Chemical Approaches to Investigate the Biochemical Crosstalk Between Histone Sumoylation, Methylation and Acetylation

Patrick M. M. Shelton

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Champak Chatterjee, Chair

Michael H. Gelb

Dustin J. Maly

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Chemical Approaches to Investigate the Biochemical Crosstalk Between Histone Sumoylation, Methylation and Acetylation

Patrick M. M. Shelton

Chair of the Supervisory Committee:

Dr. Champak Chatterjee

Chemistry

Histone post-translational modifications (PTM) within chromatin control many DNA-templated cellular processes, from DNA damage repair pathways to gene transcription. In order to understand the influence of PTM on these important processes and their associated regulatory enzymes, my graduate work in the Chatterjee research group at the University of Washington has focused on developing chemical techniques for the synthesis of full-length and site-specifically modified proteins. The biophysical and biochemical impacts of histone modification by the small ubiquitin-like modifier (SUMO) family of proteins in particular are poorly understood. Early studies associated histone H4 sumoylation with gene repression, but offered little insight into the precise mechanisms underlying this repression. Therefore, this thesis focuses on investigating mechanisms by which SUMO may control gene function. Toward this goal, I have synthesized numerous methylated and acetylated N-terminal histone peptides and full-length histone proteins to assess how sumoylation modulates the activity of lysine specific

demethylase 1 (LSD1) and histone deacetylase 1 (HDAC1) complexes associated with gene silencing.

The Chatterjee lab has previously developed a new approach for the semisynthesis of site-specifically sumoylated histone H4 (suH4) by using a 2-(aminooxy)ethanethiol-mediated expressed protein ligation strategy. During the development of the 2-(aminooxy)ethanethiol ligation method, it was serendipitously discovered that ligation products still bearing the ligation handle undergo reverse native chemical ligation (NCL) in thiol-containing buffers, resulting in isolable C-terminal α-thioesters. I have harnessed this observation in the development of a novel, robust and multifaceted technique using the C-terminal mercaptoethoxyglycinamide (MEGA) handle to synthesize valuable C-terminal peptide α-thioesters and cyclic peptides (Chapter 2). Optimized conditions allow for peptide thioesterification using most amino acids at the C-terminal position under mild conditions, and I have demonstrated compatibility with a wide range of peptide lengths. The MEGA approach is also amenable to one-pot NCL reactions by introducing a N-terminal cysteine containing peptide, and is ideally suited for the synthesis of cyclic peptides via intramolecular NCL.

Semisynthetic strategies developed by our lab have allowed assembly of simultaneously methylated, acetylated and sumoylated mononucleosome substrates for evaluating the effect of suH4 on chromatin-modifying enzymatic activity. Specifically, I have harnessed these new methodologies to demonstrate that suH4 stimulates the activity of the transcriptionally repressive enzymes LSD1 and HDAC1 when complexed with the scaffolding protein, co-repressor of RE1 silencing transcriptional factor 1 (CoREST1). I have demonstrated that suH4 stimulates the enzymatic activity of the HDAC1-CoREST1 sub-complex towards both mono- and polyacetylated nucleosomes relative to nucleosome containing wild-type H4 (Chapter 3). This stimulation was dependent on the interaction between SUMO and the intact non-

consensus-SUMO interacting motif (ncSIM) within CoREST1. Given the importance of the CoREST1-ncSIM interaction, we have further employed 2D-NMR strategies to map the important interacting residues in SUMO and quantify the binding interaction for the first time. These interactions provide new therapeutic targets in diseases arising from the misregulation of gene function.

I have also sought to more completely characterize the substrate specificity of LSD1 and to explore the effect of suH4 on LSD1-CoREST1 sub-complex activity (Chapter 4). My efforts have demonstrated that the acetylation state in H3 tail peptides finely tunes LSD1 activity towards its primary substrate, methylated H3K4 (H3K4me1/2). To corroborate these findings in the context of mononucleosomes and to assess LSD1 kinetic changes imparted by suH4 I have developed a high-throughput 384-well microplate-based demethylation assay platform. This set-up allows conservation of valuable semisynthetic materials and facilitates data collection. Using the microplate platform, I have demonstrated steady-state kinetics of LSD1-CoREST1 toward H3K4me2 and suH4 containing mononucleosomes for the first time. We hope to further delineate subtle contextual effects on the removal of H3K4 methylation by the LSD1-CoREST1 sub-complex.

Acknowledgements

This accomplishment would not have been possible if not for the support from a long list of friends, family, colleagues and mentors. First and foremost, I owe my success at the University of Washington to Dr. Champak Chatterjee. After offering me the opportunity to pursue my PhD in his lab, he has relentlessly challenged me to think more deeply and critically about my projects. Champak has helped to evolve my work ethic and love for experimental science into a philosophy on how to approach challenging chemical and biological questions. He has taught me to never accept the first explanation given and to always intellectually push myself. Most importantly, he has taught me the importance of showing up, staying positive and believing in myself.

To my lab mates and colleagues in the UW Chemistry graduate program: You have continually been a source of inspiration and motivation. When life got hard, you were there to empathize. When experiments were not going well, you were there to help troubleshoot. When a major hurdle was overcome, you were there to congratulate. Thank you for the conversation, both silly and serious, and for being kind and genuine people. Your positivity and camaraderie were indispensable.

To my family and friends in Seattle and beyond: There are too many people to thank who have helped me through my educational journey, but to everyone, thank you. In particular, my parents Mike and Marilyn have built an incredible support system for myself and my siblings and deserve all of the credit for helping me along the way and pushing me to where I am. Their support and encouragement throughout my education allowed me to focus on my goals without burden, and for that I am eternally grateful.

I dedicate this thesis to my grandfathers, James S. Shelton and John F. McPhee, two of the most influential people in my life. They were the first to instill in me the importance of and love for continuous learning, and more importantly the importance of family. My hope is that my accomplishments have and will continue to honor them.

Table of Contents

List of Figures, Schemes and Tables

Chapter 1

Chapter 2

Chapter 1

Introduction to chromatin organization, histone post-translational modifications and protein semisynthesis

1.1 Chromatin structure and organization

The nuclei of eukaryotic cells contain all of the genetic information necessary to successfully perform basic cellular processes including protein synthesis, replication, metabolism and signaling. This vast amount of information, present in the form of about three billion base pairs of deoxyribose nucleic acids (DNA), which would stretch approximately two meters from end-to-end, is packaged in an exceptionally small volume within each cell's nucleus (~500 μm^3). In order to efficiently package genomic DNA in the nucleus, it is tightly bound in a nucleoprotein complex known as chromatin. Chromatin structure was initially described as 'beads on a string' when first visualized by electron microscopy in 1974.[1,2] These studies helped to establish the nucleosome as the basic repeating unit of chromatin (**Figure 1.1**).[3,4] The nucleosome core particle (NCP) consists of an octameric protein complex containing two copies of each of the core histones, H2A, H2B, H3, and H4. Around the histone octamer, 147 base pairs (bp) of DNA are wrapped in ~1.65 super-helical turns, with typically 10-90 bp of linker DNA between the repeating NCPs.[5]

The octamer itself is composed of one H3-H4 tetramer and two H2A-H2B dimers bound together by strong electrostatic interactions between H2B and H4.[6] Each histone has a unique histone fold motif (three α-helices connected by two short disordered loops), responsible for the stabilizing interactions within the octamer core.[7] DNA affinity for the octamer varies greatly

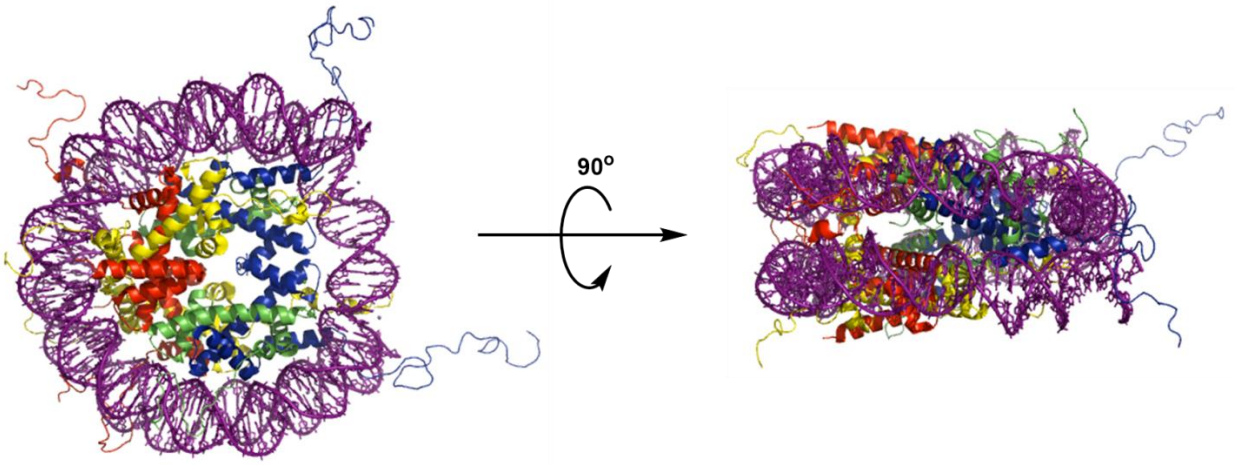


Figure 1.1. Structure of a mononucleosome. The structure of a mononucleosome at 1.9 Å resolution with individual histones and 147 bp of double-stranded DNA. The histone N-termini (tails) protrude from the globular core and are sites for post-translational modifications. Yellow = H2A, red = H2B, blue = H3, green = H4, purple = DNA. PDB: 1KX5.

depending on base pair sequence.[8] Elasticity of the DNA is essential in NCP binding, and thus intrinsically rigid stretches of DNA are heavily disfavored.[9] Protruding out from the well-defined NCP globular core, each histone contains a significant unstructured segment at its N-terminus, consisting of 20-40 amino acids. H2A and H2B also possess unstructured C-terminal stretches. These histone 'tails' extend past the DNA surface, leaving them prone to enzymatic post-translational modification (PTM) (Chapter 1.2).[10] The abundance of Lys and Arg residues in histone tails gives them a positive charge under physiological conditions, and are prominent sites of PTM. When unmodified, the tails engage in strong intra- and inter-nucleosomal contacts that are important for the formation of higher-order chromatin structures.[11] The linker histone H1, which binds to the nucleosome at entry and exit points of DNA-NCP contacts, is also critical for stabilizing compacted chromatin. Depletion of H1 causes aberrations in the formation of mitotic chromosomes during cell division, which possess the highest order of chromatin compaction.[12]

Higher-order chromatin structures form when 30 nm chromatin fibers form extended contacts, mediated by proteins such as the zinc-finger containing CCCTC-binding factor (CTCF) and protein complexes such as cohesin.[13,14] These proteins help engineer topologically associating domains (TADs) within chromatin that stretch ~800 kbps.[15] Interestingly, TAD boundaries are regions that display high transcriptional activity (also known as euchromatin) and tend to be located on the interior of the nucleus. Euchromatin typically appears as loosely compacted segments of DNA, sometimes lacking histone octamers and accessible to transcriptional machinery. Conversely, transcriptionally suppressed chromatin, or heterochromatin, is heavily compacted chromatin that and accumulates around the edges of the nucleus.[16] The formation and maintenance of euchromatic and heterochromatic states is key to the regulation of gene transcription, and both structures are heavily influenced by histone PTMs.

1.2 Histone post-translational modifications

The complex structure of chromatin and the fact that drastic changes in cellular processes arise from small changes in the DNA (e.g. DNA methylation) and protein (e.g. PTMs) landscape of the nucleosome has given rise to an extensive field of study focused on heritable changes in gene function that are independent of the underlying DNA sequence. This new field, termed epigenetics, largely focuses on how histone PTMs influence the activity of many enzymes responsible for carrying out essential cellular functions. While the primary modification on DNA is the transcriptionally repressive methylation of cytosine bases in CpG dinucleotide sequences, the solvent-exposed N-termini of histones can modified by a vast array of chemical groups and small proteins, accounting for diverse regulatory mechanisms and the so-called ‘histone code’ hypothesis. (**Figure 1.2**).[17–20]

On top of the ~24,000 annotated protein-coding genes in the human genome, the greatest source of functional diversity in the proteome comes from PTMs. Estimates of chemically unique

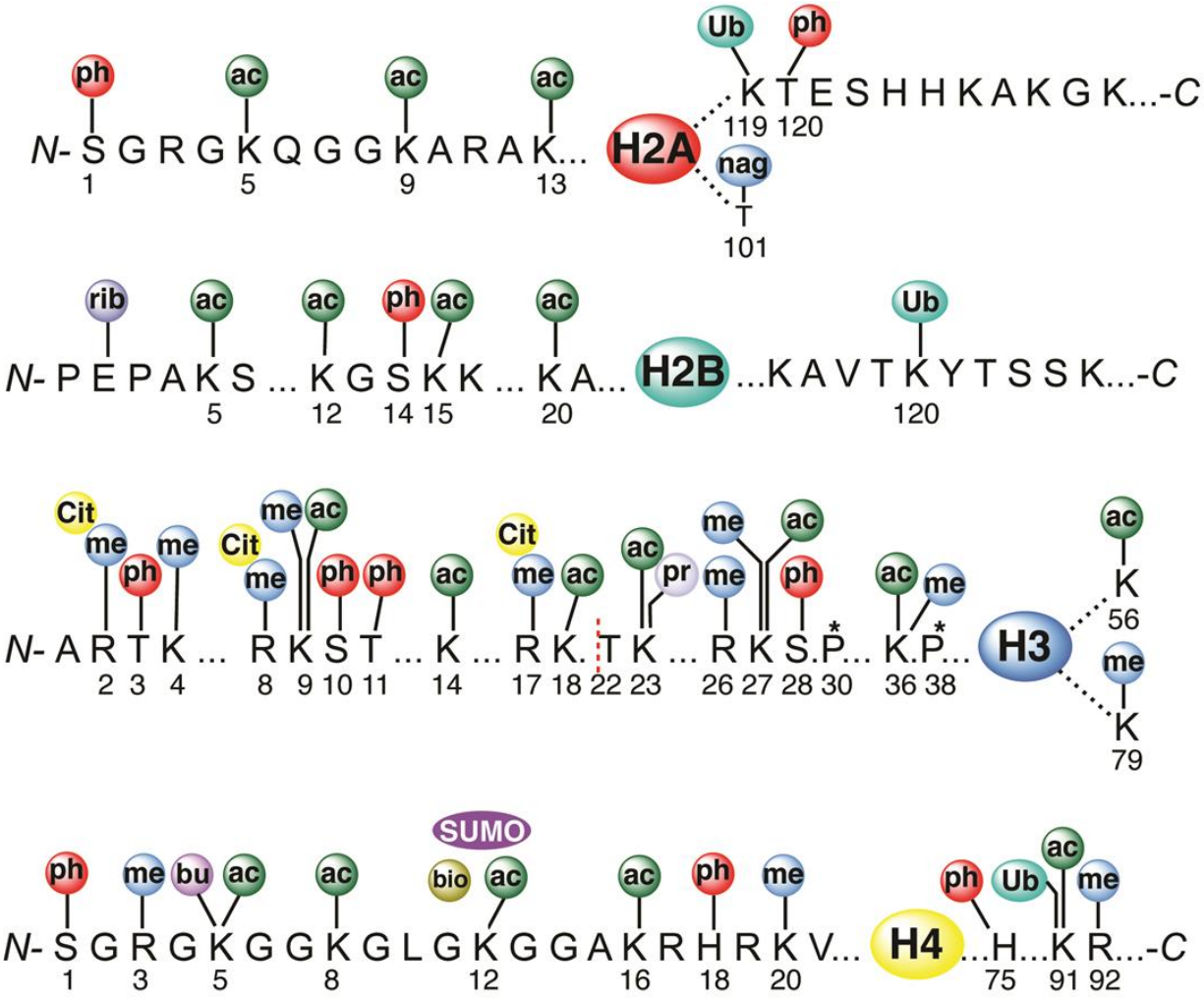


Figure 1.2. Schematic representation of histone modifications at the N- and C-termini and within globular domains. Chemical groups are indicated as follows: ac = acetyl; cit = citrullyl; bio = biotinyl; but = butyryl; me = methyl; nag = β-*N*-acetylglucosaminyl; ph = phosphoryl; pr = propionyl; rib = ADP-ribosyl; SUMO, SUMOyl; and Ub, ubiquityl. * = H3P30 and H3P38 cis-trans isomerization. The dashed line N-terminal to H3T22 indicates the primary site of proteolysis by cathepsin-L. Adapted from ref. 20.

protein species in humans exceed 1,000,000 when accounting for various post-translationally modified isoforms.[21] The reversibility of most PTMs imparts additional means for rapid response to environmental stimuli, which is necessary for cellular homeostasis (Chapter 1.3).

Technological improvements in mass spectrometry techniques have facilitated the identification of over 700 differentially modified isoforms of the core histones in HeLa cells.[22] Histone PTMs

include, but are not limited to small molecules such as methyl, phosphoryl, ADP-ribosyl and various acyl groups, to modification by the ~9 kDa protein, ubiquitin.[23] This complexity is a result of the presence of multiple modifiable residues on the histone tails and their propensity to be modified in multiple different ways. Abundant basic amino acids residues in the histone tails, such as Lys and Arg, are extensively modified by methyl and/or acyl groups. Histone methylation may manifest in multiple forms that impart drastically different outcomes on downstream cellular processes. For example, Lys residues may be mono-, di- or trimethylated (me1/2/3), which have been shown to signal distinct transcriptional functions.[24] In addition, Arg methylation can occur on the same or different ε-amines, known as symmetric or asymmetric methylation, resulting in distinct chemical and physical PTM landscapes.[25] Lys can also be modified by an array of acyl groups. Acetylation is by far the most common of these lysine acylation events and is known to occur on most H3 and H4 tail Lys residues.[26] However, proteomic studies have indicated that histone Lys are also modified by propionyl, butyryl, hydroxybutyryl, crotonyl, succinyl and malonyl acyl moieties among others.[27,28] Thus far, it is unclear what the vast diversity of Lys acylation event means for the fine-tuned regulation of chromatin-templated processes, but functional studies are beginning to unravel their differential impacts on chromatin structure and function.[29]

Phosphorylation of histone Ser, Thr and Tyr -OH groups is another highly dynamic histone PTM. Like histone methylation and acetylation, phosphorylation occurs mainly in the tails regions but has also been found within the globular core of H2B, H3 and H4. First discovered as a modification in rat thymus histones in 1966, phosphorylation has been linked to an astounding number of important cellular functions.[30] A well-recognized feature of DNA damage within chromatin is phosphorylation of the H2A variant, H2A.X.[31] Phosphorylation of H2A.X occurs rapidly over several mega bases in response to DNA damage. This phosphorylation is thought to create a signaling platform for the recruitment and retention of DNA damage repair factors

such as mediator of DNA damage checkpoint protein 1 (MDC1).[31–33] In addition to DNA damage response, histone phosphorylation is also associated with transcription and chromatin compaction during cell division and apoptosis.[34]

Many ongoing research efforts seek to understand the relationship between histone PTMs and gene expression. Transcription in eukaryotic organisms begins with transcription factor (TF) binding upstream of the core promoter region at the gene transcription start site (TSS). This leads to recruitment of adapter complexes, which results in binding of general TFs to the promoter, and finally formation of the pre-initiation complex (PIC).[35] Transcription ultimately depends on the ability of the PIC and subsequently RNA polymerase II to access DNA. This DNA accessibility is the result of prior histone PTMs and their influence on chromatin structure and recruitment of the proper enzymatic complexes. It has been well established that histone acetylation is associated with transcription activation due to its high abundance at promoter and intragenic regions of actively transcribed genes. For years it was thought that the primary means of transcriptional activation was through global acetylation of histone Lys. The high density of positively charged Lys and Arg residues in the histone N-terminal tails form strong interactions with the DNA sugar-phosphate backbone. Acetylation of lysines neutralize this positive charge and suppresses the interaction, resulting in a less compact chromatin structure that allows binding of transcription machinery to the DNA.[36] However, a report by Shogren-Knaack and Peterson in 2016 demonstrated that even a single acetylation mark at H4K16 is sufficient for chromatin decompaction by eliminating its interaction with the acidic patch within H2A/H2B dimers.[37] These results suggested that even small changes in the histone PTM landscape can have large effects on chromatin structure and gene expression. In a similar fashion, acetylation of H3K56, located at the nucleosomal DNA entry and exit points, prevents histone-DNA contacts. The weakened interactions between H3 and DNA due to K56ac increases DNA accessibility, and is strongly correlated with active transcription.[38]

While histone acetylation contributes primarily to transcription activation, the effect of other histone modifications varies depending on the site and degree of modification. For instance H3K4me2/3, H3K36me2/3, and H2BK120ub appear to be activating marks, while H3K9me2/3, H3K27me2/3, H4K20me, H2AK119ub, and H4K12 sumoylation (acylation with the small ubiquitin-like modifier, su) are repressive.[35,39] There are of course exceptions to these observations. For instance, H2Bub is highly-activating within gene bodies, but inhibits Pol II recruitment when located in promoter regions.[40] Crosstalk between PTMs is critically important in the establishment of active or repressed transcription. Therefore, the spatio-temporal control of histone PTMs by their associated regulatory enzymes is another major research focus.

1.3 Readers, writers and erasers of histone post-translation modifications

The discovery of new histone PTMs and the unravelling of their effect on chromatin structure and function have resulted in constant and sustained efforts to characterize the enzymes that regulate them. Enzymes responsible for installing PTMs are commonly referred to as *writers*, while enzymes that are responsible for removing them are referred to as *erasers*. Another class of proteins that contain specialized binding domains recognizing specific PTMs, are referred to as *readers*. With the exception of Arg methylation, all PTMs are known to be reversible to the unmodified amino acid form, though the site-specificity of most of the writers and erasers, and how specificities may change in various contexts largely remains unknown.

Long before the discovery of the first histone acetyl transferase (HAT) enzyme, Allfrey et al projected the importance of histone acetylation and methylation in regulating RNA synthesis.[41] The discovery of the first HAT, General control of amino acid synthesis protein 5 (Gcn5) in yeast, started and expanded the search for other writer and eraser enzymes such as histone deacetylases (HDACs), histone methyltransferases (HMTs) and histone demethylases (HDMs).[42] Since the discovery of Gcn5, hundreds of enzymes have been reported for the

placement, removal and recognition of all histone PTMs, aside from Arg demethylation. Surprisingly, methyl groups from arginine residues are removed in the form of Arg deimination, resulting in a new amino acid termed citrulline.[43]

Many chromatin modifying enzymes, including writers and erasers, are associated with expanded protein complexes that modulate enzyme substrate and site specificity. For example, Gcn5 acetylates K9 and K14 of histone H3 in the context of free H3, but as part of the Spt-Ada-Gcn5-acetyltransferase (SAGA) complex in yeast, its substrate scope expands to H3K56 and H2B Lys residues.[36,44] A human HAT, the CREB-binding protein/protein 300 (CBP/p300), is known to play a pivotal role in transcription. In addition to histone acetylation, p300 is known to interact with hundreds of other proteins including the androgen receptor (AR) transcription factor where acetylation significantly influences its transcription activity.[45] The histone acetylation eraser, HDAC1, is primarily studied in the context of one of several repressive enzyme complexes including the nucleosome remodeling and deacetylase (NuRD) and lysine specific demethylase (LSD1), co-repressor of RE1-silencing transcription factor (CoREST1) and HDAC1 (LCH) complexes (Chapter 3.1).[46,47] These protein complexes repress transcription through complimentary enzymatic activities that remove activating histone marks.

Methylation of histones, which can have multiple, site-dependent effects on transcription is regulated by an especially large families of HMTs and HDMs.[48,49] The first identified HMT, Su(var)3-9, is responsible for installing methyl groups at H3K9, a mark found in constitutive and facultative heterochromatin and associated with transcription repression.[50] The trimethylation of histone H3 at lysine 4 of (H3K4me3) is catalyzed by the SET1 [su(var)3-9, enhancer-of-zeste and trithorax] HMT in yeast, which co-localizes to the 5’ end of active genes associated with RNA Pol II. While still associated with active transcription, H3 K4me1 is not found in promoters but in the enhancer regions of various genes.[51]

The long in vivo half-life of methylated histones resulted in initial speculation that this modification was irreversible.[52] Then, in 2004 Shi et al reported the discovery of the first histone demethylase which they termed LSD1 (KDM1A, BHC110, AOF2).[53] Further studies showed that LSD1 uses a catalytic flavin adenine dinucleotide (FAD) cofactor to specifically demethylate H3 K4me1/2, in a transcriptionally repressive act.[54] Like Gcn5 and HDAC1, LSD1 exists in larger protein complexes, such as the LCH complex, and is known to change its substrate specificity based on its binding partners.[55,56] Since the discovery of LSD1, members of a much larger family of demethylating enzymes containing an Fe^{2+}/α-ketoglutarate-dependent jumonji (JmjC) demethylase domain have been described.[57] Unlike the chemically-limited nature of lysine demethylation by FAD-dependent amine oxidases, JmjC demethylases can demethylate trimethylated residues (Chapter 4.1).

Due to the vast number of potential histone PTMs, human cells have evolved protein domains to read the various modifications. The most abundant reader of acetylation marks is the highly conserved bromodomain which is observed in proteins functioning as transcriptional activators, chromatin remodelers and histone methyltransferases.[58] Readers of histone methylation vary widely in structure and recognition mechanism. Examples of methyl reader domains include plant homeo domain (PHD) fingers, tudor domains and chromodomains.[58] For instance, the presence of a tandem tudor domain in JMJD2 targets this enzyme to actively transcribed regions of chromatin by recognizing methylation on H3K4, which further facilitates gene activation by removing repressive H3K9me3 marks.[59]

1.4 Modification of histones by ubiquitin-like proteins

The Chatterjee lab is interested in studying the physiological relevance of histone modification by ubiquitin (Ub) and ubiquitin-like proteins (Ubls). Ub is a small, 76-amino acid protein that is

highly conserved in eukaryotes.[60] K63-linked polyubiquitylation of substrate proteins is commonly associated with 26S proteasome-mediated degradation of that substrate.[61] In contrast, monoubiquitylation is primarily involved in non-proteolytic functions, and was first discovered as a modification of histone H2A.[62] Ub is involved in many cellular pathways, including protein degradation, DNA double-stranded break repair and gene regulation.[63] A cascade of enzymatic reactions involving the E1, E2, and E3 Ub ligases deposit Ub onto substrate lysine residues. There is not a known consensus sequence for promoting ubiquitylation, and the E3 ligases responsible for enzymatic ubiquitylation have not been identified for most substrates. These limitations hinder site-specific enzymatic protein ubiquitylation and studies aimed at delineating the effect of site-specific modification.[64]

Of particular interest to the Chatterjee lab is histone modification by the Ubl SUMO family of proteins. SUMO isoforms, ranging from 92-96 amino acids in length, share a similar three-dimensional β-grasp fold and enzymatic cascade for substrate conjugation with Ub.[65] To date, five distinct human SUMO isoforms have been reported with the first, SUMO1, discovered by Melchior et al in 1997 as a modification of the Ran GTPase, RanGAP1.[66] Shortly after, genes encoding SUMO2/3 isoforms were reported.[67,68] SUMO2 and SUMO3 share 95% sequence identity, differing by just two additional amino acids in the N-terminus and SUMO2 being one amino acid longer. SUMO2/3 share ~46% sequence identity with SUMO1.[69] SUMO4 shares ~86% sequence identity with SUMO2/3, but is reportedly unable to form covalent attachments due to the inhibition of maturation from the pro-protein form, caused by a C-terminal proline residue not found in the other isoforms. Thus, the biological roles of SUMO4 remain elusive.[70,71] The recently annotated SUMO5 shares ~87% sequence identity with SUMO1, but shares sequence similarities to all of SUMO1-3 and so far has been demonstrated to be important in the regulation of promyelocytic leukemia protein (PML) nuclear condensates [72] To date,

SUMO1-3 are the most widely studied isoforms since they are constitutively expressed in cells.[67]

Histone ubiquitylation both promotes and inhibits gene transcription dependent on the site of modification. Monoubiquitylation of H2A mark regions of low transcription activity, and possibly suppresses transcription through its proclivity to bind linker histone H1.[73] Conversely, intragenic H2BK120ub promotes transcription through its propensity to sterically inhibit chromatin compaction and through activation of the H3K79 methyltransferase, disruptor of telomeric silencing 1 (hDOT1).[74,75] The roles of SUMO in transcription vary with the isoform type. Although there is no known role of SUMO1 modification of histones, SUMO1 has been associated with gene activation through modification of various transcription factors. For example, modification of p53 by SUMO1 has been shown to increase its transcription activity.[76,77] Interestingly, in a subsequent study by Müller and Schmidt, co-expression of p53 and the SUMO1 E3 ligase, protein inhibitor of activated STAT protein 1 (PIAS1), resulted in decreased p53 transcriptional activity in a SUMO-independent manner.[78] Like p53, modification of the AR transcription factor by SUMO1 increased AR-dependent transcriptional activity on promoters without effecting its transrepressing functions.[79] SUMO2/3 paralogues on the other hand have been demonstrated to modify histone H2B and H4, and so far have only been associated with transcriptional repression,.[80,81] In 2003, Shiio and Eisenmann demonstrated this effect by co-transfecting HeLa cells with a GAL4 DNA binding domain (DBD)-Ubc9 fusion and a tetrameric GAL4 promoter and luciferase reporter fusion.[82] Ubc9 was previously demonstrated to be a SUMO ligase sufficient for substrate sumoylation, and is still the only known E2 SUMO ligase.[83] Co-transfection with each construct resulted in significantly reduced luciferase reporter signal output. Three years later, Berger et al demonstrated the same phenomenon in *S. cerevisiae*, as abrogation of H2B sumoylation via lysine mutations downregulated expression of *TRP3, SUC2,* and *GAL1* genes.[84] Biophysical studies by the Chatterjee lab indicated that site-

specifically chemically sumoylated 12-mer nucleosome arrays adopt an open chromatin structure, contrary to the commonly held association of chromatin compaction with gene repression.[85] With these biophysical results in hand, Chapters 3 and 4 will focus on unravelling the biochemical and mechanistic contributions of histone sumoylation toward gene repression.

Protein targets of sumoylation often carry a ψKXE motif in their primary sequence, where ψ is a bulky aliphatic residue and X is any amino acid.[86] This motif is also found in an inverted form, and post-translational phosphorylation of Ser residues in some sequences are important for SUMO conjugation.[87] Proteomics studies have indicated that ~50% of all SUMO targets possess a consensus sumoylation motif.[88] Notably, the reported sites of histone sumoylation are not a part of a SUMO consensus motifs.[89] Sumoylated proteins also engage in noncovalent protein-protein interactions through a conserved binding cleft between the β2-sheet and α1-helix of each SUMO isoform (**Figure 1.3A,B**).[90] This shallow, rectangular binding pocket is lined with positively charged Lys and Arg residues. It is now known that many SUMO interacting proteins contain one or more putative SUMO interacting motifs (SIM) that bind this region.[91] SIMs generally bind with micromolar affinity and minimally contain a hydrophobic core with a so-called 'consensus sequence' of (V/I)-(V/I)-X-(V/I) or (V/I)-X-(V/I)-(V/I), where X is typically another aliphatic residue, Ser, Glu or Asp (**Figure 1.C**).[87,92] Patches of acidic residues often flank SIMs that may aid in SUMO binding, but no conservation in these SIM-adjacent regions has been established. SIMs are found in all known SUMO E3 ligases and some SUMO substrates, suggesting importance in SUMO conjugation.[93]

The SUMO proteins, their intracellular pathways and writer/eraser proteins are even less well-characterized than those of Ub, and warrant further study. Thus, semisynthetic access to chemically-defined substrates modified by Ub or Ubl proteins is crucial to enabling in vitro study of these processes.

