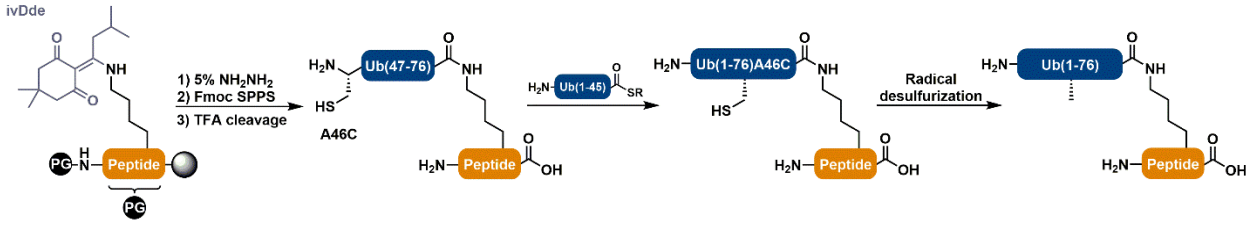


Figure 1.7. Peptide ubiquitylation by branched Fmoc-SPPS and NCL. A peptidyl Lys orthogonally protected with the ivDde group is selectively deprotected on the solid phase, and C-terminal residues of Ub sequentially coupled from the ε-amine, terminating in an A46C mutation. Following acidolytic cleavage and global deprotection, ligation with a Ub(1-45) C-terminal α-thioester affords the ubiquitylated peptide with a Ub A46C mutation. Free radical desulfurization yields the native Ub sequence. R = $CH_2CH_2SO_3H$. ivDde, 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)-3-methylbutyl; PG, acid-labile protecting group; TFA, trifluoroacetic acid.

by Ramage and coworkers in 1989,[91] and was followed by studies demonstrating that synthetic Ub is identical to natural Ub in recognition by anti-Ub antibodies,[92] X-ray crystal structure,[93] and as a Ub ligase substrate.[94] To improve on the ~4% yield of total Ub synthesis, recombinant methods, particularly those involving NCL, have been widely utilized. Like histones, Ub has no native Cys residues, so desulfurization permits temporary Cys-to-Ala mutations for NCL. This enabled Brik and coworkers to apply NCL to a synthesis of Ub directly from the ε-amine of Lys on resin-bound peptides (**Figure 1.7**).[95] Peptides were synthesized on the solid phase with the Lys of interest orthogonally protected by the hydrazine-labile 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)-3-methylbutyl (ivDde) group. After assembly of the main chain and protection of the N-terminus, ivDde was deprotected and Ub(46-76) was sequentially coupled to the Lys ε-amine with an Ala46Cys mutation. Separately, the Ub(1-45) peptide was synthesized as a C-terminal thioester. Ligation and subsequent desulfurization yielded native, ubiquitylated peptides.

In another peptide ubiquitylation which avoided the need to synthesize half of Ub by SPPS, Muir and coworkers developed a route to synthesize full-length H2B ubiquitylated at Lys 120 (ubH2B) (**Figure 1.8**).[74] The H2B(117-125) peptide was synthesized by Fmoc-SPPS, and Lys120 orthogonally protected with the 4-methyltrityl (Mtt) group. The N-terminal Cys of the peptide was coupled as Thz. Following selective deprotection of Lys120, Cys was coupled to the ε-amine. The purified peptide was ligated to recombinant Ub(1-75) α-thioester, then Thz deprotected to reveal Cys. After a second ligation to recombinant H2B(1-116) α-thioester, desulfurization yielded ubH2B with a single Gly-to-Ala mutation at position 76 of Ub. This

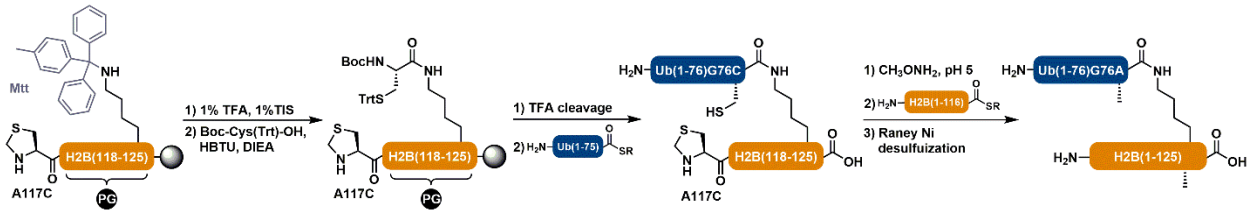


Figure 1.8. Synthesis of ubH2B by ligation on side chain Cys. The H2B(117-125) peptide with N-terminal thiazolidine is synthesized by Fmoc-SPPS. Lys corresponding to position 120, orthogonally protected with the Mtt group, is selectively deprotected on the solid phase, and Cys coupled to the Lys ε-amine. Acidolytic cleavage and global deprotection followed by ligation with a Ub(1-75) C-terminal α-thioester affords the ubiquitylated peptide with a Ub G76C mutation. Thiazolidine deprotection, ligation to H2B(1-116) C-terminal α-thioester, and subsequent desulfurization yields ubH2B with a Ub G76A mutation. R = $CH_2CH_2SO_3H$. Boc, *tert*-Butoxycarbonyl; DIEA, diispropylethylamine; HBTU, *O*-(Benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate; Mtt, 4-methyltrityl; PG, acid-labile protecting group; TFA, trifluoroacetic acid; Trt, trityl.

mutation may inhibit processes involving direct recognition of the isopeptide bond, such as hydrolysis by some DUBs,[96,97] however no such differences were observed in this study.

To streamline synthetic ubiquitylation of peptides, Brik and coworkers developed a 5-mercaptolysine derivative (also referred to as δ-mercaptolysine or δ-thiolysine) which may be incorporated at any position in a synthetic peptide and enables direct ligation of a Ub(1-76) C-terminal α-thioester with the ε-amine of the Lys side chain.[98] Desulfurization renders this method traceless, making it appealing for structural studies of Ub dimers or chains (**Figure 1.9a**). This Lys analog was applied to the impressive synthesis of the seven possible isomeric forms of di-Ub,[99] and in the total synthesis of Lys48- and Lys63-linked tetra-Ub chains.[100] Derivatives of 4-mercaptolysine (also referred to as γ-mercaptolysine or γ-thiolysine) have also been applied to di-Ub synthesis with a clever twist. Positioning the sulfhydryl group at the γ-position allows it to participate in NCL via a 6-membered ring intermediate with either the ε-amine or the α-amine if Lys is the N-terminal peptidyl amino acid (**Figure 1.9b**). Thus, Cys is not required and two distinct NCL steps may be achieved, provided that the amines can be deprotected individually.[101,102] The synthesis of these Lys derivatives, however, is relatively low-yielding and requires many synthetic steps. For instance, the synthesis of γ-thiolysine as originally reported required 21 synthetic steps and resulted in ~5% overall yield.[101] An alternative route has improved the synthesis to 8 steps, and the ~5% overall yield is due mainly to the low-yielding (~11%) initial L-lysine chlorination step, which can be performed at large scale with low-

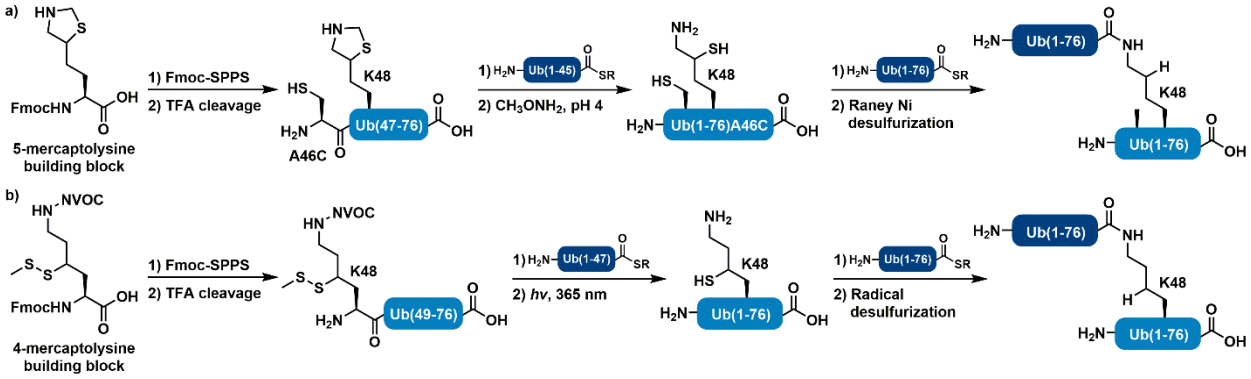


Figure 1.9. Di-Ub synthesis utilizing mercaptolysine derivatives. a) Fmoc-SPPS incorporates thiazolidine-protected 5-mercaptolysine into a Ub chain at position 48, with an N-terminal A46C mutation. Acidolytic cleavage followed by ligation to Ub(1-45) α-thioester gives full-length Ub with an A46C mutation. Thiazolidine deprotection, ligation to Ub(1-76) ε-thioester, and subsequent desulfurization yields mutation-free K48-linked di-Ub. b) Fmoc-SPPS incorporates *N*-NVOC and *S*-disulfide protected 4-mercaptolysine at the N-terminus of a Ub chain at position 48. Acidolytic cleavage and ligation to Ub(1-47) α-thioester under reducing conditions affords full-length Ub. Irradiation at 365 nm deprotects the K48 ε-amine for ligation with Ub(1-76) α-thioester. Subsequent desulfuridation yields mutation-free K48-linked di-Ub. R = $CH_2CH_2SO_3H$. Fmoc, 9- fluorenylmethoxycarbonyl; NVOC, *ortho*-nitroveratryloxycarbonyl; PG, acid-labile protecting group; TFA, trifluoroacetic acid.

cost reagents.[103] A derivative of 5-mercaptolysine has been synthesized over 8 steps in ~50% overall yield, but yield is highly dependent on the identity of the thiol protecting group.[100,104] Despite synthetic drawbacks, use of these Lys analogs in NCL has proven simple and useful, such as in and amber suppression system[105] and in the synthesis of ubiquitylated α-synuclein. Studies with this substrate revealed the inhibitory role of Ub in the formation of toxic α-synuclein aggregates and progression of Parkinson's disease.[106]

An alternative method for synthesizing native ubiquitylated peptides makes use of a small molecule ligation auxiliary. NCL requires a 1,2- or 1,3- aminothiol moiety to permit favorable *S*-to-*N* acyl shift through a 5 or 6-membered ring intermediate. The ε-amine of lysine naturally lacks this feature, so Lys derivatives and auxiliary molecules have been devised to temporarily provide it. A diverse array of N^{α}-auxiliary molecules have been developed, and unlike mercaptolysine they do not rely strictly on desulfurization for their removal.[107] A photocleaveable ligation auxiliary developed by Chatterjee and coworkers was installed on a peptidyl Lys side chain on the solid phase in the form of an *N*-substituted Gly (the native C-terminal residue of Ub) coupled at the ε-amine.[108] After ligation with a Ub(1-75) α-thioester, UV irradiation simultaneously removed the auxiliary and the 2-nitrobenzyl group protecting the N-terminal peptidyl Cys, which was then

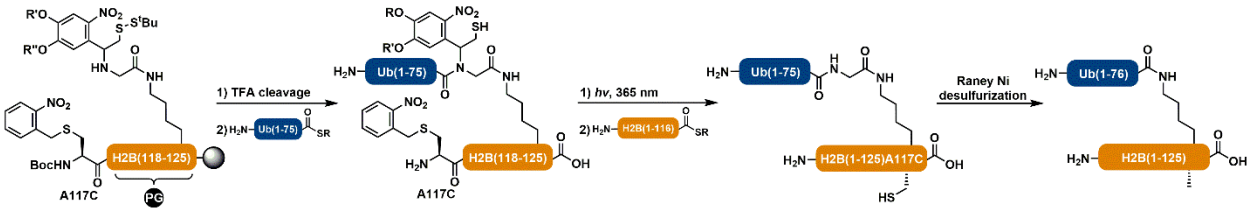


Figure 1.10. Synthesis of ubH2B by N$^{\alpha}$-auxiliary assisted side-chain ligation. H2B(117-125) is synthesized by Fmoc-SPPS with *N*-Boc and *S*-(2-nitrobenzyl) protected Cys at the N-terminus. Lys corresponding to position 120, orthogonally protected with the Mtt group, is selectively deprotected on the solid phase. Bromoacetic acid is coupled to the ε-amine, followed by nucleophilic displacement of bromine by the photo-protected N$^{\alpha}$-auxiliary, which forms the C-terminal Gly of Ub. Acidolytic cleavage, then ligation to Ub(1-75) α-thioester affords the ubiquitylated peptide retaining the ligation auxiliary. Irradiation at 365 nm removes the auxiliary, and deprotects the N-terminal Cys for ligation to H2B(1-116) α-thioester, giving ubH2B with A117C mutation in H2B. Desulfurization yields native ubH2B. R = $CH_2CH_2SO_3H$, R' = $CH_2CH_2CH_2C(O)NH_2CH_3$, R'' = CH_3. Boc, *tert*-Butoxycarbonyl; PG, acid-labile protecting group; TFA, trifluoroacetic acid.

utilized for a second ligation with the recombinant N-terminal fragment of H2B (described above) to generate ubH2B (**Figure 1.10**).[109] The photolabile auxiliary in this synthesis exemplifies the utility of small molecule-assisted NCL, yet similar to the mercaptolysine derivatives, suffers from a long, low-yielding synthesis (~8% yield, 8 steps).[108] In Chapter 2 I describe our application of a synthetically straightforward ligation auxiliary based on 2-(aminooxy)ethanethiol to the ubiquitylation and SUMOylation of peptides.

While many methods aim to generate the native isopeptide linkage between substrate Lys and Ub, others generate non-native linkages due to ease of synthesis, or to prevent hydrolysis by DUBs. In an example of a non-hydrolyzable di-Ub linkage, one Ub containing a Gly76Cys mutation and a second Ub bearing a Lys-to-Cys mutation are reacted with dichloroacetone.[110] For proteins that do not specifically recognize the isopeptide bond, these linkages give facile access to substrates for binding studies. This particular study utilized Lys29-linked tetra-Ub chains affixed to a resin to pull down interacting partners of this specific Ub chain topology, and identified a novel interaction with yeast Ufd3, a member of the Ub chain degradation pathway.[111] Other methods to make non-hydrolyzable mimic linkages include the azide-alkyne click reaction[112], or oxime formation.[113] Alternatively, a hydrolyzable linker may be generated by forming an amide with allylamine at the C-terminus of Ub, through an intein or similar method, which reacts with a Lys-to-Cys mutant substrate via the thiol-ene 'click' reaction.[114] Finally, disulfide-directed ubiquitylation is a

simple, easily applied method useful for proteins lacking natural Cys residues. An intein fusion is used to generate Ub with a C-terminal cystamine amide, which is activated as a reactive, asymmetric disulfide with 2-thio(5-nitropyridine). Reaction with a substrate protein bearing a single Lys-to-Cys mutation readily forms a disulfide linkage at the site of ubiquitylation.[56] This strategy was recently used to generate nucleosomal arrays containing SUMOylated H4 (suH4), a PTM generally considered to be repressive (**Figure 1.11**).[115] The compaction of these arrays in the presence of divalent cations was evaluated relative to unmodified and H4K16Ac-modified arrays. Surprisingly, suH4 did not permit compaction of the arrays. The implications of this finding and our efforts to elucidate the repressive mechanism of suH4 will be discussed in Chapter 4.

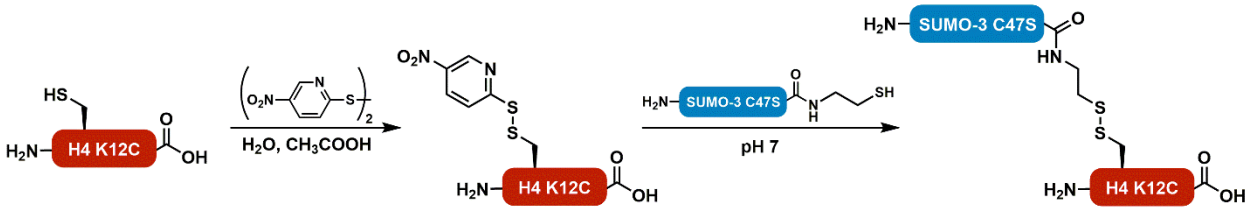


Figure 1.11. Disulfide-directed H4 SUMOylation. Recombinant H4 K12C is reacted with 2,2'-dithiobis(5-nitropyridine) to form an activated disulfide. Disulfide exchange with recombinantly-produced SUMO-3 C47S C-terminal cystamine amide gives suH4 with a disulfide linkage at H4 K12.

1.8 References

(1) Beadle, G. W.; Tatum, E. L. Genetic control of biochemical reactions in neurospora. *Proc. Natl. Acad. Sci. U. S. A.* **1941**, *27* (11), 499–506.

(2) Waddington, C. H. The epigenotype. *Endeavour* **1942**, *1*, 18–20.

(3) Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1964**, *51* (5), 786–794.

(4) Holliday, R.; Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science.* **1975**, *187* (4173), 226–232.

(5) Olins, A. L.; Olins, D. E. Spheroid chromatin units (v bodies). *Science.* **1974**, *183* (4122), 330–332.

(6) Woodcock, C. L. F.; Safer, J. P.; Stanchfield, J. E. Structural repeating units in chromatin. I. Evidence for their general occurrence. *Exp. Cell Res.* **1976**, *97*, 101–110.

(7) Kornberg, R. D. Chromatin structure: A repeating unit of histones and DNA. *Science.* **1974**, *184* (4139), 868–871.

(8) Richmond, T. J.; Finch, J. T.; Rushton, B.; Rhodes, D.; Klug, A. Structure of the nucleosome core particle at 7 Å resolution. *Nature* **1984**, *311*, 532–537.

(9) Li, G.; Zhu, P. Structure and organization of chromatin fiber in the nucleus. *FEBS Lett.* **2015**, *589*, 2893–2904.

(10) Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, *389*, 251–260.

(11) Mariño-Ramírez, L.; Kann, M. G.; Shoemaker, B. A.; Landsman, D. Histone structure and nucleosome stability. *Expert Rev. Proteomics* **2005**, *2* (5), 719–729.

(12) Struhl, K.; Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **2013**, *20* (3), 267–723.

(13) Böhm, L.; Crane-Robinson, C. Proteases as structural probes for chromatin: The domain structure of histones. *Biosci. Rep.* **1984**, *4*, 365–386.

(14) Pepenella, S.; Murphy, K. J.; Hayes, J. J. Intra- and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure. *Chromosoma* **2014**, *123* (1–2), 3–13.

(15) Luger, K.; Dechassa, M. L.; Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. cell Biol.* **2012**, *13* (7), 436–447.

(16) Schalch, T.; Duda, S.; Sargent, D. F.; Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **2005**, *436*, 138–141.

(17) Woodcock, C. L.; Ghosh, R. P. Chromatin higher-order structure and dynamics. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (5), a000596.

(18) Robinson, P. J. J.; Rhodes, D. Structure of the “30 nm” chromatin fibre: A key role for the linker histone. *Curr. Opin. Struct. Biol.* **2006**, *16*, 336–343.

(19) Dorigo, B.; Schalch, T.; Kulangar, A.; Duda, S.; Schroeder, R. R.; Richmond, T. J. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science.* **2004**, *306*, 1571–1573.

(20) Simpson, R. T. Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* **1978**, *17* (25), 5524–5531.

(21) Maresca, T. J.; Freedman, B. S.; Heald, R. Histone H1 is essential for mitotic chromosome architecture and segregation in Xenopus laevis egg extracts. *J. Cell Biol.* **2005**, *169* (6), 859–869.

(22) Dekker, J.; Heard, E. Structural and functional diversity of topologically associating domains. *FEBS Lett.* **2015**, *589*, 2877–2884.

(23) Rodriguez, A.; Bjerling, P. The links between chromatin spatial organization and biological function. *Biochem. Soc. Trans.* **2013**, *41* (6), 1634–1639.

(24) Handy, D. E.; Castro, R.; Loscalzo, J. Epigenetic modifications: Basic mechanisms and role in

cardiovascular disease. *Circulation* **2011**, *123* (19), 2145–2156.

(25) Hegyi, H.; Kalmar, L.; Horvath, T.; Tompa, P. Verification of alternative splicing variants based on domain integrity, truncation length and intrinsic protein disorder. *Nucleic Acids Res.* **2011**, *39* (4), 1208–1219.

(26) Ponomarenko, E. A.; Poverennaya, E. V.; Ilgisonis, E. V.; Pyatnitskiy, M. A.; Kopylov, A. T.; Zgoda, V. G.; Lisitsa, A. V.; Archakov, A. I. The size of the human proteome: The width and depth. *Int. J. Anal. Chem.* **2016**, *2016*, 7436849.

(27) Tian, Z.; Tolić, N.; Zhao, R.; Moore, R. J.; Hengel, S. M.; Robinson, E. W.; Stenoien, D. L.; Wu, S.; Smith, R. D.; Paša-Tolić, L. Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.* **2012**, *13* (10), R86.

(28) Arnaudo, A. M.; Garcia, B. A. Proteomic characterization of novel histone post-translational modifications. *Epigenetics Chromatin* **2013**, *6* (1), 24.

(29) Jeppesen, P.; Turner, B. M. The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell* **1993**, *74* (2), 281–289.

(30) Allis, C. D.; Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **2016**, *17* (8), 487–500.

(31) Shahbazian, M. D.; Grunstein, M. Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* **2007**, *76*, 75–100.

(32) Shogren-Knaak, M.; Ishii, H.; Sun, J.-M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science.* **2006**, *311* (5762), 844–847.

(33) Tessarz, P.; Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (11), 703–708.

(34) Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **1999**, *399* (6735), 491–496.

(35) Petty, E.; Pillus, L. Balancing chromatin remodeling and histone modifications in transcription. *Trends Genet.* **2015**, *29* (11), 621–629.

(36) Metzger, E.; Imhof, A.; Patel, D.; Kahl, P.; Hoffmeyer, K.; Friedrichs, N.; Müller, J. M.; Greschik, H.; Kirfel, J.; Ji, S.; et al. Phosphorylation of histone H3T6 by PKCβI controls demethylation at histone H3K4. *Nature* **2010**, *464* (7289), 792–796.

(37) Patel, D. J.; Wang, Z. Readout of epigenetic modifications. *Annu. Rev. Biochem.* **2013**, *82*, 81–118.

(38) Sanchez, R.; Zhou, M. The PHD finger: A versatile epigenome reader. *Trends Biochem. Sci.* **2011**, *36* (7), 364–372.

(39) Strahl, B. D.; Allis, C. D. The language of covalent histone modifications. *Nature* **2000**, *403* (6765), 41–45.

(40) Badeaux, A. I.; Shi, Y. Emerging roles for chromatin as a signal integration and storage platform. *Nat. Rev. Mol. Cell Biol.* **2013**, *14* (4), 211–224.

(41) Li, B.; Carey, M.; Workman, J. L. The role of chromatin during transcription. *Cell* **2007**, *128* (4), 707–719.

(42) Knezetic, J. A.; Luse, D. S. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell* **1986**, *45* (1), 95–104.

(43) Izban, M. G.; Luse, D. S. Transcription on nucleosomal templates by RNA polymerase II in vitro: Inhibition of elongation with enhancement of sequence- specific pausing. *Genes Dev.* **1991**, *5*, 683–696.

(44) Kireeva, M. L.; Walter, W.; Tchernajenko, V.; Bondarenko, V.; Kashlev, M.; Studitsky, V. M. Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription. *Mol. Cell* **2002**, *9* (3), 541–552.

(45) Zhang, T.; Cooper, S.; Brockdorff, N. The interplay of histone modifications - Writers that read. *EMBO Reports.* **2015**, *16* (11), 1467–1481.

(46) Batta, K.; Zhang, Z.; Yen, K.; Goffman, D. B.; Franklin Pugh, B. Genome-wide function of H2B ubiquitylation in promoter and genic regions. *Genes Dev.* **2011**, *25* (21), 2254–2265.

(47) Cheung, P.; Tanner, K. G.; Cheung, W. L.; Sassone-Corsi, P.; Denu, J. M.; Allis, C. D. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol. Cell* **2000**, *5* (6), 905–915.

(48) Sawicka, A.; Seiser, C. Histone H3 phosphorylation - A versatile chromatin modification for different occasions. *Biochimie* **2012**, *94* (11), 2193–2201.

(49) Fischle, W.; Mootz, H. D.; Schwarzer, D. Synthetic histone code. *Curr. Opin. Chem. Biol.* **2015**, *28*, 131–140.

(50) Dhall, A.; Chatterjee, C. Chemical approaches to understand the language of histone modifications. *ACS Chem. Biol.* **2011**, *6* (10), 987–999.

(51) Krieger, D. E.; Levine, R.; Merrifield, R. B.; Vidali, G.; Allfrey, V. G. Chemical studies of histone acetylation: Substrate specificity of a histone deacetylase from calf thymus nuclei. *J. Biol. Chem.* **1974**, *249*, 332–334.

(52) Krieger, D. E.; Vidali, G.; Erickson, B. W.; Allfrey, V. G.; Merrifield, R. B. The synthesis of diacetylated histone H4-(1-37) for studies on the mechanism of histone deacetylation. *Bioorg. Chem.* **1979**, *8*, 409–427.

(53) Kervabon, A.; Mery, J.; Parello, J. Enzymatic deacetylation of a synthetic peptide fragment of histone H4. *FEBS Lett.* **1979**, *106* (1), 93–96.

(54) Carmen, A. A.; Milne, L.; Grunstein, M. Acetylation of the yeast histone H4 N terminus regulates its binding to heterochromatin protein SIR3. *J. Biol. Chem.* **2002**, *277* (7), 4778–4781.

(55) Howard, C. J.; Yu, R. R.; Gardner, M. L.; Shimko, J. C.; Ottesen, J. J. Chemical and biological tools for the preparation of modified histone proteins. *Top. Curr. Chem.* **2015**, *363*, 193–226.

(56) Chatterjee, C.; McGinty, R. K.; Fierz, B.; Muir, T. W. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat. Chem. Biol.* **2010**, *6* (4), 267–269.

(57) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007**, *128* (5), 1003–1012.

(58) Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiold, L.; Liu, C.-F. A direct method for site-specific protein acetylation. *Angew. Chemie* **2011**, *50* (41), 9611–9614.

(59) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Genetically encoding Nε-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **2008**, *4* (4), 232–234.

(60) Nguyen, D. P.; Alai, M. M. G.; Kapadnis, P. B.; Neumann, H.; Chin, J. W. Genetically encoding Nε-methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.* **2009**, *131*, 14194–14195.

(61) Wang, Y.-S.; Wu, B.; Wang, Z.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P.-J.; Moe, Y. N.; Russell, D. H.; Liu, W. R. A genetically encoded photocaged Nε-methyl-l-lysine. *Mol. Biosyst.* **2010**, *6* (9), 1557–1560.

(62) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angew. Chemie* **2008**, *47* (34), 6399–6401.

(63) Dawson, P. E.; Kent, S. B. H. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.

(64) Dawson, P. E.; Muir, T. W.; Kent, S. B. H. Synthesis of proteins by native chemical ligation. *Science.* **1994**, *266*, 776–778.

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

(66) Shah, N. H.; Muir, T. W. Inteins: Nature's gift to protein chemists. *Chem. Sci.* **2014**, *5* (2), 446–461.

(67) Blanco-Canosa, J. B.; Nardone, B.; Albericio, F.; Dawson, P. E. Chemical protein synthesis using a second-generation N-acylurea linker for the preparation of peptide-thioester precursors. *J. Am. Chem. Soc.* **2015**, *137* (22), 7197–7209.

(68) Zheng, J.; Tang, S.; Huang, Y.; Liu, L. Development of new thioester equivalents for protein chemical synthesis. *Acc. Chem. Res.* **2013**, *46* (11), 2475–2484.

(69) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. Bis(2-sulfanylethyl)amino native peptide ligation. *Org. Lett.* **2010**, *12* (22), 5238–5241.

(70) Erlich, L. A.; Kumar, K. S. A.; Haj-Yahya, M.; Dawson, P. E.; Brik, A. N-methylcysteine-mediated total chemical synthesis of ubiquitin thioester. *Org. Biomol. Chem.* **2010**, *8* (10), 2392–2396.

(71) Zheng, J.; Chen, X.; Tang, S.; Chang, H.; Wang, F.; Zuo, C. A new method for synthesis of peptide thioesters via irreversible N-to-S acyl transfer. *Org. Lett.* **2014**, *16*, 4908–4911.

(72) Shelton, P. M. M.; Weller, C. E.; Chatterjee, C. A facile N-mercaptoethoxyglycinamide (MEGA) linker

approach to peptide thioesterification and cyclization. *J. Am. Chem. Soc.* **2017**, *139* (11), 3946–3949.

(73) Fang, G.-M.; Wang, J.-X.; Liu, L. Convergent chemical synthesis of proteins by ligation of peptide hydrazides. *Angew. Chemie* **2012**, *51* (41), 10347–10350.

(74) McGinty, R. K.; Köhn, M.; Chatterjee, C.; Chiang, K. P.; Pratt, M. R.; Muir, T. W. Structure activity analysis of semisynthetic nucleosomes: Mechanistic insights into the stimulation of Dot1L by ubiquitylated histone H2B. *ACS Chem. Biol.* **2009**, *4* (11), 958–968.

(75) Hirel, P.-H.; Schmitter, J.-M.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (21), 8247–8251.

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

(77) Marblestone, J. G.; Edavettal, S. C.; Lim, Y.; Lim, P.; Zuo, X.; Butt, T. R. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **2006**, *15*, 182–189.

(78) Xia, X.; Longo, L. M.; Blaber, M. Mutation choice to eliminate buried free cysteines in protein therapeutics. *J. Pharm. Sci.* **2015**, *104*, 566–576.

(79) Wan, Q.; Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: A powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chemie* **2007**, *119*, 9408–9412.

(80) Yan, L. Z.; Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **2001**, *123*, 526–533.

(81) Keung, A. J.; Joung, J. K.; Khalil, A. S.; Collins, J. J. Chromatin regulation at the frontier of synthetic biology. *Nat. Rev. Genet.* **2015**, *16* (3), 159–171.

(82) Sharp, P. M.; Li, W.-H. Molecular evolution of ubiquitin genes. *Trends Ecol. Evol.* **1987**, *2* (11), 328–332.

(83) Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **2009**, *78*, 477–513.

(84) Goldknopf, I. L.; French, M. F.; Musso, R.; Busch, H. Presence of protein A24 in rat liver nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (12), 5492–5495.

(85) Hicke, L. Protein regulation by monoubiquitin. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 195–201.

(86) Clague, M. J.; Heride, C.; Urbé, S. The demographics of the ubiquitin system. *Trends Cell Biol.* **2015**, *25* (7), 417–426.

(87) Amerik, A. Y.; Hochstrasser, M. Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta* **2004**, *1695*, 189–207.

(88) Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **1998**, *67*, 425–479.

(89) Weller, C. E.; Pilkerton, M. E.; Chatterjee, C. Chemical strategies to understand the language of ubiquitin signaling. *Biopolymers* **2014**, *101* (2), 144–155.

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

(91) Ramage, R.; Green, J.; Ogunjobi, O. M. Solid phase peptide synthesis of ubiquitin. *Tetrahedron Lett.* **1989**, *30* (16), 2149–2152.

(92) Ramage, R.; Green, J.; Muir, T. W.; Ogunjobi, O. M.; Love, S.; Shaw, K. Synthetic, structural and biological studies of the ubiquitin system: The total chemical synthesis of ubiquitin. *Biochem. J.* **1994**, *299*, 151–158.

(93) Alexeev, D.; Bury, S. M.; Turner, M. A.; Ogunjobi, O. M.; Muir, T. W.; Ramage, R.; Sawyer, L. Synthetic, structural and biological studies of the ubiquitin system: Chemically synthesized and native ubiquitin fold into identical three-dimensional structures. *Biochem. J.* **1994**, *299*, 159–163.

(94) Love, S. G.; Muir, T. W.; R, R. R.; Shaw, K. T.; Alexeev, D.; Sawyer, L.; Kelly, S. M.; Price, N. C.; Arnold, J. E.; Mee, M. P.; et al. Synthetic, structural and biological studies of the ubiquitin system: Synthesis and crystal structure of an analogue containing unnatural amino acids. *Biochem. J.* **1997**, *323*, 727–734.

(95) Kumar, K. S. A.; Spasser, L.; Ohayon, S.; Erlich, L. A.; Brik, A. Expeditious chemical synthesis of ubiquitinated peptides employing orthogonal protection and native chemical ligation. *Bioconjug. Chem.* **2011**, *22*, 137–143.

(96) Pickart, C. M.; Kasperek, E. M.; Beal, R.; Kim, A. Substrate properties of site-specific mutant ubiquitin protein (G76A) reveal unexpected mechanistic features of ubiquitin-activating enzyme (E1). *J. Biol. Chem.* **1994**, *269* (10), 7115–7123.

(97) Hodgins, R. R. W.; Ellison, K. S.; Ellison, M. J. Expression of a ubiquitin derivative that conjugates to protein irreversibly produces phenotypes consistent with a ubiquitin deficiency. *J. Biol. Chem.* **1992**, *267* (13), 8807–8812.

(98) Ajish Kumar, K. S.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A. Highly efficient and chemoselective peptide ubiquitylation. *Angew. Chemie* **2009**, *121*, 8234–8238.

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

(100) El Oualid, F.; Merkx, R.; Ekkebus, R.; Hameed, D. S.; Smit, J. J.; de Jong, A.; Hilkmann, H.; Sixma, T. K.; Ovaa, H. Chemical synthesis of ubiquitin, ubiquitin-based probes, and diubiquitin. *Angew. Chemie* **2010**, *49*, 10149–10153.

(101) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. Dual native chemical ligation at lysine. *J. Am. Chem. Soc.* **2009**, *131*, 13592–13593.

(102) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. *Chem. Commun.* **2010**, *46*, 7199–7201.

(103) Merkx, R.; de Bruin, G.; Kruithof, A.; van den Bergh, T.; Snip, E.; Lutz, M.; El Oualid, F.; Ovaa, H. Scalable synthesis of γ-thiolysine starting from lysine and a side by side comparison with δ-thiolysine in non-enzymatic ubiquitination. *Chem. Sci.* **2013**, *4*, 4494–4498.

(104) Haj-Yahya, M.; Kumar, K. S. A.; Erlich, L. A.; Brik, A. Protecting group variations of δ-mercaptolysine useful in chemical ubiquitylation. *Biopolymers* **2010**, *94* (4), 504–510.

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

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

(107) Offer, J. Native chemical ligation with Nα acyl transfer auxiliaries. *Biopolymers* **2010**, *94* (4), 530–541.

(108) Chatterjee, C.; McGinty, R. K.; Pellois, J.-P.; Muir, T. W. Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chemie* **2007**, *46*, 2814–2818.

(109) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008**, *453* (7196), 812–816.

(110) Yin, L.; Krantz, B.; Russell, N. S.; Deshpande, S.; Wilkinson, K. D. Nonhydrolyzable diubiquitin analogues are inhibitors of ubiquitin conjugation and deconjugation. *Biochemistry* **2000**, *39*, 10001–10010.

(111) Russell, N. S.; Wilkinson, K. D. Identification of a novel 29-linked polyubiquitin binding protein, Ufd3, using polyubiquitin chain analogues. *Biochemistry* **2004**, *43*, 4844–4854.

(112) Weikart, N. D.; Mootz, H. D. Generation of site-specific and enzymatically stable conjugates of recombinant proteins with ubiquitin-like modifiers by the Cu(I)-catalyzed azide-alkyne cycloaddition. *ChemBioChem* **2010**, *11*, 774–777.

(113) Shanmugham, A.; Fish, A.; Luna-Vargas, M. P. A.; Faesen, A. C.; Oualid, F. El; Sixma, T. K.; Ovaa, H. Nonhydrolyzable ubiquitin-isopeptide isosteres as deubiquitinating enzyme probes. *J. Am. Chem. Soc.* **2010**, *132*, 8834–8835.

(114) Valkevich, E. M.; Guenette, R. G.; Sanchez, N. A.; Chen, Y.-C.; Ge, Y.; Strieter, E. R. Forging isopeptide bonds using thiol-ene chemistry: Site-specific coupling of ubiquitin molecules for studying the activity of isopeptidases. *J. Am. Chem. Soc.* **2012**, *134*, 6916–6919.

(115) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. SUMOylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837.

Chapter 2

Auxiliary-mediated synthesis of native and protease-resistant ubiquitylated peptides

2.1 Introduction

The small, 76-residue protein ubiquitin (Ub) is highly conserved in eukaryotes[1] and plays an important role in countless cellular functions.[2] As described in Chapter 1, ubiquitin-like proteins (Ubls), including small ubiquitin-like modifier (SUMO) proteins, are conjugated by their C-termini to ε-amines of substrate Lys residues by a series of Ubl-specific E1, E2, and E3 ligases.[3] Recent methods coupling immunoaffinity-based enrichment of ubiquitylated or SUMOylated species with proteomics have identified hundreds of previously unknown targets.[4,5] Thus, research into the functional consequences of substrate Ubl modification is ongoing. Histone H4 SUMOylation at Lys12 (suH4) is one such mark that remains to be fully characterized, although several studies have linked it to transcriptional repression.[6,7] Meaningful study of suH4 and other Ubl-modified proteins requires the ability to follow their dynamic regulation and subcellular localization in living cells,[8,9] as well as the ability to examine direct biochemical and biophysical consequences of their post-translational modification (PTM) *in vitro*.[10] The overall goal of my thesis project is to interrogate cross-talk between H4 SUMOylation and H3 acetylation *in vitro*, using mononucleosome substrates reconstituted from purified, site-specifically modified histones and DNA. A semisynthetic strategy to access suH4 is necessary, in part because the E3 ligase responsible for site-specific H4 SUMOylation is unknown, rendering enzymatic SUMOylation inaccessible. A method that yields the native isopeptide linkage is preferred, due to its stability under reducing enzymatic assay conditions, in contrast to a recently utilized disulfide-directed H4 SUMOylation strategy.[11] In addition, the importance of the specific length, geometry, and other properties of the isopeptide bond to the interacting partners of suH4 is simply unknown. Utilizing the native linkage ensures the accuracy of the *in vitro* system. In this chapter I will describe the

development of a mild, efficient method for the site-specific modification of peptides with Ubls by a native isopeptide linkage.

There are various semisynthetic methods for site-specific protein ubiquitylation, as reviewed in Chapter 1. Most involve separate ligations between a synthetic peptide and longer, recombinant protein fragments: one to link Ub to the Lys side chain, and one or more to complete the main chain protein.[12–14] The ability to purify Ub and truncated substrate proteins from *E. coli* for native chemical ligation (NCL) significantly increases the potential yield of the overall synthesis. In addition, most chemical manipulations can be limited to the small peptide fragment on the solid phase, which greatly facilitates purification and reduces the likelihood of side reactions. The N- and C-termini of peptides may bear protected or masked functional groups so that ligation chemistry between fragments may occur in any desired order. For ligation on the Lys side chain, a sulfhydryl group must be positioned such that it can perform NCL. Classical NCL involves two basic steps: *capture* and *transfer*.[15] The capture step involves the chemoselective reaction between Cys thiol and thioester to join two unprotected peptide chains. The pK_a of the Cys sulfhydryl group, ~8.5, ensures that a large fraction is anionic under typical NCL conditions (pH 7-8). In contrast, the pK_a of the Lys ε-amine conjugate acid is ~10.5, meaning that it is largely positively charged under the same conditions, and that of the N-terminal α-amine is ~8.[16] Thus, non-specific direct amidation is unlikely to occur. With peptide starting materials typically in low mM concentrations, nucleophilic attack by the N-terminal Cys side chain on a peptide bearing a C-terminal α-thioester results in transthioesterification that positions the α-amine to react with the new thioester by virtue of its now-high local concentration. This *S*-to-*N* acyl transfer step results in a thermodynamically stabile amide bond, and only occurs on a Cys with a correctly-positioned nearby amine. In the case of the Lys ε-amine, unnatural incorporation of a sulfhydryl group necessitates desulfurization to yield the native isopeptide linkage for mercaptolysine derivatives, yet this method is incompatible with other Cys residues in the protein or peptide.[13,14] For ligation auxiliaries, a *release* step to remove the auxiliary group is dependent upon the nature of the group itself. In this way, auxiliaries may be designed such that their removal conditions are compatible with other functional groups

in the protein, or even the folded state of the protein. We recognized that, despite a lack of Cys residues in histones, our protein/peptide ubiquitylation method would be most useful if it were widely applicable to other protein substrates, which excludes desulfurization as an option. Because mercaptolysine derivatives require desulfurization to achieve the native linkage and are synthetically challenging to access, we pursued a ligation auxiliary-based strategy.

There are many forms of NCL auxiliaries, and the vast majority have been developed to perform traceless ligation at the α-amine at either non-Cys sites, or at X-Cys junctions in which X is a hindered, β-branched amino acid.[17] We considered that Ub, and in fact most Ubls, terminates in a C-terminal Gly-Gly sequence, the ideal unhindered position for ligation with an N^{α}-auxiliary (**Table 2.1**).[18] Coupling bromoacetic acid to the desired Lys ε-amine creates a site for facile installation of the auxiliary amine by nucleophilic displacement of bromine on the solid phase. After ligation and auxiliary removal, this becomes the C-terminal Gly of Ub. This strategy was beautifully demonstrated by Chatterjee and coworkers using a photolabile, 2-nitrobenzyl-based auxiliary.[19] In fact, many auxiliaries have been developed based on the N^{α}-benzyl scaffold, and different substitutions on the phenyl ring enable different removal conditions.[17] Yet they still suffer from sometimes harsh removal conditions, or exhibit significantly impaired ligation kinetics that ultimately limit their broad applicability.[17,19] Smaller, hindrance-free auxiliaries based on 2-aminothiol

Table 2.1. C-terminal sequence alignment of human ubiquitin-like modifier proteins.

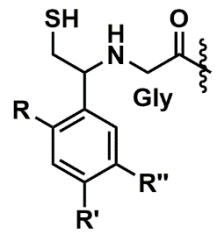
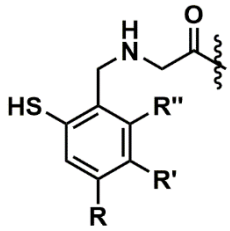
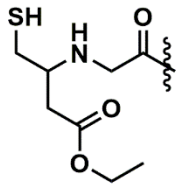
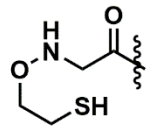
Ubl																					Function
Ubiquitin	-	-	S	T	L	H	L	V	L	R	L	R	**G**	**G**	M	Q	I	F	V	-	-	Protein degradation
SUMO-1	-	-	D	V	I	E	V	Y	Q	E	Q	T	**G**	**G**	H	S	T	V				Nuclear localization
SUMO-2	-	-	D	T	I	D	V	F	Q	Q	Q	T	**G**	**G**	V	Y						Transcriptional regulation
SUMO-3	-	-	D	T	I	D	V	F	Q	Q	Q	T	**G**	**G**	V	P	E	S	S	-	-	Transcriptional regulation
Nedd8	-	-	S	V	L	H	L	V	L	A	L	R	**G**	**G**	G	G	L	R	Q	-	-	Modifies cullin family ligases
FAT10	-	-	N	L	L	F	L	A	S	Y	C	I	**G**	**G**								Apoptosis, signal transduction
Fub1	-	-	T	T	L	E	V	A	G	R	M	L	**G**	**G**	K	V	H	G	S	-	-	T-cell activation
ISG15	-	-	S	T	V	F	M	N	L	R	L	R	**G**	**G**	G	T	E	P	G	-	-	Immune response

*Precursor sequences shown are proteolytically processed C-terminal to the Gly-Gly motif (black).

are simpler to access synthetically and exhibit rapid kinetics of ligation, but removal conditions are harsh (**Table 2.2**).

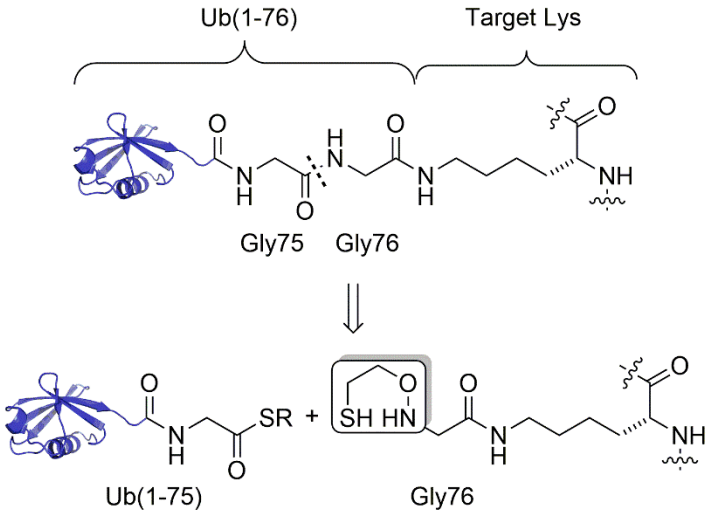
A 2-(aminooxy)ethanethiol-based ligation auxiliary developed by Kent and coworkers in 1996, however, was removed by mild reduction of the *N-O* bond with zinc dust in acidic HPLC buffer (**Table 2.2**).[27] The kinetics of peptidyl ligation were rapid, although rearrangement of the *S*-linked intermediate was slow compared to Cys-based ligation, except at Gly-Gly junctions. This was attributed to the slightly less favorable formation of a 6- (rather than 5-) membered ring intermediate. The auxiliary was later repurposed

Table 2.2. N^{α}-auxiliaries for native chemical ligation at Gly-Gly junctions.

	Substituent groups	Removal	Ref
	R = H, R' = OCH_3, R'' = H	HF or TFA	20,21
	R = OCH_3, R' = OCH_3, R'' = H	TFA	22
	R = NO_2, R' = H, R'' = H	*hv*	23
	R = NO_2, R' = $O(CH_2)_3CONH_2CH_3$, R'' = OCH_3	*hv*	19
	R = H, R' = H, R'' = H	TCEP, morpholine, 40 °C	24
	R = OCH_3, R' = OCH_3, R'' = H	TFMSA/TFA	25
	R = OCH_3, R' = OCH_3, R'' = OCH_3	TFA	22
	n/a	TCEP, morpholine, 60 °C	26
	n/a	Zn, H_2O/CH_3CN, 0.1% TFA	27

TCEP = tris(2-carboxyethyl)phosphine; TFA = trifluoroacetic acid; TFMSA = trifluoromethanesulfonic acid.

by Liu and coworkers for ligations with the one-residue removed primary N-terminal α-amine, in a method analogous to side-chain assisted ligation.[28,29] However, use of this auxiliary was never extended to full-length proteins. We identified 2-(aminooxy)ethanethiol as a ligation auxiliary that fit our ideal criteria: good ligation and acyl transfer kinetics at Gly-Gly junctions, readily achievable synthesis, and mild, efficient removal conditions. In this chapter I demonstrate successful utilization of the auxiliary to synthesize ubiquitylated and SUMOylated peptides, as well as this method's compatibility with protein thiol groups (**Scheme 2.1**). This facile synthetic method will greatly expand the scope of native ubiquitylated and SUMOylated proteins and peptides available for mechanistic studies.

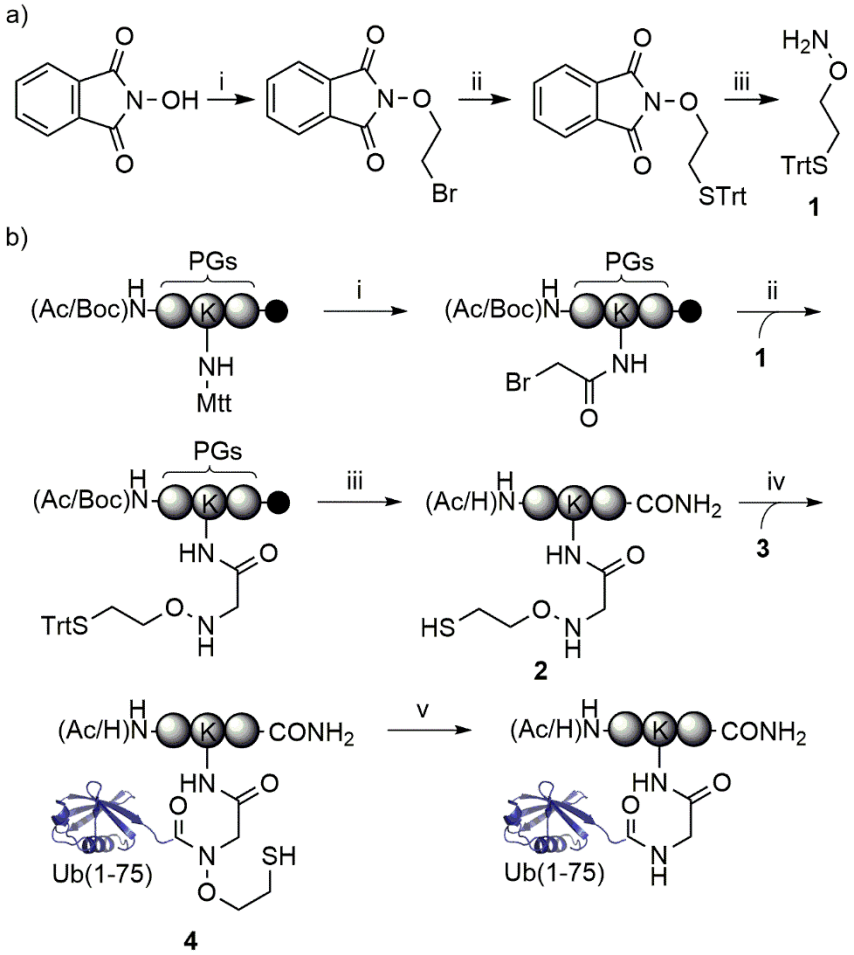


Scheme 2.1. Site-specific peptide ubiquitylation.

2.2 Results and discussion

2.2.1 Synthesis and application of the ligation auxiliary

We began by synthesizing a trityl-protected form of 2-(aminooxy)ethanethiol starting from *N*-hydroxyphthalimide. In contrast to the auxiliary as reported by Kent and Liu in which the thiol is protected by a disulfide or a *p*-methylbenzyl group, we reasoned that the trityl group would be ideal for our application as the peptidyl auxiliary will be deprotected during acidic cleavage from the resin and ready for ligation



Scheme 2.2. Synthesis and application of the ligation auxiliary. a) i) $BrCH_2CH_2Br$, Et_3N, DMF, 18 h, 25 °C, 64%; ii) TrtSH, NaH, DMF, 2 h, 25 °C, 79%; iii) H_2NNH_2, $CHCl_3$, 2.5 h, 25 °C, 90%. b) i) 1. 1% TFA, 1% TIS in CH_2Cl_2; 2. $BrCH_2COOH$, DIC, DMF; ii) **1**, 1 M in DMSO; iii) Reagent K; iv) Ub(1-75)-α-thioester (**3**), 6 M Gn-HCl, 100 mM Na_2HPO_4, 10 mM TCEP, pH 7.5; v) 6 M Gn-HCl, Zn, pH 3.0, 37 °C, 24 h. Ac = acetyl, Boc = tert-butyloxycarbonyl, DIC = N,N′-diisopropylcarbodiimide, Fmoc = 9-fluorenylmethoxycarbonyl, Gn-HCl = guanidinium hydrochloride, Mtt = 4-methyltrityl, PG = protecting group, Reagent K = 82.5:5:5:5:2.5, TFA: thioanisole: H_2O: phenol: 1,2-ethanedithiol (v/v), TCEP = tris(2-carboxyethyl)phosphine, TFA = trifluoroacetic acid, TIS = triisopropylsilane, Trt = trityl.

immediately following purification.[27,28] A facile 3-step synthetic scheme afforded the suitably protected auxiliary, **1**, in multi-gram quantities and 46% overall yield (**Scheme 2.2a**) for application in 9-fluorenylmethoxycarbonyl (Fmoc) based solid-phase peptide synthesis (SPPS). To test the utility of the auxiliary for peptide ubiquitylation, we first considered an internal peptide sequence from Ub itself. There are seven Lys residues in Ub, at positions 6, 11, 27, 29, 33, 48 and 63, and all are found to be modified by ubiquitin *in vivo*.[30] Therefore, we first chose a short N-terminally acetylated and C-terminally amidated tripeptide, QKE, containing Lys63 of Ub, as a test substrate for auxiliary-mediated ubiquitylation. The peptide was synthesized on rink amide resin with an orthogonal 4-methyltrityl (Mtt) protecting group at the Lys ε-amine targeted for modification.[31] After chain assembly, the Lys side-chain was selectively deprotected by 1% TFA in dichloromethane and the free ε-amine was subsequently coupled with bromoacetic acid. The auxiliary was readily introduced via direct nucleophilic displacement of bromine with a 1 M solution of **1** in DMSO for 24 h at room temperature. Importantly, the unreacted auxiliary could be filtered away from the solid-phase and reused at least four times without appreciable losses in incorporation. Finally, TFA-mediated cleavage from the resin afforded the auxiliary containing peptide QK$^{(aux)}$E, **2**, in 38% purified yield based on initial resin loading (**Scheme 2.2b**).

Heterologously expressed human Ub(1-75)-α-thioester, **3**, was obtained by thiolysis of the corresponding GyrA intein fusions with mercaptoethanesulfonic acid (MES).[19] Ligation to the auxiliary-containing peptide was initiated by mixing 2.5 mM of **2** and 0.5 mM of 3 in 6 M Gn-HCl, 100 mM Na_2HPO_4, 10 mM TCEP, pH 7.5 at room temperature. Although auxiliary-mediated ligation onto a secondary amine was previously shown to be slower than ligation to the primary amine in Cys,[12] the enhanced nucleophilicity of the aminooxy group at pH 7.5, due to the alpha effect, led to complete acyl transfer within 12 h.[32] The ligation product was then purified by reversed phase high performance liquid chromatography (RP-HPLC) to obtain the auxiliary-containing branched protein QK$^{Ub(aux)}$E, **4**. Surprisingly, and contradictory to previous results with short peptides,[27] our initial attempts to reductively cleave the ligation auxiliary from **4** with activated Zn in acidic HPLC buffers or water proved largely unsuccessful (**Table 2.3**, entries 1, 2). The inclusion of 6 M

Table 2.3. Reductive removal of the ligation auxiliary.

No.	Product	Reagent	pH	temp (°C)	time (h)	Yield (%)[a]
1	$QK^{Ub}E$	Zn, 1:1 AcN:H_2O	1	22	24	5
2	$QK^{Ub}E$	Zn, H_2O	1	22	24	12
3	$QK^{Ub}E$	Zn, 6M Gn-HCl	1	22	24	28
4	$QK^{Ub}E$	Zn, 6M Gn-HCl	1	37	24	54
5	$QK^{Ub}E$	Raney-Nickel[b]	7	37	17	5
6	$QK^{Ub}E$	In, 6 M Gn-HCl	3	37	24	31
7	$QK^{Ub}E$	Zn, 6 M Gn-HCl	3	37	24	80
8	$QK^{Su}E$	Zn, 6 M Gn-HCl	3	37	24	75
9	$KAK^{Ub}I$	Zn, 6 M Gn-HCl	3	37	24	80
10	$KAK^{Su}I$	Zn, 6 M Gn-HCl	3	37	24	60

[a]Yields were calculated from purified product weights and relative signal intensities in ESI-MS. [b]As per the procedure reported by McGinty et al.[12]

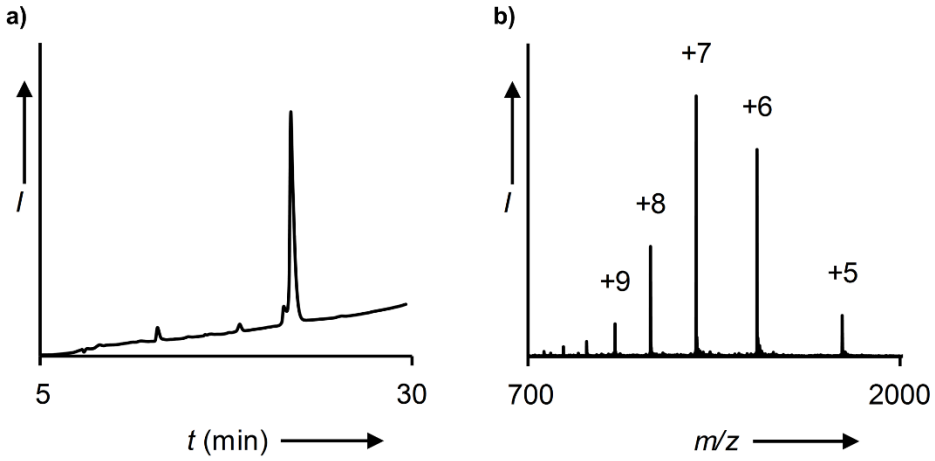


Figure 2.1. Synthesis of $QK^{Ub}E$. a) C18 analytical RP-HPLC chromatogram of reduced and purified $QK^{Ub}E$. a) ESI-MS of reduced $QK^{Ub}E$. Calculated *m/z* [M+H]+ 8,992.0 Da, observed 8,992.2 ± 1.8 Da.

Gn-HCl during the reduction step overcame this issue and led to greater removal of the β-mercaptoethanol group resulting in the native isopeptide linkage at Lys (**Figure 2.1** and **Table 2.3**, entry 3). The requirement for a denaturant to facilitate efficient electron transfer from Zn to the low-energy N-O bond in **4** is consistent with previous biophysical studies which showed that Ub retains its tertiary structure in solution at pH ~1 (**Figure 2.2**).[33]

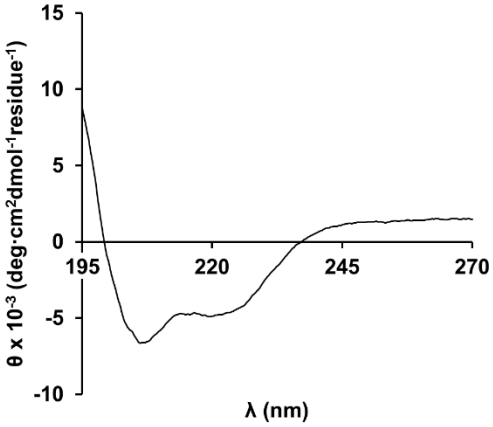


Figure 2.2. Structure of Ub(1-75)-MES (3) in HPLC buffer. Circular Dichroism spectrum of Ub(1-75)-MES (**3**) at 0.1 mg/mL in 50:50 H_2O: CH_3CN, 0.1% TFA (v/v), pH ~2.

We next sought to identify suitable reagents and conditions that maximize the yields of reduced products (**Table 2.3**, entries 4-7). Several methods have been reported for the conversion of *N*,*N*-disubstituted hydroxylamines to the corresponding amines, including reduction with Raney Ni[34] and indium metal.[35] However, we observed that treatment of **4** with activated Zn in 6 M Gn-HCl at pH 3.0 and 37 °C for 24 h yielded the best results with minimal side-products (**Table 2.3**, entry 7). A small amount of hydrolyzed Ub(1-75) was the major side-product observed during auxiliary removal. An *N*-to-*S* acyl shift of the amide backbone that would yield a thioester susceptible to hydrolysis has previously been demonstrated at Gly-Cys junctions[36] and with *N*-methylated Cys at low pH and elevated temperatures.[37] However, for **4** the hydrolysis product is more likely a result of direct nucleophilic attack by water on the protonated tertiary amide. Evidence toward this comes from our observation that alkylation of the auxiliary thiol post-ligation, which precludes an *N*-to-*S* acyl shift, did not significantly affect the small degree of hydrolysis (**Figure 2.3**).

2.2.2 Substrate scope and linkage specificity

With a robust methodology for peptide ubiquitylation in hand we proceeded to test its scope for peptide SUMOylation with a recombinant human SUMO-3(2-91)-α-thioester and the peptide $QK^{(aux)}E$. Ligation proceeded with similar kinetics as for Ub and was complete in 12 h (**Figure 2.4**). Further, alkaline hydrolysis

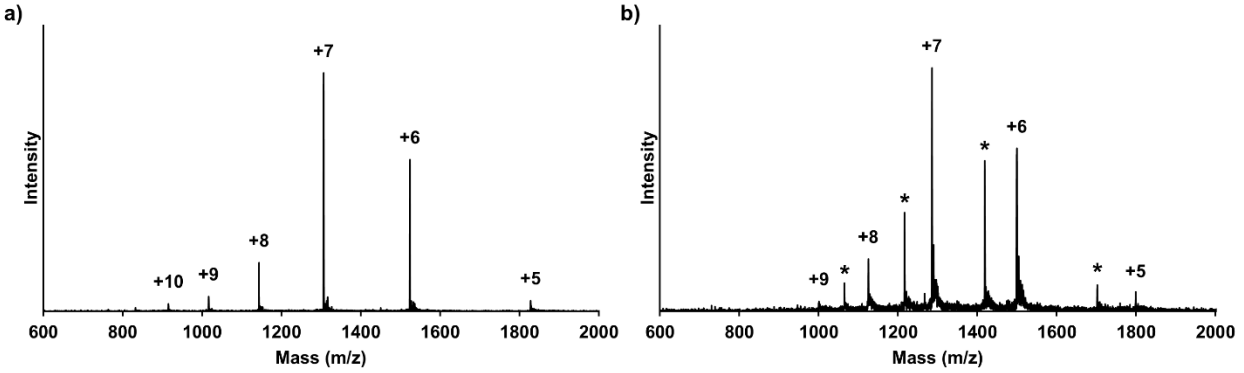


Figure 2.3. Partial hydrolysis of *S*-alkylated QK$^{Ub(aux)}$E under reducing conditions. a) ESI-MS of purified *S*-alkylated QK$^{Ub(aux)}$E. b) ESI-MS of products formed during auxiliary reduction at pH 1.0 for 48 h. Calculated for QK^{Ub}E, *m/z* [M+H]$^+$ 8,992 Da, observed 8,992 Da (major species). Calculated for Ub(1-75)-COOH, *m/z* [M+H]$^+$ 8,507.7 Da and observed 8,508 ± 2 Da. Asterisks indicate the Ub(1-75)-COOH hydrolysis product.

of the ligation product at each time point confirmed that the *S*-to-*N* acyl shift had taken place (**Figure 2.5**). A previously reported ligation of SUMO to a secondary amine required 7 days to achieve 60% yield,[19] which highlights the advantage of employing the aminooxy group for acyl transfer. As expected, reductive removal of the pendant auxiliary group afforded the isopeptide-linked QK^{Su}E in good yield while retaining Cys47 in SUMO (**Table 2.3**, entry 8).