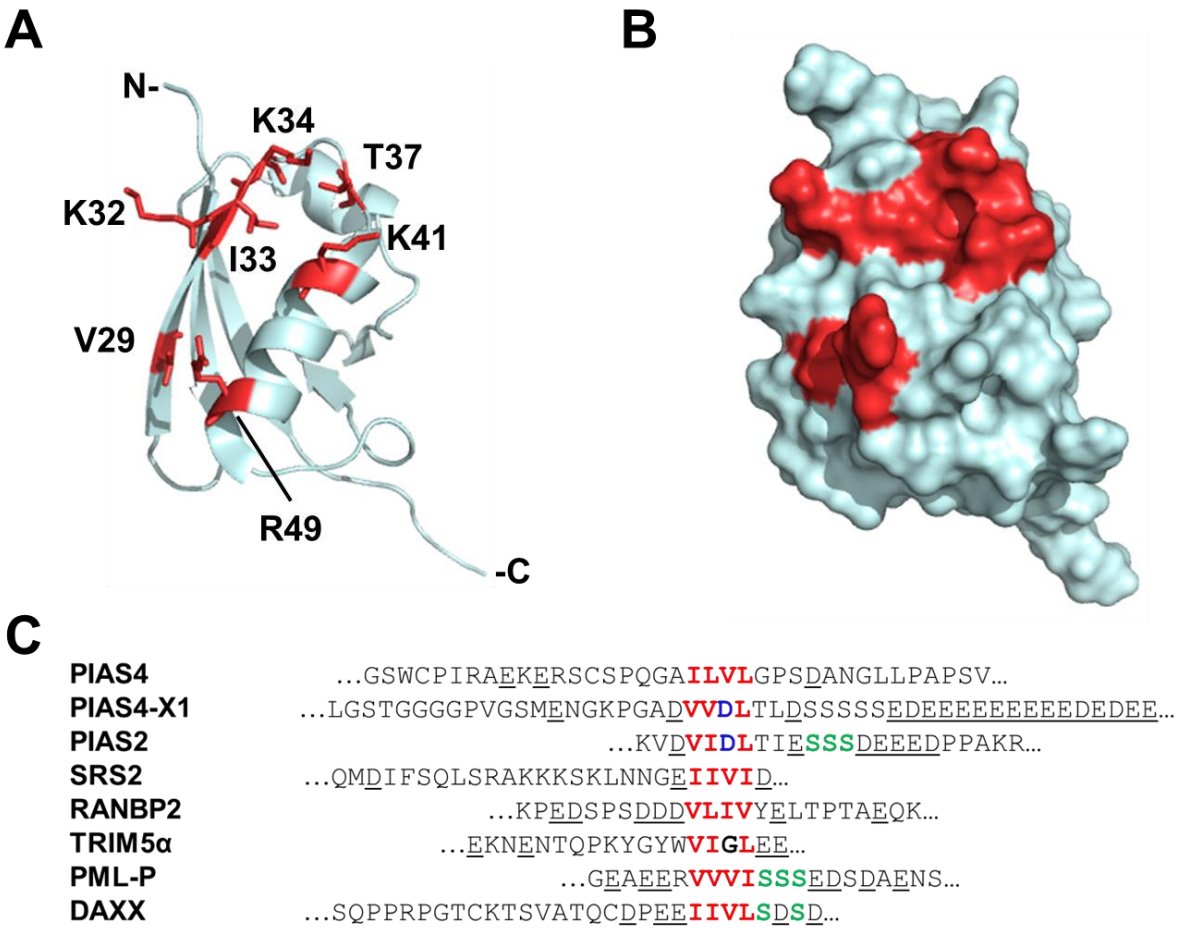


Figure 1.3. The SUMO-SUMO interacting motif (SIM) binding cleft and consensus SIM sequences. (A) Cartoon representation of SUMO3(C47S) structure. Important SIM bind cleft residues, as originally described in reference 90, are highlighted in red. (B) Surface representation of SUMO3(C47S) with matching highlighted SIM binding cleft residues. (C) Sequence alignment of consensus SIMs from selected proteins. Residues in red indicate important bulky aliphatic SIM positions. Serine residues in green indicate sites of post-translational phosphorylation known to enhance SIM binding. Underlined residues indicate SIM-adjacent acidic patch positions. PDB: 1U4A. Sequence alignment by C.J.A. Leonen.

1.5 Chemical modification of histones with small molecules

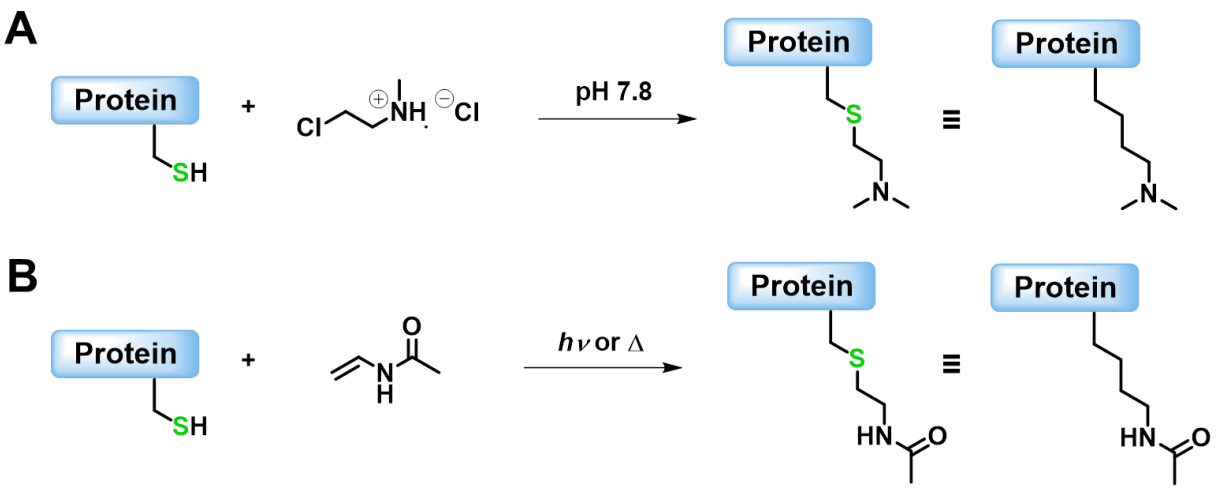
With the rapid increase in histone PTMs being discovered, efforts to delineate their roles in various cellular processes have been undertaken, both in isolation and in complex cellular

environments. However, a drawback to cell-based assays in elucidating PTM-specific contributions to a question of interest is the heterogeneity of chromatin PTMs. This PTM diversity cannot be completely controlled and may influence the activity of the PTM in question. In vitro assays with chemically defined substrates avoid these complications, and research into the functionality of histone PTM combinations has relied on solid-phase peptide synthesis (SPPS) and protein semisynthesis. 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.[65] However, peptide substrates do not accurately reflect in vivo processes in all cases. This was demonstrated in early work by Allfrey et al in which they found HDACs could not deacetylate short H4(15-21)K16ac tail peptides.[94] However, later studies by Allrey et al and Parello et al showed that an extended H4(1-37)K12,16ac peptide or an N-terminally acetylated H4(14-21) peptide were substrates for the same HDAC isoform.[95,96]

While minimal systems, such as modified peptides, can be useful for establishing the fundamental properties of chromatin-modifying enzymes, the results often do not translate to the cellular context. It is therefore useful to incorporate modified substrates into physiologically relevant systems to more accurately study the interplay between histone PTMs and their writer, eraser and reader proteins. This requires the synthesis of full-length modified histones. Many different approaches to generating full-length histones with defined PTMs have been reported. The simplest method is amino acid substitution through mutagenesis to mimic a particular modification. For instance, Gln resembles acetylated Lys, Arg resembles unacetylated Lys, and Glu resembles phosphorylated Ser. Amino acid substitutions are also amenable to high-yielding purification from *E. coli*, which is advantageous for biochemical and biophysical characterization techniques.[97]

Despite their usefulness as structural mimics, amino acid substitutions are limited for understanding enzymatic activity of potential writer and eraser proteins. High-yielding, cysteine-based methods can generate better PTM mimics that are substrates for enzymatic processing. For these techniques, histones are purified from *E. coli* with a Lys-to-Cys mutation at the desired site of post-translational modification. The only naturally occurring Cys residue in histones is C110 in H3. The structurally inconsequential mutation of C110 to Ala allows site-selective cysteine reactivity at other sites in the histone.[75] In 2007, Shokat et al reported an elegant approach for Cys alkylation and demonstrated its utility by modifying H3 with mono-, di- or trimethylated electrophilic ethylamines (**Scheme 1.1A**).[98] Alternatively, the thiol-ene 'click' chemistry has been used to alkylate Cys with N-vinylacetamide and used as an acetylated Lys mimic. In this reaction, a thermal- or UV light-inducible initiator generates a Cys thiyl radical that yields the anti-Markovnikov thioether product through addition across the double bond (**Scheme 1.1B**).[99] The authors showed that the H4K16ac mimic in nucleosomal arrays inhibited compaction to the same extent as native H4K16, and that deacetylation of the residue was only mildly hindered. Recent reports, however, suggest that reader proteins may not recognized thialysine analogs with the same fidelity as the native substrate, and thus results should be interpreted with caution when using these mimics.[100]

Amber suppression is another powerful alternative for PTM incorporation that uses ribosomal incorporation of biomimetics or native versions of modified amino acids into proteins, most robustly in *E. coli*.[101] During translation, the amber stop codon is recognized by an engineered tRNA co-expressed with its specific orthogonal aminoacyl tRNA synthetase. The synthetase is engineered to accept an unnatural amino acid. Chin et al have reported successful use of amber suppression to incorporate non-natural Nε-acetyl-L-lysine into rat super oxide dismutase and Nε-*tert*-butoxycarbonyl (Boc)-Nε-methyl-L-lysine into histone H3.[102,103] Additional control over site-specific modification comes from incorporation of selenocysteine. Hydrogen peroxide-



Scheme 1.1. Cysteine-based analogs of modified lysine. (A) Cysteine reaction with an electrophilic ethylamine, such as (2-chloroethyl)dimethylamine, yields a methyl lysine mimic. (B) Thiol-ene 'click' reaction between cysteine and N-vinylacetamide yields an acetyl lysine mimic. Adapted from Caroline Weller, Ph.D. thesis.

mediated oxidation of selenocysteine to dehydroalanine allows further reaction with N-acetylated or N-methylated 2-aminoethanethiol for installing acetyl and methyl Lys mimics.[104]

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 empirically. Further, incorporation of more complex modifications or of multiple different marks into the same histone by any of the techniques discussed thus far is non-trivial. In these cases, the ability to join chemically modified peptide fragments with native linkages is especially useful.

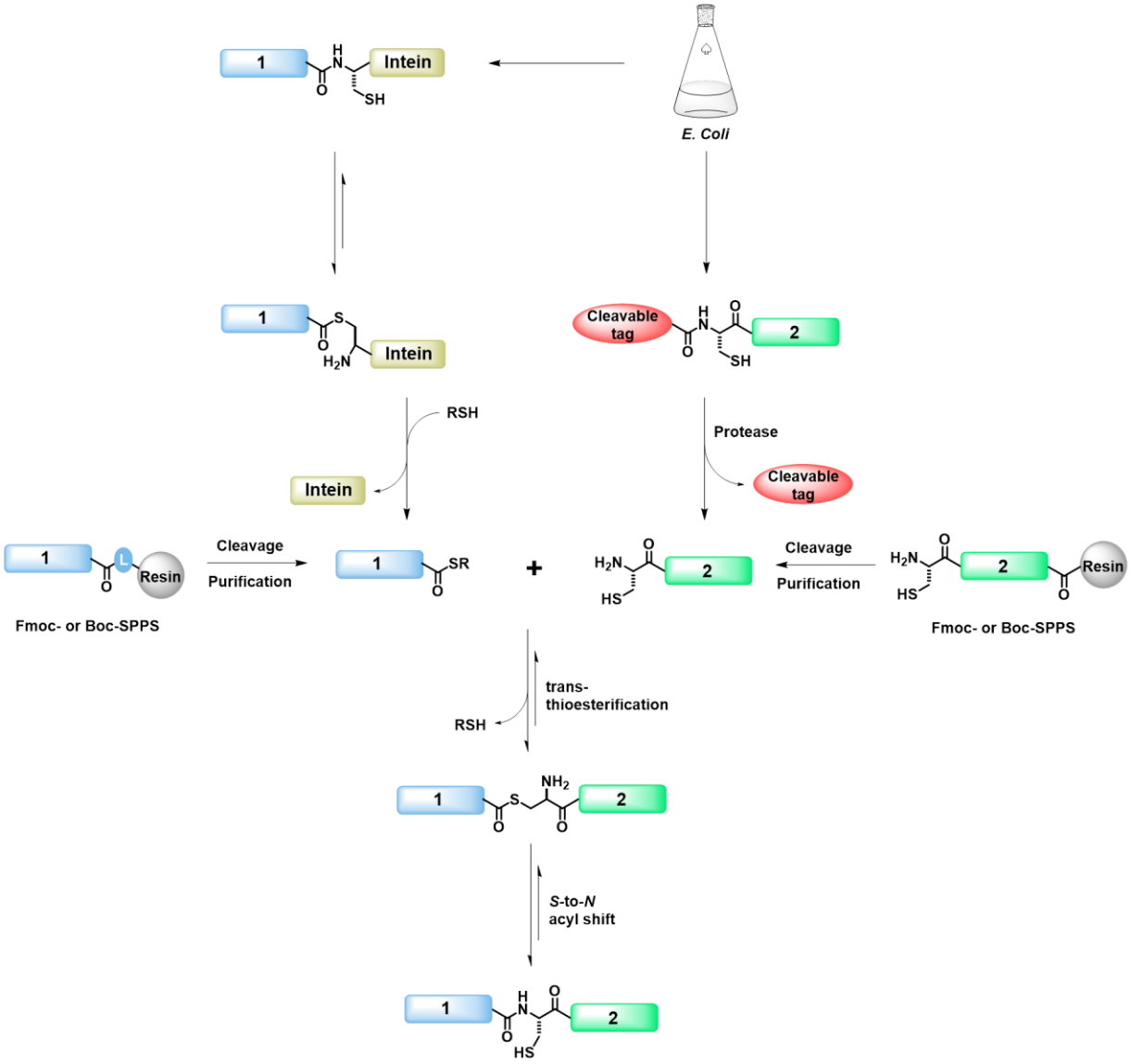
1.6 Modification of histones through Native Chemical Ligation

Native chemical ligation is a collection of strategies, the first reported by Kent et al, used to join C-terminally activated and N-terminal cysteine-containing peptide fragments with native amide linkages.[105] Since its discovery, NCL has become an essential tool in protein synthesis and semisynthesis, giving rise to dozens of derivative ligation techniques and hundreds of examples

utilizing the various strategy.[106] NCL relies on the transthioesterification reaction between a C-terminal peptide thioester and the thiol of an N-terminal cysteinyl peptide (**Scheme 1.2**). Thermodynamically favorable *S*-to-*N* acyl shift at the ligation junction provides the native linkage between peptide species. Given the utility of NCL, ligation techniques involving recombinant N-terminal cysteinyl protein truncates or C-terminal protein thioesters generated through intein-mediated thioesterification, known as expressed protein ligation (EPL), have found wide application in the synthesis of large and site-specifically modified proteins (**Scheme 1.2)**.

The semisynthesis of Ubl-modified proteins presents a distinctive challenge for applications of native chemical ligation, since Ubl substrates are typically modified through an isopeptide linkage at Lys side-chains. Chemical strategies have been developed to circumvent these issues, and their utility is typically demonstrated by synthesizing di-Ub and comparing the biological activity of various isomeric linkages (e.g. di-UbK6, 11, 27, 29, 33, 48, and 63).[107] The small and soluble nature of Ubl proteins make them quite amenable to chemical manipulation and many approaches have been used to synthesize Ub. The first total chemical synthesis of Ub by Fmoc-SPPS was performed by Ramage et al in 1989, and was followed by studies demonstrating that synthetic Ub is identical to natural Ub in structure and antibody recognition.[108–110]

To improve on the low yield of Ub total synthesis, recombinant methods, particularly those involving NCL and EPL, have been developed. Like histones, Ub has no native cysteine residues, so desulfurization permits temporary Ala-to-Cys mutations for NCL.[111] This allowed Brik et al to apply NCL to the synthesis of Ub directly from the ε-amine of Lys on resin-bound peptides.[112] In another approach to peptide ubiquitylation, which avoided the need for SPPS to synthesize large portions of Ub, Muir et al developed EPL strategy to synthesize full-length H2B



Scheme 1.2. Native Chemical Ligation and Expressed Protein Ligation. Scheme depicting the semisynthesis of proteins by NCL or EPL fragments generated through solid-phase peptide synthesis or heterologous expression in *E. coil.* The presence of a C-terminal α-thioester in fragment 1 and a N-terminal cysteine in fragment 2 results in a reversible trans-thioesterification reaction when incubated together. Favorable *S*-to-*N* acyl shift isomerizes the thioester bond between fragments 1 and 2 to the amide bond, which is less prone to the reverse shift, resulting in a stable semisynthetic protein. Adapted from Abhinav Dhall, Ph.D. thesis.

ubiquitylated at Lys120 (ubH2B).[113] The H2B(117-125) peptide was synthesized by Fmoc-SPPS, and Lys120 was orthogonally protected with the 4-methyltrityl (Mtt) group. The N-terminal Cys of the peptide was coupled as thiazolidine (Thz). Following selective deprotection

of K120, Cys was coupled to the ε-amine. The purified peptide was ligated to recombinant Ub(1-75) α-thioester, then Thz deprotection unmasked the N-terminal cysteine. A final ligation step to recombinant H2B(1-116) α-thioester, followed by desulfurization, yielded ubH2B with the G76A mutation.

To further simplify the synthesis ubiquitylation of peptides, Brik and coworkers developed a 5-mercaptolysine (δ-mercaptolysine) derivative 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.[114] Desulfurization renders this method traceless, making it appealing for structural studies of Ub dimers or chains. This Lys analog was applied to the impressive synthesis of the seven possible isomeric forms of di-Ub, and in the total synthesis of K48- and K63-linked tetra-Ub chains.[115,116] Derivatives of 4-mercaptolysine (γ-mercaptolysine) have also been applied to di-Ub syntheses.[117,118]

The synthesis of these Lys derivatives 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.[118] A recent alternative route has improved the synthesis to 8 steps but still in only ~5% overall yield.[119] A derivative of 5-mercaptolysine has been synthesized over 8 steps in ~50% overall yield, but its yield is highly dependent on the identity of the thiol protecting group.[115,119]

Despite the significant attention that Ub synthesis and semiysthesis has garnered, very few reports describe the synthesis or semisynthesis of SUMO isoforms and sumoylated proteins, presumably due the difference in length between SUMO and Ub. In 2016, Melnyk et al reported the first total synthesis of SUMO1 utilizing the δ-mercaptolysine strategy. In 2019, the same

group synthesized SUMO2/3 utilizing sequential bis-sulfanylethylamine (SEA) ligations (Chapter 2.1), and demonstrated the structural importance of the sole SUMO2/3 cysteine residue.[120–122] Lang et al have very recently reported an elegant method for site-specific protein sumoylation and ubiquitylation using a combination of genetic code expansion and chemoenzymatic techniques.[123] First, amber suppression was used to incorporate the unnatural amino acid, N^6-glycylglycyl-L-lysine at the desired site of modification in a protein of interest (POI), using an azide moiety to mask the side-chain primary amine. Staudinger reduction conditions were subsequently used to unmask the free amine. Concurrently, Ub or SUMO proteins were expressed with the C-terminal sortase recognition motifs, LP(A)LTGG. Incubation with sortase A then resulted in a thioester-linked Ubl-sortase conjugate, which was further ligated to the genetically modified POI to generate the native isopeptide linkage. Using this technique, the authors generated ubiquitylated proliferating cell nuclear antigen (PCNA) and sumoylated green fluorescent protein (GFP).

The Chatterjee lab was the first group to report the semisynthesis of sumoylated histones.[124,125] Our synthetically straightforward EPL-approach takes advantage of a previously reported ligation auxiliary based on 2-(aminooxy)ethanethiol, originally reported by Kent et al as an alternative to Cys for NCL at Gly-Gly junctions.[105] By conjugating the 2-(aminooxy)ethanethiol handle to histone peptide lysine residues, NCL is enabled with recombinant SUMO thioesters (Chapter 3.2). Metallic zinc was used to cleave the labile N-O bond, and a final ligation of an activated peptide C-terminus to recombinant, N-terminal cysteine-containing histone truncates provided full-length and natively-linked sumoylated histones. During the development of this novel and traceless ligation approach, Caroline Weller noticed that incubating full-length SUMO-histone ligation products, still bearing the ligation auxiliary in buffered solutions containing the thiol sodium mercaptoethane sulfonate (MESNa), resulted in reverse chemical ligation and stable MES thioester formation. Chapter 2 will describe how we have harnessed this

serendipitous discovery to develop a novel and multifaceted peptide α-thioesterification strategy.

1.7 References

(1) Olins, A. L.; Olins, D. E. *Science (80-.).* **1974**, *183* (4122), 330–332.

(2) Woodcock, C. L. F.; Safer, J. P.; Stanchfield, J. E. *Exp. Cell Res.* **1976**, *97* (1975).

(3) Kornberg, R. D. *Science (80-.).* **1974**, *184* (4139), 868–871.

(4) Richmond, T. J.; Finch, J. T.; Rushton, B.; Rhodes, D.; Klug, A. *Nature* **1984**, *311* (5986), 532–537.

(5) Li, G.; Zhu, P. *FEBS Lett.* **2015**, *589* (20), 2893–2904.

(6) Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389* (6648), 251–260.

(7) Mariño-Ramírez, L.; Kann, M. G.; Shoemaker, B. A.; Landsman, D. *Expert Rev. Proteomics* **2005**, *2* (5), 719–729.

(8) Gottesfeld, J. M.; Luger, K. *Biochemistry* **2001**, *40* (37), 10927–10933.

(9) Struhl, K.; Segal, E. *Nat. Struct. Mol. Biol.* **2013**, *20* (3), 267–273.

(10) Crane-Robinson, C.; Bohm, L. *Biosci. Rep.* **1984**, *4* (5), 365–386.

(11) Luger, K.; Dechassa, M. L.; Tremethick, D. J. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (7), 436–447.

(12) Maresca, T. J.; Freedman, B. S.; Heald, R.; Maresca, T. J.; Freedman, B. S.; Heald, R. *J. Cell Biol.* **2005**, *169* (6), 859–869.

(13) Dean, A. *Brief. Funct. Genomics* **2011**, *10* (1), 3–10.

(14) Kagey, M. H.; Newman, J. J.; Bilodeau, S.; Zhan, Y.; Orlando, D. A.; Van Berkum, N. L.; Ebmeier, C. C.; Goossens, J.; Rahl, P. B.; Levine, S. S.; Taatjes, D. J.; Dekker, J.; Young, R. A. *Nature* **2010**, *467* (7314), 430–435.

(15) Dekker, J.; Heard, E. *FEBS Lett.* **2015**, *589* (20), 2877–2884.

(16) Rodriguez, A.; Bjerling, P. *Biochem. Soc. Trans.* **2013**, *41* (6), 1635–1639.

(17) Jones, P. A. *Nat. Rev. Genet.* **2012**, *13* (7), 484–492.

(18) Jenuwein, T.; Allis, C. D. *Science (80-.).* **2001**, *293* (5532), 1074–1080.

(19) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403* (January).

(20) Dhall, A.; Chatterjee, C. *ACS Chem. Biol.* **2011**, *6* (10), 987–999.

(21) Ponomarenko, E. A.; Poverennaya, E. V.; Ilgisonis, E. V.; Pyatnitskiy, M. A.; Kopylov, A. T.; Zgoda, V. G.; Lisitsa, A. V.; Archakov, A. I. *Int. J. Anal. Chem.* **2016**, *2016*.

(22) 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. *Genome Biol.* **2012**, *13* (10).

(23) Nadal, S.; Raj, R.; Mohammed, S.; Davis, B. G. *Curr. Opin. Chem. Biol.* **2018**, *45*, 35–47.

(24) Greer, E. L.; Shi, Y. *Nat. Rev. Genet.* **2012**, *13* (5), 343–357.

(25) Di Lorenzo, A.; Bedford, M. T. *FEBS Lett.* **2011**, *585* (13), 2024–2031.

(26) Sterner, D. E.; Berger, S. L. *Microbiol. Mol. Biol. Rev.* **2000**, *64* (2), 435–459.

(27) Olsen, C. A. *Angew. Chemie - Int. Ed.* **2012**, *51* (16), 3755–3756.

(28) Xie, Z.; Dai, J.; Dai, L.; Tan, M.; Cheng, Z.; Wu, Y.; Boeke, J. D.; Zhao, Y. *Mol. Cell. Proteomics* **2012**, *11* (5), 100–107.

(29) Rousseaux, S.; Khochbin, S. *Cell J.* **2015**, *17* (1), 1–6.

(30) Stevely, W. S.; Stocken, L. A. *Biochem. J.* **1966**, *100*, 20–21.

(31) Rogakou, E. P.; Pilch, D. R.; Orr, A. H.; Ivanova, V. S.; Bonner, W. M. *J. Biol. Chem.* **1998**, *273* (10), 5858–5868.

(32) Stucki, M.; Clapperton, J. A.; Mohammad, D.; Yaffe, M. B.; Smerdon, S. J.; Jackson, S. P. *Cell* **2005**, *123* (7), 1213–1226.

(33) Iacovoni, J. S.; Caron, P.; Lassadi, I.; Nicolas, E.; Massip, L.; Trouche, D.; Legube, G. *EMBO J.* **2010**, *29* (8), 1446–1457.

(34) Rossetto, D.; Avvakumov, N.; Côté, J. *Epigenetics* **2012**, *7* (10), 1098–1108.

(35) Li, B.; Carey, M.; Workman, J. L. *Cell* **2007**, *128* (4), 707–719.

(36) Shahbazian, M. D.; Grunstein, M. *Annu. Rev. Biochem.* **2007**, *76* (1), 75–100.

(37) Shogren-Knaak, M.; Ishii, H.; Sun, J.-M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L. *Science (80-.).* **2006**, *311* (5762), 844–847.

(38) Tessarz, P.; Kouzarides, T. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (11), 703–708.

(39) Zhang, T.; Cooper, S.; Brockdorff, N. *EMBO Rep.* **2015**, *16* (11), 1467–1481.

(40) Batta, K.; Zhang, Z.; Yen, K.; Goffman, D. B.; Franklin Pugh, B. *Genes Dev.* **2011**, *25* (21), 2254–2265.

(41) Allfrey, G.; Faulkner, R.; Mirsky, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **1964**, *51* (1938), 786–794.

(42) Myers, P. L.; Skvirsky, R. C.; Greenberg, M. L.; Greer, H. *Mol. Cell. Biol.* **1986**, *6* (9), 3150–3155.

(43) Klose, R. J.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (4), 307–318.

(44) Grant, P. A.; Duggan, L.; Côté, J.; Roberts, S. M.; Brownell, J. E.; Candau, R.; Ohba, R.; Owen-Hughes, T.; Allis, C. D.; Winston, F.; Berger, S. L.; Workman, J. L. *Genes Dev.* **1997**, *11* (13), 1640–1650.

(45) Fu, M.; Wang, C.; Reutens, A. T.; Wang, J.; Angeletti, R. H.; Siconolfi-Baez, L.; Ogryzko, V.; Avantaggiati, M. L.; Pestell, R. G. *J. Biol. Chem.* **2000**, *275* (27), 20853–20860.

(46) Humphrey, G. W.; Wang, Y.; Russanova, V. R.; Hirai, T.; Qin, J.; Nakatani, Y.; Howard, B. H. *J. Biol. Chem.* **2001**, *276* (9), 6817–6824.

(47) Xue, Y.; Wong, J.; Moreno, G. T.; Young, M. K.; Côté, J.; Wang, W. *Mol. Cell* **1998**, *2* (6), 851–861.

(48) Mosammaparast, N.; Shi, Y. *Annu. Rev. Biochem.* **2010**, *79* (1), 155–179.

(49) Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z. W.; Schmid, M.; Opravil, S.; Mechtier, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T. *Nature* **2000**, *406* (6796), 593–599.

(50) Schotta, G.; Ebert, A.; Krauss, V.; Fischer, A.; Hoffmann, J.; Rea, S.; Jenuwein, T.; Dorn, R.; Reuter, G. *EMBO J.* **2002**, *21* (5), 1121–1131.

(51) Black, J. C.; Van Rechem, C.; Whetstine, J. R. *Mol. Cell* **2012**, *48* (4), 491–507.

(52) Byvoet, P.; Shepherd, G. R.; Hardin, J. M.; Noland, B. J. *Arch. Biochem. Biophys.* **1972**, *148*, 558–567.

(53) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. *Cell* **2004**, *119*, 941–953.

(54) Forneris, F.; Binda, C.; Vanoni, M. A.; Mattevi, A.; Battaglioli, E. *FEBS Lett.* **2005**, *579* (10), 2203–2207.

(55) Metzger, E.; Wissmann, M.; Yin, N.; Müller, J. M.; Schneider, R.; Peters, A. H. F. M.; Günther, T.; Buettner, R.; Schüle, R. *Nature* **2005**, *437* (September), 436–439.

(56) Shi, Y.-J.; Matson, C.; Lan, F.; Iwase, S.; Baba, T.; Shi, Y. *Mol. Cell* **2005**, *19* (6), 857–864.

(57) Klose, R. J.; Kallin, E. M.; Zhang, Y. *Nat. Rev. Genet.* **2006**, *7* (9), 715–727.

(58) Teske, K. A.; Hadden, M. K. *Eur. J. Med. Chem.* **2017**, *136*, 14–35.

(59) Hillringhaus, L.; Yue, W. W.; Rose, N. R.; Ng, S. S.; Gileadi, C.; Loenarz, C.; Bello, S. H.; Bray, J. E.; Schofield, C. J.; Oppermann, U. *J. Biol. Chem.* **2011**, *286* (48), 41616–41625.

(60) Sharp, P. M.; Li, W. *Trends Ecol. Evol.* **1987**, *2* (11).

(61) Finley, D. *Annu. Rev. Biochem.* **2009**, *78* (1), 477–513.

(62) Goldknopf, I. R. A. L.; French, M. F.; Busch, H. *Proc. Natl. Acad. Sci. USA* **1977**, *74* (12), 5492–5495.

(63) Hicke, L. *Nat. Rev. Mol. Cell Biol.* **2001**, *2* (3), 195–201.

(64) Weller, C. E.; Pilkerton, M. E.; Chatterjee, C. *Biopolymers* **2014**, *101* (2), 144–155.

(65) Ploegh, H. L.; van der Veen, A. G. *Annu. Rev. Biochem.* **2012**, *81*.

(66) Mahajan, R.; Delphin, C.; Guan, T.; Gerace, L.; Melchior, F. *Cell* **1997**, *88* (1), 97–107.

(67) Lapenta, V.; Chiurazzi, P.; Van Der Spek, P.; Pizzuti, A.; Hanaoka, F.; Brahe, C. *Genomics* **1997**, *40* (2), 362–366.

(68) I, A. C.; Mannen, H.; Steven, S.; Carolina, N. *Biochem. Mol. Biol. Int.* **1998**, *46* (6), 1161–1174.

(69) Saitoh, H.; Hinchey, J. *J. Biol. Chem.* **2000**, *275* (9), 6252–6258.

(70) Bohren, K. M.; Nadkarni, V.; Song, J. H.; Gabbay, K. H.; Owerbach, D. *J. Biol. Chem.* **2004**, *279* (26), 27233–27238.

(71) Owerbach, D.; McKay, E. M.; Yeh, E. T. H.; Gabbay, K. H.; Bohren, K. M. *Biochem. Biophys. Res. Commun.* **2005**, *337* (2), 517–520.

(72) Liang, Y. C.; Lee, C. C.; Yao, Y. L.; Lai, C. C.; Schmitz, M. L.; Yang, W. M. *Sci. Rep.* **2016**, *6*, 1–15.

(73) Weake, V. M.; Workman, J. L. *Mol. Cell* **2008**, *29* (6), 653–663.

(74) Fierz, B.; Chatterjee, C.; McGinty, R. K.; Bar-Dagan, M.; Raleigh, D. P.; Muir, T. W. *Nat. Chem. Biol.* **2011**, *7* (2), 113–119.

(75) Chatterjee, C.; McGinty, R. K.; Fierz, B.; Muir, T. W. *Nat. Chem. Biol.* **2010**, *6* (4), 267–269.

(76) Rodriguez, M. S.; Desterro, J. M. P.; Lain, S.; Midgley, C. A.; Lane, D. P.; Hay, R. T. *EMBO J.* **1999**, *18* (22), 6455–6461.

(77) Gostissa, M.; Hengstermann, A.; Fogal, V.; Sandy, P.; Schwarz, S. E.; Scheffner, M.; Del Sal, G. *EMBO J.* **1999**, *18* (22), 6462–6471.

(78) Schmidt, D.; Müller, S. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (5), 2872–2877.

(79) Poukka, H.; Karvonen, U.; Jinne, O. A.; Palvimo, J. J. *Proc Natl Acad Sci USA* **2000**, *97* (26), 14145–14150.

(80) Hendriks, I. A.; D'Souza, R. C. J.; Yang, B.; Verlaan-De Vries, M.; Mann, M.; Vertegaal, A. C. O. *Nat. Struct. Mol. Biol.* **2014**, *21* (10), 927–936.

(81) Chandrasekharan, M. B.; Huang, F.; Sun, Z.; Allis, C. D.; Chandrasekharan, M. B.; Huang, F.; Sun, Z. *Proc Natl Acad Sci U S A* **2009**, *106* (39), 16686–16691.

(82) Shiio, Y.; Eisenman, R. N. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230.

(83) Desterro, J. M. P.; Thomson, J.; Hay, R. T. *FEBS Lett.* **1997**, *417* (3), 297–300.

(84) 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.; Johnson, E. S.; Berger, S. L. *Genes Dev.* **2006**, *20* (8), 966–976.

(85) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837.

(86) Rodriguez, M. S.; Dargemont, C.; Hay, R. T. *J. Biol. Chem.* **2001**, *276* (16), 12654–12659.

(87) Flotho, A.; Melchior, F. *Annu. Rev. Biochem.* **2013**, *82*, 357–385.

(88) Matic, I.; Schimmel, J.; Hendriks, I. A.; van Santen, M. A.; van de Rijke, F.; van Dam, H.; Gnad, F.; Mann, M.; Vertegaal, A. C. O. *Mol. Cell* **2010**, *39* (4), 641–652.

(89) Lamoliatte, F.; McManus, F. P.; Maarifi, G.; Chelbi-Alix, M. K.; Thibault, P. *Nat. Commun.* **2017**, *8*.

(90) Chupreta, S.; Holmstrom, S.; Subramanian, L.; Iñiguez-lluhí, J. A.; In, J. A. *Mol. Cell. Biol.* **2005**, *25* (10), 4272–4282.

(91) Song, J.; Durrin, L. K.; Wilkinson, T. a; Krontiris, T. G.; Chen, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (40), 14373–14378.

(92) Namanja, A. T.; Li, Y. J.; Su, Y.; Wong, S.; Lu, J.; Colson, L. T.; Wu, C.; Li, S. S. C.; Chen, Y. *J. Biol. Chem.* **2012**, *287* (5), 3231–3240.

(93) Cappadocia, L.; Lima, C. D. *Chem. Rev.* **2018**, *118* (3), 889–918.

(94) Vidali, G.; Gershey, E. L.; Allfrey, V. G. *J. Biol. Chem.* **1974**, *249* (24), 332–334.

(95) Krieger, D. E.; Vidali, G.; Erickson, B. W.; Allfrey, V. G.; Merrifield, R. B. *Bioorg. Chem.* **1979**, *8* (4), 409–427.

(96) Kervabon, A.; Mery, J.; Parello, J. *FEBS Lett.* **1979**, *106* (1), 93–96.

(97) Qi, Y.-K.; Ai, H.-S.; Li, Y.-M.; Yan, B. *Front. Chem.* **2018**, *6* (February), 1–11.

(98) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. *Cell* **2007**, *128* (5), 1003–1012.

(99) Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiöld, L.; Liu, C. F. *Angew. Chemie - Int. Ed.* **2011**, *50* (41), 9611–9614.

(100) Chen, Z.; Notti, R. Q.; Ueberheide, B.; Ruthenburg, A. J. *Biochemistry* **2018**, *57* (3), 300–304.

(101) Davis, L.; Chin, J. W. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (3), 168–182.

(102) Nguyen, D. P.; Garcia Alai, M. M.; Kapadnis, P. B.; Neumann, H.; Chin, J. W. *J. Am.*

Chem. Soc. **2009**, *131* (40), 14194–14195.

(103) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4* (4), 232–234.

(104) Guo, J.; Wang, J.; Lee, J. S.; Schultz, P. G. *Angew. Chemie - Int. Ed.* **2008**, *47* (34), 6399–6401.

(105) Canne, L. E.; Bark, S. J.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118* (25), 5891–5896.

(106) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J. C. M.; Melnyk, O. *Chem. Rev.* **2019**, *119* (12).

(107) Hershko, A.; Ciechanover, A. *Annu. Rev. Biochem.* **1998**, *67*, 425–479.

(108) Ramage, R.; Green, J.; Ogunjobi, O. M. *Tetrahedron Lett.* **1989**, *30* (16), 2149–2152.

(109) Ramage, R.; Green, J.; Muir, T. W.; Ogunjobi, O. M.; Love, S.; Shaw, K. *Biochem. J.* **1994**, *299* (1), 151–158.

(110) Alexeev, D.; Bury, S. M.; Turner, M. A.; Ogunjobi, O. M.; Muir, T. W.; Ramage, R.; Sawyer, L. *Biochem. J.* **1994**, *299* (1), 159–163.

(111) Wan, Q.; Danishefsky, S. J. *Angew. Chemie - Int. Ed.* **2007**, *46* (48), 9248–9252.

(112) Kumar, K. S. A.; Spasser, L.; Ohayon, S.; Erlich, L. A.; Brik, A. *Bioconjug. Chem.* **2011**, *22* (2), 137–143.

(113) McGinty, R. K.; Köhn, M.; Chatterjee, C.; Chiang, K. P.; Pratt, M. R.; Muir, T. W. *ACS Chem. Biol.* **2009**, *4* (11), 958–968.

(114) Ajish Kumar, K. S.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A. *Angew. Chemie - Int. Ed.* **2009**, *48* (43), 8090–8094.

(115) El Oualid, F.; Merkx, R.; Ekkebus, R.; Hameed, D. S.; Smit, J. J.; De Jong, A.; Hilkmann, H.; Sixma, T. K.; Ovaa, H. *Angew. Chemie - Int. Ed.* **2010**, *49* (52), 10149–10153.

(116) Kumar, K. S. A.; Spasser, L.; Erlich, L. A.; Bavikar, S. N.; Brik, A. *Angew. Chemie - Int. Ed.* **2010**, *49* (48), 9126–9131.

(117) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. *Chem. Commun.* **2010**, *46* (38), 7199–7201.

(118) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. *J. Am. Chem. Soc.* **2009**, *131* (38), 13592–13593.

(119) Merkx, R.; De Bruin, G.; Kruithof, A.; Bergh, T. Van Den; Snip, E.; Lutz, M.; El Oualid, F.; Ovaa, H. *Chem. Sci.* **2013**, *4* (12), 4494–4498.

(120) Bouchenna, J.; Sénéchal, M.; Drobecq, H.; Stankovic-Valentin, N.; Vicogne, J.; Melnyk, O. *Bioconjug. Chem.* **2019**.

(121) Hou, W.; Zhang, X.; Li, F.; Liu, C. *Org. Lett.* **2011**, *13* (3), 2009–2012.

(122) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. *Org. Lett.* **2010**, *12* (22), 5238–5241.

(123) Fottner, M.; Brunner, A. D.; Bittl, V.; Horn-Ghetko, D.; Jussupow, A.; Kaila, V. R. I.; Bremm, A.; Lang, K. *Nat. Chem. Biol.* **2019**, *15* (3), 276–284.

(124) 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.; Chatterjee, C. *Nat. Commun.* **2016**, *7*, 1–10.

(125) Weller, C. E.; Huang, W.; Chatterjee, C. *Chembiochem* **2014**, *15* (9), 1263–1267.

Chapter 2

The MEGA approach to peptide α-thioesterification, ligation and cyclization

2.1 Introduction

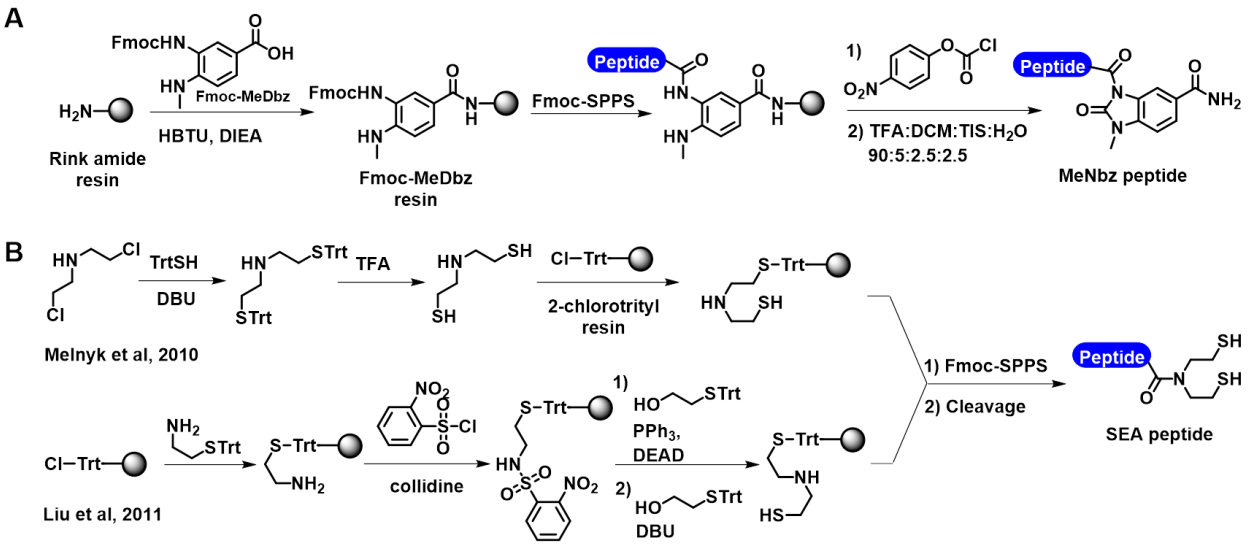
Peptide α-thioesters are critical components in protein semisynthesis strategies that use native chemical (NCL) or expressed protein ligation (EPL). In contrast with the relative ease of obtaining N-terminal Cys-containing polypeptides and proteins for NCL and EPL via solid-phase peptide synthesis (SPPS) or heterologous expression in *Escherichia coli*,[1,2] the preparation of C-terminal peptide α-thioesters using 9-fluorenylmethoxy carbonyl SPPS (Fmoc-SPPS) chemistry is limited by their intrinsic reactivity with the organic amine bases used for Fmoc-deprotection. While peptide α-thioesters may be synthesized with various C-terminal thiol resin linkers using the *tert*-butyloxycarbonyl (Boc-) α-amine protecting group strategy,[3,4] applications of Boc-chemistry have several limitations including incompatibility with phosphorylated and glycosylated amino acids,[5,6] as well the common use of hydrogen fluoride (HF) gas for cleavage of the peptide from the solid phase.[7] Although Kent and coworkers have recently reported trifluoromethanesulfonic acid (TFMSA) as an alternative to HF, the general utility and functional group compatibility of TFMSA have not yet been described.[8] Thus, simple and robust strategies employing reagents from Fmoc-SPPS for generating peptide α-thioesters are highly desirable for simplifying protein semisynthesis by chemical ligation and for incorporating various PTMs in proteins.

Several strategies for generating peptide α-thioesters by Fmoc- chemistry have previously been reported. However, these approaches often require multistep chemical manipulations after

SPPS,[9–11] or modification of the solid-phase with a chemical linker prior to peptide assembly,[12–15] each with their own inherent limitations and synthetic challenges. One example of a solid-phase linker approach is the diaminobenzoic acid (Dbz), and the second generation methyl Dbz (MeDbz) strategies, pioneered by Dawson et al (**Scheme 2.1A**).[9,11] This strategy first incorporates Fmoc-MeDbz on to Rink amide resin, followed by standard SPPS protocols to extend to peptide chains from the primary amine of the MeDbz linker. The resin-linked MeDbz peptide is then converted to the *N*-acylbenzimidazolinone (Nbz) via double acylation of the linker amines with *p*-nitrophenylchloroformate, and subsequently purified as the Nbz peptide. The authors showed that C-terminally activated peptides derived from this approach were well-suited for direct thioesterification by incubation with an external thiol, and could be applied to further NCL and cyclization reactions. Despite the reported applicability of this method,[16–18] the Chatterjee lab was unable to produce a 6-mer Nbz peptide derived from the histone H3 N-terminal tail (ARTKQT) using the Dbz linker. Our results suggested that any single peptide thioesterification strategy is unlikely to be applicable to all peptide sequences, and the continued development of thioesterification approaches are necessary to truly expand the scope of protein targets accessible by NCL.

Alternative thioesterification strategies also utilize modified C-terminal amino acids,[19] or strongly acidic conditions and elevated temperatures with Cys to favor intramolecular *N*-to-*S* acyl shift of the backbone amide bond,[20] followed by transthioesterification with external thiols. Functionalized resins containing alkyl thiols that are suitably positioned for nucleophilic attack at the C-terminal amide bond, also known as crypto-thioesters, hold promise due to the minimal chemical manipulation required post-SPPS.[21–28] For example, the bis(2-sulfanylethyl)amino (SEA) approach, reported by both Melnyk et al and Liu et al in close succession, has proven to be a robust method for the synthesis of peptide thioesters (**Scheme 2.1B**).[27,28] The reported strategies use differing methods to install a bis(2-sulfanylethyl)amine handle on the solid phase,

at which point SPPS is used to prepare peptides from the secondary amine. Cleavage from the solid support provides the SEA peptide crypto-thioesters, which can be converted to isolable thioesters via incubation in thiol containing buffers, or used directly in NCL. However, the complex chemistry required to install many crypto-thioester handles limits their accessibility to a handful of laboratories. Therefore, efforts to expand the utility of NCL would benefit from facile and high-yielding Fmoc-based strategies that produce peptide α-thioesters. Based on the serendipitous discovery by Caroline Weller in our lab that chemical ligation products bearing the 2-(aminooxy)ethanethiol handle can undergo reverse native chemical ligation under specific conditions (Chapter 1.6), we wondered if this chemistry could be harnessed into a functional thioesterification strategy.



Scheme 2.1. MeDbz and SEA approaches toward peptide α-thioesterification. (A) Dawson's MeDbz linker approach toward MeNbz peptides. (B) Melnyk and Liu's approach toward SEA peptides

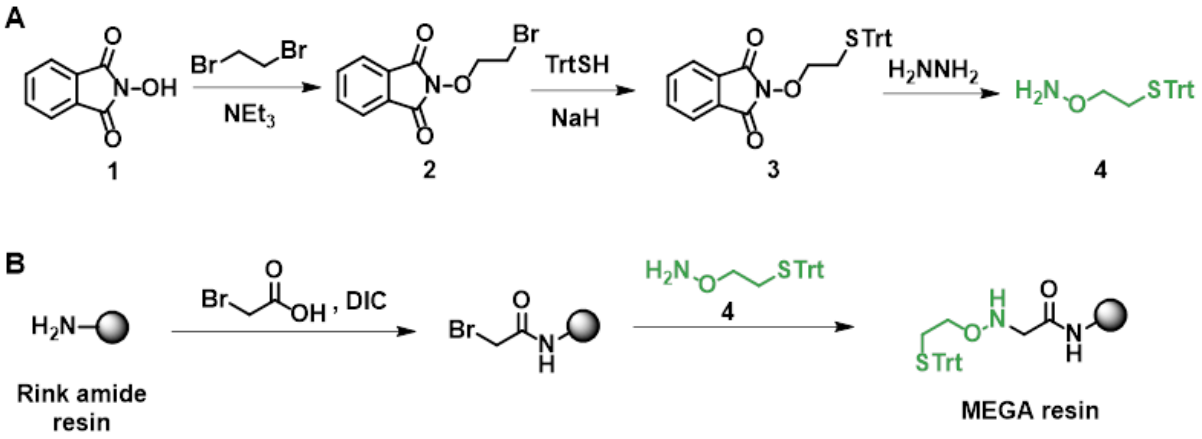
In this chapter, I will report the development of a novel and robust peptide α-thioesterification approach based on the mercaptoethoxyglycinamide (MEGA) solid-phase linker.[29] This approach requires minimal solid-phase preparation and is compatible with a wide-range of C-terminal

amino acids and peptide lengths. The versatile MEGA linker and MEGA-based peptides are also amenable one-pot NCL reactions, greatly simplifying the chemical ligation work flow, and are ideally suited for the construction of cyclic peptide scaffolds.

2.2 Results and Discussion

2.2.1 Synthesis of MEGA resin, AWKG-MEGA peptide and initial α-thioesterification screen

The proposed thioesterification strategy first required preparation of a functionalized solid support utilizing the 2-(aminooxy)ethanethiol linker that was suitable for constructing peptides and would undergo the desired thioesterification reaction. Preparation of the 2-(aminooxy)ethanethiol molecule was carried out in three previously reported synthetic steps to yield multi-gram quantities of the MEGA precursor (**Scheme 2.2A**).[30,31] Briefly, sub-stoichiometric amounts of *N*-hydroxyphthalimide (**1**) was first reacted with 1,2-dibromoethane to yield the monosubstituted bromoethoxyphthalimide product (**2**). Compound **2** was purified by recrystallization after aqueous work up. Triphenylmethanethiol was subsequently used to displace the second bromide in the presence of anhydrous sodium hydride to produce compound **3**. After purification of **3** by column chromatography, the final deprotection was performed using aqueous hydrazine to release the free amine of S-trityl protected 2-(aminooxy)ethanethiol (**4**). Following synthesis of the small molecule, the resin-bound MEGA handle was assembled in two steps prior to peptide synthesis. Commercially available Rink amide resin was first condensed with bromoacetic acid via diisopropylcarbodiimide (DIC)-mediated coupling. The bromoacetylated resin was then incubated with protected 2-(aminooxy)ethanethiol in a simple S_N2 reaction to obtain the MEGA-linked resin (**Scheme 2.2B**).



Scheme 2.2. Synthesis of 2-(aminooxy)ethanethiol and MEGA resin. (A) Synthesis of 2-(aminooxy)ethanethiol. (B) Synthesis of MEGA resin.

Next, we explored conditions for coupling the first amino acid to the secondary amine of MEGA. We elected to prepare the AWKG-MEGA model peptide for initial peptide synthesis and thioesterification studies. This short peptide sequence was designed to have good solubility in aqueous buffers through incorporation of a charged Lys residue and to be spectrophotometrically quantifiable by the presence of the Trp aromatic system. Gly, being the simplest amino acid that lacks steric bulk or reactive side-chain functional groups, was chosen as the first C-terminal amino acid to evaluate. We found that using DIC with the additive ethyl(hydroxyimino)cyanoacetate (Oxyma) sufficed for coupling Gly to the MEGA resin. Loading of the amino acid was evaluated by quantitative Kaiser test, which routinely indicated >80% coupling efficiency of Gly to the secondary amine.[32] Upon attaching the first Gly residue to the MEGA resin, standard SPPS procedures were used to extend the peptide chain. The AWKG-MEGA peptide was cleaved from the solid support and finally purified by reverse-phase high-performance liquid chromatography (RP-HPLC) (**Figure 2.S1A,B**). The lack of significant side-product formation observed in the crude peptide prior to bulk purification indicated that the relatively labile N-O bond in the MEGA auxiliary was largely stable throughout peptide synthesis and cleavage procedures (**Figure 2.S2A**).

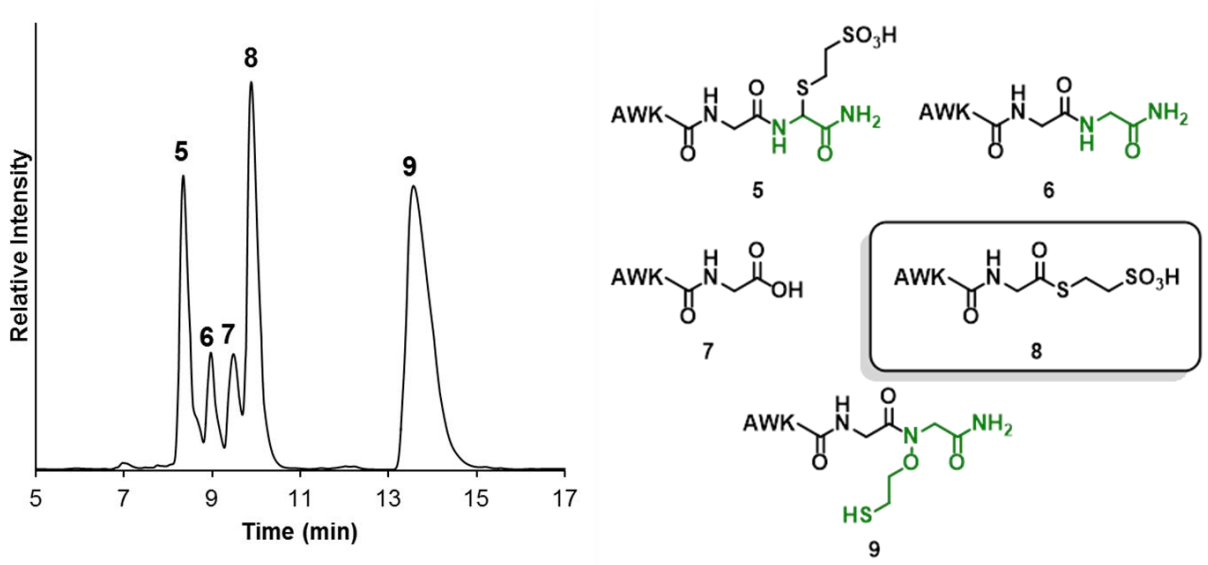


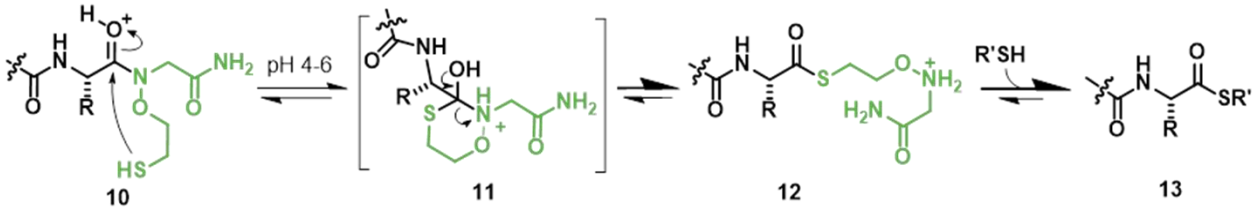
Figure 2.1. Initial thioesterification attempt. AWKG-MEGA peptide was incubated in thiol containing buffer for 24 h at 25 °C. RP-HPLC chromatogram (left) and structures (right) describe the observed product distribution. Buffer: 100 mM Na_2HPO4 (pH 7.4), 200 mM MESNa. RP-HPLC: C18, 10-30% CH_3CN in water, 30 min gradient.

The purified AWKG-MEGA peptide was then tested in an initial screen to assess the extent of product thioester formation. In an attempt to emulate typical chemical ligation conditions, the AWKG-MEGA peptide was incubated in a buffered phosphate solution (pH 7.4), containing the aliphatic thiol mercaptoethane sulfonate (MES, MESNa for the sodium salt). The reaction was subsequently analyzed by RP-HPLC and electron spray ionization mass spectrometry (ESI-MS) to evaluate product distribution (**Figure 2.1**). While we were pleased to observe formation of the desired MES peptide thioester **8**, this first reaction was plagued by low thioester yield (26%) sluggish consumption of the initial MEGA peptide **9**, as well as significant side-product formation. Thioester hydrolysis (**7**), N-O bond cleavage in the starting MEGA peptide (**6**), as well as an uncharacterized MEGA-MES peptide adduct that coincided with N-O cleavage (**5**) were the main products of undesired reactivity. We speculate that the unknown product **5** is the result of base-mediated elimination adjacent to the MEGA amide nitrogen, causing N-O bond cleavage and intermediate imine formation. This highly reactive species could then conceivably

undergo a reaction with MESNa to give the MES adduct **5**. The results from this experiment, while encouraging, necessitated extensive optimization of the MEGA approach.

2.2.2 Thioesterification reaction optimizations

Mechanistically, we envisioned that thioester formation from the N-oxyamide **10** could proceed through an *N*-to-*S* acyl shift and transthioesterification to form the stable thioester **13** (**Scheme 2.3**). Collapse of the tetrahedral intermediate **11** would provide the initial thioester **12**, and addition of an excess of external thiol could then promote thiol exchange to generate the final product. Although MEGA thioesterification proceeds through a 6-membered cyclic intermediate, rather than the 5-membered intermediate for N-alkylated Cys[19] or SEA strategies,[27,28] the good aminooxy leaving group may enable thioesterification at mildly acidic pH and room temperature for the AWKG-MEGA peptide. This is in contrast with pH ≈ 1 required for thioesterification from N-alkylated Cys.[19]



Scheme 2.3. Proposed mechanism for the MEGA peptide thioesterification reaction

We sought to optimize thioesterification conditions to find a favorable balance between the nucleophilicity of the MEGA handle thiolate with the electrophilicity of the C-terminal amino acid carbonyl moiety. These two properties are primarily influenced by reaction pH, as more basic conditions lead to deprotonation of the MEGA sulfhydryl and enhanced thiolate reactivity. Conversely, more acidic conditions promote carbonyl oxygen protonation, thereby increasing electrophilicity of the carbonyl carbon. Hence, we first screened the AWKG-MEGA thioesterification reaction with varying pH. Thioester yields were best under mildly acidic

conditions (pH 4.0−6.0), and no reaction was observed at pH < 4. Mildly basic reaction conditions, pH 7.0−8.0, yielded a mixture of thioester and hydrolyzed products at elevated temperatures (**Figure 2.S3A**). We note that when other aliphatic thiols were used in this reaction, they failed to provide any of the corresponding α-thioester (**Figure 2.S3B-D**). Aromatic thiols were not tested alone in thioesterfication reactions because, despite their generally lower pK_a and increased nucleophilicity, they were recently shown by the Chatterjee lab to cleave N-O bonds.[33]

We also found that the thioesterification reaction temperature significantly influenced product yield. Under an optimized reaction pH using the AWKG-MEGA peptide, increasing the reaction temperature from 25 to 37 °C led to a concomitant increase in thioester yield from 26% to 60% (**Figure 2.S3E**). However, further increasing reaction temperature for this substrate led to lower thioester yield, simultaneously resulting in increased thioester hydrolysis. The addition of organic co-solvents to the reaction was also tested to increase the solution dielectric constant, thus theoretically enhancing MEGA thiolate nucleophilicity by reducing its solvation. To this end, we added 10% (v/v) acetonitrile (MeCN), dimethylformamide (DMF) or dimethylsulfoxide (DMSO) to the AWKG-MEGA thioesterification reaction (**Figures 2.S3F-H**). We found that adding MeCN to the reaction offered no increase in thioester yield, yet was not detrimental to product formation. On the other hand, adding either DMF or DMSO resulted in a mixture of mono- and diformylated species or solely MEGA peptide disulfide formation, respectively. Thus, the addition of these organic co-solvents was not beneficial toward MEGA reactivity.

2.2.3 Scope and limitations of the MEGA approach

A series of 4-mer peptides with the sequence AWKX-MEGA was synthesized to assess the C-terminal amino acid compatibility of MEGA. Consistent with previously reported conditions for coupling onto secondary amines, coupling reagents such as hexafluorophosphate

azabenzotriazole tetramethyl uronium (HATU) or bis(trichloromethyl) carbonate (BTC) were required for amino acids other than Gly or Ala.[34] However, this did not result in significant racemization during resin loading, as seen from the good purity of crude peptides released from the resin (**Figure 2.S2**). Acidolytic cleavage from the resin and purification by RP-HPLC yielded good quantities of AWKX-MEGA peptides from each of the 20 naturally occurring C-terminal amino acids (**Table 2.1, Figure 2.S1**).

AWKX-MEGA peptides were subsequently tested for the production of isolable α-thioesters. We observed that the nature of the C-terminal amino acid, thiol nucleophile, pH, and reaction temperature significantly influenced thioester yields (**Scheme 2.3**, **Figure 2.2**, **Table 2.2**). As mentioned above, transthioesterification was particularly sensitive to the identity of the external thiol, and MESNa was the only thiol that provided appreciable amounts of any peptide thioester. AWKX-MEGA peptides exhibited slow thioesterification kinetics at room temperature when X was not Gly or Ala. However, like the AWKG-MEGA peptide, elevated temperatures improved yields for all peptide thioesters (**Table 2.2, Figure 2.S4**). Unlike AWKG-MEGA, other C-terminal amino acids required temperatures ≥50 °C for efficient thioesterification.

Steric hindrance in β-branched C-terminal amino acids drastically reduces reaction rates during NCL,[4] and so were inclined to test AWKI-MEGA, AWKV-MEGA and AWKT-MEGA in thioesterification reactions. Ile, possessing the most steric hindrance, displayed extremely slow reaction kinetics leading to limited thioester production (18%) over 48 h (**Table 2.2, entry 18, Figure 2.S5GG**)**.** Val produced a modest 26% yield, while Thr generated significantly more of the corresponding α-thioester (**Table 2.2, entries 11,17**). Overall, β-branching in the C-terminal amino acid proved to be a far greater obstacle toward efficient thioesterification than the absolute bulk of the side chain, as demonstrated by the good yields for AWKW, AWKL, and

Table 2.1. Synthesis of AWKX-MEGA Peptides

Entry	X	Crude Purity [%][a]	Isolated Yield [%][b]	Calcd MW	Obsd MW[c]
1	K	85	51	664.8	664.6
2	Q	85	47	664.8	664.8
3	V	75	47	635.8	635.5
4	G	71	45	593.7	593.4
5	A	77	44	607.7	607.4
6	L	79	43	649.8	649.5
7	S	72	42	623.7	623.6
8	E	78	41	665.8	665.8
9	N	58	41	650.8	650.7
10	M	69	35	667.9	667.7
11	W	65	35	722.9	722.8
12	F	89	33	683.8	683.6
13	H	77	31	672.8	673.2
14	C	44	31	639.8	639.5
15	I	67	31	648.8	649.2
16	Y	62	29	699.8	699.6
17	P	81	23	632.8	633.3
18	R	40	20	692.9	692.5
19	T	79	19	637.8	637.5
20	D	62	18	651.8	651.4
21	D-A	82	56	607.7	607.4
22	D-C	58	43	639.8	639.5

[a]Purity of crude peptide based on RP-HPLC peak integration at 280 nm. [b]0.05 mmol scale.
[c]Based on ESI-MS [$M+H^+$] ion peak.

Table 2.2. Optimized thioesterification conditions for AWKX-MEGA peptides

Entry	X	Temperature (°C)	pH	Reaction Time (h)	Yield (%)[a]
1	Y	70	4.5	48	70.4
2	W	70	4.5	48	68.8
3	M	70	4.5	48	68.7
4	L	70	4.5	48	67.8
5	K	70	4.5	48	67.5
6	F	70	4.5	48	65.8[b]
7	G	37	5.6	72	60.5
8	D-A	70	5.6	24	57.1[b]
9	Q	50	4.5	72	48.8
10	A	70	5.6	24	45.2[b]
11	T	70	4.5	48	43.1
12	S	50	4.5	48	42.8
13	R	70	5.6	72	41.1
14	D-C	50	4.5	48	38.3[b]
15	D	50	5.6	8	32.5
16	C	50	4.5	48	32.5[b]
17	V	70	5.6	72	26.3
18	I	70	4.5	72	18.0

[a]Based on RP-HPLC peak integration at 280 nm. [b]average yield ± 2.8%, n=3

AWKF thioesters in comparison with AWKV and AWKI (**Table 2.2, entries 2,4,6,17,18**).

Surprisingly, although the AWKD-MEGA peptide showed good initial kinetics of thioesterification, we observed a dramatic drop in recovered thioester product after >8 h (**Figure 2.S5**). ESI-MS characterization of the reaction products at 24 h suggested the formation of a C-terminal aspartic anhydride, followed by its hydrolysis to aspartate (**Figure 2.S6**). This suggests that some C-terminal amino acids with nucleophilic side chains may cyclize upon incubation at elevated temperatures for extended periods of time, which is a known issue for Asn and Glu.[35] Therefore, AWKN-MEGA and AWKE-MEGA were also tested for their thioesterification propensity, but resulted only in diastereomers of α- and δ-glutamic thioesters and asparaginimide products, respectively (**Figure 2.S7**). Despite these results, peptides with C-terminal Lys and Gln residues produced corresponding thioesters in good yield (**Table 2.2, entries 5,9**). Lastly, we note that using Pro as the C-terminal amino acid resulted in no amount of the corresponding MES thioester, but also very little consumption of the starting MEGA peptide (**Figure 2.S4HH**). The distinct absence of reactivity for this species may suggest a particular lack of conformational freedom at its C-terminus, preventing the initial intramolecular *N*-to-*S* acyl shift.