We further explored the substrate scope of ligation with a KAKI peptide sequence, which contains both Lys27 and Lys29 of Ub. Installation of the ligation auxiliary at the Lys29 position in KAKI led to similar yields for ubiquitylation and SUMOylation of this peptide (**Table 2.3**, entries 9 and 10). However, unlike the N-terminally acetylated QKE peptide, the reaction of Ub(1-75)- and SUMO-3(2-91)-α-thioesters with KAKI could proceed through nucleophilic attack by either the peptide N-terminus, the Lys27 ε-amine, or the auxiliary alkoxyamine. In order to test the precise site of ligation, KAK^{Ub}I and KAK^{Su}I were assayed with the ubiquitin C-terminal hydrolase L3 (UCH-L3) and the Sentrin-specific protease 1 (SENP1), which are cysteine proteases that remove Ub and SUMO, respectively, from diverse cellular targets (**Figure 2.6a**).[38,39]

Complete hydrolysis of the isopeptide linkage was observed in both cases and the full-length Ub(1-76) and SUMO-3(2-92) proteins were observed after 3.5 and 8 h, respectively (**Figure 2.6b-c**). As the C-terminal

Gly in the full-length proteins could only be derived from conjugation at Lys29 (**Scheme 2.2**), these results indicated that auxiliary-mediated ligation of Ub and SUMO to the KAKI peptide occurred at the desired Lys29 side-chain. Complete hydrolysis of KAK^{Ub}I by UCH-L3 suggests that our chemical strategy does not interfere with the correct folding of Ub in the final ubiquitylated product. Furthermore, X-ray crystal structures of Ub bound to UCH-L3 reveal extensive protein-protein interactions that may only exist in the native folded form of Ub.[40] Thus, our methodology yields site-specifically ubiquitylated and SUMOylated peptides that adopt a native fold and are suitable for *in vitro* biochemical studies.

2.2.3 Protease resistance of the auxiliary-retaining species

When subjecting semisynthetic ubiquitylated proteins to assays in complex mixtures, such as cell lysates, one must contend with the presence of ~100 deubiquitylating enzymes (DUBs) that may hydrolyze the isopeptide linkage.[41] In thinking of this problem, we noted that retention of the ligation auxiliary in the final ubiquitylated peptide yields an *N*-alkylated Gly76. The enhanced proteolytic stability of polymeric *N*-substituted glycines, also known as peptoids, and peptoid-peptide hybrid molecules has previously been demonstrated.[42] Therefore, we wondered if simply retaining the ligation auxiliary in the final ubiquitylated product would inhibit proteolysis of the isopeptide bond. Satisfyingly, incubation of the unreduced ligation product containing the auxiliary group, QK$^{Ub(aux)}$E, with UCH-L3 for 3.5 h at 37 °C led to no observable hydrolysis of the isopeptide linkage (**Figure 2.7a**, lane 3). In comparison, the reduced ligation product, QK^{Ub}E, was completely hydrolyzed to produce Ub(1-76) under identical conditions (**Figure 2.7a**, lane 1). To test the scope of DUB resistance, we also assayed KAK$^{Ub(aux)}$I with the ubiquitin specific proteases Usp2 and Usp5, which structurally are vastly different from UCH-L3.[43] We observed that the ligation auxiliary also inhibited the removal of Ub from peptide targets by these enzymes (**Figure 2.8**). Recently, *N*-methylation of the isopeptide bond was demonstrated to be a means to inhibit UCH-L3 activity.[44] This required the challenging chemical synthesis of an *N*-methylated branched polypeptide and a final desulfurization step to generate wild-type Ub. We demonstrate that *N*-alkylation of Gly76 offers an alternate and facile means

to access both protease resistant and wild-type isopeptide-linked ubiquitylated peptides from a single synthetic route.

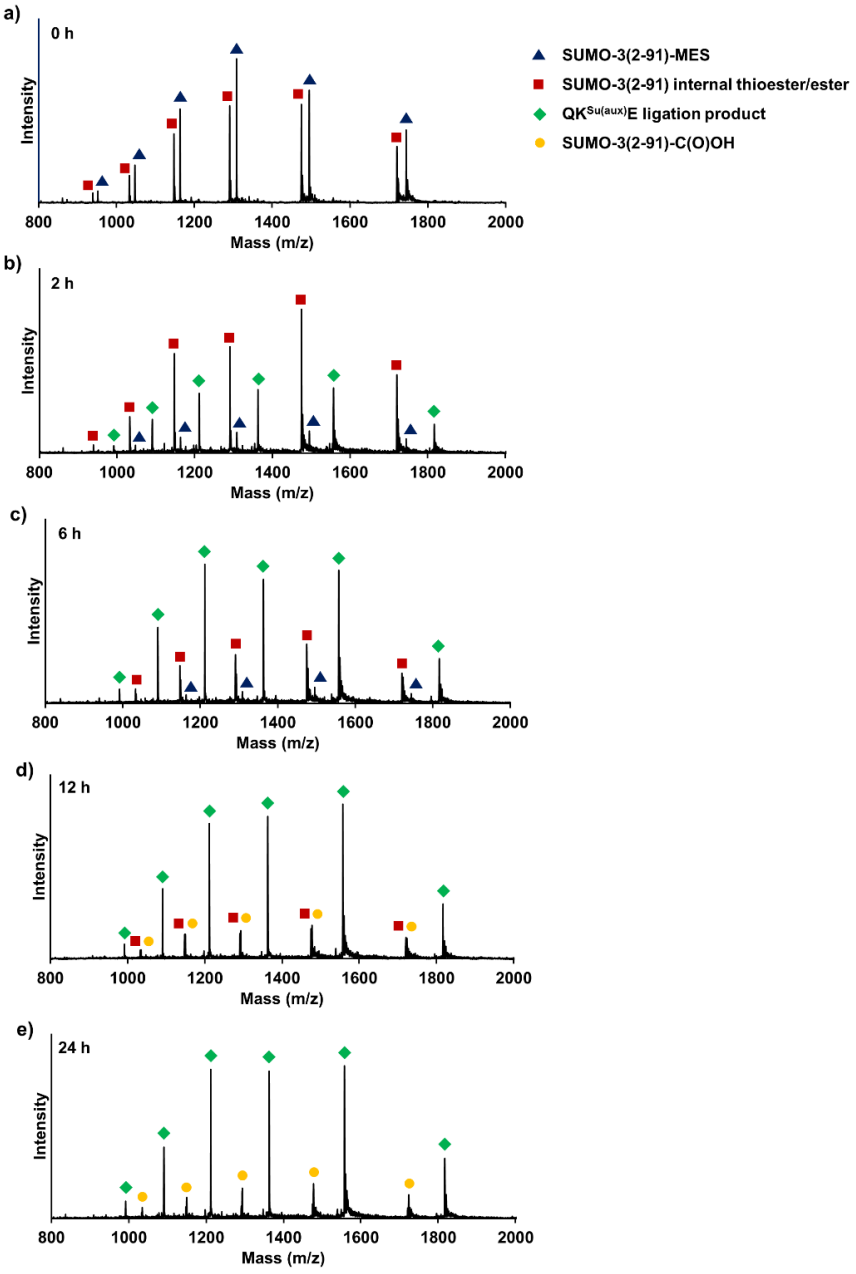


Figure 2.4. Time-course of auxiliary-mediated peptide SUMOylation. LC-ESI-MS analysis of the ligation reaction between QK$^{(aux)}$E (**2**) and the SUMO-3(2-91)-MES α-thioester at a) 0 h, b) 2 h, c) 6 h, d) 12 h, and e) 24 h. Earlier time-points show the formation of an intramolecular ester/thioester of SUMO-3(2-91) that remains competent for ligation with QK$^{(aux)}$E.

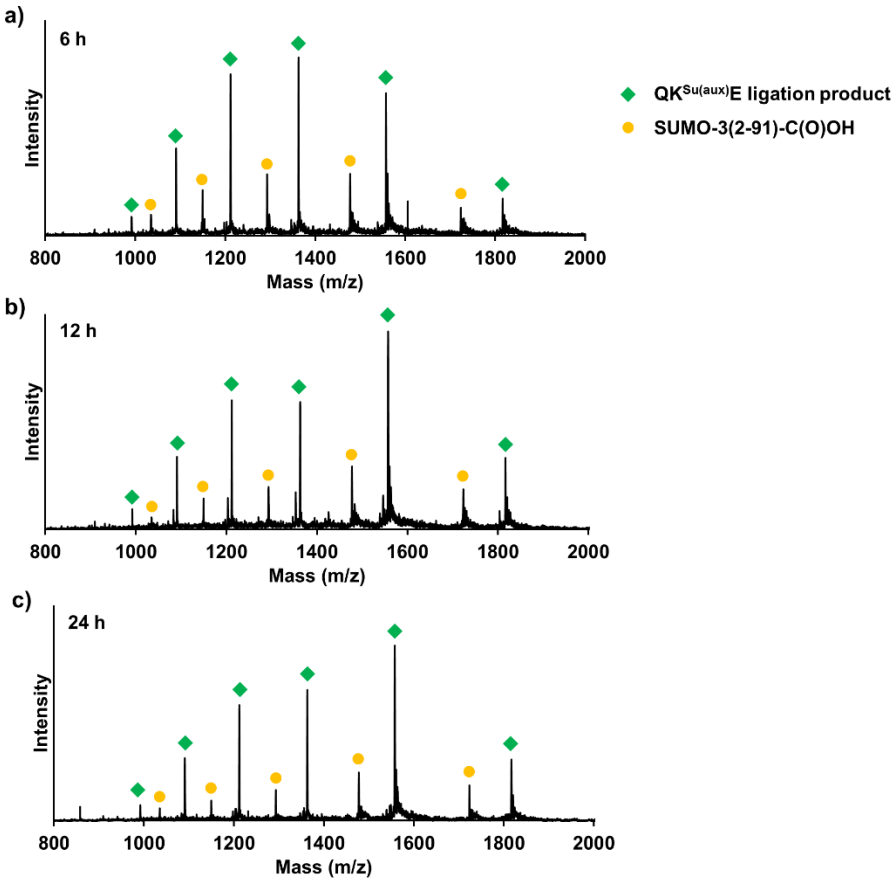


Figure 2.5. Time-course of *S*-to-*N* acyl shift and isopeptide bond formation. LC-ESI-MS analysis of the base hydrolysis products after allowing ligation between QK$^{(aux)}$E (**2**) and the SUMO-3(2-91)-MES α-thioester for a) 6 h, b) 12 h, and c) 24 h.

Interestingly, non-reducing denaturing polyacrylamide gel electrophoresis of the products obtained from incubating $QK^{Ub(aux)}E$ with UCH-L3 in the presence of ~6 mM dithiothreitol (DTT) revealed a new higher molecular weight band (**Figure 2.7a**, lane 3). Reverse phase liquid chromatography of the assay products followed by electrospray ionization mass spectrometry (LC-ESI-MS) revealed a species corresponding to disulfide-linked $QK^{Ub(aux)}E$ and UCH-L3 (**Figure 2.7b**). This higher molecular-weight species was also observed upon incubation of $KAK^{Ub(aux)}I$ with UCH-L3 (**Figure 2.7c**). In keeping with a disulfide-linkage, the adduct was not observed when assay products were boiled with 50 mM DTT prior to gel electrophoresis (**Figure 2.7a**, lane 5), or when the active site mutant UCH-L3(C95A) was employed in assays (**Figure 2.7a**, lane 4). Disulfide formation is however not the sole mechanism for UCH-L3 resistance as unlinked $QK^{Ub(aux)}E$ remained intact while $QK^{Ub}E$ was effectively hydrolyzed. A higher molecular weight band was also observed during the deubiquitylation of $QK^{Ub}E$ by UCH-L3, however, Ub has no intrinsic thiol groups and the mass of this species corresponded to an enzyme-bound Ub(1-76) acyl-intermediate (**Figure 2.9**).

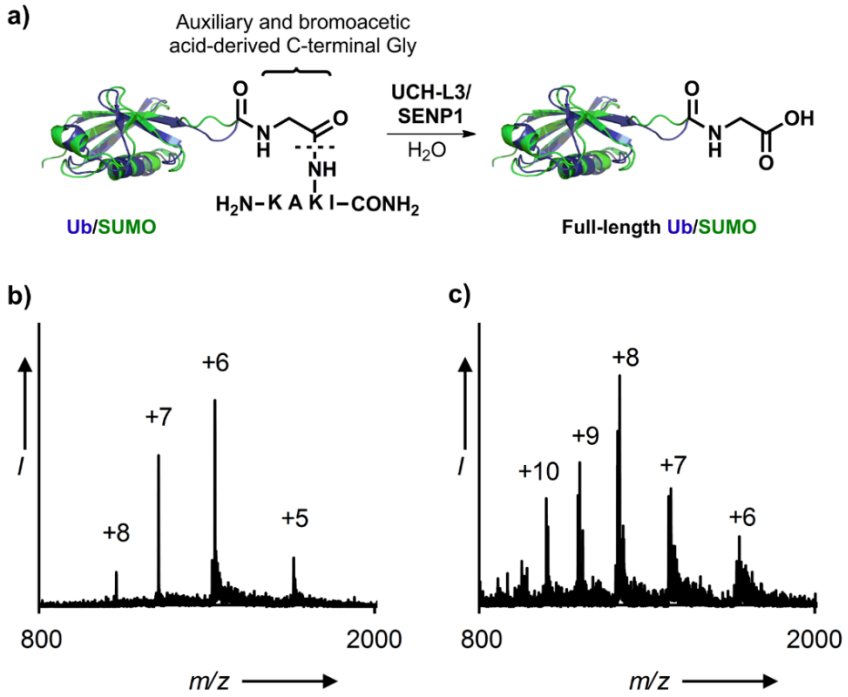


Figure 2.6. Testing the site of Ub and SUMO linkage. a) Hydrolysis of KAK^{Ub}I and KAK^{Su}I by UCH-L3 and SENP1, respectively. b) ESI-MS of KAK^{Ub}I assayed with UCH-L3. Calculated *m/z* $[M+H]^+$ 8,565.8 Da, observed 8,566.4 ± 0.8 Da. c) ESI-MS of KAK^{Su}I assayed with SENP1. Calculated *m/z* $[M+H]^+$ 13,394.6 Da, observed 13,394.6 ± 2.1 Da.

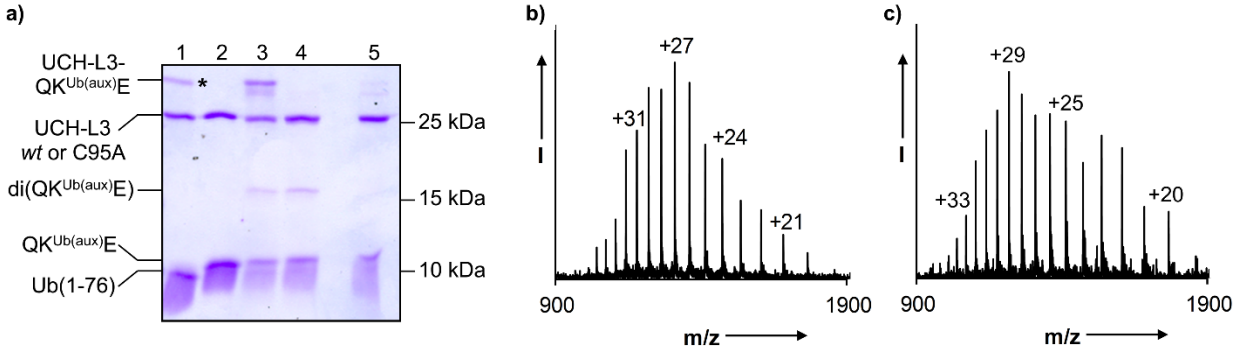


Figure 2.7. Protease-resistant ubiquitylated peptide and DUB adduct formation. a) Coomassie stained 18% SDS-PAGE gel of DUB activity towards QK^{Ub}E and QK$^{Ub(aux)}$E. Lane 1 = UCH-L3 + QK^{Ub}E; lane 2 = UCH-L3(C95A) + QK^{Ub}E; lane 3 = UCH-L3 + QK$^{Ub(aux)}$E; lane 4 = UCH-L3(C95A) + QK$^{Ub(aux)}$E; lane 5 = UCH-L3 + QK$^{Ub(aux)}$E + 50 mM DTT, 95 °C, 5 min. An asterisk indicates the Ub(1-76)-UCH-L3 acyl enzyme intermediate. b) ESI-MS of the observed UCH-L3-QK$^{Ub(aux)}$E adduct. Calculated *m/z* $[M+H]^+$ 35,247.5 Da, observed 35,252.5 ± 5.6 Da. c) ESI-MS of the observed UCH-L3-KAK$^{Ub(aux)}$I adduct. Calculated *m/z* $[M+H]^+$ 35,260.8 Da, observed 35,278.4 ± 16.3 Da.

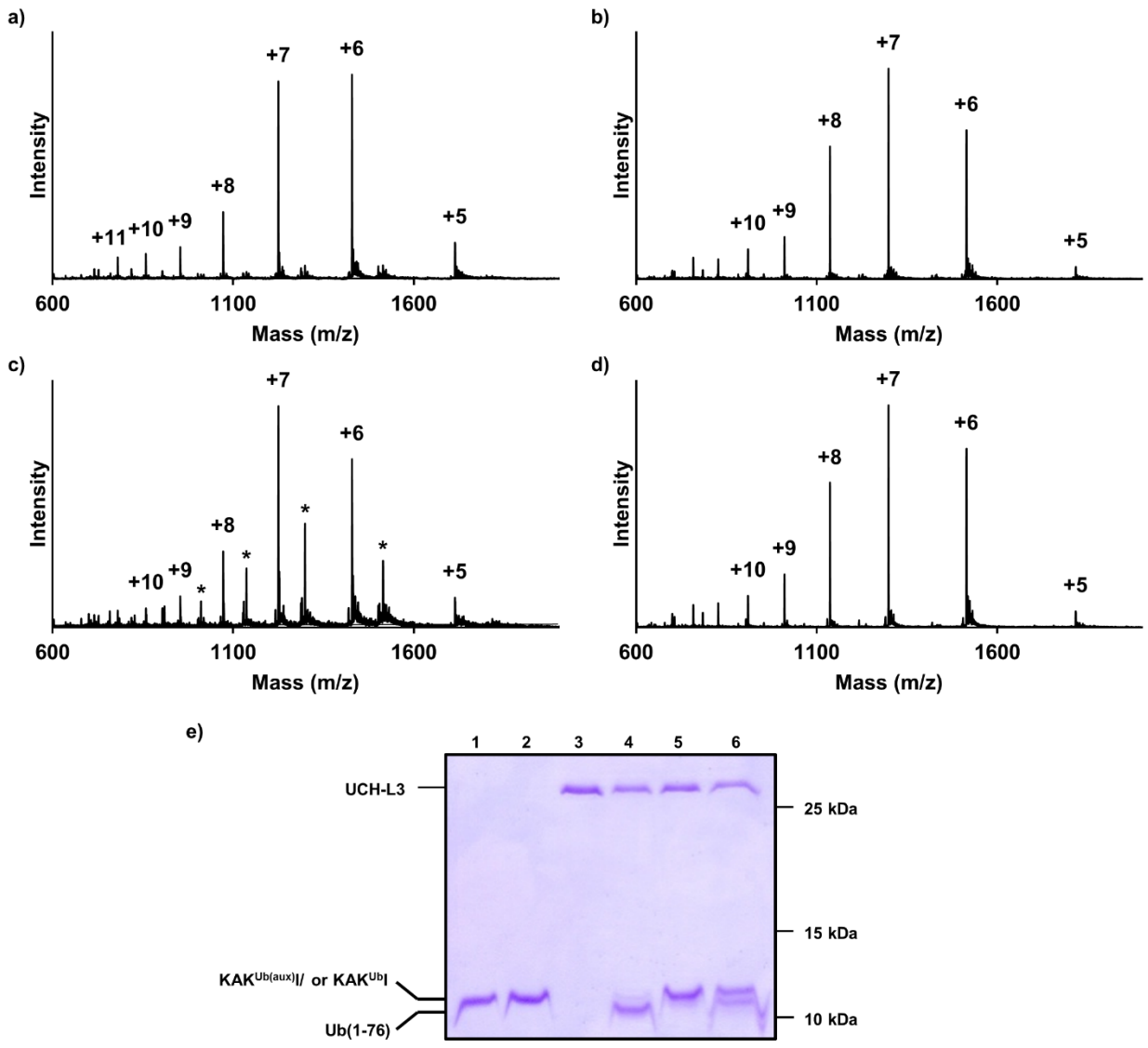


Figure 2.8. DUB resistance of the auxiliary-retaining ligation product. a) LC-ESI-MS of KAK^{Ub}I assayed with Usp2. Calculated for Ub(1-76)-COOH *m/z* [M+H]$^+$ 8,565.8 Da, observed *m/z* [M+H]$^+$ 8,568.4 ± 2.5 Da. b) LC-ESI-MS of KAK$^{Ub(aux)}$I assayed with Usp2. Calculated for KAK$^{Ub(aux)}$I *m/z* [M+H]$^+$ 9,081.1 Da, observed *m/z* [M+H]$^+$ 9,084.8 ± 2.6 Da. c) LC-ESI-MS of KAK^{Ub}I assayed with Usp5/IsoT. Calculated for Ub(1-76)-COOH *m/z* [M+H]$^+$ 8,565.8 Da, observed *m/z* [M+H]$^+$ 8,568.4 ± 3.0 Da. Asterisks indicate unreacted KAK^{Ub}I. d) LC-ESI-MS of KAK$^{Ub(aux)}$I assayed with Usp5/IsoT. Calculated for KAK$^{Ub(aux)}$I *m/z* [M+H]$^+$ 9,081.1 Da, observed *m/z* [M+H]$^+$ 9,084.9 ± 3.1 Da. e) Coomassie stained 15% SDS-PAGE gel of UCH-L3 activity towards KAK^{Ub}I and KAK$^{Ub(aux)}$I. Lane 1 = pure KAK^{Ub}I; lane 2 = pure KAK$^{Ub(aux)}$I; lane 3 = pure UCH-L3; lane 4 = UCH-L3 + KAK^{Ub}I; lane 5 = UCH-L3 + KAK$^{Ub(aux)}$I; lane 6= UCH-L3 + KAK^{Ub}I + KAK$^{Ub(aux)}$I. All assays were undertaken for 6 h at 37 °C. For lane 6, KAK^{Ub}I was added after 3 h of UCH-L3 incubation with KAK$^{Ub(aux)}$I followed by an additional 3 h of incubation at 37 °C.

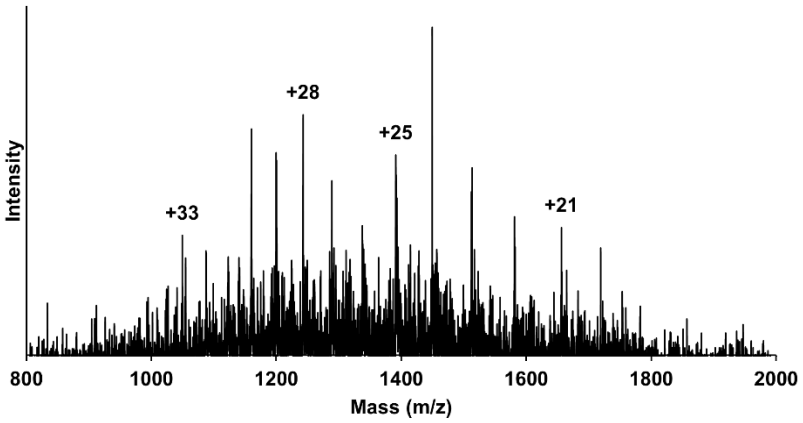


Figure 2.9. Detection of a Ub(1-76) acyl-UCH-L3 intermediate. LC-ESI-MS of QK^{Ub}E assay with UCH-L3. Calculated *m/z* $[M+2(O)+H]^+$ 34,766.6 Da, found 34,768.6 ± 10.1 Da.

2.3 Conclusion and outlook

We have demonstrated the successful application of a synthetically facile and readily removable ligation auxiliary toward the ubiquitylation and SUMOylation of various peptide targets. Our methodology significantly advances current ubiquitylation strategies, as it is orthogonal to Cys residues in proteins. When employed in combination with native chemical ligation at Cys this will greatly expand the substrate scope of protein targets. Furthermore, post-ligation the auxiliary mimics an *N*-substituted Gly76 and inhibits several deubiquitylating enzymes. Our preliminary results also indicate that the auxiliary group has utility as a probe in pull-down assays to identify ubiquitin C-terminal hydrolases. Finally, the mild conditions for removal of the auxiliary after ligation do not inhibit protein recognition by specific proteases such as UCH-L3 and SENP1, which is promising for studies of the biochemical effects of ubiquitylation and SUMOylation. However, use of denaturant in the reductive auxiliary removal step may be prohibitive for some proteins. To overcome this limitation, in Chapter 3 I describe an alternative method for auxiliary removal that permits one-pot ligation and auxiliary removal under non-denaturing conditions.

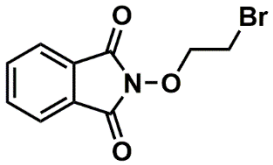
2.4 Experimental procedures

2.4.1 General methods

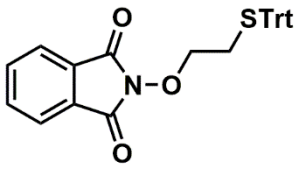
Rink-amide resin (0.72 mmol/g substitution) was purchased from Chem-Impex (Wood Dale, IL). Standard Fmoc-L-amino acids were purchased from AGTC Bioproducts (Wilmington, MA) or AnaSpec (Fremont, CA). All other chemical reagents were purchased from Sigma-Aldrich Chemical Company (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). DNA synthesis and gene sequencing were performed by Integrated DNA Technologies (Coralville, IA) and Genewiz (South Plainfeld, NJ), respectively. Plasmid min-prep, PCR purification and gel extraction kits were purchased from Qiagen (Valencia, CA). Restriction enzymes were purchased from New England BioLabs (Ipswitch, MA) or Fermentas (Thermo Fisher Scientific, Philadelphia, PA). Chitin beads for purification of intein-CBD fusion proteins were purchased from New England BioLabs. Ni-NTA resin for purification of His_6-UCH-L3 and His_6-UCH-L3(C95A) was purchased from Thermo Scientific (Waltham, MA). Analytical reversed-phase HPLC (RP-HPLC) was performed on a Varian (Palo Alto, CA) ProStar HPLC with a Grace (Deerfield, IL) C18 column (5 micron, 150 x 4.6 mm) employing 0.1% TFA in water (A) and 90% CH_3CN, 0.1% TFA in water (B) as the mobile phases. Typical analytical gradients were 0-73% B over 30 min at a flow rate of 1 mL/min. Preparative scale purifications were conducted on a Grace-Vydac C18 column (10 micron, 250 x 22 mm) at a flow rate of 9 mL/min. Mass spectrometric analysis was conducted on a Bruker (Billerica, MA) Esquire ESI-MS instrument. Analytical reversed-phase liquid chromatography-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) employing 5% CH_3CN, 1% AcOH in water (C) and CH_3CN, 1% AcOH (D) as the mobile phases. Typical analytical gradients were 0-50% D over 40 min at a flow rate of 0.2 mL/min. NMR spectra were recorded on Bruker Avance AV-300, AV-301, AV-500, or DRX-499 instruments. Circular dichroism measurements were performed on a JASCO (Easton, MD) J-720 spectropolarimeter.

2.4.2 Synthesis of the ligation auxiliary

Ligation auxiliary **1** was prepared over 3 steps from *N*-hydroxyphthalimide following previous reports, with minor modifications.[28]

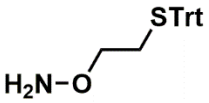


***N*-(2-bromoethoxy)phthalimide.** *N*-(2-bromoethoxy)phthalimide was synthesized as previously reported.[27] Triethylamine (10.7 mL, 76.7 mmol) was added dropwise to a stirring solution of *N*-hydroxyphthalimide (0.85 M in DMF, 41 mL, 34.9 mmol). To this solution 1,2-dibromoethane (12.6 mL, 146.2 mmol) was added dropwise. The resulting mixture was stirred at room temperature protected from light. Reaction progress was monitored by silica gel TLC. After 18 h, solids were filtered and washed with DMF. Product was precipitated from the filtrate by the addition of 350 mL water, vacuum-filtered and washed with water. The solids were dissolved in 200 mL ethyl acetate and washed with 1 N HCl (2 x 100 mL), water (1 x 100 mL), and saturated NaCl (1 x 100 mL), and dried over anhydrous $MgSO_4$. Volatiles were removed *in vacuo*. The resulting solid was recrystallized from 95% EtOH to give *N*-(2-bromoethoxy)phthalimide as white, needle-like crystals (5.7 g, 64%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.81 (m, 4H), 4.47 (t, 2H, *J* = 6.88 Hz), 3.64 (t, 2H, *J* = 6.88 Hz) (**Figure 2.10**). ^{13}C NMR (499 MHz, $CDCl_3$): δ 163.49, 134.81, 128.79, 123.78, 77.27, 26.89 (**Figure 2.11**). ESI-MS calculated *m/z* $[M+Na]^+$ 291.96 Da, observed 293.1 Da.



***N*-(2-(tritylthio)ethoxy)phthalimide.** Sodium hydride (0.133 g, 5.54 mmol) was added to a solution of triphenylmethanethiol (1.23 g, 4.45 mmol) in anhydrous DMF (18 mL) under N_2 and stirred at room temperature until bubbling was no longer observed. This solution was then added to a stirring mixture of *N*-(2-bromoethoxy)phthalimide (0.2 M in anhydrous DMF, 18 mL, 3.70 mmol) and stirred at room temperature under N_2 until no further change was observed by TLC. A second, identical sodium hydride and triphenylmethanethiol solution was added to the reaction mixture, and the reaction proceeded until TLC revealed that starting material had been consumed (~1.5 h). The reaction mixture was quenched with water until no further bubbling was observed. Precipitate was removed by vacuum filtration and the filtrate concentrated under high vacuum. The residue

was purified by silica gel (70-230 mesh) column chromatography (90:10 hexane: ethyl acetate) to give compound *N*-(2-(tritylthio)ethoxy)phthalimide (1.43 g, 79%). ^{1}H NMR (301 MHz, DMF-D_7): δ 7.89 (m, 4H), 7.44-7.16 (15H), 3.83 (t, 2H, *J* = 7.30 Hz), 2.65 (t, 2H, *J* = 7.30 Hz) (**Figure 2.12**). ^{13}C NMR (499 MHz, DMF-D_7): δ 163.51, 144.90, 135.10, 129.72, 129.17, 128.34, 127.16, 123.48, 76.45, 67.23, 30.11 (**Figure 2.13**). ESI-MS calculated *m/z* $[M+Na]^+$ 488.13 Da, observed 488.2 Da.



***O*-(2-(tritylthio)ethyl)hydroxylamine (1).** Hydrazine hydrate (0.3 mL, 4.81 mmol) was added to a stirring solution of *N*-(2-(tritylthio)ethoxy)phthalimide (58.9 mM in $CHCl_3$, 27 mL, 1.59 mmol). The reaction proceeded at room temperature for 2 h with vigorous stirring. Reaction progress was monitored by silica gel TLC. After 2 h, solids formed were filtered and rinsed with $CHCl_3$. The filtrate was washed with basified water (pH 8, 3 x 30 mL) and saturated NaCl (1 x 30 mL), and dried over anhydrous $MgSO_4$. Solvent was removed *in vacuo* to yield compound **1** (0.48 g, 90%). ^{1}H NMR (301 MHz, $CDCl_3$): δ 7.41-7.03 (15H), 5.23 (s, 2H), 3.46 (t, 2H, *J* = 6.32 Hz), 2.36 (t, 2H, *J* = 6.32 Hz) (**Figure 2.14**). ^{13}C NMR (500 MHz, DMSO-D_6): δ 144.49, 129.10, 128.01, 126.70, 72.59, 65.93, 31.36 (**Figure 2.15**). Due to the poor ionization of **1** during ESI-MS it was converted to the acetamide derivative. The mass of the derivatized compound was confirmed by ESI-MS, calculated *m/z* $[M+Na]^+$ 415.5 Da, observed 415.2 Da.

2.4.3 Solid-phase peptide synthesis

Synthesis of Ac-QKE-Rink-amide resin

The peptide Ac-QKE-$CONH_2$ was manually synthesized on a 0.5 mmol scale employing standard 9-fluorenylmethoxycarbonyl (Fmoc)-based N^{α}-deprotection chemistry. From Rink-amide resin (0.7 g, 0.72 mmol/g) each amino acid was coupled in 4 molar excess based on resin loading. Deprotection of the Fmoc group was achieved with 20% piperidine in DMF (5 mL, 3 x 10 min). Coupling reactions were undertaken for a minimum of 1 hour with a mixture of Fmoc-amino acid (2 mmol), *O*-(benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HBTU, 1.97 mmol) and *N,N*-diisopropylethylamine (DIEA, 4 mmol) in DMF. Coupling efficiency was monitored by the Kaiser test[45] and additional couplings performed

until a negative test was obtained. Lys was orthogonally protected at the ε-NH_2 position with the 4-methyltrityl (Mtt) group. Following Fmoc-deprotection, the N-terminal Gln was acetylated by reaction with DIEA (20-fold molar excess) and acetic anhydride (40-fold molar excess) for 3 x 15 min. A test portion of the resin was cleaved with a cocktail consisting of trifluoroacetic acid (TFA): triisopropylsilane (TIS): H_2O (95:2.5:2.5 v/v) for 1 hour, then precipitated and washed twice with cold diethyl ether. The peptide was purified by C18 analytical RP-HPLC with a gradient of 0-40% B over 30 min, and characterized by ESI-MS. Calculated *m/z* $[M+H]^+$ 445.2 Da, observed 445.4 Da.

Synthesis of Fmoc-KAKI-Rink-amide resin

The peptide H_2N-KAKI-$CONH_2$ was synthesized similarly to Ac-QKE-$CONH_2$. The N-terminal Lys was protected at the ε-NH_2 position with the *tert*-butoxycarbonyl (Boc) group. The interior Lys was orthogonally protected with the Mtt protecting group. Prior to Fmoc deprotection, a test portion of the resin was cleaved with a cocktail consisting of TFA: TIS: H_2O (95:2.5:2.5 v/v) for 1 hour, then precipitated and washed 2 times with cold diethyl ether. The peptide was purified by C18 analytical RP-HPLC with a gradient of 0-73% B over 30 min, and characterized by ESI-MS. Calculated *m/z* $[M+H]^+$ 680.4 Da, observed 680.5 Da.

Attachment of the ligation auxiliary

Resin bound peptides were each treated with a solution of 1% TFA, 1% TIS in DCM to deprotect the Mtt protecting group and expose the Lys ε-amine. The peptidyl resin was then coupled to bromoacetic acid (8-fold molar excess) with *N,N'*-Diisopropylcarbodiimide (DIC, 8-fold molar excess) in DMF for 45 min at room temperature. The coupling was repeated once and a test portion of the resin was cleaved to confirm coupling. Subsequently, dry peptidyl resin was placed in a solution containing 9 equivalents of auxiliary **1** (1 M in DMSO) and shaken for 24 hours at room temperature. Completion of the displacement was judged by test cleavages and subsequent ESI-MS analysis. The auxiliary solution was then washed from the resin with DMF, lyophilized to remove DMF and stored in DMSO at -80 ºC for re-use. In the case of KAKI, the N-terminal Fmoc was then deprotected with 20% piperidine in DMF (5 mL, 2 x 15 min). Resin was reacted at

20 µL/mg with Reagent K (TFA: thioanisole: H_2O: phenol: 1,2-ethanedithiol 82.5:5:5:5:2.5 v/v)[46] for 1.5 hours at room temperature, then precipitated and washed 2 times with cold diethyl ether. Dry peptides were dissolved in RP-HPLC buffer A and purified by C18 analytical RP-HPLC with a gradient of 0-40% B over 30 min. This yielded 38% and 42% of the peptide-auxiliary conjugates Ac-QK$^{(aux)}$E-$CONH_2$ (**2**) and H_2N-KAK$^{(aux)}$I-$CONH_2$, respectively (**Figure 2.16**).

ESI-MS of Bromoacetylated Ac-QKE-$CONH_2$. Calculated *m/z* $[M+H]^+$ 565.2 Da, observed 565.7 Da.

ESI-MS of Bromoacetylated Fmoc-KAKI-$CONH_2$. Calculated *m/z* $[M+H]^+$ 800.3 Da, observed 802.4 Da.

ESI-MS of Ac-QK$^{(aux)}$E-$CONH_2$ (**2**). Calculated *m/z* $[M+H]^+$ 578.3 Da, observed 578.4 Da.

ESI-MS of H_2N-KAK$^{(aux)}$I-$CONH_2$. Calculated *m/z* $[M+H]^+$ 591.4 Da, observed 591.6 Da.

2.4.4 Molecular cloning of Ub(1-75) and SUMO-3(1-91)

The plasmid pUb(1-75) containing the partial human ubiquitin gene, *ub(1-75)*, was a kind gift from the Muir lab at Princeton University.[19] The human SUMO-3 gene, *Smt3(1-92)*, was kindly provided by Dr. Brian Houck-Loomis at Rockefeller University. This was cloned upstream of the *intein-cbd* gene in pTXB1 to obtain the plasmid pSUMO-3(1-92). The plasmid pSUMO-3(1-91), which lacks the terminal Gly of SUMO-3, was prepared from this template by site-directed mutagenesis (QuikChange kit, Agilent Technologies, Santa Clara, CA) with the following primers; Forward: 5'-ATC GAC GTG TTC CAG CAG CAG ACG GGA TGC ATC ACG GGA GAT GCA CTA GTT GCC-3' and Reverse: 5'-GGC AAC TAG TGC ATC TCC CGT GAT GCA TCC CGT CTG CTG CTG GAA CAC GTC GAT-3'. The desired gene sequences were confirmed by sequencing with the T7 forward primer (Genewiz).

2.4.5 Overexpression and purification of Ub(1-75)-MES (3) and SUMO-3(2-91)-MES

E. coli BL21(DE3) cells were transformed with the plasmids pUb(1-75) and pSUMO3(1-91). Cells were grown in 6 L Luria-Bertani medium supplemented with 100 µg/mL of Ampicillin at 37 ºC with shaking at 250 rpm until OD_{600} ~0.6-0.8. Overexpression of the desired fusion proteins was induced by the addition of 0.3 mM isopropyl *β*-D-1-thiogalactopyranoside (IPTG) and cells were grown for an additional 6 h at 25 ºC. The

cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in lysis buffer: PBS, pH 7.2 containing 1 mM 2-mercaptoethanesulfonic acid sodium salt (MESNa). Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The lysate supernatant was passed through a 0.45 µm filter then applied to a 30 mL chitin column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 12 h at 4 ºC. The column was then washed with 20 column volumes (CV) of lysis buffer followed by 2 CV of PBS, pH 7.75. Ub(1-75)-MESNa and SUMO-3(2-91)-MESNa were cleaved from their respective intein-CBD fusions by incubation with 1.5 CV of PBS, pH 7.75 containing 100 mM MESNa for 72 h at 4 ºC. The eluted Ub(1-75)- and SUMO-3(2-91)-α-thioesters were purified by C18 preparative RP-HPLC employing a gradient of 30-60% B over 60 min. Fractions containing the desired thioesters were identified by ESI-MS (**Figure 2.17**). We observed that the N-terminal Met of SUMO-3 (but not Ub) is consistently processed *in vivo*, leading to the SUMO-3(2-91)-α-thioester product.

ESI-MS for Ub(1-75)-MES (**3**). Calculated *m/z* $[M+H]^+$ 8,632.8 Da, observed 8,633.5 ± 1.8 Da.

ESI-MS for SUMO-3(2-91)-MES. Calculated *m/z* $[M+H]^+$ 10,461.6 Da, observed 10,461.3 ± 1.4 Da.

2.4.6 Expressed protein ligation of auxiliary-containing peptides and protein α-thioesters

In a typical small-scale reaction, the purified auxiliary-conjugated peptide **2** (0.35 mg, 0.6 µmol) and protein thioester **3** (1.0 mg, 0.12 µmol) were separately dissolved in 120 µL of a buffer consisting of 6 M Gn-HCl, 0.1 M Na_2HPO_4, and 10 mM TCEP, pH 7.5. The two solutions were mixed and ligation allowed with gentle shaking at room temperature for 24 h. The kinetics of ligation were monitored by withdrawing 3 µL aliquots from the reaction mixture at specific time-points, diluting into 27 µL H_2O, 0.1% formic acid, and analyzing by LC-ESI-MS. The extent of the *S*-to-*N* acyl shift leading to the desired amide-linked product was also monitored at each time-point by withdrawing 3 µL aliquots from the reaction mixture, adding 3 µL 1 M NaOH, and incubating on ice for 30 sec. These conditions led to the hydrolysis of non-rearranged thioester-linked ligation products. The hydrolysis reaction was quenched with 25 µL H_2O, 0.1% formic acid, and analyzed by LC-ESI-MS employing a gradient of 0-50% D, 40 min. The ratio of peak heights in ESI-MS of the hydrolyzed Ub(1-75)-COOH product to the desired ligation product was considered an indicator of *S*-to-*N*

acyl rearrangement. The final rearranged ligation product was purified by C18 analytical RP-HPLC employing a gradient of 30-60% B over 30 min to yield 0.8 mg of **4** in 80% overall yield. In each ligation, the conversion to product was greater than 70%, with the major side product being the hydrolyzed Ub(1-75)-COOH, or SUMO-3(2-91)-COOH protein (**Figure 2.18**).

ESI-MS of $QK^{Ub(aux)}E$ (**4**). Calculated *m/z* $[M+H]^+$ 9,068.0 Da, observed 9,068.0 ± 1.4 Da.

ESI-MS of $KAK^{Ub(aux)}I$. Calculated *m/z* $[M+H]^+$ 9,081.1 Da, observed 9,080.9 ± 1.6 Da.

ESI-MS of $QK^{Su(aux)}E$. Calculated *m/z* $[M+H]^+$ 10,896.8 Da, observed 10,896.2 ± 0.4 Da.

ESI-MS of $KAK^{Su(aux)}I$. Calculated *m/z* $[M+H]^+$ 10,909.9 Da, observed 10,909.4 ± 1.3 Da.

2.4.7 Auxiliary removal by reductive cleavage of the *N-O* bond (Table 2.4)

Metallic Zn/In in acidified Gn-HCl solutions

In a typical reaction, 5 mg of freshly activated metallic Zn[47] (or In)[35] was added to 200 µL of degassed 6 M Gn-HCl, pH 3 containing 0.2 mg purified ligation product. Degassing was accomplished by 3 freeze-thaw cycles under nitrogen, and was found necessary to reduce the generation of oxidized side products. The reduction proceeded at 37 ºC under nitrogen with gentle shaking for 24 h. After this time, the reaction mixture was briefly centrifuged at 13,000 rpm to pellet the Zn (or In) and the supernatant containing reduced products was removed. The pelleted Zn (or In) was washed twice with 150 µL of 6 M Gn-HCl, 50 mM EDTA, pH 3. The combined supernatant and washes were reduced with 10 mM TCEP at pH 7.5 and subsequently purified by C18 analytical RP-HPLC with a gradient of 30-60% B over 30 min (**Figure 2.19**).

ESI-MS of $QK^{Ub}E$. Calculated *m/z* $[M+H]^+$ 8,992.0 Da, observed 8,992.2 ± 1.8 Da.

ESI-MS of $KAK^{Ub}I$. Calculated *m/z* $[M+H]^+$ 9,005.1 Da, observed 9,005.7 ± 1.9 Da.

ESI-MS of $QK^{Su}E$. Calculated *m/z* $[M+H]^+$ 10,820.8 Da, observed 10,821.4 ± 2.5 Da.

ESI-MS of $KAK^{Su}I$. Calculated *m/z* $[M+H]^+$ 10,833.9 Da, observed 10,833.0 ± 1.9 Da.

Zn in acidified HPLC solvents or water

Following a previous report,[27] 5 mg of freshly activated metallic Zn was added to 200 µL of degassed 50:50 H_2O: CH_3CN, 0.1% TFA, pH 1, or degassed H_2O, 0.1% TFA, pH 1 containing 0.2 mg purified $QK^{Ub(aux)}E$ (**4**). The reaction proceeded at room temperature under nitrogen with gentle shaking for 24 hours. After this period the reaction mixture was briefly centrifuged at 13,000 rpm to pellet the Zn and the supernatant containing reduced products was removed. The pelleted Zn was washed twice with 150 µL of 6 M Gn-HCl, 50 mM EDTA, pH 7-8. The combined supernatant and washes were subsequently analyzed by C18 analytical RP-HPLC with a gradient of 30-60% B over 30 min followed by ESI-MS.

Raney nickel

Freshly-prepared Raney nickel[12] was added to a 560 µL solution of 6 M Gn-HCl, 0.2 M Na_2HPO_4, 35 mM TCEP, pH 7 containing 0.2 mg $QK^{Ub(aux)}E$ (**4**). The reaction proceeded with gentle shaking at room temperature for 6.5 h at which point a second equal amount of fresh Raney nickel was added. The reaction proceeded for an additional 1.5 h, for a total reaction time of 8 hours. The reaction mixture was then briefly centrifuged at 13,000 rpm to pellet the Ni. The supernatant was removed and the Ni was washed 2 times with 0.75 mL of a solution containing 6 M Gn-HCl, 0.2 M Na_2HPO_4, 35 mM TCEP, pH 7. The combined supernatant and washes were analyzed by LC-ESI-MS employing a gradient of 0-50% D, 40 min.

2.4.8 *S*-alkylation of $QK^{Ub(aux)}E$

Alkylation of the pendant thiol in the ligation product $QK^{Ub(aux)}E$ was performed under conditions established for selectively alkylating protein sulfhydryl groups.[48] Briefly, 0.4 mg of the ligation product was dissolved in 93 µL of alkylation buffer containing 1 M HEPES, 4 M Gn-HCl, and 10 mM methionine at pH 7.8. To this was added 2 µL of 1 M DTT in alkylating buffer. The resulting mixture was incubated at 37 °C for 20 min after which time 5 µL of a 1 M solution of *N*-(2-chloroethyl)-*N,N*-dimethylammonium chloride in alkylation buffer was added. The reaction was incubated at 25 °C for 2 h. After this, an additional 1 µL of 1 M DTT solution was added followed by 30 min incubation at 25 °C to reduce any residual disulfides. Finally, an additional 5 µL of *N*-(2-chloroethyl)-*N,N*-dimethylammonium chloride in alkylation buffer was added and the

reaction incubated for 2 h to ensure complete alkylation of all thiol groups. The reaction mixture was quenched with 5 µL of β-mercaptoethanol and purified by C18 analytical RP-HPLC employing a gradient of 30-60% B over 30 min. *S*-alkylated QK$^{ub(aux)}$E was characterized by ESI-MS. Calculated *m/z* $[M+H]^+$ 9,139.2 Da, observed 9,139.9 ± 1.9 Da.

2.4.9 Hydrolysis of *S*-alkylated QK$^{Ub(aux)}$E

To test the mechanism of hydrolysis of ligation products during auxiliary removal, 0.2 mg of the *S*-alkylated QK$^{Ub(aux)}$E peptide was dissolved in 200 µL of 6 M Gn-HCl at pH 1 with 6.5 mg of activated Zn. The reaction was incubated at 37 °C for 48 h to test for hydrolysis arising from direct nucleophilic attack by water. After 48 h, the reaction was purified by C18 analytical RP-HPLC employing a gradient of 30-60% B over 30 min, and characterized by ESI-MS.

2.4.10 UCH-L3 and SENP1 hydrolysis assays

Pure Ub or SUMO-conjugated peptides were assayed with recombinant human ubiquitin C-terminal hydrolase L3 (UCH-L3, Boston Biochem, Boston, MA) or the catalytic domain of sentrin-specific protease 1 (SENP1, Boston Biochem), respectively. In a typical reaction, 0.42 nmol of UCH-L3 or 0.25 nmol of SENP1 was pre-activated in 25 µL buffer containing 50 mM Tris, 150 mM NaCl, 12 mM DTT, pH 8 for 20 min at 25 °C. To the reduced activated enzyme were then added 25 µL of a solution containing 1.68 nmol of the ubiquitylated or SUMOylated peptide in 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.5. The resulting mixture was incubated for 3.5 h (UCH-L3) or 8 h (SENP1) at 37 °C. Assays were quenched by the addition of 1 µL formic acid and analyzed by LC-ESI-MS employing a gradient of 0-75% D, 40 min (**Figure 2.20**).

2.4.11 Molecular cloning of UCH-L3 and UCH-L3(C95A)

The wild-type *uchL3* gene was obtained by PCR from Addgene plasmid 22564[49] with the following primers; Forward: 5'-GGG AAT TCC ATA TGG AGG GTC AAC GCT GGC TGC CGC TG-3' and Reverse: 5'-GGG AAT TCC TCG AGC TAT GCT GCA GAA AGA GCA ATC GCA TT-3'. The PCR product was cloned with

an N-terminal hexahistidine tag (His_6-) between the NdeI and XhoI restriction sites in the pET15b vector to obtain the plasmid pUCH-L3. The UCH-L3(C95A) mutant was generated by site-directed mutagenesis of pUCH-L3 with the primers; Forward: 5'-TTC ATG AAG CAA ACA ATC AGC AAT GCC GCC GGA ACA ATT GGA CTG ATT CAT GCT ATT-3' and Reverse: 5'-AAT AGC ATG AAT CAG TCC AAT TGT TCC GGC GGC ATT GCT GAT TGT TTG CTT CAT GAA-3' to generate the plasmid pUCH-L3(C95A). The desired gene sequences were confirmed by sequencing with the T7 forward primer (Genewiz).

2.4.12 Overexpression and purification of His_6-UCH-L3 and His_6-UCH-L3(C95A)

E. coli BL21(DE3) cells transformed with the plasmids pUCH-L3 or pUCH-L3(C95A) were grown in 1 L Luria-Bertani medium supplemented with 100 µg/mL Ampicillin at 37 ºC with shaking at 250 rpm until OD_{600} ~0.6. Overexpression of UCH-L3 or UCH-L3(C95A) was induced by the addition of 0.3 mM IPTG and growing for an additional 5 h at 25 ºC. The cells were harvested by centrifugation at 7,000x*g* for 15 min and the cell pellet was resuspended in a lysis buffer containing 50 mM Tris, 150 mM NaCl, and 10% glycerol at pH 7.2. The cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The supernatant was filtered through a 0.45 µm membrane and then applied to a 2 mL Ni^{2+}-column pre-equilibrated with lysis buffer. Protein binding was achieved by flowing the lysate supernatant through the column thrice at 0.5 mL/min. The column was the washed with 60 column volumes of lysis buffer. Proteins were eluted from the column with lysis buffer containing 500 mM imidazole and fractions containing UCH-L3/UCH-L3(C95A) were combined and concentrated using 10 kDa MWCO centrifugal filters (Sartorius). Stock enzyme concentrations were determined by comparison with known BSA standards from coomassie-stained SDS-PAGE gels. Structure was confirmed by circular dichroism (**Figure 2.21**).

ESI-MS for His_6-UCH-L3. Calculated *m/z* $[M+H]^+$ 28,215.7 Da, found 28,215.6 ± 5.5 Da.

ESI-MS for His_6-UCH-L3(C95A). Calculated *m/z* $[M+H]^+$ 28,183.6 Da, found 28,189.1 ± 6.1 Da.

2.4.13 Assays to test DUB resistance

Enzymatic assays were conducted with purified recombinant human His_6-UCH-L3, His_6-UCH-L3(C95A), and the commercially sourced catalytic domain of ubiquitin specific peptidase 2 (Usp2, Boston Biochem) and ubiquitin specific peptidase 5/isopeptidase T (Usp5/IsoT, Boston Biochem). The substrates employed were $QK^{Ub}E$, $KAK^{Ub}I$, $QK^{Ub(aux)}E$, and $KAK^{Ub(aux)}I$. Assays were performed identically as the previously described UCH-L3 hydrolysis assay. Reaction products were analyzed after resolving on 18% SDS-PAGE and staining with Coomassie blue. To see disulfide-linked products, assay samples were neither boiled nor reduced prior to gel loading. A replicate of the UCH-L3 + $QK^{Ub(aux)}E$ reaction was reduced with 50 mM DTT and boiled at 95 °C to confirm the disulfide linkage. All DUB Assays were also analyzed by LC-ESI-MS employing a gradient of 0-75% D, 40 min.

2.5 Product characterization and supplemental data

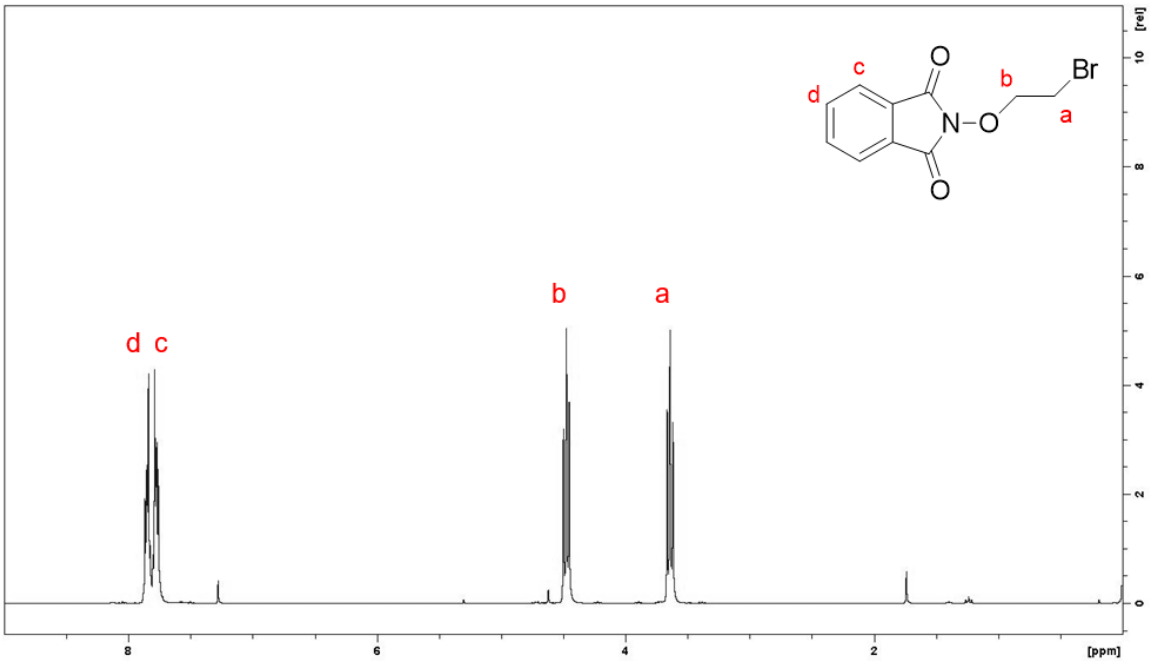


Figure 2.10. ^{1}H-NMR of *N*-(2-bromoethoxy)phthalimide. $CHCl_3$ solvent peak observed at 7.28 ppm (s). Peak at 1.74 ppm (s) identified as H_2O.

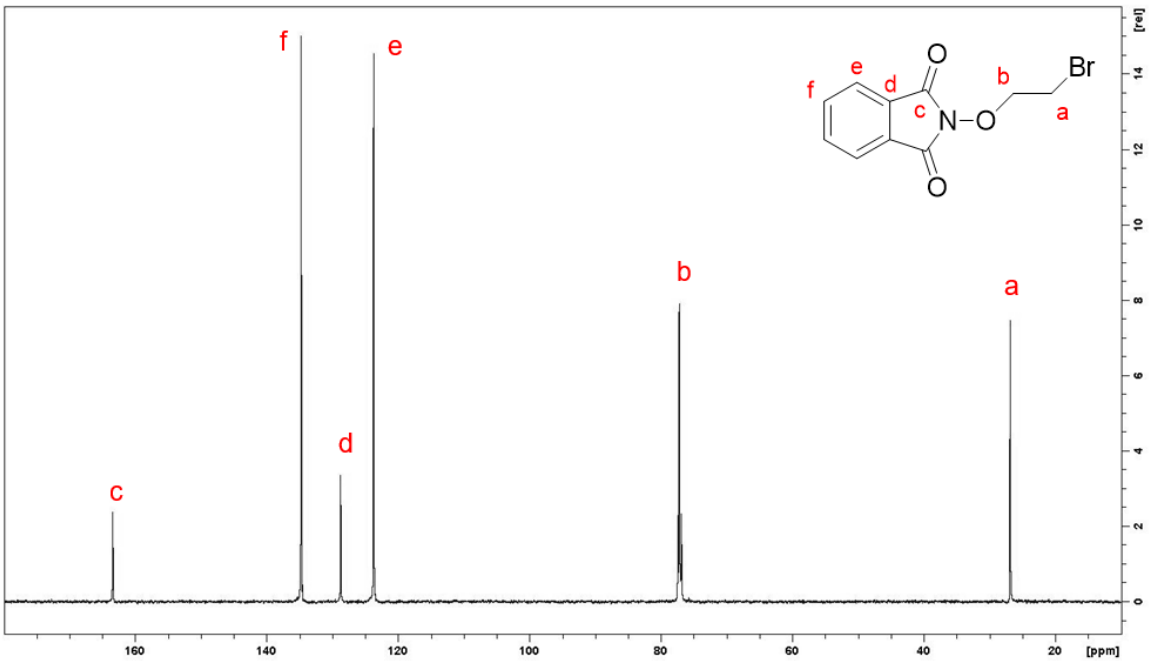


Figure 2.11. ^{13}C-NMR of *N*-(2-bromoethoxy)phthalimide. $CHCl_3$ solvent peak identified at 77.22 ppm (t).

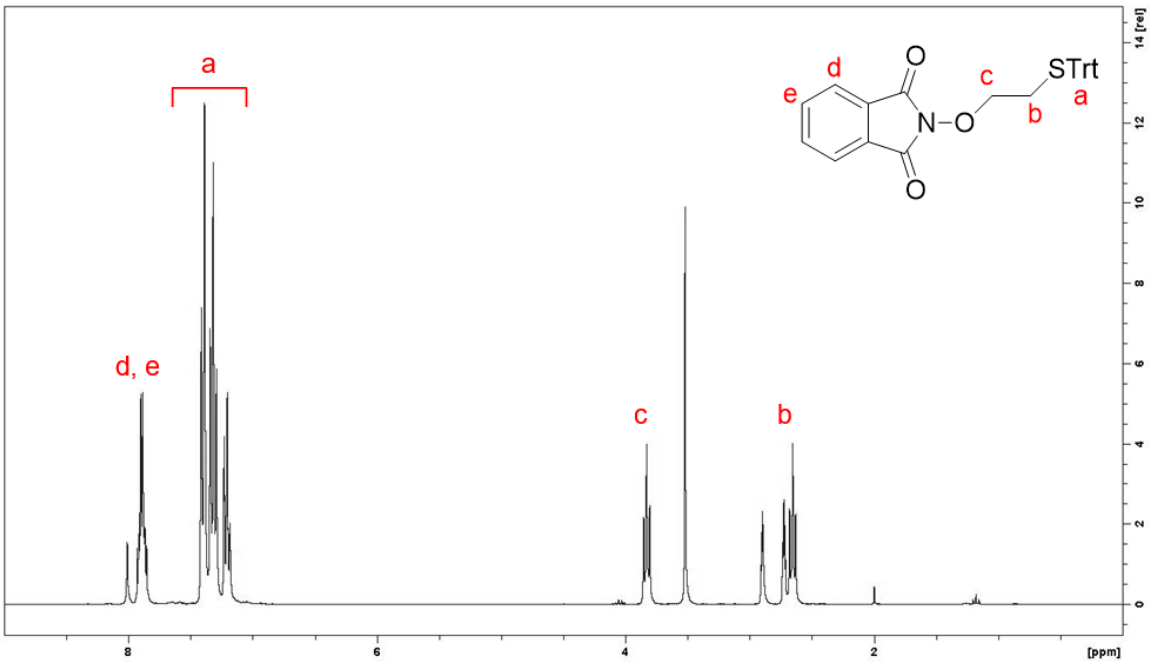


Figure 2.12. ^{1}H-NMR of *N*-(2-(tritylthio)ethoxy)phthalimide. DMF solvent peaks observed at 8.01 (s), 2.90 (m) and 2.72 (m) ppm. Peak at 3.52 ppm (s) identified as H_2O. Peaks at 4.04 (q), 2.00 (s) and 1.18 (t) identified as residual ethyl acetate.

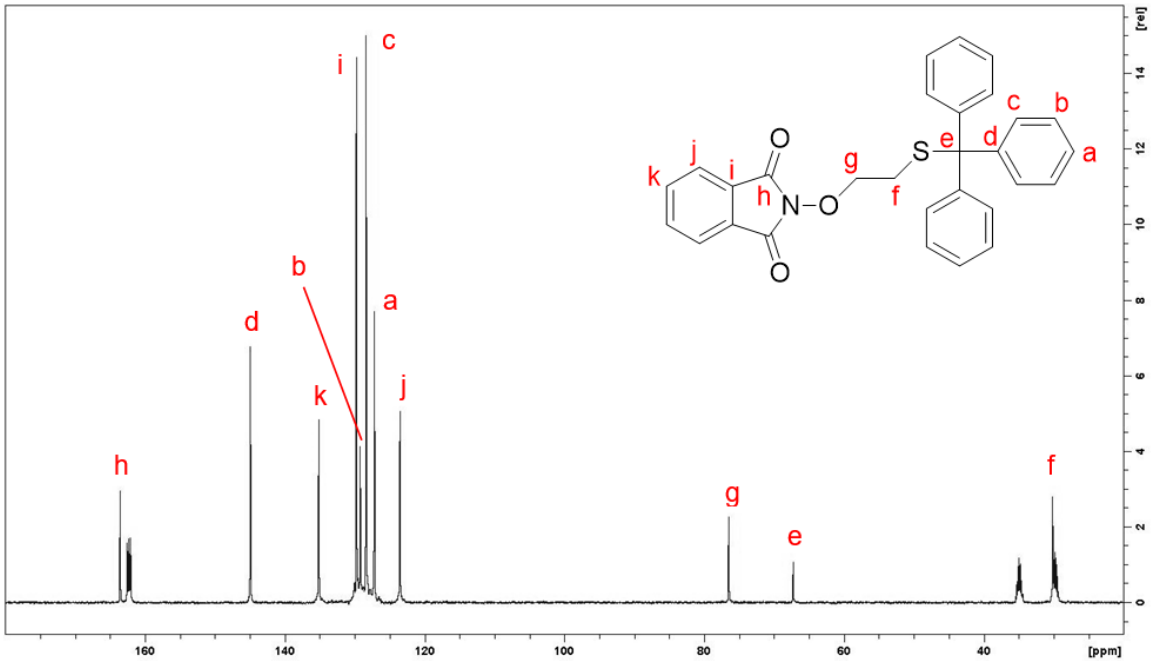


Figure 2.13. ^{13}C-NMR of *N*-(2-(tritylthio)ethoxy)phthalimide. DMF solvent peaks observed at 162.28 (t), 34.90 (m) and 29.78 (m) ppm.