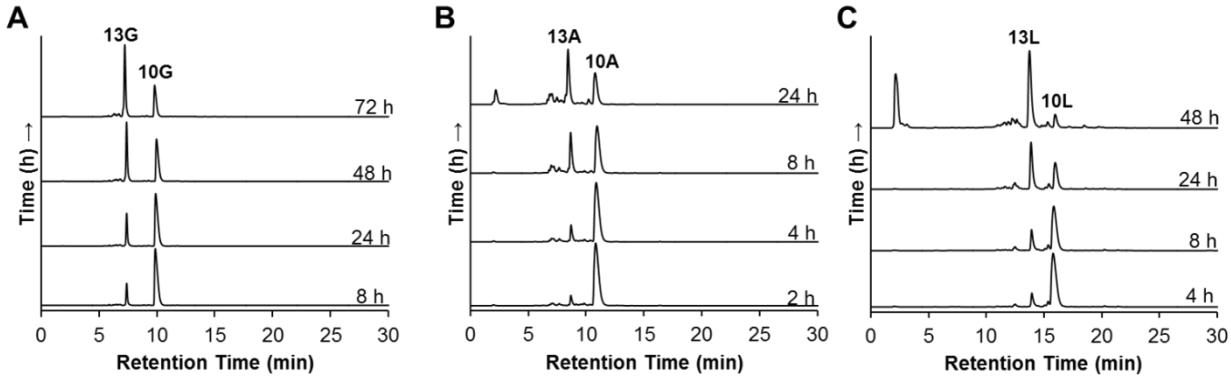


Figure 2.2. Example AWKX-MEGA thioesterification time-courses. (A-C) C18 RP-HPLC time-course of thioester formation for AWKG-MEGA (A), AWKA-MEGA (B), and AWKL-MEGA (C) peptides. Buffer: 100 mM NaH_2PO_4, 200−400 mM MESNa, 25−50 mM TCEP. RP-HPLC: 10−60% CH_3CN in water, 30 min gradient.

To test the utility of MEGA for longer sequences, the 35-mer peptide p53(1−35)-MEGA was prepared via microwave-assisted automated SPPS on a Liberty Blue peptide synthesizer (CEM corporation, Matthews, NC) (**Figure 2.S8A**). Initial attempts revealed that 90 °C coupling and deprotection cycles with 20% (v/v) piperidine in DMF led to significant N−O bond cleavage and reduced yields. We addressed this issue by decreasing coupling and deprotection temperatures to 50 °C, and employed 5% (w/v) piperazine and 0.1 M HOBt in DMF as a less basic mixture for Fmoc-deprotection. Under these optimized conditions N−O bond cleavage was completely eliminated from the crude peptide product and pure p53(1−35)-MEGA peptide was obtained in 22% isolated yield (**Figures 2.S8B,C**). The final peptide was subjected to thioesterification and generated the corresponding MES thioester in 75% yield (**Figures 2.S8D,E**). The fact that MEGA-linked resin is fully compatible with microwave-assisted SPPS may facilitate the synthesis of otherwise challenging sequences in higher yields and purity.[36]

2.2.4 MEGA-enabled one-pot native chemical ligation

With the successful synthesis of thioesters of varying lengths, we next investigated the scope for one-pot NCL between AWKX-MEGA peptides and an N-terminal Cys-containing peptide, CASW (**16**) (**Figures 2.3A,2.S9**). First, AWKXMEGA thioesterification was undertaken with optimized conditions (**Table 2.2**), followed by the addition of CASW at pH 7.5. NCL proceeded rapidly at room temperature to generate the peptide AWKXCASW in 1 h with minimal thioester hydrolysis (**Figure 2.3B-D**). The ability to directly use thioesters generated with MEGA in NCL, without intermediate purification, is particularly attractive for higher overall yields of ligation products

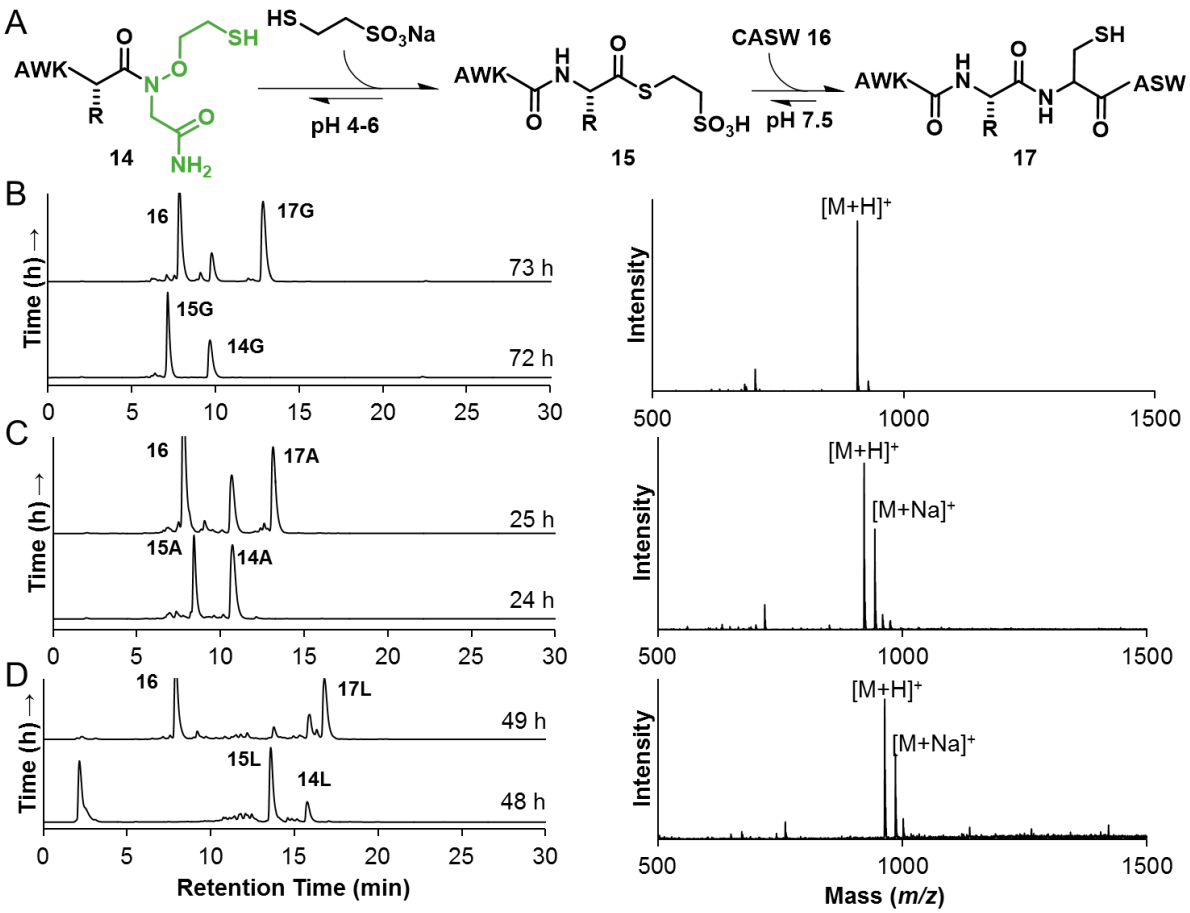


Figure 2.3. One-pot thioesterification of AWKX-MEGA (14) and native chemical ligation (NCL) with CASW (16). (A) Scheme for one-pot NCL of AWKX-MEGA peptides. (B) C18 RP-HPLC time-course of NCL for AWKG-MEGA (left) and ESI-MS of isolated AWKGCASW (right). Calcd. $[M+H]^+$ 907.6 Da, obsd. 907.9 Da. (C) C18 RP-HPLC time-course of NCL for AWKA-MEGA (left) and ESI-MS of isolated AWKACASW (right). Calcd. $[M+H]^+$ 921.9 Da, obsd. 921.7 Da. (D) C18 RP-HPLC time-course of NCL for AWKL-MEGA (left) and ESI-MS of isolated AWKLCASW (right). Calcd. $[M+H]^+$ 963.2 Da, obsd. 963.8 Da. RP-HPLC: C18 analytical, 10-60% B, 30 min gradient.

2.2.5 MEGA-enabled peptide cyclization

Following successful one-pot intermolecular NCL with MEGA, we envisioned its application for intramolecular NCL to access head-to-tail cyclized peptides. Cyclic peptides are useful scaffolds in therapeutic discovery efforts, and the lack of free N- and C-termini in cyclic peptides confers resistance to exopeptidases and enhances cell membrane permeability.[37] To this end, we synthesized two short, N-terminal cysteine containing peptides, CASHEW-MEGA and CRGD(D-F)-MEGA (**Figures 2.S10**). Cyclic CRGD(D-F) binds the integrin αVβ3 receptor with nanomolar

affinity.[38] Both peptides were prepared by automated SPPS in overall isolated yields of 56% and 45%, respectively (**Figure 2.S11,S12**). The peptides were subjected to one-pot thioesterification and cyclization, and both reactions proceeded to completion in <8h with no significant side products detected by HPLC and MS (**Figure 2.S13**). Although the synthesis of cyclic peptides by MEGA requires the presence of an N-terminal thiol for ligation, the ability to desulfurize Cys,[39] and other thiol-containing Cys surrogates,[40] post-cyclization will provide facile access to a wide range of cyclic peptides that may not contain Cys residues in their primary sequences.

Encouraged by our results with short cyclic peptides, we applied the MEGA strategy toward a more challenging target - the Sunflower Trypsin Inhibitor-1 (SFT-1) peptide. SFT-1 is a 14-mer cyclic peptide that is structurally constrained by three Pro residues and an extensive H-bond network while also containing one internal disulfide.[41] SFT-1 is the smallest Bowman−Birk type trypsin inhibitor, and analogs of SFT-1 are active against a range of proteases including matriptase and kallikreins (**Figure 2.4A**). One analog of SFT-1, the I10G mutant, was previously prepared using manual peptide synthesis and Boc-chemistry.[42] To simplify access to SFT-1(I10G), we synthesized the linear sequence 14-mer peptide CFPDGRCTKSIPPG-MEGA by automated SPPS (**Figures 2.4B**, **2.S14**). We note that the SFT-1(I10G) mutant retains the potent trypsin inhibitory property of the naturally occurring compound, but replacement of the β-branched Ile by Gly allows simpler access to the cyclic scaffold through intramolecular NCL.[42] Additionally, Ile10 is opposite to the trypsin binding interface within the cyclic peptide and therefore is non-essential for the inhibitory activity of SFT-1.[41] The purified peptide was incubated in optimal thioesterification buffer for 24 h at 50 °C to provide the cyclized product in 30% isolated yield after RP-HPLC purification (**Figure 2.S15**). The cyclized peptide was then quantitatively oxidized to the disulfide form by incubation in ammonium bicarbonate buffer overnight (**Figure 2.4C**).

Finally, we tested SFT-1 (I10G) for inhibitory activity against bovine Trypsin by employing the well-characterized substrate *N*α-benzoyl-L-arginine-4-nitroanilide (BAPNA). Trypsin hydrolyzes

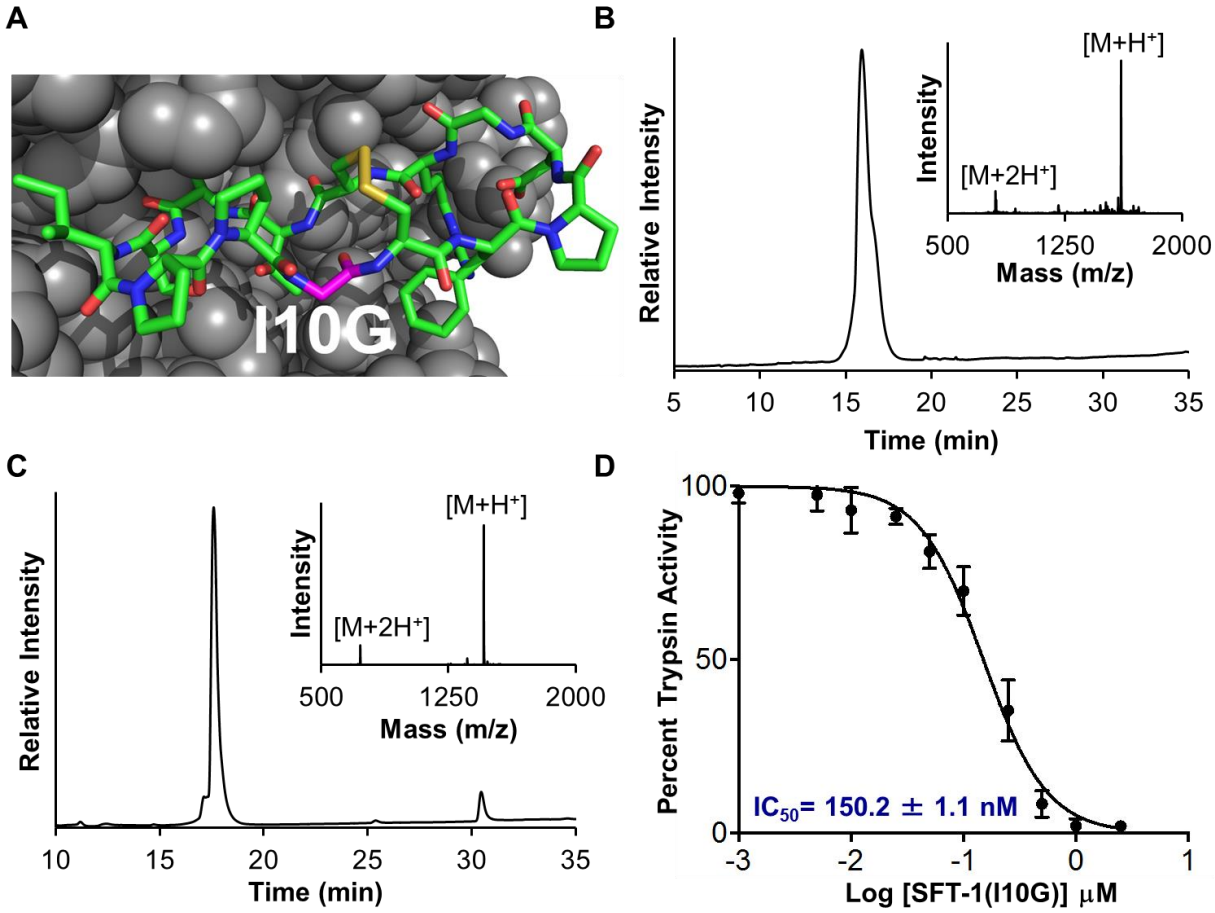


Figure 2.4. One-pot synthesis and trypsin inhibitory activity of SFT-1 (I10G). (A) Model of SFT-1(I10G) (stick representation) based on the X-ray structure of SFT-1 bound to bovine trypsin (gray spheres). PDB code 1SFI. (B) C18 RP-HPLC of SFT-1(I10G)-MEGA peptide. Inset is the ESI-MS of SFT-1(I10G)-MEGA; calcd $[M+H]^+$ 1609.9 Da, obsd 1609.8 ± 0.2 Da. (C) C18 RP-HPLC of cyclized and oxidized SFT-1 (I10G) product. Inset is the ESI-MS of purified SFT-1 (I10G); calcd $[M+H]^+$ 1458.7 Da, obsd 1458.6 ± 0.3 Da. (D) Dose−response curve for the inhibition of bovine trypsin activity by SFT-1 (I10G). RP-HPLC: 0−73% CH_3CN in water, 30 min gradient.

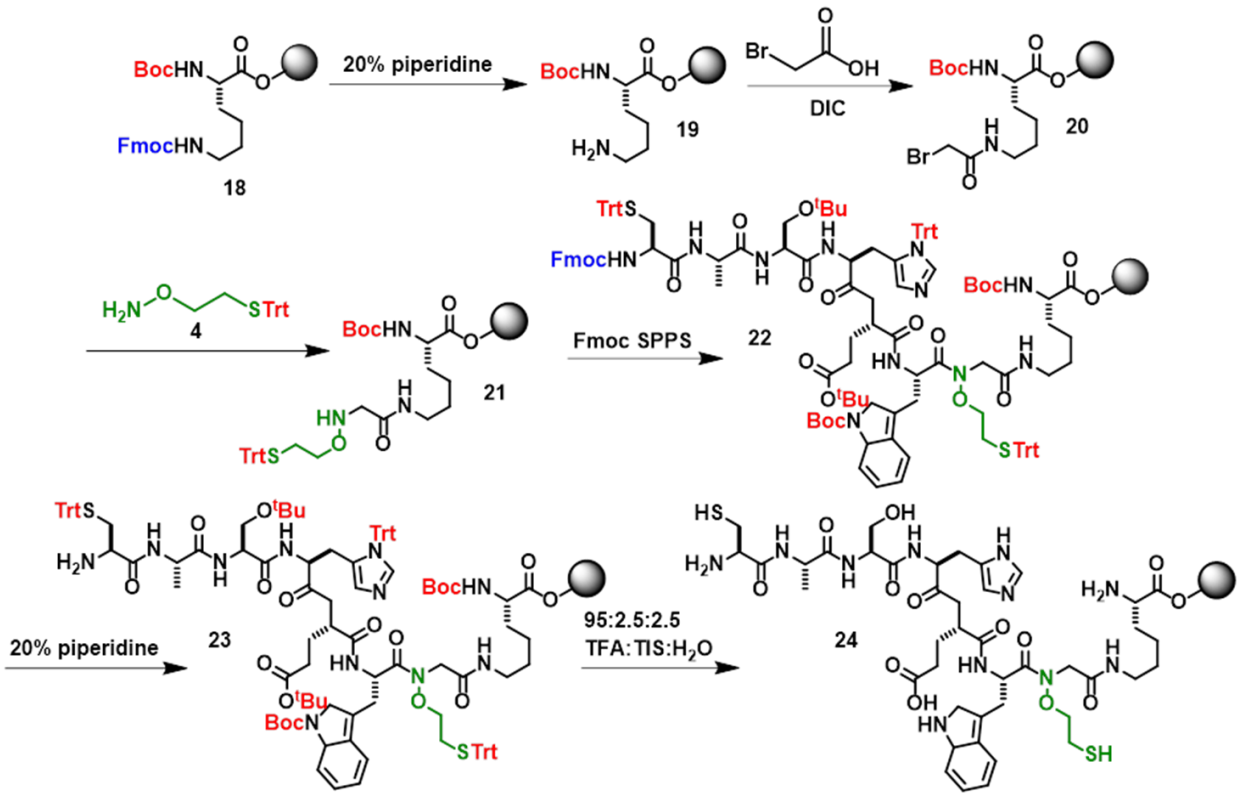
BAPNA to release the yellow colored 4-nitroaniline, which is easily detected by its absorbance at 410 nm. Assays were conducted in a 96-well plate format with 100 nM Trypsin, 500 μM BAPNA, and varying concentration of SFT-1 (I10G). Consistent with previous reports, we

observed a robust dose−response curve with IC_{50} = 150.2 ± 1.1 nM (**Figure 2.4D**). Thus, our Fmoc-compatible and automated approach simplified access to a bioactive cyclic peptide.

To further simplify the MEGA concept, we envisioned that peptide thioesters and cyclic peptides could be prepared directly from the solid phase without intermediate purification of the MEGA peptide. Several previous reports describe completely on-resin techniques for obtaining cyclic peptides with non-native linkages through Asp/Glu side-chain ligation, ring-closing metathesis, and various click chemistries.[3,43–46] However, only a select few strategies describe on-resin techniques for preparing head-to-tail cyclic peptides containing solely native amide linkages and using Fmoc-SPPS-based protocols.[47–49] In one report, Olsen et al utilized the MeDbz linker to optimize conditions for Dbz activation followed by global on-resin side-chain deprotection of N-terminal Cys containing peptides.[49] Incubation of the resulting resin in mixed acetonitrile and phosphate buffer solutions led to intramolecular NCL and release of a cyclic peptide from the solid phase. However, the on-resin cyclization of MeNbz peptides suffers from limitations including low yields and the propensity for cyclic peptide dimerization as a side-product. The dearth of available on-resin techniques and room for improvement on current strategies led us to explore MEGA as an on-resin strategy for α-thioester and cyclic peptide synthesis.

Solid phase preparation must be carefully considered when designing an on-resin thioesterification/cyclization platform. For utmost simplification, the resin must be compatible with Fmoc-SPPS. However, in order to obtain completely deprotected peptides, the resin linkage must be stable to Fmoc-SPPS cleavage conditions (i.e. TFA) to remove the acid-labile support. Hence, we employed a resin typically used in Boc-SPPS, Boc-Lys(ε-Fmoc) 4-(hydroxymethyl)phenylacetamidomethyl (PAM) resin (**18**), to prepare a resin-linked CASHEW-MEGA model peptide cyclization system (**Scheme 2.4**). This resin was chosen because the Fmoc-protected ε-amine of the resin-bound Lys residue could be deprotected and coupled to

activated Fmoc-amino acids, which afforded spacing the elongating peptide chain from the solid support thereby maximizing reactive site accessibility. Additionally, this resin is stable to TFA due to the PAM linkage, and no off-target coupling to the α-amine of Lys from SPPS was expected due to its Boc-protection. Fmoc deprotection of resin **18** was first performed, followed by Lys ε-amine bromoacetylation and subsequent reaction with the MEGA handle **4** to give MEGA PAM resin, **21**. Following standard SPPS of the CASHEW sequence, base-labile and acid-labile protecting groups were sequentially removed to provide the fully deprotected, resin-bound CASHEW-MEGA peptide, **24**. We note that attempts to synthesize the same peptide on 4-methylbenzhydrylamine (MBHA) and aminomethyl polyethyleneglycol (AM-PEGA) resins were unsuccessful. Due to the stability of the peptide-resin linkage to cleavage by TFA, options to monitor the progress of SPPS were limited. Quantitative Fmoc-loading determination (after initial Trp coupling),[50] suggested resin loadings between 0.20-0.24 mmol/g, or ~50% of the reported PAM resin loading (0.4-0.6 mmol/g). The resin-bound MEGA peptide was then subjected to optimized thioesterification conditions for C-terminal Trp (**Table 2.2**), and the production of the cyclic CASHEW peptide was monitored over time via RP-HPLC and ESI-MS (**Figure 2.5A**). We were pleased to observe increasing amounts of released cyclic CASHEW peptide (**27**) through 48 h of incubation. Of particular interest, we observed the build-up of CASHEW-MES thioester **26** at early reaction time-points and successive consumption of the thioester via intramolecular NCL (**Figure 2.5B**).. Additionally, no amount of cyclic peptide was observed in the absence of the external thiol, MESNa. Taken together, these observation led us to regard overall process to be an ‘on-resin thiolysis’ reaction, releasing the peptide thioester from the resin, followed by efficient head-to-tail peptide cyclization (**Figure 2.5B**). Finally, we



Scheme 2.4. Preparation of Lys(ε-CASHEW-MEGA)-PAM resin. Red text = acid-labile protecting groups. Blue text = base-labile protecting groups. Green text = MEGA thioesterification handle.

also demonstrated the ability to form peptide thioesters directly from the resin by synthesizing Lys(ε-ASHEW-MEGA)-PAM resin. By subjecting this material to optimized thioesterfication conditions, we were able to isolate the stable ASHEW-MES thioester (**Figure 2.S16**).

2.2.6. Epimerization studies

Finally, we addressed the potential for C-terminal epimerization during thioesterification. The D-epimers of AWKA-MEGA and AWKC-MEGA were synthesized and converted to their respective thioesters (**Tables 2.1, 2.2, Figures 2.S1B,Y, Figure 2.S17**). Cys was chosen, as it is particularly prone to racemization in its activated form.[51] All diastereomeric thioesters displayed different RP-HPLC retention times upon co-injection, and to our delight, an average of <1% epimerization was observed throughout the time-course of thioesterification. Thus, the MEGA

linker approach leads to minimal epimerization for Ala and Cys, and hence good overall thioesterification yields.

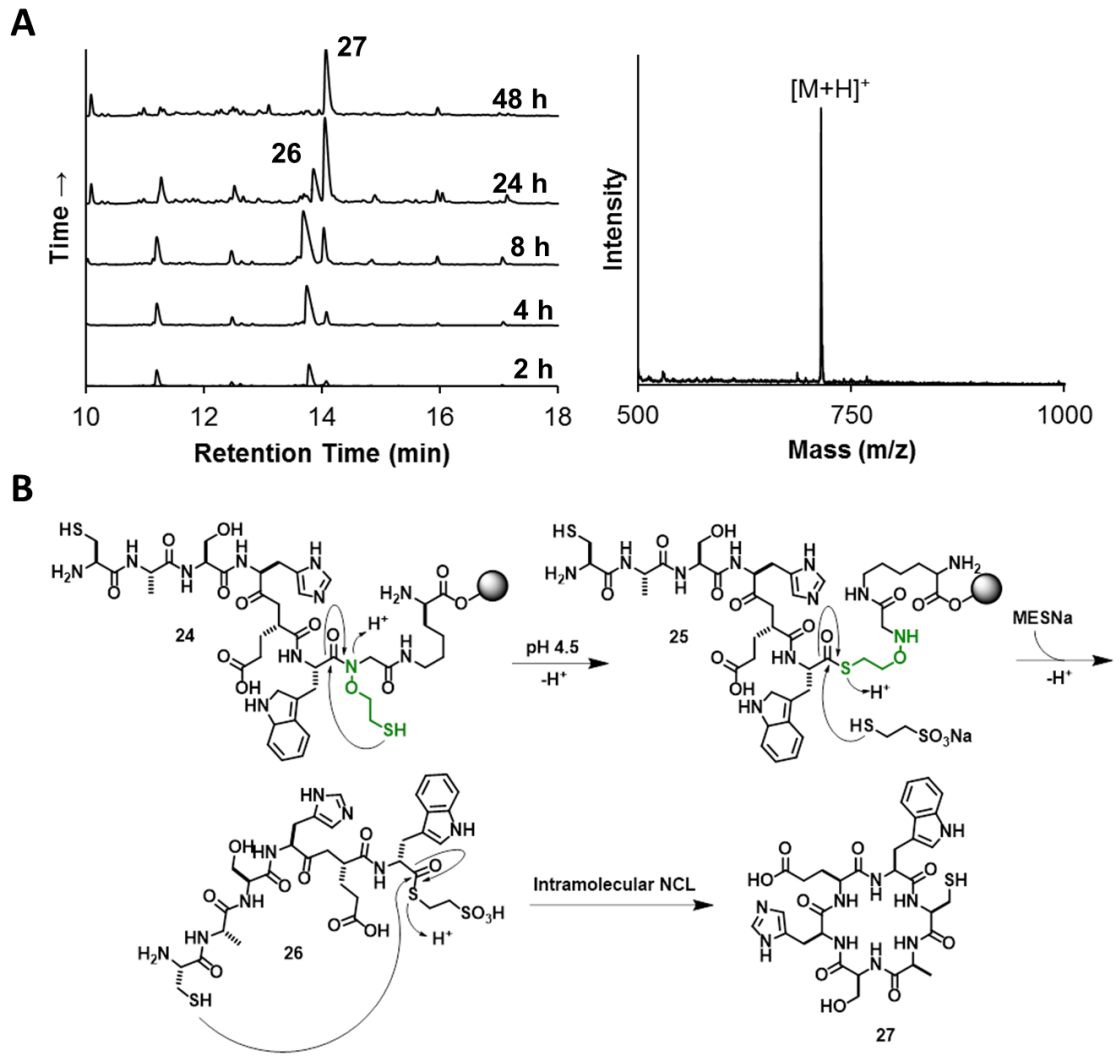


Figure 2.5. On-resin thiolysis and cyclization of CASHEW-MEGA. (A) C18 RP-HPLC time-course of on-resin thiolysis and cyclization (left) and ESI-MS of isolated cyclic CASHEW (right). Calcd. $[M+H]^+$ 713.8 Da, obsd. 714.3 Da. RP-HPLC: 0−73% CH_3CN in water, 30 min gradient. (B) Proposed mechanism of MEGA-mediated on-resin thiolysis and intramolecular NCL.

We also assessed the extent of epimerization during peptide cyclization. C-terminally activated His residues are particularly prone to epimerization due to the proximity of the imidazole side chain to the backbone α-hydrogen.[42] Therefore, we synthesized two epimeric peptides, CLAS(D-H)-MEGA and CLAS(L-H)-MEGA, and measured their degree of epimerization during one-pot cyclization (**Figure 2.S18**). The diastereomeric cyclic CLASH peptides were well-resolved by RP-HPLC (**Figure S19A−D**), and we found that CLAS(L-H)-MEGA epimerized <2% after 24h (**Figure S19E**).

2.3 Conclusions and outlook

In conclusion, we have harnessed the serendipitous discovery that NCL products containing the ligation auxiliary 2-(aminooxy)ethanethiol may undergo thiolysis under specific conditions to develop a straightforward and robust strategy for accessing peptide α-thioesters. The MEGA linker approach, which employs inexpensive and readily available starting materials, affords latent peptide α-thioesters applicable to one-pot inter- and intramolecular NCL reactions. MEGA is compatible with a wide-range of C-terminal amino acids and peptide lengths, and is particularly useful for the synthesis of cyclic peptides. Compatibility with microwave-assisted automated SPPS, demonstrated by our synthesis of the p53 N-terminal 35-mer peptide, and the development of on-resin protocols are especially exciting for future applications of the MEGA linker strategy.

2.4 Experimental procedures

2.4.1 General methods

Rink amide resin (0.30-60 mmol/g substitution) was purchased from Chem-Impex (Wood Dale, IL). Standard Fmoc-L-amino acids were purchased from MilliporeSigma (Burlington, MA) or AnaSpec (Fremont, CA). All other chemical reagents were purchased from Sigma-Aldrich Chemical Company (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). Solid phase peptide synthesis (SPPS) was performed manually, or on a Liberty Blue Automated Microwave Peptide Synthesizer (CEM Corporation, Matthews, NC).[52] Analytical reversed-phase HPLC (RP-HPLC) was performed on a Varian (Palo Alto, CA) ProStar HPLC with a Grace-Vydac (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. Semi-preparative scale purifications were conducted on a Grace-Vydac 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.

2.4.2 Solid-phase peptide synthesis

Synthesis of MEGA resin. Rink amide resin (0.25 mmol, 0.33 mmol/g) was allowed to swell in a reaction vessel in 50:50 (v/v) DMF:DCM for 30 min followed by Fmoc-deprotection with 20% piperidine in DMF for 25 min. The resin was thoroughly washed by consecutive 30 sec DMF, DCM and DMF flow washes. Bromoacetic acid (2.5 mmol) and diisopropylcarbodiimide (DIC, 2.5 mmol) were dissolved in 4 mL DMF and added to the resin. The mixture was agitated with N_2(g) for 45 min, then an additional 45 min with fresh coupling reagents. The resin was thoroughly washed and dried under vacuum. *O*-(2-(tritylthio)ethylhydroxylamine (1.75 mmol) was dissolved in 7 mL of 50:50 (v/v) sieve-dried DMSO:DMF, added to the resin in a 20 mL

scintillation vial and shaken for 24-48 h. The resin was filtered, washed and dried under vacuum to give MEGA resin.

Synthesis of peptide-MEGA – first amino acid coupling. MEGA resin (0.05 mmol) was swelled in 0.5 mL sieve-dried DMF in a 20 mL scintillation vial. 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxid hexafluorophosphate (HATU, 0.49 mmol), Fmoc-amino acid (0.5 mmol) and diisopropylethylamine (DIEA, 0.5 mmol) were dissolved in 2.0 mL dry DMF, added to MEGA resin and mixed for 24 h at 25 °C.

Peptide elongation, cleavage and purification. Standard Fmoc-SPPS protocols were employed to extend the peptide chain.[53] The peptide was cleaved from the resin in by mixing with 95:2.5:2.5 (v/v) TFA:H_2O:triisopropylsilane (TIS) for 1-2 h. The crude peptide was precipitated from solution by mixing with 10 volumes of cold diethyl ether and centrifuged for 2 min at 3,500 rpm. The supernatant was discarded and the peptide lyophilized. Preparative scale RP-HPLC was used to purify the crude peptide and provide pure peptide-MEGA after lyophilization.

Microwave-assisted synthesis of MEGA peptides. The MEGA resin was loaded with the amino acid of choice as described above for the first amino acid coupling. The pre-loaded MEGA resin was place in a Liberty Blue peptide synthesizer (CEM) and automated peptide synthesis was conducted with 50 °C deprotection steps using a 5% piperazine/0.1 M hydroxybenzotriazole (HOBt) solution in DMF. 50 °C amino acid couplings were used for all coupling steps.[36]

2.4.3 MEGA peptide α-thioesterification, ligation and cyclization

MEGA peptide α-thioesterifications. Peptide-MEGA (0.5 μmol) was dissolved in 500 μL thioesterification buffer consisting of sodium 2-mercaptoethanesulfonate (MESNa, 200-400

mM), sodium phosphate (NaH_2PO4, 100 mM) and tris(2-carboxyethyl)phosphine (TCEP, 25-50 mM) at pH 4-6. The reactions were agitated for 24-72 h at 37-70 °C except where noted in **Table 2.2**. The peptide-MES α-thioester was purified by analytical or semi-preparative scale RP-HPLC.

MEGA peptide ligations. Peptide-MEGA thioesterification was performed as described above. The N-terminal Cys peptide, CASW (1.25 μmol), was dissolved in 50 μL NaH_2PO4 (200 mM), TCEP (400 mM) buffer and added directly to the thioesterification reaction vessel. The solution pH was adjusted to 7.5 by litmus and the reaction mixed for 8-24 h. The reaction was analyzed by RP-HPLC after reduction of the assay mixture with additional TCEP (25-50 mM final concentration).

MEGA peptide cyclizations. N-terminal Cys peptide (0.5 μmol) was dissolved in MESNa (200-400 mM), NaH_2PO4 (100 mM) and TCEP (25-50 mM) buffer at pH 4-6. Reactions were allowed to proceed for 8-72 h at 50-70 °C. Reaction mixtures were analyzed by C18 analytical RP-HPLC following prior treatment with additional TCEP (25-50 mM final concentration).

2.4.4 Synthesis of cyclized and oxidized SFT-1(I10G)

Synthesis of SFT-1(I10G). The peptide sequence CFPDGRCTKSIPPG-MEGA was prepared on Rink amide resin (0.05 mmol) via automated peptide synthesis as described above. The crude peptide was cleaved from the resin in a 95:2.5:2.5 (v/v) TFA:H_2O:TIS solution and precipitated with 10 volumes of cold diethyl ether and lyophilized. The linear peptide was purified by C18 preparative RP-HPLC (15-40% B, 60 min) and lyophilized (15.4 mg, 20%).

SFT-1(10G) cyclization. SFT-1(I10G)-MEGA (5 mg, 3.1 μmol) was cyclized as described above at 70 °C for 24 h. The cyclic peptide was purified by C18 semi-preparative RP-HPLC (10-50% B, 45 min gradient) (1.5 mg, 33%).

Oxidation of Cyclic SFT-1(I10G). Cyclic SFT-1(I10G) was oxidized to form the disulfide bridge by incubating the peptide in 100 mM $(NH_4)HCO_3$ at 25 °C overnight (0.2mg/mL). SFT-1 (I10G) was purified by C18 analytical RP-HPLC (10-50% B, 30 min gradient) with a quantitative yield.

Trypsin inhibition assays. The inhibition of bovine trypsin by the Sunflower Trypsin Inhibitor-1 analog, SFT-1(I10G), was measured spectrophotometrically. The hydrolysis of N(α)-benzoyl-L-arginine 4-nitroanilide (BAPNA) by bovine trypsin to generate the yellow colored 4-nitroaniline was followed at 410 nm in the presence of varying concentrations of SFT-1(I10G).[42] Briefly, 500 μM BAPNA and 1 nM to 2.5 μM SFT-1(I10G) were incubated for 5 min in a 96-well plate in the presence of assay buffer consisting of 50 mM Tris and 20 mM $CaCl_2$ at pH 8.0. BAPNA hydrolysis was initiated by the addition of 100 nM trypsin and the reaction allowed to proceed for 20 min at 27 °C. The final Abs_{410} was measured using a BioTek Synergy 4 microplate reader and plotted in GraphPad Prism. All experiments were undertaken in triplicate and the average value reported with error being the standard deviation from the mean.

2.4.5 On-resin cyclization

Synthesis of Lys(εCASHEW-MEGA)-PAM resin. Boc-Lys(ε-Fmoc) 4-(hydroxymethyl) phenylacetamidomethyl resin (~0.1 mmol, 0.5-0.6 mmol/g) was swelled in 50:50 DMF:DCM for 30 min followed by Lys ε-Fmoc deprotection with 20% piperidine in DMF for 25 min. MEGA resin was prepared from the free Lys ε-amine as described above. First amino acid coupling and peptide chain elongation procedures were performed as described above to prepare Boc-Lys(εCASHEW-MEGA)-PAM. Resin loading after initial Fmoc-Trp-OH coupling indicated 0.24

mmol/g substitution via dibenzofulvene absorbance at 305 nm after Fmoc deprotection. After SPPS, final Fmoc deprotection was performed with 20% piperidine in DMF for 25 min. The resin was washed thoroughly with DMF and DCM and transferred to a scintillation vial. Acid labile side-chains were then removed by incubating the resin with 10 vol. of 92.5:2.5:2.5:2.5 TFA:TIS:H_2O:ethane dithiol for 2 h. Resin collected and thoroughly washed with DMF and DCM and then dried on the lyophilizer.

On-resin CASHEW-MEGA cyclization. Lys(εCASHEW-MEGA)-PAM (5 μmol, 25 mg), was incubated in 400 μL buffer containing 100 mM NaH_2PO_4, 400 mM MESNa, 25 mM TCEP, pH 5.6 at 70 °C for 48 h. 20 μL aliquots were analyzed via C18 analytical RP-HPLC (0-73% B, 30 min gradient) at 2, 4, 8, 24 and 48 h with prior reduction using TCEP (final concentration 50 mmol).

2.5 Product characterization and supplemental data

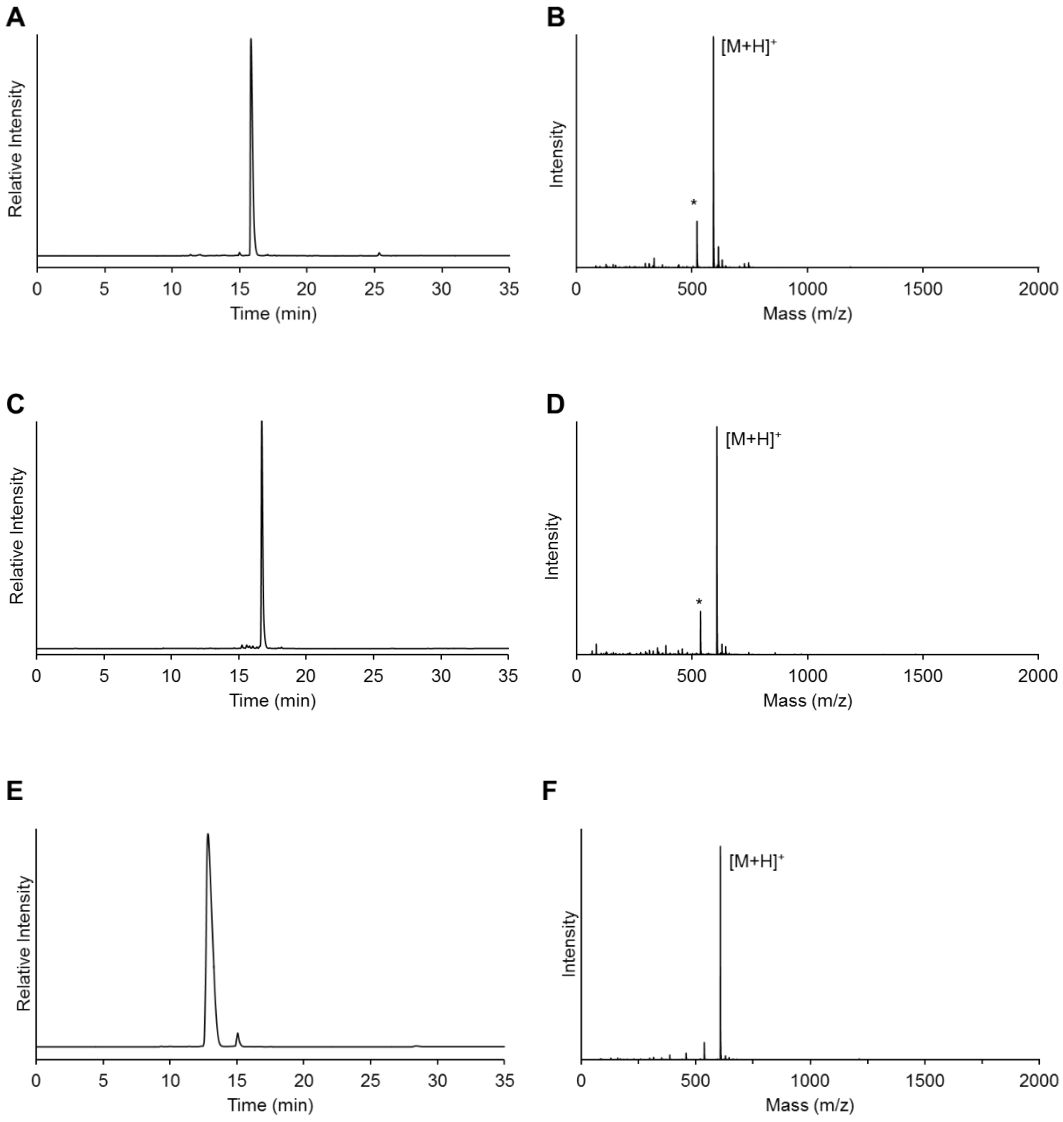
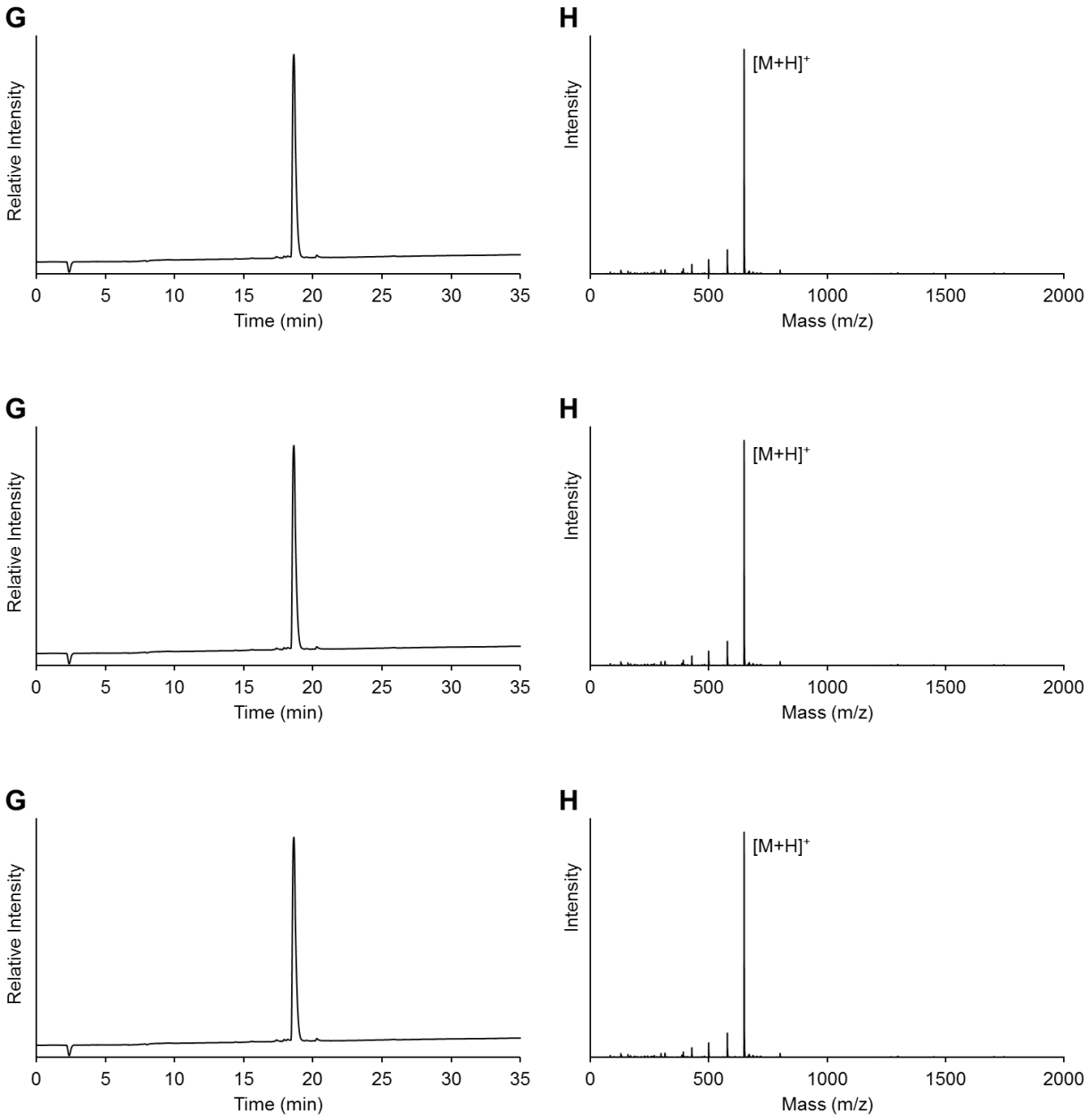
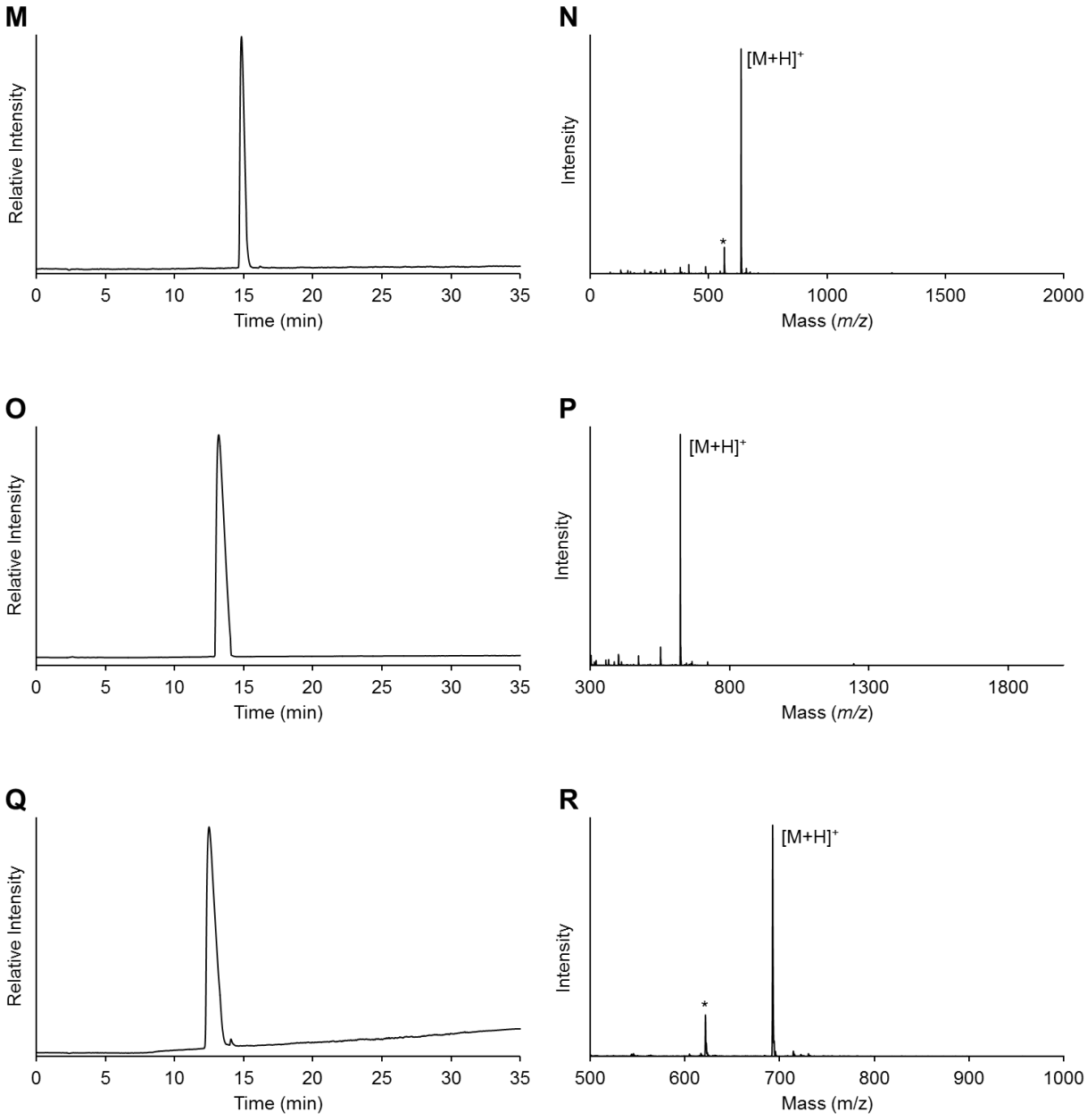


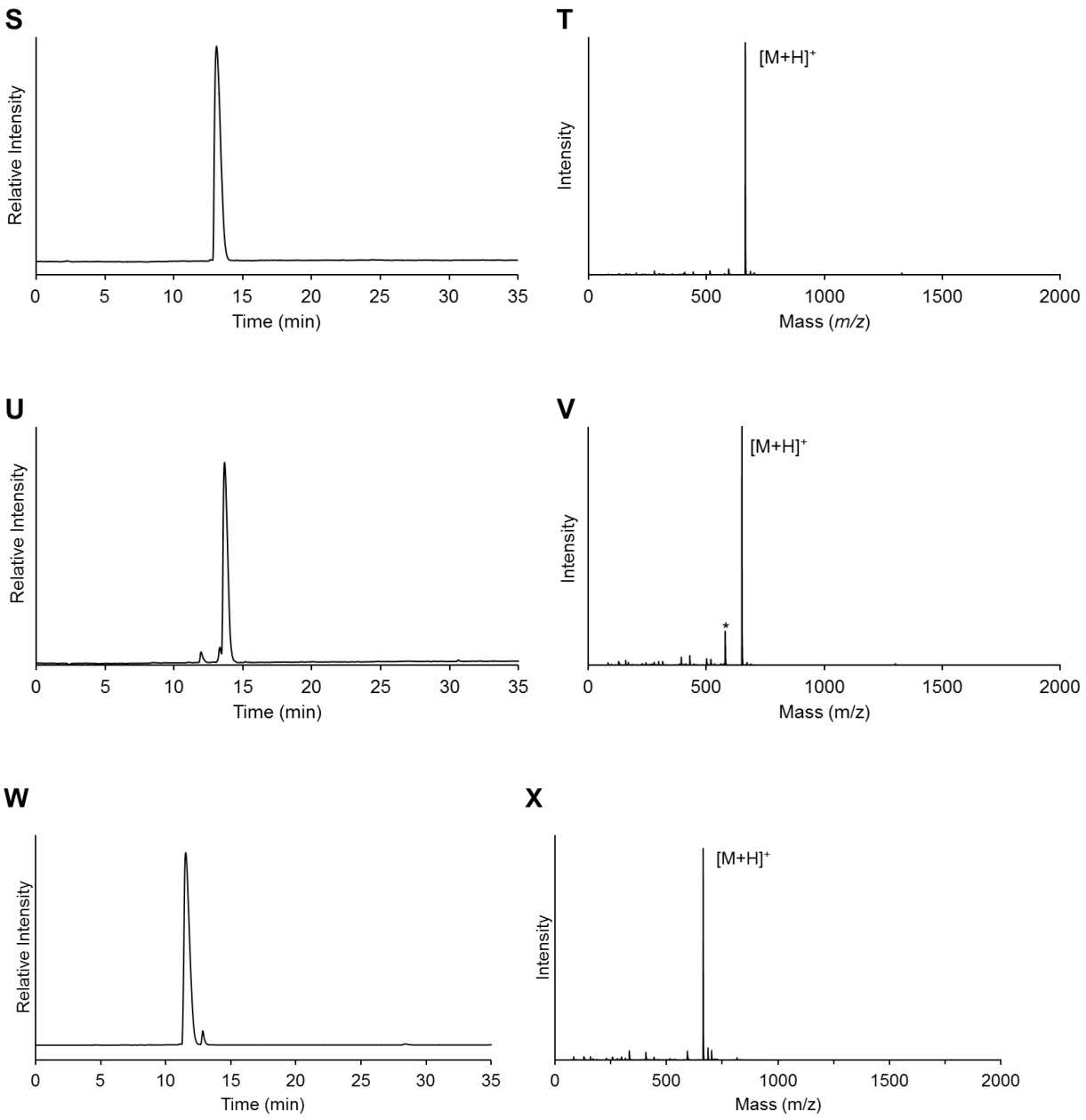
Figure 2.S1. AWKX-MEGA characterization. (A) Purified AWK**G**-MEGA. (B) ESI-MS of purified AWK**G**-MEGA. Calcd. $[M+H]^+$ 592.7 Da, obsd. 592.3 Da. (C) Purified AWK**A**-MEGA. (D) ESI-MS of purified AWK**A**-MEGA. Calcd. $[M+H]^+$ 606.7 Da, obsd. 606.5 Da. (E) Purified AWK**(D-A)**-MEGA (F) ESI-MS of purified AWK**(D-A)**-MEGA. Calcd. $[M+H]^+$ 607.7 Da, obsd. 607.4 Da. * = MS-fragmentation of full-length peptide corresponding to loss of N-terminal Ala. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min gradient.



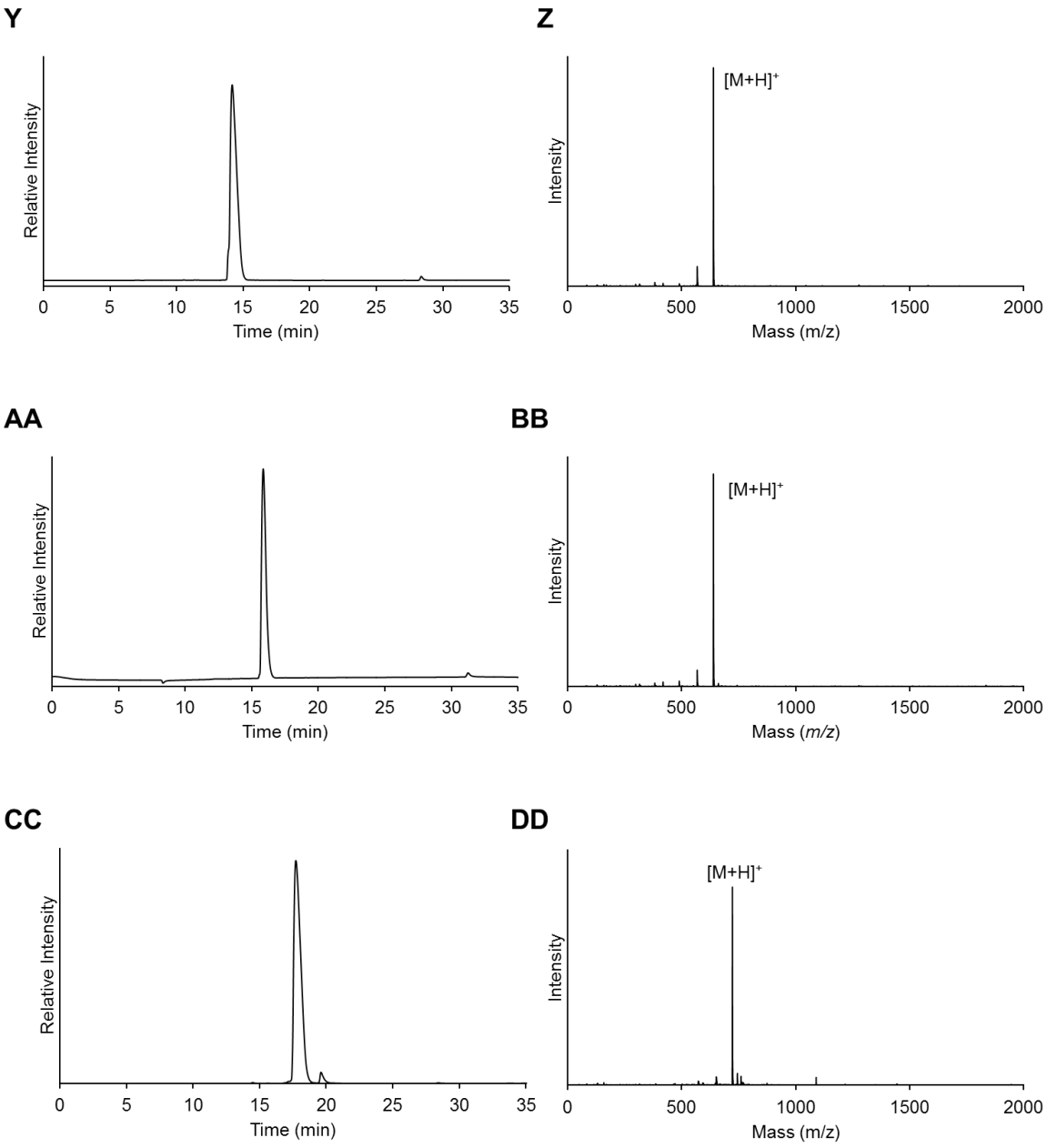
Continued Figure 2.S1. AWKX-MEGA characterization. (G) Purified AWK**L**-MEGA. (H) ESI-MS of purified AWK**L**-MEGA. Calcd. $[M+H]^+$ 648.8 Da, obsd. 648.5 Da. (I) Purified AWK**F**-MEGA. (J) ESI-MS of purified AWK**F**-MEGA. Calcd. $[M+H]^+$ 682.8 Da, obsd. 682.6 Da. (K) Purified AWK**V**-MEGA. (L) ESI-MS of purified AWK**V**-MEGA. Calcd. $[M+H]^+$ 634.8 Da, obsd. 634.4 Da. RP-HPLC spectra performed on C18 analytical column, 0-73% B, 30 min gradient.



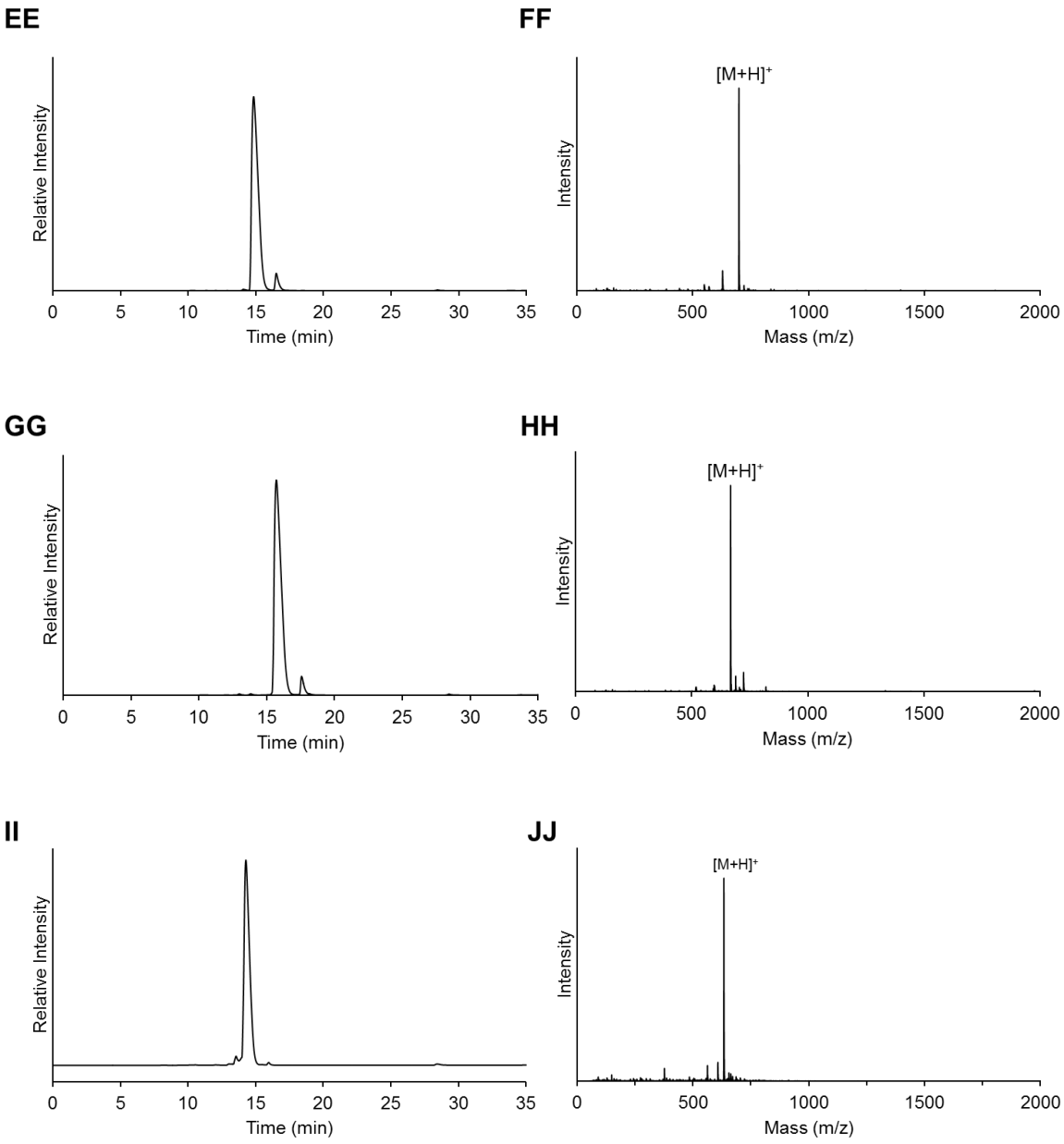
Continued Figure 2.S1. AWKX-MEGA characterization. (M) Purified AWK**T**-MEGA. (N) ESI-MS of purified AWK**T**-MEGA. Calcd. $[M+H]^+$ 636.8 Da, obsd. 636.5 Da. (O) Purified AWK**S**-MEGA. (P) ESI-MS of purified AWK**S**-MEGA. Calcd. $[M+H]^+$ 622.7 Da. obsd. 622.6 Da. (Q) Purified AWK**R**-MEGA. (R) ESI-MS of purified AWK**R**-MEGA. Calcd. $[M+H]^+$ 691.9 Da, obsd. 691.5 Da. * = MS-fragmentation of full-length peptide corresponding to loss of N-terminal Ala. RP-HPLC spectra performed on C18 analytical column, 0-73% B, 30 min gradient.