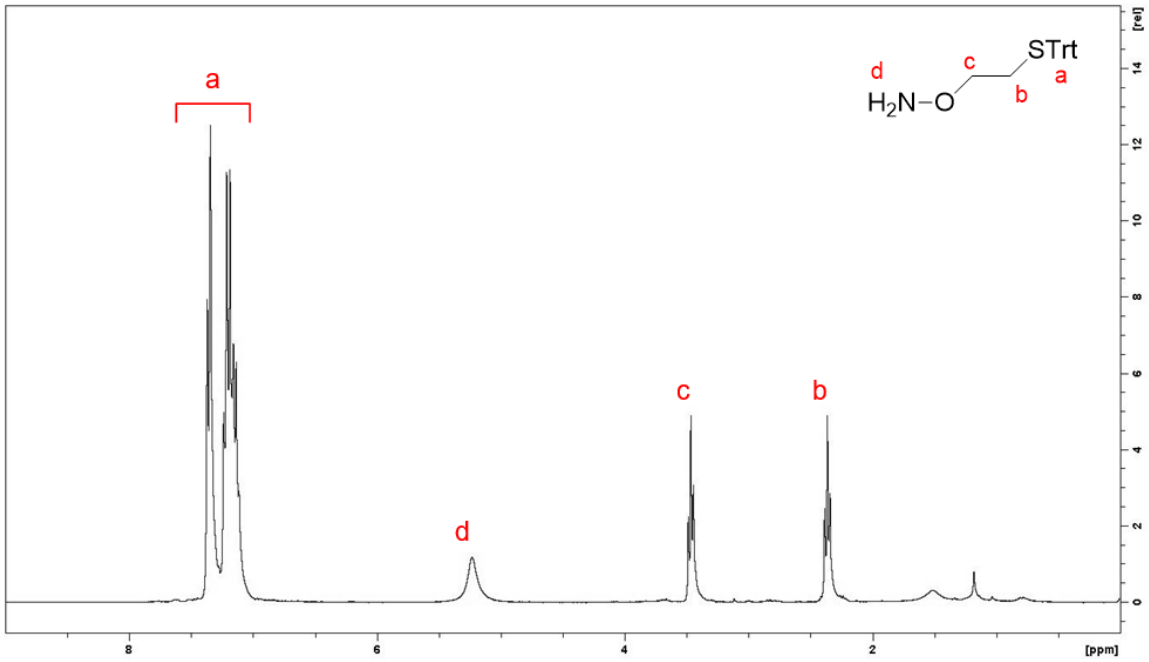


Figure 2.14. ^{1}H-NMR of *O*-(2-(tritylthio)ethyl)hydroxylamine (1). Peak at 1.51 (s) identified as H_2O.

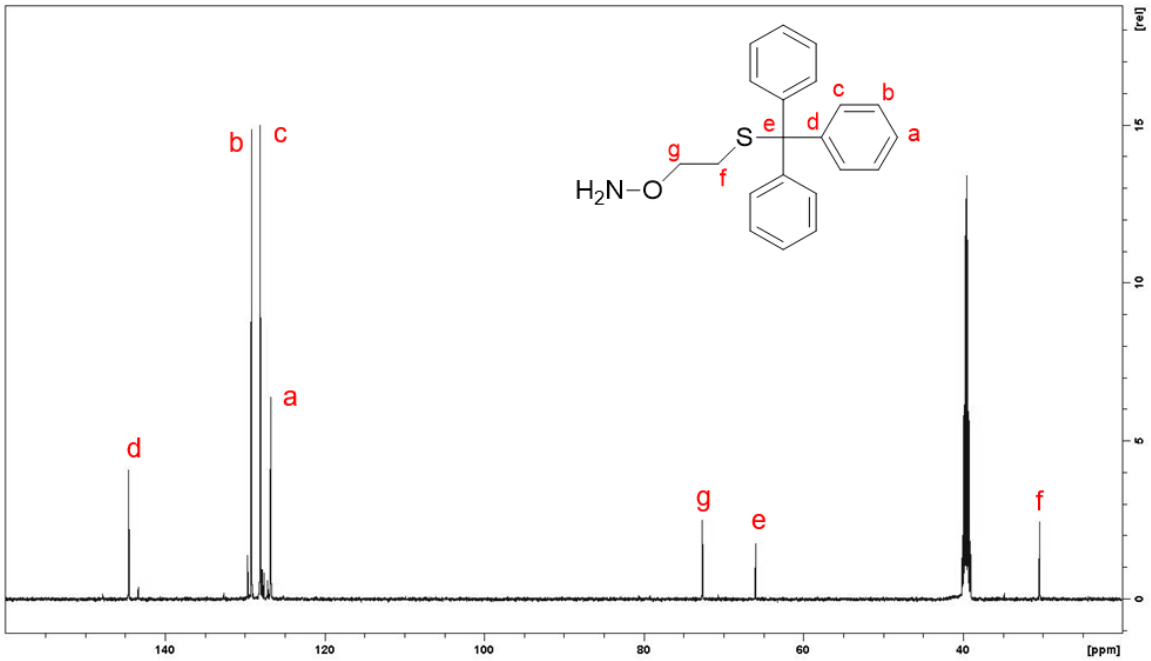


Figure 2.15. ^{13}C-NMR of *O*-(2-(tritylthio)ethyl)hydroxylamine (1). DMSO solvent peak observed at 40.54 (m) ppm.

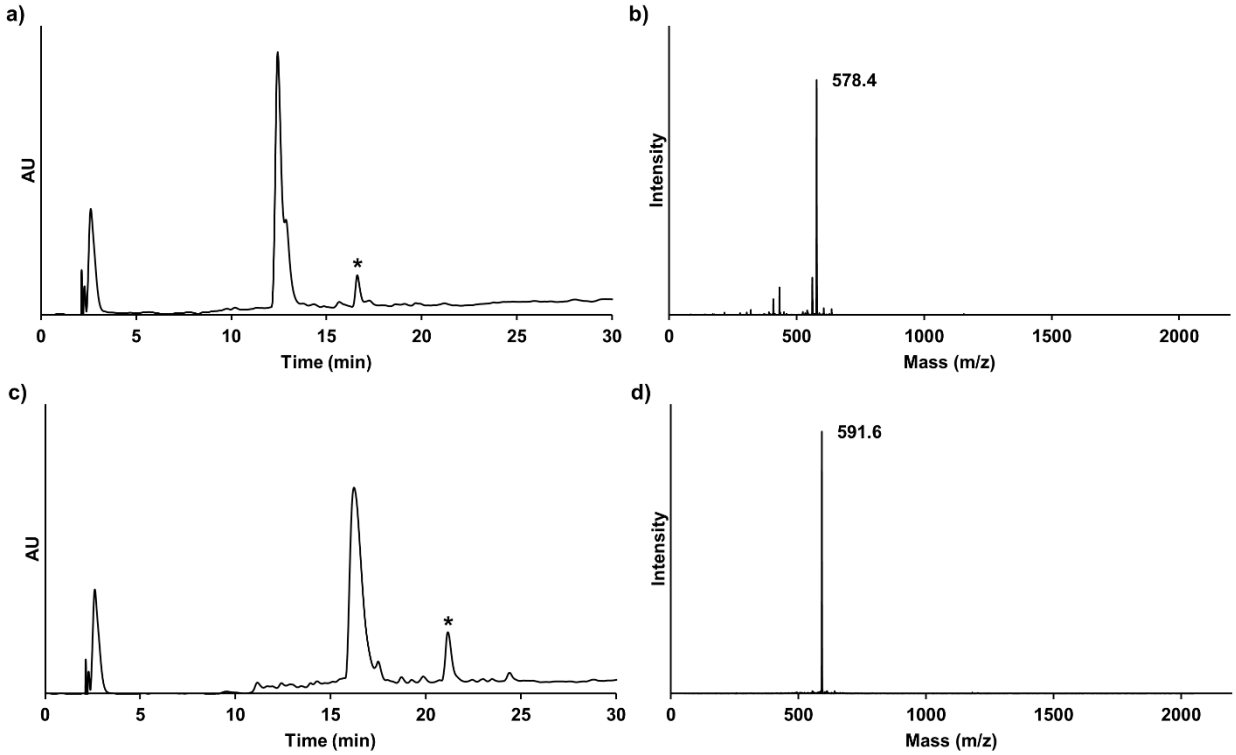


Figure 2.16. Purification of QK$^{(aux)}$E (2) and KAK$^{(aux)}$I. a) C18 analytical RP-HPLC chromatogram of QK$^{(aux)}$E (**2**), gradient of 0-40% B, 30 min. b) ESI-MS of purified QK$^{(aux)}$E. c) C18 analytical RP-HPLC chromatogram of KAK$^{(aux)}$I, gradient of 0-40% B, 30 min. d) ESI-MS of purified KAK$^{(aux)}$I. Asterisks indicates the disulfide forms of QK$^{(aux)}$E and KAK$^{(aux)}$I.

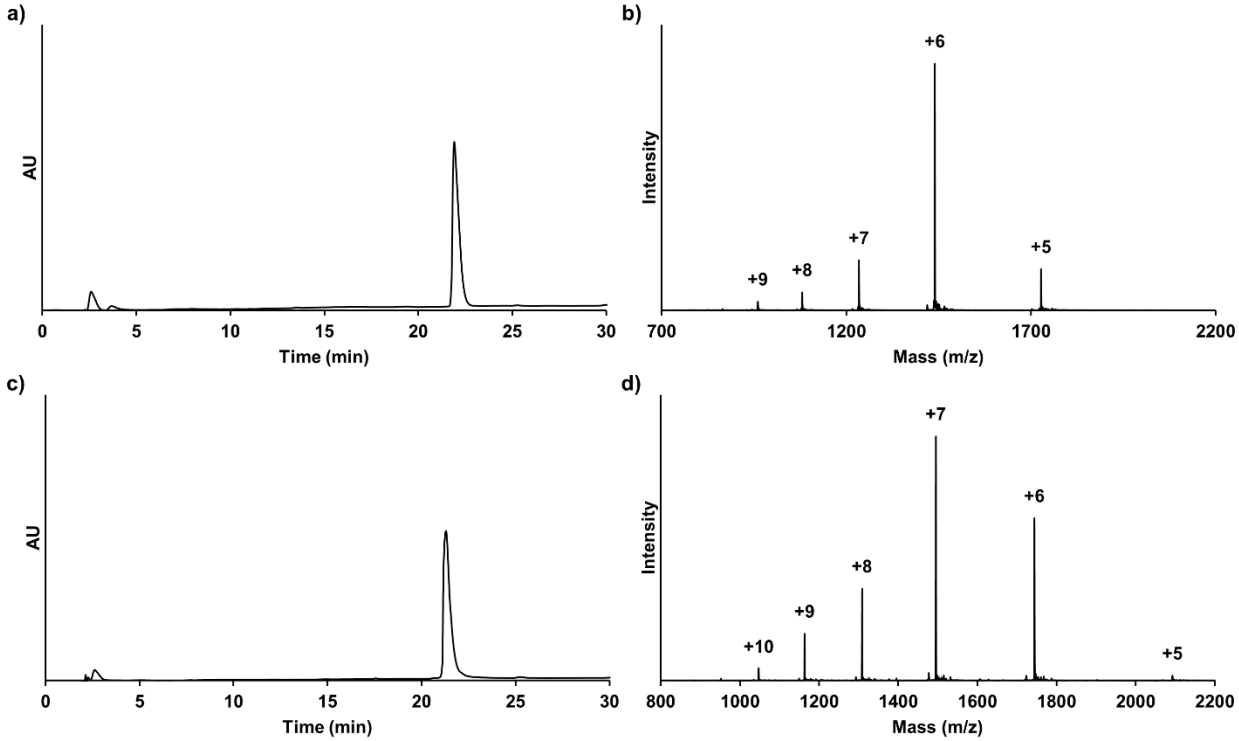


Figure 2.17. Purification of Ub(1-75)-MES (3) and SUMO-3(2-91)-MES. a) C18 analytical RP-HPLC chromatogram of purified Ub(1-75)-MES (**3**), gradient of 0-73% B, 30 min. b) ESI-MS of purified Ub(1-75)-MES. c) C18 analytical RP-HPLC chromatogram of purified SUMO-3(2-91)-MES, gradient of 0-73% B, 30 min. d) ESI-MS of purified SUMO-3(2-91)-MES.

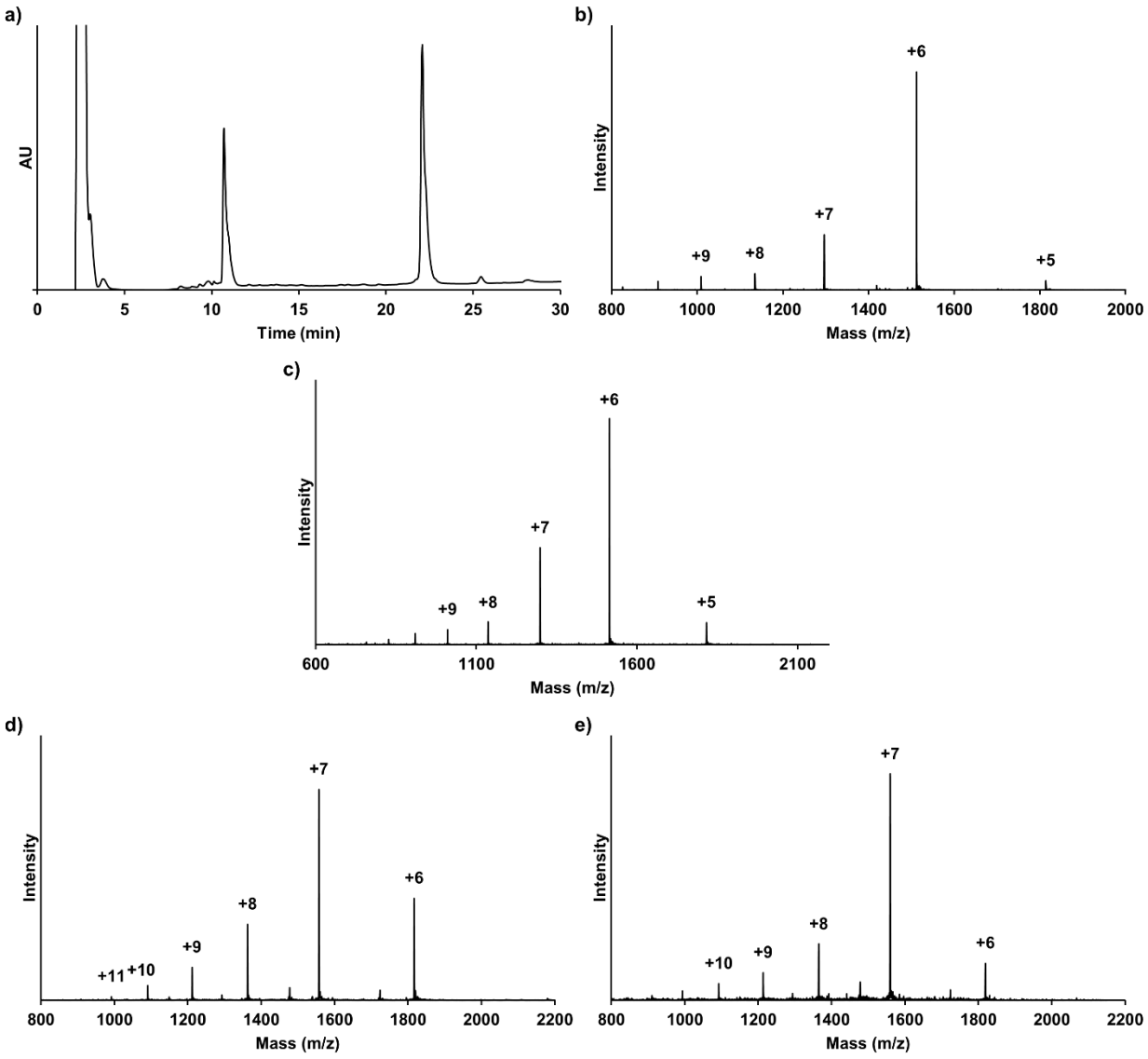


Figure 2.18. Expressed protein ligation and purification of QK$^{(aux)}$E (2) and KAK$^{(aux)}$I to Ub(1-75)-MES and SUMO-3(2-91)-MES α-thioesters. a) C18 analytical RP-HPLC chromatogram of the QK$^{Ub(aux)}$E ligation reaction after 12 h, gradient of 0-73% B, 30 min. b) ESI-MS of the RP-HPLC peak at 22 min corresponding to QK$^{Ub(aux)}$E (**4**). c) ESI-MS of purified KAK$^{Ub(aux)}$I. d) ESI-MS of QK$^{Su(aux)}$E. e) ESI-MS of KAK$^{Su(aux)}$I.

Table 2.4. Optimization of reaction conditions for reductive removal of the ligation auxiliary.

Product	Reagents	pH	Temp. (°C)	Time (h)	Yield (%)
$QK^{Ub}E$	Zn, 6 M Gn-HCl	1	37	48	66
$QK^{Su}E$	Zn, 6 M Gn-HCl	1	37	48	61
$KAK^{Ub}I$	Zn, 6 M Gn-HCl	1	37	48	59
$KAK^{Su}I$	Zn, 6 M Gn-HCl	1	37	48	59
$KAK^{Ub}I$	Zn, 6 M Gn-HCl	3	37	48	53
$QK^{Ub}E$	Zn, 6 M Gn-HCl	1	22	48	19
$QK^{Ub}E$	Zn, 6 M Gn-HCl	3	22	48	28
$QK^{Ub}E$	Zn, 6 M Gn-HCl	3	22	24	24
$QK^{Ub}E$	In, 6 M Gn-HCl	3	22	24	23

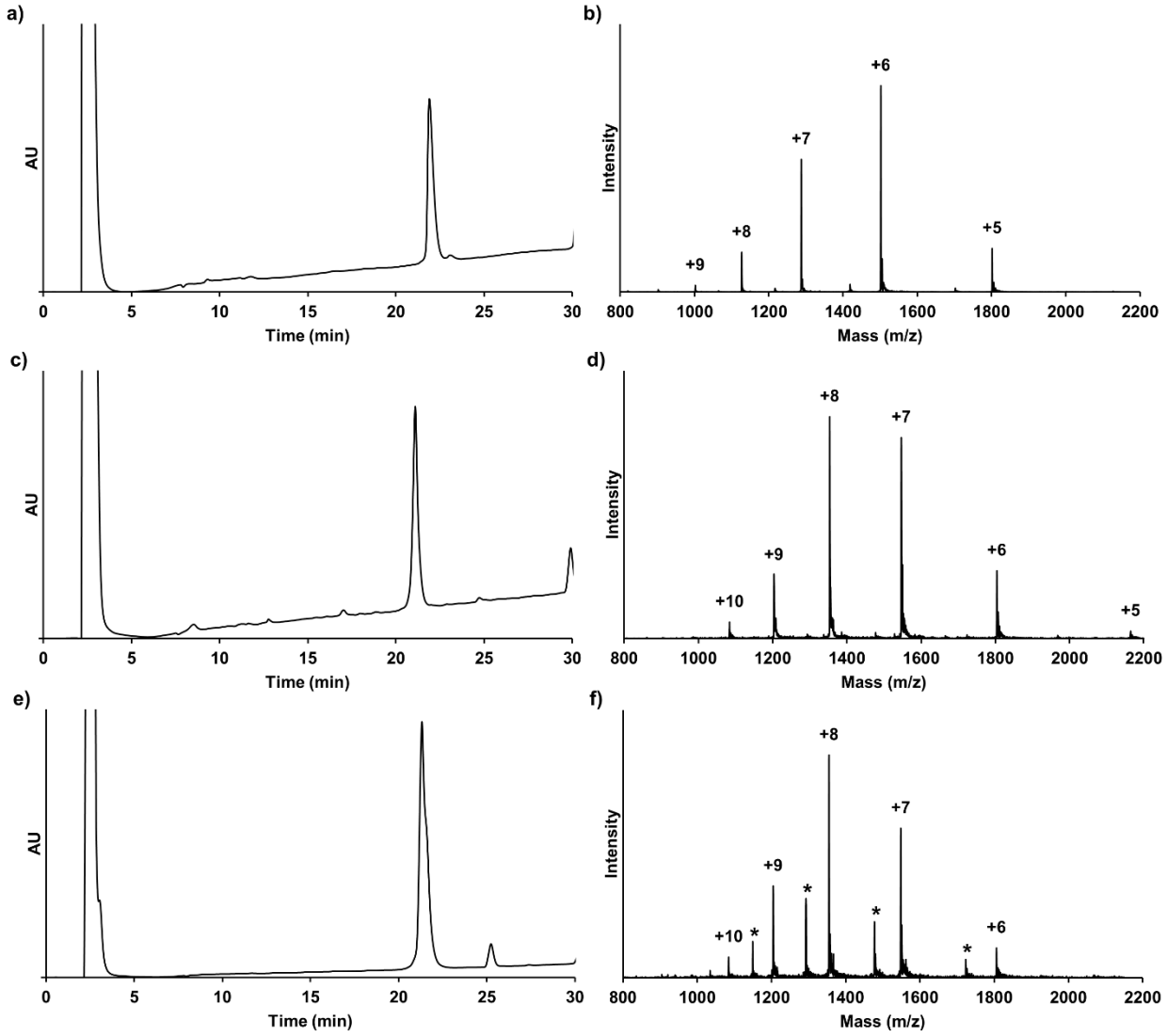


Figure 2.19. Purification of KAK^{Ub}I, QK^{Su}E, and KAK^{Su}I. a) C18 analytical RP-HPLC chromatogram of purified KAK^{Ub}I, gradient of 0-73% B, 30 min. R_t = 22 min. b) ESI-MS of purified KAK^{Ub}I. c) C18 analytical RP-HPLC chromatogram of purified QK^{Su}E, gradient of 0-73% B, 30 min. R_t = 21 min. d) ESI-MS of purified QK^{Su}E. e) C18 analytical RP-HPLC chromatogram of purified KAK^{Su}I, gradient of 0-73% B, 30 min. R_t = 21 min. f) ESI-MS of purified KAK^{Su}I. The co-eluting minor species indicated by asterisks is SUMO-3(2-91)-COOH.

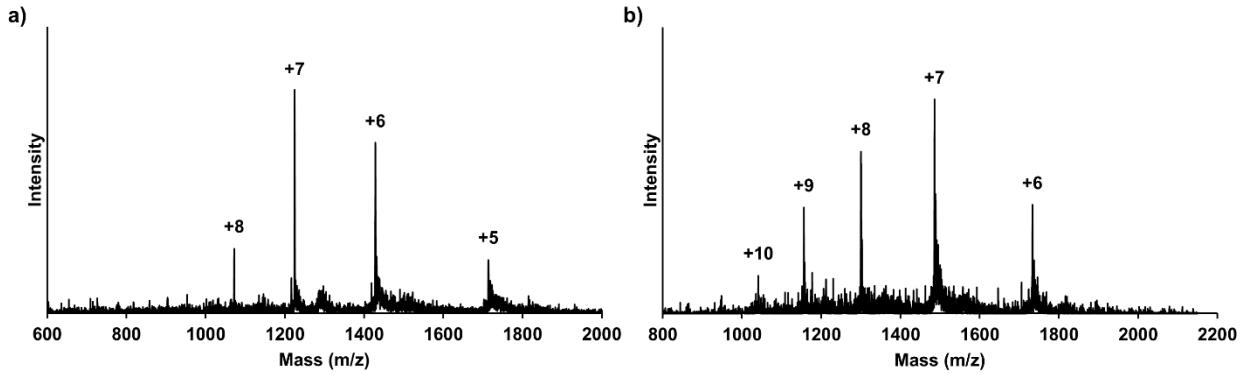


Figure 2.20. UCH-L3 and SENP1 hydrolysis assays. a) LC-ESI-MS of $QK^{Ub}E$ assay with UCH-L3. Calculated for Ub(1-76)-COOH *m/z* $[M+H]^+$ 8,565.8 Da, observed 8,565.5 ± 1.4 Da. b) LC-ESI-MS of $QK^{Su}E$ assay with SENP1. Calculated for SUMO-3(2-91)-COOH *m/z* $[M+H]^+$ 10,394.6 Da, observed 10,395.9 ± 2.6 Da.

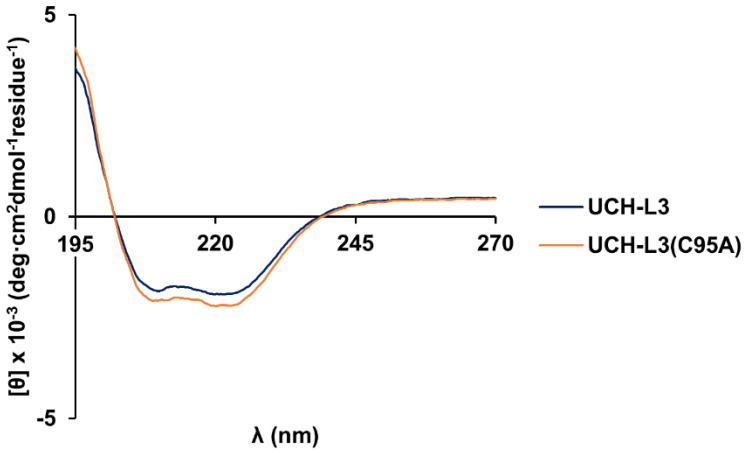


Figure 2.21. Characterization of wild-type and mutant UCH-L3. Circular Dichroism spectrum of UCH-L3 and UCH-L3(C95A) at 0.4 mg/mL in 50 mM NaH_2PO_4, pH 7.2.

2.6 References

Portions of this chapter have been published as:

Weller, C. E.; Huang, W.; Chatterjee, C. Facile Synthesis of Native and Protease-Resistant Ubiquitylated Peptides. *ChemBioChem* **2014**, *15* (9), 1263–1267.

(1) Sharp, P. M.; Li, W.-H. Molecular evolution of ubiquitin genes. *Trends Ecol. Evol.* **1987**, *2* (11), 328–332.

(2) Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **1998**, *67*, 425–479.

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

(4) Xu, G.; Paige, J. S.; Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* **2010**, *28* (8), 868–873.

(5) Becker, J.; Barysch, S. V; Karaca, S.; Dittner, C.; Hsiao, H.-H.; Berriel Diaz, M.; Herzig, S.; Urlaub, H.; Melchior, F. Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat. Struct. Mol. Biol.* **2013**, *20* (4), 525–531.

(6) Shiio, Y.; Eisenman, R. N. Histone SUMOylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230.

(7) Nathan, D.; Ingvarsdottir, K.; Sterner, D. E.; Bylebyl, G. R.; Dokmanovic, M.; Dorsey, J. A.; Whelan, K. A.; Krsmanovic, M.; Lane, W. S.; Meluh, P. B.; et al. Histone SUMOylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes Dev.* **2006**, *20*, 966–976.

(8) Ayaydin, F.; Dasso, M. Distinct in vivo dynamics of vertebrate SUMO paralogues. *Mol. Biol. Cell* **2005**, *15*, 5208–5218.

(9) van Wijk, S. J. L.; Fiskin, E.; Putyrski, M.; Pampaloni, F.; Hou, J.; Wild, P.; Kensche, T.; Grecco, H. E.; Bastiaens, P.; Dikic, I. Fluorescence-based sensors to monitor localization and functions of linear and K63-linked ubiquitin chains in cells. *Mol. Cell* **2012**, *47*, 797–809.

(10) Weller, C. E.; Pilkerton, M. E.; Chatterjee, C. Chemical strategies to understand the language of ubiquitin signaling. *Biopolymers* **2014**, *101* (2), 144–155.

(11) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. SUMOylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837.

(12) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008**, *453* (7196), 812–816.

(13) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. *Chem. Commun.* **2010**, *46*, 7199–7201.

(14) Hejjaoui, M.; Haj-Yahya, M.; Kumar, K. S. A.; Brik, A.; Lashuel, H. A. Towards elucidation of the role of ubiquitination in the pathogenesis of Parkinson's disease with semisynthetic ubiquitinated α-

synuclein. *Angew. Chemie* **2011**, *50*, 405–409.

(15) Fotouhi, N.; Galakatos, N. G.; Kemp, D. S. Peptide synthesis by prior thiol capture. 6. Rates of the disulfide bond forming capture reaction and demonstration of the overall strategy by synthesis of the C-terminal 29-peptide sequence of BPTI. *J. Org. Chem.* **1989**, *54*, 2803–2817.

(16) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pK values of the ionizable groups of proteins. *Protein Sci.* **2006**, *15*, 1214–1218.

(17) Offer, J. Native chemical ligation with Nα acyl transfer auxiliaries. *Biopolymers* **2010**, *94* (4), 530–541.

(18) Welchman, R. L.; Gordon, C.; Mayer, R. J. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (8), 599–609.

(19) Chatterjee, C.; McGinty, R. K.; Pellois, J.-P.; Muir, T. W. Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chemie* **2007**, *46*, 2814–2818.

(20) Botti, P.; Carrasco, M. R.; Kent, S. B. H. Native chemical ligation using removable Nα-(1-phenyl-2-mercaptoethyl) auxiliaries. *Tetrahedron Lett.* **2001**, *42*, 1831–1833.

(21) Tchertchian, S.; Hartley, O.; Botti, P. Synthesis of Nα-(1-phenyl-2-mercaptoethyl) amino acids, new building blocks for ligation and cyclization at non-cysteine sites: Scope and limitations in peptide synthesis. *J. Org. Chem.* **2004**, *69* (8), 9208–9214.

(22) Macmillan, D.; Anderson, D. W. Rapid synthesis of acyl transfer auxiliaries for cysteine-free native glycopeptide ligation. *Org. Lett.* **2004**, *6* (25), 4659–4662.

(23) Kawakami, T.; Aimoto, S. A photoremovable ligation auxiliary for use in polypeptide synthesis. *Tetrahedron Lett.* **2003**, *44*, 6059–6061.

(24) Loibl, S. F.; Harpaz, Z.; Seitz, O. A type of auxiliary for native chemical peptide ligation beyond cysteine and glycine junctions. *Angew. Chemie* **2015**, *54*, 15055–15059.

(25) Kawakami, T.; Akaji, K.; Aimoto, S. Peptide bond formation mediated by 4,5-dimethoxy-2-mercaptobenzylamine after periodate oxidation of the N-terminal serine residue. *Org. Lett.* **2001**, *3* (9), 1403–1405.

(26) Harpaz, Z.; Loibl, S.; Seitz, O. Native chemical ligation at a base-labile 4-mercaptobutyrate Nα-auxiliary. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 1434–1437.

(27) Canne, L. E.; Bark, S. J.; Kent, S. B. H. Extending the applicability of native chemical ligation. *J. Am. Chem. Soc.* **1996**, *118*, 5891–5896.

(28) Li, J.; Cui, H. K.; Liu, L. Peptide ligation assisted by an auxiliary attached to amidyl nitrogen. *Tetrahedron Lett.* **2010**, *51* (13), 1793–1796.

(29) Ajish Kumar, K. S.; Harpaz, Z.; Haj-Yahya, M.; Brik, A. Side-chain assisted ligation in protein synthesis. *Bioorganic Med. Chem. Lett.* **2009**, *19*, 3870–3874.

(30) Komander, D.; Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **2012**, *81*, 203–229.

(31) Sax, B.; Dick, F.; Tanner, R.; Gosteli, J. 4-Methyltrityl (Mtt): a new protecting group for the side chain protection of Asn and Gln in solid-phase peptide synthesis. *Pept. Res.* **1992**, *5* (4), 245–246.

(32) Castro, E. A. Kinetics and mechanisms of reactions of thiol, thiono and dithio analogues of carboxylic esters with nucleophiles. *Chem. Rev.* **1999**, *99*, 3505–3524.

(33) Cary, P. D.; King, D. S.; Crane-Robinson, C.; Bradbury, E. M.; Rabbani, A.; Goodwin, G. H.; Johns, E. W. Structural studies on two high-mobility-group proteins from calf thymus, HMG-14 and HMG-20 (ubiquitin), and their interaction with DNA. *Eur. J. Biochem.* **1980**, *112*, 577–580.

(34) Denmark, S. E.; Stolle, A.; Dixon, J. A.; Guagnano, V. Tandem inter [4+2]/intra [3+2] cycloadditions. 6. The bridged mode. *J. Am. Chem. Soc.* **1995**, *117*, 2100–2101.

(35) Cicchi, S.; Bonanni, M.; Cardona, F.; Revuelta, J.; Goti, A. Indium-mediated reduction of hydroxylamines to amines. *Org. Lett.* **2003**, *5* (10), 1773–1776.

(36) Kang, J.; Reynolds, N. L.; Tyrrell, C.; Dorin, J. R.; Macmillan, D. Peptide thioester synthesis through N→S acyl-transfer: Application to the synthesis of a β-defensin. *Org. Biomol. Chem.* **2009**, *7*, 4918–4923.

(37) Erlich, L. A.; Kumar, K. S. A.; Haj-Yahya, M.; Dawson, P. E.; Brik, A. N-methylcysteine-mediated total chemical synthesis of ubiquitin thioester. *Org. Biomol. Chem.* **2010**, *8* (10), 2392–2396.

(38) Johnston, S. C.; Riddle, S. M.; Cohen, R. E.; Hill, C. P. Structural basis for the specificity of ubiquitin C-terminal hydrolases. *EMBO J.* **1999**, *18* (14), 3877–3887.

(39) Shen, L. N.; Dong, C.; Liu, H.; Naismith, J. H.; Hay, R. T. The structure of SENP1-SUMO-2 complex suggests a structural basis for discrimination between SUMO paralogues during processing. *Biochem. J.* **2006**, *397*, 279–288.

(40) Maiti, T. K.; Permaul, M.; Boudreaux, D. A.; Mahanic, C.; Mauney, S.; Das, C. Crystal structure of the catalytic domain of UCHL5, a proteasome-associated human deubiquitinating enzyme, reveals an unproductive form of the enzyme. *FEBS J.* **2011**, *278*, 4917–4926.

(41) Reyes-Turcu, F. E.; Ventii, K. H.; Wilkinson, K. D. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu. Rev. Biochem.* **2009**, *78*, 363–397.

(42) Stawikowski, M.; Stawikowska, R.; Jaśkiewicz, A.; Zabłotna, E.; Rolka, K. Examples of peptide-peptoid hybrid serine protease inhibitors based on the trypsin inhibitor SFTI-1 with complete protease resistance at the P1-P1′ reactive site. *ChemBioChem* **2005**, *6*, 1057–1061.

(43) Komander, D. Chapter 6. Mechanism, specificity, and structure of the deubiquitinases. In *Conjugation and deconjugation of ubiquitin family modifiers*; Groettrup, M., Ed.; Landes Bioscience: Austin, TX, 2010; pp 69–87.

(44) Haj-Yahya, M.; Eltarteer, N.; Ohayon, S.; Shema, E.; Kotler, E.; Oren, M.; Brik, A. N-methylation of isopeptide bond as a strategy to resist deubiquitinases. *Angew. Chemie* **2012**, *51*, 11535–11539.

(45) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* **1970**, *34* (2), 595–598.

(46) King, D. S.; Fields, C. G.; Fields, G. B. A cleavage method which minimizes side reactions following

Fmoc solid phase peptide synthesis. *Int. J. Pept. Protein Res.* **1990**, *36*, 255–266.

(47) Yamamura, S.; Toda, M.; Hirata, Y. Modified Clemmensen reduction: Cholestane. *Org. Synth.* **1973**, *53*, 289.

(48) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007**, *128* (5), 1003–1012.

(49) Sowa, M. E.; Bennett, E. J.; Gygi, S. P.; Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **2009**, *138*, 389–403.

Chapter 3

Chemical ubiquitylation of folded proteins enabled by aromatic thiol-mediated *N*-*O* bond cleavage

3.1 Introduction

The reversible conjugation of proteins with ubiquitin (Ub) is a post-translational modification (PTM) conserved in all eukaryotic organisms.[1] The family of ubiquitin-like proteins (Ubls) consists of about 25 proteins, a majority of which can be conjugated with protein targets, either at specific Lys side-chain amines or at the protein N-terminus.[2] As discussed in Chapter 2, only a small fraction of proteins are ubiquitylated at any given time,[2] and for Ub alone, there exist over 600 ligases that attach it to protein substrates in humans.[3] This complicates the isolation or enzymatic generation of ubiquitylated substrates for *in vitro* biochemical studies aimed at understanding ths dynamic and transient PTM.

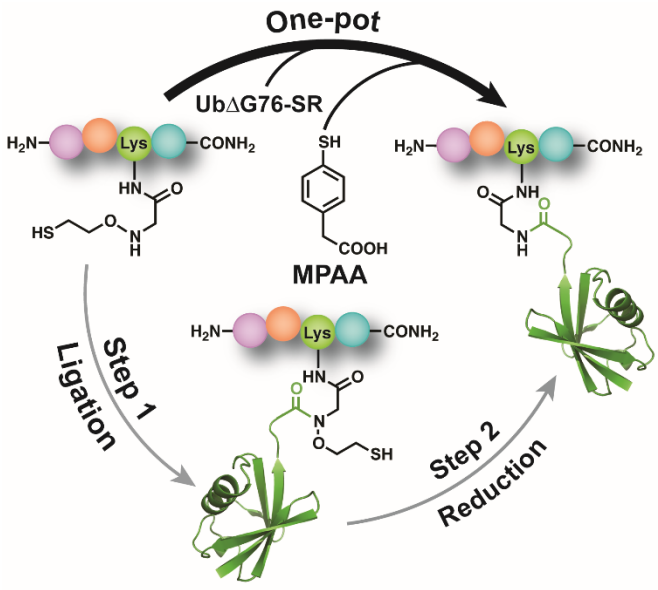
Over the past few decades, a handful of chemical strategies have emerged that enable access to site-specifically ubiquitylated proteins or close analogs thereof.[4,5] Often, these approaches (reviewed in Chapter 1) rely on challenging multi-step syntheses, and a desulfurization step that is incompatible with Cys residues.[6–8] Alternatively, N^{α}-ligation auxiliaries (reviewed in Chapter 2) may be appended to the Lys ε-amine during solid-phase peptide synthesis (SPPS) via the C-terminal Ubl Gly residue, and used for ligation with the truncated, recombinant Ubl C-terminal α-thioester.[9] However, many auxiliaries have harsh removal conditions, or are obtained through complex and low-yielding syntheses.[9,10] In an attempt to overcome some of these limitations, we reported a peptide ubiquitylation strategy (described in Chapter 2) with the temporary ligation auxiliary, 2-(aminooxy)ethanethiol, which employed reduction with metallic Zn in the terminal step (**Scheme 3.1**).[11,12] Unfortunately, we found that efficient *N*-*O* bond reduction required both harsh denaturants and strongly acidic conditions. Therefore, our approach was limited to peptides or proteins amenable to refolding from the denatured state.

We realized that adapting our ligation strategy to native, non-denaturing conditions would significantly broaden access to Ubl-modified proteins for *in vitro* biochemical studies. The vast majority of native chemical ligation (NCL) or expressed protein ligation (EPL) reactions in the literature take place under strongly denaturing conditions, such as 8 M urea or 6 M guanidine-HCl (Gn-HCl). Because the ultimate goal of protein semisynthesis is the utilization of products in relevant assays, methods thus far have been limited to proteins with the ability to properly refold *in vitro*, such as ubiquitin,[13] phospholipase A2,[14] histones,[15] nitrophorin 4,[16] and more. Many other, often larger, proteins do not spontaneously re-fold, and attempts at refolding *in vitro* lead to improperly-folded yet stable intermediate structures that are prone to aggregation through hydrophobic interactions. It is theoretically possible to refold any protein from a denatured state, but identifying the proper conditions is time-consuming and often low-yielding.[17] Thus, there is a need for non-denaturing ligation methodology.

There have been few reports of ECL under non-denaturing conditions, and no instances to our knowledge that take place on the Lys side chain. Many involve intein-mediated reactions. Protein trans splicing, for example, utilizes split inteins to catalyze ligation between protein domains expressed separately as fusions with the N- or C-terminal intein fragments.[18] When incubated together, the reconstituted intein catalyzes their ligation and its own self-removal. Similarly, peptides or proteins containing N-terminal Cys may be added directly to buffer containing a C-terminal α-thioester protein from an intein fusion thiolysis reaction.[19] The choice of unhindered ligation sites, such as unstructured linker regions between protein domains, is crucial in both cases. However, even ligations at unhindered sites proceed slowly between folded substrates. In 1999, Muir and coworkers reported the first non-denaturing ligation between differentially isotopically labeled domains of the Abelson protein tyrosine kinase-SH(3,2) domain pair for NMR studies.[20] This ligation, at a Gly-Cys junction, reached 70% yield only after 4 days of reaction, even in the presence of ~150 mM each of thiophenol and benzyl mercaptan, additives known to enhance ECL reaction rate.

We anticipated that ligation between folded proteins may be slow, but the unstructured C-terminal tail and di-Gly motif of Ubls, and the tendency of substrate Lys side chains to be solvent exposed, were promising. Even so, we considered a thiol additive for the reaction. Alkyl thioesters, such as 2-mercaptoethanesulfonate (MES), are often applied to the synthesis of proteinyl or peptidyl C-terminal α-thioesters. The relative stability of these thioesters during purification and other manipulations is useful, but also inhibits their reactivity towards ligation. Thiol additives, such as thiophenol, have been used to overcome this problem.[14,21] When present in large excess, thiophenol performs an initial transthioesterification with the alkyl thioester, and creates a much more reactive thioester for NCL due to its superior leaving group ability. However, the efficacy of thiophenol is hampered by its low water solubility. The readily soluble 4-mercaptophenylacetic acid (MPAA) is a widely employed, and significantly more effective, rate enhancer.[22]

To overcome slow ligation kinetics under non-denaturing conditions, we therefore included MPAA in our initial test ligations. In addition to greatly enhancing the rate of ligation, we found that MPAA also effected auxiliary removal without the need for zinc and acidic, denaturing conditions. This led to the discovery of a novel aromatic thiol-mediated *N*-*O* bond cleavage reaction that enables traceless, one-pot, chemical ubiquitylation of folded proteins at physiological pH (**Scheme 3.1**). Mechanistic investigation of this new reaction implicates a disulfide radical anion as the reductive species that cleaves *N*-*O* bonds. The semisynthesis of full-length human histone H2B modified by the small ubiquitin-like modifier protein, SUMO-3, demonstrates the complete compatibility of this reaction with thiol side-chains in folded proteins and significantly expands the practical scope of chemical ubiquitylation.



Scheme 3.1. Aromatic thiol-mediated one-pot traceless native chemical ubiquitylation. MPAA = 4-mercaptophenylacetic acid. UbΔG76-SR = Ub(1-75)-α-thioester with 2-mercaptoethanesulfonic acid. PDB code 1UBQ (Ub).

3.2 Results and discussion

3.2.1 4-mercaptophenylacetic acid-mediated *N-O* bond cleavage

We previously reported the successful application of the auxiliary 2-aminooxyethanethiol toward peptide ubiquitylation.[12] The utility of this auxiliary group lies in its high-yielding 3-step synthesis and easy incorporation in various peptide substrates. However, two challenges in removing the auxiliary and producing a wild-type amide linkage were the requirement for pH 3, and the necessity of chaotropes such as 6 M Gn-HCl that unfold ubiquitin to allow reduction of the *N-O* bond by metallic zinc. Although such a strategy is compatible with proteins that may be refolded from the denatured state, its broad utility is limited. Moreover, the electrophilic character of the nascent disubstituted amide bond in the ligation product led to a small amount of hydrolysis over time, which was exacerbated at the low pH required for efficient *N-O* bond reduction. Indeed, initial efforts to reduce the *N-O* bond with Zn at higher, near-physiological pH was quite low-yielding, and led to side products, precipitation of protein at the Zn surface, and a high rate of competing hydrolysis.

In an effort to reduce the amount of hydrolyzed Ub(1-75)-COOH side-product and to increase the rate of transthioesterification between auxiliary-bearing peptides and the Ub(1-75)-α-thioester, we tested the aromatic thiol MPAA as a ligation additive (**Scheme 3.1**). The excellent leaving group ability of MPAA renders its protein thioesters more reactive toward transthioesterification, the first and rate-limiting step in native chemical ligation.[22,23] In a typical ligation reaction with 0.5 mM Ub(1-75)-α-thioester and 5 mM of auxiliary-bearing peptide (KAK^{aux}I) in a buffer consisting of 50 mM tris, 150 mM NaCl, and 200 mM MPAA at pH 7.3, we observed good ligation kinetics, and completion of the ligation reaction within 24 h. To our surprise, when the reaction was allowed to continue for 48 h, the final ligation product was altogether missing the ligation auxiliary (**Figure 3.1a**). This unexpected result was consistently reproducible, although the slow kinetics of product formation necessitated up to 48 hours to achieve 50-70% yields (**Figure 3.1b**). Additional controls revealed MPAA to be the critical component required for *N-O* bond cleavage, and re-purification of the commercial compound by high performance liquid chromatography (HPLC) did not inhibit the reaction (**Table 3.1**, entries 1-8). The necessity of a free sulfhydryl group in MPAA was seen from the fact that pure disulfide-linked MPAA dimer did not undertake *N-O* bond cleavage (**Table 3.1**, entry 9).

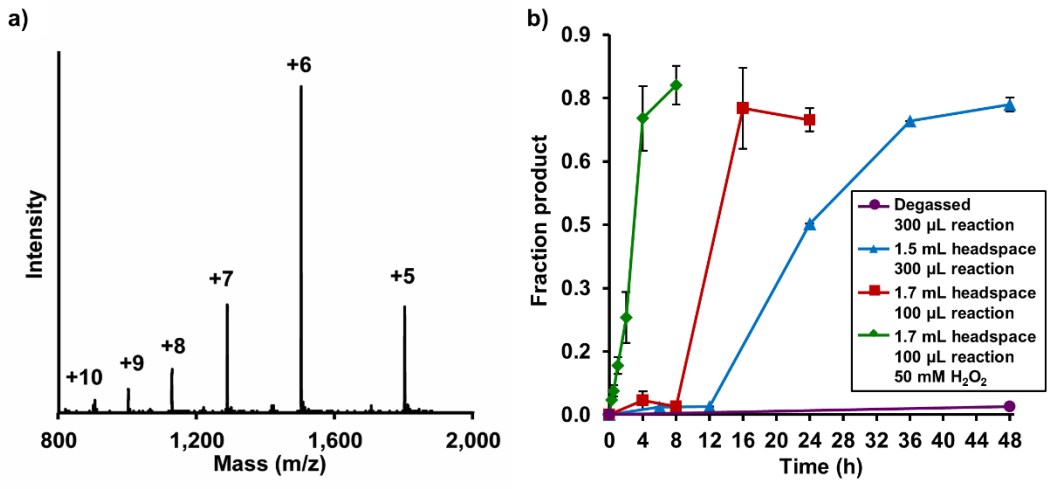


Figure 3.1. *N-O* bond cleavage in the native chemical ligation product KAK$^{Ub(aux)}$I. a) ESI-MS spectrum of the final ligation product of Ub(1-75)-α-thioester with KAK^{aux}I. Calculated for KAK$^{Ub(aux)}$I, *m/z* $[M+H]^+$ 9,081.1 Da. Observed KAK^{Ub}I, *m/z* $[M+H]^+$ 9,004.8 ± 2.7 Da. b) Time-course of *N-O* bond cleavage and KAK^{Ub}I formation from the Ub(1-75)-α-thioester and KAK^{aux}I in a buffer consisting of 200 mM MPAA, 100 mM NaH_2PO_4 at pH 7.3 under the indicated conditions. Error bars represent the standard deviation from three independent measurements.

Table 3.1. Requirements for auxiliary removal under non-denaturing conditions.

		Reaction Conditions	Yield
Buffer	**1**	200 mM MPAA, 10 mM TCEP, 50 mM tris, 150 mM NaCl, pH 7.5	61%
	2	200 mM MPAA, 10 mM TCEP, 100 mM Na_2HPO_4, pH 7.5	58%
Additives	**3**	10 mM TCEP, 50 mM tris, 150 mM NaCl, pH 7.5	n.d.
	4	200 mM MPAA, 50 mM tris, 150 mM NaCl, pH 7.5	63%
[MPAA]	**5**	100 mM MPAA, 100 mM Na_2HPO_4, pH 7.3	56%
	6	50 mM MPAA, 100 mM Na_2HPO_4, pH 7.3	55%
	7	5 mM MPAA, 100 mM Na_2HPO_4, pH 7.3	2%
Contaminant	**8**	200 mM HPLC-purified[a] MPAA, 100 mM Na_2HPO_4, pH 7.3	72%
Oxidized MPAA	**9**	100 mM MPAA disulfide, 100 mM Na_2HPO_4, pH 7.3	n.d.
pH	**10**	200 mM MPAA, 100 mM Na_2HPO_4, pH 6.0	2%
	11	200 mM MPAA, 100 mM Na_2HPO_4, pH 7.3	73%
	12	200 mM MPAA, 100 mM Na_2HPO_4, pH 8.0	58%
Oxygen	**13**	200 mM MPAA, 100 mM Na_2HPO_4, pH 7.3, degassed[b]	2%
	14	200 mM MPAA, 100 mM Na_2HPO_4, pH 8.5, degassed[b]	5%

[a]MPAA was dissolved in 30% buffer B and purified by C18 preparative HPLC employing a gradient of 30-80% B over 60 min. [b]Buffer was subjected to three cycles of freeze-thaw degassing under argon and kept under argon atmosphere. n.d.= no detectable *N-O* bond cleavage.

3.2.2 Comparison of aromatic and aliphatic thiols for *N-O* bond reduction

To ascertain the generalizability of the *N-O* bond cleavage reaction, we tested a range of aromatic and aliphatic thiols with the ubiquitylated ligation product $KAK^{Ub(aux)}I$, bearing the auxiliary at the site of ligation. With the exception of 4-hydroxythiophenol, all aromatic thiols tested undertook *N-O* bond cleavage (**Table 3.2**, entries 1-5). Interestingly, 4-hydroxythiophenol inhibited MPAA mediated *N-O* bond cleavage when equal amounts of both were present in the reaction mixture. In contrast, none of the aliphatic thiols led to *N-O* bond reduction under identical buffer conditions (**Table 3.2**, entries 6-10). This suggests that near a neutral pH the *N-O* bond is compatible with aliphatic thiol additives commonly employed in native chemical

Table 3.2. Aromatic and aliphatic thiol pK_a values and corresponding yields in *N-O* bond cleavage assays.

Entry	Thiol	pKa[22,24]	Yield (%)
1	4-Nitrothiophenol	4.5	59%
2	3-Mercaptobenzoic acid	5.8	81%
3	4-Mercaptophenylacetic acid	6.6	73%
4	4-Aminothiophenol	6.9	73%
5	4-Hydroxythiophenol	7.0	n.d.
6	2,2,2-Trifluoroethanethiol	7.6	n.d.
7	L-Glutathione	9.1	n.d.
8	D,L-Dithiothreitol	9.2, 10.1	n.d.
9	2-Mercaptoethanesulfonic acid	9.2	n.d.
10	2-Methyl-2-propanethiol	11.2	n.d.

n.d. = no detectable *N-O* bond cleavage

ligations, such as 2-mercaptoethanesulfonic acid (MESNa), and that it is also stable to the biological reducing agent glutathione. One key difference between the two compound classes tested is that the aromatic thiols are >50% deprotonated at pH 7.3, while the aliphatic thiols are largely protonated. The importance of a deprotonated thiolate species was suggested by the fact that *N-O* bond cleavage by MPAA was dramatically reduced at pH 6.0 (**Table 3.1**, entries 10-12).

3.2.3 Mechanistic investigation of the *N-O* bond cleavage reaction in ubiquitylated peptides

Examples of bioactive compounds with chemically labile *N-O* bonds include pro-drug forms of the duocarmycin and CC-1065 class of antitumor agents.[25] These were proposed to undergo conversion to a biologically active form upon cleavage of the *N-O* bond by nucleophilic thiols within the tumor microenvironment. Because we observed an increase in *N-O* bond cleavage with increasing pH, a nucleophilic mechanism may in principle also be invoked for the aromatic thiols. However, freeze-thaw degassing the reaction mixtures under an argon atmosphere sufficed to inhibit the reaction with MPAA at pH 7.3-8.5 (**Figure 3.1b** and **Table 3.1**, entries 13-14), ruling out a purely nucleophilic mechanism and suggesting a key role for dissolved oxygen. In support of the latter, increasing the reaction headspace, and

thereby the ratio of molecular oxygen to thiol from 1:1 to 3.5:1 (mol:mol), resulted in halving the reaction time to 24 h (**Figure 3.1b**).

Molecular oxygen may act as a terminal electron acceptor and favor the formation of aromatic thiyl radicals from aromatic thiolates.[26] This process is more prevalent for aromatic rather than aliphatic thiols, owing to their relatively weaker *S-H* bonds, and thiophenols have been established as excellent hydrogen atom donors in organic synthesis.[27] One indication that thiyl radicals may be present under the reaction conditions was our observation of the oxidized MPAA disulfide species. To test the possibility that aromatic thiyl radicals are spontaneously formed in buffered aqueous solutions at pH 7.3, we incubated a range of aromatic thiols with the co-factor nicotinamide adenine dinucleotide (NADH) (**Figure 3.2**). Thiyl radicals react with NADH to yield the NAD$^{\bullet}$ radical and the consumption of NADH is readily detected by a decrease in absorbance at 340 nm.[28] In buffers that favored *N-O* bond reduction, we also observed a dramatic decrease in NADH concentration. Importantly, and consistent with their inability to reduce the *N-O* bond,

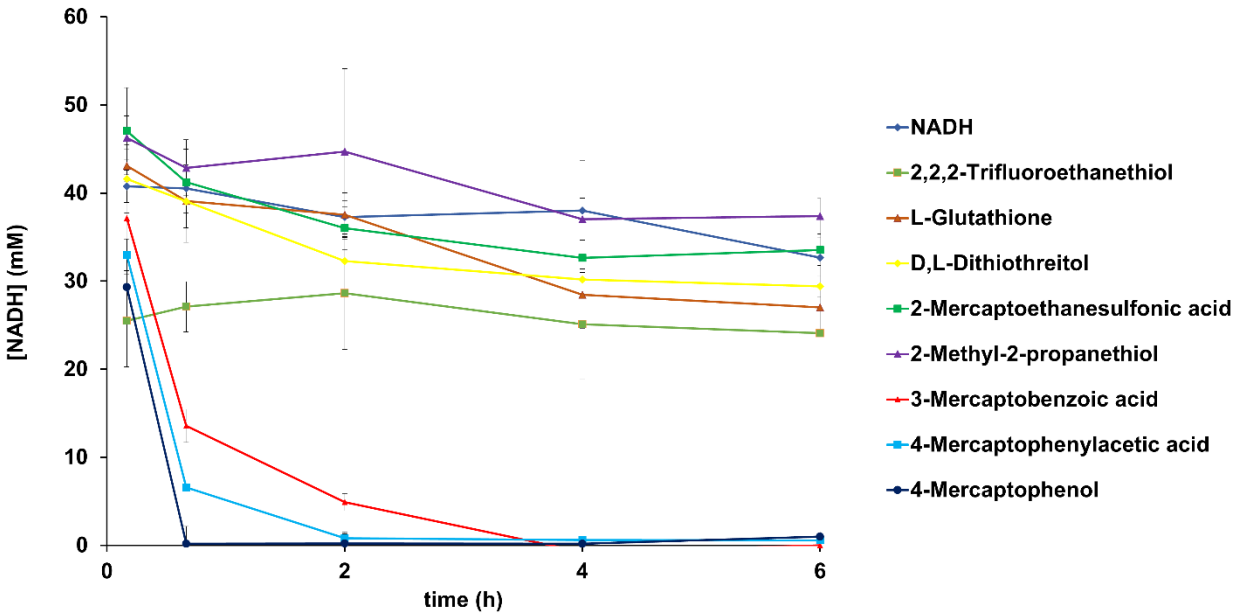


Figure 3.2. Time course of NADH oxidation by aromatic and aliphatic thiols. NADH was dissolved at 40 mM in a solution of 100 mM Na_2HPO_4, pH 7.3, containing 200 mM of the indicated thiol. NADH was detected by absorbance at 340 nm. NADH concentration at each time point was calculated from a standard curve of NADH in buffer.

we did not observe similar oxidation of NADH in reactions with aliphatic thiols 6-10 in **Table 3.2** over a 6 h time-course. The disparity in reducing nature of aromatic and aliphatic thiols was further seen by their reaction with methyl viologen (MV^{2+}). A rapid increase in absorption at 605 nm characteristic of the single-electron transfer reduction product $MV^{\bullet+}$ was observed only with aromatic thiols, suggesting the formation of a strongly reducing species in solution.[29]

Electron paramagnetic resonance (EPR) is a widely employed technique to detect species with unpaired electrons. However, thiyl radicals are generally not directly detectable by EPR in solution due to the large spin-orbit coupling constant of sulfur, which leads to fast relaxation of the electron spin.[30] Hence we attempted indirect EPR detection of the MPAA radical in solution at room temperature by spin-trapping with the compound 5,5-dimethyl-1-pyrroline-*N*-oxide (DMPO).[31] Gratifyingly, we observed the formation of nitroxide radicals in a solution containing 50 mM MPAA and 100 mM DMPO dissolved in 50% (v/v) aqueous DMF (**Figure 3.3a**). This was attributed to the addition of a thiyl radical into DMPO. The spectrum exhibited 6 lines characteristic of DMPO-trapped thiyl radicals,[22] which could be simulated with a_N = 14.32 G and a_H = 16.31 G (**Figure 3.3b**), giving $a_N < a_H$ as expected for DMPO-thiyl radical adducts in aqueous solution.[32,33] The inclusion of 50 mM Na_2HPO_4, pH 7.5 in the reaction altered the appearance of the EPR spectrum (**Figure 3.3c**), but controls lacking MPAA did not show EPR signal under these conditions. Furthermore, trapping of the MPAA thiyl radical was confirmed by ESI-MS (**Figure 3.3d**). Importantly, alkylation of MPAA *in situ* with 70 mM 2-Iodoacetamide for 1.5 hours precluded the appearance of an EPR signal, providing further evidence for a thiyl radical as the reactive species (**Figure 3.3e**). In support of MPAA thiyl radical formation in the presence of 50 mM Na_2HPO_4, we recapitulated the spectrum observed in **Figure 3.3c** by adding stock Na_2HPO_4 to 50 mM final concentration in the 50% (v/v) aqueous DMF mixture (**Figure 3.4**). Next, we included 50 mM H_2O_2 in the solution containing 50 mM MPAA and 100 mM DMPO dissolved in 50% (v/v) aqueous DMF in an attempt to enhance MPAA radical formation. We observed a spectrum identical to that without H_2O_2 present, yet with much greater signal intensity, and again observed the mass of the DMPO-MPAA adduct (**Figure 3.5a**). However, we did not see any significant nitroxide radicals

forming in the absence of MPAA, and any DMPO-OH adduct expected to arise from H_2O_2 alone was only observable by ESI-MS (**Figure 3.5b**). Finally, if the MPAA radical is crucial for *N*-*O* bond cleavage, we expected its trapping by DMPO to inhibit the reaction. Indeed, we found that the addition of 1 M DMPO to a reaction containing 100 mM MPAA inhibited product formation over the course of 24 h.

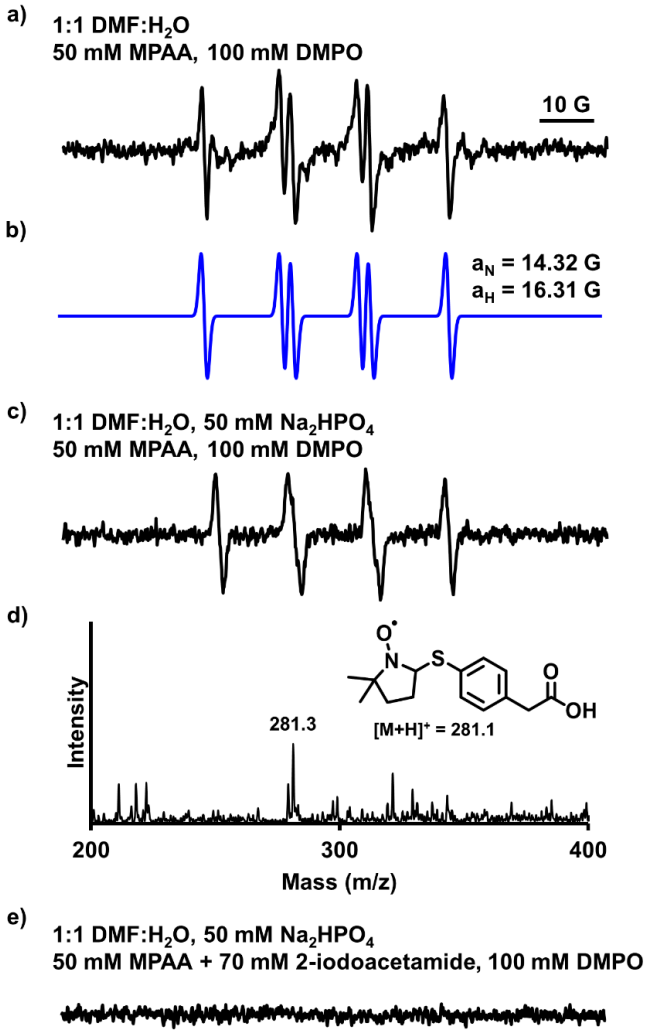


Figure 3.3. EPR spectra of DMPO/$^{\bullet}$*S*-Ar adduct. a) Spectrum obtained upon incubating 50 mM MPAA and 100 mM DMPO in a 1:1 water: DMF mixture at 25 °C. b) Computer simulation of the spectrum observed in (a) with hyperfine splitting constants a_N = 14.32 G and a_H = 16.31 G.[32] c) Spectrum obtained upon incubating 50 mM MPAA and 100 mM DMPO in 50 mM Na_2HPO_4 at pH 7.5, in a 1:1 water: DMF mixture at 25 °C. d) ESI-MS spectrum obtained by LC-ESI-MS analysis of the reaction components in (c). Inset shows the proposed radical adduct. e) EPR scan obtained upon pre-incubating 50 mM MPAA with 70 mM 2-Iodoacetamide for 1.5 h followed by 100 mM DMPO in 50 mM Na_2HPO_4 at pH 7.5, in a 1:1 water: DMF mixture at 25 °C. Incubation of 50 mM MPAA and 100 mM DMPO in 50 mM Na_2HPO_4 at pH 7.5, in a 1:1 water: DMF mixture at 25 °C for 1.5 h without the addition of 2-Iodoacetamide resulted in a spectrum similar to that seen in (c). Spectrometer settings: microwave power, 20 mW; modulation amplitude, 1.0 G; time constant, 163 ms; scan rate, 0.6 G/s.

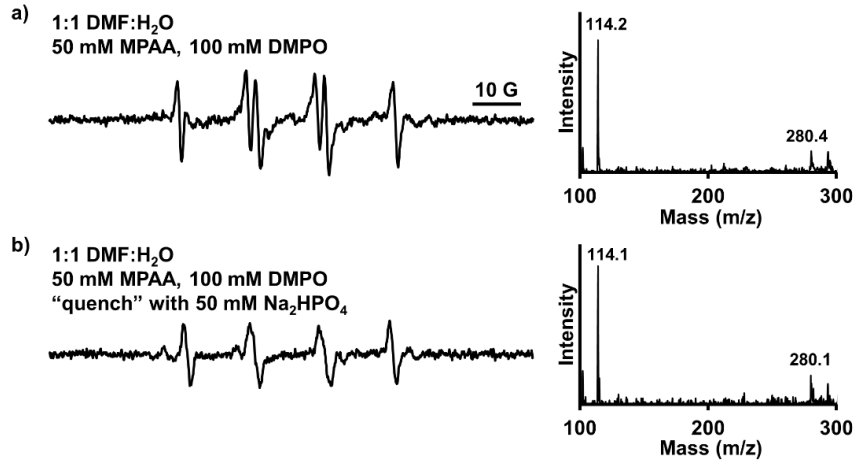


Figure 3.4. EPR spectra of DMPO/•*S*-Ar adduct in buffered and unbuffered conditions. a) EPR spectrum obtained upon incubating 100 mM DMPO in a 1:1 water: DMF mixture at 25 °C, and corresponding spectrum obtained from LC-ESI-MS analysis of the reaction mixture. b) EPR spectrum obtained upon incubating 100 mM DMPO in a 1:1 water: DMF mixture at 25 °C followed by "quenching" with 600 mM Na_2HPO_4, pH 7.5, for a final concentration of 50 mM Na_2HPO_4, and corresponding spectrum obtained from LC-ESI-MS analysis of the reaction mixture. Mass at 114 m/z identified as unreacted DMPO. Spectrometer settings: microwave power, 20 mW; modulation amplitude, 1.0 G; time constant, 163 ms; scan rate, 0.6 G/s. LC-ESI-MS analysis employed a gradient of 30-100% D over 10 min. Spectra are averages of the ESI-MS signal between R_t = 6-7 min.

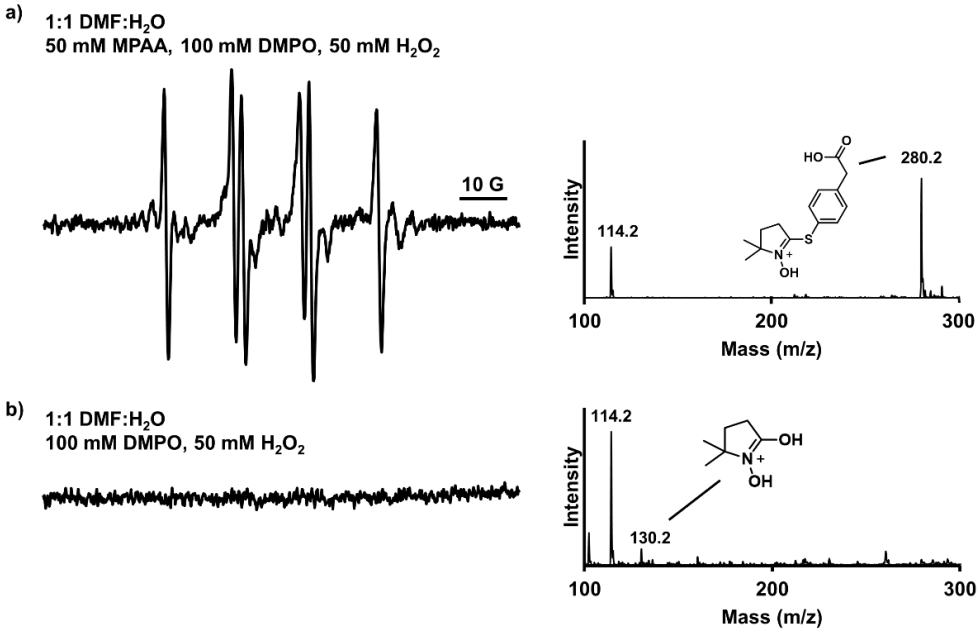
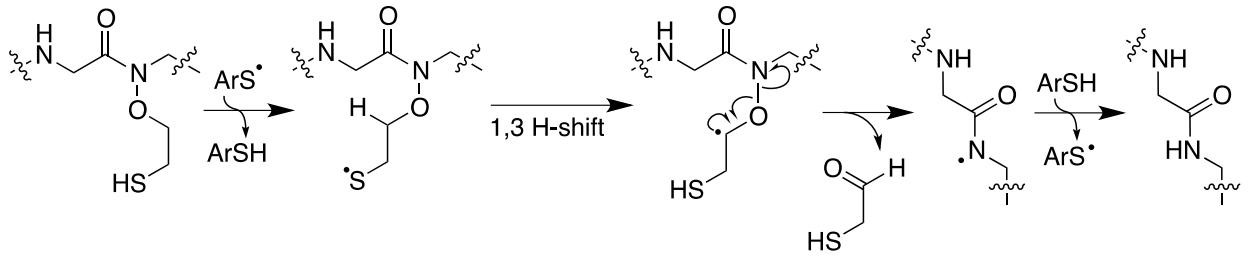


Figure 3.5. EPR spectra of DMPO/•*S*-Ar adduct in the presence of H_2O_2. a) EPR spectrum obtained upon incubating 50 mM MPAA, 100 mM DMPO, and 50 mM H_2O_2 in a 1:1 water: DMF mixture at 25 °C, and corresponding spectrum obtained from LC-ESI-MS analysis of the reaction mixture. b) EPR spectrum obtained upon incubating 100 mM DMPO and 50 mM H_2O_2 in a 1:1 water: DMF mixture at 25 °C, and corresponding spectrum obtained from LC-ESI-MS analysis of the reaction mixture. Mass at 114 m/z identified as unreacted DMPO. Spectrometer settings: microwave power, 20 mW; modulation amplitude, 1.0 G; time constant, 163 ms; scan rate, 0.6 G/s. LC-ESI-MS analysis employed a gradient of 30-100% D over 10 min. Spectra are averages of the ESI-MS signal between R_t = 6-7 min.

One potential pathway for radical mediated *N*-*O* bond cleavage is by formation of a thiyl radical in the auxiliary (**Scheme 3.2**). A 1,3-sigmatropic rearrangement of the thiyl radical would result in a carbon-centered radical adjacent to the low energy *N*-*O* bond, which favors its homolysis.[34] In order to test this mechanism we alkylated the auxiliary thiol in the ligation product with *N*-(2-chloroethyl)-*N,N*-dimethylammonium chloride, thereby precluding formation of a thiyl radical. Upon treatment with 200 mM MPAA at pH 7.3 we still observed efficient *N*-*O* bond cleavage in the *S*-alkylated product, indicating that a substrate-derived thiyl radical is not essential for the reaction to proceed. An alternative pathway for *N*-*O* bond cleavage is direct reduction by a reducing agent. In thinking of reducing species that are generated by a combination of aromatic thiolates and thiyl radicals, we considered the possibility of a disulfide radical anion. The formation of this high-energy species has been observed with small molecule thiols such as cysteine and glutathione,[35] and in the active site of the enzyme ribonucleotide reductase.[36] The presence of an unpaired electron in an antibonding σ* orbital renders the disulfide radical anion a strongly reducing yet transient species, in equilibrium with the dissociated radical and thiolate forms.[37]



Scheme 3.2. Proposed mechanism of *N*-*O* bond cleavage from an auxiliary-based thiyl radical.

We surmised that slow formation of the thiyl radical by molecular oxygen in the absence of added radical initiators, along with the transient nature of the disulfide radical anion, together contribute to the slow kinetics of *N*-*O* bond cleavage. An initial attempt to increase the rate of MPAA-mediated *N*-*O* bond cleavage by including the water-soluble radical initiator, 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) in our reactions proved unfruitful. This is likely due to the rapid quenching of the carbon-centered radical prior to the formation of significant amounts of the disulfide radical anion and is consistent with the

reported inhibition of radical-mediated desulfurization of cysteine residues by MPAA, which also employs VA-044.[24]

We also considered that superoxide, formed en route to the thiyl radical, may act as a reducing agent.[38] To test this possibility, we utilized the well-established superoxide generating system consisting of xanthine oxidase (XO) and its substrate hypoxanthine.[39] XO catalyzes the conversion of hypoxanthine first to xanthine and then to uric acid, and superoxide is released at each step. We performed this reaction in the presence of $KAK^{Ub(aux)}I$, with hypoxanthine at 4-fold excess relative to the auxiliary-containing test substrate. By monitoring the appearance of uric acid at an absorbance of 290 nm, we observed that all of the hypoxanthine was converted to uric acid within the first minute of reaction, producing a burst of superoxide. However, even after 24 h no cleavage of the auxiliary was observed, strengthening our hypothesis that a disulfide radical anion is the likely reductant (**Table 3.3**, entry 1).

Known methods to generate disulfide radical anions from thiols or disulfides in solution include flash photolysis,[37] pulse radiolysis,[40] or cyclic voltammetry,[41] all of which are technically challenging in the presence of folded protein substrates. Therefore, we wondered if a mild oxidant, such as hydrogen peroxide

Table 3.3. Conditions for auxiliary removal in the presence of hydrogen peroxide[a] or superoxide.

	Reaction Conditions	Time (h)	Yield
1	100 mM Na_2HPO_4, pH 7.5, 0.32 mM hypoxanthine, xanthine oxidase[b]	24	n.d.
2	200 mM MPAA, 50 mM H_2O_2, 100 mM Na_2HPO_4, pH 7.3	4	79%
3	200 mM MPAA, 50 mM EDTA, 100 mM Na_2HPO_4, pH 7.3	24	4%
4	200 mM MPAA, 1 mM $FeCl_2$, 100 mM Na_2HPO_4, pH 7.3	24	87%
5	50 mM H_2O_2, 100 mM Na_2HPO_4, pH 7.3	24	n.d.
6	50 mM H_2O_2, 1 mM $FeCl_2$, 100 mM Na_2HPO_4, pH 7.3	24	n.d.
7	200 mM MPAA, 50 mM H_2O_2, 100 mM Na_2HPO_4, pH 7.3, degassed[c]	4	68%

[a]Reaction conditions were evaluated for auxiliary removal from the ligation product $KAK^{Ub(aux)}I$. $KAK^{Ub(aux)}I$ (10 nmol, ~0.1 mg) was dissolved in 100 µL of buffer. To this was added 0.51 µL of a 30% (9.8 M) solution of H_2O_2 for a final concentration of 50 mM H_2O_2. The reaction was incubated in a 1.5 mL microcentrifuge tube at 25 °C. [b]$KAK^{Ub(aux)}I$ was dissolved at 0.04 mM in 100 µL of 50 mM Na_2HPO_4, pH 7.5 in the presence of 0.32 mM hypoxanthine and 0.14 u/mL xanthine oxidase at 25 °C for 24 h. [c]Buffer was subjected to three cycles of freeze-thaw degassing under argon and kept under argon atmosphere. n.d.= no detectable *N-O* bond cleavage.

(H_2O_2), would facilitate the formation of thiyl radicals by Fenton chemistry (**Figure 3.6a**).[42,43] Indeed, the inclusion of 50 mM H_2O_2 with 200 mM MPAA led to a significant increase in the rate of product formation, requiring only 4 h to attain maximal conversion with no detectable amounts of undesired protein oxidation (**Figure 3.1b** and **Table 3.3**, entry 2). The chelation of free metal ions with 50 mM EDTA effectively inhibited the reaction, indicating the key role for trace metal ions in generating thiyl radicals (**Table 3.3**, entry 3). We consistently found that trace metals in buffer components were sufficient for the reaction to proceed. However, if this is not the case, we found that $FeCl_2$ may be added to at least 1 mM without deleterious effect on the reaction yield (**Table 3.3**, entry 4). Importantly, 50 mM H_2O_2 alone or mixed with 1 mM $FeCl_2$ did not yield detectable product in the absence of MPAA, proving that Fenton chemistry alone cannot undertake *N*-*O* bond cleavage (**Table 3.3**, entries 5-6). Finally, freeze-thaw degassing a solution of 50 mM H_2O_2 and 200 mM MPAA failed to prevent reductive chemistry, which confirmed our hypothesis that H_2O_2 can favor thiyl radical formation even in the absence of molecular oxygen (**Figure 3.6a** and **Table 3.3**, entry 7).

3.2.4 Computational studies of *N*-*O* bond cleavage

We next undertook *ab initio* quantum chemistry calculations to interrogate the feasibility of disulfide radical anion formation and its reactivity toward the *N*-*O* bond. The relative redox potentials ($\Delta E°$) for (1) electron transfer between various thiols and hydrogen peroxide (**Figure 3.6b**), and (2) subsequent electron transfer between the disulfide radical anion and a model diglycine compound, **1** (**Figure 3.6c**), were computed. $\Delta E°$ values for the first step were obtained by calculating the free energy changes of the redox reactions using the Gaussian 09 program package.[44] Equilibrium geometries of all species were located via geometry optimization and the thermal corrections were evaluated at the B3LYP/6-31G* level of theory, with the solvent effect modeled using the polarizable continuum model (PCM). The electronic energies at the equilibrium geometries were computed at the B3LYP/6-311++G** level of theory and the results are summarized in **Table 3.4**. Our calculations revealed that protonated aliphatic thiols do not favor disulfide radical anion formation (**Table 3.4**, entries 5-8). In contrast, deprotonated aromatic thiolates can be oxidized

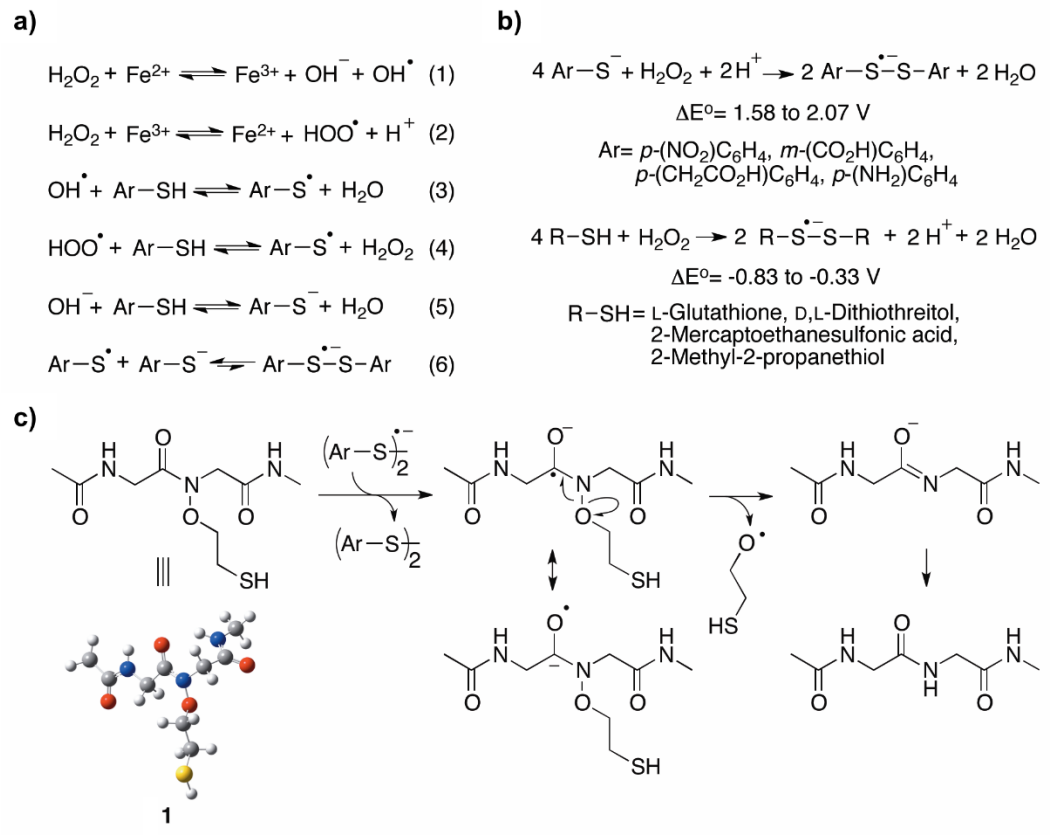


Figure 3.6. Formation of disulfide radical anions and their role in *N-O* bond cleavage. a) Production of aromatic thiyl radicals mediated by trace-metal-catalyzed Fenton chemistry (1-4) and their combination with aromatic thiolates to form disulfide radical anions (5-6). b) Net chemical equations for the formation of disulfide radical anions from aromatic thiolates and aliphatic thiols at pH 7.3. The calculated range of standard redox potentials is indicated for compounds from each class of molecules. c) Proposed mechanism for disulfide radical anion-mediated *N-O* bond cleavage in the model compound **1**.

by hydrogen peroxide to form disulfide radical anions (**Table 3.4**, entries 1-4). For MPAA, the calculated ΔE° of 2.07 V (**Table 3.4**, entry 3) indicates that disulfide radical anion formation is thermodynamically favored (according to the Nernst equation). However, the rate of formation is not predictable by computation, and disulfide radical anions are known to exist in an equilibrium that favors the dissociated radical and thiolate species,[33] which likely underlies the slow kinetics of reduction. Next, we focused on electron transfer from the MPAA disulfide radical anion to **1** (**Figure 3.6c**). Based on the calculated electron densities in the highest occupied molecular orbital (HOMO) of the disulfide radical anion and the lowest unoccupied molecular orbital (LUMO) in the diglycine peptide, two possible pathways exist. We found both

pathways to be thermodynamically permissible, with $\Delta E°$ of 0.53 V for the production of a β-mercaptoethyloxyl radical, and $\Delta E°$ of 0.04 V for the production of the β-mercaptoethanolate anion. When up to 9 explicit H_2O molecules in varying combinations were included in these calculations, most instances within the test set favored the β-mercaptoethanolate anion. Finally, we also examined the thermodynamics of direct electron transfer from the MPAA thiolate to **1** and found this process to be energetically unfavorable, which suggests that the thiolate form alone cannot act as a reductant.

3.2.5 Mechanistic studies with a model diglycine compound

In order to unambiguously identify the β-mercaptoethanol predicted by our proposed mechanism (**Figure 3.6c**), we synthesized the *S*-trityl-protected form of the model dipeptide, compound **2** (**Figure 3.7a**). Compound **2** was sparingly soluble in water and hence subjected to reduction with 200 mM MPAA in a buffer consisting of 100 mM Na_2HPO_4, pH 7.3 in 50% (v/v) aqueous DMF. Consistent with results obtained from protein substrates, and the detection of a MPAA radical under these conditions, we observed cleavage of the *N-O* bond over 24 h. The *S*-trityl-protected β-mercaptoethanol was isolated and confirmed by NMR (**Figure 3.7b**). Surprisingly, in the presence of 50 mM H_2O_2 and 200 mM MPAA, complete *N-O* bond cleavage in 20 mM of **2** was observed in 10 min. This may reflect the greater accessibility of the labile bond in **2** than in ubiquitylated peptides, or more productive electron transfer arising from a smaller number of

Table 3.4. Calculated redox potentials for disulfide radical anion formation from aliphatic and aromatic thiols at pH 7.3.

Entry	Thiol	$\Delta E°$ (V)
1	4-Nitrothiophenol	1.58
2	3-Mercaptobenzoic acid	1.98
3	4-Mercaptophenylacetic acid	2.07
4	4-Aminothiophenol	2.03
5	L-Glutathione	-0.40
6	D,L-Dithiothreitol	-0.33
7	2-Mercaptoethanesulfonic acid	-0.43
8	2-Methyl-2-propanethiol	-0.83

competing amide bonds than in ubiquitin. That trityl protection of the auxiliary thiol did not prevent *N-O* bond cleavage underscores the fact that electron transfer to the *N-O* bond occurs directly from the reducing species. As expected, sequestration of trace metal ions by treatment of the buffer with a metal-chelating resin also inhibited *N-O* bond cleavage in the model compound.

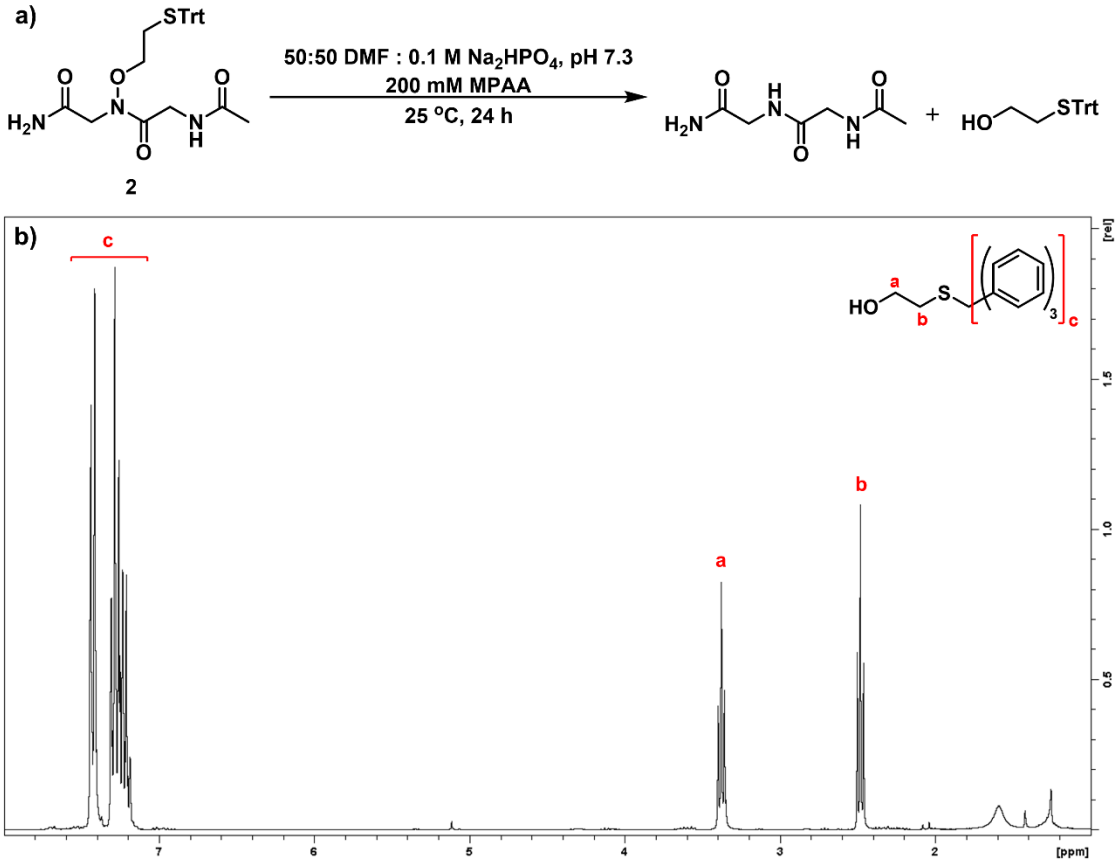


Figure 3.7. *N-O* bond reduction in model dipeptide 2. a) Formation of 2-(tritylthio)ethanol during MPAA-mediated *N-O* bond cleavage. b) ^{1}H-NMR of 2-(tritylthio)ethanol. Peak at 1.59 ppm (s) identified as H_2O.

3.2.6 A one-pot ligation and reduction strategy for native chemical ubiquitylation

Current chemical ubiquitylation methodologies are not optimal for native folded proteins.[24,45] Therefore, an efficient one-pot method to perform native ubiquitylation is highly desirable. With this in mind, we sought to improve the yield of one-pot auxiliary-mediated ubiquitylation by limiting the formation of Ub(1-75)-COOH. Slow hydrolysis of both the Ub(1-75)-α-thioester and the disubstituted amide in the ligation product contribute to this undesired side-product, which can in principle be alleviated by enhancing the kinetics of both ligation and *N*-*O* bond reduction. Surprisingly, we found that conducting the ligation reaction without MPAA significantly avoided Ub(1-75)-COOH formation by preventing premature auxiliary removal in the starting materials. The subsequent addition of 200 mM MPAA to the crude ligation mixture and incubation for an additional 24 h at 25 °C generated the final reduced ubiquitylated peptide in 82% overall yield. This represents a 20% higher yield over reactions where MPAA was added at the start of ligation.

3.2.7 Synthesis of full-length SUMOylated histone H4

As an initial test of the auxiliary's utility in the context of folded proteins, we incorporated non-denaturing MPAA mediated auxiliary removal into the semisynthesis of full-length SUMOylated human histone H4 (suH4). Although histone SUMOylation was first reported over a decade ago, very little is known regarding its functional role in human chromatin.[46] Access to quantities of H4 site-specifically conjugated with the C-terminus of SUMO-3 at Lys12 is crucial for biochemical investigations of the role for SUMOylation in regulating chromatin structure and function.[15] We therefore devised a synthetic strategy for suH4 (**Figure 3.8a**). The H4(1-14) peptidyl hydrazide was synthesized using 9-fluorenylmethoxycarbonyl (Fmoc) chemistry on the solid phase, with Gly92 of SUMO-3 and the ligation auxiliary attached to Lys12. Following release from the solid-phase and global deprotection, the peptide was ligated to the SUMO-3(2-91)C47S-α-thioester to yield the SUMOylated peptide. One key challenge we anticipated was conversion of the hydrazide to a thioester without cleaving the *N*-*O* bond. However, we found the auxiliary to be completely stable to both diazotization and thioester formation, which employed $NaNO_2$ at pH 3.0 followed by displacement of the resulting azide with 100 mM MPAA.[47] Complete retention of the auxiliary through these

steps highlights its utility in diverse native chemical ligation strategies. The SUMOylated H4 peptide α-thioester was then reacted with a truncated H4(15-102) protein containing the A15C mutation at its N-terminus to facilitate native chemical ligation. Ligation proceeded over 24 h to afford 2.1 mg of the ligated product, retaining the ligation auxiliary, in 66% purified yield. The ligation product was then dissolved in a buffer consisting of 100 mM Na_2HPO_4, 200 mM MPAA, pH 7.3, and *N*-*O* bond cleavage allowed at 25 °C over 24 h to yield the reduced compound (**Figure 3.8b**). Importantly, Cys15 in H4 was unaffected by MPAA-mediated auxiliary removal, demonstrating the compatibility of this reaction with folded proteins containing Cys residues. In the terminal step, the full-length SUMOylated histone H4 A15C mutant was desulfurized to yield the desired suH4 in 41% yield over the last two steps (**Figure 3.8c**).

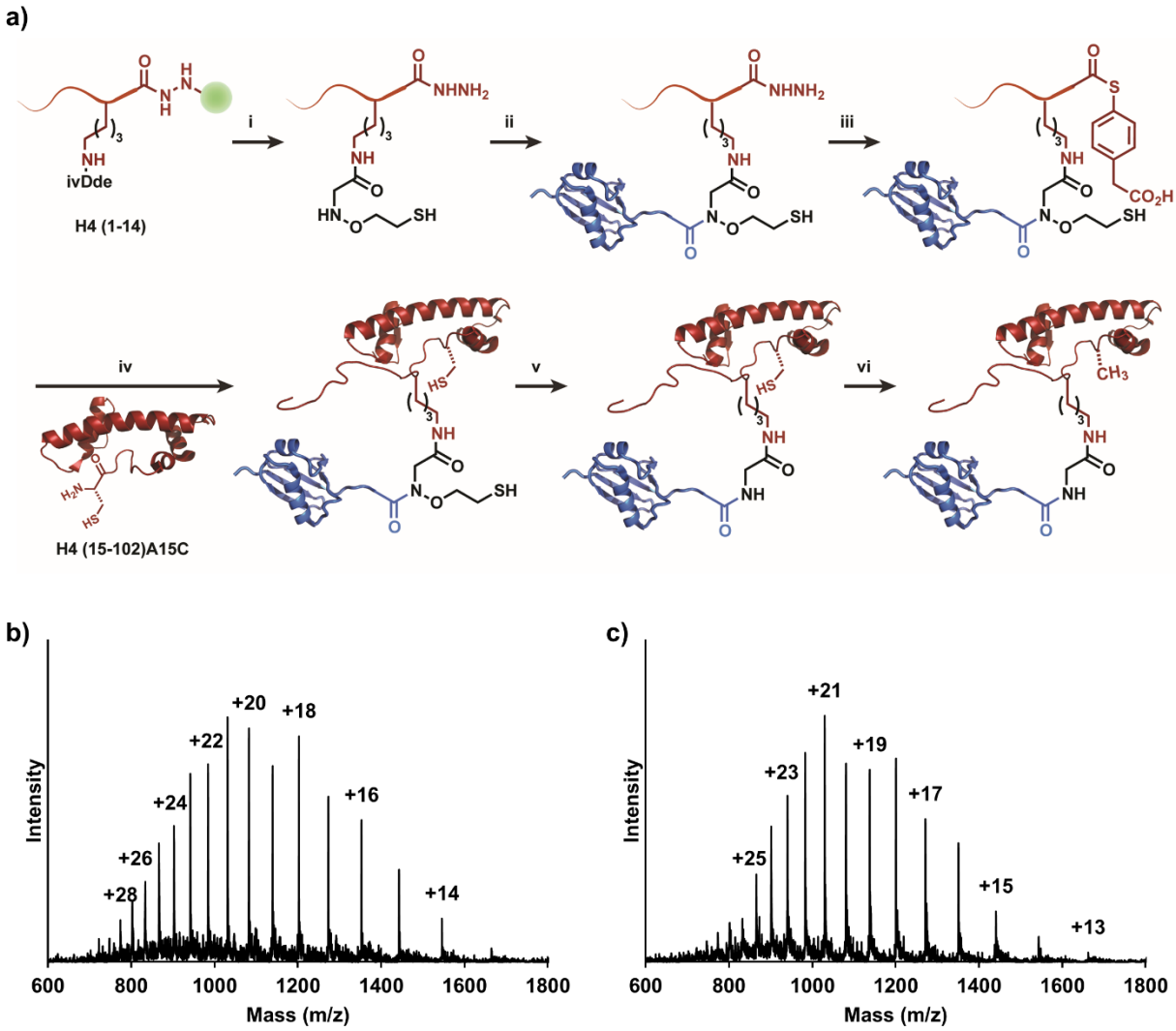


Figure 3.8. Semisynthesis of full-length SUMOylated histone H4. a) i, Site-specific coupling of the ligation auxiliary to H4(1-14) Lys12 followed by acidolytic release of the fully unprotected peptidyl hydrazide from the solid-phase. ii, Expressed protein ligation of H4(1-14)^{aux}-C(O)NHNH$_2$ with SUMO-3(2-91)C47S-α-thioester to generate SUMOylated peptide hydrazide, H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$. iii, Conversion of H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$ to the C-terminal MPAA α-thioester. iv, Native chemical ligation of the H4(1-14)$^{Su(C47S)(aux)}$-C(O)SR MPAA thioester with the H4(15-102)A15C truncant protein to yield full-length SUMOylated H4(A15C), with retention of the ligation auxiliary. v, Selective removal of the ligation auxiliary with 200 mM MPAA under non-denaturing conditions to yield SUMOylated H4(A15C). vi, Radical mediated desulfurization of Cys15 in H4 to generate full-length wild-type suH4. b) ESI-MS of purified H4(A15C)$^{Su(C47S)}$. Calculated *m/z* [M+H]$^+$ 21,628.7 Da, observed 21,634.0 ± 6.1 Da. c) ESI-MS of purified suH4. Calculated *m/z* [M+H]$^+$ 21,596.7 Da, observed 21,601.0 ± 6.2 Da. ivDde = 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)-3-methylbutyl. PDB code for SUMO-3, 1U4A. PDB code for H4, 1KX5.

3.2.8 Synthesis of full-length SUMOylated histone H2B

The ultimate goal of our chemical strategy is complete compatibility with native folded proteins containing Cys residues. We envision future applications wherein a suitably protected ligation auxiliary is directly incorporated in target proteins by employing an amber suppression strategy, which may be purified from producer strains prior to native ubiquitylation/SUMOylation. Having demonstrated that MPAA can mediate auxiliary removal from SUMOylated histone H4 under non-denaturing conditions, we sought to perform (1) auxiliary deprotection, (2) SUMOylation and (3) auxiliary removal on a second model protein without intermediate denaturation and purification steps. We were particularly interested in histone H2B as it is SUMOylated at its C-terminal Lys120 (suH2B)[48] and genetic experiments suggest that SUMOylation recapitulates the genomic occupancy of H2B K120 ubiquitylation.[49] However, similar to suH4, the role of suH2B in chromatin regulation also awaits *in vitro* biochemical investigation.

Toward the semisynthesis of suH2B, we first generated full-length histone H2B bearing a protected ligation auxiliary as the entry point for testing our methodology. To ensure that the auxiliary protecting group could be removed under native conditions, we synthesized a 2-nitrobenzyl protected form (**3**) from 2-nitrobenzyl chloride and *N*-(2-bromoethoxy)phthalimide[12] (**Figure 3.9a**). The protected auxiliary **3** was incorporated at Lys120 of the H2B(117-125) C-terminal peptide with an Ala to Cys mutation at position 117 (**Figure 3.9b**). After acidolytic cleavage from the solid phase, the peptide **4** was ligated via its N-terminal Cys to an H2B(1-116)-α-thioester to generate full-length H2B(A117C), **5** (**Figure 3.10a**). The product **5** was folded by dialysis into 50 mM Na_2HPO_4, pH 7.5 (**Figure 3.10b**), and all subsequent steps were performed under folded conditions. First, complete deprotection of the auxiliary thiol was achieved by irradiation with 365 nm light for 3.5 h in the presence of ascorbic acid, semicarbazide, and DTT. The folded state of the photo-deprotected protein was confirmed by circular dichroism (CD) (**Figure 3.10c-d**). The deprotected H2B(A117C)^{aux}, **6**, was then ligated under non-denaturing conditions to the SUMO-3(2-91)C47S-α-thioester over 48 h to yield the ligation product H2B(A117C)$^{Su(C47S)aux}$, **7** (**Figure 3.10e-f**). The SUMOylated product **7** was then subjected to MPAA-mediated auxiliary removal for 24 h, yielding the ligation product lacking the

ligation auxiliary, H2B(A117C)$^{Su(C47S)}$ (**8**), in 15-30% yield over 2 steps (**Figure 3.10e, g**). Importantly, we observed no precipitation of the H2B species throughout these multiple manipulations. The final folded state of the reduced ligation product **8** was confirmed by size exclusion chromatography and CD.

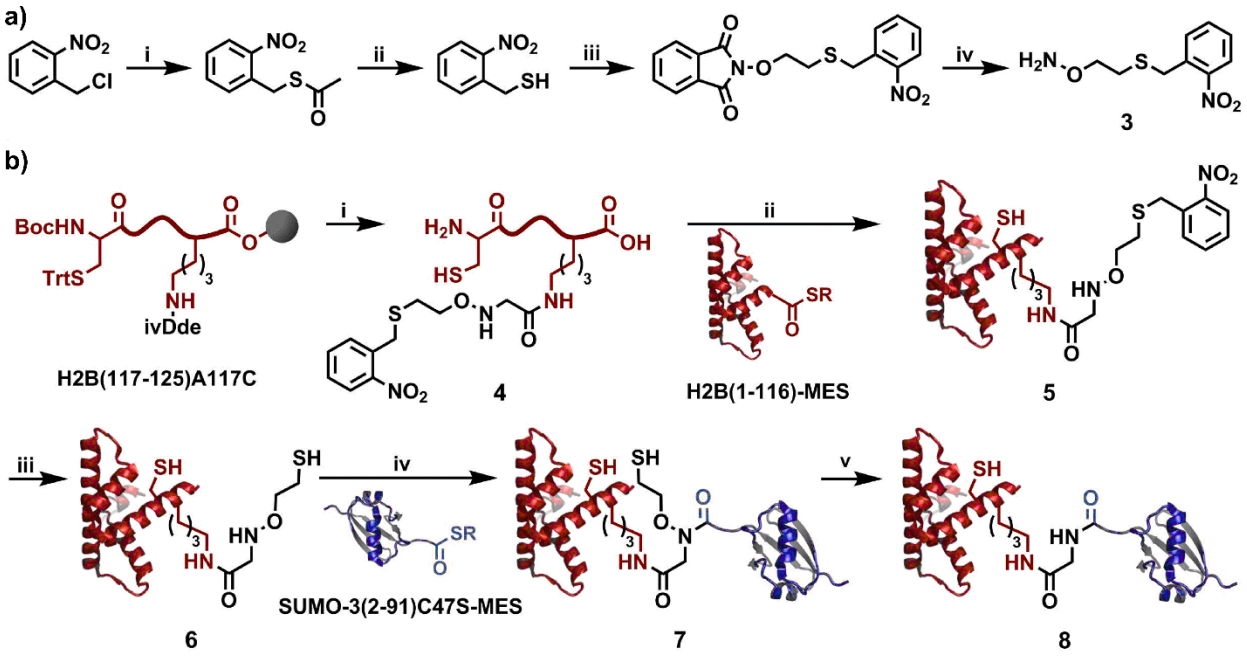


Figure 3.9. Semisynthesis of full-length SUMOylated histone H2B. a) Synthesis of photo-protected auxiliary **3**. i, $CH_3C(O)SH$, K_2CO_3, THF, 8h, 25 °C. ii, HCl, CH_3OH, 6 h, 60 °C, 75% (2 steps). iii, *N*-(2-bromoethoxy)phthalimide, Et_3N, DMSO, 4 h, 25 °C, 74%. iv, H_2NNH_2, CH_3Cl, 1 h, 25 °C, 98%. b) i, Site-specific coupling of **3** to H2B(117-125)A117C Lys120 followed by acidolytic release of the unprotected peptide, **4**, from the solid-phase. ii, Expressed protein ligation of **4** with H2B(1-116)-α-thioester to generate full-length H2B(A117C) with protected auxiliary at Lys120, **5**. iii, photolytic removal of the auxiliary protecting group to give H2B(A117C) with unprotected auxiliary at Lys120, **6**. iv, Expressed protein ligation of **6** with SUMO-3(2-91)C47S-α-thioester to generate SUMOylated H2B(A117C) **7**, with retention of the ligation auxiliary. v, Selective removal of the ligation auxiliary with 150 mM MPAA under non-denaturing conditions to yield SUMOylated H2B(A117C) **8**. ivDde = 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)-3-methylbutyl. PDB codes 1KXF (H2B) and 1U4A (SUMO-3).

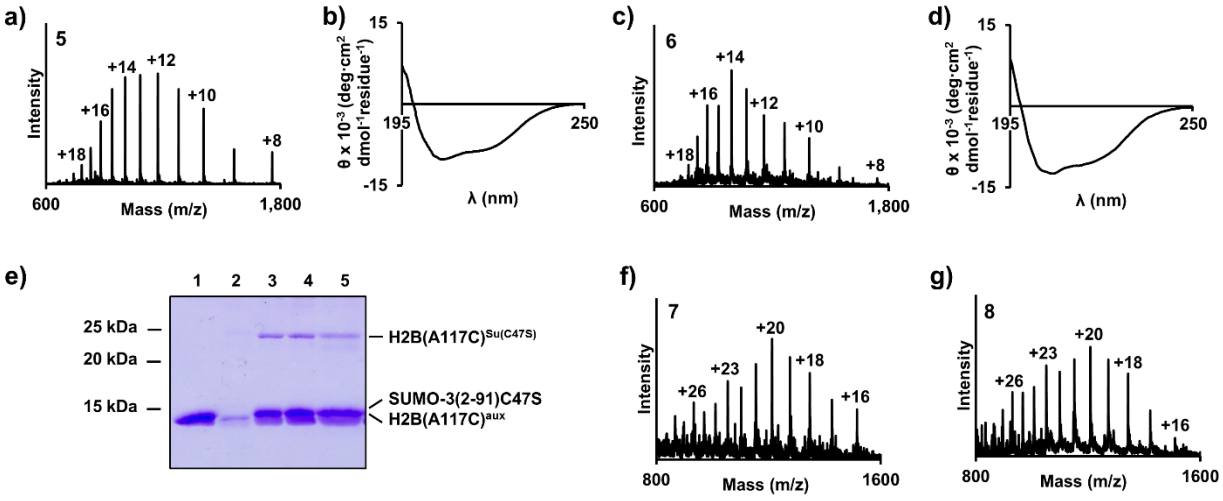


Figure 3.10. Photo-deprotection and SUMOylation of folded histone H2B. a) ESI-MS spectrum of H2B(A117C)^{photoaux} (**5**). Calculated *m/z* $[M+H]^+$ 14,059.2 Da, observed 14,062.7 ± 2.8 Da. b) Circular Dichroism spectrum of **5** in 50 mM Na_2HPO_4, pH 7.5. c) ESI-MS spectrum of H2B(A117C)^{aux} (**6**). Calculated *m/z* $[M+H]^+$ 13,924.1 Da, observed 13,925.8 ± 2.6 Da. d) Circular Dichroism spectrum of **6** in 50 mM Na_2HPO_4, pH 7.5. e) Coomassie-stained 15% SDS-PAGE gel of ligation between H2B(A117C)^{aux} (**6**) and SUMO3(2-91)C47S-α-thioester under non-denaturing conditions. Lane 1: SUMO3(2-91)C47S-MES. Lane 2: H2B(A117C)^{aux}. Lane 3: 24 h ligation reaction. Lane 4: 48 h ligation reaction. Lane 5: 24 h MPAA-mediated *N-O* bond cleavage auxiliary removal reaction. f) ESI-MS spectrum of the ligation product, H2B(A117C)$^{Su(C47S)aux}$ (**7**). Calculated *m/z* $[M+H]^+$ 24,225.6 Da, observed 24,230.9 ± 4.0 Da. g) ESI-MS spectrum of the final product, H2B(A117C)$^{Su(C47S)}$ (**8**). Calculated *m/z* $[M+H]^+$ 24,149.6 Da, observed 24,153.1 ± 3.2 Da.

3.3 Conclusion and outlook

Access to homogenous protein substrates site-specifically modified by Ubls is critical for biophysical and biochemical investigations aimed at understanding their many diverse functions. To this end, we applied the 2-(aminooxy)ethanethiol ligation auxiliary to the SUMOylation of folded proteins under non-denaturing conditions. Removal of the auxiliary requires reductive cleavage of an *N-O* bond, which is a frequently encountered moiety in organic chemistry. Several methods exist for its cleavage, including $TiCl_3$,[50] catalytic hydrogenation,[51] Na/Hg amalgams,[52] and SmI_2.[53] More recently, neutral organic super-electron donors were demonstrated to reduce *N-O* bonds in Weinreb amides.[54] However, the application of any of these reagents to folded proteins in aqueous buffers is extremely challenging. Our discovery of an unprecedented *N-O* bond reductive cleavage reaction sets the stage for new applications that would benefit from the controlled reversal of this low-energy bond, including non-denaturing protein semisynthesis.

Both experimental and computational investigations support our hypothesis that *N*-*O* bond cleavage involves the formation of a transient disulfide radical anion species. Multiple observations toward this include the requirement for an oxidant, the necessity for trace metal ions, and the detection of aromatic thiyl radicals. Importantly, our ability to readily control the timing of *N*-*O* bond cleavage is particularly useful as demonstrated by the one-pot synthesis of the ubiquitylated peptide, KAK^{Ub}I. As highlighted in our syntheses of the full-length SUMOylated human histones H4 and H2B, the extremely mild reductive strategy may also be applied toward the modification of native folded proteins by Ubls in aqueous buffers. Indeed, the semisynthesis of native suH4 will, for the first time, permit detailed biochemical study of this poorly understood modification, as detailed in Chapter 4.

One logical caveat to this strategy is a potential incompatibility with proteins containing structural disulfide bonds, which exist primarily in secreted or membrane proteins. However, the scope of this method still extends to the vast majority of intracellular proteins.[55] Finally, having established the stability of the ligation auxiliary to the intracellular reductant glutathione, our immediate future efforts in this area are focused on an amber-codon-suppression strategy to incorporate the auxiliary into natively folded proteins that are inaccessible by fragment-based semisynthetic approaches.

3.4 Experimental procedures

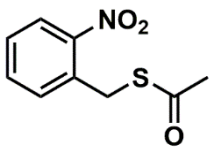
3.4.1 General Methods

Rink-amide resin (0.46 mmol/g substitution) was purchased from Chem-Impex (Wood Dale, IL). 2-chlorotrityl chloride resin (1.33 mmol/g substitution) and Wang resin (0.4-0.6 mmol/g substitution) were purchased from EMD Millipore. Standard Fmoc-L-amino acids were purchased from AGTC Bioproducts (Wilmington, MA) or AnaSpec (Fremont, CA). All other chemical reagents were purchased from Sigma-Aldrich Chemical Company (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). DNA synthesis and gene sequencing were performed by Integrated DNA Technologies (Coralville, IA) and Genewiz (South Plainfeld, NJ), respectively. Plasmid mini-prep, PCR purification and gel extraction kits were purchased from Qiagen (Valencia, CA). Restriction enzymes were purchased from New England BioLabs (Ipswitch, MA) or Fermentas (Thermo Fisher Scientific, Philadelphia, PA). Chitin beads for purification of intein-CBD fusion proteins were purchased from New England BioLabs. Ni-NTA resin for purification of His_6-tagged proteins was purchased from Thermo Scientific (Waltham, MA). Solid phase peptide synthesis (SPPS) was performed on a Liberty Blue Automated Microwave Peptide Synthesizer (CEM Corporation, Matthews, NC). Analytical reversed-phase HPLC (RP-HPLC) was performed on a Varian (Palo Alto, CA) ProStar HPLC with a Grace-Vydac (Deerfield, IL) C4 or C18 column (5 micron, 150 x 4.6 mm) employing 0.1% TFA in water (A) and 90% CH_3CN, 0.1% TFA in water (B) as the mobile phases. Typical analytical gradients were 0-73% B over 30 min at a flow rate of 1 mL/min. Preparative scale purifications were conducted on a Grace-Vydac C4 or C18 column (10 micron, 250 x 22 mm) at a flow rate of 9 mL/min. Semi-preparative scale purifications were conducted on a Grace-Vydac C4 or C18 column (5 micron, 250 x 10 mm) at a flow rate of 3.5 mL/min. Mass spectrometric analysis was conducted on a Bruker (Billerica, MA) Esquire ESI-MS instrument. Analytical reversed-phase liquid chromatography-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) employing 5% CH_3CN, 1% AcOH in water (C) and CH_3CN, 1% AcOH (D) as the mobile phases. Typical analytical gradients were 0-100% D over 40 min at a flow rate of 0.2 mL/min. NMR spectra were recorded on Bruker Avance AV-300,

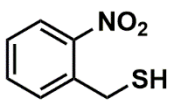
AV-301, or AV-500 instruments. Circular dichroism measurements were performed on a JASCO (Easton, MD) J-720 spectropolarimeter. Size exclusion chromatography (SEC) was performed on a Bio-Rad Bio-Sil SEC 250 column (5 micron, 300 x 7.8 mm) at a flow rate of 1 mL/min.

3.4.2 Synthesis of the ligation auxiliary

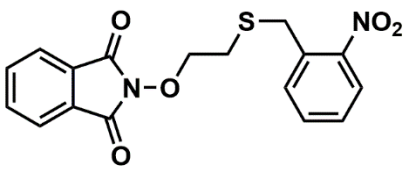
The ligation auxiliary *O*-(2-(tritylthio)ethyl)hydroxylamine was prepared over 3 steps from *N*-hydroxyphthalimide as described in Chapter 2.[12]



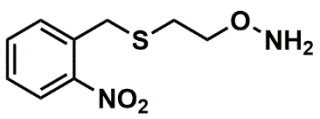
***S*-(2-nitrobenzyl)ethanethioate.** 4.04 g of 2-nitrobenzyl chloride (23.6 mmol) and 3.86 g of potassium carbonate, in separate flasks, were suspended in 24 mL and 48 mL of THF, respectively. To the stirring suspension of potassium carbonate was added 2 mL (30.5 mmol) of thioacetic acid. After 30 min the nitrobenzyl chloride solution was cannulated into this mixture. Reaction progress was followed by TLC. After 8 h the starting material had been consumed and volatiles were stripped. The crude material was dissolved in methylene chloride, washed with bicarbonate solution, then water, and the organic phase dried *in vacuo* to give 5.87 g of solids, which were carried on to the next reaction without further purification.



(2-nitrobenzyl)thiol. Crude *S*-(2-nitrobenzyl)ethanethioate (5.87 g) was dissolved in 100 mL of methanol. To this solution was added 24 mL of concentrated hydrochloric acid (36%, 23.7 mmol), and the reaction allowed to proceed at 60 °C for 6 h, after which TLC indicated complete hydrolysis of the starting material. Methanol was removed *in vacuo*, and the resulting material purified by column chromatography with 90:10 hexane: ethyl acetate to give 2.30 g of (2-nitrobenzyl)thiol (75%, 2 steps). ^{1}H NMR (300 MHz,$CDCl_3$): δ 8.03 (d, 1H, *J* = 8.16), 7.61-7.44 (3H), 4.04 (d, 2H, *J* = 8.46), 2.17 (t, 1H, *J* = 8.46) (**Figure 3.11**). ^{13}C NMR (301 MHz, $CDCl_3$): δ 147.89, 137.00, 133.74, 131.58, 128.26, 125.35, 26.45 (**Figure 3.12**).



***N*-(2-((2-nitrobenzyl)thio)ethoxy)phthalimide.** 2.3 g of (2-nitrobenzyl)thiol (14.2 mmol) and 5.73 g of *N*-(2-bromoethoxy)phthalimide (21.3 mmol, prepared as described in Chapter 2[12]) were dissolved in 25 mL DMSO. To this stirring solution was added 3 mL (21.3 mmol) of triethylamine, and the reaction allowed to proceed for 4 h at room temperature. After TLC indicated complete consumption of (2-nitrobenzyl)thiol, 300 mL of water was added and the suspension extracted three times with methylene chloride. The organic layer was dried over anhydrous sodium sulfate and removed *in vacuo* to give crude product, which was recrystallized from a mixture of methylene chloride: hexane to give 3.74 g of *N*-(2-((2-nitrobenzyl)thio)ethoxy)phthalimide as pale yellow crystals (74%). ^{1}H NMR (500 MHz,$CDCl_3$): δ 7.99 (d, 1H, *J* = 8.15), 7.86-7.80 (4H), 7.56-7.44 (3H), 4.32 (t, 2H, *J* = 7.00), 4.21 (s, 2H), 2.87 (t, 2H, *J* = 7.00) (**Figure 3.13**). ^{13}C NMR (500 MHz, $CDCl_3$): δ 163.45, 148.64, 134.65, 133.82, 133.16, 132.15, 128.85, 128.41, 125.50, 123.65, 77.56, 33.77, 29.39 (**Figure 3.14**).



***O*-(2-((2-nitrobenzyl)thio)ethyl)hydroxylamine (3).** 2 g (5.58 mmol) of *N*-(2-((2-nitrobenzyl)thio)ethoxy)phthalimide was dissolved in 100 mL chloroform. To this stirring solution was added 1.04 mL hydrazine hydrate (50% solution, 8.4 mmol), and the reaction monitored by TLC. After 1 h the reaction was complete. Solids were removed by vacuum filtration and the organic layer washed three times with 200 mL water. Solvent was removed *in vacuo* to give 1.25 g of *O*-(2-((2-nitrobenzyl)thio)ethyl)hydroxylamine as a yellow oil (98%). ^{1}H NMR (300 MHz,$CDCl_3$): δ 7.99 (d, 1H, *J* = 7.91), 7.61-7.41 (3H), 5.43 (s, 2H), 4.13 (s, 2H), 3.80 (t, 2H, *J* = 6.41), 2.69 (t, 2H, *J* = 6.41) (**Figure 3.15**). ^{13}C NMR (500 MHz, $CDCl_3$): δ 148.84, 134.18, 132.97, 131.98, 128.19, 125.33, 74.56, 33.62, 30.35 (**Figure 3.16**). ESI-MS calculated *m/z* $[M+H]^+$ 229.1 Da, observed 228.9 Da.

3.4.3 Solid-phase peptide synthesis

Synthesis of Boc-KAKI-Rink-amide resin

The peptide H_2N-KAKI-C(O)NH_2 was synthesized by microwave-assisted SPPS on a 0.1 mmol scale employing standard Fmoc-based N^{α}-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 sec 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 Lys targeted for ubiquitylation was orthogonally protected with the 1-(4,4-dimethyl-2,6-dioxocyclohexylidine)-3-methylbutyl (ivDde) protecting group. The peptide was protected at the α-NH_2 position with the *tert*-butoxycarbonyl (Boc) group by reaction with di-*tert*-butyl dicarbonate (0.8 mmol) and DIEA (1.6 mmol) in DMF for 2 h.

Synthesis of BocHN-H4(1-14)-2-chlorotrityl hydrazine resin

The peptide BocHN-SGRGKGGKGLGKGG-C(O)$NHNH_2$ corresponding to the first 14 N-terminal residues of the human histone H4 protein was synthesized by SPPS on a 0.25 mmol scale employing standard Fmoc-based N^{α}-deprotection chemistry. Briefly, 2-chlorotrityl hydrazine resin was prepared by reacting 2-chlorotrityl chloride resin (1.33 mmol/g) in a 10% solution of hydrazine in DMF at 30 °C for 30 min.[47] The reaction was repeated one time with fresh hydrazine solution. The resin was then treated with a 10% methanol in DMF solution for 10 min to cap any unreacted sites on the resin. The first amino acid, Gly, was coupled in 4 molar excess. The coupling reaction containing Fmoc-Gly-OH (1.0 mmol), *O*-(6-chlorobenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HCTU, 0.95 mmol) , and DIEA (2.0 mmol) proceeded for 60 min at 30 °C. From glycinyl 2-chlorotrityl hydrazine resin each remaining amino acid was coupled in 5 molar excess based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 20% piperidine in DMF for 3 min at 75 °C. Coupling reactions were undertaken for 5 min at 75 °C with a mixture of Fmoc-amino acid (1.31 mmol), HBTU (1.28 mmol) and DIEA (2.75 mmol) in DMF. For Arg, an additional coupling reaction was performed for 25 min at 75 °C. The Lys at position 12

was orthogonally protected with the ivDde protecting group. The peptide was protected at the α-NH_2 position with Boc group by reaction with di-*tert*-butyl dicarbonate (2.0 mmol) and DIEA (4.0 mmol) in DMF for 2 h.

Synthesis of BocHN-H2B(117-125, A117C)-Wang resin

The peptide BocHN-CVTKYTSAK-C(O)OH corresponding to the last 9 C-terminal residues of the human histone H2B protein, with an Ala to Cys mutation at position 117, was synthesized by SPPS on a 0.1 mmol scale employing standard Fmoc-based N^{α}-deprotection chemistry. From Fmoc-Lys(Boc) Wang resin (0.21 g, 0.4-0.6 mmol/g) each amino acid was coupled in 6 molar excess based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 20% piperidine in DMF for 3 min at 75 °C. Coupling reactions were undertaken for 5 min at 75 °C with a mixture of Fmoc-amino acid (0.6 mmol), *N,N'*-diisopropylcarbodiimide (DIC, 0.6 mmol), ethyl (hydroxyimino)cyanoacetate (Oxyma, 0.6 mmol) and DIEA (0.06 mmol) in DMF. Cys was coupled for 10 min at 50 °C. The Lys at position 120 was orthogonally protected with the ivDde protecting group. The peptide was Boc protected at the α-NH_2 position by reaction with di-*tert*-butyl dicarbonate (0.8 mmol) and DIEA (1.6 mmol) in DMF for 2 hours.

Attachment of the ligation auxiliary

Deprotection of the ivDde group was achieved by reacting resin bound peptide with a solution of 5% hydrazine in DMF for 5 min. This deprotection was repeated three times. The peptidyl resin was then coupled to bromoacetic acid (8-fold molar excess) with DIC (8-fold molar excess) in DMF for 45 min at room temperature. The coupling was repeated once. Subsequently, dry peptidyl resin was placed in a solution containing 9 equivalents of auxiliary (0.25 M in DMSO). The auxiliary *O*-(2-(tritylthio)ethyl)hydroxylamine ("aux") was incubated with H_2N-KAKI-C(O)NH_2 and BocHN-H4(1-14)-C(O)$NHNH_2$, and the photo-labile auxiliary *O*-(2-((2-nitrobenzyl)thio)ethyl)hydroxylamine ("photoaux") was incubated with BocHN-H2B(117-125, A117C)-C(O)OH. The reactions were shaken for 24 hours at room temperature. Completion of the displacement was judged by test cleavage and subsequent ESI-MS analysis. Peptide was cleaved and deprotected by reaction of resin at 20 μL/mg with Reagent K (TFA: thioanisole: H_2O: phenol: 1,2-

ethanedithiol 82.5:5:5:5:2.5 v/v)[56] for 1.5 hours at room temperature, then precipitated and washed 2 times with cold diethyl ether. Dry peptide was dissolved in RP-HPLC buffer A and purified by C18 preparative and semi-preparative RP-HPLC with a gradient of 0-25% B for KAK^{aux}I and 0-50% B for H4(1-14)^{aux}-C(O)NHNH$_2$ and H2B(117-125, A117C)^{photoaux}-C(O)OH. This yielded 38% of the peptide-auxiliary conjugate KAK^{aux}I, 9% of the peptide-auxiliary conjugate H4(1-14)^{aux}-C(O)NHNH$_2$, and 25% of the peptide-auxiliary conjugate H2B(117-125, A117C)^{photoaux}-C(O)OH based on initial resin loading (**Figure 3.17**).

ESI-MS of KAK^{aux}I. Calculated *m/z* [M+H]$^+$ 591.4 Da, observed 591.8 Da.

ESI-MS of H4(1-14)^{aux}-CONHNH$_2$. Calculated *m/z* [M+H]$^+$ 1,363.6 Da, observed 1,363.8 Da.

ESI-MS of H2B(117-125, A117C)^{photoaux}-C(O)OH (**4**). Calculated *m/z* [M+H]$^+$ 1,268.6 Da, observed 1,269.0 Da.

3.4.4 Molecular cloning of SUMO-3(2-91)C47S, Ub(1-75), and H4(15-102)A15C

The plasmid pTXB1-Ub(1-76)-AvaDNAE-AAFN-His$_6$ containing the human ubiquitin gene, *ub(1-76)*, was a kind gift from the Muir lab at Princeton University,[57] and was used to generate the plasmid pTXB1-Ub(1-75)-AvaDNAE-AAFN-His$_6$, which lacks the C-terminal Gly of ubiquitin. The plasmid pTXB1-SUMO3(1-92)C47S, containing the human SUMO-3 gene *Smt3(1-92)* with a C47S mutation,[15] was used to generate the plasmid pTXB1-SUMO3(1-91)C47S, which lacks the C-terminal Gly of SUMO-3. The plasmid pET15b-His$_6$-[TEV]-H4, containing the human histone H4 gene with a Tobacco Etch Virus (TEV) protease cleavage sequence between the His$_6$ tag and the N-terminus of H4,[15] was used to generate the plasmid pET15b-His$_6$-[TEV]-H4(15-102)A15C, which lacks the first 14 residues of histone H4 and bears the mutation A15C. Modified plasmids were prepared from their respective templates by site-directed mutagenesis (QuikChange kit, Agilent Technologies, Santa Clara, CA) with the following primers:

Primer	**DNA Sequence (5'- to -3')**
hSUMO-3(1-91)C47S-FP	ATCGACGTGTTCCAGCAGCAGACGGGATGCATCACGGGAGATGCACT AGTTGCC
hSUMO-3(1-91)C47S-RP	GGCAACTAGTGCATCTCCCGTGATGCATCCCGTCTGCTGCTGGAACAC GTCGAT
hUb(1-75)-AvaDnaE-AAFN-H6-FP	CTGCACCTGGTACTCCGTCTCAGAGGTTGCCTGAGCTATGATACCGAA GTGCTG

hUb(1-75)-AvaDnaE-AAFN-H6-RP	CAGCACTTCGGTATCATAGCTCAGGCAACCTCTGAGACGGAGTACCAG GTGCAG
hH4(15-102)A15C-FP	GGGAATTC*CATATG*GAAAACCTGTACTTCCAGTGCAAACGTCACCGTA AAGTTCTG (*NdeI*)
hH4-RP	GCCCGC*GGATCC*TCAACCACCGAAACCGTACAGGGTACGACCC (*BamHI*)

The desired gene sequences were confirmed by sequencing with the T7 forward primer (Genewiz).

3.4.5 Overexpression and purification of Ub(1-75)-MES

E. coli BL21(DE3) cells were transformed with the plasmid pTXB1-Ub(1-75)-AvaDNAE-AAFN-His_6. Cells were grown in 3 L Luria-Bertani medium supplemented with 100 µg/mL of Ampicillin at 37 °C with shaking at 250 rpm until OD_{600} reached ~0.6. Overexpression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cells were grown for an additional 18 h at 16 °C. The cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in 45 mL lysis buffer: 50 mM Na_2HPO_4, 300 mM NaCl, 5 mM imidazole, pH 8. Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The lysate supernatant was passed through a 0.45 µm filter then applied to a 6 mL Ni-NTA column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 1 h at 4 °C. The column was then washed with lysis buffer containing increasing concentrations of imidazole: 5 mM (5 Column Volumes, CV), 20 mM (5 CV), 50 mM (2.5 CV), 250 mM (5 CV). Pooled fractions containing Ub(1-75)-intein-His_6 were dialyzed into 2 L thiolysis buffer (100 mM Na_2HPO_4, 150 mM NaCl, 1 mM EDTA, 1 mM MESNa, pH 7.2) for 1 h at 4 °C twice. Thiolysis buffer containing 1 M MESNa was added to the dialyzed fraction pool for a final concentration of 100 mM MESNa. The thiolysis reaction was incubated at 30 °C for 18 h, after which Ub(1-75)-intein-His_6 was no longer present. The eluted Ub(1-75)-MES was purified by C18 preparative RP-HPLC employing a gradient of 30-60% B over 60 min. Fractions containing Ub(1-75)-MES were identified by ESI-MS. Typical yields were 5-7 mg/L of cell culture (**Figure 3.18**).