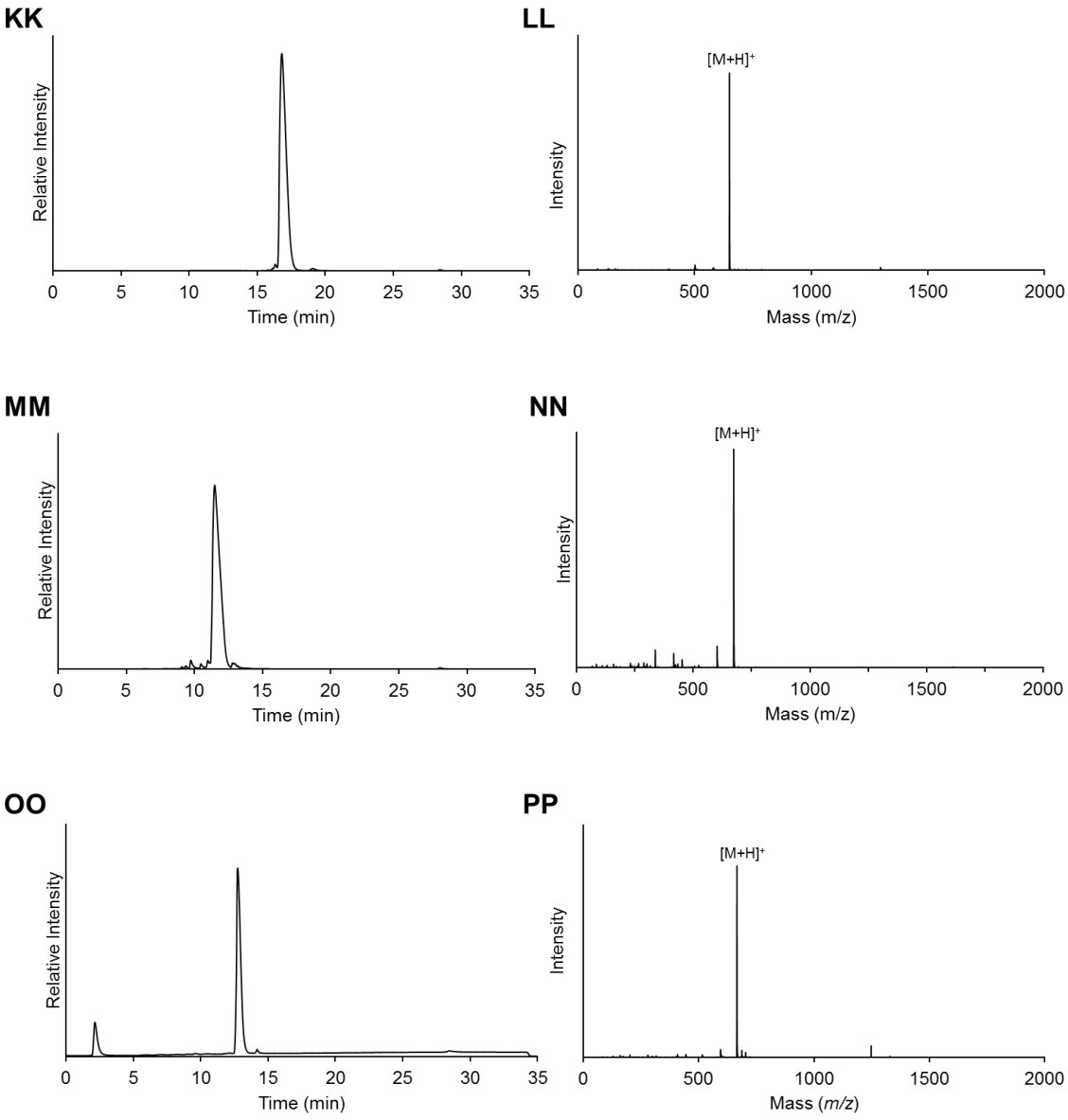
Continued Figure 2.S1. AWKX-MEGA characterization. (S) Purified AWK**Q**-MEGA. (T) ESI-MS of purified AWK**Q**-MEGA. Calcd. $[M+H]^+$ 663.8 Da, obsd. 663.8 Da. (U) Purified AWK**D**-MEGA. (V) ESI-MS of purified AWK**D**-MEGA. Calcd. $[M+H]^+$ 650.8 Da, obsd. 650.4 Da. (W) Purified AWK**K**-MEGA. (X) ESI-MS of purified AWK**K**-MEGA. Calcd. $[M+H]^+$ 664.8 Da, obsd. 664.6 Da * = MS-fragmentation of full-length peptide corresponding to loss of N-terminal Ala. RP-HPLC spectra performed on C18 analytical column, 0-73% B, 30 min gradient.



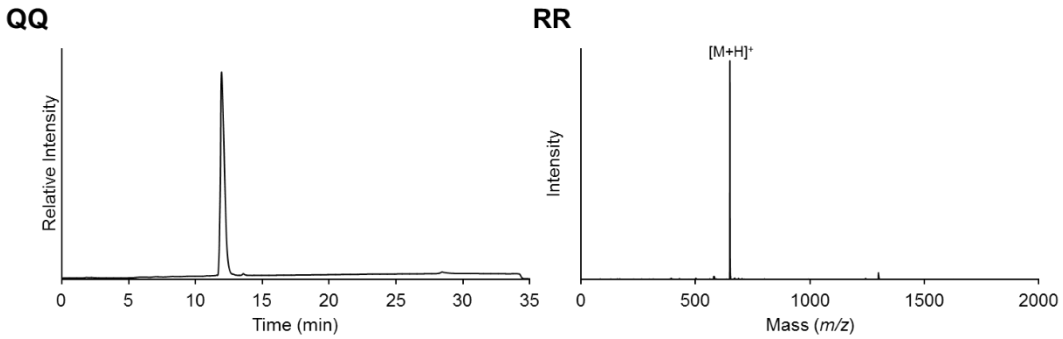
Continued Figure 2.S1. AWKX-MEGA characterization. (Y) Purified AWK**C**-MEGA. (Z) ESI-MS of purified AWK**C**-MEGA. Calcd. $[M+H]^+$ 639.8 Da, obsd. 639.5 Da. (AA) Purified AWK**(D-C)**-MEGA. (BB) ESI-MS of purified AWK AWK**(D-C)**-MEGA. Calcd. $[M+H]^+$ 639.8 Da, obsd. 639.5 Da. (CC) Purified AWK**W**-MEGA (DD) ESI-MS of purified AWK**W**-MEGA. Calcd. $[M+H]^+$ 722.9 Da, obsd. 722.8 Da. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min.



Continued Figure 2.S1. AWKX-MEGA characterization. (EE) Purified AWK**Y**-MEGA. (FF) ESI-MS of purified AWK**Y**-MEGA. Calcd. $[M+H]^+$ 699.8 Da, obsd. 699.6 Da. (GG) Purified AWK**M-**MEGA. (HH) ESI-MS of purified AWK AWK**M-**MEGA. Calcd. $[M+H]^+$ 667.9 Da, obsd. 667.7 Da. (II) Purified AWK**P**-MEGA. (JJ) ESI-MS of purified AWK**P**-MEGA. Calcd. $[M+H]^+$ 632.8 Da, obsd. 634.0 Da. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min.



Continued Figure 2.S1. AWKX-MEGA characterization. (KK) Purified AWK**I**-MEGA. (LL) ESI-MS of purified AWK**I**-MEGA. Calcd. $[M+H]^+$ 648.8 Da, obsd. 649.5 Da. (MM) Purified AWK**H-**MEGA. (NN) ESI-MS of purified AWH AWK**H-**MEGA. Calcd. $[M+H]^+$ 672.8 Da, obsd. 673.9 Da. (OO) Purified AWK**E**-MEGA. (PP) ESI-MS of purified AWK**E**-MEGA. Calcd. $[M+H]^+$ 664.8 Da, obsd. 665.4 Da. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min.



Continued Figure 2.S1. AWKX-MEGA characterization. (QQ) Purified AWK**N**-MEGA. (RR) ESI-MS of purified AWK**N**-MEGA. Calcd. $[M+H]^+$ 649.7 Da, obsd. 651.4 Da. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min.

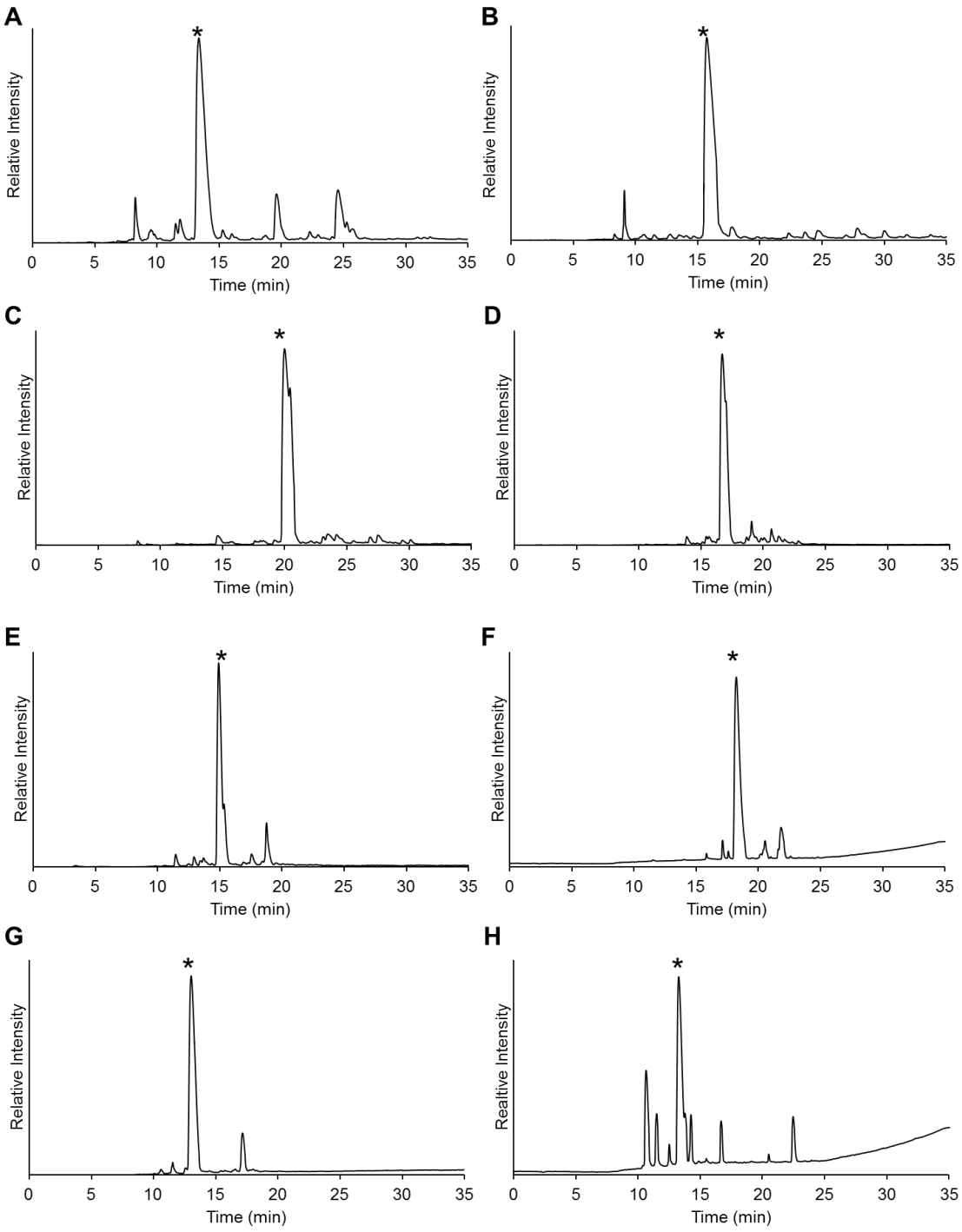
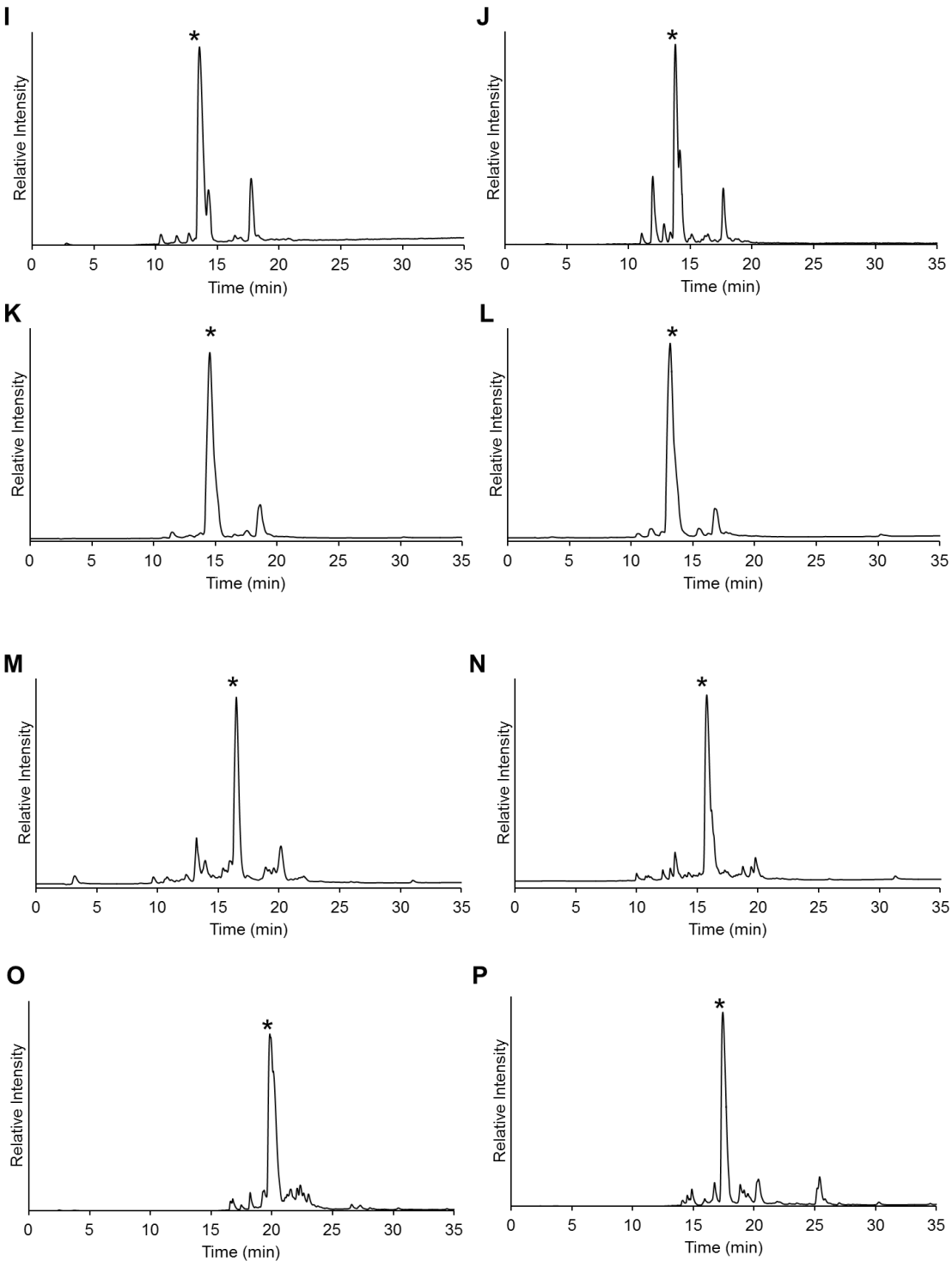
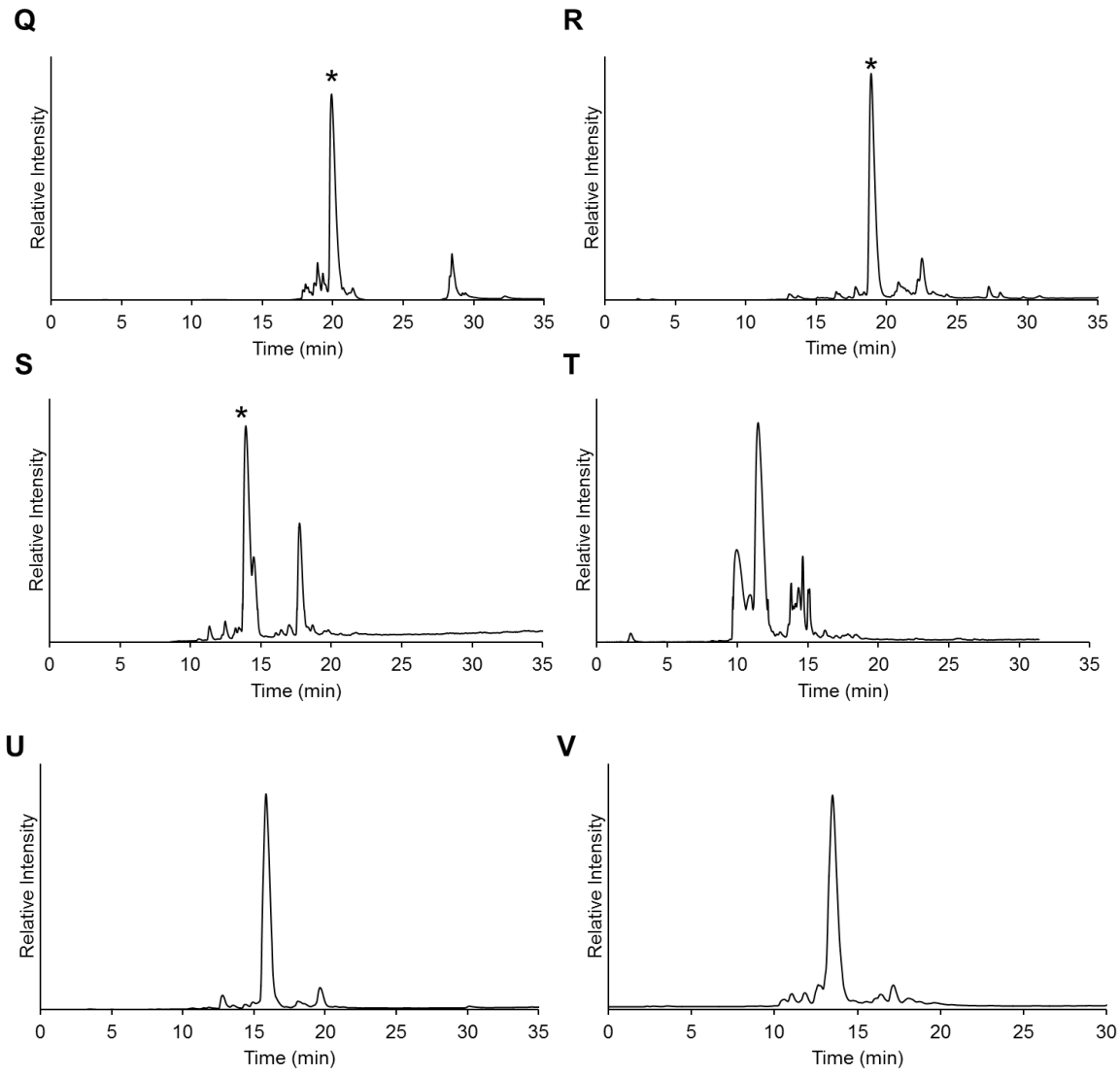


Figure 2.S2. RP-HPLC spectra of crude AWKX-MEGA peptides after TFA-cleavage from resin. (A) AWK**G**-MEGA (B) AWK**A**-MEGA (C) AWK**L**-MEGA (D) AWK**V**-MEGA (E) AWK**D**-MEGA (F) AWK**F**-MEGA (G) AWK**Q**-MEGA (H) AWK**R**-MEGA. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min gradient. * = AWK**X**-MEGA peptide.



Continued Figure 2.S2. RP-HPLC spectra of crude AWKX-MEGA peptides after TFA-cleavage from resin. (I) AWK**S**-MEGA (J) AWK**T**-MEGA (K) AWK**(D-A)**-MEGA (L) AWK**K**-MEGA (M) AWK**C**-MEGA (N) AWK**(D-C)-**MEGA (O) AWKW-MEGA (P) AWKY-MEGA. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min gradient. * = AWK**X**-MEGA peptide.



Continued Figure 2.S2. RP-HPLC spectra of crude AWKX-MEGA peptides after TFA-cleavage from resin. (Q) AWK**M**-MEGA (R) AWK**I**-MEGA (S) AWK**E**-MEGA (T) AWK**N**-MEGA (U) AWK**P**-MEGA (V) AWK**H**-MEGA.RP-HPLC performed on C18 analytical column, 0-73% B, 30 min gradient. * = AWK**X**-MEGA peptide.

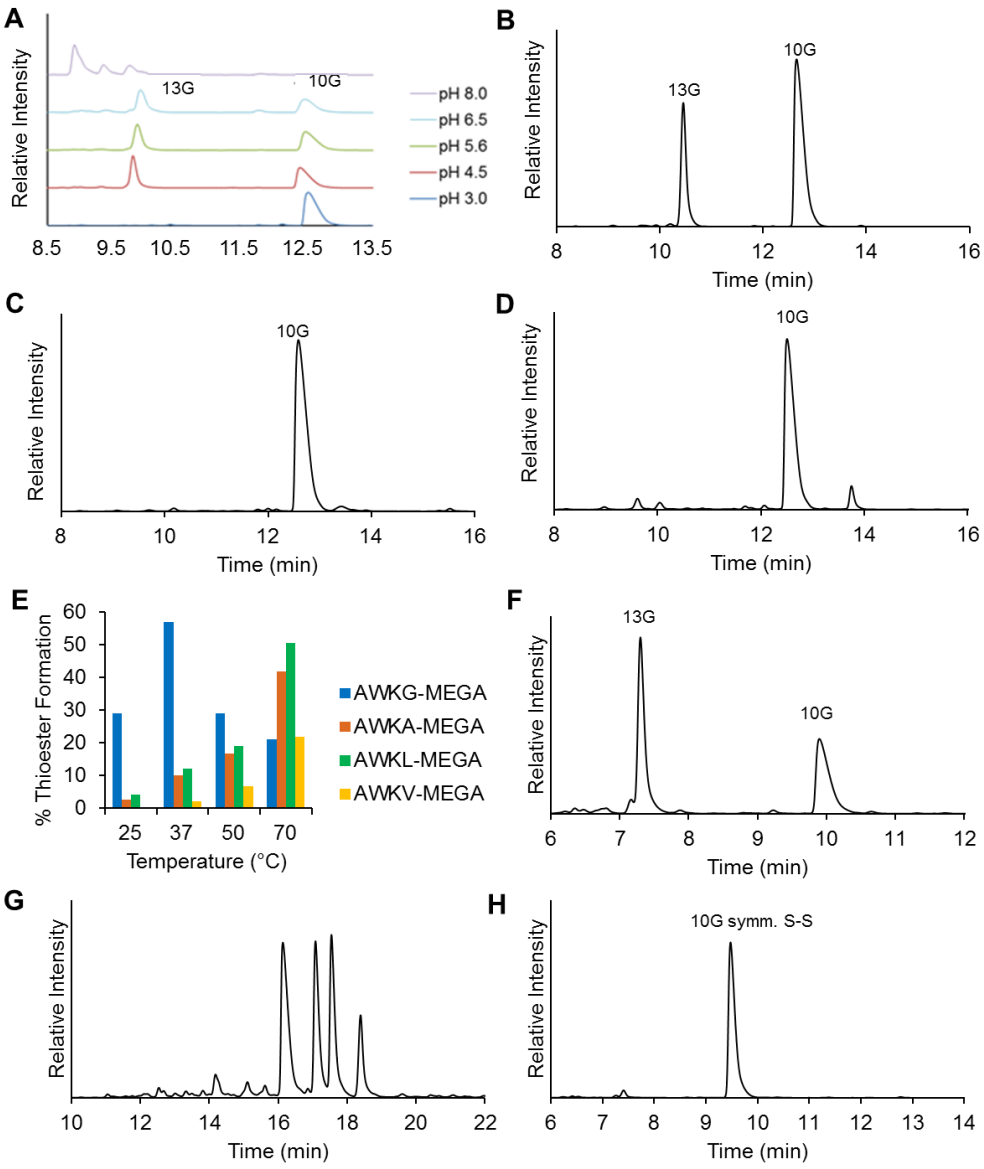


Figure 2.S3. Optimization of thioesterification conditions with AWKG-MEGA. (A) 24 h RP-HPLC time-points of AWK**G**-MEGA thioesterification with MESNa at varying pH and 25 °C. (B) 24 h RP-HPLC time-point of AWK**G**-MEGA thioesterification, 100 mM MESNa and 100 mM mercaptophenylacetic acid, pH 5.6, 25 °C. (C) 24 h RP-HPLC time-point of AWK**G**-MEGA thioesterification, 200 mM 2,2,2-trifluoroethanethiol, pH 5.6, 25 °C. (D) 24 h RP-HPLC time-point of AWK**G**-MEGA thioesterification, 200 mM 3-mercaptopropionic acid, pH 5.6, 25 °C. (E) Histogram shows the percentage of AWK**X**-MES thioester formed at 24 h at various temperatures and pH 5.6 for the given AWK**X**-MEGA peptides. Percent thioester formation was determined by RP-HPLC peak integration at 280 nm. (F) 24 h RP-HPLC time-point of AWKG-MEGA thioesterification under optimized condition with 10% (v/v) MeCN. (G) 24 h RP-HPLC time-point of AWKG-MEGA thioesterification under optimized condition with 10% (v/v) DMF. (H) 24 h RP-HPLC time-point of AWKG-MEGA thioesterification under optimized condition with 10% (v/v) DMSO.

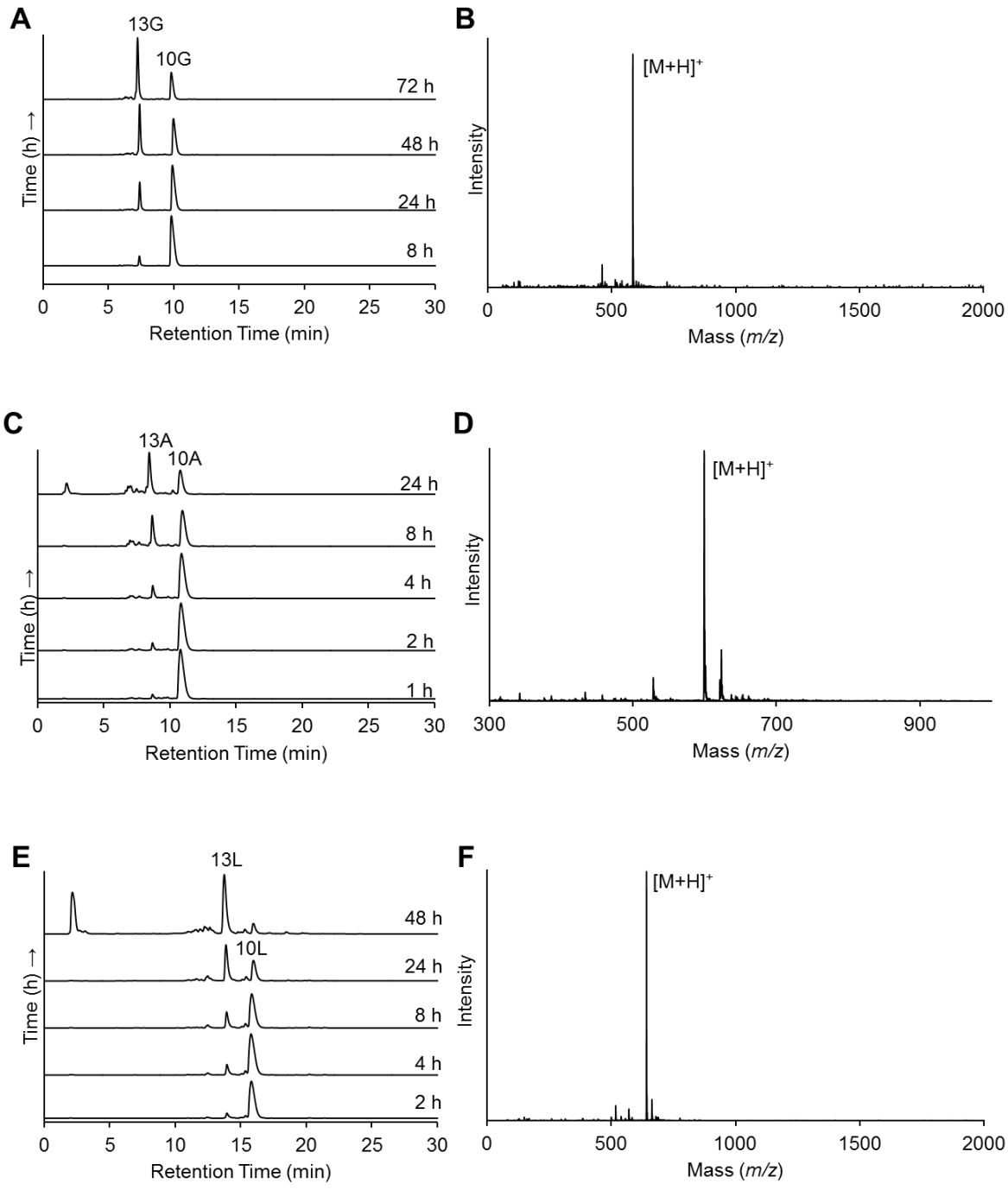
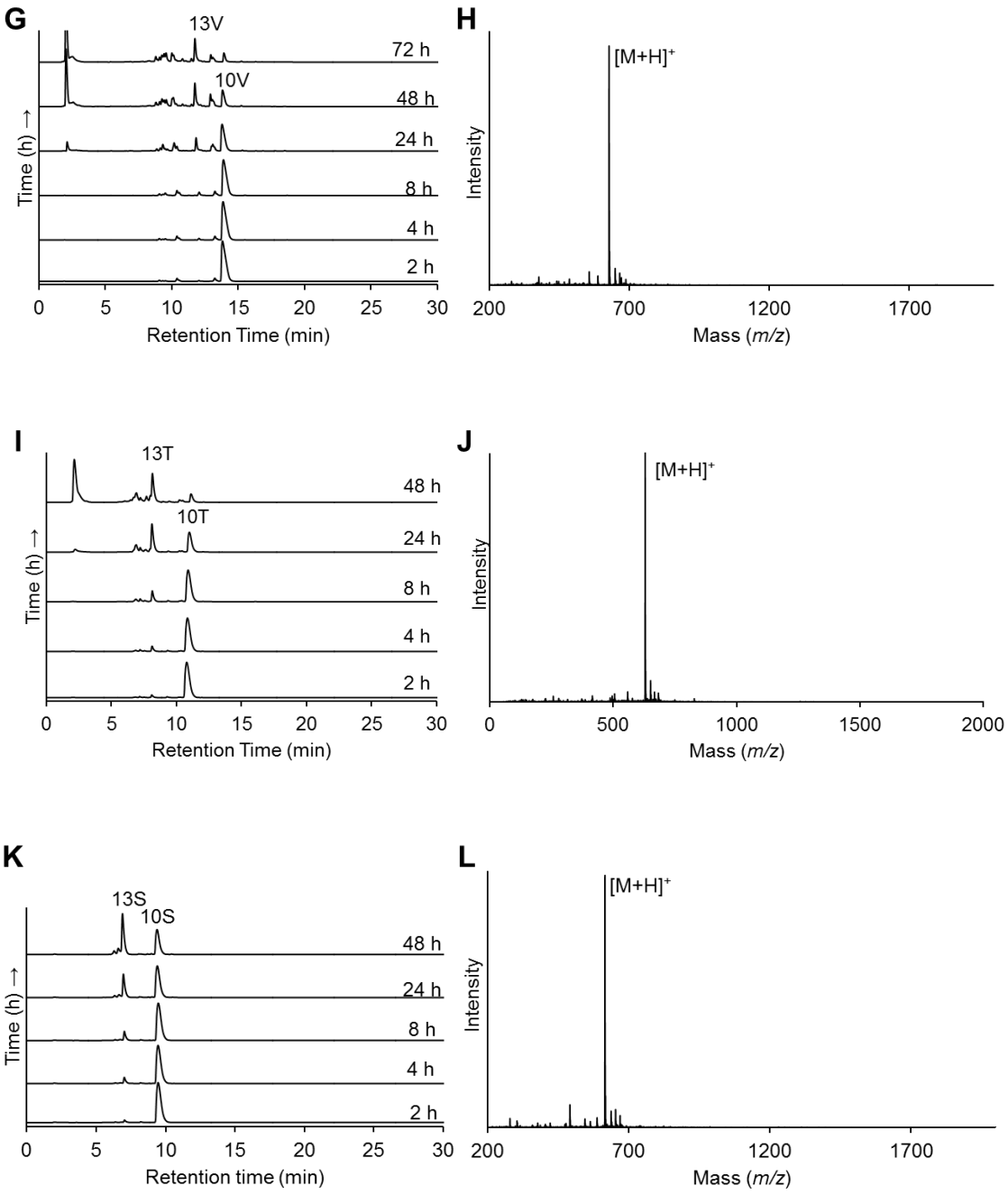
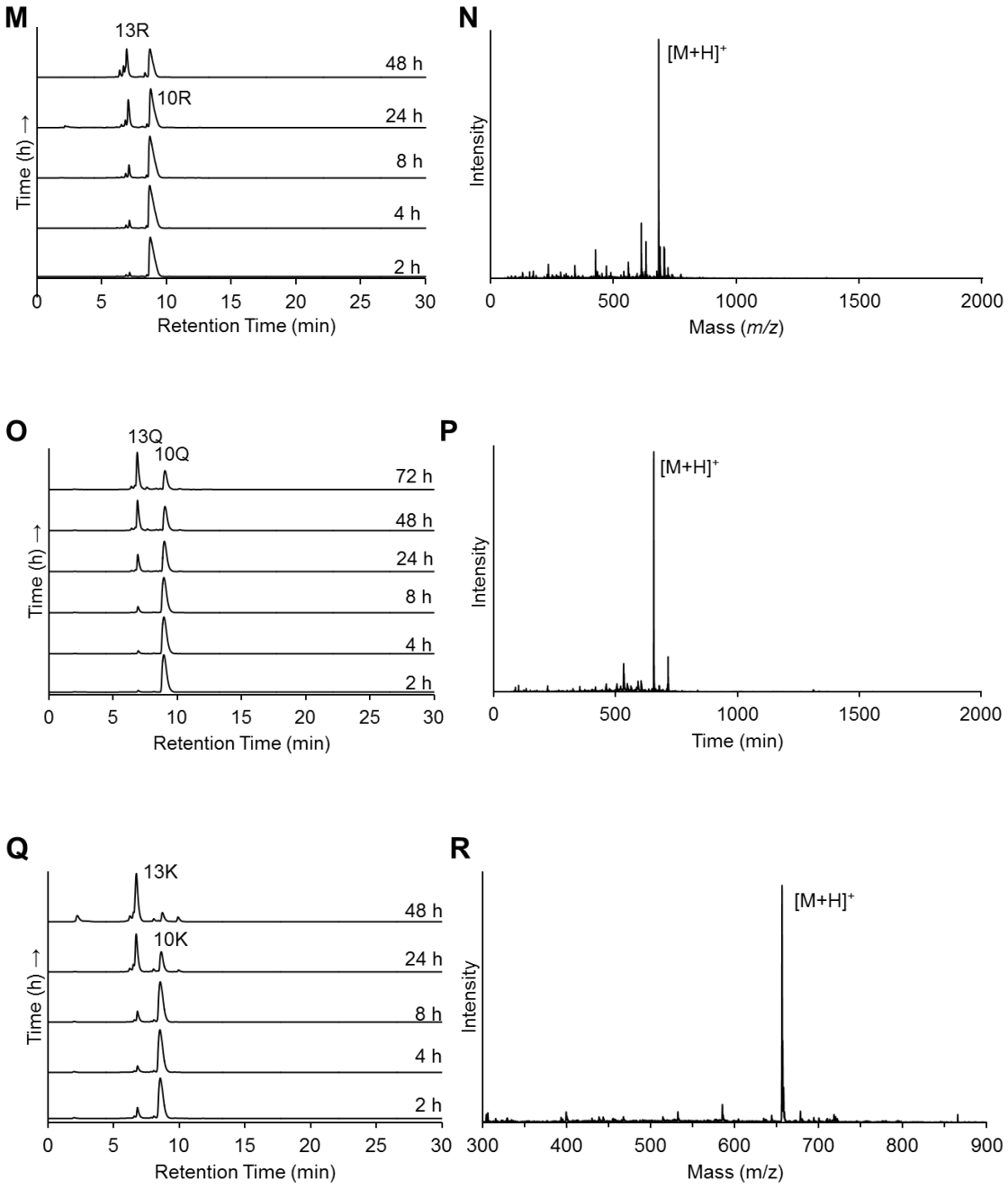