ESI-MS for Ub(1-75)-MES. Calculated *m/z* $[M+H]^+$ 8,632.8 Da, observed 8,632.9 ± 1.8 Da.

3.4.6 Non-denaturing expressed protein ligation of $KAK^{aux}I$ and Ub(1-75)-MES α-thioester

Ub(1-75)-MES α-thioester (0.038 μmol, 0.4 mg) and KAK^{aux}I (0.38 μmol, 0.23 mg) were dissolved in 200 μL of reaction buffer containing 50 mM tris, 150 mM NaCl, 10 mM TCEP, 200 mM 4-mercaptophenylacetic acid (MPAA), pH 7.3-7.5. The reaction proceeded at 25 °C for a total of 48 h. After incubation, the sample was treated with 50 mM TCEP at 4 °C for 30 min, acidified to pH ~3 with formic acid, and extracted once with diethyl ether to remove the majority of MPAA. The sample was analyzed by LC-ESI-MS employing a gradient of 0-100% D over 40 min.

3.4.7 Generation of KAK$^{Ub(aux)}$I for *N-O* bond cleavage tests by expressed protein ligation of KAK^{aux}I and Ub(1-75)-MES α-thioester

In a typical reaction, purified KAK^{aux}I (0.35 mg, 0.6 μmol) and Ub(1-75)-MES (0.12 μmol) were dissolved in 300 μL of a buffer consisting of 6 M Gn-HCl, 100 mM Na_2HPO_4, and 10 mM TCEP, pH 7.3. Ligation proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C18 analytical RP-HPLC employing a gradient of 30-50% B over 30 min. Typical yields were 60-80% of purified product (**Figure 3.19**). ESI-MS of KAK$^{Ub(aux)}$I. Calculated *m/z* $[M+H]^+$ 9,081.1 Da, observed 9,082.3 ± 2.1 Da.

3.4.8 Requirements for auxiliary removal under non-denaturing conditions

KAK$^{Ub(aux)}$I was subjected to various assay conditions to evaluate the reaction components necessary for auxiliary removal via *N-O* bond cleavage. KAK$^{Ub(aux)}$I (30 nmol, ~0.3 mg) was dissolved in 300 μL buffer and incubated in a 1.5 mL microcentrifuge tube for 48 h at 25 °C, after which the sample was combined with an equal volume 200 mM TCEP, 6 M Gn-HCl, 100 mM Na_2HPO_4, pH 7.3 and incubated at 4 °C for 30 min. The sample was then treated with 50 mM TCEP at 4 °C for 30 min, acidified to pH ~3 with formic acid, and extracted once with diethyl ether to remove the majority of MPAA. The sample was analyzed by LC-ESI-MS employing a gradient of 0-100% D over 40 min. Percent yield was determined by the percent auxiliary-free ligation product KAK^{Ub}I relative to other species present, determined by ESI-MS signal intensity for R_t = 14-17 min (**Figure 3.20**).

3.4.9 Time course of auxiliary removal

Condition 1

Auxiliary-containing ligation product (30 nmol, ~0.3 mg) was dissolved in 300 µL of 200 mM MPAA, 100 mM Na_2HPO_4, pH 7.3, and incubated in a 1.5 mL microcentrifuge tube at 25 °C for up to 48 h.

Condition 2

Auxiliary-containing ligation product (10 nmol, ~0.1 mg) was dissolved in 100 µL of 200 mM MPAA, 100 mM Na_2HPO_4, pH 7.3, and incubated in a 1.5 mL microcentrifuge tube at 25 °C for up to 24 h.

Condition 3

Auxiliary-containing ligation product (10 nmol, ~0.1 mg) was dissolved in 100 µL of 200 mM MPAA, 100 mM Na_2HPO_4, pH 7.3. To this was added 0.51 µL of a 30% (9.8 M) solution of H_2O_2 for a final concentration of 50 mM H_2O_2. The reaction was incubated in a 1.5 mL microcentrifuge tube at 25 °C for up to 8 h.

Analysis of time-points for auxiliary removal

An equal volume of 200 mM TCEP, 6 M Gn-HCl, 100 mM Na_2HPO_4, pH 7.5 was added to each sample, which was then incubated at 4 °C for 30 min. The solution pH was adjusted to 2-3 with formic acid, and one extraction with 0.5 mL diethyl ether was performed. Reaction progress was analyzed by LC-ESI-MS employing a gradient of 0-100% D over 40 min. Percent yield was determined by the percent auxiliary-free ligation product $KAK^{Ub}I$ relative to other species present, determined by ESI-MS signal intensity for R_t = 14-17 min.

3.4.10 Dependence of auxiliary removal on reduced thiol

The disulfide form of MPAA was generated by dissolving MPAA at 200 mM in 100 mM Na_2HPO_4, pH 8.0, in a tube with headspace filled with air equal to 10 times the liquid volume. The solution was shaken vigorously at room temperature for 48 h, after which time the pH was lowered to 1-2 with HCl causing the

MPAA species to precipitate. Acetonitrile was added until the precipitate dissolved (~40% of total volume). The MPAA disulfide, 2,2'-(disulfanediylbis(4,1-phenylene))diacetic acid, was purified by C18 preparative RP-HPLC employing a gradient of 30-80% B over 60 min. ^{1}H NMR (301 MHz, CD_3CN): δ 7.16 (d, 4H, *J* = 8.34 Hz), 6.92 (d, 4H, *J* = 8.34 Hz), 3.26 (s, 4H) (**Figure 3.21**). ESI-MS calculated *m/z* $[M-H]^-$ 333.4 Da, observed 333.9 Da.

3.4.11 Thiyl radical detection by oxidation of NADH

To investigate the formation of thiyl radicals under auxiliary-removal conditions, reduced NADH was dissolved at a final concentration of 40 mM in solutions containing 200 mM thiol and 100 mM Na_2HPO_4, pH 7.3. The solutions were incubated at 25 °C, protected from light. Absorbance at 340 nm was measured at various time points. A decrease in absorbance, due to oxidation of NADH, suggested the presence of thiyl radicals.[28]

To further confirm the ability of aromatic thiols to perform single-electron transfer reactions, solutions were prepared containing 200 mM of each thiol and 100 mM Na_2HPO_4, pH 7.3, and these solutions were added to dry aliquots of the radical indicator, methyl viologen (MV^{2+}), for a final concentration of 20 mM MV^{2+}. Both MV^{2+} and its two-electron reduction product (MV°) have absorbance maxima less than 400 nm. The single-electron reduction product $MV^{\bullet+}$, however, has a strong characteristic absorbance at 605-610 nm.[58] The resulting deep purple color was observed immediately upon mixing methyl viologen with the aromatic thiol solutions, but no color change occurred with aliphatic thiols, even after 24 h.

3.4.12 EPR experiments

A 200 mM MPAA stock solution was prepared in DMF. Aliquots from the stock solution were diluted with DMF and 50 mM Na_2HPO_4 at pH 7.5 to a final 1:1 water-DMF mixture. Then, 100 mM 5,5-Dimethyl-1-pyrroline *N*-oxide (DMPO) was added, the samples were vortexed, transferred to a flat cell and EPR spectra were recorded at room temperature on a Bruker EMX spectrometer equipped with a high sensitivity cavity

and operating at 9.65 GHz and 100 KHz field modulation. MPAA alkylation was performed by incubating 50 mM MPAA with 70 mM 2-Iodoacetamide in 50 mM Na_2HPO_4 at pH 7.5, in a 1:1 water-DMF for 1.5 h at 25 °C before DMPO addition. Parallel controls by pre-incubating 50 mM MPAA alone in 50 mM Na_2HPO_4 at pH 7.5, in a 1:1 water-DMF for 1.5 h at 25 °C before DMPO addition were also run (**Figure 3.22**). Computer simulation was performed using Winsim program from P.E.S.T.[32]

3.4.13 Detection of DMPO adducts by mass spectrometry

Reactions were prepared identically to the EPR experiments. Aliquots of 5 µL were analyzed by LC-ESI-MS employing a gradient of 30-100% D over 10 min. Adduct spectra were generated by averaging the ESI-MS signal between 6-7 min. R_t = 3-5 min for DMPO; R_t = 6-7 min for the DMPO-OH and DMPO-MPAA adducts; R_t = 9-10 min for MPAA. The DMPO-MPAA or DMPO-OH adduct radical may exist in the radical, oxidized, and reduced states when observed by ESI-MS,[59] and the oxidized form appeared to be most prevalent for reactions of DMPO with H_2O_2 or MPAA (**Figure 3.23**).

3.4.14 MPAA-mediated auxiliary removal in the presence of a radical quencher

$KAK^{Ub(aux)}I$ (50 µg, ~5 nmol) was dissolved in 58 µL of 100 mM MPAA, 100 mM Na_2HPO_4, pH 7.3 containing 1 M DMPO and incubated at 25 °C for 24 h. The reaction was then treated with TCEP as above, acidified to pH ~3 with formic acid, and analyzed by LC-ESI-MS employing a gradient of 0-100% D over 40 min. Percent yield was determined by the percent auxiliary-free ligation product $KAK^{Ub}I$ relative to other species present, determined by ESI-MS signal intensity for R_t = 14-17 min.

3.4.15 *S*-alkylation of $KAK^{Ub(aux)}I$ and MPAA-mediated auxiliary removal from the alkylated species

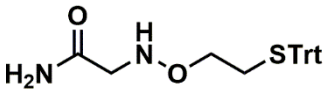
Alkylation of the pendant thiol in the ligation product $KAK^{Ub(aux)}I$ was performed under conditions established for selectively alkylating protein sulfhydryl groups.[60] Briefly, 1.2 mg of the ligation product was dissolved in 280 µL of alkylation buffer containing 1 M HEPES, 4 M Gn-HCl, and 10 mM Methionine at pH 7.8. To this was added 14 µL of 1 M DTT in alkylation buffer. The resulting mixture was incubated at 25 °C for 20 min

after which time 31 µL of a 1 M solution of *N*-(2-chloroethyl)-*N,N*-dimethylammonium chloride in alkylation buffer was added. The reaction was incubated at 25 °C for 2 h. Next, an additional 3.4 µL of 1 M DTT solution was added followed by a 30 min incubation at 25 °C to reduce any residual disulfides. Finally, an additional 31 µL of *N*-(2-chloroethyl)-*N,N*-dimethylammonium chloride in alkylation buffer was added and the reaction incubated for 2 h to ensure complete alkylation of all thiol groups. The reaction mixture was quenched with 10 µL of β-mercaptoethanol and purified by C18 analytical RP-HPLC employing a gradient of 30-50% B over 30 min. *S*-alkylated KAK$^{Ub(aux)}$I was characterized by ESI-MS. Calculated *m/z* $[M+H]^+$ 9,152.3 Da, observed 9,154.0 ± 2.3 Da (**Figure 3.24**). To evaluate whether MPAA-mediated auxiliary removal is dependent on the pendent thiol of the auxiliary, *S*-alkylated KAK$^{Ub(aux)}$I (0.135 mg) was dissolved in 0.135 mL buffer containing 200 mM MPAA, 100 mM Na_2HPO_4, pH 7.3, and incubated at 25 °C for 48 h. The reaction was reduced with TCEP and analyzed by LC-ESI-MS as above. The expected product KAK^{Ub}I was observed. Calculated *m/z* $[M+H]^+$ 9,005.1 Da, observed 9,004.8 ± 2.7 Da.

3.4.16 Effect of superoxide on *N-O* bond cleavage in KAK$^{Ub(aux)}$I

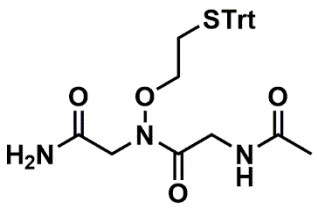
To rule out superoxide-mediated *N-O* bond cleavage, KAK$^{Ub(aux)}$I was subjected to the xanthine oxidase/hypoxanthine reaction. Molecular oxygen is reduced to superoxide as hypoxanthine is oxidized to xanthine, and subsequently to uric acid.[61] KAK$^{Ub(aux)}$I was dissolved at 0.04 mM in 100 µL of 50 mM Na_2HPO_4, pH 7.5 in the presence of 0.08, 0.16, or 0.32 mM hypoxanthine and 0.14 u/mL xanthine oxidase (Sigma-Aldrich) at 25 °C for 24 h. Progress of the enzymatic reaction was monitored by an increase in the absorbance of uric acid at 290 nm (**Figure 3.25**).[62] Cleavage of the *N-O* bond was evaluated by LC-ESI-MS employing a gradient of 0-100% D over 40 min.

3.4.17 Synthesis of a model diglycine compound



2-((2-(tritylthio)ethoxy)amino)acetamide. To a stirring solution of *O*-(2-(tritylthio)ethyl)hydroxylamine (1.74 g, 5.19 mmol) in anhydrous DMF (15 mL) under a flow of argon was added a solution containing iodoacetamide (482 mg, 2.60 mmol) in

anhydrous DMF (10 mL). To this mixture was added DIEA (0.91 mL, 5.19 mmol) and the reaction stirred at room temperature until TLC revealed that starting material had been consumed (~20 h). Ice-cold, basified H_2O (250 mL, pH ~8) was added to the reaction mixture in DMF, causing a cloudy, white precipitate to form. The DMF/H_2O solution was extracted with ethyl acetate until no further precipitate was visible (~500 mL), then the ethyl acetate layer washed 3 times with an equal volume of basified H_2O. Solvent was removed *in vacuo* to yield crude product, which was purified by silica gel (70-230 mesh) column chromatography (10:90 hexane: ethyl acetate, 0.1% triethylamine) to give compound 2-((2-(tritylthio)ethoxy)amino)acetamide (0.88 g, 86%). ^{1}H NMR (301 MHz, $CDCl_3$): δ 7.43-7.19 (15H), 6.35 (s, 1H), 5.77 (t, 1H, *J* = 5.01 Hz), 5.43 (s, 1H), 3.56 (t, 2H, *J* = 6.46 Hz), 3.46 (d, 2H, *J* = 5.15 Hz), 2.44 (t, 2H, *J* = 6.46 Hz) (**Figure 3.26**). ^{13}C NMR (500 MHz, $CDCl_3$): δ 172.59, 144.69, 129.61, 127.97, 126.77, 72.30, 66.77, 54.85, 30.90 (**Figure 3.27**). ESI-MS calculated *m/z* $[M+Na]^+$ 415.5 Da, observed 415.2 Da.



2-acetamido-*N*-(2-amino-2-oxoethyl)-*N*-(2-(tritylthio)ethoxy)acetamide (2). A solution of 2-((2-(tritylthio)ethoxy)amino)acetamide (900 mg, 2.29 mmol), 1-hydroxybenzotriazole (930 mg, 6.88 mmol), *N*-acetylglycine (806 mg, 6.88 mmol), and DIEA (1.2 mL, 6.88 mmol) in 23 mL DMF was prepared over an ice bath. To this stirring solution was added *N,N'*-dicyclohexylcarbodiimide (1.42 g, 6.88 mmol). The reaction was stirred on ice for 90 min, then allowed to warm to room temperature overnight. The reaction proceeded for 24 h, at which time TLC revealed that starting material had been consumed. The white precipitate formed over the course of the reaction was removed by vacuum filtration. To the filtrate was added ice-cold H_2O (500 mL) causing a cloudy, white precipitate to form. The DMF/H_2O solution was extracted with ethyl acetate until no further precipitate was visible (~1.5 L), then the ethyl acetate layer washed 3 times with an equal volume of H_2O. Solvent was removed *in vacuo* to yield crude product, which was purified by silica gel (70-230 mesh) column chromatography (90:10 ethyl acetate: methanol, 0.1% triethylamine) to give model dipeptide **2**, 2-acetamido-*N*-(2-amino-2-oxoethyl)-*N*-(2-(tritylthio)ethoxy)acetamide (0.74 g, 66%). ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.43-7.21 (15H), 6.18 (t, 1H), 6.09

(s, 1H), 5.38 (s, 1H), 4.13 (s, 2H), 4.09 (d, 2H, *J* = 4.75 Hz), 3.53 (t, 2H, *J* = 6.55 Hz), 2.52 (t, 2H, *J* = 6.60 Hz), 2.03 (s, 3H) (**Figure 3.28**). ^{13}C NMR (500 MHz, DMF-D_7): δ 174.12, 171.74, 170.41, 146.67, 131.47, 130.13, 128.89, 81.13, 74.71, 68.87, 52.06, 43.37, 23.90 (**Figure 3.29**). ESI-MS calculated *m/z* $[M+Na]^+$ 514.6 Da, observed 514.3 Da.

3.4.18 *N-O* bond cleavage and characterization of the cleaved auxiliary in a model diglycine compound

Model dipeptide **2** was dissolved at 40 mM in 2.5 mL DMF. To this solution was added 2.5 mL of 100 mM Na_2HPO_4, pH 7.3 containing 400 mM MPAA. The reaction was incubated at 25 °C for a total of 24 h. At the end of 24 h, an equal volume of 100 mM Na_2HPO_4, pH 7.3 was added. A white precipitate formed and was collected by centrifugation, then washed 2 times with buffer. The dried precipitate was dissolved in ethyl acetate and passed through a silica plug with 100% ethyl acetate. Solvent was removed *in vacuo* to yield a single spot by TLC, identified as 2-(tritylthio)ethanol. ^{1}H NMR (301 MHz, $CDCl_3$): δ 7.31-7.19 (15H), 3.38 (t, 2H, *J* = 6.19 Hz), 2.48 (t, 2H, *J* = 6.19 Hz). ESI-MS calculated *m/z* $[M+Na]^+$ 343.4 Da, observed 343.3 Da.

Various conditions were evaluated for auxiliary removal in 100 µL reaction volumes of 50:50 DMF: 100 mM Na_2HPO_4, pH 7.3 containing 20 mM **2**, and incubated at 25 °C. Starting material disappearance (R_f = 0.5) and 2-(tritylthio)ethanol appearance (R_f = 0.9) were monitored by TLC in 90:10 ethyl acetate: methanol (**Table 3.5**).

3.4.19 Computational studies of *N-O* bond cleavage

Relative redox potential (ΔE°) values were obtained by calculating the free energy changes of redox reactions using the Gaussian 09 program package.[44] Equilibrium geometries of all species were located via geometry optimization and the thermal corrections were evaluated at the B3LYP/6-31G* level of theory, with the solvent effect modeled using the polarizable continuum model (PCM). The electronic energies at the equilibrium geometries were computed at the B3LYP/6-311++G** level of theory. Electron densities in

the highest occupied molecular orbital (HOMO) of the disulfide radical anion and the lowest unoccupied molecular orbital (LUMO) in the diglycine peptide were found (**Figure 3.30**). These values were used to calculate two reaction pathways, resulting in production of a β-mercaptoethyloxyl radical, or a β-mercaptoethanolate anion. Calculations were performed with up to 9 explicit H_2O molecules in varying combinations (**Table 3.6**).

3.4.20 One-pot ligation and auxiliary removal

Ub(1-75)-α-thioester (0.019 μmol, 0.2 mg) and $KAK^{aux}I$ (0.19 μmol, 0.12 mg) were dissolved in 100 μL of reaction buffer containing 50 mM tris, 150 mM NaCl, 0.5 mM MPAA, pH 7.3, and incubated at 25 °C for 24 h. Then, 200 uL of buffer containing 300 mM MPAA, 50 mM tris, 150 mM NaCl, pH 7.3, was added and the reaction incubated at 25 °C for 24 h, for a total of 48 h reaction time. After incubation the sample was treated with 50 mM TCEP at 4 °C for 30 min, acidified to pH ~3 with formic acid, and extracted one time with diethyl ether to remove the majority of MPAA. The sample was analyzed by LC-ESI-MS employing a gradient of 0-100% D over 40 min.

3.4.21 Overexpression and purification of SUMO-3(2-91)C47S-MES

E. coli BL21(DE3) cells were transformed with the plasmid pTXB1-SUMO-3(2-91)C47S. Cells were grown in 6 L Luria-Bertani medium supplemented with 100 μg/mL of Ampicillin at 37 °C with shaking at 250 rpm until OD_{600} ~0.6-0.8. Overexpression of the desired fusion proteins was induced by the addition of 0.3 mM IPTG and cells were grown for an additional 4 h at 25 °C. The cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in lysis buffer: PBS, pH 7.2 containing 1 mM 2-mercaptoethanesulfonic acid sodium salt (MESNa). Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The lysate supernatant was passed through a 0.45 μm filter then applied to a 30 mL chitin column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 12 h at 4 °C. The column was then washed with 20 column volumes (CV) of lysis buffer followed by 2 CV of PBS, pH 7.75. SUMO-3(2-91)C47S-MES was cleaved from its intein-CBD fusion by incubation with 1.5 CV of

PBS, pH 7.75 containing 100 mM MESNa for 72 h at 4 °C. The eluted α-thioester was purified by C18 preparative RP-HPLC employing a gradient of 30-60% B over 60 min. Fractions containing the desired thioester were identified by ESI-MS. We observed that the N-terminal Met of SUMO-3 is consistently processed *in vivo*, leading to the SUMO-3(2-91)-α-thioester product. Typical yield is 4-5 mg/L of cell culture. ESI-MS for SUMO-3(2-91)C47S-MES. Calculated *m/z* $[M+H]^+$ 10,444.7 Da, observed 10,445.8 ± 3.6 Da (**Figure 3.31**).

3.4.22 Expressed protein ligation of H4(1-14)$^{(aux)}$-C(O)NHNH$_2$ and SUMO-3(2-91)C47S-MES

Purified H4(1-14)$^{(aux)}$-C(O)NHNH$_2$ (15.6 mg, 11.6 μmol) and SUMO-3(2-91)C47S-MES (20.1 mg, 1.9 μmol) were dissolved in 7.2 mL of a buffer consisting of 6 M Gn-HCl, 100 mM Na_2HPO_4, and 10 mM TCEP, pH 7.3. Ligation proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C18 preparative RP-HPLC employing a gradient of 25-50% B over 60 min to give 12.6 mg (56%). ESI-MS of H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$. Calculated *m/z* $[M+H]^+$ 11,666.1 Da, observed 11,666.9 ± 2.8 Da (**Figure 3.32**).

3.4.23 Overexpression and purification of TEV protease

E. coli BL21(DE3) cells containing the plasmid pRK793-His$_6$-TEV[15] were grown in 1 L LB supplemented with 100 μg/mL of Ampicillin at 37 ºC with shaking at 250 rpm until OD_{600} ~0.6. Overexpression was induced by the addition of 0.3 mM IPTG and cells were grown for an additional 6 h at 25 °C. The cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in 15 mL lysis buffer: 20 mM tris, 150 mM NaCl, pH 7.2. Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The lysate supernatant was passed through a 0.45 μm filter then applied to ~5 mL Ni-NTA column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 3 h at 4 °C. The column was then washed thoroughly with lysis buffer containing 200 mM imidazole. His$_6$-TEV was eluted with lysis buffer containing 500 mM imidazole, and dialyzed into 4 L of 20 mM tris, 150 mM NaCl, 1 mM DTT, pH 7.5 for 18 h at 4 °C.

3.4.24 Overexpression and purification of H4(15-102)A15C

E. coli BL21(DE3) cells containing pET15b-His_6-[TEV]-H4(15-102)A15C were grown in 3 L 2x YT medium supplemented with 100 µg/mL of Ampicillin at 37 °C with shaking at 250 rpm until OD_{600} reached ~0.6. Overexpression was induced by the addition of 0.3 mM IPTG and cells were grown for an additional 1.5 h at 37 °C. The cells were harvested by centrifugation at 7,000x*g* for 15 min. H4(15-102)A15C was purified using a previously established protocol.[15] Cells were resuspended in wash buffer (20 mM tris, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 7.5, 1% triton X-100) and lysed by sonication on ice. Inclusion bodies were pelleted by centrifugation at 20,000x*g* for 20 min and washed twice with wash buffer. Inclusion bodies were then dissolved in extraction buffer (6 M Gn-HCl, 20 mM tris, 1 mM 2-mercaptoethanol, pH 7.5) and applied to Ni-NTA resin. Column binding proceeded overnight at 4 °C, after which the resin was washed with 10 CV extraction buffer containing 25 mM imidazole. The protein was eluted with 3 x 1 CV extraction buffer containing 400 mM imidazole, then dialyzed into water containing 1 mM DTT. After dialysis, 10x cleavage buffer was added for final concentrations of 50 mM tris, 1 mM EDTA, 10 mM DTT, 10 mM L-cysteine, pH 6.9. Purified TEV protease was added to ~20% of the final volume, and the cleavage reaction proceeded overnight at 37 °C. The reaction was then dialyzed back into extraction buffer, incubated overnight at 4 °C with Ni-NTA resin to remove the his-tagged TEV protease and cleaved H4 N-terminal His_6 tag, and the column flow-through containing H4(15-102)A15C purified by C4 preparative RP-HPLC employing a gradient of 40-70% B over 60 min. Typical yields were 3-4 mg/L of cell culture. ESI-MS of H4(15-102)A15C. Calculated *m/z* $[M+H]^+$ 10,071.8 Da, observed 10,075.3 ± 4.8 Da (**Figure 3.33**).

3.4.25 Expressed protein ligation of H4(1-14)$^{Su(C47S)(aux)}$-C(O)$NHNH_2$ and H4(15-102)A15C

Ligation was accomplished by first converting the C-terminal hydrazide of H4(1-14)$^{Su(C47S)(aux)}$-C(O)$NHNH_2$ to an acyl azide with $NaNO_2$ via the diazotization reaction, as described previously.[47] Subsequent addition of MPAA served to both quench the remaining $NaNO_2$ and generate a highly reactive C-terminal thioester for the ligation reaction. Purified H4(1-14)$^{Su(C47S)(aux)}$-C(O)$NHNH_2$ (1.7 mg, 0.15 µmol) was dissolved at 1 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl, pH 3, and kept at -20 °C for a minimum of 20 min. To this solution

was added 4.5 µL of a 500 mM solution of $NaNO_2$ in water. The reaction was briefly mixed, then kept at -20 °C for 15 min. Then, a solution of H4(15-102)A15C (3 mg, 0.3 µmol) dissolved at 1.25 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl, 200 mM MPAA, pH 6.5, was added to the reaction. The mixture was allowed to come to room temperature, and the pH adjusted with 3 M NaOH to 6.8-7.0 in order to form the MPAA α-thioester. The ligation reaction proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C4 semi-preparative RP-HPLC employing a gradient of 30-70% B over 45 min to give 2.1 mg (66%). ESI-MS of H4(A15C)$^{Su(C47S)(aux)}$. Calculated *m/z* $[M+H]^+$ 21,704.9 Da, observed 21,711.0 ± 5.9 Da (**Figure 3.34**).

3.4.26 MPAA-mediated auxiliary removal from H4(A15C)$^{Su(C47S)(aux)}$

Purified H4(A15C)$^{Su(C47S)(aux)}$ (1.4 mg, 64 nmol) was dissolved in 7 mL of 200 mM MPAA, 100 mM Na_2HPO_4, pH 7.3, and incubated in a 15 mL conical tube at 25 °C for 24 h. After incubation, the reaction volume was reduced with a 10,000 MWCO spin concentrator (GE Healthcare Life Sciences, Pittsburgh, PA) and buffered TCEP, pH 7.3 was added to a concentration of 50 mM and the reaction incubated at 4 °C for 30 min. Ligation product was purified by C4 analytical RP-HPLC employing a gradient of 30-70% B over 30 min. ESI-MS of H4(A15C)$^{Su(C47S)}$. Calculated *m/z* $[M+H]^+$ 21,628.7 Da, observed 21,634.0 ± 6.1 Da (**Figure 3.35a**).

3.4.27 Desulfurization of H4(A15C)$^{Su(C47S)}$

Purified H4(A15C)$^{Su(C47S)}$ (1.4 mg, 65 nmol) was dissolved at 65 µM in 100 mM Na_2HPO_4, 6 M Gn-HCl, 500 mM TCEP, 100 mM MESNa, pH 7.5. To this solution was added 2-methyl-2-propanethiol to a concentration of 280 mM and radical initiator 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) to a concentration of 10 mM. The reaction was incubated at 37 °C for 24 h, and the product purified by C4 analytical RP-HPLC employing a gradient of 30-70% B over 30 min to give 0.6 mg (41%). ESI-MS of suH4. Calculated *m/z* $[M+H]^+$ 21,596.7 Da, observed 21,601.0 ± 6.2 Da (**Figure 3.35b**).

3.4.28 Overexpression and purification of H2B(1-116)-MES

E. coli BL21(DE3) cells were transformed with the plasmid pTXB1-H2B(1-116)-GyrA-His_6.[63] Cells were grown in 4 L Luria-Bertani medium supplemented with 100 µg/mL of Ampicillin at 37 °C with shaking at 250 rpm until OD_{600} reached ~0.6. Overexpression was induced by the addition of 0.35 mM IPTG and cells were grown for an additional 4 h at 25 ºC. The cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in 40 mL of 1X PBS. Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The pellet was washed with 40 mL of 50 mM tris, 200 mM NaCl, 1% triton X-100, pH 7.5. The pellet was then dissolved in 40 mL extraction buffer: 6 M urea, 10 mM tris, 200 mM NaCl, pH 7.5. After centrifugation at 20,000x*g* for 15 min the supernatant was applied to a 5 mL Ni-NTA column pre-equilibrated with extraction buffer. Proteins were bound to the column over a period of 1 h at 4 °C. The column was then washed with extraction buffer containing increasing concentrations of imidazole: 20 mM (10 CV), 50 mM (10 CV), 500 mM (5 CV). Pooled fraction containing H2B(1-116)-GyrA-His_6 were dialyzed into 2 L thiolysis buffer (10 mM tris, 200 mM NaCl, 2 M urea, 1 mM MESNa, 1 mM EDTA, pH 7.5) for 1 h at 4 °C twice. Thiolysis buffer containing 1 M MESNa was added to the dialyzed fraction pool for a final concentration of 100 mM MESNa. The thiolysis reaction was incubated at 4 °C for 24 h, after which H2B(1-116)-GyrA-His_6 was no longer present. The reaction containing H2B(1-116)-MES was dialyzed into 2 L of 4 M Gn-HCl, 50 mM Na_2HPO_4, pH 7.5, for 1 h at 4 °C twice, then applied to 5 mL Ni-NTA resin pre-equilibrated in the same buffer. The byproduct GyrA-His_6 was bound to the column over a period of 30 min at 4 °C. The product H2B(1-116)-MES was eluted and purified by C18 preparative RP-HPLC employing a gradient of 35-65% B over 60 min. Fractions containing H2B(1-116)-MES were identified by ESI-MS. Typical yields were 1-2 mg/L of cell culture. ESI-MS of H2B(1-116)-MES. Calculated *m/z* $[M+H]^+$ 12,933.8 Da, observed 12,936.0 ± 3.3 Da (**Figure 3.36**).

3.4.29 Expressed protein ligation of H2B(117-125, A117C)^{photoaux} (4) and H2B(1-116)-MES α-thioester

H2B(1-116)-MES α-thioester (0.31 μmol, 4 mg) and peptide **4** (6.6 μmol, 8.5 mg) were dissolved in 500 μL of reaction buffer containing 6 M Gn-HCl, 100 mM Na_2HPO_4, 10 mM EDTA, 5 mM TCEP, pH 7.5. The reaction proceeded at 25 °C for 6 h. After incubation the sample was treated with 50 mM TCEP at 4 °C for 30 min. Ligation product was purified by C4 analytical RP-HPLC employing a gradient of 30-65% B over 30 min, and yielded 2.6 mg (60%). ESI-MS of H2B(A117C)^{photoaux} (**5**). Calculated *m/z* $[M+H]^+$ 14,059.2 Da, observed 14,062.7 ± 2.8 Da.

3.4.30 Refolding and photo-deprotection of H2B(A117C)^{photoaux} (5)

5 was dissolved at 0.4 mg/mL in 6 M Gn-HCl, 50 mM Na_2HPO_4, pH 7.5, and 1.5 mL was dialyzed into 1 L of 50 mM Na_2HPO_4, pH 7.5, at 4 °C, protected from light, for incubations of 3, 12, and 3 h, with fresh buffer for each incubation. The presence of secondary structure due to folding was confirmed by circular dichroism. To this solution, 3X photo-deprotection buffer (50 mM Na_2HPO_4, 12 mM semicarbazide, 15 mM ascorbic acid, 1.5 mM DTT, pH 6) was added to a final concentration of 1X. The sample was placed in a quartz cuvette with 1 cm path length and irradiated at 365 nm for 3.5 h with a hand-held 4 W Hg lamp (UVP, Upland, CA; measured irradiance = 1.8 mW cm^{-2}) 2 cm from light source. Irradiation effected complete removal of the *o*-nitrobenzyl group and unmasking of the auxiliary thiol. Deprotection was confirmed by LC-ESI-MS analysis employing a gradient of 0-100% D over 15 min. ESI-MS of H2B(A117C)^{aux} (**6**). Calculated *m/z* $[M+H]^+$ 13,924.1 Da, observed 13,925.8 ± 2.6 Da.

3.4.31 Expressed protein ligation of H2B(A117C)^{aux} (6) and SUMO-3(2-91)C47S-MES-α-thioester

Photo-deprotected **6** was dialyzed into 1 L of 50 mM Na_2HPO_4, 0.5 mM TCEP, pH 7.5, at 4 °C for incubations of 3 and 12 h. Separately, SUMO-3(2-91)C47S-MES was dissolved at 1.45 mg/mL in 50 mM Na_2HPO_4, pH 7.5, 1% DMF (**Figure 3.37**). The solution containing **6** (~ 11.6 μM) was combined 3:1 with the SUMO-3(2-91)C47S-MES solution for final concentrations of 8.6 uM H2B(1-117)^{aux} and 26.2 uM SUMO-3(2-91)C47S-MES (3 eq). To this mixture was added 200 mM TCEP, 20 mM MPAA, 50 mM Na_2HPO_4, pH 7.5, to final concentrations of 2 mM TCEP and 0.2 mM MPAA. The reaction was incubated under Ar at 22

°C for 48 h to give H2B(A117C)$^{Su(C47S)aux}$ (**7**) in 30-60% yield. Ligation was visualized by SDS-PAGE and confirmed by LC-ESI-MS analysis employing a gradient of 0-100% D over 15 min. Yield was determined by SDS-PAGE band intensity relative to BSA standards of known concentration. ESI-MS of H2B(A117C)$^{Su(C47S)aux}$ (**7**). Calculated *m/z* $[M+H]^+$ 24,225.6 Da, observed 24,230.9 ± 4.0 Da.

3.4.32 MPAA-mediated auxiliary removal from H2B(A117C)$^{Su(C47S)aux}$ (7)

A solution of 600 mM MPAA, 50 mM Na_2HPO_4, pH 7.5 was passed through a 0.45 µm filter, then added directly to the ligation reaction containing the ligation product, **7**, for a final concentration of 150 mM MPAA. The reaction was placed in a container with headspace filled with air equal to 10 times the liquid volume and incubated at 22 °C for 24 h to give H2B(A117C)$^{Su(C47S)}$ (**8**) in 15-30% yield over 2 steps. Final product was visualized by SDS-PAGE and confirmed by LC-ESI-MS analysis employing a gradient of 0-100% D over 15 min. Yield was determined by SDS-PAGE band intensity relative to BSA standards of known concentration. Folding was confirmed by SEC purification of the final product in 50 mM Na_2HPO_4, pH 7.5, at 4 °C, followed by circular dichroism (**Figure 3.38**). ESI-MS of H2B(A117C)$^{Su(C47S)}$ (**8**). Calculated *m/z* $[M+H]^+$ 24,149.6 Da, observed 24,153.1 ± 3.2 Da.

3.4.33 SENP1 hydrolysis assay

SEC-purified H2B(A117C)$^{Su(C47S)}$ (**8**) was assayed with the catalytic domain of sentrin-specific protease 1 (SENP1, Boston Biochem). SENP1 (0.05 nmol) was pre-activated in 10 µL buffer containing 50 mM tris, 150 mM NaCl, 12 mM DTT, pH 8 for 20 min at 25 °C. To the reduced SENP1 was then added 10 µL of a solution containing 0.5 nmol of **8** in 50 mM tris, 150 mM NaCl, 1 mM DTT, pH 7.5. The resulting mixture was incubated for 24 h at 37 °C. The assay was quenched by the addition of 6X Laemmli buffer containing 300 mM DTT and boiled for 5 min, then run on an 18% SDS-PAGE gel at 200 V for 1.5 h and stained with Coomassie (**Figure 3.39**).

3.5 Product characterization and supplemental data

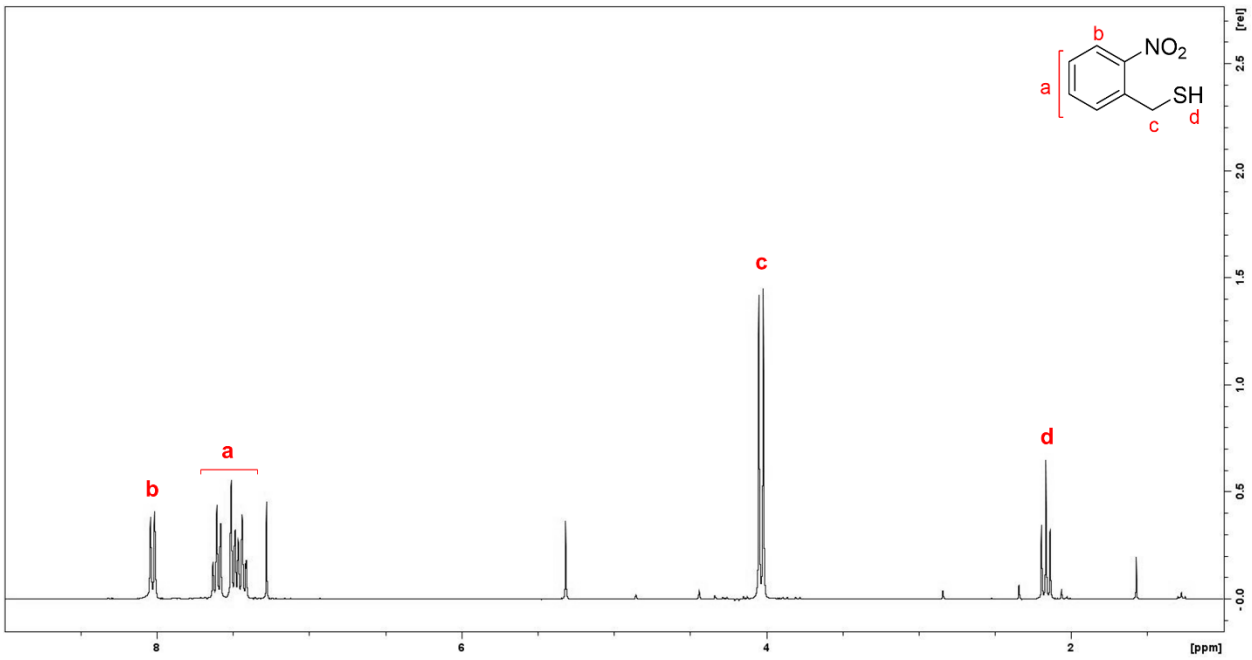


Figure 3.11. ^{1}H-NMR of (2-nitrobenzyl)thiol.

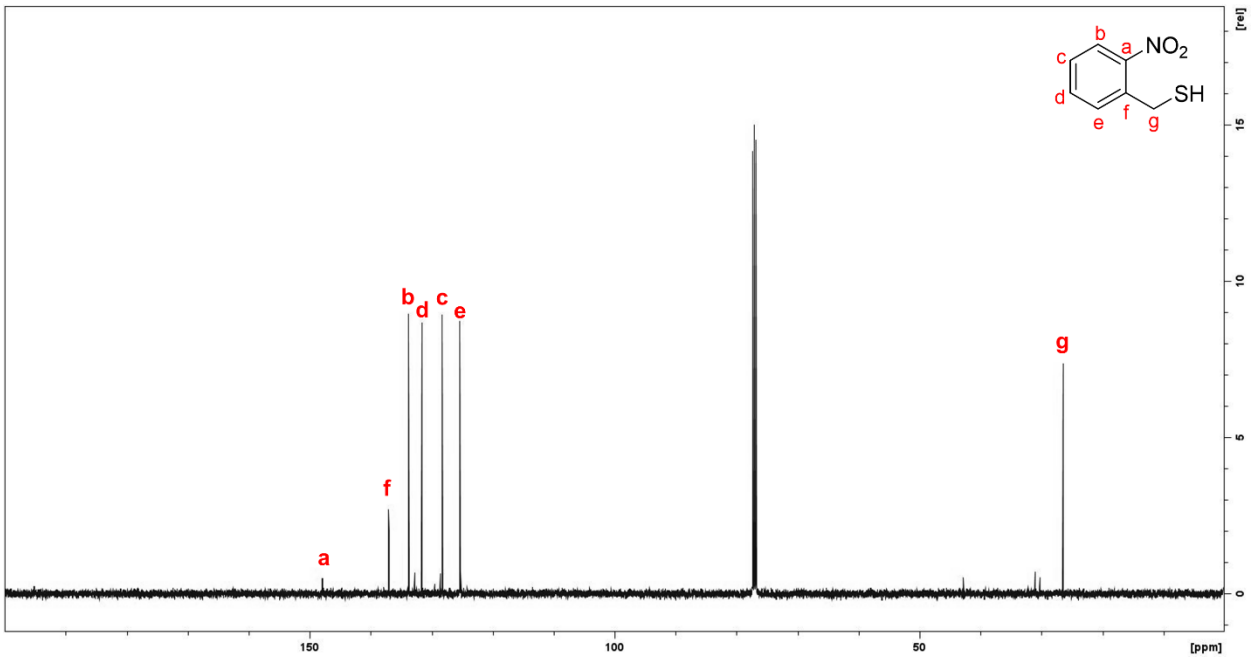


Figure 3.12. ^{13}C-NMR of (2-nitrobenzyl)thiol.

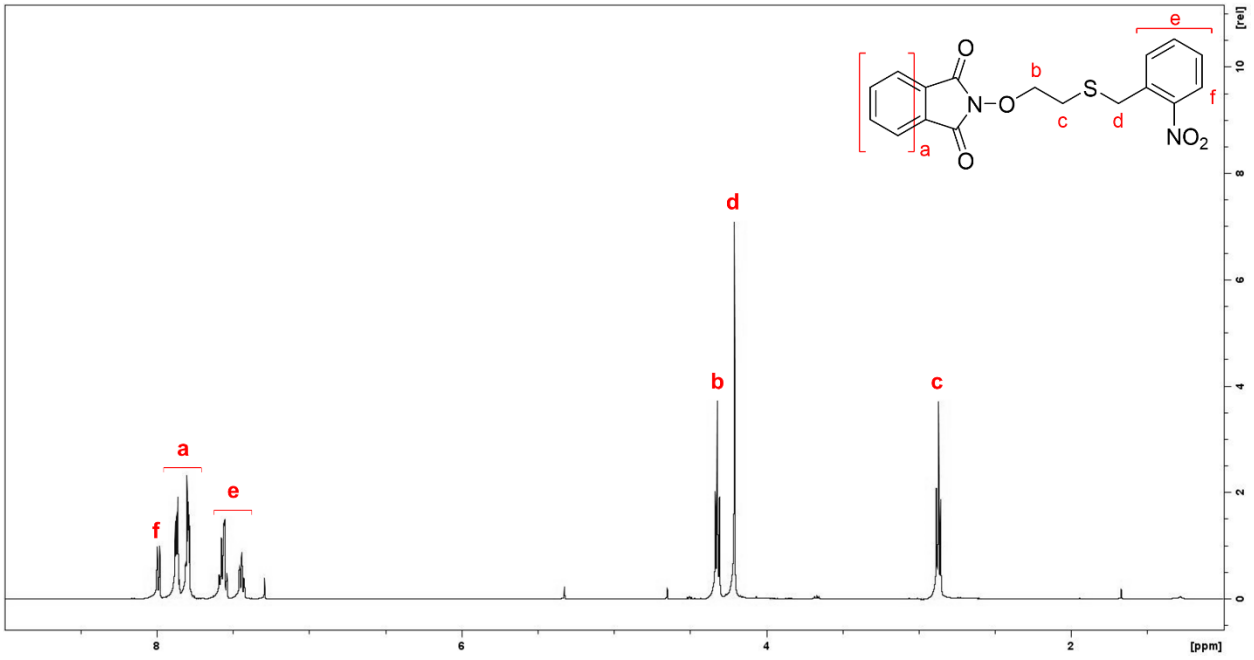


Figure 3.13. ^{1}H-NMR of *N*-(2-((2-nitrobenzyl)thio)ethoxy)phthalimide.

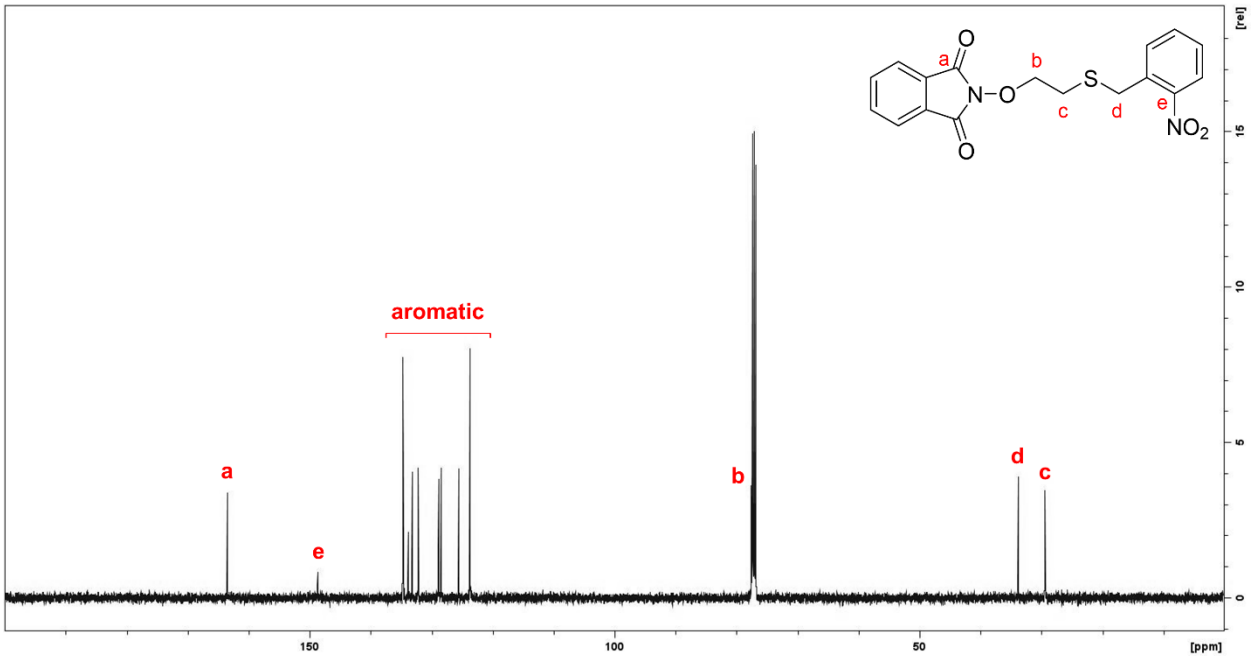


Figure 3.14. ^{13}C-NMR of *N*-(2-((2-nitrobenzyl)thio)ethoxy)phthalimide.

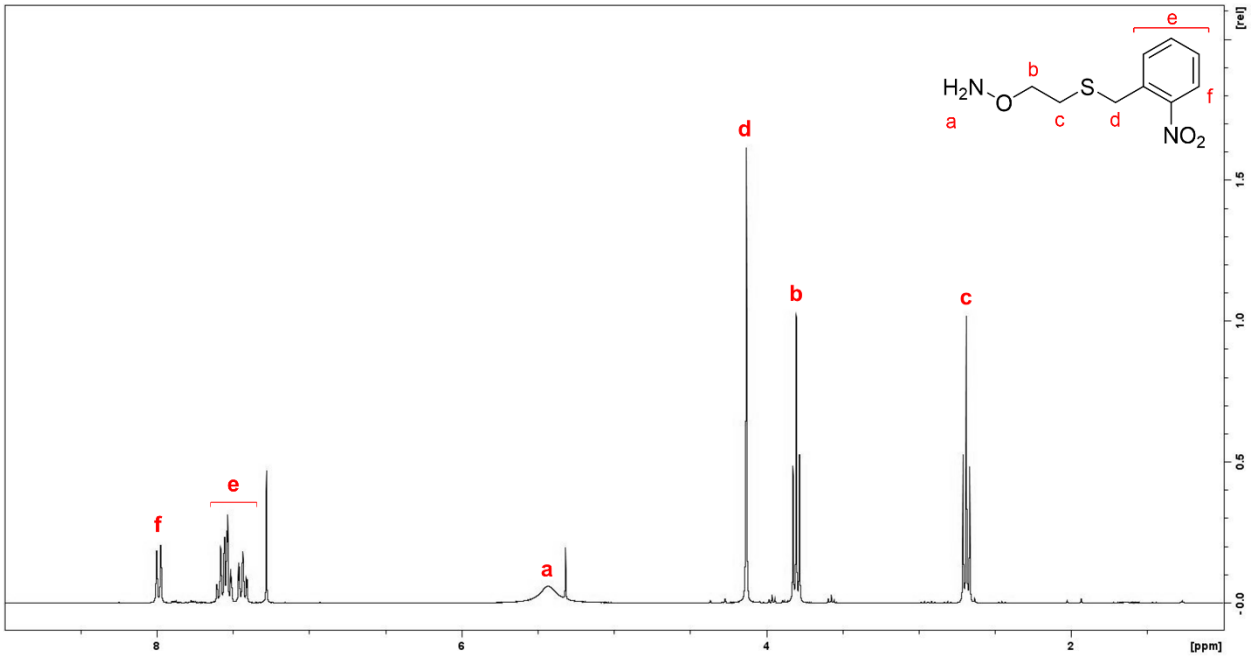


Figure 3.15. ^{1}H-NMR of *O*-(2-((2-nitrobenzyl)thio)ethyl)hydroxylamine (3).

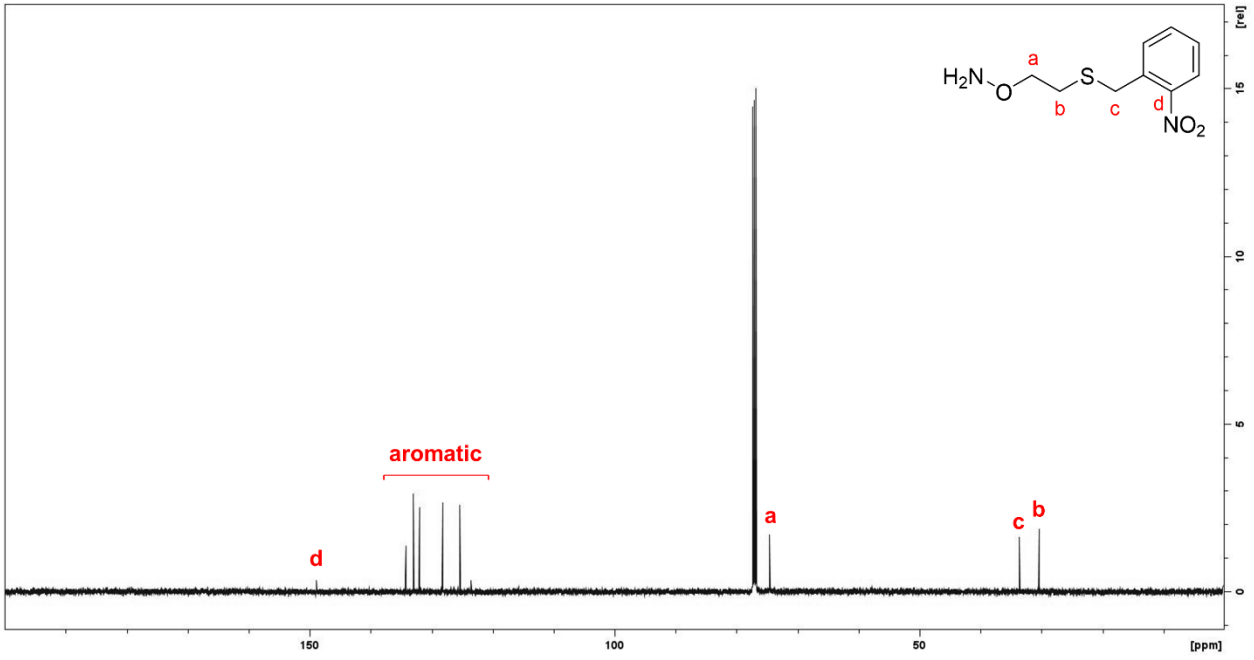


Figure 3.16. ^{13}C-NMR of *O*-(2-((2-nitrobenzyl)thio)ethyl)hydroxylamine (3).

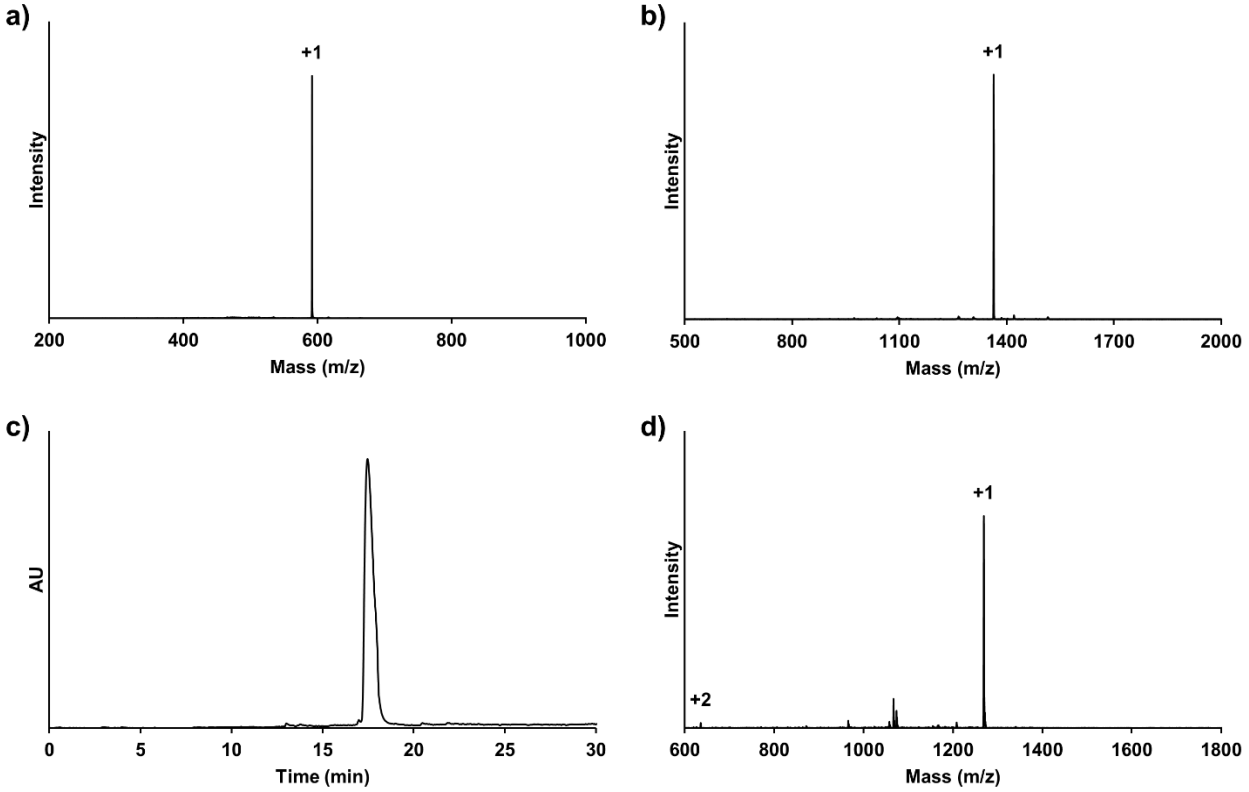


Figure 3.17. Purification of auxiliary-containing peptides. a) ESI-MS of purified of KAK^{aux}I. b) ESI-MS of purified H4(1-14)^{aux}-C(O)NHNH$_2$. c) C18 analytical RP-HPLC chromatogram of purified H2B(117-125, A117C)^{photoaux} (**4**), gradient of 0-50% B, 30 min. d) ESI-MS of purified H2B(117-125, A117C)^{photoaux} (**4**).

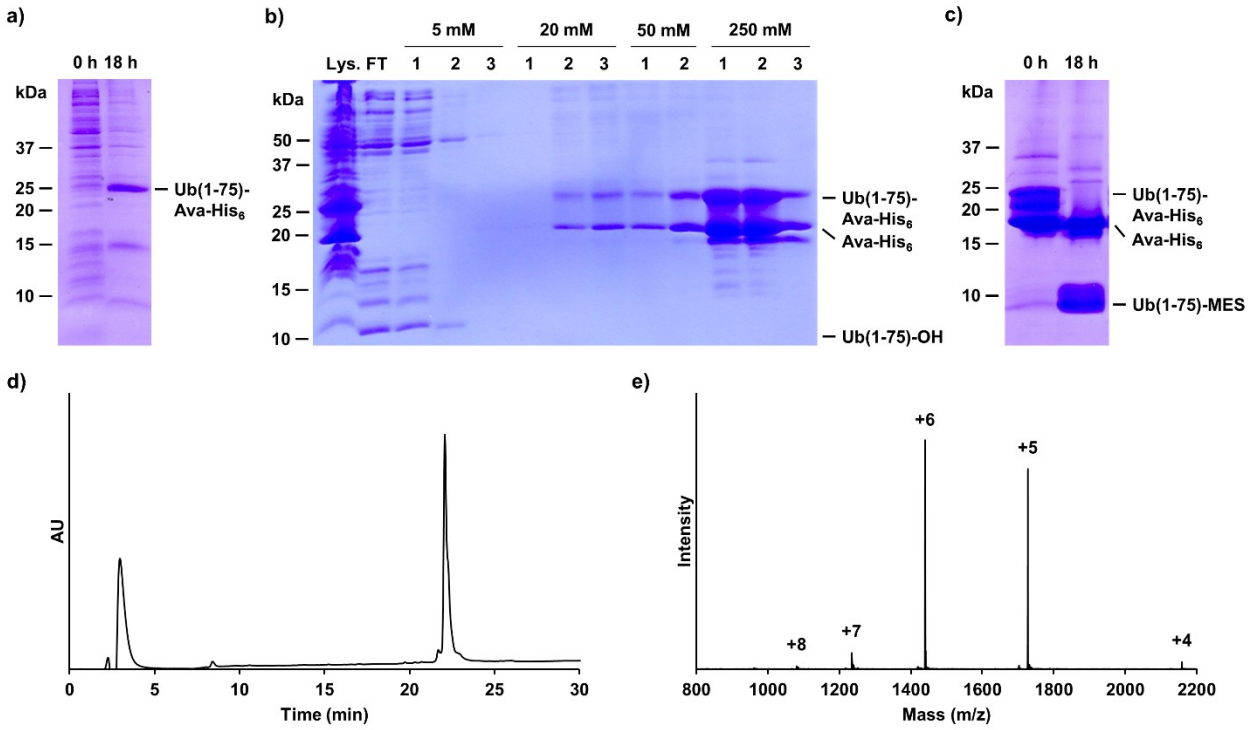


Figure 3.18. Purification of Ub(1-75)-MES. Coomassie stained 15% SDS-PAGE gel of a) IPTG-induced Ub(1-75)-Ava-His_6 overexpression; b) Ub(1-75)-Ava-His_6 purification by Ni-NTA column; and c) Thiolysis of pooled fractions containing Ub(1-75)-Ava-His_6. d) C18 analytical RP-HPLC chromatogram of purified Ub(1-75)-MES α-thioester, gradient of 0-73% B, 30 min. b) ESI-MS of purified Ub(1-75)-MES α-thioester. Lys = cell lysate. FT = column flow-through after protein binding.

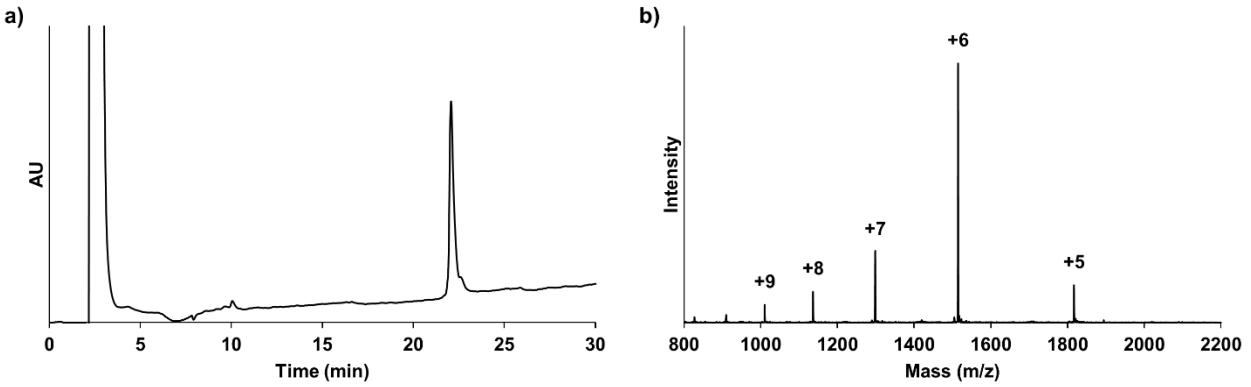


Figure 3.19. Purification of $KAK^{Ub(aux)}I$. a) C18 analytical RP-HPLC chromatogram of purified $KAK^{Ub(aux)}I$, gradient of 0-73% B, 30 min. b) ESI-MS of purified $KAK^{Ub(aux)}I$.

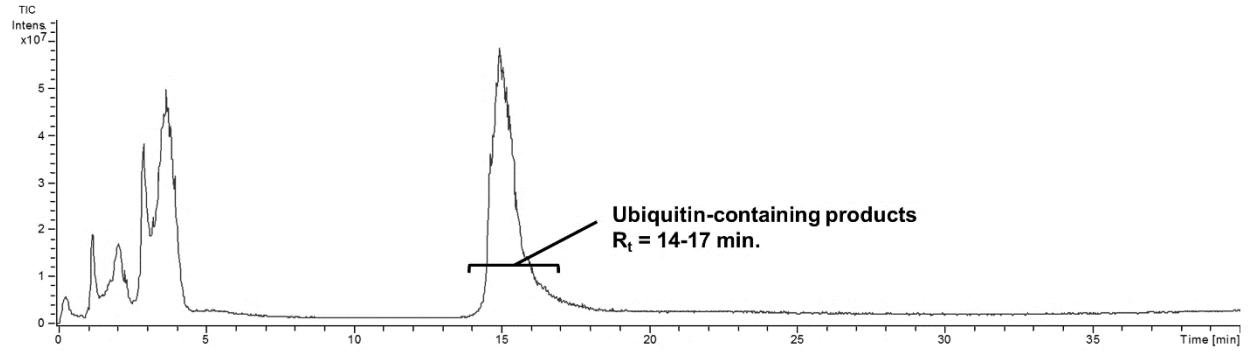


Figure 3.20. Sample LC-ESI-MS trace of an *N-O* bond cleavage test. Typical total ion chromatogram (TIC) of an *N-O* bond cleavage test on KAK$^{Ub(aux)}$I substrate. The sample is injected on a gradient of 0-100% D over 40 min. Ubiquitin-containing species co-elute between 14 and 17 min. ESI-MS signal is averaged over 14-17 min, and this spectrum used to quantify yield.