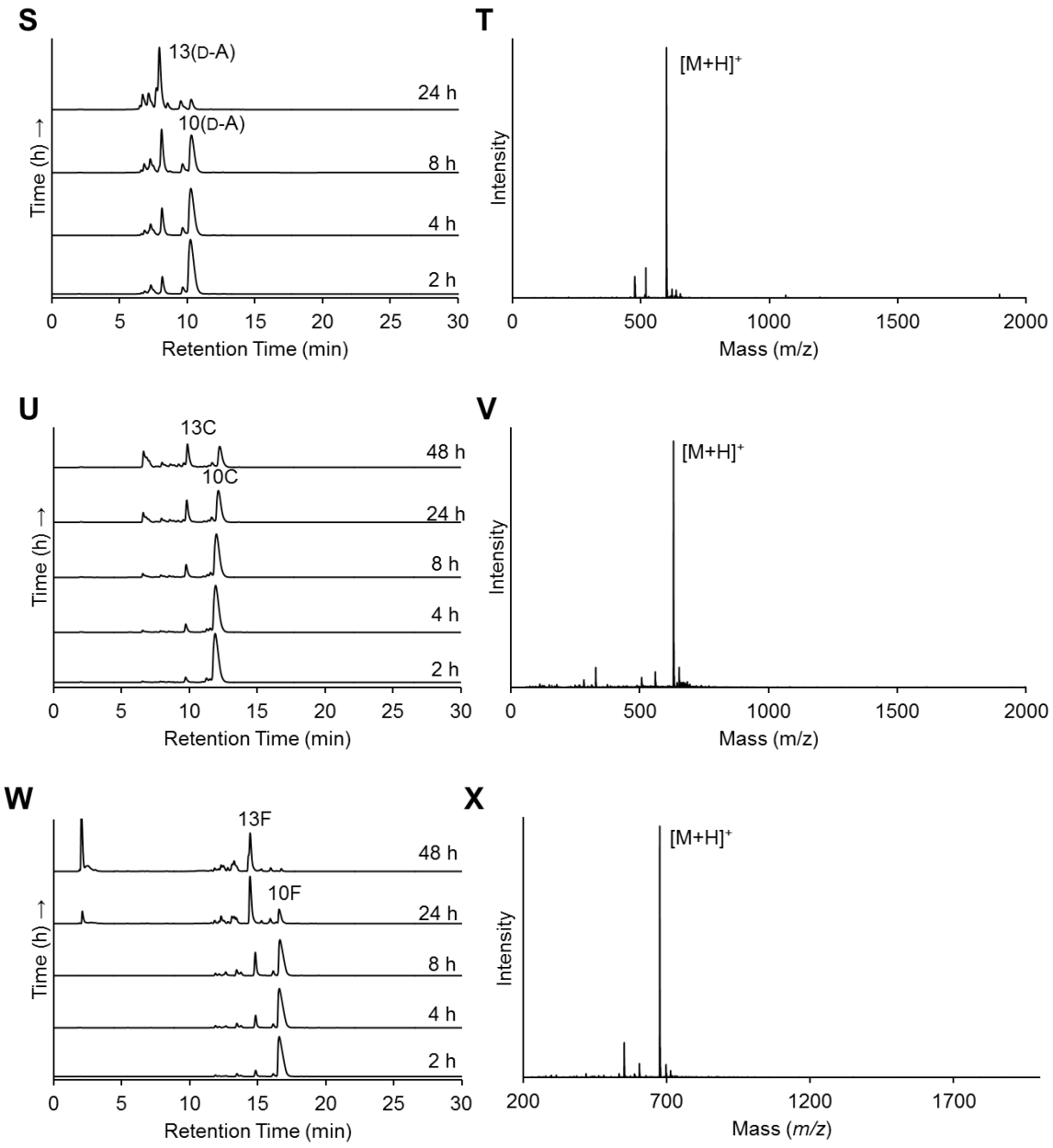
Figure 2.S4. AWKX-MEGA thioesterification. (A) AWK**G**-MEGA thioesterification time-course. (B) ESI-MS of AWK**G**-MES thioester. Calcd. $[M+H]^+$ 584.7 Da, obsd. 584.1 Da. (C) AWK**A**-MEGA thioesterification time-course. (D) ESI-MS of AWK**A**-MES thioester. Calcd. $[M+H]^+$ 598.7 Da, obsd. 598.3 Da. (E) AWK**L**-MEGA thioesterification time-course. (F) ESI-MS of AWK**L**-MES thioester. Calcd. $[M+H]^+$ 640.8 Da, obsd. 640.4 Da. **10** = AWK**X**-MEGA, **13** = AWK**X**-MES thioester. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.



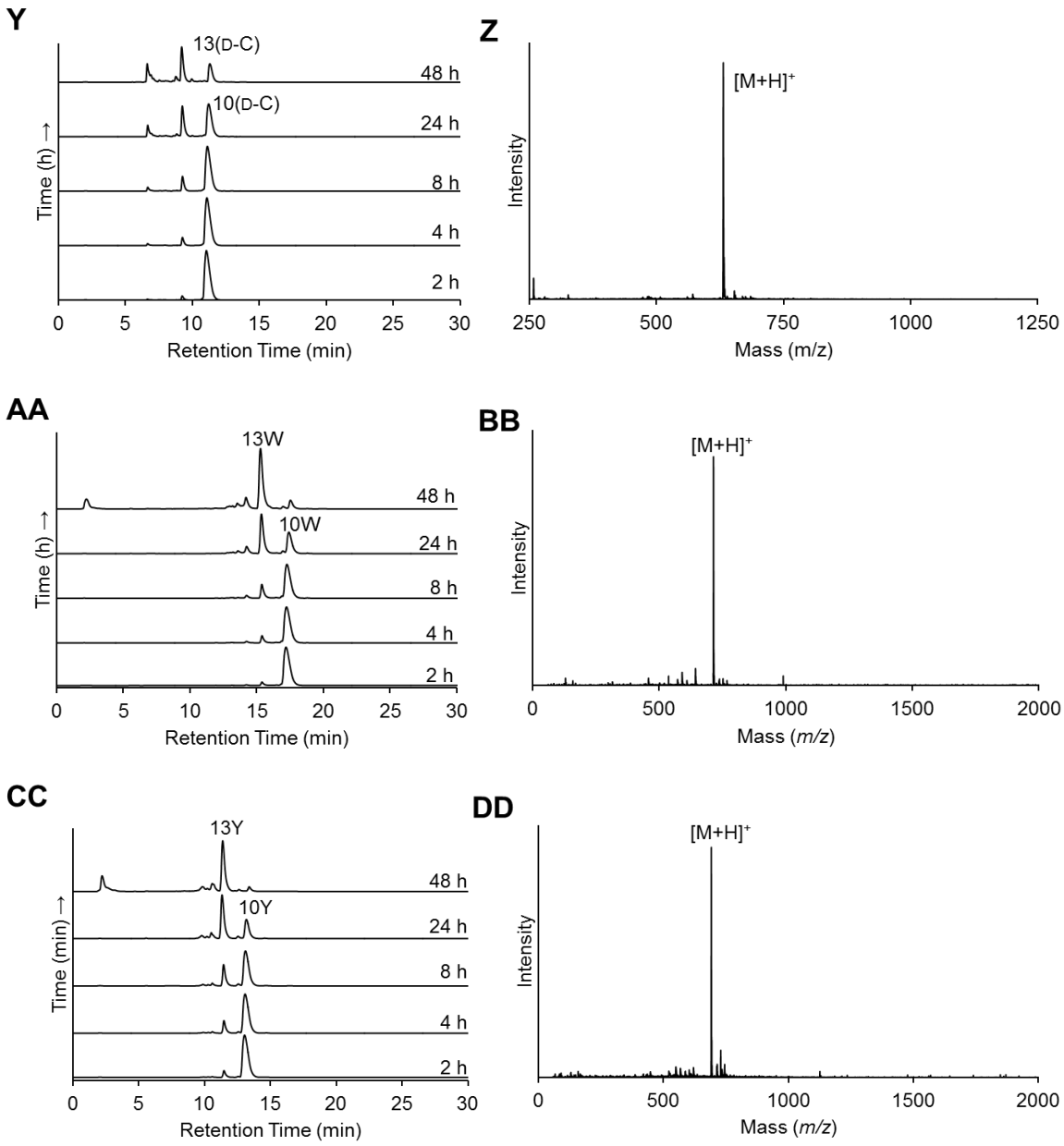
Continued Figure 2.S4. AWKX-MEGA thioesterification. (G) AWK**V**-MEGA thioesterification time-course. (H) ESI-MS of AWK**V**-MES thioester. Calcd. $[M+H]^+$ 626.8 Da, obsd. 626.4 Da. (I) AWK**T**-MEGA thioesterification time-course. (J) ESI-MS of AWK**T**-MES thioester. Calcd. $[M+H]^+$ 628.8 Da, obsd. 628.7 Da. (K) AWK**S**-MEGA thioesterification time-course. (L) ESI-MS of AWK**S**-MES thioester. Calcd. $[M+H]^+$ 614.7 Da, obsd. 614.4 Da. **10** = AWK**X**-MEGA, **13** = AWK**X**-MES thioester. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.



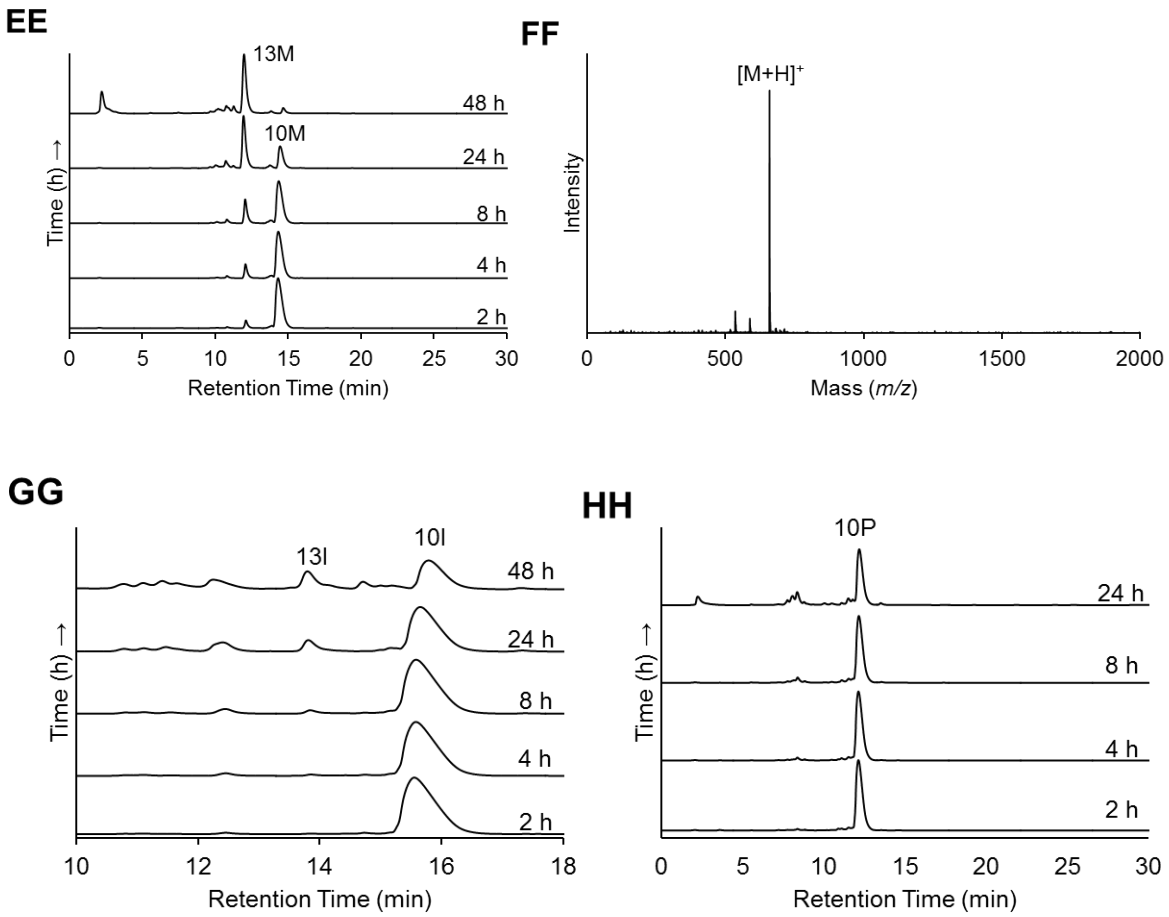
Continued Figure 2.S4. AWKX-MEGA thioesterification. (M) AWK**R**-MEGA thioesterification time-course. (N) ESI-MS of AWK**R**-MES thioester. Calcd. $[M+H]^+$ 683.8 Da, obsd. 683.5 Da. (O) AWK**Q**-MEGA thioesterification time-course. (P) ESI-MS of AWK**Q**-MES thioester. Calcd. $[M+H]^+$ 655.8 Da, obsd. 655.3 Da. (Q) AWK**K**-MEGA thioesterification time-course. (R) ESI-MS of AWK**K**-MES thioester. Calcd. $[M+H]^+$ 656.8 Da, obsd. 656.8 Da. **10** = AWK**X**-MEGA, **13** = AWK**X**-MES thioester. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.



Continued Figure 2.S4. AWKX-MEGA thioesterification. (S) AWK**(D-A)**-MEGA thioesterification time-course. (T) ESI-MS of AWK**(D-A)**-MES thioester. Calcd. $[M+H]^+$ 599.7 Da, obsd. 599.8 Da. (U) AWK**C**-MEGA thioesterification time-course. (V) ESI-MS of AWK**C**-MES thioester. Calcd. $[M+H]^+$ 631.8 Da, obsd. 631.4 Da. (W) AWK**F**-MEGA thioesterification time-course. (X) ESI-MS of AWK**F**-MES thioester. Calcd. $[M+H]^+$ 675.8 Da, obsd. 675.3 Da. **10** = AWK**X**-MEGA, **13** = AWK**X**-MES thioester. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.



Continued Figure 2.S4. AWKX-MEGA thioesterification. (Y) AWK**(D-C)**-MEGA thioesterification time-course. (Z) ESI-MS of AWK**(D-C)**-MES thioester. Calcd. $[M+H]^+$ 631.8 Da, obsd. 631.1 Da. (AA) AWK**W-**MEGA thioesterification time-course. (BB) ESI-MS of AWK**W-**MES thioester. Calcd. $[M+H]^+$ 714.9 Da, obsd. 714.7 Da. (CC) AWK**Y-**MEGA thioesterification time-course. (DD) ESI-MS of AWK**Y**-MES thioester. Calcd. $[M+H]^+$ 691.8 Da, obsd. 691.6 Da. **10** = AWK**X**-MEGA, **13** = AWK**X**-MES thioester. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.



Continued Figure 2.S4. AWKX-MEGA thioesterification. (EE) AWK**M-**MEGA thioesterification time-course. (FF) ESI-MS of AWK**M**-MES thioester. Calcd. $[M+H]^+$ 659.9 Da, obsd. 659.6 Da. (GG) AWK**I-**MEGA thioesterification time-course. (HH) AWK**P-**MEGA thioesterification time-course. **10** = AWK**X**-MEGA, **13** = AWK**X**-MES thioester. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.

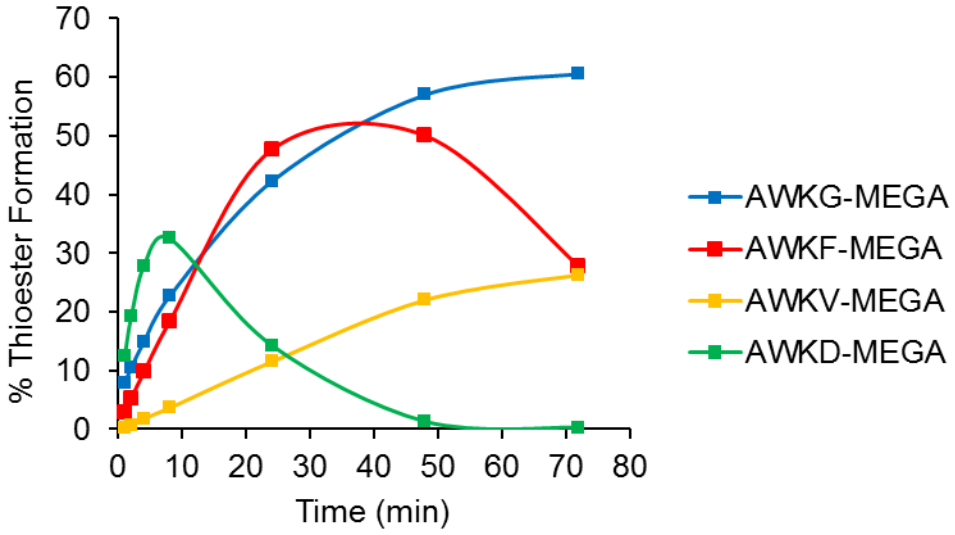


Figure 2.S5. Time-course of thioester formation for AWKX-MEGA peptides. Reactions performed under optimized temperatures at pH 5.6 for each peptide. The final percent thioester formed at each time-point was determined by RP-HPLC peak integration at 280 nm.

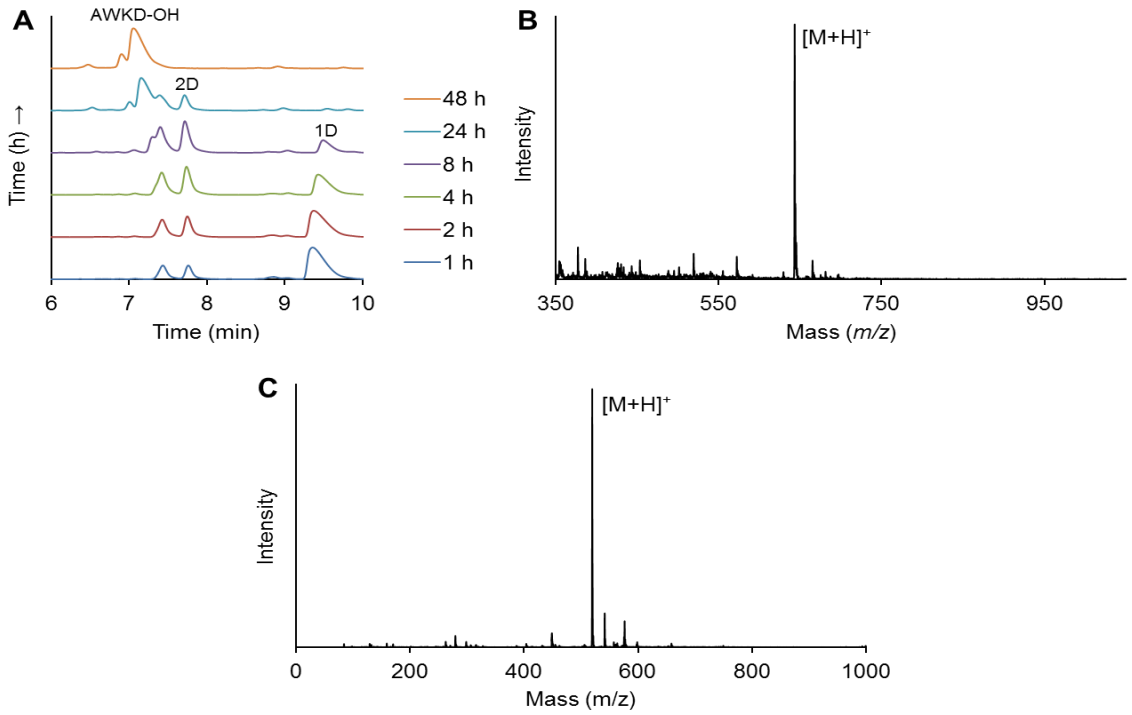


Figure 2.S6. AWKD-MEGA thioesterification (A) AWK**D**-MEGA thioesterification time-course. (B) ESI-MS of AWK**D**-MES thioester. Calcd. $[M+H]^+$ 642.7 Da, obsd. 642.3 Da. (C) ESI-MS of AWK**D**-OH by-product. Calcd. $[M+H]^+$ 518.5 Da; obsd. 518.4 Da. **1** = AWK**D**-MEGA, **2** = AWK**D**-MES thioester. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.

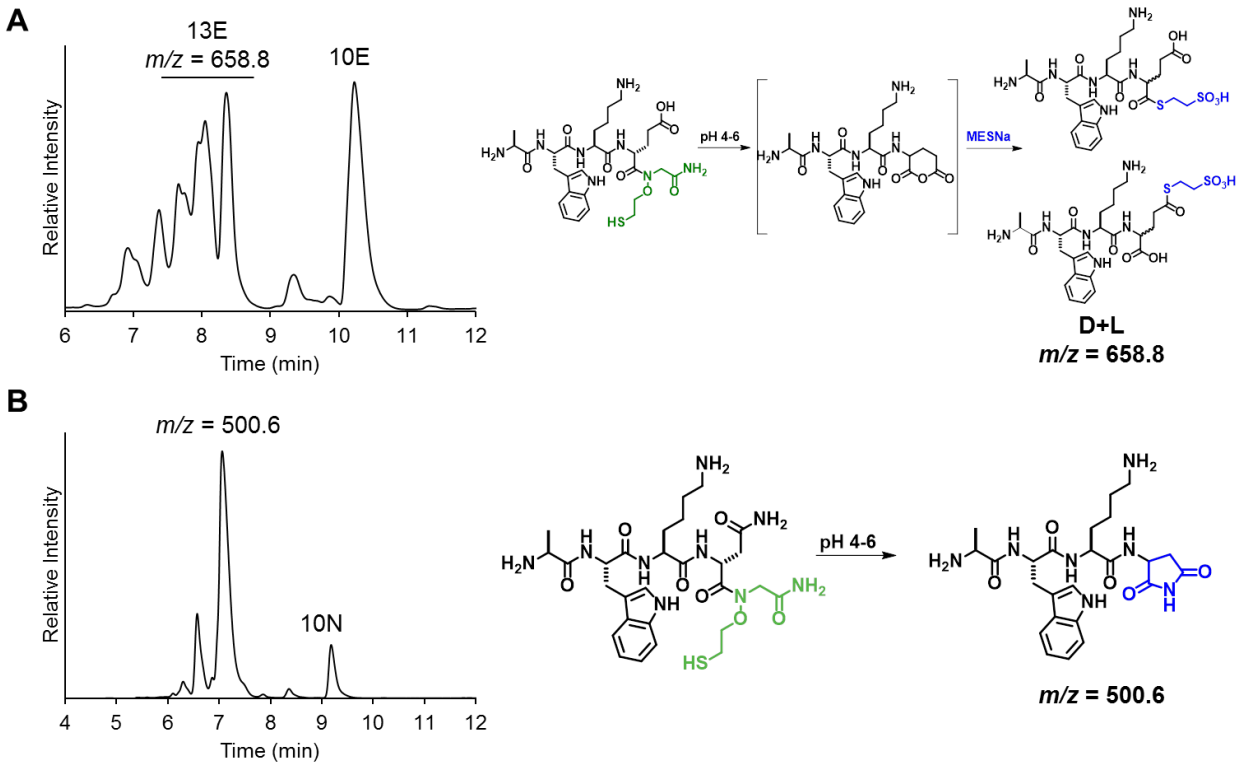


Figure 2.S7. AWKE/N-MEGA thioesterification. (A) AWK**E-**MEGA thioesterification RP-HPLC 48 time-point, pH 5.6, 37 °C (left), and scheme depicting formation of epimeric of AWKE-MES and AWKE-(δ-MES) products (right) (B) AWK**N-**MEGA thioesterification 8 h time-point, pH 4.5, 50 °C (left), and scheme depicting formation of AWKN asparaginimide product (right). **10** = AWK**X**-MEGA, **13** = AWK**X**-MES thioester. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.

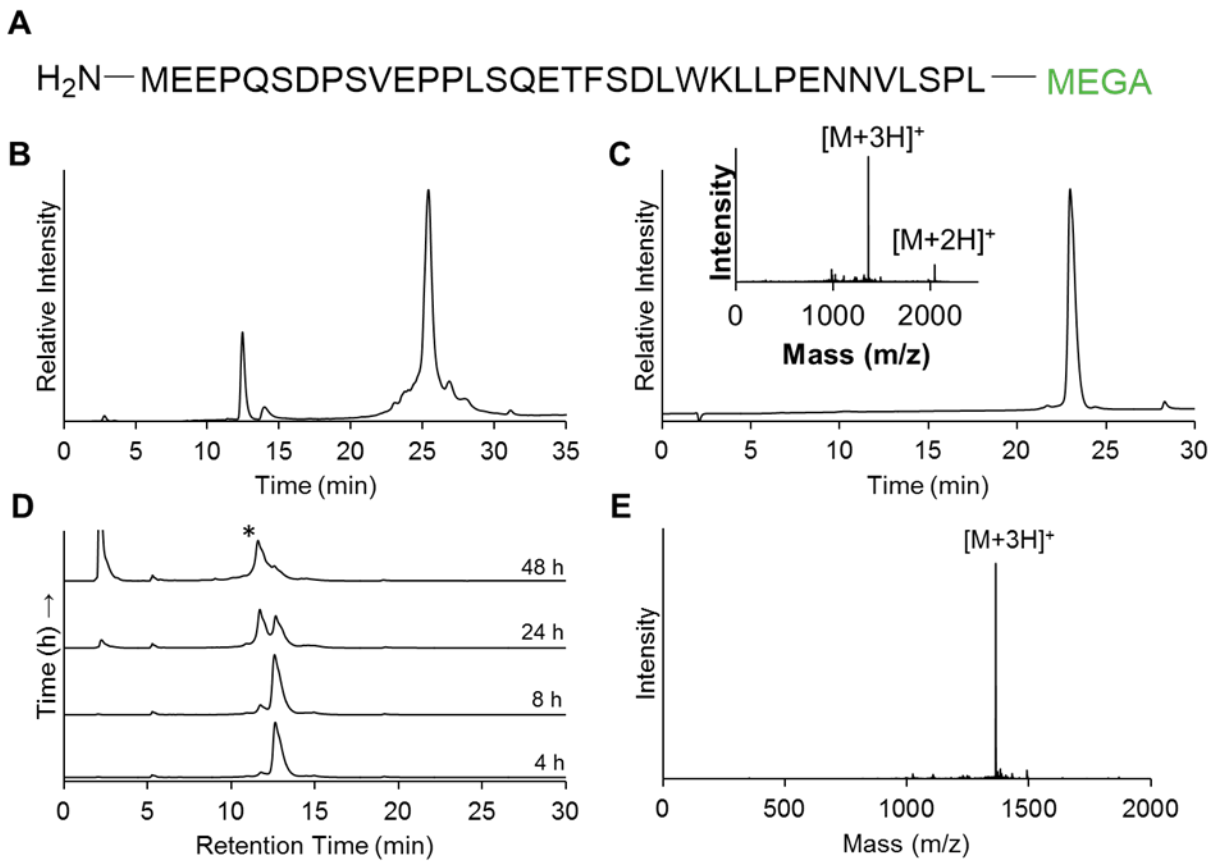


Figure 2.S8. Synthesis and thioesterification of p53(1-35)-MEGA. (A) Amino acid sequence of p53(1-35)-MEGA. (B) RP-HPLC of crude p53(1-35)-MEGA peptide product after TFA cleavage from solid support (0-73% B, 30 min gradient). (C) RP-HPLC of pure p53(1-35)-MEGA (0-73% B, 30 min gradient). Inset is ESI-MS of purified p53(1-35)-MEGA. Calcd. $[M+H]^+$ 4,098.6 Da, obsd. 4,097.7 ± 0.7 Da. (D) Thioesterification time-course of p53(1-35)-MEGA (C18 RP-HPLC, 30-80% B, 30 min gradient). Buffer: 400 mM MESNa, 100 mM NaH_2PO_4, 25 mM TCEP, pH 4.5, 70 °C. (E) ESI-MS of isolated p53(1-35) C-terminal thioester; Calcd. $[M+H]^+$ 4,091.6 Da, obsd. 4,090.9 ± 2.2 Da. * = p53(1-35)-MEGA thioester.

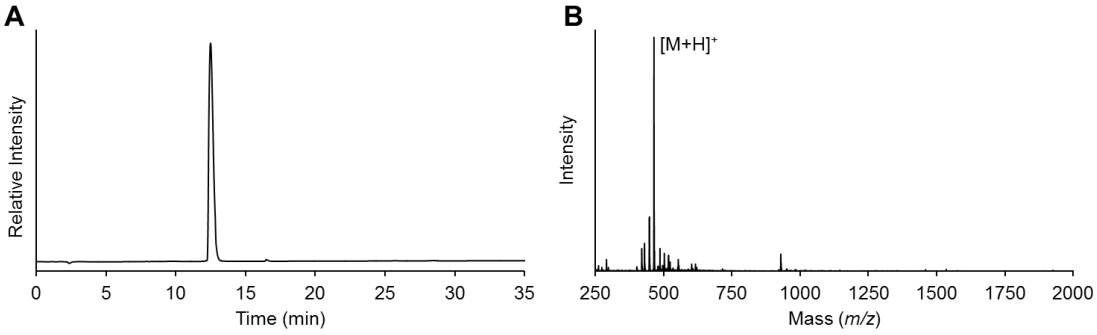


Figure 2.S9. Purified CASW peptide. (A) C18 analytical RP-HPLC chromatogram of purified CASW, 0-73% B, 30 min gradient. (B) ESI-MS of purified CASW. Calcd. $[M+H]^+$ 464.2 Da, obsd. 464.2 Da.