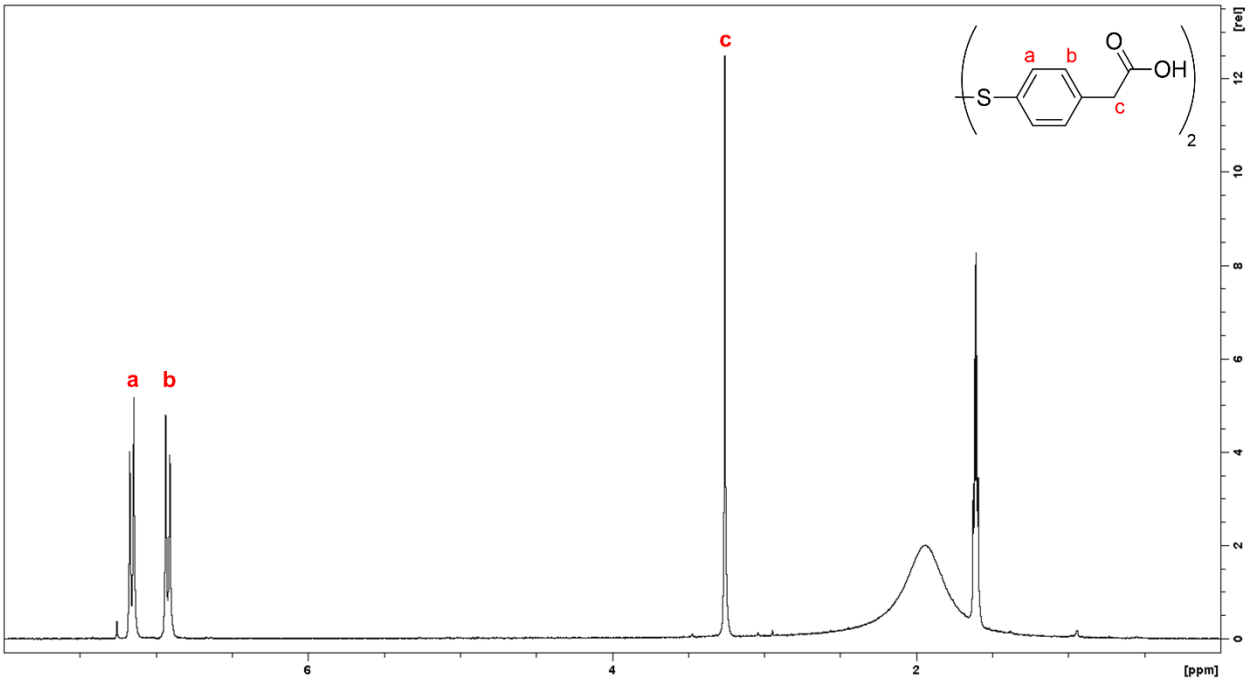


Figure 3.21. ^{1}H-NMR of the disulfide-linked MPAA dimer 2,2'-(disulfanediylbis(4,1-phenylene))diacetic acid. Peak at 1.95 ppm (s) identified as H_2O. CD_3CN solvent peak observed at 1.61 ppm (m).

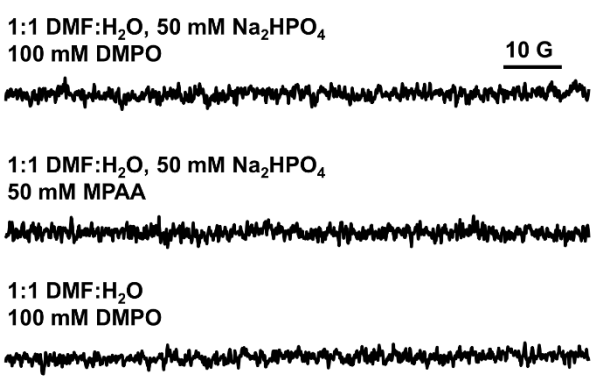


Figure 3.22. Control EPR spectra for DMPO/$^{\bullet}$*S*-Ar adduct detection. a) Spectrum obtained upon incubating 100 mM DMPO in 50 mM Na_2HPO_4 at pH 7.5, in a 1:1 water-DMF mixture at 25 °C. b) Spectrum obtained upon incubating 50 mM MPAA in 50 mM Na_2HPO_4 at pH 7.5, in a 1:1 water-DMF mixture at 25 °C. c) Spectrum obtained upon incubating 100 mM DMPO in a 1:1 water-DMF mixture at 25 °C. Spectrometer settings: microwave power, 20 mW; modulation amplitude, 1.0 G; time constant, 163 ms; scan rate, 0.6 G/s.

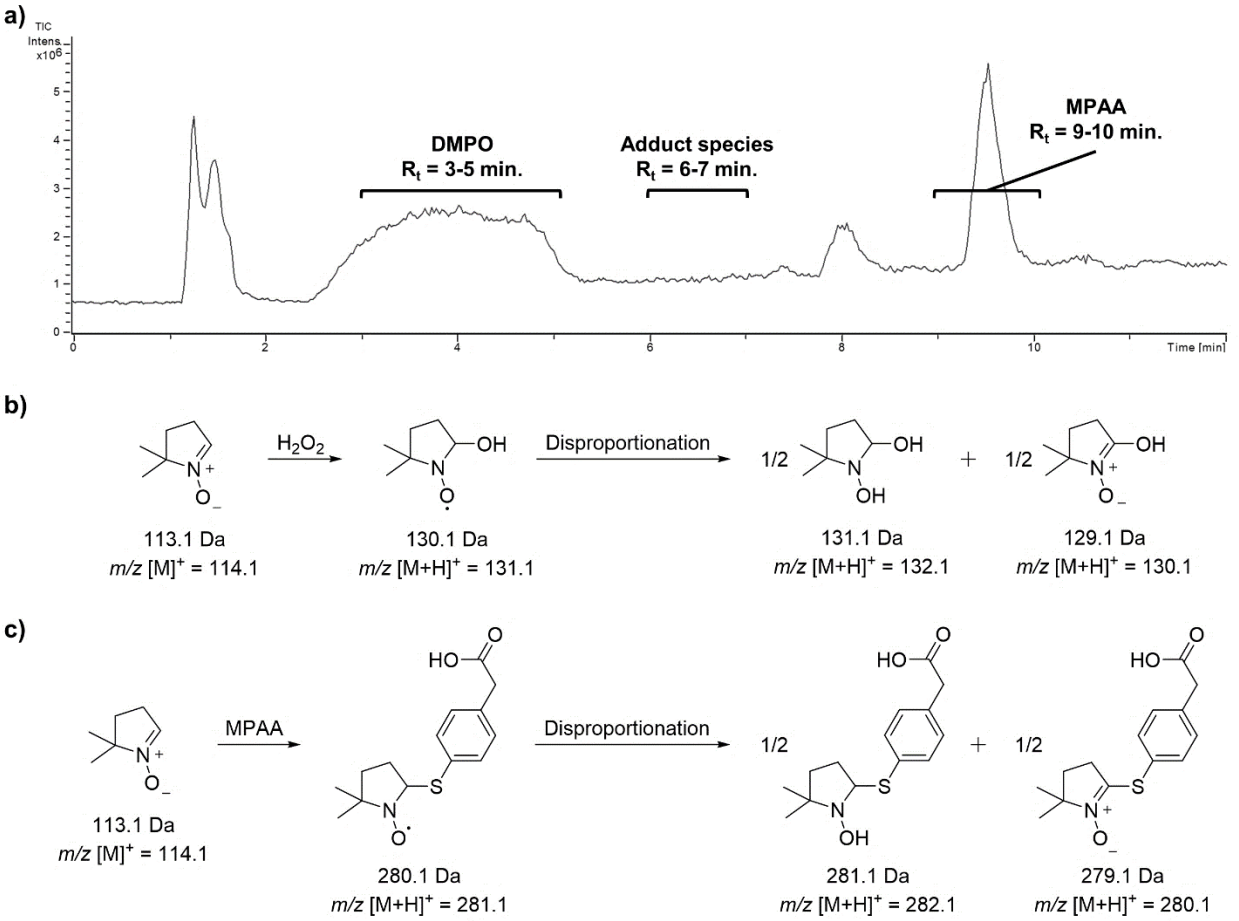


Figure 3.23. Detection of DMPO adducts by LC-ESI-MS. a) Typical total ion chromatogram of a DMPO reaction containing MPAA. The sample is injected on a gradient of 30-100% D over 10 min. R_t = 3-5 min for DMPO; R_t = 6-7 min for the DMPO-OH and DMPO-MPAA adducts; R_t = 9-10 min for MPAA. b) Expected products of DMPO reaction with hydroxyl radical. c) Expected products of DMPO reaction with MPAA thiyl radical.

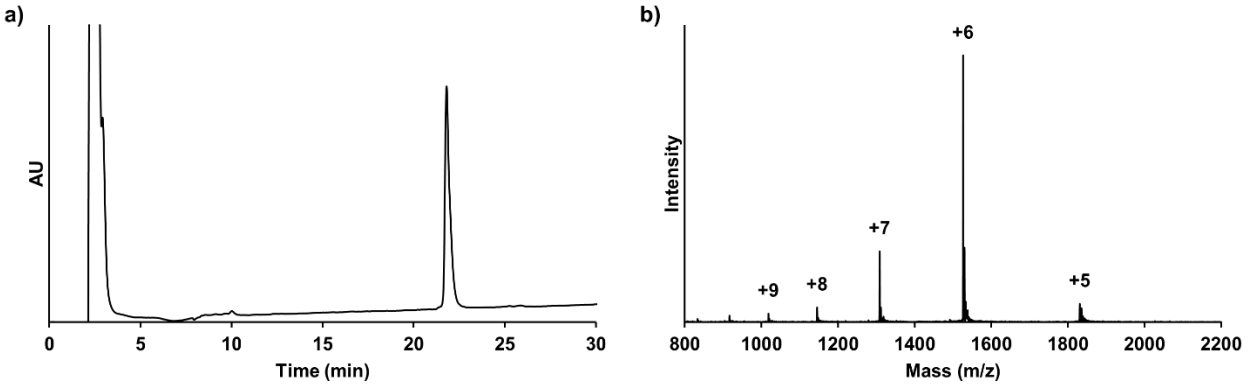


Figure 3.24. Purification of *S*-alkylated KAK$^{Ub(aux)}$I. a) C18 analytical RP-HPLC chromatogram of purified *S*-alkylated KAK$^{Ub(aux)}$I, gradient of 0-73% B, 30 min. b) ESI-MS of purified *S*-alkylated KAK$^{Ub(aux)}$I.

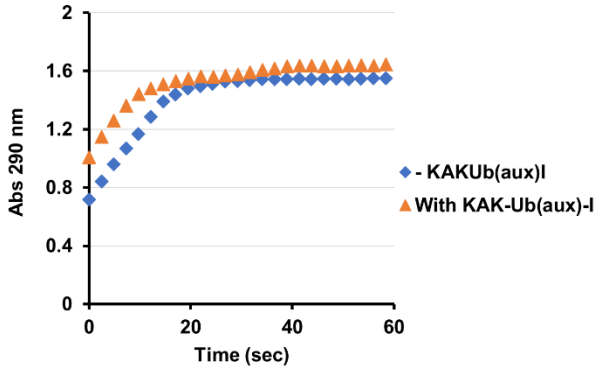


Figure 3.25. Formation of uric acid from the xanthine oxidase/hypoxanthine reaction in the presence or absence of auxiliary-containing substrate. Increase in absorbance at 290 nm over time in a reaction containing 0.16 mM hypoxanthine and 0.14 u/mL xanthine oxidase, with (red triangle) or without (blue diamond) 0.04 mM KAK$^{Ub(aux)}$I.

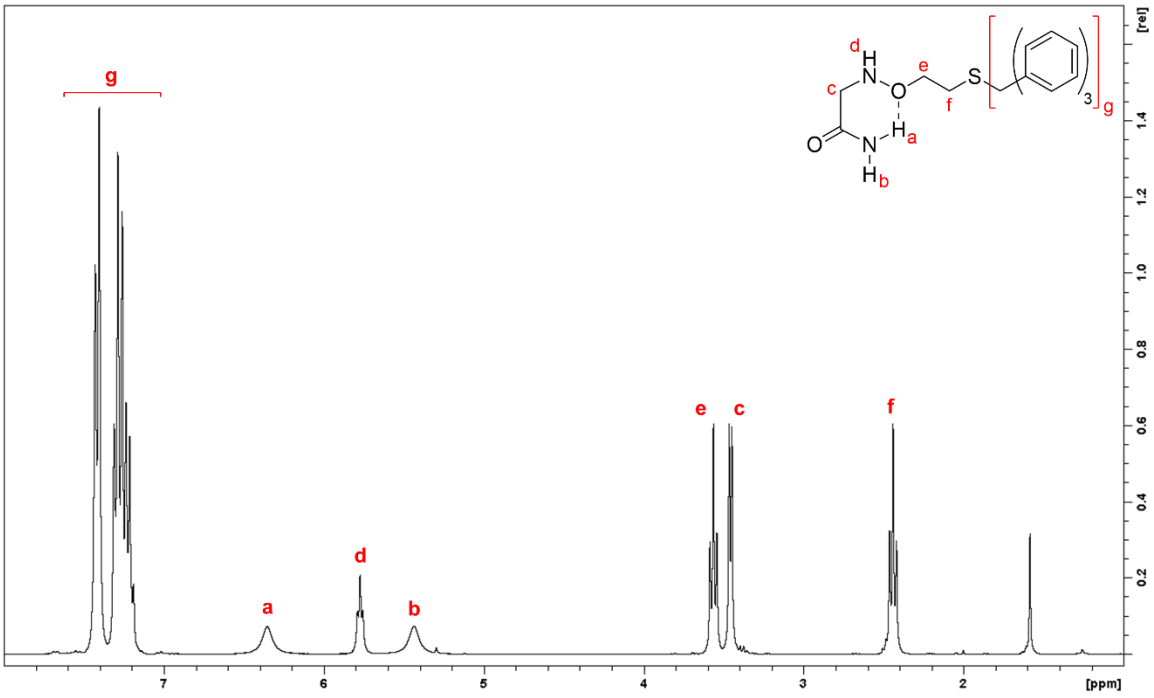


Figure 3.26. ^{1}H-NMR of 2-((2-(tritylthio)ethoxy)amino)acetamide. Peak at 1.58 ppm (s) identified as H_2O.

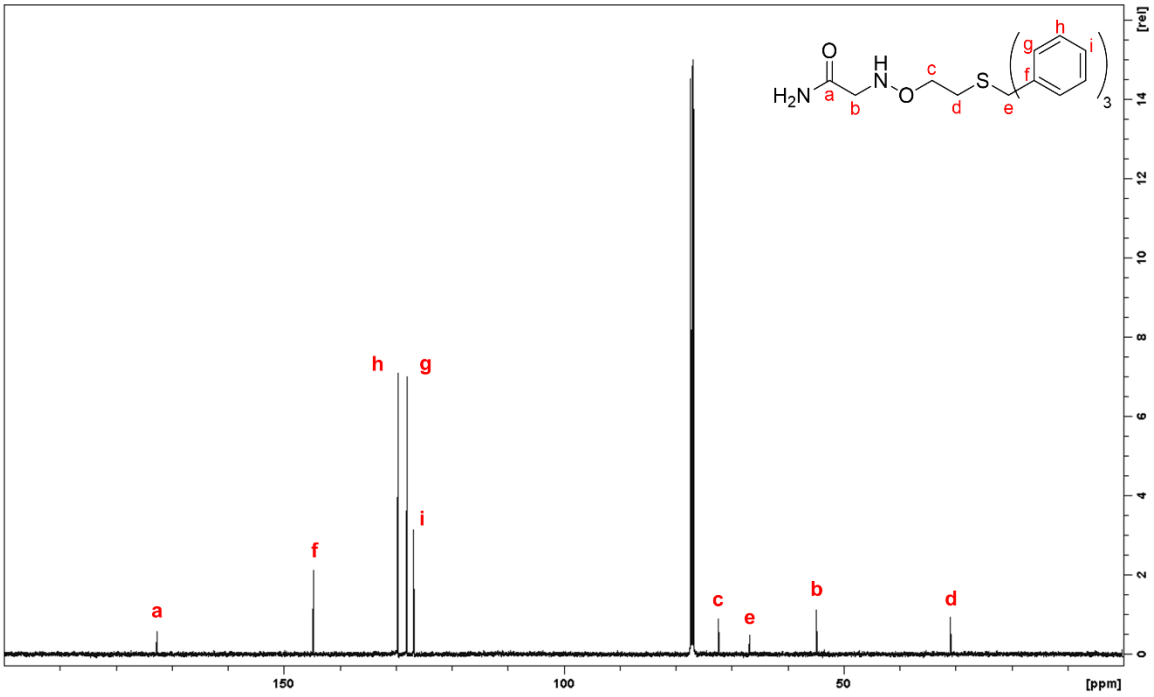


Figure 3.27. ^{13}C-NMR of 2-((2-(tritylthio)ethoxy)amino)acetamide. $CDCl_3$ solvent peak identified at 77.05 ppm (t).

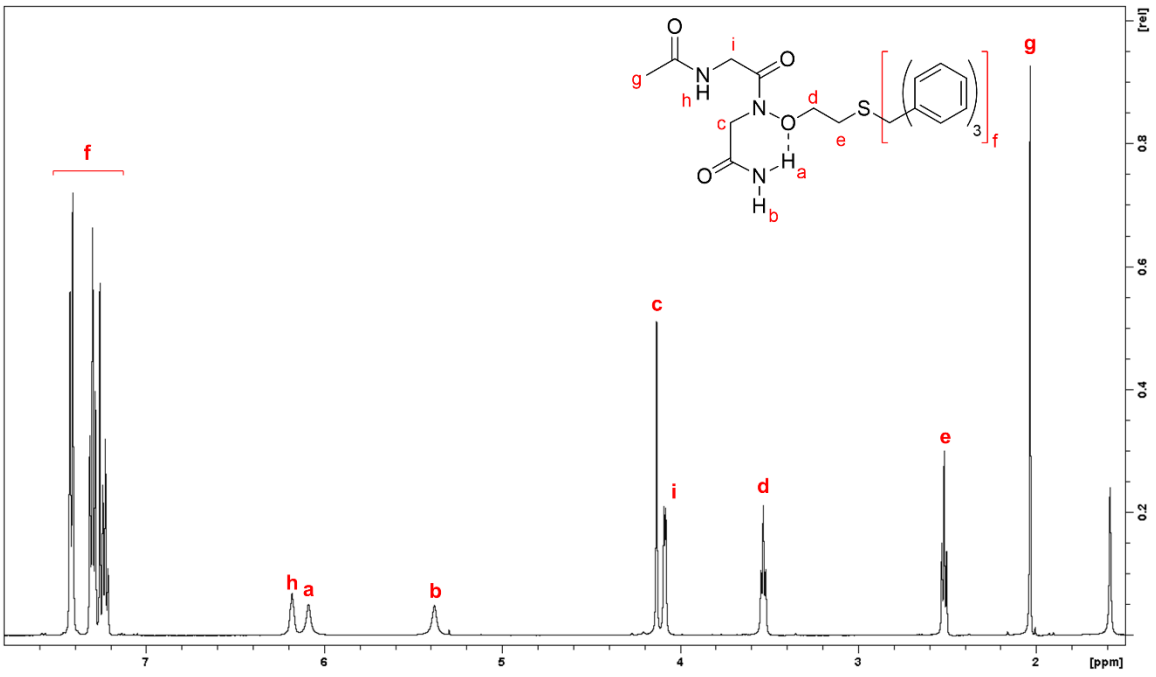


Figure 3.28. ^{1}H-NMR of model dipeptide 2, 2-acetamido-*N*-(2-amino-2-oxoethyl)-*N*-(2-(tritylthio)ethoxy) acetamide. Peak at 1.58 ppm (s) identified as H_2O.

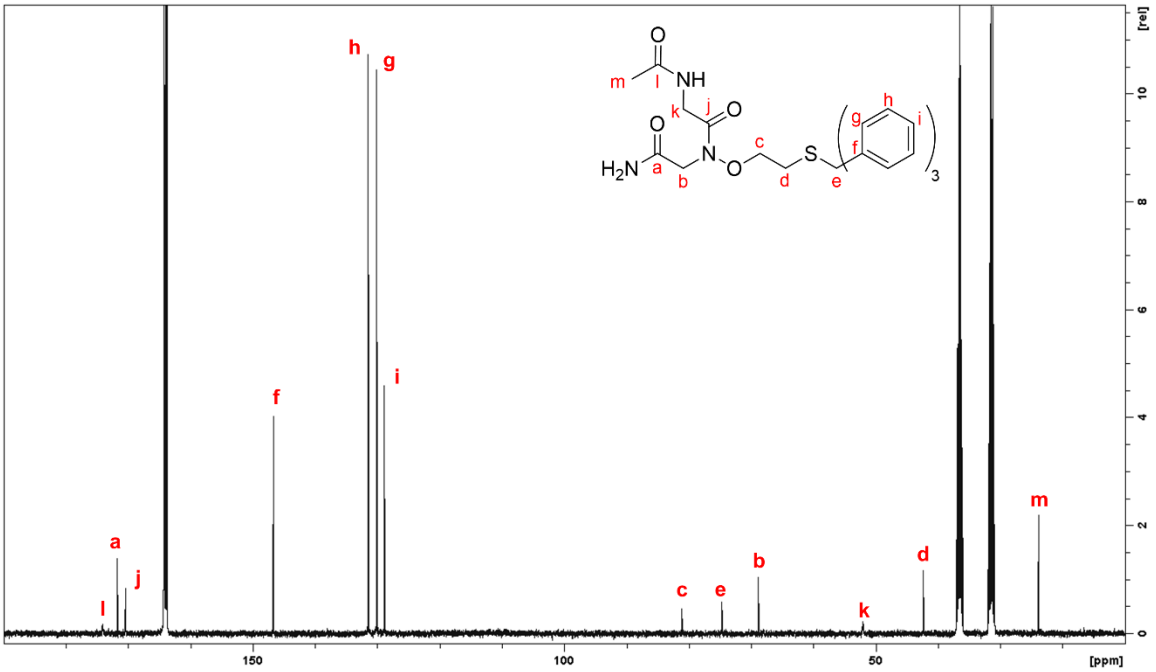


Figure 3.29. ^{13}C-NMR of model dipeptide 2, 2-acetamido-*N*-(2-amino-2-oxoethyl)-*N*-(2-(tritylthio)ethoxy) acetamide. DMF-D_7 solvent peaks observed at 164.04 (t), 36.55 (m), and 31.44 (m) ppm.

Table 3.5. Conditions for auxiliary removal from model dipeptide 2.

Reaction Conditions	10 min	24 h
Small molecule only	n.r.	n.r.
200 mM MPAA	n.r.	Reaction complete
50 mM H_2O_2	n.r.	n.r.
200 mM MPAA, 50 mM H_2O_2	Reaction complete	No change from 10 min.
200 mM MPAA, Chelex-treated[a]	n.r.	Minor progress
200 mM MPAA, 50 mM H_2O_2, Chelex-treated[a]	n.r.	~50% complete
	n.r.	
200 mM MPAA (pH 8), degassed[b]	n.r.	Minor progress
	n.r.	
200 mM 4-mercaptophenol	n.r.	n.r.
200 mM 4-mercaptophenol, 50 mM H_2O_2	n.r.	n.r.
200 mM 4-mercaptophenol, 200 mM MPAA	n.r.	n.r.

[a]Chelex-100 resin was swelled in water for 30 min. prior to use. DMF and buffer solutions were passed 2 times through the resin to remove divalent metal cations. [b]Buffer was subjected to three cycles of freeze-thaw degassing under Argon and the reaction kept under Argon atmosphere. n.r. = no reaction.

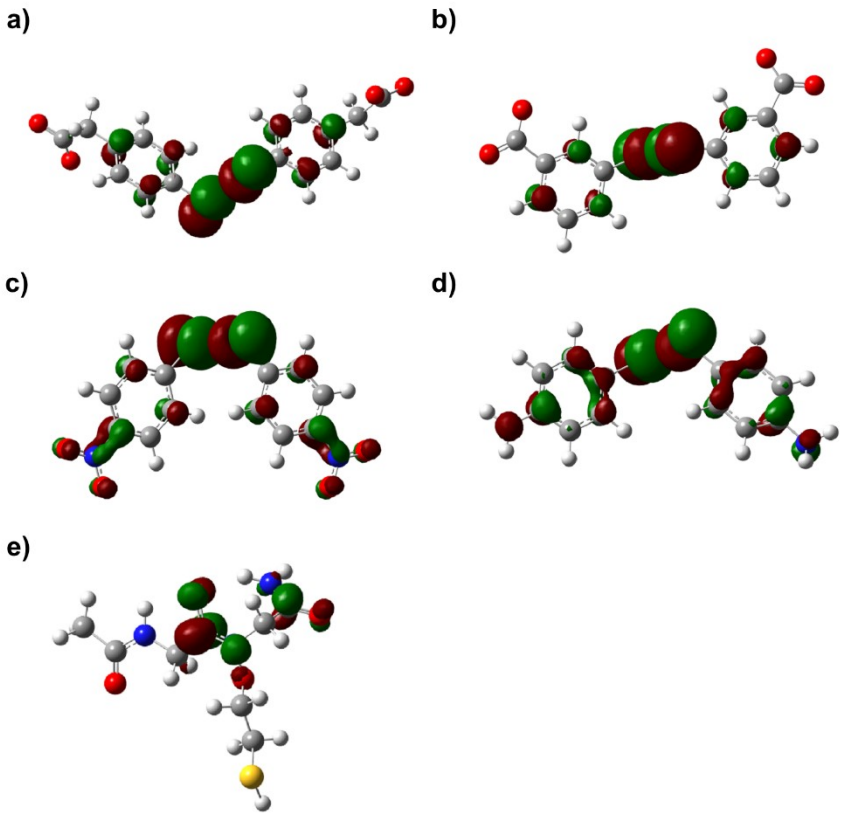


Figure 3.30. Calculated HOMOs of reducing agents and LUMO of the model peptide substrate. a) HOMO of MPAA disulfide radical anion. b) HOMO of 3-Mercaptobenzoic acid disulfide radical anion. c) HOMO of *p*-nitrothiophenol disulfide radical anion. d) HOMO of p-aminothiophenol radical anion. e) LUMO of model diglycine peptide 1.

Table 3.6. Relative stability of products following electron transfer from MPAA disulfide radical anion to model diglycine compound (2)

Pathway 1		Pathway 2		$\Delta G_{(1-2)}$ (eV)[a]
$GlyGlyN^-$ + xH_2O	$^{\bullet}OSH$ + xH_2O	$GlyGlyN^{\bullet}$ + xH_2O	^-OSH + xH_2O	
x = 0	x = 0	x = 0	x = 0	-0.49
0	1	1	0	-0.18
0	2	2	0	0.00
0	3	3	0	0.08
1	1	1	1	-0.25
1	2	2	1	-0.06
1	3	3	1	0.02
2	2	2	2	-0.09
2	3	3	2	-0.01
3	2	3	2	-0.24
3	2	5	0	-0.31
3	3	5	1	-0.14
4	3	5	2	-0.45
4	2	5	1	-0.67
4	1	5	0	-0.78
6	3	7	2	0.00
6	2	7	1	-0.22
7	2	7	2	-0.18
7	0	7	0	-0.35

[a]Relative energy at equilibrium geometry calculated for products of the electron transfer reaction from the MPAA disulfide radical anion to the model diglycine compound GlyGlyN-OCH_2CH_2SH (**2**) to yield a $GlyGlyN^-$ anion and $^{\bullet}OSH$ radical (Pathway 1), or $GlyGlyN^{\cdot}$ radical and ^-OSH anion (Pathway 2). Various numbers of explicit water molecules surrounding the products were included in the equilibrium geometry energy calculations where allowed. Values for $\Delta G_{(1-2)}$ reflect the energy at equilibrium geometry for the products of Pathway 1 relative to the products of Pathway 2.

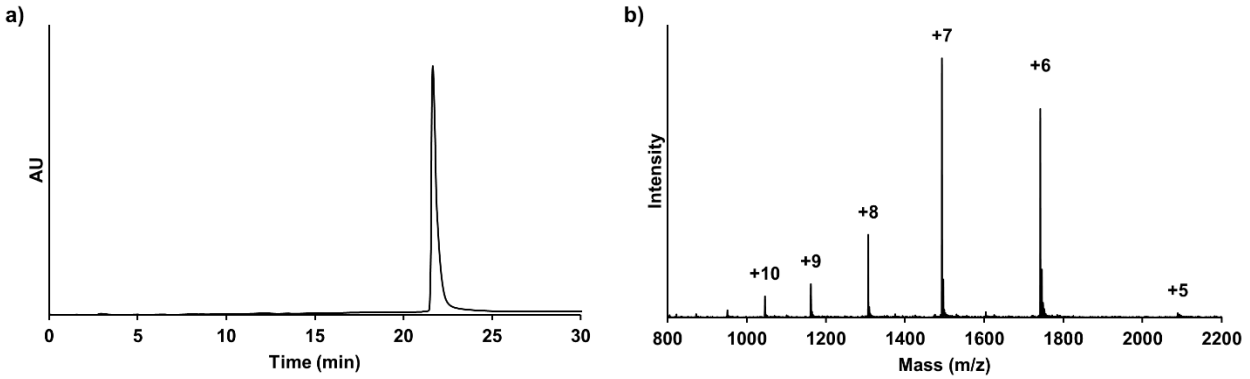


Figure 3.31. Purification of SUMO-3(2-91)C47S-MES. a) C18 analytical RP-HPLC chromatogram of purified SUMO-3(2-91)C47S-MES, gradient of 0-73% B, 30 min. b) ESI-MS of purified SUMO-3(2-91)C47S-MES.

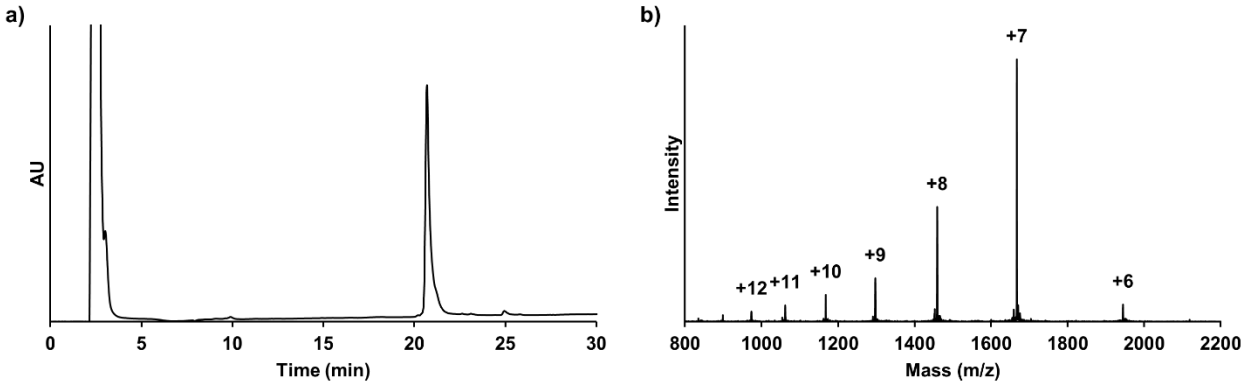


Figure 3.32. Purification of H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$. a) C18 analytical RP-HPLC chromatogram of purified H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$, gradient of 0-73% B, 30 min. b) ESI-MS of purified H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$.

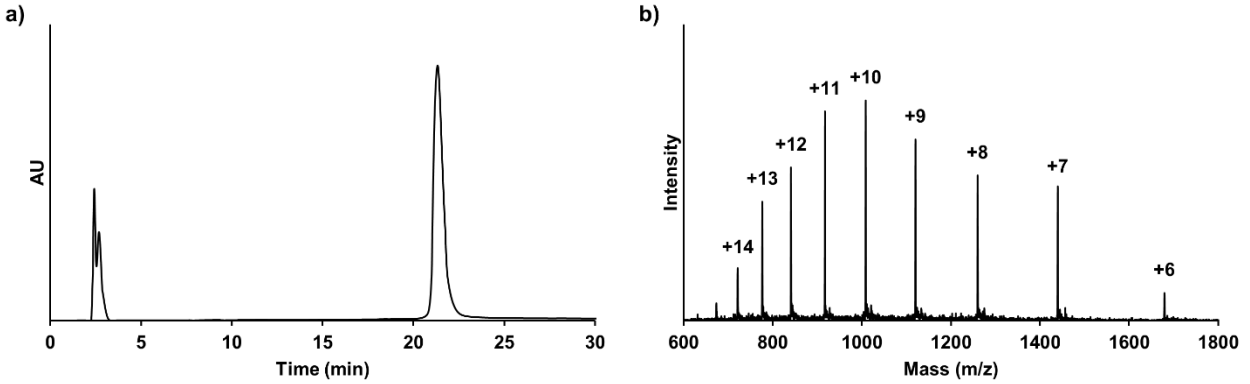


Figure 3.33. Purification of H4(15-102)A15C. a) C4 analytical RP-HPLC chromatogram of purified H4(15-102)A15C, gradient of 0-100% B, 30 min. b) ESI-MS of purified H4(15-102)A15C.

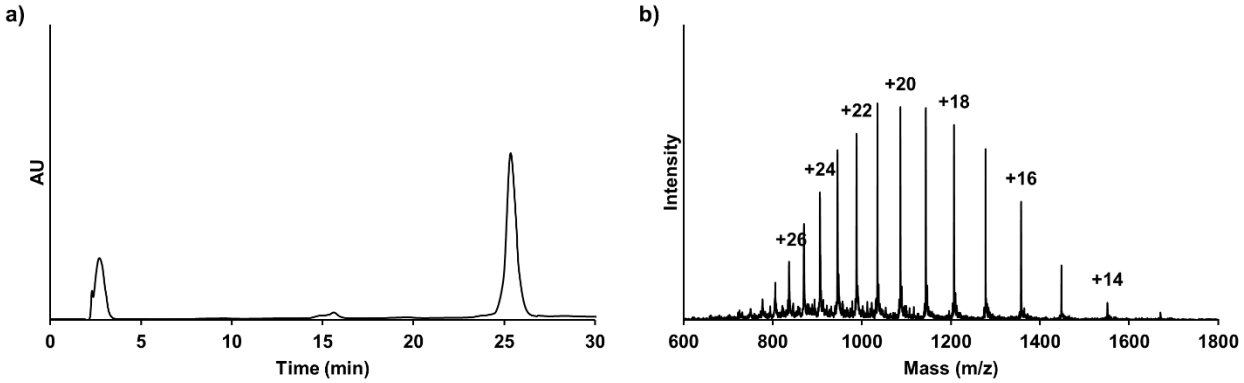


Figure 3.34. Purification of H4(A15C)$^{Su(C47S)(aux)}$. a) C4 analytical RP-HPLC chromatogram of purified H4(A15C)$^{Su(C47S)(aux)}$, gradient of 0-73% B, 30 min. b) ESI-MS of purified H4(A15C)$^{Su(C47S)(aux)}$.

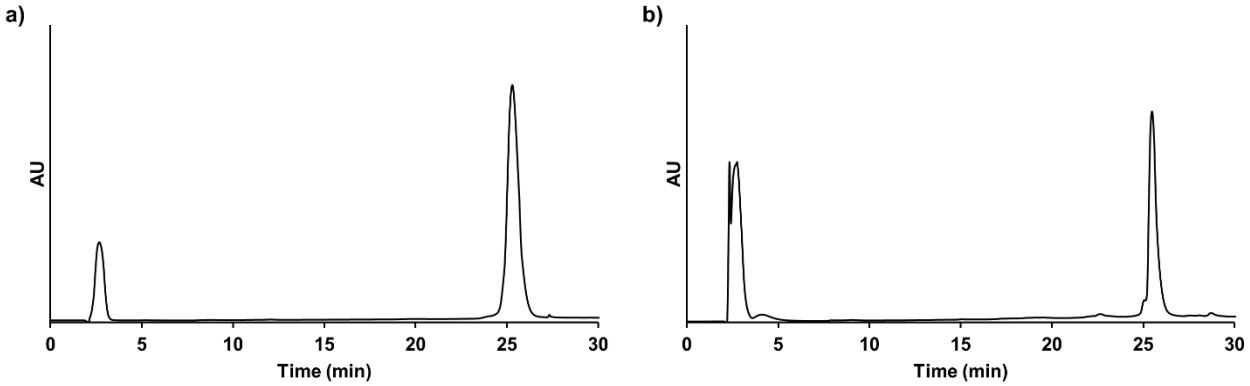


Figure 3.35. Purification of H4(A15C)$^{Su(C47S)}$ and suH4. a) C4 analytical RP-HPLC chromatogram of purified H4(A15C)$^{Su(C47S)}$, gradient of 0-73% B, 30 min. b) C4 analytical RP-HPLC chromatogram of purified suH4, gradient of 0-73% B, 30 min.

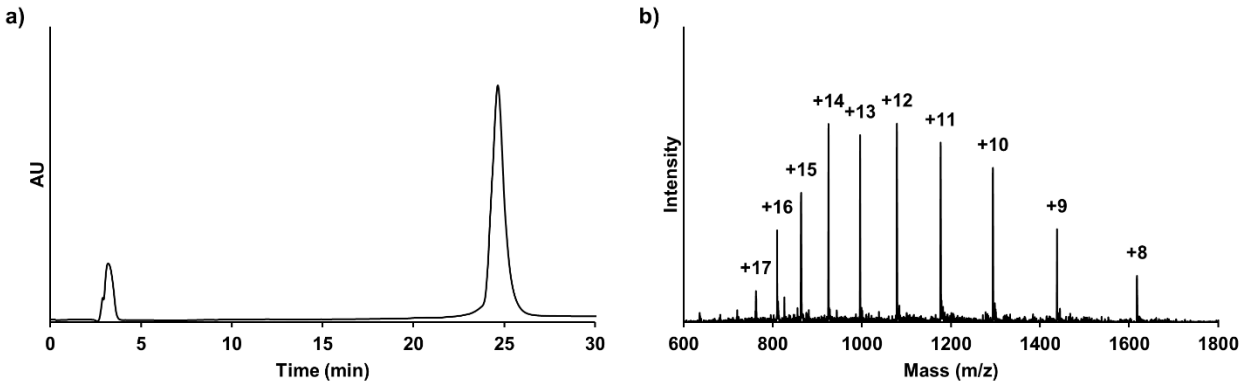


Figure 3.36. Purification of H2B(1-116)-MES. a) C4 analytical RP-HPLC chromatogram of purified of H2B(1-116)-MES, gradient of 0-73% B, 30 min. b) ESI-MS of purified of H2B(1-116)-MES.

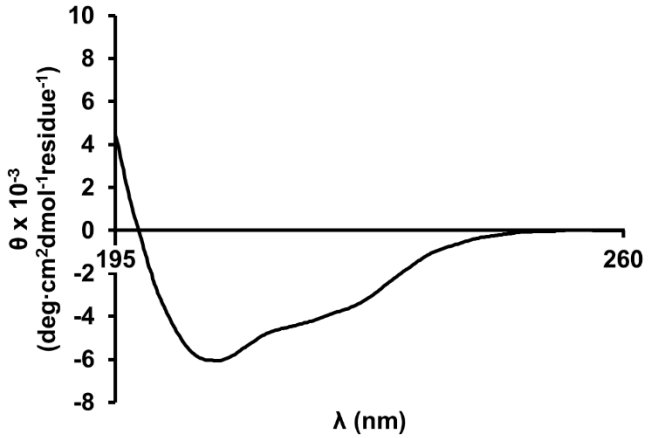


Figure 3.37. Refolded SUMO-3(2-91)C47S-MES. Circular Dichroism spectrum of SUMO-3(2-91)C47S-MES at 0.3 mg/mL in 50 mM Na_2HPO_4, pH 7.5.

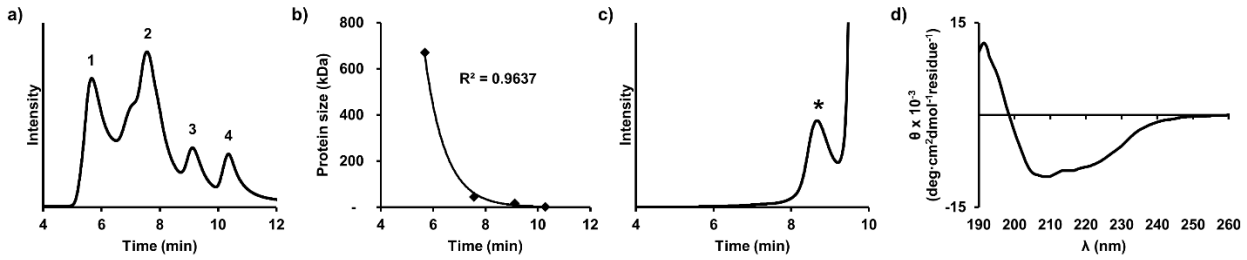


Figure 3.38. Size exclusion chromatography and circular dichroism of H2B(A117C)$^{Su(C47S)}$ (8). a) Size exclusion UV chromatogram of protein standards in 50 mM Na_2HPO_4, pH 7.5, at 4 °C, run at 1 mL/min. Protein standards: 1) 670,000 Da, R_t = 5.69 min; 2) 44,000 Da, R_t = 7.56 min; 3) 17,000 Da, R_t = 9.11 min; 4) 1,350 Da, R_t = 10.28 min. b) Protein standard data fit to an exponential equation. c) Size exclusion UV chromatogram of the crude MPAA-mediated *N-O* bond cleavage auxiliary removal reaction. Calculated mass of the asterisked peak (R_t = 8.55 min) is 18,000 Da. Large signal beginning at R_t = 9.5 min. attributed to MPAA. d) Circular Dichroism spectrum of the ~20 kDa SEC peak in 50 mM Na_2HPO_4, pH 7.5.

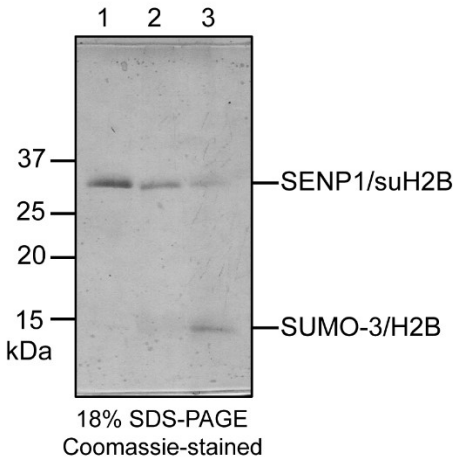


Figure 3.39. SENP1 hydrolysis of H2B(A117C)$^{Su(C47S)}$ (8). Coomassie-stained 18% SDS-PAGE gel, run at 200 V for 1.5 h showing products of the SENP1 hydrolysis assay. Lane 1 = SENP1 alone, 2 = H2B(A117C)$^{Su(C47S)}$ (**8**) alone, 3 = SENP1 and **8** incubated for 24 h.

3.6 References

Portions of this chapter have been published as:

Weller, C. E., Dhall, A., Ding, F., Linares, E., Whedon, S. D., Senger, N. A., Tyson, E. L., Bagert, J. D., Li, X., Augusto, O., and Chatterjee, C. Aromatic thiol-mediated cleavage of N-O bonds enables chemical ubiquitylation of folded proteins. *Nature Communications*. **2016**, *7*, 12979.

(1) Sharp, P. M.; Li, W.-H. Molecular evolution of ubiquitin genes. *Trends Ecol. Evol.* **1987**, *2* (11), 328–332.

(2) Danielsen, J. M. R.; Sylvestersen, K. B.; Bekker-Jensen, S.; Szklarczyk, D.; Poulsen, J. W.; Horn, H.; Jensen, L. J.; Mailand, N.; Nielsen, M. L. Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol. Cell. proteomics* **2011**, *10* (3), M110.003590.

(3) Li, W.; Bengtson, M. H.; Ulbrich, A.; Matsuda, A.; Reddy, V. A.; Orth, A.; Chanda, S. K.; Batalov, S.; Joazeiro, C. A. P. Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PLoS One* **2008**, *3* (1), e1487.

(4) Weller, C. E.; Pilkerton, M. E.; Chatterjee, C. Chemical strategies to understand the language of ubiquitin signaling. *Biopolymers* **2014**, *101* (2), 144–155.

(5) Abeywardana, T.; Pratt, M. R. Using chemistry to investigate the molecular consequences of protein ubiquitylation. *ChemBioChem* **2014**, *15* (11), 1547–1554.

(6) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F. Synthesis of K48-linked diubiquitin using dual native chemical ligation at lysine. *Chem. Commun.* **2010**, *46*, 7199–7201.

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

(8) Wan, Q.; Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: A powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chemie* **2007**, *119*, 9408–9412.

(9) Chatterjee, C.; McGinty, R. K.; Pellois, J.-P.; Muir, T. W. Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chemie* **2007**, *46*, 2814–2818.

(10) Offer, J. Native chemical ligation with Nα acyl transfer auxiliaries. *Biopolymers* **2010**, *94* (4), 530–541.

(11) Canne, L. E.; Bark, S. J.; Kent, S. B. H. Extending the applicability of native chemical ligation. *J. Am. Chem. Soc.* **1996**, *118*, 5891–5896.

(12) Weller, C. E.; Huang, W.; Chatterjee, C. Facile synthesis of native and protease-resistant ubiquitylated peptides. *ChemBioChem* **2014**, *15* (9), 1263–1267.

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

(14) Canne, L. E.; Botti, P.; Simon, R. J.; Chen, Y.; Dennis, E. A.; Kent, S. B. H. Chemical protein synthesis by solid phase ligation of unprotected peptide segments. *J. Am. Chem. Soc.* **1999**, *121*, 8720–8727.

(15) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. SUMOylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837.

(16) He, C.; Kulkarni, S. S.; Thuaud, F.; Bode, J. W. Chemical synthesis of the 20 kDa heme protein nitrophorin 4 by α-ketoacid-hydroxylamine (KAHA) ligation. *Angew. Chemie* **2015**, *54*, 1–7.

(17) Hartl, F. U.; Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **2009**, *16* (6), 574–581.

(18) Lockless, S. W.; Muir, T. W. Traceless protein splicing utilizing evolved split inteins. *Proc. Natl. Acad. Sci.* **2009**, *106* (27), 10999–11004.

(19) Evans, T. C.; Benner, J.; Xu, M.-Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.* **1998**, *7*, 2256–2264.

(20) Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. W. Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 388–393.

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

(22) Johnson, E. C. B.; Kent, S. B. H. Insights into the mechanism and catalysis of the native chemical ligation reaction. *J. Am. Chem. Soc.* **2006**, *128* (20), 6640–6646.

(23) Wang, C.; Guo, Q.-X.; Fu, Y. Theoretical analysis of the detailed mechanism of native chemical ligation reactions. *Chem. - An Asian J.* **2011**, *6* (5), 1241–1251.

(24) Thompson, R. E.; Liu, X.; Alonso-Garcia, N.; Pereira, P. J. B.; Jolliffe, K. A.; Payne, R. J. Trifluoroethanethiol: An additive for efficient one-pot peptide ligation-desulfurization chemistry. *J. Am. Chem. Soc.* **2014**, *136*, 8161–8164.

(25) Jin, W.; Trzupek, J. D.; Rayl, T. J.; Broward, M. A.; Vielhauer, G. A.; Weir, S. J.; Hwang, I.; Boger, D. L. A unique class of duocarmycin and CC-1065 analogues subject to reductive activation. *J. Am. Chem. Soc.* **2007**, *129* (49), 15391–15397.

(26) Fava, A.; Reichenbach, G.; Peron, U. Kinetics of the thiol-disulfide exchange. II. Oxygen-promoted free-radical exchange between aromatic thiols and disulfides. *J. Am. Chem. Soc.* **1967**, *89* (25), 6696–6700.

(27) Denes, F.; Pichowicz, M.; Povie, G.; Renaud, P. Thiyl radicals in organic synthesis. *Chem. Rev.* **2014**, *114*, 2587–2693.

(28) Forni, L. G.; Willson, R. L. Thiyl and phenoxyl free radicals and NADH: Direct observation of one-electron oxidation. *Biochem. J.* **1986**, *240*, 897–903.

(29) Zhao, R.; Lind, J.; Merbnyi, G.; Eriksen, T. E. Kinetics of one-electron oxidation of thiols and

hydrogen abstraction by thiyl radicals from α-amino C-H bonds. *J. Am. Chem. Soc.* **1994**, *116* (9), 12010–12015.

(30) Jeschke, G. EPR techniques for studying radical enzymes. *Biochim. Biophys. Acta* **2005**, *1707*, 91–102.

(31) Augusto, O.; Bonini, M. G.; Trindade, D. F. Spin trapping of glutathiyl and protein radicals produced from nitric oxide-derived oxidants. *Free Radic. Biol. Med.* **2004**, *36* (10), 1224–1232.

(32) Duling, D. R. Simulation of multiple isotropic spin-trap EPR spectra. *Journal of magnetic resonance, Series B*. 1994, pp 105–110.

(33) Mile, B.; Rowlands, C. C.; Sillman, P. D.; Fildes, M. The EPR spectra of thiyl radical spin adducts produced by photolysis of disulfides in the presence of 2,4,6-tri-tert-butylnitrosobenrene and 5,5-dimethyl-1-pyrroline N-oxide. *J. Chem. Soc. Perkin Trans. 2* **1992**, 1431–1437.

(34) Wu, M.; Begley, T. P. β-scission of the N-O bond in alkyl hydroxamate radicals: A fast radical trap. *Org. Lett.* **2000**, *2* (10), 1345–1348.

(35) Caspari, G.; Granzow, A. The flash photolysis of mercaptans in aqueous solution. *J. Phys. Chem.* **1970**, *74* (4), 836–839.

(36) Lawrence, C. C.; Bennati, M.; Obias, H. V; Bar, G.; Griffin, R. G.; Stubbe, J. High-field EPR detection of a disulfide radical anion in the reduction of cytidine 5′-diphosphate by the E441Q R1 mutant of Escherichia coli ribonucleotide reductase. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 8979–8984.

(37) Tung, T.-L.; Stone, J. A. The formation and reactions of disulfide radical anions in aqueous solution. *Can. J. Chem.* **1975**, *53* (21), 3153–3157.

(38) Petlicki, J.; van de Ven, T. G. M. The equilibrium between the oxidation of hydrogen peroxide by oxygen and the dismutation of peroxyl or superoxide radicals in aqueous solutions in contact with oxygen. *J. Chem. Soc. Faraday Trans.* **1998**, *94*, 2763–2767.

(39) Kuppusamy, P.; Zweier, J. L. Characterization of free radical generation by xanthine oxidase: Evidence for hydroxyl radical generation. *J. Biol. Chem.* **1989**, *264* (17), 9880–9884.

(40) Purdie, J. W.; Gillis, H. A.; Klassen, N. V. Pulse radiolysis of penicillamine in aqueous solution: The thiyl radical and the disulphide radical anion. *Chem. Commun.* **1971**, No. 19, 1163–1165.

(41) Antonello, S.; Daasbjerg, K.; Jensen, H.; Taddei, F.; Maran, F. Formation and cleavage of aromatic disulfide radical anions. *J. Am. Chem. Soc.* **2003**, *125*, 14905–14916.

(42) Goldstein, S.; Meyerstein, D.; Czapski, G. The Fenton reagents. *Free Radic. Biol. Med.* **1993**, *15*, 435–445.

(43) Paulsen, C. E.; Carroll, K. S. Cysteine-mediated redox signaling: Chemistry, biology, and tools for discovery. *Chem. Rev.* **2013**, *113*, 4633–4679.

(44) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; G. Scalmani, V. B.; Petersson, G. A.; Nakatsuji, H.; Li, X.; et al. Gaussian 09, Revision A.02. Gaussian, Inc.: Wallingford CT 2014.

(45) Moyal, T.; Hemantha, H. P.; Siman, P.; Refua, M.; Brik, A. Highly efficient one-pot ligation and desulfurization. *Chem. Sci.* **2013**, *4*, 2496–2501.

(46) Shiio, Y.; Eisenman, R. N. Histone SUMOylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230.

(47) Zheng, J.; Tang, S.; Qi, Y.; Wang, Z.; Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **2013**, *8* (12), 2483–2495.

(48) Hendriks, I. A.; D'Souza, R. C. J.; Yang, B.; Verlaan-de Vries, M.; Mann, M.; Vertegaal, A. C. O. Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat. Struct. Mol. Biol.* **2014**, *21* (10), 927–936.

(49) Chandrasekharan, M. B.; Huang, F.; Sun, Z.-W. Ubiquitination of histone H2B regulates chromatin dynamics by enhancing nucleosome stability. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (39), 16686–16691.

(50) Mattingly, P. G.; Miller, M. J. Titanium trichloride reduction of substituted N-hydroxy-2-azetidinones and other hydroxamic acids. *J. Org. Chem.* **1980**, *45* (3), 410–415.

(51) Denmark, S. E.; Stolle, A.; Dixon, J. A.; Guagnano, V. Tandem inter [4+2]/intra [3+2] cycloadditions. 6. The bridged mode. *J. Am. Chem. Soc.* **1995**, *117*, 2100–2101.

(52) Keck, G.; Fleming, S.; Nickell, D.; Weider, P. Mild and efficient methods for the reductive cleavage of nitrogen-oxygen bonds. *Synth. Commun.* **1979**, *9* (4), 281–286.

(53) Keck, G. E.; Wager, T. T.; McHardy, S. F. Reductive cleavage of N-O bonds in hydroxylamines and hydroxamic acid derivatives using samarium diiodide. *Tetrahedron* **1999**, *55*, 11755–11772.

(54) Cutulic, S. P. Y.; Murphy, J. A.; Farwaha, H.; Zhou, S.-Z.; Chrystal, E. Metal-free reductive cleavage of N-O bonds in weinreb amides by an organic neutral super-electron donor. *Synlett* **2008**, *14*, 2132–2136.

(55) Sevier, C. S.; Kaiser, C. a. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **2002**, *3* (11), 836–847.

(56) King, D. S.; Fields, C. G.; Fields, G. B. A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis. *Int. J. Pept. Protein Res.* **1990**, *36*, 255–266.

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

(58) Watanabe, T.; Honda, K. Measurement of the extinction coefficient of the methyl viologen cation radical and the efficiency of its formation by semiconductor photocatalysis. *J. Phys. Chem.* **1982**, *86*, 2617–2619.

(59) Guo, Q.; Qian, S. Y.; Mason, R. P. Separation and identification of DMPO adducts of oxygen-centered radicals formed from organic hydroperoxides by HPLC-ESR, ESI-MS and MS/MS. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 862–871.

(60) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G.

J.; Shokat, K. M. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007**, *128* (5), 1003–1012.

(61) Chen, J.; Rogers, S. C.; Kavdia, M. Analysis of kinetics of dihydroethidium fluorescence with superoxide using xanthine oxidase and hypoxanthine assay. *Ann. Biomed. Eng.* **2013**, *41* (2), 327–337.

(62) Jezewska, M. M. Xanthine accumulation during hypoxanthine oxidation by milk xanthine oxidase. *Eur. J. Biochem.* **1973**, *36*, 385–390.

(63) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008**, *453* (7196), 812–816.

Chapter 4

Biochemical investigation of cross-talk between histone H4 SUMOylation and histone H3 acetylation

4.1 Introduction

Modification of proteins with the small ubiquitin-like modifier (SUMO) is important in a variety of intracellular processes. SUMOylation can antagonize other Lys modifications at the same site,[1] alter substrate intracellular distribution,[2] prevent binding between proteins[3] or between transcription factors and DNA,[4] reduce the toxicity of unfolded proteins by inhibiting their aggregation,[5] facilitate multiprotein complex formation,[6] and more. SUMO is conjugated to the Lys ε-amine of substrates by an ATP-dependent enzymatic cascade resembling that of ubiquitin (Ub), described in Chapter 1.[7] In contrast to Ub, there is only one E1 and one E2 SUMO ligase known. The SUMO E1 enzyme is a heterodimer of SUMO-activating enzyme subunits 1 and 2 (SAE1/SAE2). The E1 utilizes ATP to activate the C-terminus of SUMO as an adenylate, and subsequently forms a C-terminal SUMO thioester with the active-site Cys of SAE2. The Ub fold domain of SAE2 then interacts with the SUMO E2 enzyme, Ubc9, and SUMO is transferred to Ubc9, again as a C-terminal thioester with the active-site Cys. Ubc9 may then interact with an E3 ligase to facilitate transfer of SUMO to a substrate Lys ε-amine, or transfer SUMO directly to the substrate itself without the need for an E3. Ubc9 is capable of binding to a consensus motif, ΨKXE, where Ψ is a large, hydrophobic amino acid, usually L/I/V, X is any amino acid, and K is the Lys to be SUMOylated.[8,9] This enables a degree of substrate specificity by Ubc9, even in the absence of an E3. The largest family of known E3 SUMO ligases is the siz/PIAS-RING (SP-RING)-containing protein inhibitor of activated STAT (PIAS) family, consisting of PIAS1, PIAS3, PIASxα, PIASxβ and PIASy in humans.[10] They act in an analogous manner to Ub E3 RING-type ligases, by holding substrate and SUMO-charged Ubc9 thioester in a position favorable to nucleophilic attack by the substrate Lys. An example of a different class of E3 is the nucleoporin Ran binding protein 2 (RanBP2), which is neither HECT- nor RING-type. RanBP2 binds both SUMO and Ubc9

in SUMO-charged Ubc9, and holds them in a conformation that favors catalysis. In this manner it stimulates Ubc9 activity without making contact with substrate proteins, and its location at the nuclear pore has spurred the hypothesis that it may be responsible for nucleocytoplasmic transport-coupled SUMOylation.[7] Other notable E3 ligases are the polycomb 2 protein (PC2) and topoisomerase I-binding arginine/serine rich protein (TOPORS),[10] and SUMO E3 activity continues to be discovered as an unexpected secondary activity of known proteins, including histone deacetylase 4 (HDAC4)[11] and the DNA replication-related element-binding *factor (*DREF) transcription factor.[12] Some, but not all, E3 ligases display target specificity, and some display SUMO isoform specificity.[7]

There are three functional paralogs of SUMO in humans: SUMO-1, SUMO-2, and SUMO-3. SUMO-2 and -3 share 97% sequence identity and are collectively referred to as SUMO-2/3. They each share 47% sequence identity with SUMO-1.[10] Like Ub, each SUMO paralog contains a C-terminal Gly-Gly motif that must be proteolytically revealed following expression of the immature SUMO precursor. This is carried out by the same enzymes responsible for substrate de-SUMOylation, the SUMO/sentrin-specific protease (SENP) family. Of the six known SENPs, SENP1 and 2 are active towards SUMO-1 and -2/3. SENP3, 5, 6, and 7 are only active toward SUMO-2/3.[13] Another SUMO paralog, SUMO-4, contains a Pro residue near its C-terminus that significantly retards proteolytic maturation by SENPs, thus it has been found to modify an extremely limited number of substrates.[14] The SUMO paralogs share many substrates, yet exhibit functional differences. For example, modification by SUMO is often monomeric, but can be polymeric. SUMO polymers form via Lys11 of SUMO-2/3, and the addition of SUMO-1 effectively terminates the chain.[15] Their intracellular distribution and dynamics also differ. In HeLa cells expressing YFP fused to SUMO-1, -2, or -3, all paralogs were localized to the nucleus and PML bodies.[16] However, SUMO-1 showed unique localization to the nuclear periphery and nucleolus, whereas SUMO-2/3 accumulated on chromosomes during post-mitotic nuclear reformation much more rapidly than SUMO-1. Further, fluorescent recovery after photobleaching (FRAP) experiments revealed that conjugation and deconjugation of SUMO-2/3 is more dynamic than that of SUMO-1. SUMO-2/3 has also been shown to be

more abundant in its free, unconjugated form than SUMO-1, and significantly more abundant under heat stress conditions, whereas SUMO-1 is not similarly upregulated.[17]

Some functional divergence can be accounted for by differences in the ability of SUMO-binding proteins to recognize the different SUMO paralogs. Proteins interact with SUMO through SUMO-interacting motifs (SIMs), typically with a dissociation constant (K_d) in the low micromolar range.[18] All Ub-like proteins (Ubls) share the β-grasp fold structural motif, which is made up of five antiparallel β-strands overlaid with a single α-helix. The differences between SUMO-1 and SUMO-2/3 are located primarily in the α-helix (α_1) and in the second β-strand (β_2). The SIM consists of a loosely-conserved, linear sequence that fits in the groove between α_1 and β_2 and extends the β-sheet of SUMO. The SIM can bind in either the parallel or antiparallel orientation.[19] Key to the SIM consensus sequence is a hydrophobic core, typically V/I-V/I-X-V/I (where X is an acidic residue). This core is often flanked by a stretch of acidic residues, commonly Asp or Glu, that interacts with a basic patch on the surface of SUMO adjacent to the α_1-β_2 groove.[20] The sequence composition of SIMs is extremely diverse, and they can bind SUMO paralogs equally, such as the SIM of PIASxα,[20] or favor binding to one paralog. For instance, SUMOylation of MBD1 strengthens its binding interaction with MBD1-containing chromatin-associated factor 1 (MCAF1), which shows a preference for SUMO-2/3: it binds SUMO-2/3 with a K_d of 1.3 μM, and SUMO-1 with a K_d of 13.9 μM.[21] Because the K_d of SIM-SUMO binding is relatively high, it is possible that specific interactions of SUMOylated proteins are enabled by simultaneous low-affinity interaction at a second site.[10]

In the last decade, the SUMOylation of chromatin and chromatin-associated proteins and its effect on gene regulation has been a subject of increasing research and interest. Pioneering studies by Shiio and Eisenman revealed endogenous H4 modification by SUMO-1 in human B-lymphocytes. Co-expression of FLAG-tagged core histones and HA-tagged SUMO-1 or SUMO-3 in HEK293T cells showed that H4 is capable of modification by both paralogs, and is the primary site of histone SUMOylation.[22] Later studies by Berger and coworkers found that in *S. cerevisiae* all four core histones are modified by the yeast SUMO,

Smt3, which shares ~50% sequence identity with both human SUMO-1 and SUMO-2/3, and confirmed that SUMOylated histones were present within chromatin.[23] At endogenous expression levels of histones and SUMO, they estimated that ~5% of each histone is SUMOylated at any given time. While they were able to locate the site of H4 SUMOylation to the N-terminal tail region, a subsequent proteomic study identified Lys12 at the major site of H4 SUMOylation.[24] Further proteomic studies confirmed H4 Lys12, and indeed sites on all core histones, as substrates of SUMOylation,[1] and demonstrated that H4 is modified by SUMO-1 and SUMO-2/3 in roughly equal amounts.[25] The consistent identification of H4 in independent studies of SUMOylation targets indicates its legitimacy as a SUMOylation substrate, and raises the question of its effect on chromatin structure and function.

Upon its initial discovery, SUMOylated H4 (suH4) was associated with transcriptional repression. SUMO-H4 and SUMO-H2B fusions were expressed in *S. cerevisiae*, and induction of the GAL1 gene was monitored when medium was switched to a galactose carbon source. Interestingly, induction was reduced two-fold relative to cells lacking a SUMO fusion protein, and 'double-affinity' chromatin immunoprecipitation (ChDIP) experiments revealed an inverse correlation between SUMO and histone acetylation at that promoter.[23] FLAG-tagged SUMO-1 and SUMO-3 H4 fusions were also employed in pull-down experiments. From HEK293T lysate, both fusion proteins immunoprecipitated histone deacetylase 1 (HDAC1), an enzyme strongly associated with repression, which supports a model of opposing histone SUMOylation and acetylation.[22] Most studies of SUMOylated chromatin to date, however, involve methods that have been unable to distinguish between SUMOylation of histones or of chromatin associated proteins. Thus, they have shown an indirect correlation between histone SUMOylation and a transcriptional outcome. For instance, in HEK293T cells a GAL4-Ubc9 fusion repressed expression of a luciferase reporter gene under control of a GAL4-dependent promoter, and showed dependence on Ubc9 catalytic activity.[22] This experiment also showed a reduction in local acetylation by chromatin immunoprecipitation (ChIP), but did not identify the SUMOylated proteins at chromatin.

Although sometimes difficult to parse, SUMO ChIP-Seq data correlated with RNA-Seq data can be informative. Two separate studies, one in yeast and one in human fibroblasts, surprisingly found that SUMO localized to the promoters of highly and constitutively active genes. Manley and coworkers found that in yeast, SUMO was located only at active promoters, although a previous study found that SUMOylated histones were enriched in telomeric regions.[23,26] Ubc9 was not present at these active promoters, and deletion of Ubc9 caused only a slight decrease in transcription at these sites. However, expression of some inducible genes, such as ARG1, resulted in Ubc9 localization and significant promoter SUMOylation. In these cases, Ubc9 deletion caused a significant increase in transcription, indicating that chromatin SUMOylation may represent a means to turn off the temporary activation of genes. In human fibroblast experiments, SUMO-1 and -2 were also highly localized to active promoters, and correlated with RNA polymerase II (Pol II) localization.[27] SUMO was also found at repressed genes, especially developmentally regulated genes. Of note was the observation that sires of active SUMOylation, in which SUMO, Ubc9 and PIASy co-localized, occurred primarily at promoters controlling protein and histone biogenesis genes, such as those for histones, ribosomal proteins, rRNA (transcribed by RNA polymerase I), and tRNA (transcribed by RNA polymerase III). As in yeast, genes under active SUMOylation were upregulated when the SUMOylation machinery was disturbed. This indicates that SUMO plays a maintenance role in tempering expression of those genes. The stark differences in SUMO-correlated transcriptional activity between different classes of genes could have many explanations. One possibility is that SUMO-mediated transcriptional repression requires active SUMOylation at chromatin, perhaps due to the rapid rate of SUMO conjugation and deconjugation.[16] This implies that SUMO on chromatin without active local SUMOylation activity may be conjugated to chromatin associated factors prior to their deposition on chromatin. Further support of this theory is the heat shock response, which results in such expansive SUMOylation of chromatin that it depletes the entire nuclear pool of free SUMO-2/3.[28] Under these conditions, PIAS1 and SUMO-2/3 accumulate significantly and preferentially at active, Pol II-occupied promoters, and lead to increased promoter-proximal pausing and temporary repression of those genes.[29] Still, the mechanism of

SUMO-mediated transcriptional repression could be very different in each case, and the role of H4 SUMOylation in these processes remains to be elucidated.

Transcriptional repression due to suH4 may result from steric occlusion of the promoter by SUMO, by preventing the approach of transcriptional machinery and enzymes that deposit activating marks. For instance, the transcription-activating SAGA complex, which contains the Gcn5 histone acetyltransferase, is also the major deubiquitinase for Ub at H2B Lys120 (ubH2B), which is repressive when located in promoter regions.[30] This complex does not have deSUMOylating activity, and suH4 could potentially inhibit its histone acetylation. Repression may also result from *trans* effects, whereby SUMO recruits effector proteins or complexes that cause a change in chromatin and establish a compacted, heterochromatic state. Until recently, it was unknown whether suH4 would permit chromatin compaction. As a comparison, Ub has a similar size and structure to SUMO, and Lys119 ubiquitylation of H2A (ubH2A) is a repressive mark that permits chromatin compaction.[31,32] In contrast, ubH2B, which is an active mark when found within gene bodies, does not allow higher-order compaction.[33] Studies in our lab with suH4 have revealed that SUMO disrupts internucleosomal interactions, and that fully SUMOylated arrays are refractory to compaction.[34] However, sub-stoichiometric, 50% occupancy of suH4 does permit compaction, although the formation of higher-order structures is moderately impaired. Hence, sub-stoichiometric suH4 could allow heterochromatin formation. Alternatively, a high local density of suH4 could initiate repression, after which de-SUMOylation would permit compaction.

In support of a trans effect of suH4, Muir and coworkers have demonstrated that cross-talk exists between histone ubiquitylation and methylation: ubH2B activates the methyltransferase hDot1L toward H3 K79 methylation.[35] Indeed, SUMO could similarly stimulate repressive enzymes upon binding, or simply enhance their local action through a recruitment mechanism. In fact, the repressive effect of SUMO is known to be dependent on its surface basic patch, which is important for binding to many SIMs.[36,37] There are many examples of transcription factor (TF) SUMOylation, most of which result in repression of the TF

target gene.[38] The mechanism of this effect is not known in all cases, but TF SUMOylation has been shown to recruit repressor complexes. For example, SUMOylated TF Sp3 recruits heterochromatin factors such as HP1α, the H3 K9 histone methyltransferase SETDB1, the H4 K20 histone methyltransferase Suv4-20h, and the nucleosome remodeling ATPase Mi-2 in complex with several other factors.[39] In addition, the activity of TF Elk-1 becomes inhibitory toward transcription when Elk-1 is SUMOylated, and SUMOylated Elk-1 was found to bind HDAC2 in a pull-down.[40] The study further showed that SUMO alone, when fused to GAL4, could recruit HDAC2 and repress a luciferase reporter gene under control of the GAL1 promoter. Recently, Gill and coworkers reported gene-specific recruitment of the corepressor of RE1 silencing transcription factor (CoREST), lysine specific demethylase 1 (LSD1), and HDAC1 complex by SUMO-2/3, but not SUMO-1.[41] CoREST was initially identified as a corepressor for the RE1-silencing transcription factor (REST, or the neuron-restrictive silencer factor, NRSF), which binds a DNA sequence known as RE1 (or neuron-restrictive silencer element, NRSE). REST silences neuronal genes in non-neuronal cells by recruiting several HDAC-containing complexes.[42,43] In the aforementioned study, the authors found that CoREST mediated repression of the sodium voltage-gated channel alpha subunits 1 and 3 (SCN1A and SCN3A), but not 2 (SCN2A2), and that this repression, as well as CoREST localization to the gene promoters, was dependent on a functional SUMO-2/3-binding SIM in CoREST. Interestingly, only SCN2A2 has been confirmed to be REST-dependent,[44] thus SUMO-2/3 at chromatin may represent an alternative mechanism of CoREST recruitment.

The CoREST complex, like other repressive complexes, has interchangeable subunits based on a common core scaffold.[45–47] In addition to LSD1 and HDAC1, the proteins BHC80 and BRAF35 are sometimes, but are not always, present. BRAF35 is important for REST-mediated repression of neuronal genes.[48] The role of BHC80 is unclear, although it may suppress LSD1 activity.[49] Different compositions of the complex can thus have different activities. For instance, *in vitro* assays with the LSD1-HDAC1-CoREST complex and HeLa-extracted mononucleosomes have shown an interdependence between deacetylation and demethylation activity,[50] in which deacetylation enhances demethylation and vice versa. CoREST contains

a C-terminal ELM2 (Egl-27 and MTA1 homology 2) domain, which is sufficient for binding to HDAC1/2 (**Figure 4.1a**).[50] It also contains two SANT (SWI-SNF, ADA, N-CoR, and TFIIIB) domains. The SANT2 domain binds DNA and is critical for LDS1-mediated H3 K4me$_{1/2}$ demethylation of nucleosomes.[49,51] The SANT1 domain contributes to HDAC1/2 binding and may bind DNA as well, based on homology with the SANT2 domain.[50,52] Between the SANT domains is a linker region that tightly binds LSD1, with a K_d of ~16 nM.[51,53] The SIM of CoREST is also located in the linker region, between SANT1 and the LSD1-binding

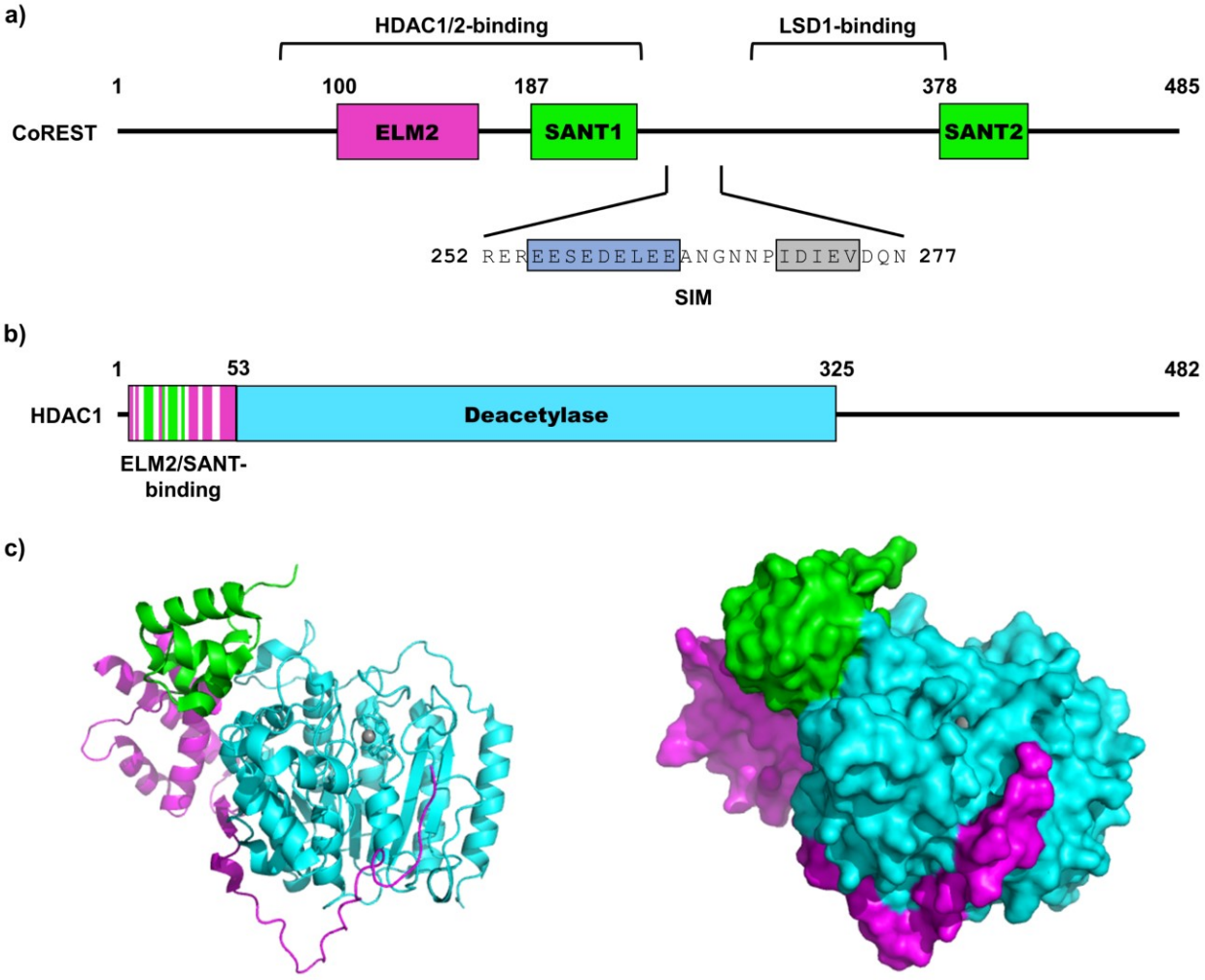


Figure 4.1. Domains of CoREST and HDAC1. a) Domain structure of CoREST, with HDAC1/2, LSD1 binding, and SUMO-interacting motif (SIM) indicated. The SIM hydrophobic core is indicated in gray, and the acidic patch indicated with blue.[41] b) Domain structure of HDAC1. Sites important for binding ELM2 and SANT domains are indicated in purple and green, respectively.[52] c) Crystal structure of HDAC1 (blue) in complex with the ELM2-SANT domains (purple and green, respectively) of metastasis associated protein 1 (MTA1). Catalytic Zn^{2+} ion is indicated in gray. PDB code 5ICN.

region.[41] To test the hypothesis that suH4 mediates transcriptional repression by cross-talk, we investigated whether suH4 could facilitate nucleosome deacetylation by the HDAC1-CoREST dimeric complex.

HDAC1 is a class I histone deacetylase, as are HDAC2, 3, and 8. They are ubiquitously expressed in all tissue types, and are primarily nuclear – HDAC1 and 2 are almost exclusively nuclear.[54] HDAC1, 2, and 3 are usually localized to chromatin, although they are excluded from mitotic chromosomes by an unknown mechanism. Despite their name, HDACs are not limited to histones and act on a wide variety of substrates with diverse function.[55] Although relatively promiscuous, they do exhibit some degree of substrate preference with regard to residues immediately surrounding acetyllysine.[56] The HDACs are extremely important clinical targets, and due to their substrate diversity HDAC inhibitors have been used as treatment for Alzheimer's disease, heart disease, HIV infection, and most commonly, cancer.[57] Developing HDAC- and even HDAC class-specific inhibitors remains a challenge due to strong similarities between HDAC active sites. These Zn^{2+}-dependent enzymes catalyze deacetylation of a substrate Lys ε-amine. Two Asp and one His residue hold Zn^{2+} in the active site, where it coordinates to and activates a water molecule for nucleophilic attack on the side chain amide.[58] Therefore, many inhibitors, notably hydroxamic acids, act by forming stable chelates with the catalytic zinc ion.[59]

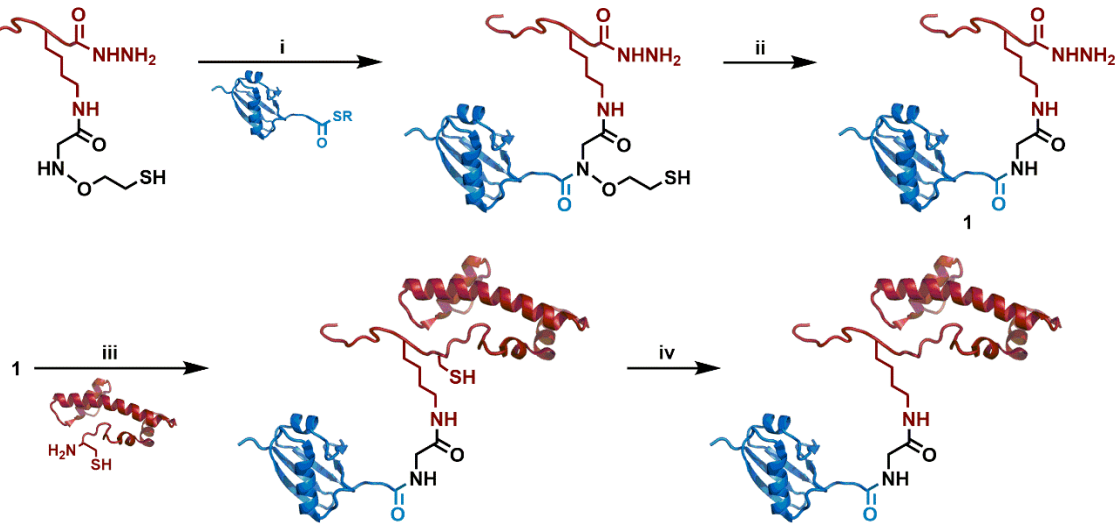
HDAC1 and HDAC2 share 85% sequence identity, with most differences located in the unstructured C-terminal region (**Figure 4.1b**). While this region has no deacetylase activity, it is heavily post-translationally modified and required for full HDAC1 activity.[60] Modification in this region also accounts for differences in HDAC1 and HDAC2 regulation. For example, HDAC1, but not HDAC2, can be acetylated in its C-terminal domain by p300, which significantly reduces its deacetylation activity and thus transcriptional repressive ability.[61,62] Some functions, however, are redundant, and studies have revealed that RNAi knock-down of HDAC1 or 2 results in compensatory upregulation of the other.[63] HDAC1 and 2 are also incorporated nearly interchangeably into many transcriptional repressor complexes, including CoREST, NuRD, Sin3, and MIDAC.[64] With the exception of Sin3, HDAC1/2 binds to an ELM2-SANT motif in one of the core complex

proteins (**Figure 4.1c**), however the binding constant for this interaction has not been reported. Evidence suggests that ELM2 domain binding can stimulate HDAC1 activity. Experiments with HeLa-extracted mononucleosomes and HDAC1 showed that HDAC1 activity increased slightly, and to the same degree, when either full-length CoREST or CoREST truncated C-terminal to the ELM2 domain were included in the assay.[50] HDAC1/2 is most often found in complex with other proteins *in vivo*, and its target specificity is controlled in part by its binding partners. By this mechanism, suH4 may recruit CoREST and direct HDAC1 toward local histone deacetylation. In this chapter, I detail studies that confirm an inhibitory role of suH4 by *in vitro* transcription experiments. I also demonstrate, by *in vitro* enzymatic assays with the dimeric HDAC1-CoREST complex, that suH4 recruits HDAC1 in a CoREST-SIM-dependent manner, and stimulates deacetylation of H3 K14Ac in semisynthetic mononucleosome substrates, thus confirming the feasibility of this mechanism *in vivo*.

4.2 Results and discussion

4.2.1 Transcription from SUMOylated chromatin *in vitro*

Studies that link suH4 with transcriptional repression have thus far only shown correlation, not a definitive causative effect. To prove that SUMOylation of chromatin has a direct repressive effect, we collaborated with the Roeder lab at Rockefeller University to perform *in vitro* transcription assays from a chromatinized template containing suH4. Wild-type (wt) human histones were expressed and purified from *E. coli*, and native, semisynthetic suH4 was prepared by a synthetic strategy described previously,[65] utilizing the native chemical ligation auxiliary 2-(aminooxy)ethanethiol, discussed in Chapters 2 and 3 (**Scheme 4.1**).[66–68] Octamers containing wt or SUMOylated H4 were formed by combining the four core histones in equimolar amounts in denaturing buffer, then folding by dialysis into a high-salt buffer.[69] The resulting octamers were purified by size exclusion chromatography (**Figure 4.2a**), then incorporated into chromatin on a circular plasmid DNA template. The plasmid lacks any strong nucleosome positioning sequences, and in the presence of the histone chaperone NAP1 and chromatin remodelers ACF1 and ISWI nucleosomes were positioned regularly on the template, each occupying ~150 bp DNA (**Figure 4.2b**).



Scheme 4.1. Semisynthesis of SUMOylated H4 (suH4). i) SUMO-3(2-91)C47S-MES, 6 M Gn-HCl, 100 mM Na_2HPO_4, 10 mM TCEP, pH 7.3, 25 °C, 24 h. ii) 6 M Gn-HCl, pH 3, Zn, 37 °C, 24 h. iii) a) 6 M Gn-HCl, 200 mM Na_2HPO_4, 10 mM $NaNO_2$, pH 3, -20 °C, 15 min, b) H4(15-102)A15C, 6 M Gn-HCl, 200 mM Na_2HPO_4, 130 mM MPAA, pH 6.9, 25 °C, 24 h. iv) 6 M Gn-HCl, 100 mM Na_2HPO_4, 500 mM TCEP, 100 mM MESNa, 280 mM *t*-BuSH, 10 mM VA-044, 37 °C, 24 h.

The plasmid template contains five GAL4 binding sites, the adenovirus major late core promoter, and a 400 bp G-less cassette.[70] Transcription of the cassette is dependent on supplemented GAL4-VP16 activator, p300 and acetyl-CoA, and transcriptional machinery from a human cell nuclear extract. Transcription in the presence of [α-^{32}P] CTP from chromatin containing wtH4 or suH4 was monitored by autoradiography. We observed a striking decrease in transcription from suH4 chromatin relative to wt chromatin (**Figure 4.3a**). There are several possible mechanisms that may be responsible for this effect. First, SUMO may inhibit crucial co-transcriptional chromatin remodeling processes. We believe this is unlikely, however, due to the successful incorporation of suH4 containing octamers into the chromatin template, and the known activating effect of ubH2B, a modification of similar size, within gene bodies.[71] Another possibility is that suH4 in the promoter region inhibits binding of the GAL4-VP16 activator, or the approach of the HAT p300, despite the open chromatin structure promoted by suH4.[34] In this system, acetylation of the H3 and H4 tails by p300 is necessary for robust transcription.[72] Conversely, suH4 could promote recruitment of repressive chromatin-modifying complexes from the nuclear lysate, such as those containing HDACs.[65]

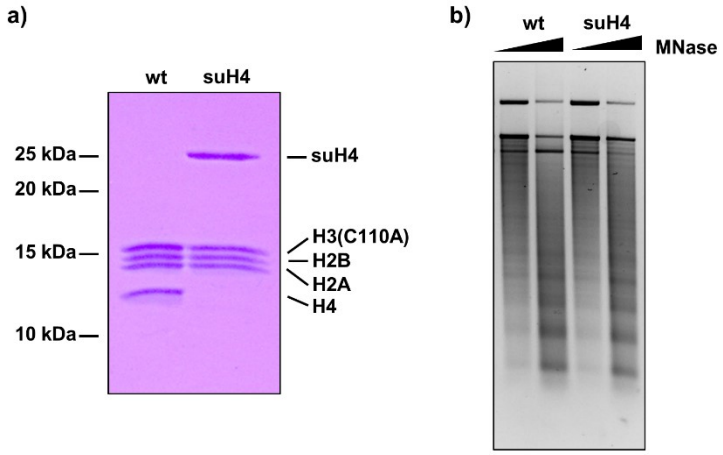


Figure 4.2. Generation of chromatin for *in vitro* transcription assays. a) Coomassie-stained 15% SDS-PAGE gel of wt and suH4 containing histone octamers. b) Ethidium bromide-stained 1.25% agarose gel of micrococcal nuclease-digested wt and suH4 chromatinized template DNA.

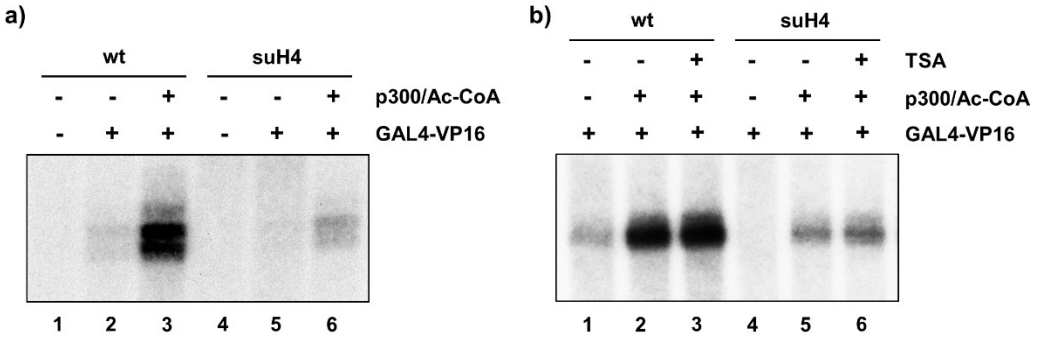


Figure 4.3. Transcription assays with wt or suH4 containing chromatin *in vitro*. a) Autoradiographs of 5% denaturing polyacrylamide gels of [α-^{32}P]-labeled transcripts from *in vitro* transcription assays with wt or suH4 containing chromatinized template. b) Assays were repeated in the presence of trichostatin A (TSA).

To test whether suH4 affected HDAC activity, we repeated the assay in the presence of trichostatin A (TSA), a nanomolar inhibitor of class I and II HDACs.[55] In suH4, but not wt, chromatin, TSA treatment led to a small increase in transcription relative to the untreated reaction (**Figure 4.3b**). Therefore, HDAC recruitment likely contributes to transcriptional repression by suH4 in this system. However, further studies must be performed to confirm this result, and to identify other sources of repression in this system.

4.2.2 Semisynthesis of H3 K14Ac substrate for *in vitro* deacetylation assays

To test the hypothesis that H4 SUMOylation leads to transcriptional repression by recruiting repressive complexes, we opted to examine the CoREST complex as a model system. Repression of the SCN1A and SCN3A genes by CoREST is known to be dependent on both SUMO-2/3 expression and a functional SIM in CoREST.[41] We therefore wondered if this represents direct recruitment of the CoREST complex by SUMO-2/3 located at chromatin, which would lead to the removal of H3 K4me$_2$ and histone acetyl marks by LSD1 and HDAC1/2, respectively. Recent *in vitro* studies in our lab have revealed that this mechanism is indeed possible, as evidenced by a 2-fold stimulation of intranucleosomal demethylation by the LSD1-CoREST dimeric complex in suH4 containing MN relative to wt MN.[65] Thus, the same stimulation may be seen with deacetylation by the HDAC1-CoREST complex.

As a substrate for deacetylation assays, we first synthesized H3 with a γ-sulfide mimic of acetyllysine at position 14.[73] HDAC1 is known to deacetylate the N-terminal tails of all four core histones. Because H3 tail acetylation is crucial to active transcription, and H3 K14 is in fact a primary target of many acetyltransferases, including Gcn5,[74] we considered this a reasonable initial test. We first purified recombinant H3 with a K14C mutation in its N-terminal tail region, then performed a thiol-ene 'click' reaction between the H3 mutant and *N*-vinylacetamide using the water-soluble radical initiator 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) (**Figure 4.4a-b**). Because we intended to monitor deacetylation by western blot, it was necessary to test the ability of H3 K14$_S$Ac to be recognized by primary α-acetyllysine antibodies. As a positive control, we hyper-acetylated wt H3 with acetic anhydride in acetate

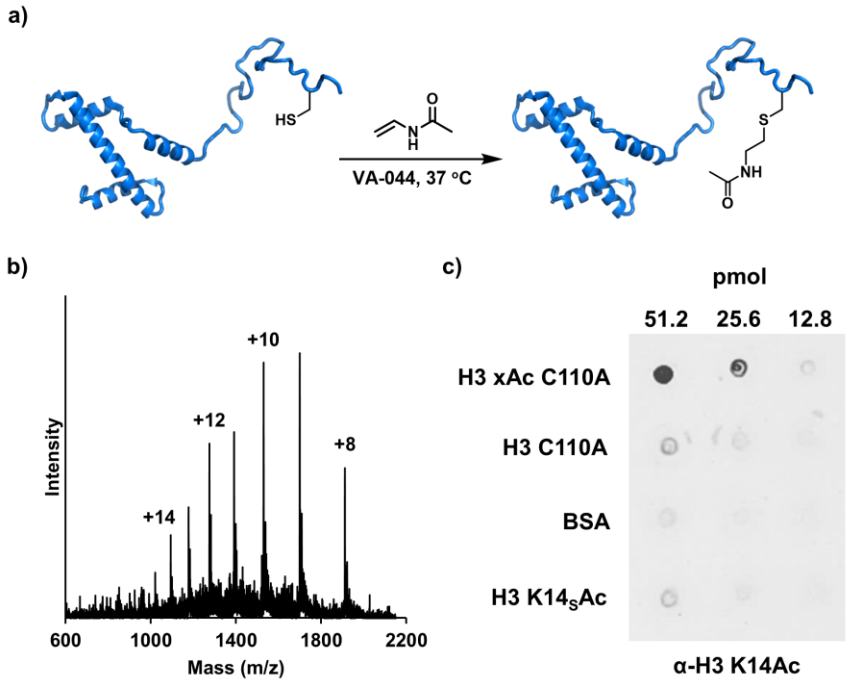


Figure 4.4. Synthesis and antibody recognition of H3 K14$_S$Ac. a) Schematic of thiol-ene 'click' reaction between H3 K14C, C110A and *N*-vinylacetamide to generate H3K14$_S$Ac. b) ESI-MS of purified H3K14$_S$Ac. c) Immunodot assay for detection of H3K14$_S$Ac by α-H3 K14Ac antibody.

buffer under basic conditions. By immunodot assay, several pan-acetyllysine antibodies were unable to bind H3 K14$_S$Ac, so we attempted the assay with two different commercial H3 K14Ac-specific antibodies. Unfortunately, and contrary to previous reports with γ-sulfide mimics of acetyllysine,[73] both antibodies failed to detect the acetyl H3 mimic (**Figure 4.4c**). We therefore used a semisynthetic approach to access native H3 K14Ac.

Semisynthetic H3 K14Ac was generated by first synthesizing a 28-mer H3 N-terminal tail peptide, H3(1-28, K14Ac)-C(O)NHNH$_2$, with C-terminal hydrazide, by 9-fluorenylmethoxycarbonyl based solid-phase peptide synthesis (Fmoc-SPPS). Separately, we expressed and purified H3(29-135) bearing an A29C mutation at its C-terminus. The H3 N-terminal peptide hydrazide was oxidized to a C-terminal azide by $NaNO_2$ at pH 3 and -20 °C. This species was converted to a C-terminal α-thioester *in situ* by the addition of 4-mercaptophenylacetic acid (MPAA), then H3(29-135)A29C added, the pH increased to 6.9, and the ligation

allowed to proceed for 24 h at room temperature. Following RP-HPLC purification, the ligation product underwent free-radical desulfurization to yield the native H3 K14Ac (**Figure 4.5**).

4.2.3 Deacetylation of mononucleosomes by the HDAC1-CoREST dimeric complex

With native, semisynthetic H3 K14Ac and suH4 in hand, we generated octamers containing H3 K14Ac and either wt or SUMOylated H4. We then assembled octamers into mononucleosomes (MN) using 147 bp 601 DNA, reported by Widom and coworkers, which contains a strong nucleosome positioning sequence (**Figure 4.6a**).[75] These MN were subjected to deacetylation in assays containing full-length, recombinant HDAC1 and CoREST, and were stable over the course of the reaction (**Figure 4.6b**). As expected, the kinetics of HDAC1 deacetylation were significantly slower on MN substrates than on free H3 K14Ac (**Figure 4.7**).

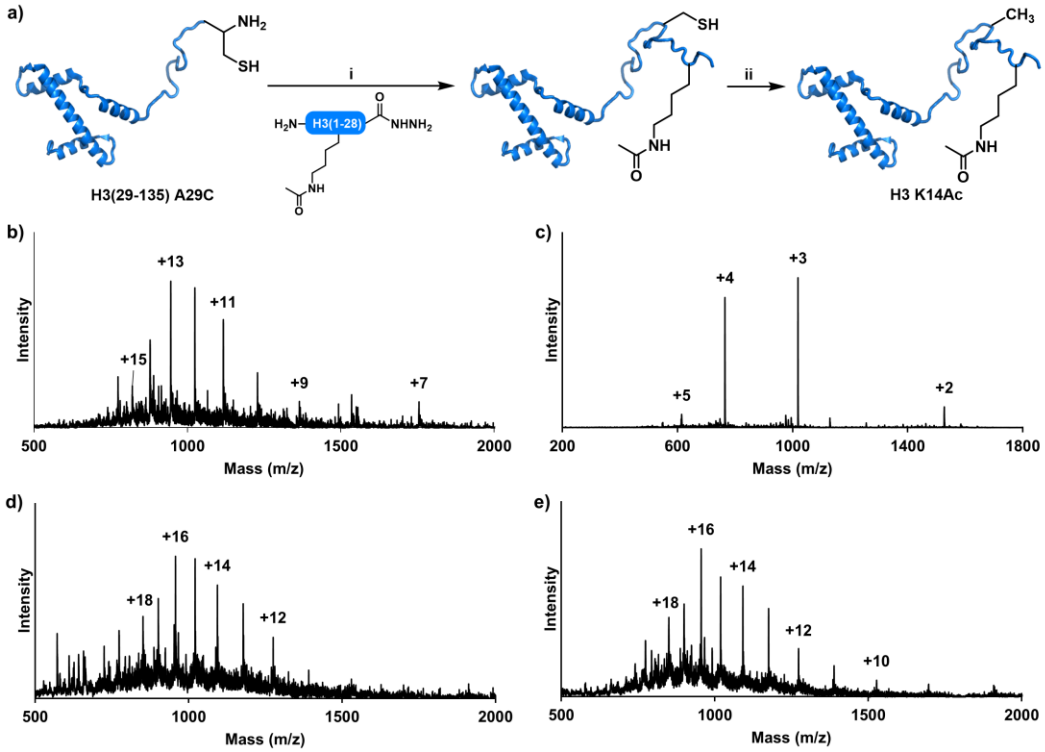


Figure 4.5. Synthesis of H3 K14Ac. a) Schematic of H3 K14 Ac semisynthesis. i) (1) H3(1-28, K14Ac)-C(O)NHNH$_2$, 6 M Gn-HCl, 200 mM Na_2HPO_4, 10 mM $NaNO_2$, pH 3, -20 °C, 15 min, (2) H3(29-135)A29C C110A, 6 M Gn-HCl, 200 mM Na_2HPO_4, 130 mM MPAA, pH 6.9, 25 °C, 24 h. ii) 6 M Gn-HCl, 100 mM Na_2HPO_4, 500 mM TCEP, 100 mM MESNa, 280 mM *t*-BuSH, 10 mM VA-044, 37 °C, 24 h. b) ESI-MS of purified H3(29-135) A29C, C110A. c) ESI-MS of purified H3(1-28, K14Ac)-C(O)NHNH$_2$. d) ESI-MS of purified H3 K14Ac, A29C, C110A. e) ESI-MS of purified H3 K14C, C110A.

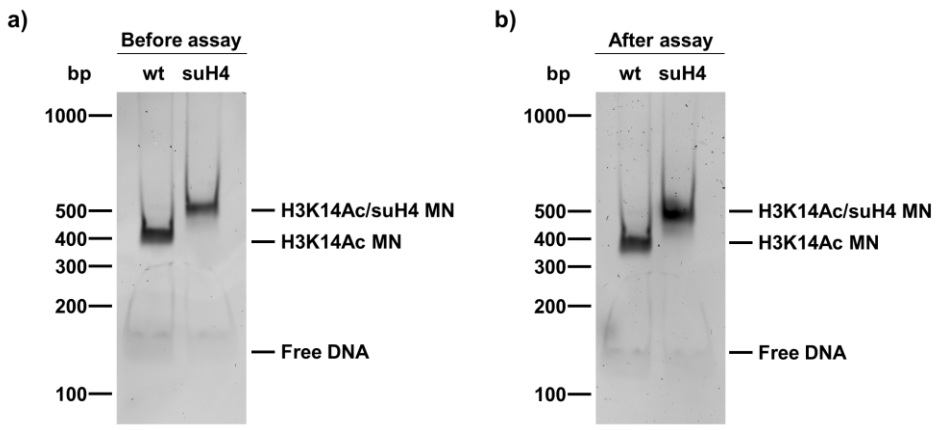


Figure 4.6. Characterization of mononucleosomes used in deacetylation assays and their stability toward assay conditions. a) Ethidium bromide-stained 5% TBE gel of mononucleosomes (MN) reconstituted with H3 K14Ac and either wild-type (wt) or suH4. b) Ethidium bromide-stained 5% TBE gel of MN from (a) after incubation in deacetylation assay buffer at 25 °C for 6 h.

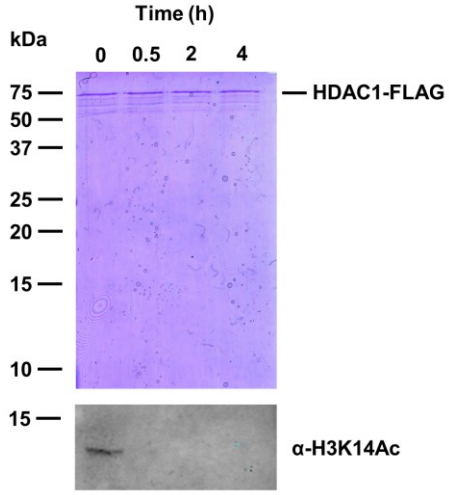


Figure 4.7. Deacetylation of H3 K14Ac by HDAC1. Coomassie-stained 15% SDS-PAGE gel and western blot for H3 K14Ac. 0.5 μM H3 K14Ac was reacted with 0.1 μM HDAC1 under typical mononucleosome deacetylation assay conditions.

We observed that suH4 did not stimulate the activity of HDAC1 alone (**Figure 4.8a,b**). The small initial rate enhancement relative to un-SUMOylated MN at 1 h may be due to the ability of SUMO to limit interactions between MNs,[34] which could facilitate access of HDAC1 to the site of acetylation. The rate of deacetylation of wt MN in the presence of CoREST was also not statistically different than in the same assay with CoREST omitted (**Figure 4.8a,b**). This contrasts with a previous report by Shiekhattar and coworkers, in which deacetylation of nucleosomal H3 K9/K14Ac by HDAC1 was enhanced in the presence of CoREST.[50] Their assay, however, employed nucleosomes purified from HeLa cells, which almost certainly contained other PTMs that may inhibit the enzymatic activity of HDAC1.[56] In fact, the inclusion of LSD1 in their assay resulted in further enhancement of deacetylation. We may expect that, even if CoREST binding does not significantly stimulate HDAC1 activity, it may localize HDAC1 to the MN through its DNA-binding SANT2 domain and thereby enhance deacetylation. We observed this effect previously in H3 K4 demethylation assays containing CoREST and LSD1 at 200 mM NaCl, in which demethylation at 2 h was 2-fold greater with CoREST present.[76] Like LSD1, HDAC1 does not contain a DNA-binding domain. Further, it is thought to act on chromatin *in vivo* only when recruited as part of a complex.[77] However, it was recently reported that recombinant FLAG-HDAC1 can be immunoprecipitated by immobilized MN reconstituted from HeLa-purified octamers and Widom 601 or mouse mammary tumor virus (MMTV) promoter DNA.[78] The immunoprecipitation was performed in buffer containing 150 mM NaCl, which is similar to the 163 mM ionic strength of our deacetylation assays. The same study repeated the pull-down with FLAG-LSD1 and found no binding. Therefore, nonspecific association of HDAC1 with MN may explain the lack of deacetylation rate enhancement by CoREST.

We next included CoREST in deacetylation assays with suH4 containing MN, and observed a ~2-fold enhancement of deacetylation relative to un-SUMOylated MN over the course of 6 h (**Figure 4.8a,c**). The difference was significant, with $p < 0.05$ for each time point. To confirm that the rate enhancement was due to SUMO, we repeated the assay with a CoREST I270A, I272A, and V274A mutant (CoREST3A). These mutations in the SIM of CoREST abolish binding to SUMO-2/3,[41] and are located in the unstructured linker

region between the SANT domains, which is not involved in HDAC1 binding.[50] The mutations similarly do not affect the binding of CoREST3A to DNA.[65] We found no difference in the deacetylation rate between SUMOylated and un-SUMOylated MN with CoREST3A (**Figure 4.8a,d**). This shows that a functional SIM in CoREST is required for stimulation of the HDAC1-CoREST complex towards MN by suH4, and strengthens the case that suH4 may mediate transcriptional repression via a trans effect.

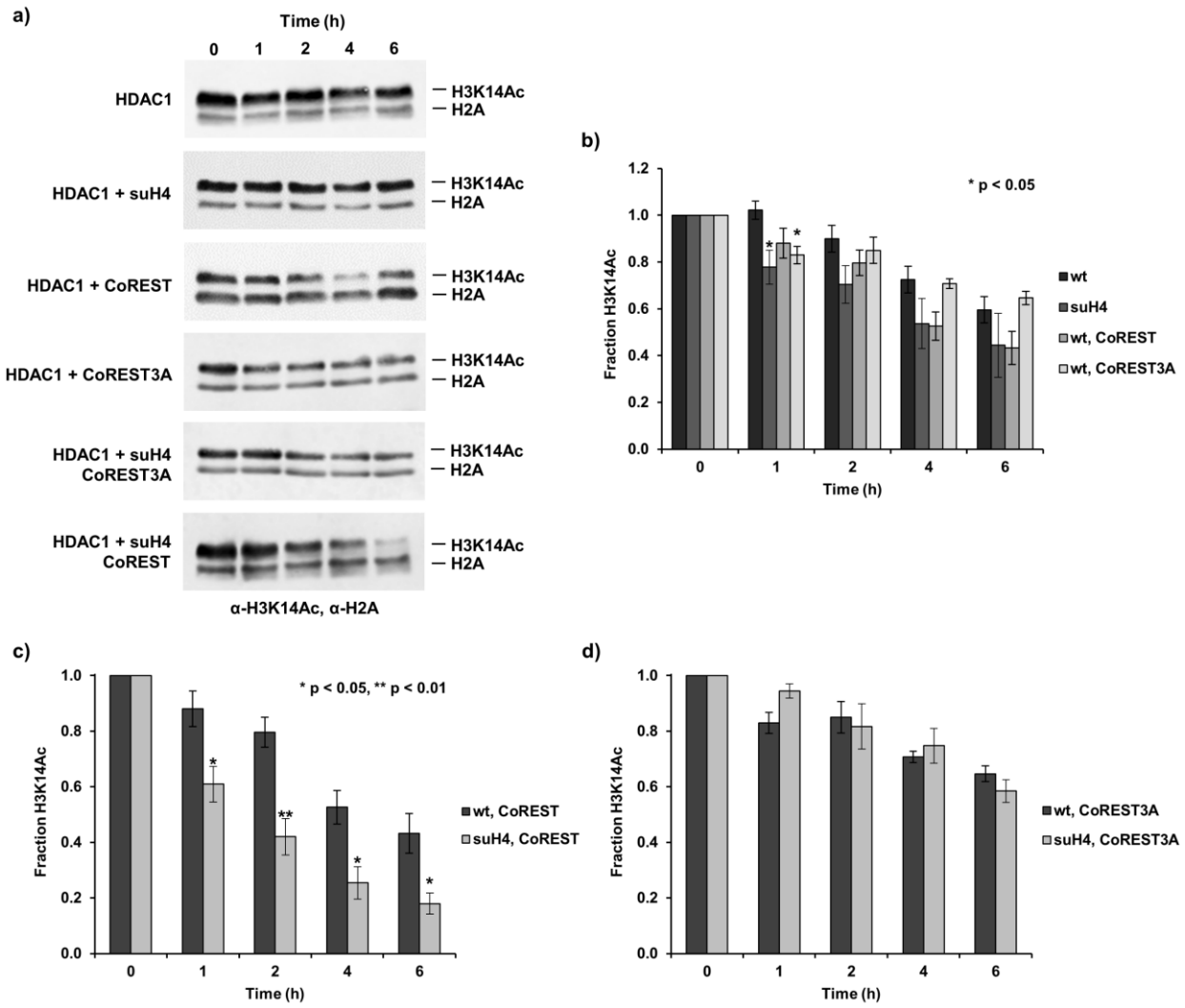


Figure 4.8. Effect of SUMOylated H4 on HDAC1 activity toward mononucleosomes. a) Western blots showing the time-course of mononucleosome (MN) deacetylation with HDAC1 or the HDAC1-CoREST/CoREST3A dimeric complex. H2A signal was a loading control. b) MNs containing H3 K14Ac and suH4 (dark grey) were assayed with HDAC1. MNs containing wt H4 were assayed with HDAC1 (darkest grey), the HDAC1-CoREST complex (light grey), or the HDAC1-CoREST3A complex (lightest grey). c) MNs containing H3 K14Ac and wt H4 (dark grey) or suH4 (light grey) were assayed with the HDAC1-CoREST complex. d) MNs containing H3 K14Ac and wt H4 (dark grey) or suH4 (light grey) were assayed with the HDAC1-CoREST3A complex. In (b), (c), and (d), the signal from a H3 K14Ac-specific antibody in (a) was quantified and normalized to signal from a H2A-specific antibody using NIH ImageJ software. $n \geq 4$, error bars show standard error of the mean. Asterisks denote significance calculated with a student's two-tailed t-test, * $p < 0.05$, ** $p < 0.01$.

4.3 Conclusion and outlook

Even decades after the first report of SUMO,[79] we are still constantly discovering new insights into its function, and uncovering new pathways that it regulates. The conservation of SUMO in all eukaryotes underscores its significance, and due to the diversity of its targets dysregulation of SUMO has been implicated in pathologies ranging from Alzheimer's disease to heart disease.[10,80] Interestingly, the SUMO pathway is often upregulated in cancer, which may be explained by its involvement in DNA damage repair and stress-response pathways.[81]

Studying the effect of site-specific SUMO modification *in vitro* is challenging, because unlike Ub, the small number of SUMO E3 enzymes identified hampers our ability to enzymatically generate a homogenous substrate pool. Hence, we utilized a semisynthetic approach to synthesize suH4, developed in Chapters 2 and 3. To understand the role of suH4 in transcriptional repression, we incorporated suH4 containing octamers into a chromatinized template for *in vitro* transcription. We observed significant repression that may be partially HDAC-dependent, and further experimentation will reveal the altered post-translational modification (PTM) pattern in the presence of suH4. Ongoing experimentation in our lab will determine the ability of the enzyme p300 to acetylate chromatin in the presence of varying degrees of suH4 occupancy, and reveal whether the repressive effect is dependent on steric occlusion of the promoter region.

We also performed *in vitro* deacetylation assays on H3 K14Ac and suH4 containing MN with the HDAC1-CoREST dimeric complex. Similar to our recent work with the LSD1-CoREST complex,[65] we found an approximately 2-fold increase in deacetylation when CoREST contained a functional SIM. We hypothesize that this effect is due to a recruitment mechanism that increases the residence time of HDAC1 near the MN, and thereby increases its local concentration. It will be interesting to also test whether this holds true for other sites of acetylation, such as H4 K16, either on the same or different histone H4 as SUMO, within the same MN. Further, cross-talk with SUMO may not be limited to deacetylation, as the NuRD complex, which has both deacetylase and ATP-dependent nucleosome remodeling activity, contains a SIM in its

MTA1 subunit.[82] Thus, our characterization of the mechanistic relationship between histone SUMOylation and deacetylation will inform future studies, and we continue to actively pursue the potential cross-talk and mechanisms by which suH4 mediates transcriptional repression.

4.4 Experimental procedures

4.4.1 General Methods

Fmoc-Lys(Boc)-Wang resin was purchased from Sigma-Aldrich Chemical Company (St. Louis, MO), and 2-chlorotrityl chloride resin was purchased from AnaSpec (Fremont, CA). Standard Fmoc-L-amino acids were purchased from EMD Millipore (Billerica, MA), AGTC Bioproducts (Wilmington, MA), or AnaSpec. All other chemical reagents were purchased from Sigma-Aldrich or Fisher Scientific (Pittsburgh, PA). DNA synthesis and gene sequencing were performed by Integrated DNA Technologies (Coralville, IA) and Genewiz (South Plainfeld, NJ), respectively. Plasmid mini-prep, PCR purification and gel extraction kits were purchased from Qiagen (Valencia, CA). 147 bp 601 DNA PCR enzymes and reagents were purchased from New England BioLabs (Ipswitch, MA). Chitin beads for purification of intein-CBD fusion proteins were purchased from New England BioLabs. Ni-NTA resin for purification of His_6-tagged proteins was purchased from Thermo Scientific (Waltham, MA). Anti-FLAG M2 affinity gel was purchased from Sigma-Aldrich. Solid phase peptide synthesis (SPPS) was performed on a Liberty Blue Automated Microwave Peptide Synthesizer (CEM Corporation, Matthews, NC). Centrifugal filtration units were from Sartorius (Goettingen, Germany), Slide-A-Lyzer dialysis cassettes were from Pierce (Rockford, IL), and SpectraPor dialysis membrane was from Spectrum Labs (Rancho Dominguez, CA). Analytical reversed-phase HPLC (RP-HPLC) was performed on a Varian (Palo Alto, CA) ProStar HPLC with a Grace-Vydac (Deerfield, IL) C4 or C18 column (5 micron, 150 x 4.6 mm) employing 0.1% TFA in water (A) and 90% CH_3CN, 0.1% TFA in water (B) as the mobile phases. Typical analytical gradients were 0-73% B over 30 min at a flow rate of 1 mL/min. Preparative scale purifications were conducted on a Grace-Vydac C4 or C18 column (10 micron, 250 x 22 mm) at a flow rate of 9 mL/min. Semi-preparative scale purifications were conducted on a Grace-Vydac C4 or C18 column (5 micron, 250 x 10 mm) at a flow rate of 3.5 mL/min. Mass spectrometric analysis

was conducted on a Bruker (Billerica, MA) Esquire ESI-MS instrument. Size-exclusion chromatography was performed on an AKTA FPLC system (GE Healthcare, Little Chalfont, UK) equipped with a P-920 pump and UPC-900 monitor. Mononucleosome gels were visualized using a GE Typhoon FLA 9000 Biomolecular Imager (GE). Blots were visualized using an Odyssey IR Fluorescent Imaging System (LI-COR Biosciences, Lincoln, NE).

4.4.2 Solid phase peptide synthesis

Synthesis of H_2N-H3(1-28, K14Ac)-C(O)$NHNH_2$

The peptide H_2N-ARTKQTARKSTGGK(Ac)APRKQLATKAARKS-C(O)$NHNH_2$ corresponding to the first 28 N-terminal residues of the human histone H3 protein was synthesized by microwave-assisted SPPS on a 0.1 mmol scale employing standard 9-fluorenylmethoxycarbonyl (Fmoc)-based N^{α}-deprotection chemistry. Briefly, 2-chlorotrityl hydrazine resin was prepared by reacting 2-chlorotrityl chloride resin (1.52 mmol/g) in a 10% solution of hydrazine in DMF at 30 °C for 30 min.[83] The reaction was repeated one time with fresh hydrazine solution. The resin was then treated with a 10% methanol in DMF solution for 10 min to cap any unreacted sites on the resin. The first amino acid, Ser, was coupled in 4 molar excess. The coupling reaction containing Fmoc-Ser(*t*-butyl)-OH (0.4 mmol), *O*-(6-Chlorobenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HCTU, 0.38 mmol) , and DIEA (0.8 mmol) proceeded for 60 min at 30 °C. From *t*-butyl-serinyl 2-chlorotrityl hydrazine resin each remaining amino acid was coupled in 5 molar excess based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 20% piperidine in DMF for 3 min at 75 °C. Coupling reactions were undertaken for 5 min at 75 °C with a mixture of Fmoc-amino acid (0.5 mmol), HBTU (0.49 mmol) and DIEA (1.0 mmol) in DMF. For Arg, an additional coupling reaction was performed for 25 min at 75 °C. The Lys at position 14 was coupled as Fmoc-Lys(Ac)-OH. Peptide was cleaved and deprotected by reaction of resin at 20 µL/mg with standard cleavage cocktail (TFA: H_2O: triisopropylsilane: anisole 92.5:2.5:2.5:2.5 v/v) for 1.5 hours at room temperature, then precipitated and washed 2 times with cold diethyl ether. Dry peptide was dissolved in RP-HPLC buffer A and purified by C18 preparative RP-HPLC with a gradient of 0-50% B. This yielded 11% of the peptide H_2N-H3(1-28,

K14Ac)-C(O)NHNH$_2$ based on initial resin loading. ESI-MS of H$_2$N-H3(1-28, K14Ac)-C(O)NHNH$_2$. Calculated *m/z* [M+H]$^+$ 3,052.5 Da, observed 3,054.3 ± 1.5 Da.

Synthesis of BocHN-H4(1-14)-2-chlorotrityl hydrazine resin

The peptide BocHN-SGRGKGGKGLGKGG-C(O)NHNH$_2$ corresponding to the first 14 N-terminal residues of the human histone H4 protein was synthesized by microwave-assisted SPPS on a 0.25 mmol scale employing standard Fmoc-based N$^\alpha$-deprotection chemistry. The 2-chlorotrity hydrazine resin was prepared as above. The first amino acid, Gly, was coupled in 4 molar excess. The coupling reaction containing Fmoc-Gly-OH (1.0 mmol), HCTU (0.95 mmol) , and DIEA (2.0 mmol) proceeded for 60 min at 30 °C. From glycinyl 2-chlorotrityl hydrazine resin each remaining amino acid was coupled in 5 molar excess based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 20% piperidine in DMF for 3 min at 75 °C. Coupling reactions were undertaken for 5 min at 75 °C with a mixture of Fmoc-amino acid (1.31 mmol), HBTU (1.28 mmol) and DIEA (2.75 mmol) in DMF. For Arg, an additional coupling reaction was performed for 25 min at 75 °C. The Lys at position 12 was orthogonally protected with the ivDde protecting group. The peptide was protected at the α-NH$_2$ position with the Boc group by reaction with di-*tert*-butyl dicarbonate (2.0 mmol) and DIEA (4.0 mmol) in DMF for 2 hours.

Attachment of the ligation auxiliary

Deprotection of the ivDde group was achieved by reacting resin bound peptide with a solution of 5% hydrazine in DMF for 5 min. This deprotection was repeated three times. The peptidyl resin was then coupled to bromoacetic acid (8-fold molar excess) with *N,N'*-Diisopropylcarbodiimide (DIC, 8-fold molar excess) in DMF for 45 min at room temperature. The coupling was repeated once. Subsequently, dry peptidyl resin was placed in a solution containing 9 equivalents of auxiliary (0.5 M in DMSO) and shaken for 24 hours at room temperature. The ligation auxiliary *O*-(2-(tritylthio)ethyl)hydroxylamine was prepared over 3 steps from *N*-hydroxyphthalimide as described previously.[67] Completion of the displacement was judged by test cleavage and subsequent ESI-MS analysis. Peptide was cleaved and deprotected by

reaction of resin at 20 µL/mg with Reagent K (TFA: thioanisole: H_2O: phenol: 1,2-ethanedithiol 82.5:5:5:5:2.5 v/v)[84] for 1.5 hours at room temperature, then precipitated and washed 2 times with cold diethyl ether. Dry peptide was dissolved in RP-HPLC buffer A and purified by C18 preparative RP-HPLC with a gradient of 0-50% B. This yielded 9% of the peptide-auxiliary conjugate H4(1-14)^{aux}-C(O)NHNH$_2$ based on initial resin loading. ESI-MS of H4(1-14)^{aux}-C(O)NHNH$_2$. Calculated *m/z* $[M+H]^+$ 1,363.6 Da, observed 1,363.8 Da (**Figure 4.9a**).

Synthesis of 3xFLAG

The peptide H_2N-MDYKDHDGDYKDHDIDYKDDDDK-C(O)OH (3xFLAG) was synthesized by microwave-assisted SPPS on a 0.1 mmol scale employing standard Fmoc-based N^{α}-deprotection chemistry. From Fmoc-Lys(Boc)-Wang resin (0.4-0.6 mmol/g) each remaining amino acid was coupled in 5 molar excess based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 20% piperidine in DMF for 3 min at 75 °C. Coupling reactions were undertaken for 5 min at 75 °C with a mixture of Fmoc-amino acid (0.5 mmol), HBTU (0.49 mmol) and DIEA (1.0 mmol) in DMF. Peptide was cleaved and deprotected by reaction of resin at 20 µL/mg with standard cleavage cocktail (TFA: H_2O: triisopropylsilane: anisole 92.5:2.5:2.5:2.5 v/v) for 1.5 hours at room temperature, then precipitated and washed 2 times with cold diethyl ether. Dry peptide was dissolved in RP-HPLC buffer A and purified by C18 preparative RP-HPLC with a gradient of 0-40% B. This yielded 19% of the peptide 3xFLAG based on initial resin loading. ESI-MS of 3xFLAG. Calculated *m/z* $[M+H]^+$ 2,861.9 Da, observed 2,863.3 ± 3.0 Da (**Figure 4.10**).

4.4.3 Molecular cloning of H3 K14C, C110A and H3(29-135)A29C, C110A

The plasmid pET3a-hH3 C110A, containing the human H3 gene with C110A mutation, was used to generate the plasmid pET3a-hH3 K14C, C110A. The plasmid pET15b-His_6-[TEV]-H3(7-135)A7C, C110A, containing the truncated human histone H3 gene with a Tobacco Etch Virus (TEV) protease cleavage sequence between the His_6 tag and the N-terminus of H3,[65] was used to generate the plasmid pET15b-His_6-[TEV]-H3(29-135)A29C, C110A, which lacks the first 28 residues of histone H3 and bears the mutation

A29C. The modified plasmids were prepared from template by site-directed mutagenesis (QuikChange kit, Agilent Technologies, Santa Clara, CA) with the following primers:

Primer	**DNA Sequence (5'- to -3')**
hH3(29-135)A29C-FP	CCGGCTACCGGCGGCGTGAAAAAG
hH3-RP	GCACTGGAAGTACAGGTTTTCCATATGGCTGCC
hH3-K14C-FP	CAGACGGCTCGGAAATCCACCGGCGGTTGCGCGCCACGCAAGCAGC TGGCTACCAAG
hH3-K14C-RP	CTTGGTAGCCAGCTGCTTGCGTGGCGCGCAACCGCCGGTGGATTTCC GAGCCGTCTG

The desired gene sequences were confirmed by sequencing with the T7 forward primer (Genewiz).

4.4.4 Overexpression and purification of TEV protease

E. coli BL21(DE3) cells containing the plasmid pRK793-His_6-TEV[34] were grown in 1 L LB supplemented with 100 µg/mL of Ampicillin at 37 °C with shaking at 250 rpm until OD_{600} ~0.6. Overexpression was induced by the addition of 0.3 mM IPTG and cells were grown for an additional 6 h at 25 °C. The cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in 15 mL lysis buffer: 20 mM tris, 150 mM NaCl, pH 7.2. Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The lysate supernatant was passed through a 0.45 µm filter then applied to ~5 mL Ni-NTA column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 3 h at 4 °C. The column was then washed thoroughly with lysis buffer containing 200 mM imidazole. His_6-TEV was eluted with lysis buffer containing 500 mM imidazole, and dialyzed into 4 L of 20 mM tris, 150 mM NaCl, 1 mM DTT, pH 7.5 for 18 h at 4 °C.

4.4.5 Overexpression and purification of H3(29-135)A29C, C110A and H4(15-102)A15C

E. coli BL21(DE3) cells containing pET15b-His_6-[TEV]-H4(15-102)A15C or pET15b-His_6-[TEV]-H3(29-135)A29C, C110A were grown in 3 L 2xYT (yeast extract, tryptone) medium supplemented with 100 µg/mL of Ampicillin at 37 °C with shaking at 250 rpm until OD_{600} reached ~0.6. Overexpression was induced by

the addition of 0.3 mM IPTG and cells were grown for an additional 1.5 h at 37 °C. The cells were harvested by centrifugation at 7,000x*g* for 15 min. H4(15-102)A15C and H3(29-135)A29C, C110A were purified using a previously established protocol.[68] Cells were resuspended in wash buffer (20 mM tris, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 7.5, 1% triton X-100) and lysed by sonication on ice. Inclusion bodies were pelleted by centrifugation at 20,000x*g* for 20 min and washed twice with wash buffer. Inclusion bodies were then dissolved in extraction buffer (6 M Gn-HCl, 20 mM tris, 1 mM 2-mercaptoethanol, pH 7.5) and applied to Ni-NTA resin. Column binding proceeded overnight at 4 °C, after which the resin was washed with 10 CV extraction buffer containing 25 mM imidazole. The protein was eluted with 3 x 1 CV extraction buffer containing 400 mM imidazole, then dialyzed into water containing 1 mM DTT. After dialysis, 10x cleavage buffer was added for final concentrations of 50 mM tris, 1 mM EDTA, 10 mM DTT, 10 mM L-cysteine, pH 6.9. Purified TEV protease was added to 20% of the final volume, and the cleavage reaction proceeded overnight at 37 °C. The reaction was then dialyzed back into extraction buffer, incubated overnight at 4 °C with Ni-NTA resin to remove the His_6-tagged TEV protease and cleaved H3 or H4 N-terminal His_6-tag, and the column flow-through containing H3(29-135)A29C, C110A or H4(15-102)A15C purified by C4 preparative RP-HPLC employing a gradient of 30-80% B (H3) or 40-70% B (H4) over 60 min. Typical yields were 3-4 mg/L of cell culture. ESI-MS of H3(29-135)A29C, C110A. Calculated *m/z* $[M+H]^+$ 12,277.7 Da, observed 12,282.3 ± 7.1 Da. ESI-MS of H4(15-102)A15C. Calculated *m/z* $[M+H]^+$ 10,071.8 Da, observed 10,075.3 ± 4.8 Da (**Figure 4.9b**).

4.4.6 Overexpression and purification of SUMO-3(2-91)C47S-MES

E. coli BL21(DE3) cells were transformed with the plasmid pTXB1-SUMO-3(2-91)C47S. Cells were grown in 6 L Luria-Bertani medium supplemented with 100 µg/mL of Ampicillin at 37 °C with shaking at 250 rpm until OD_{600} ~0.6-0.8. Overexpression was induced by the addition of 0.3 mM IPTG and cells were grown for an additional 4 h at 25 °C. The cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in lysis buffer: PBS, pH 7.2 containing 1 mM 2-mercaptoethanesulfonic acid sodium salt (MESNa). Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The lysate supernatant

was passed through a 0.45 μm filter then applied to a 30 mL chitin column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 12 h at 4 °C. The column was then washed with 20 column volumes (CV) of lysis buffer followed by 2 CV of PBS, pH 7.75. SUMO-3(2-91)C47S-MES was cleaved from its intein-CBD fusion by incubation with 1.5 CV of PBS, pH 7.75 containing 100 mM MESNa for 72 h at 4 °C. The eluted α-thioester was purified by C18 preparative RP-HPLC employing a gradient of 30-60% B over 60 min. Fractions containing the desired thioester were identified by ESI-MS. We observed that the N-terminal Met of SUMO-3 is consistently processed *in vivo*, leading to the SUMO-3(2-91)-α-thioester product. Typical yield is 4-5 mg/L of cell culture. ESI-MS for SUMO-3(2-91)C47S-MES. Calculated *m/z* $[M+H]^+$ 10,444.7 Da, observed 10,445.8 ± 3.6 Da.

4.4.7 Expressed protein ligation of H4(1-14)$^{(aux)}$-C(O)NHNH$_2$ and SUMO-3(2-91)C47S-MES

Purified H4(1-14)$^{(aux)}$-C(O)NHNH$_2$ (15.6 mg, 11.6 μmol) and SUMO-3(2-91)C47S-MES (20.1 mg, 1.9 μmol) were dissolved in 7.2 mL of a buffer consisting of 6 M Gn-HCl, 100 mM Na_2HPO_4, and 10 mM TCEP, pH 7.3. Ligation proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C18 preparative RP-HPLC employing a gradient of 25-50% B over 60 min to give 12.6 mg (56%). ESI-MS of H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$. Calculated *m/z* $[M+H]^+$ 11,666.1 Da, observed 11,667.4 ± 3.4 Da (**Figure 4.9c**).

4.4.8 Zn mediated auxiliary removal from H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$

Metallic Zn was freshly activated by stirring in a solution of 5% HCl for 5 min followed by washing with water, ethanol, and diethyl ether, and dried over vacuum. Zn powder (1 g) was added to 9 mL of degassed 6 M Gn-HCl, pH 3 containing 9.3 mg purified H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$. Degassing was accomplished by 3 freeze-thaw cycles under Ar. The reduction proceeded at 37 °C under Ar with gentle shaking for 24 h. The reaction mixture was briefly centrifuged at 13,000 rpm to pellet Zn, and supernatant containing reduced products was removed. The pelleted Zn was washed twice with 0.5 mL of 6 M Gn-HCl, 50 mM EDTA, pH 3. The combined supernatant and washes were purified by C18 semi-preparative RP-HPLC with a gradient

of 25-50% B over 45 min to give 5.5 mg of reduced product (60%). ESI-MS of H4(1-14)$^{Su(C47S)}$-C(O)NHNH$_2$. Calculated *m/z* $[M+H]^+$ 11,589.9 Da, observed 11,591.4 ± 3.9 Da (**Figure 4.9d**).

4.4.9 Expressed protein ligation of H4(1-14)$^{Su(C47S)}$-C(O)NHNH$_2$ and H4(15-102)A15C

Ligation was accomplished by first converting the C-terminal hydrazide of H4(1-14)$^{Su(C47S)}$-C(O)NHNH$_2$ to an acyl azide with $NaNO_2$ via the diazotization reaction, as described previously.[83] Subsequent addition of 4-mercaptophenylacetic acid (MPAA) served to both quench the remaining $NaNO_2$ and generate a highly reactive C-terminal thioester for the ligation reaction. Purified H4(1-14)$^{Su(C47S)}$-C(O)NHNH$_2$ (5.5 mg, 0.474 μmol) was dissolved at 1 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl, pH 3, and kept at -20 °C for a minimum of 20 min. To this solution was added 3.8 μL of a 500 mM solution of $NaNO_2$ in water. The reaction was briefly mixed, then kept at -20 °C for 15 min. Then, a solution of H4(15-102)A15C (14.3 mg, 1.42 μmol) dissolved at 1.25 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl, 200 mM MPAA, pH 6.5, was added to the reaction. The mixture was brought to room temperature, and the pH adjusted with 3 M NaOH to 6.8-7.0. The ligation reaction proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C18 semi-preparative RP-HPLC employing a gradient of 30-70% B over 45 min to give 3.5 mg (28%). ESI-MS of H4(A15C)$^{Su(C47S)}$. Calculated *m/z* $[M+H]^+$ 21,628.7 Da, observed 21,634.0 ± 9.1 Da (**Figure 4.9e**).

4.4.10 Expressed protein ligation of H3(1-28, K14Ac)-C(O)NHNH$_2$ and H3(29-135)A29C, C110A

Ligation was accomplished by first converting the C-terminal hydrazide of H3(1-28, K14Ac)-C(O)NHNH$_2$ to an acyl azide with $NaNO_2$ via the diazotization reaction, as described previously.[83] Purified H3(1-28, K14Ac)-C(O)NHNH$_2$ (6 mg, 1.97 μmol) was dissolved at 5.3 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl, pH 3, and kept at -20 °C for a minimum of 20 min. To this solution was added 84 μL of a 500 mM solution of $NaNO_2$ in water. The reaction was briefly mixed, then kept at -20 °C for 15 min. Then, 400 μL of solution containing 800 mM MPAA, 200 mM Na_2HPO_4, 6 M Gn-HCl, pH 6.5, at -20 °C, was added to the reaction. Immediately following this addition, 422 uL of a solution containing H3(29-135)A29C, C110A (6 mg, 0.49 μmol) dissolved at 1.2 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl, pH 6.5, was added to the reaction. The

mixture was brought to room temperature, and the pH adjusted with 3 M NaOH to 6.8-7.0. The ligation reaction proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C18 semi-preparative RP-HPLC employing a gradient of 35-70% B over 45 min to give 2.2 mg (29%). ESI-MS of H3(1-135, K14Ac)A29C, C110A. Calculated *m/z* $[M+H]^+$ 15,298.8 Da, observed 15,301.7 ± 3.8 Da.

4.4.11 Desulfurization of H3(1-135, K14Ac)A29C, C110A and H4(A15C)$^{Su(C47S)}$

Purified H3(1-135, K14Ac)A29C, C110A or H4(A15C)$^{Su(C47S)}$ was dissolved at 90 μM in 100 mM Na_2HPO_4, 6 M Gn-HCl, 500 mM TCEP, 100 mM MESNa, pH 7.5. To this solution was added 2-methyl-2-propanethiol to a concentration of 280 mM and radical initiator 2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) to a concentration of 10 mM. The reaction was incubated at 37 °C for 24 h, and the product purified by C18 analytical RP-HPLC employing a gradient of 35-70% B (H3) or 30-70% B (H4) over 30 min to give 89% and 71%, respectively. ESI-MS of H3(1-135, K14Ac)C110A (H3 K14Ac). Calculated *m/z* $[M+H]^+$ 15,266.8 Da, observed 15,272.1 ± 7.0 Da. ESI-MS of H4$^{Su(C47S)}$ (suH4). Calculated *m/z* $[M+H]^+$ 21,596.7 Da, observed 21,602.9 ± 5.5 Da (**Figure 4.9f**).

4.4.12 Overexpression and purification of H2A, H2B, H3 C110A, H3 K14C C110A, and H4

Full-length human histone genes in the pET3a vector were a generous gift from Dr. Peter Moyle at the University of Queensland, Australia: hH2A 2-A (*HIST2H2AA3*), hH2B (*HIST1H2BK*), hH3 C110A (*HIST2H3C*), and hH4 (*HIST1H4c*). hH3 K14C C110A was cloned from hH3 C110A vector as described above. *E. coli* BL21(DE3) cells containing a pET3a-histone plasmid were grown in 6 L of 2xYT medium at 37 °C until OD_{600} reached ~0.6-0.8. Protein expression was induced by the addition of 0.3 mM IPTG, and cells were grown for an additional 2.5 h at 37 °C. Cells were harvested by centrifugation at 7,000x*g*, resuspended in 50 mM tris,150 mM NaCl, pH 7.5, and lysed by sonication on ice. The lysate was centrifuged at 20,000xg to separate insoluble inclusion bodies containing histone protein. Histones were extracted with buffer containing 6 M Gn-HCl, 10 mM tris, pH 7.5. The solubilized histones were purified away from other proteins and cellular debris by size exclusion chromatography on a Superdex S-200 column. Fractions

containing the desired histones were identified by 15% SDS-PAGE analysis. Proteins were further purified to homogeneity by RP-HPLC (**Figure 4.11**).

ESI-MS of H2A. Calculated *m/z* $[M+H]^+$ 13,961.2 Da, observed 13,968.0 ± 6.1 Da.

ESI-MS of H2B. Calculated *m/z* $[M+H]^+$ 13,759.9 Da, observed 13,763.1 ± 4.5 Da.

ESI-MS of H3 C110A. Calculated *m/z* $[M+H]^+$ 15,225.7 Da, observed 15,229.2 ± 3.8 Da.

ESI-MS of H3 K14C C110A. Calculated *m/z* $[M+H]^+$ 15,199.7 Da, observed 15,203.0 ± 2.8 Da.

ESI-MS of H4. Calculated *m/z* $[M+H]^+$ 11,237.1 Da, observed 11,239.1 ± 3.4 Da.

4.4.13 Synthesis of H3 K14$_S$Ac

Performed a thiol-ene 'click' reaction between Cys at position 14 of H3 K14C, C110A to mimic an acetyllysine residue. Dissolved H3 K14C, C110A at 0.125 mM in 6 M Gn-HCl, 200 mM NaOAc, 15 mM L-glutathione, 50 mM *N*-vinylacetamide, pH 5. Freeze-thaw degassed thrice under argon atmosphere, then added dimethylsulfide to 100 mM, and radical initiator VA-044 to 50 mM. Incubated at 37 °C for 2.5 h, then purified by C4 analytical RP-HPLC employing a gradient of 15-80% B over 30 min to give H3 K14$_S$Ac in 16% yield. ESI-MS of H3 K14$_S$Ac. Calculated *m/z* $[M+H]^+$ 15,284.8 Da, observed 15,287.3 ± 2.6 Da.

4.4.14 H3 K14$_S$Ac detection by commercial α-H3K14Ac antibodies

Generated hyperacetylated H3 C110A as a positive control for immunodot assays. Dissolved 1 mg of H3 C110A in 450 µL of 1 M NH_4OAc, 8 M urea, 100 mM NH_4HCO_3, pH 8. Added 50 uL acetic anhydride, nutated at 25 °C for 1 h, then purified by C4 analytical RP-HPLC employing a gradient of 35-70% B over 30 min. ESI-MS of H3(xAc) C110A. Observed *m/z* $[M+H]^+$ 15,813, 15,855, and 15,897 Da, corresponding to 14-16 acetylation events (**Figure 4.12**).

Protein samples were dissolved at 51.2 µM in buffer containing 8 M urea, 100 mM Na_2HPO_4, 10 mM tris, pH 8, then 1 µL was spotted onto nitrocellulose membrane (0.2 µm, Bio-Rad) and let dry completely. Membranes were blocked in 5% nonfat dry milk powder in PBST buffer for 1 h at room temperature, then

incubated for 1 h at room temperature in 5% nonfat dry milk powder in PBST buffer containing primary α-H3K14Ac antibody (Abcam 52946, 1:2500 dilution; or Active Motif 11709001, 1:5000 dilution). Membranes were subsequently washed and incubated with goat α-rabbit secondary antibody (LI-COR 926-32211) at 1:15,000 dilutions in PBST buffer containing 4% nonfat dry milk powder for 1 h at room temperature. Membranes were then washed and visualized by IR fluorescence.

4.4.15 Overexpression and purification of CoREST and CoREST3A

The pET28b-His_6-CoREST and pET28b- His_6-CoREST3A plasmids containing the human CoREST1 gene (*RCOR1*) were a kind gift from Dr. Grace Gill.[41] For protein overexpression, *E. coli* BL21(DE3) plasmid-containing cells were grown at 37 °C in 2xYT medium containing 25 μg/mL kanamycin to an OD_{600} of 0.5, then cooled at 16 °C for 1 h. Protein expression was induced by the addition of 0.2 mM IPTG to the growth media and the cells allowed to grow at 16 °C for 6 h. At the end of the induction period, the cells were harvested by centrifugation at 7,000x*g*, resuspended in 50 mM Tris, 500 mM NaCl, 20 mM imizadole, 0.2 mM phenylmethylsulfonyl fluoride (PMSF), 10% glycerol, pH 8, and lysed by sonication on ice. The lysate was the centrifuged at 20,000x*g* and the supernatant passed through a 0.45 μm filter and applied to 10 mL Ni-NTA resin. Lysate was incubated with the column at 4 °C for 1.5 h, after which the column was washed with 10 CV (column volumes) lysis buffer containing 25 mM imidazole and 0.1% Triton X-100. The column was further washed with 5 CV buffer containing 100 mM imidazole. His_6-CoREST or His_6-CoREST3A was eluted with 4 CV buffer containing 250 mM imidazole. Elution fractions were analyzed on a 12% SDS-PAGE gel, and pure fractions dialyzed against 50 mM HEPES, 300 mM NaCl, 10% glycerol, pH 8, for 3 h at 4 °C. After dialysis, samples were concentrated with 30,000 MWCO centrifugal concentrators at 4 °C. Concentration was determined by comparing the band intensity of His_6-CoREST or His_6-CoREST3A on a Coomassie-stained 12% SDS-PAGE gel relative to BSA standards of known concentration (**Figure 4.13**).

4.4.16 Purification of HDAC1

The human HDAC1 gene with C-terminal FLAG tag and downstream IRES and GFP coding sequence was transduced by lentiviral vector into HEK293F suspension cells. After confirming GFP expression by flow

cytometry, cell cultures were expanded and cells harvested. HDAC1 was purified as previously reported, with some modifications.[61] Cells were resuspended at 100 mg/mL in lysis buffer containing 20 mM tris, 500 mM KCl, 5 mM $MgCl_2$, 1 mM PMSF, 0.1% IGEPAL CA-630, 1x protease inhibitor cocktail (Roche), 10% glycerol, pH 7.4. The cell suspension was frozen on dry ice, thawed in a room temperature water bath, and nutated for 30 min at 4 °C. The lysate was clarified by centrifuging at 13,000 rpm for 30 min at 4 °C. The supernatant was applied to a column of 0.4 mL anti-FLAG M2 agarose resin per 1 mL lysate, then the column nutated with lysate at 4 °C for 2 h. The column was then washed with 10 CV lysis buffer, followed by 4 CV elution buffer containing 20 mM tris, 150 mM KCl, 5 mM $MgCl_2$, 1 mM PMSF, 0.1% IGEPAL CA-630, 10% glycerol, pH 7.4. Following wash steps, 3 CV of elution buffer containing 0.8 mg/mL 3xFLAG peptide was added to the column, and the column was nutated at 4 °C for 1.5 h. Eluate was collected from the column, and a 10,000 MWCO centrifugal concentrator used to exchange buffer to 50 mM HEPES, 150 mM KCl, 10% glycerol, pH 8, and finally to concentrate the sample. Western blot with HDAC1-specific primary antibody confirmed identity of the eluted protein. Concentration was determined by comparing the band intensity of HDAC1-FLAG on a Coomassie-stained 12% SDS-PAGE gel relative to BSA standards of known concentration. Activity was confirmed with the Fluor de Lys HDAC fluorometric activity assay (Enzo Life Sciences, Inc, Farmingdale, NY) (**Figure 4.14**).

HDAC1 was purified from eukaryotic cell culture rather than bacterial culture due to previous reports indicating little protein in the soluble fraction, and little enzymatic activity, when purified from *E. coli*.[85,86] However, purification from human cells introduces the risk that other HDAC1-binding proteins, including enzymes, may be co-purified.[86] Extensive and stringent washes were included in the purification, and western blot performed to detect LSD1 and CoREST (**Figure 4.15**). In a blot containing 0.25 µg of HDAC1, only trace signals from LSD1 and CoREST were detected. In assays with mononucleosomes (MNs), however, western blotting for histone H4 revealed deSUMOylation activity over the course of the assay (**Figure 4.16a**). We considered that trace amounts of a deSUMOylating enzyme (SENP) may be present, as HDAC1 is known to interact with SENP1.[87] Reaction of 0.5 µM purified HDAC1 with 1 µM of a SUMO-3

C-terminal 7-amido-4-methylcoumarin derivative (SUMO-3 AMC, Boston Biochem) in MN assay buffer revealed a moderate amount of SENP activity (**Figure 4.16b**). SENP2 catalytic domain (Boston Biochem) was included as a control. Purified HDAC1 was treated with 1 mM *N*-ethylmaleimide (NEM) at room temperature for 5 min, after which NEM was quenched by the addition of 100 mM DTT. NEM and DTT were removed by buffer exchange in a 10,000 MWCO centrifugal concentrator, then reduced SENP activity confirmed by reaction with SUMO-3 AMC and suH4 containing MNs (**Figure 4.16c**). HDAC1 deacetylation activity was re-confirmed by Fluor de Lys assay, and by reaction with suH4 containing MNs.

Western blots (excluding MN assays, described below) were performed with the following conditions. Samples were run on 12% SDS-PAGE gels and proteins subsequently transferred to Immunoblot PVDF membrane (0.2 µm, Bio-Rad) for 90 min at 100 V and 4 °C, in blot buffer (3 g tris, 14.4 g glycine, 100 mL methanol per 1 L). Membranes were blocked in 5% BSA in PBS buffer for 1 h at room temperature, then incubated for 1 h at room temperature in 5% BSA in PBST buffer containing primary antibody. Membranes were washed and incubated with secondary antibody in PBST buffer containing 4% BSA for 1 h at room temperature, then washed and visualized by IR fluorescence.

α-HDAC1 (Active Motif 40967), 1:2000

α-CoREST (EMD Millipore 07-55), 1:4000

α-LSD1 (EMD Millipore ABE365), 1:4000

α-H4 (Abcam 10158), 1:4000

α-rabbit-IRDye CW800 (LI-COR 926-32211), 1:15,000

4.4.17 Histone octamer formation

Histone octamers were assembled as previously reported, excluding reducing agents.[69] Each of the four core histones was dissolved at ~4 mg/mL in an unfolding buffer containing 7 M Gn-HCl, 20 mM tris, pH 7.5. Exact concentration was determined using the 280 nm extinction coefficients: H2A, ε = 4470 $M^{-1}cm^{-1}$; H2B, ε = 7450 $M^{-1}cm^{-1}$; H3 C110A, ε = 4470 $M^{-1}cm^{-1}$; H3 K14Ac, ε = 4470 $M^{-1}cm^{-1}$; H4, ε = 5960 $M^{-1}cm^{-1}$; suH4,

ε = 7450 $M^{-1}cm^{-1}$. Histones were mixed in equimolar amounts and the resulting mixture dialyzed with a 3500 MWCO dialysis cassette into refolding buffer (3 x 1 L) containing 2 M NaCl, 10 mM tris, 1 mM EDTA, pH 7.5. Crude octamers were concentrated with 10,000 MWCO centrifugal concentrators at 4 °C, then purified by size exclusion chromatography on a Superdex S-200 column. Fractions containing pure histone octamers were identified by 15% SDS-PAGE, then combined and concentrated with 10,000 MWCO centrifugal concentrators at 4 °C. Octamer concentration was determined with the 280 nm extinction coefficients: wild-type, ε = 44,700 $M^{-1}cm^{-1}$; suH4, ε = 47,680 $M^{-1}cm^{-1}$; H3K14Ac, ε = 44,700 $M^{-1}cm^{-1}$; H3K14Ac/suH4, ε = 47,680 $M^{-1}cm^{-1}$. Glycerol was added to 10% of the total volume, and samples flash-frozen and stored at -80 °C until use (**Figure 4.17**).

4.4.18 Generation of 147 bp 601 DNA

The 147 bp 601 DNA[75] was amplified by PCR with the following primers:

Primer	**DNA Sequence (5'- to -3')**
1_147_601-FP	CTGGAGAATCCCGGTGCCGAGG
1_147_601-RP	ACAGGATGTATATATCTGACACG
1_147_601 PCR product	CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTC TAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGC CAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATC CTGT

The PCR product was purified using a QIAquick PCR purification kit, eluted in sterile water, and concentrated using a 10,000 MWCO centrifugal concentrator. Concentration of the DNA was calculated with the 260 nm extinction coefficient ε = 2.7845 $uM^{-1}cm^{-1}$.

4.4.19 Mononucleosome assembly

Pure histone octamers and 147 bp 601 DNA were combined in 10 µL of a high-salt refolding buffer consisting of 2 M KCl, 50 mM HEPES, pH 8, to a final concentration of 4.2 µM in each. After incubation at 37 °C for 15 min, samples were transferred to 30 °C, and a volume of 50 mM HEPES, pH 8 dilution buffer was added every 15 min in the following order: 3.3, 6.7, 5, 3.6, 4.7, 6.7, 10, 30, and 20 µL. MN were

incubated at 30 °C for 15 min following the final dilution, then stored on ice until use. MN were analyzed by 5% TBE gel run at 135 V for 30 min and stained with ethidium bromide.

4.4.20 Mononucleosome deacetylation assays

HDAC1 and CoREST were combined at 2.1 μM each in buffer containing 50 mM HEPES, 14 mM KCl, 163 mM NaCl, 0.2 mM DTT, 7% glycerol, pH 8, and incubated on ice for 25 min. HDAC1/CoREST solution was then combined with H3 K14Ac-containing MN for final concentrations of 200 μM MN, and 800 μM each of HDAC1 and CoREST, in reaction buffer composed of 50 mM HEPES, 100 mM KCl, 63 mM NaCl, 0.15 mM DTT, 3% glycerol, pH 8. Reactions were incubated at 25 °C. At each time point, 15 μL was removed and quenched by adding 3 μL of 6x Laemmli dye containing 300 mM DTT and 6 mM sodium butyrate, and boiling for 2 min. Samples were run on 15% SDS-PAGE gels at 200 V for 50 min, then transferred to Immunoblot PVDF membranes (0.2 μm, Bio-Rad) for 16 h at 35 V and 4 °C, in SDS-Towbin buffer (800 mg SDS, 3 g tris, 14.4 g glycine, 100 mL methanol per 1 L). Membranes were blocked in 5% BSA in PBS buffer for 6 h at 4 °C, then incubated overnight at 4 °C in 5% BSA in PBST buffer containing primary α-H2A antibody (Abcam 88770, lot #GR197571-1) at 1:2500 and primary α-H3K14Ac antibody (Abcam 52946, lot #GR149741-17) at 1:1000 dilutions. Membranes were subsequently washed and incubated with goat α-rabbit-IRDye CW800 secondary antibody (LI-COR 926-32211, lot #C60321-05) at 1:15,000 dilutions in PBST buffer containing 4% BSA for 1 h at room temperature. Membranes were then washed and visualized by IR fluorescence (**Figure 4.18**). H3K14Ac signal was normalized to H2A loading control signal, and quantified using NIH ImageJ software.[88]

4.4.21 *In vitro* transcription assays

Recombinant ACF1, ISWI, NAP1 GAL4-VP16, and p300 were purified as described previously.[72] *In vitro* transcription assays were performed as described previously.[70] Template DNA consisted of a pUC18 vector into which a G-less cassette under control of the adenoviral major late core promoter was inserted, with 5

copies of the GAL4 upstream activator sequence. Structural analysis of chromatinized template was performed by micrococcal nuclease digestion.[70]

4.5 Product characterization and supplemental data

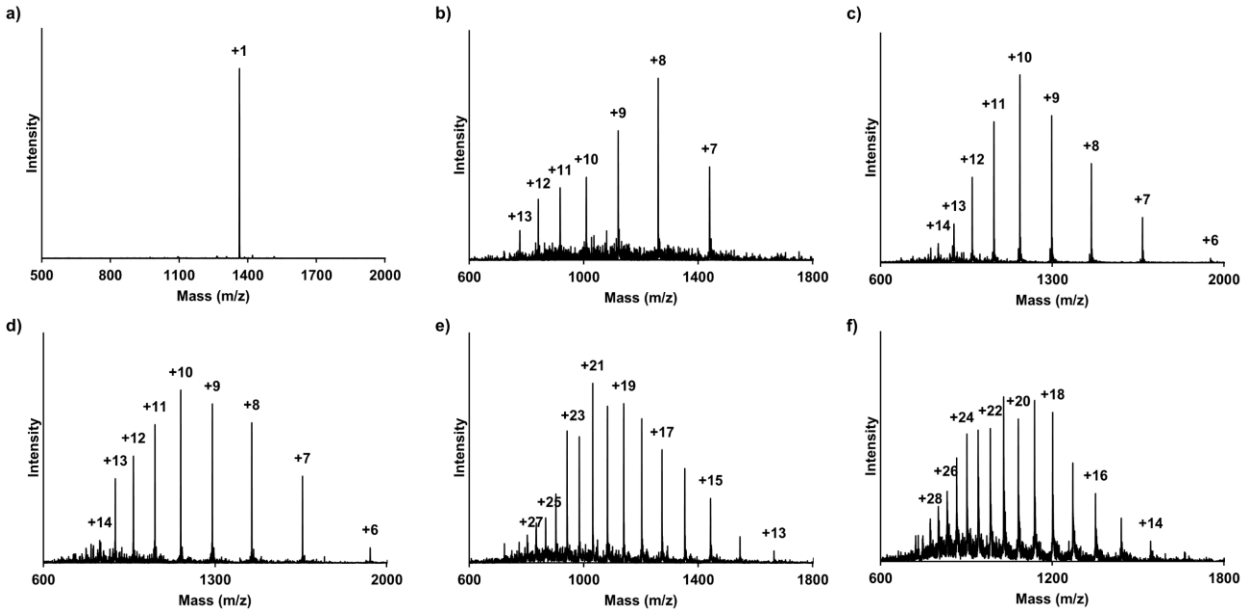


Figure 4.9. Synthesis of suH4. ESI-MS of purified products of each step in suH4 synthesis. a) H4(1-14)$^{(aux)}$-C(O)NHNH$_2$. b) H4(15-102)A15C. c) H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$. d) H4(1-14)$^{Su(C47S)}$-C(O)NHNH$_2$. e) H4(A15C)$^{Su(C47S)}$. f) H4$^{Su(C47S)}$ (suH4).

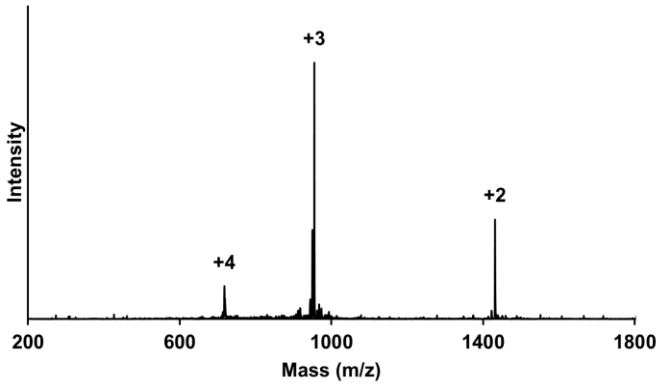


Figure 4.10. Purification of 3xFLAG peptide. ESI-MS of purified 3xFLAG peptide.

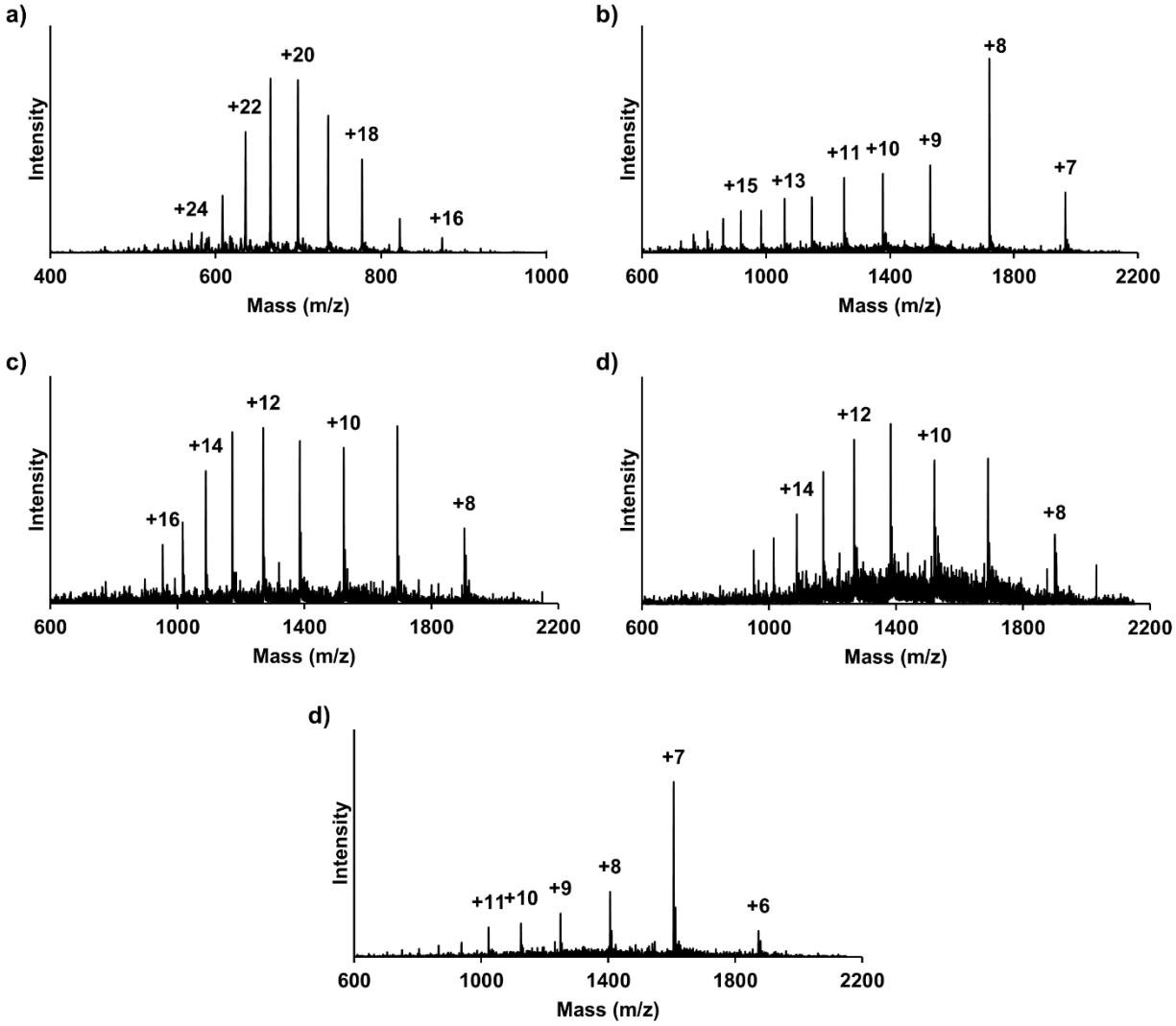


Figure 4.11. Purification of human histones. a) ESI-MS of purified H2A. b) ESI-MS of purified H2B. c) ESI-MS of purified H3 C110A. d) ESI-MS of purified H3 K14C, C110A. e) ESI-MS of purified H4.

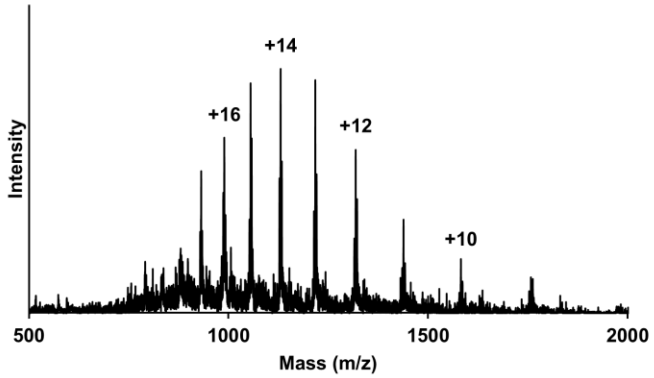


Figure 4.12. Purification of hyperacetylated H3. ESI-MS of purified H3 (xAc) C110A, where x = 14-16 acetylation events.

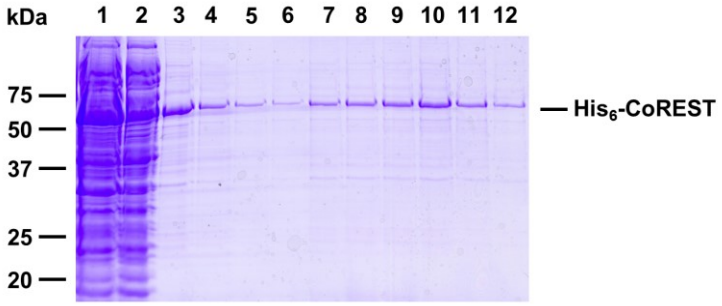


Figure 4.13. Purification of full-length CoREST. Coomassie-stained 12% SDS-PAGE gel of Ni-NTA column fractions from CoREST purification. Lane 1: column flow-through. Lanes 2-3: 50 mM imidazole washes. Lanes 4-6: 100 mM imidazole washes. Lanes 7-12: 250 mM imidazole elution fractions.

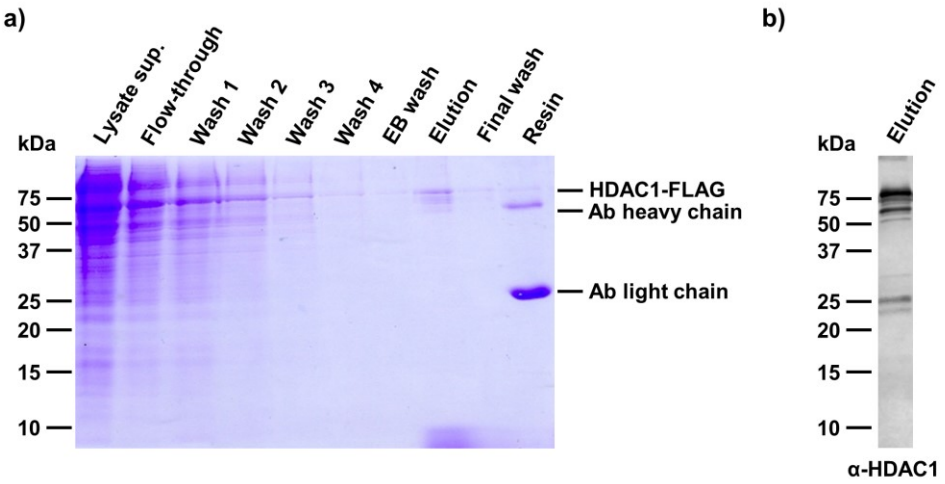


Figure 4.14. Purification of full-length HDAC1. a) Coomassie-stained 12% SDS-PAGE gel of anti-FLAG column fractions. EB = elution buffer. b) Western blot for HDAC1 in the 3xFLAG elution fraction.

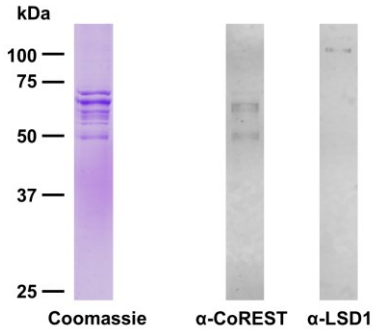


Figure 4.15. Purified HDAC1 contains minimal CoREST and LSD1. Coomassie-stained 12% SDS-PAGE gel (left) of 0.25 µg of purified HDAC1, and western blots of the same sample for CoREST and LSD1.

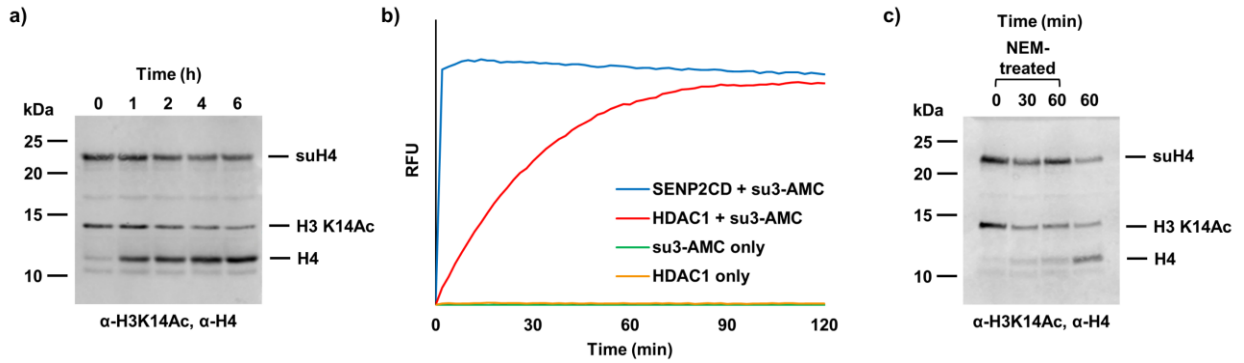


Figure 4.16. SENP activity in purified HDAC1. a) Western blot detecting H3 K14Ac and H4, showing the time course of H3 K14Ac and suH4-containing mononucleosome deacetylation by HDAC1. b) Time course of 7-amido-4-methylcoumarin release from SUMO-3 AMC, monitored by fluorescence at 380/460 nm ex/em, in the presence of no additives (green), HDAC1 (red), or SENP2 catalytic domain (blue). Also monitored HDAC1 with no probe present (yellow). c) Western blot detecting H3 K14Ac and H4, showing the time course of H3 K14Ac and suH4-containing mononucleosome deacetylation by HDAC1 with or without prior *N*-ethylmaleimide (NEM) treatment.

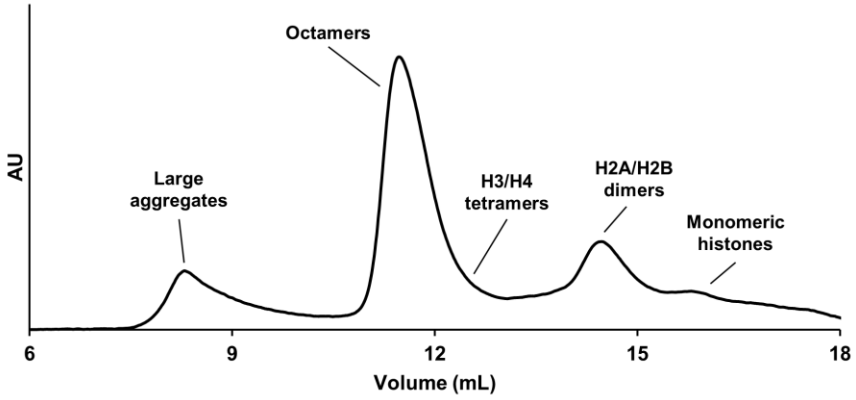


Figure 4.17. Size exclusion chromatogram of histone octamer formation. Typical chromatogram of crude, refolded histone octamer purification by size exclusion chromatography, monitored by absorbance at 214 nm. Histone octamers elute from a 24 mL Superdex S-200 column, run at 0.4 mL/min, at approximately 11 mL.

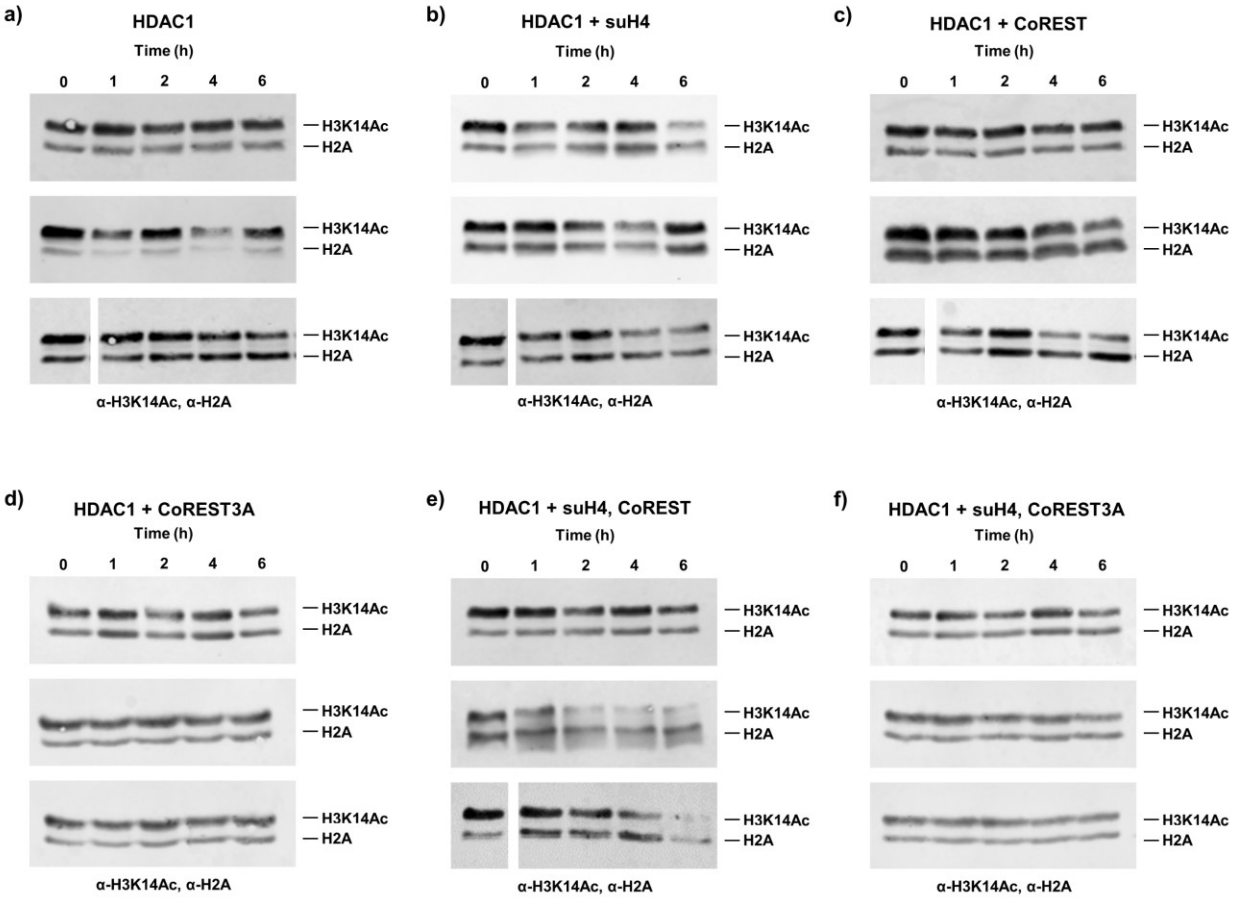


Figure 4.18. Time course of mononucleosome deacetylation by HDAC1. Western blots detecting H3 K14Ac and H2A, showing the time course of H3 K14Ac and wt or suH4-containing mononucleosome deacetylation by HDAC1. H2A was a loading control. a) Deacetylation of MN containing wt H4 by HDAC1. b) Deacetylation of MN containing suH4 by HDAC1. c) Deacetylation of MN containing wt H4 by the HDAC1-CoREST complex. d) Deacetylation of MN containing wt H4 by the HDAC1-CoREST3A complex. e) Deacetylation of MN containing suH4 by the HDAC1-CoREST complex. f) Deacetylation of MN containing suH4 by the HDAC1-CoREST3A complex.

4.6 References

Experiments in this chapter were performed in the lab of Prof. Champak Chatterjee at the University of Washington, except for *in vitro* transcription experiments, which were performed in the lab of Prof. Robert G. Roeder at The Rockefeller University. Miho Shimada performed the *in vitro* transcription experiments. Patrick M. M. Shelton (Chatterjee lab), Elizabeth L. Tyson (Chatterjee lab) and Willem J. de van der Schueren (Fred Hutchinson Cancer Research Center) contributed reagents.

(1) Hendriks, I. A.; D'Souza, R. C. J.; Yang, B.; Verlaan-de Vries, M.; Mann, M.; Vertegaal, A. C. O. Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat. Struct. Mol. Biol.* **2014**, *21* (10), 927–936.

(2) Seeler, J.-S.; Dejean, A. Nuclear and unclear functions of SUMO. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (9), 690–699.

(3) Pichler, A.; Knipscheer, P.; Oberhofer, E.; van Dijk, W. J.; Körner, R.; Olsen, J. V.; Jentsch, S.; Melchior, F.; Sixma, T. K. SUMO modification of the ubiquitin-conjugating enzyme E2-25K. *Nat. Struct. Mol. Biol.* **2005**, *12* (3), 264–269.

(4) Wu, S.-Y.; Chiang, C.-M. Crosstalk between SUMOylation and acetylation regulates p53-dependent chromatin transcription and DNA binding. *EMBO J.* **2009**, *28* (9), 1246–1259.

(5) Krumova, P.; Meulmeester, E.; Garrido, M.; Tirard, M.; Hsiao, H.-H.; Bossis, G.; Urlaub, H.; Zweckstetter, M.; Kugler, S.; Melchior, F.; et al. Sumoylation inhibits α-synuclein aggregation and toxicity. *J. Cell Biol.* **2011**, *194* (1), 49–60.

(6) Matunis, M. J.; Zhang, X.-D.; Ellis, N. A. SUMO: The glue that binds. *Dev. Cell* **2006**, *11*, 596–597.

(7) Gareau, J. R.; Lima, C. D. The SUMO pathway: Emerging mechanisms that shape specificity, conjugation and recognition. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (12), 861–871.

(8) Sampson, D. A.; Wang, M.; Matunis, M. J. The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *J. Biol. Chem.* **2001**, *276* (24), 21664–21669.

(9) Lin, D.; Tatham, M. H.; Yu, B.; Kim, S.; Hay, R. T.; Chen, Y. Identification of a substrate recognition site on Ubc9. *J. Biol. Chem.* **2002**, *277* (24), 21740–21748.

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

(11) Zhao, X.; Sternsdorf, T.; Bolger, T. A.; Evans, R. M.; Yao, T.-P. Regulation of MEF2 by histone deacetylase 4- and SIRT1 deacetylase-mediated lysine modifications. *Mol. Cell. Biol.* **2005**, *25* (19), 8456–8464.

(12) Yamashita, D.; Moriuchi, T.; Osumi, T.; Hirose, F. Transcription factor hDREF is a novel SUMO E3

ligase of Mi2α. *J. Biol. Chem.* **2016**, *291* (22), 11619–11634.

(13) Kumar, A.; Zhang, K. Y. J. Advances in the development of SUMO specific protease (SENP) inhibitors. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 204–211.

(14) Owerbach, D.; McKay, E. M.; Yeh, E. T. H.; Gabbay, K. H.; Bohren, K. M. A proline-90 residue unique to SUMO-4 prevents maturation and SUMOylation. *Biochem. Biophys. Res. Commun.* **2005**, *337* (2), 517–520.

(15) Vertegaal, A. C. O. SUMO chains: Polymeric signals. *Biochem. Soc. Trans.* **2010**, *38* (1), 46–49.

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

(17) Saitoh, H.; Hinchey, J. Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* **2000**, *275* (9), 6252–6258.

(18) Keusekotten, K.; Bade, V. N.; Meyer-Teschendorf, K.; Sriramachandran, A. M.; Fischer-Schrader, K.; Krause, A.; Horst, C.; Schwarz, G.; Hofmann, K.; Dohmen, R. J.; et al. Multivalent interactions of the SUMO-interaction motifs in RING finger protein 4 determine the specificity for chains of the SUMO. *Biochem. J.* **2014**, *457* (1), 207–214.

(19) Kerscher, O. SUMO junction - what's your function? New insights through SUMO-interacting motifs. *EMBO Rep.* **2007**, *8* (6), 550–555.

(20) Hecker, C.-M.; Rabiller, M.; Haglund, K.; Bayer, P.; Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* **2006**, *281* (23), 16117–16127.

(21) Sekiyama, N.; Ikegami, T.; Yamane, T.; Ikeguchi, M.; Uchimura, Y.; Baba, D.; Ariyoshi, M.; Tochio, H.; Saitoh, H.; Shirakawa, M. Structure of the small ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing chromatin-associated factor 1 bound to SUMO-3. *J. Biol. Chem.* **2008**, *283* (51), 35966–35975.

(22) Shiio, Y.; Eisenman, R. N. Histone SUMOylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230.

(23) Nathan, D.; Ingvarsdottir, K.; Sterner, D. E.; Bylebyl, G. R.; Dokmanovic, M.; Dorsey, J. A.; Whelan, K. A.; Krsmanovic, M.; Lane, W. S.; Meluh, P. B.; et al. Histone SUMOylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes Dev.* **2006**, *20*, 966–976.

(24) Galisson, F.; Mahrouche, L.; Courcelles, M.; Bonneil, E.; Meloche, S.; Chelbi-Alix, M. K.; Thibault, P. A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol. Cell. proteomics* **2011**, *10*, M110.004796.

(25) Becker, J.; Barysch, S. V; Karaca, S.; Dittner, C.; Hsiao, H.-H.; Berriel Diaz, M.; Herzig, S.; Urlaub, H.; Melchior, F. Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat. Struct. Mol. Biol.* **2013**, *20* (4), 525–531.

(26) Rosonina, E.; Duncan, S. M.; Manley, J. L. SUMO functions in constitutive transcription and during activation of inducible genes in yeast. *Genes Dev.* **2010**, *24*, 1242–1252.

(27) Neyret-Kahn, H.; Benhamed, M.; Ye, T.; Le Gras, S.; Cossec, J. C.; Lapaquette, P.; Bischof, O.; Ouspenskaia, M.; Dasso, M.; Seeler, J.-S.; et al. SUMOylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. *Genome Res.* **2013**, *23* (10), 1563–1579.

(28) Seifert, A.; Schofield, P.; Barton, G. J.; Hay, R. T. Proteotoxic stress reprograms the chromatin landscape of SUMO modification. *Sci. Signal.* **2015**, *8* (384), rs7.

(29) Niskanen, E. A.; Malinen, M.; Sutinen, P.; Toropainen, S.; Paakinaho, V.; Vihervaara, A.; Joutsen, J.; Kaikkonen, M. U.; Sistonen, L.; Palvimo, J. J. Global SUMOylation on active chromatin is an acute heat stress response restricting transcription. *Genome Biol.* **2015**, *16*, 153.

(30) Lang, G.; Bonnet, J.; Umlauf, D.; Karmodiya, K.; Koffler, J.; Stierle, M.; Devys, D.; Tora, L. The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements. *Mol. Cell. Biol.* **2011**, *31* (18), 3734–3744.

(31) Kalb, R.; Latwiel, S.; Baymaz, H. I.; Jansen, P. W. T. C.; Müller, C. W.; Vermeulen, M.; Müller, J. Histone H2A monoubiquitination promotes histone H3 methylation in polycomb repression. *Nat. Struct. Mol. Biol.* **2014**, *21* (6), 569–571.

(32) Jason, L. J. M.; Moore, S. C.; Ausió, J.; Lindsey, G. Magnesium-dependent association and folding of oligonucleosomes reconstituted with ubiquitinated H2A. *J. Biol. Chem.* **2001**, *276* (18), 14597–14601.

(33) Fierz, B.; Chatterjee, C.; McGinty, R. K.; Bar-Dagan, M.; Raleigh, D. P.; Muir, T. W. Histone H2B ubiquitylation disrupts local and higher order chromatin compaction. *Nat. Chem. Biol.* **2011**, *7* (2), 113–119.

(34) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. SUMOylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837.

(35) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008**, *453* (7196), 812–816.

(36) Chupreta, S.; Holmstrom, S.; Subramanian, L.; Iñiguez-Lluhí, J. A. A small conserved surface in SUMO is the critical structural determinant of its transcriptional inhibitory properties. *Mol. Cell. Biol.* **2005**, *25* (10), 4272–4282.

(37) Ullmann, R.; Chien, C. D.; Avantaggiati, M. L.; Muller, S. An acetylation switch regulates SUMO-dependent protein interaction networks. *Mol. Cell* **2012**, *46* (6), 759–770.

(38) Rouvière, J. O.; Geoffroy, M. C.; Palancade, B. Multiple crosstalks between mRNA biogenesis and SUMO. *Chromosoma* **2013**, *122* (5), 387–399.

(39) Garcia-Dominguez, M.; Reyes, J. C. SUMO association with repressor complexes, emerging routes for transcriptional control. *Biochim. Biophys. Acta* **2009**, *1789*, 451–459.

(40) Witty, J.; Aguilar-Martinez, E.; Sharrocks, A. D. SENP1 participates in the dynamic regulation of Elk-1 SUMOylation. *Biochem. J.* **2010**, *428* (2), 247–254.

(41) Ouyang, J.; Shi, Y.; Valin, A.; Xuan, Y.; Gill, G. Direct binding of CoREST1 to SUMO-2/3 contributes

to gene-specific repression by the LSD1/CoREST1/HDAC complex. *Mol. Cell* **2009**, *34* (2), 145–154.

(42) Andrés, M. E.; Burger, C.; Peral-Rubio, M. J.; Battaglioli, E.; Anderson, M. E.; Grimes, J.; Dallman, J.; Ballas, N.; Mandel, G. CoREST: A functional corepressor required for regulation of neural-specific gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9873–9878.

(43) Ooi, L.; Wood, I. C. Chromatin crosstalk in development and disease: Lessons from REST. *Nat. Rev. Genet.* **2007**, *8*, 544–554.

(44) Lunyak, V. V.; Burgess, R.; Prefontaine, G. G.; Nelson, C.; Sze, S.-H.; Chenoweth, J.; Schwartz, P.; Mandel, G.; Rosenfeld, M. C. Corepressor-Dependent Silencing of Chromosomal Regions Encoding Neuronal Genes. *Science.* **2002**, *298*, 1747–1752.

(45) You, A.; Tong, J. K.; Grozinger, C. M.; Schreiber, S. L. CoREST is an integral component of the CoREST-human histone deacetylase complex. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (4), 1454–1458.

(46) Gocke, C. B.; Yu, H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One* **2008**, *3* (9), e3255.

(47) Lakowski, B.; Roelens, I.; Jacob, S. CoREST-like complexes regulate chromatin modification and neuronal gene expression. *J. Mol. Neurosci.* **2006**, *29*, 227–239.

(48) Hakimi, M.-A.; Bochar, D. A.; Chenoweth, J.; Lane, W. S.; Mandel, G.; Shiekhattar, R. A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (11), 7420–7425.

(49) Shi, Y.-J.; Matson, C.; Lan, F.; Iwase, S.; Baba, T.; Shi, Y. Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell* **2005**, *19* (6), 857–864.

(50) Lee, M. G.; Wynder, C.; Bochar, D. A.; Hakimi, M.-A.; Cooch, N.; Shiekhattar, R. Functional interplay between histone demethylase and deacetylase enzymes. *Mol. Cell. Biol.* **2006**, *26* (17), 6395–6402.

(51) Yang, M.; Gocke, C. B.; Luo, X.; Borek, D.; Tomchick, D. R.; Machius, M.; Otwinowski, Z.; Yu, H. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol. Cell* **2006**, *23* (3), 377–387.

(52) Millard, C. J.; Watson, P. J.; Celardo, I.; Gordiyenko, Y.; Cowley, S. M.; Robinson, C. V.; Fairall, L.; Schwabe, J. W. R. Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol. Cell* **2013**, *51*, 57–67.

(53) Hwang, S.; Schmitt, A. A.; Luteran, A. E.; Toone, E. J.; McCafferty, D. G. Thermodynamic characterization of the binding interaction between the histone demethylase LSD1/KDM1 and CoREST. *Biochemistry* **2011**, *50* (4), 546–557.

(54) de Ruijter, A. J. M.; van Gennip, A. H.; Caron, H. N.; Kemp, S.; van Kuilenburg, A. B. P. Histone deacetylases (HDACs): Characterization of the classical HDAC family. *Biochem. J.* **2003**, *370*, 737–749.

(55) Dokmanovic, M.; Clarke, C.; Marks, P. A. Histone deacetylase inhibitors: Overview and perspectives. *Mol. cancer Res.* **2007**, *5* (10), 981–989.

(56) Riester, D.; Hildmann, C.; Grünewald, S.; Beckers, T.; Schwienhorst, A. Factors affecting the substrate specificity of histone deacetylases. *Biochem. Biophys. Res. Commun.* **2007**, *357*, 439–445.

(57) Millard, C. J.; Watson, P. J.; Fairall, L.; Schwabe, J. W. R. Targeting class I histone deacetylases in a “complex” environment. *Trends Pharmacol. Sci.* **2017**, *38* (4), 363–377.

(58) Seto, E.; Yoshida, M. Erasers of histone acetylation: The histone deacetylase enzymes. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a018713.

(59) Lombardi, P. M.; Cole, K. E.; Dowling, D. P.; Christianson, D. W. Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. *Curr. Opin. Struct. Biol.* **2011**, *21*, 735–743.

(60) Segre, C. V.; Chiocca, S. Regulating the regulators: The post-translational code of class I HDAC1 and HDAC2. *J. Biomed. Biotechnol.* **2011**, Article ID 690848.

(61) Qiu, Y.; Zhao, Y.; Becker, M.; John, S.; Parekh, B. S.; Huang, S.; Hendarwanto, A.; Martinez, E. D.; Chen, Y.; Lu, H.; et al. HDAC1 acetylation is linked to progressive modulation of steroid receptor-induced gene transcription. *Mol. Cell* **2006**, *22* (5), 669–679.

(62) Luo, Y.; Jian, W.; Stavreva, D. A.; Fu, X.; Hager, G.; Bungert, J.; Huang, S.; Qiu, Y. Trans-regulation of histone deacetylase activities through acetylation. *J. Biol. Chem.* **2009**, *284* (50), 34901–34910.

(63) Senese, S.; Zaragoza, K.; Minardi, S.; Muradore, I.; Ronzoni, S.; Passafaro, A.; Bernard, L.; Draetta, G. F.; Alcalay, M.; Seiser, C.; et al. Role for histone deacetylase 1 in human tumor cell proliferation. *Mol. Cell. Biol.* **2007**, *27* (13), 4784–4795.

(64) Delcuve, G. P.; Khan, D. H.; Davie, J. R. Roles of histone deacetylases in epigenetic regulation: Emerging paradigms from studies with inhibitors. *Clin. Epigenetics* **2012**, *4*, 5.

(65) Dhall, A.; Weller, C. E.; Shelton, P. M. M.; Chu, A.; Chatterjee, C. Chemically SUMOylated histone H4 stimulates intranucleosomal demethylation by the LSD1- CoREST complex. **2017**, Manuscript under review at ACS Chemical Biology.

(66) Canne, L. E.; Bark, S. J.; Kent, S. B. H. Extending the applicability of native chemical ligation. *J. Am. Chem. Soc.* **1996**, *118*, 5891–5896.

(67) Weller, C. E.; Huang, W.; Chatterjee, C. Facile synthesis of native and protease-resistant ubiquitylated peptides. *ChemBioChem* **2014**, *15* (9), 1263–1267.

(68) Weller, C. E.; Dhall, A.; Ding, F.; Linares, E.; Whedon, S. D.; Senger, N. A.; Tyson, E. L.; Bagert, J. D.; Li, X.; Augusto, O.; et al. Aromatic thiol-mediated cleavage of N–O bonds enables chemical ubiquitylation of folded proteins. *Nat. Commun.* **2016**, *7*, 12979.

(69) Luger, K.; Rechsteiner, T. J.; Flaus, A. J.; Waye, M. M.; Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **1997**, *272* (3), 301–311.

(70) An, W.; Roeder, R. G. Reconstitution and Transcriptional Analysis of Chromatin In Vitro. *Methods Enzymol.* **2004**, *377*, 460–474.

(71) Batta, K.; Zhang, Z.; Yen, K.; Goffman, D. B.; Franklin Pugh, B. Genome-wide function of H2B ubiquitylation in promoter and genic regions. *Genes Dev.* **2011**, *25* (21), 2254–2265.

(72) An, W.; Palhan, V. B.; Karymov, M. A.; Leuba, S. H.; Roeder, R. G. Selective requirements for histone H3 and H4 N termini in p300-dependent transcriptional activation from chromatin. *Mol. Cell* **2002**, *9*, 811–821.

(73) Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiold, L.; Liu, C.-F. A direct method for site-specific protein acetylation. *Angew. Chemie* **2011**, *50* (41), 9611–9614.

(74) Rosaleny, L. E.; Ruiz-García, A. B.; García-Martínez, J.; Pérez-Ortín, J. E.; Tordera, V. The Sas3p and Gcn5p histone acetyltransferases are recruited to similar genes. *Genome Biol.* **2007**, *8*, R119.

(75) Lowary, P. T.; Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **1998**, *276*, 19–42.

(76) Dhall, A.; Shelton, P. M. M.; Chu, A.; Marie-France, D. A.; Fierz, B.; Chatterjee, C. Untitled manuscript. **2017**, In preparation.

(77) Kelly, R. D. W.; Cowley, S. M. The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochem. Soc. Trans.* **2013**, *41*, 741–749.

(78) Li, X.; Yang, H.; Huang, S.; Qiu, Y. Histone deacetylase 1 and p300 can directly associate with chromatin and compete for binding in a mutually exclusive manner. *PLoS One* **2014**, *9* (4), e94523.

(79) Matunis, M. J.; Coutavas, E.; Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* **1996**, *135* (6), 1457–1470.

(80) Lee, L.; Dale, E.; Staniszewski, A.; Zhang, H.; Saeed, F.; Sakurai, M.; Fa', M.; Orozco, I.; Michelassi, F.; Akpan, N.; et al. Regulation of synaptic plasticity and cognition by SUMO in normal physiology and Alzheimer's disease. *Sci. Rep.* **2014**, *4*, 7190.

(81) Seeler, J.-S.; Dejean, A. SUMO and the robustness of cancer. *Nat. Rev. Cancer* **2017**, *17* (3), 184–197.

(82) Cong, L.; Pakala, S. B.; Ohshiro, K.; Li, D.-Q.; Kumar, R. SUMOylation and SUMO-interacting motif (SIM) of metastasis tumor antigen 1 (MTA1) synergistically regulate its transcriptional repressor function. *J. Biol. Chem.* **2011**, *286* (51), 43793–43808.

(83) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **2013**, *8* (12), 2483–2495.

(84) King, D. S.; Fields, C. G.; Fields, G. B. A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis. *Int. J. Pept. Protein Res.* **1990**, *36*, 255–266.

(85) Saito, M.; Ishikawa, F. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem.* **2002**, *277* (38), 35434–35439.

(86) Li, J.; Staver, M. J.; Curtin, M. L.; Holms, J. H.; Frey, R. R.; Edalji, R.; Smith, R.; Michaelides, M. R.;

Davidsen, S. K.; Glaser, K. B. Expression and functional characterization of recombinant human HDAC1 and HDAC3. *Life Sci.* **2004**, *74* (22), 2693–2705.

(87) Cheng, J.; Wang, D.; Wang, Z.; Yeh, E. T. H. SENP1 enhances androgen receptor-dependent transcription through deSUMOylation of histone deacetylase 1. *Mol. Cell. Biol.* **2004**, *24* (13), 6021–6028.

(88) W. S. Rasband. ImageJ, U. S. National Institutes of Health. Bethesda, Maryland, USA. https://imagej.nih.gov/ij/, 1997-2017.