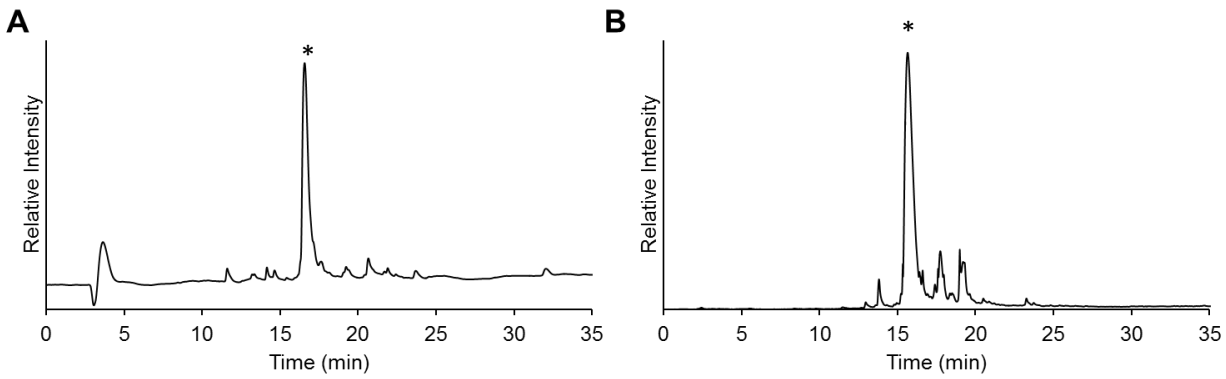


Figure 2.S10. RP-HPLC spectra of crude N-terminal Cys-containing peptides after TFA-cleavage from resin. (A) Crude CRGD(D-F)-MEGA peptide (B) Crude CASHEW-MEGA peptide. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min gradient. Asterisks indicate the desired product.

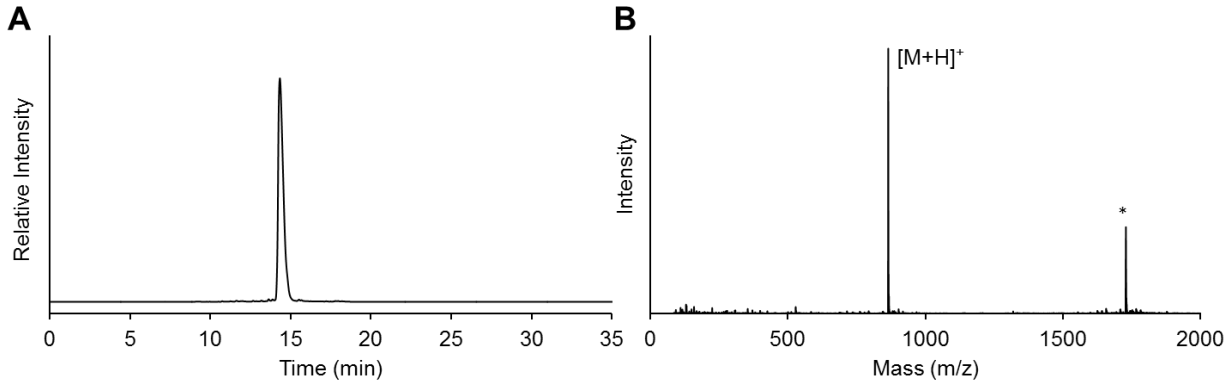


Figure 2.S11. Synthesis of CASHEW-MEGA peptide. (A) C18 analytical RP-HPLC chromatogram of purified CASHE**W**-MEGA peptide, 0-73% B, 30 min gradient. (B) ESI-MS of purified CASHE**W**-MEGA peptide. Calcd. $[M+H]^+$ 864.0 Da, obsd. 864.0 Da. * = Symmetric disulfide of CASHEW-MEGA.

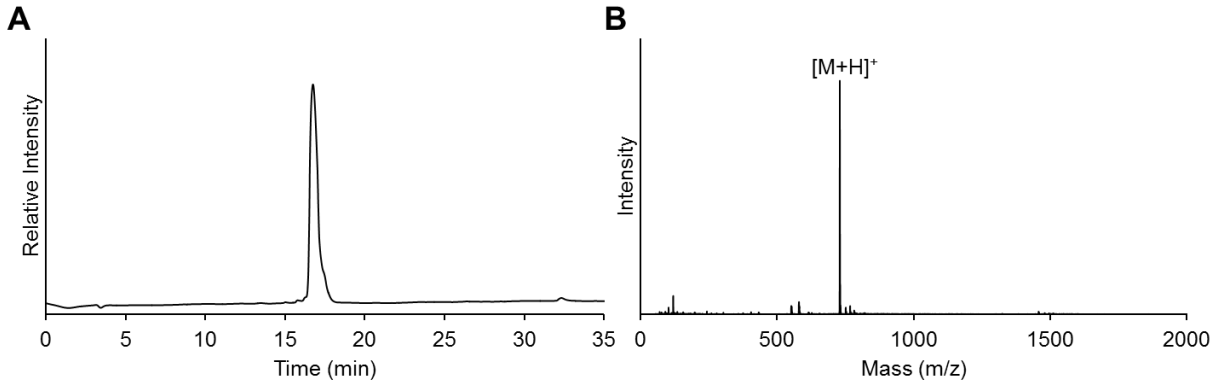


Figure 2.S12. Synthesis of CRGD(D-F)-MEGA peptide. (A) C18 analytical RP-HPLC chromatogram of purified CRGD(D-F)-MEGA peptide, 0-73% B, 30 min gradient. (B) ESI-MS of purified CRGD(D-F)-MEGA peptide. Calcd. $[M+H]^+$ 728.8 Da, obsd. 728.7 Da.

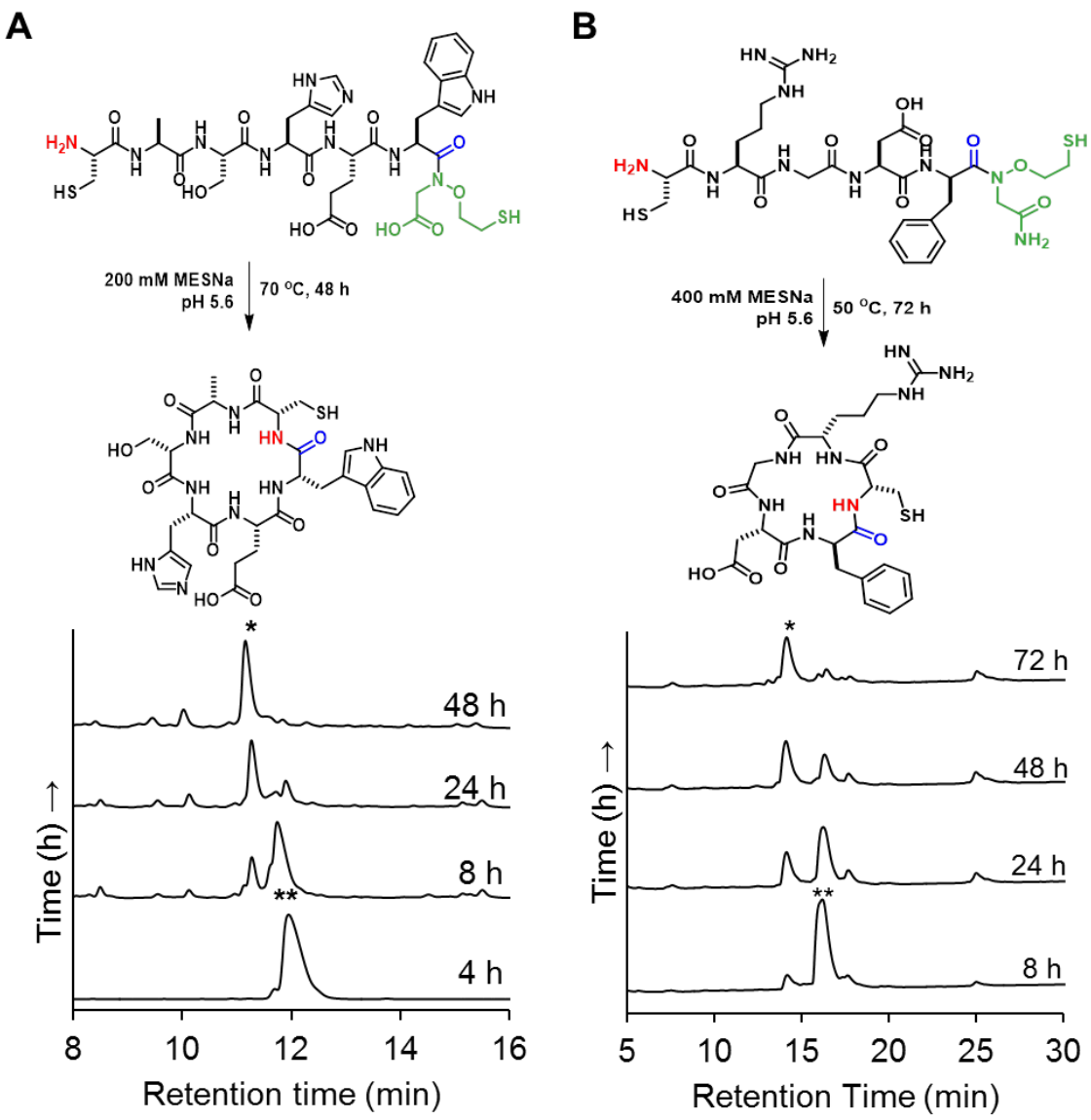


Figure 2.S13. Cyclization of CASHEW-MEGA and CRGD(D-F)-MEGA peptides. (A) Scheme depicting the cyclization reaction of CASHEW-MEGA (top) and RP-HPLC time-course of CASHEW-MEGA cyclization (bottom). (B) Scheme depicting the cyclization reaction of CRGD(D-F)-MEGA (top) and RP-HPLC time-course of CRGD(D-F)-MEGA cyclization (bottom). * = Cyclized peptide, ** = Initial MEGA peptide. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.

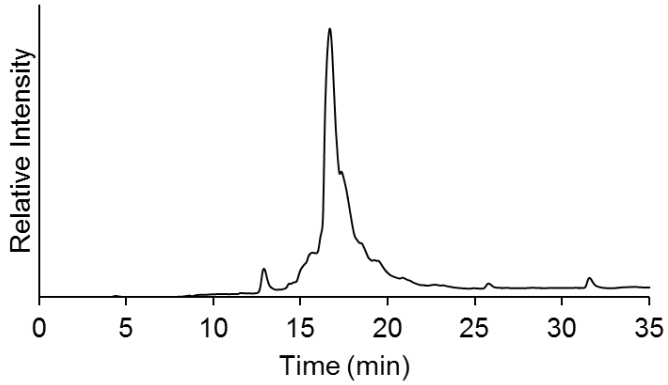


Figure 2.S14. RP-HPLC spectrum of crude SFT-1 (I10G)-MEGA peptide after TFA-cleavage from resin. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min gradient.

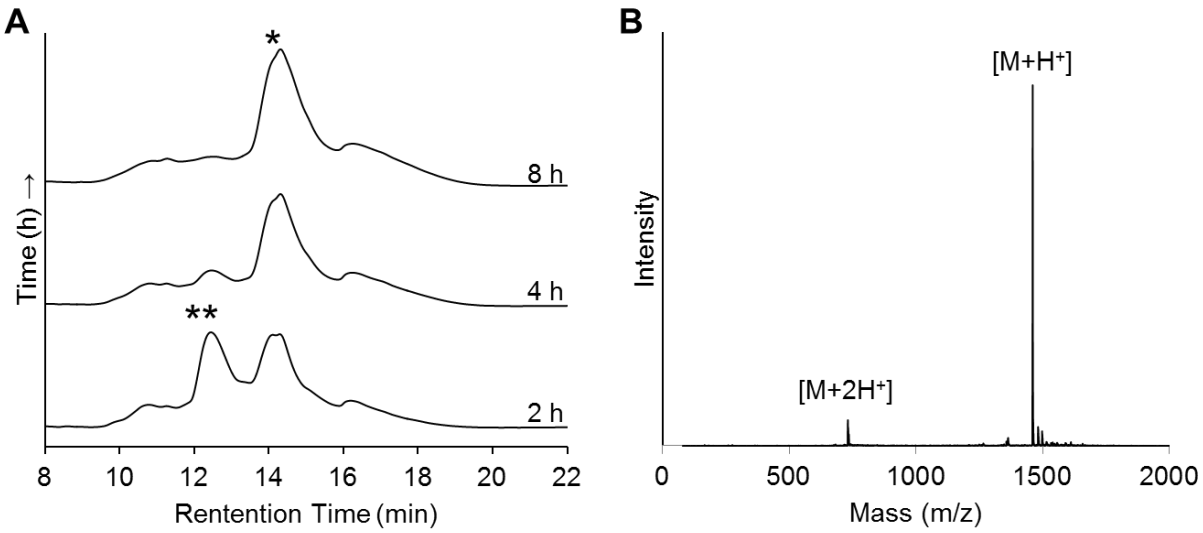


Figure 2.S15. SFT-1 (I10G)-MEGA cyclization. (A) RP-HPLC time-course of SFT-1 (I10G)-MEGA thioesterification. Buffer: 400 mM MESNa, 100 mM NaH_2PO_4, 25 mM TCEP, pH 5.6, 70 °C. * = Cyclized product, ** = SFT-1 (I10G)-MEGA peptide. (B) ESI-MS of SFT-1 (I10G) cyclized product. Calcd. $[M+H]^+$ 1,459.7 Da, obsd. 1,459.4 Da. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.

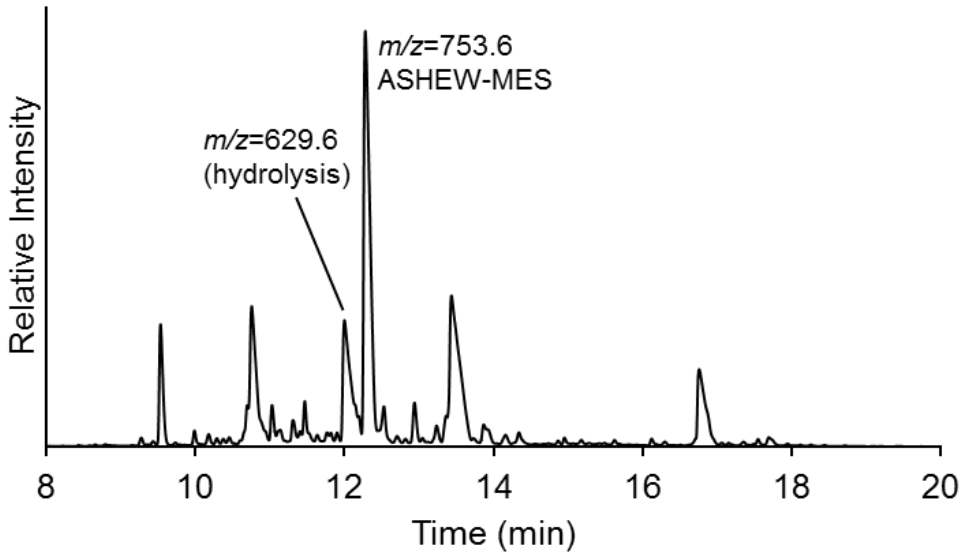


Figure S2.S16. On-resin ASHEW-MEGA thiolysis. 24 h RP-HPLC time-point of ASHEW-MEGA thiolysis, 200 mM NaH_2PO_4, 400 mM MESNa, pH 5.6, 70 °C. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min gradient.

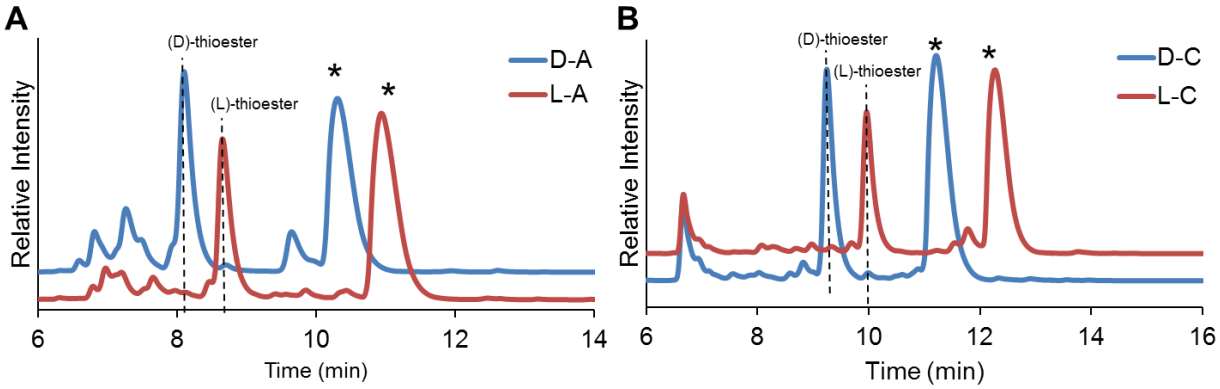


Figure 2.S17. Epimerization during MEGA-mediated thioesterification (A) Overlaid C18 analytical RP-HPLC of AWK**(D-A)**-MES and AWK**A**-MES thioesters formed at 8 h. 0.6% and 1.3% epimerization were observed for AWK**(D-A)**-MEGA and AWK**A**-MEGA, respectively. (B) Overlaid C18 analytical RP-HPLC of AWK**(D-C)**-MES and AWK**C**-MES thioesters formed at 24 h; 0.8% and 1.1% epimerization were observed for AWK**(D-C)**-MEGA and AWK**C**-MEGA, respectively. RP-HPLC: 10-60% B, 30 min gradient. * = AWK**X**-MEGA starting material.

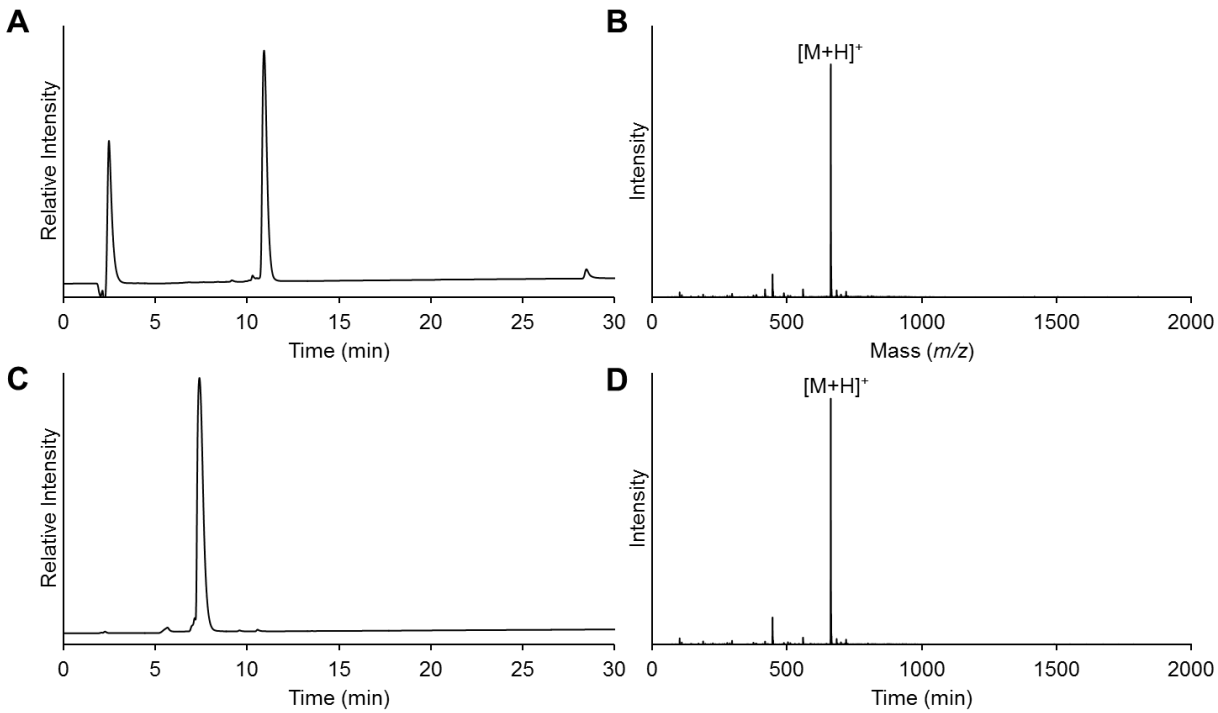


Figure 2.S18. Synthesis of CLAS(D-H)-MEGA and CLAS(L-H)-MEGA peptides (A) C18 analytical RP-HPLC chromatogram of purified CLAS(D-H)-MEGA peptide. (B) ESI-MS of purified CLAS(D-H)-MEGA peptide. Calcd. $[M+H]^+$ 662.8 Da, obsd. 662.6 Da. (C) C18 analytical RP-HPLC chromatogram of purified CLAS(L-H)-MEGA peptide. (D) ESI-MS of purified CLAS(D-H)-MEGA peptide. Calcd. $[M+H]^+$ 662.8 Da, obsd. 662.5. RP-HPLC performed on C18 analytical column, 0-73% B, 30 min gradient.

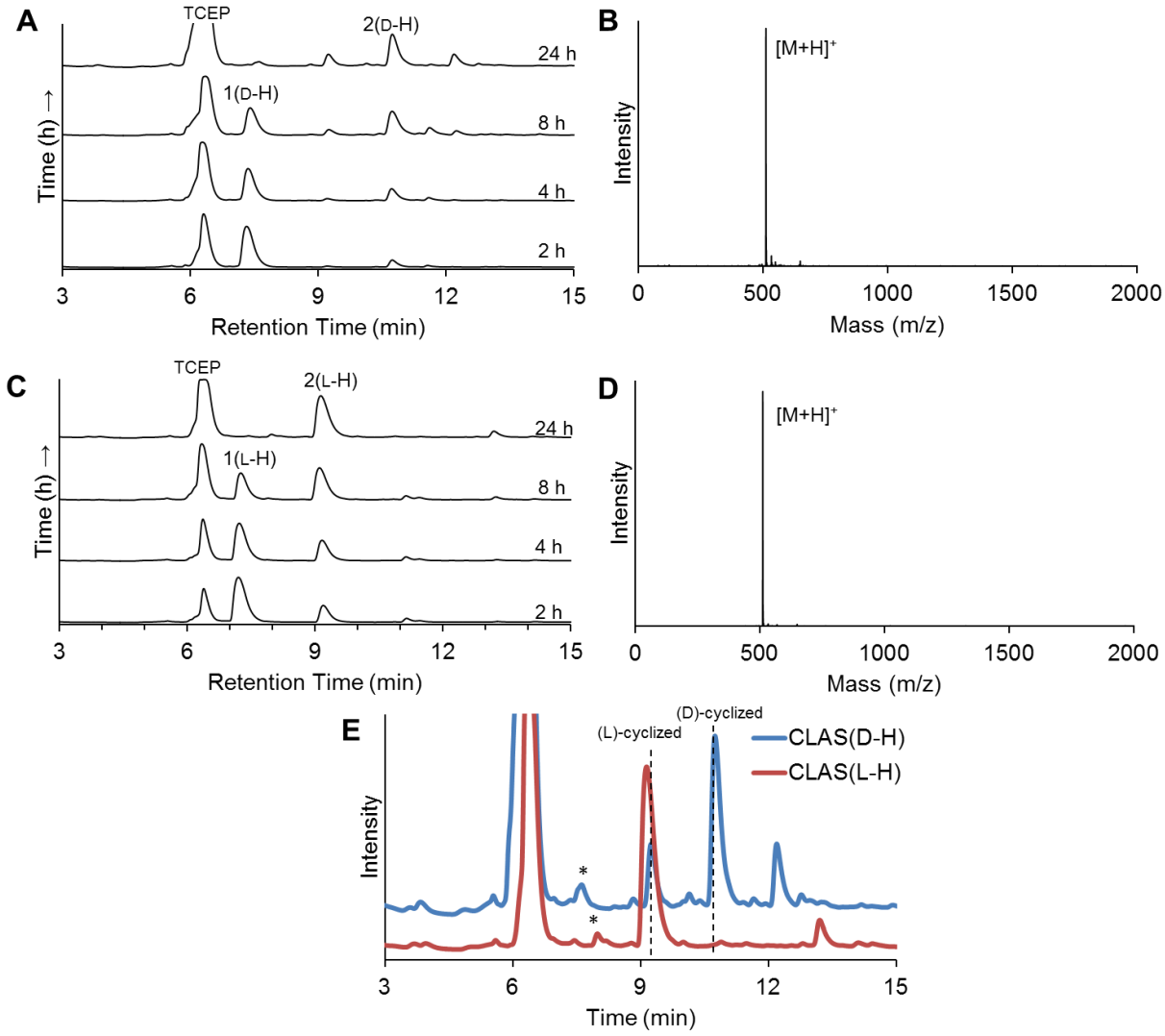


Figure 2.S19. Cyclization of CLAS(D-H)-MEGA and CLAS(L-H)-MEGA peptides (A) CLAS(D-H) cyclization time-course. (B) ESI-MS of purified cyclic CLAS(D-H) peptide. Calcd. $[M+H]^+$ 512.2 Da, obsd. 512.4 Da. (C) CLAS(L-H) cyclization time-course. (D) ESI-MS of purified cyclic CLAS(L-H) peptide. Calcd. $[M+H]^+$ 512.2 Da, obsd. 512.4 Da. (E) Overlaid C18 analytical RP-HPLC of CLAS(D-H)-MEGA and CLAS(L-H)-MEGA cyclization reactions at 24 h; 24.8% and 1.8% epimerization were observed for CLAS(D-H)-MEGA and CLAS(L-H)-MEGA, respectively. **1** = CLASH-MEGA, **2** = Cyclized product, * = Unreacted CLASH-MEGA. RP-HPLC performed on C18 analytical column, 10-60% B, 30 min gradient.

2.6 References

(1) Vila-Perelló, M.; Liu, Z.; Shah, N. H.; Willis, J. A.; Idoyaga, J.; Muir, T. W. *J. Am. Chem. Soc.* **2013**, *135* (1), 286–292.

(2) Gentle, I. E.; De Souza, D. P.; Baca, M. *Bioconjug. Chem.* **2004**, *15* (3), 658–663.

(3) Camarero, J. A.; Cotton, G. J.; Adeva, A.; Muir, T. W. *J. Pept. Res.* **1998**, *51* (4), 303–316.

(4) Tilman M . Hackeng, J. H. . G. and P. E. . D. *J. Pept. Res.* **1999**, *96* (18), 10068–10073.

(5) Laszlo Otvos, Ilona Elekes, V. M.-Y. L. *J. Pept. Protein Res.* **1989**, *34*, 129–133.

(6) Gamblin, D. P.; Scanlan, E. M.; Davis, B. G. *Chem. Rev.* **2009**, *109* (1), 131–163.

(7) Muttenthaler, M.; Albericio, F.; Dawson, P. E. *Nat. Protoc.* **2015**, *10* (7), 1067–1083.

(8) Gates, Z. P.; Dhayalan, B.; Kent, S. B. H. *Chem. Commun.* **2016**, *52* (97), 13979–13982.

(9) Blanco-Canosa, J. B.; Dawson, P. E. *Angew. Chemie - Int. Ed.* **2008**, *47*, 6851–6855.

(10) Zheng, J. S.; Tang, S.; Huang, Y. C.; Liu, L. *Acc. Chem. Res.* **2013**, *46* (11), 2475–2484.

(11) Blanco-Canosa, J. B.; Nardone, B.; Albericio, F.; Dawson, P. E. *J. Am. Chem. Soc.* **2015**, *137* (22), 7197–7209.

(12) Raz, R.; Rademann, J. *Org. Lett.* **2011**, *13* (7), 1606–1609.

(13) Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. Bin; Lin, Y.; Cui, H. K.; Liu, L. *Angew. Chemie - Int. Ed.* **2011**, *50* (33), 7645–7649.

(14) Mende, F.; Seitz, O. *Angew. Chemie - Int. Ed.* **2011**, *50*, 1232–1240.

(15) Ficht, S.; Payne, R. J.; Guy, R. T.; Wong, C. H. *Chem. - A Eur. J.* **2008**, *14*, 3620–3629.

(16) Hemantha, H. P.; Bavikar, S. N.; Herman-Bachinsky, Y.; Haj-Yahya, N.; Bondalapati, S.; Ciechanover, A.; Brik, A. *J. Am. Chem. Soc.* **2014**, *136* (6), 2665–2673.

(17) Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (17), 6657–6662.

(18) Fauvet, B.; Butterfield, S. M.; Fuks, J.; Brik, A.; Lashuel, H. A. *Chem. Commun.* **2013**, *49* (81), 9254–9256.

(19) Erlich, L. A.; Kumar, K. S. A.; Haj-Yahya, M.; Dawson, P. E.; Brik, A. *Org. Biomol. Chem.* **2010**, *8* (10), 2392–2396.

(20) Kang, J.; Reynolds, N. L.; Tyrrell, C.; Dorin, J. R.; MacMillan, D. *Org. Biomol. Chem.* **2009**, *7* (23), 4918–4923.

(21) Loibl, S. F.; Harpaz, Z.; Seitz, O. *Angew. Chemie - Int. Ed.* **2015**, *54* (50), 15055–15059.

(22) Zheng, J.; Chen, X.; Tang, S.; Chang, H.; Wang, F.; Zuo, C. *Org. Lett.* **2014**, *16*, 4908–

4911.

(23) Taichi, M.; Hemu, X.; Qiu, Y.; Tam, J. P. *Org. Lett.* **2013**, *15* (11), 2620–2623.

(24) Liu, F.; Mayer, J. P. *J. Org. Chem.* **2013**, *78* (19), 9848–9856.

(25) Sharma, R. K.; Tam, J. P. *Org. Lett.* **2011**, *13* (12), 5176–5179.

(26) Sato, Kohei; Shigenaga, Akira; Tsuji, Kohei; Sumikawa, Yoshitake; Sakamoto, Kan; Otaka, A. *ChemBioChem* **2011**, *12*, 1840–1844.

(27) Hou, W.; Zhang, X.; Li, F.; Liu, C. *Org. Lett.* **2011**, *13* (3), 2009–2012.

(28) Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. *Org. Lett.* **2010**, *12* (22), 5238–5241.

(29) Shelton, P. M. M.; Weller, C. E.; Chatterjee, C. *J. Am. Chem. Soc.* **2017**, *139* (11), 3946–3949.

(30) Weller, C. E.; Huang, W.; Chatterjee, C. *Chembiochem* **2014**, *15* (9), 1263–1267.

(31) Canne, L. E.; Bark, S. J.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118* (25), 5891–5896.

(32) Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117* (1), 147–157.

(33) 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.; Chatterjee, C. *Nat. Commun.* **2016**, *7*, 1–10.

(34) Patgiri, A.; Witten, M. R.; Arora, P. S. *Org. Biomol. Chem.* **2010**, *8* (8), 1773–1776.

(35) Villain, M.; Gaertner, H.; Botti, P. *European J. Org. Chem.* **2003**, No. 17, 3267–3272.

(36) Bacsa, B.; Kappe, C. O. *Nat. Protoc.* **2007**, *2* (9), 2222–2227.

(37) Wang, C. K.; Craik, D. J. *Biopolymers* **2016**, *106* (6), 901–909.

(38) Prante, O.; Einsiedel, J.; Haubner, R.; Gmeiner, P.; Wester, H. J.; Kuwert, T.; Maschauer, S. *Bioconjug. Chem.* **2007**, *18* (1), 254–262.

(39) Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123* (4), 526–533.

(40) Noisier, A. F. M.; Albericio, F. *Amin. Acids, Pept. Proteins* **2015**, *39*, 1–20.

(41) Luckett, S.; Garcia, R. S.; Barker, J. .; Konarev, A. .; Shewry, P. .; Clarke, A. .; Brady, R. . *J. Mol. Biol.* **1999**, *290* (2), 525–533.

(42) Quimbar, P.; Malik, U.; Sommerhoff, C. P.; Kaas, Q.; Chan, L. Y.; Huang, Y. H.; Grundhuber, M.; Dunse, K.; Craik, D. J.; Anderson, M. A.; Daly, N. L. *J. Biol. Chem.* **2013**, *288* (19), 13885–13896.

(43) Punna, S.; Kuzelka, J.; Wang, Q.; Finn, M. G. *Angew. Chemie - Int. Ed.* **2005**, *44* (15), 2215–2220.

(44) Tulla-Puche, J.; Barany, G. *J. Org. Chem.* **2004**, *69* (12), 4101–4107.

(45) Pernerstorfer, J.; Schuster, M.; Blechert, S. *Chem. Commun.* **1997**, *1* (20), 1949–1950.

(46) Aimetti, A. A.; Shoemaker, R. K.; Lin, C. C.; Anseth, K. S. *Chem. Commun.* **2010**, *46* (23), 4061–4063.

(47) Selvaraj, A.; Chen, H. T.; Ya-Ting Huang, A.; Kao, C. L. *Chem. Sci.* **2018**, *9* (2), 345–349.

(48) Cheng, X.; Hong, H.; Zhou, Z.; Wu, Z. *J. Org. Chem.* **2018**, *83* (22), 14078–14083.

(49) Gless, B. H.; Olsen, C. A. *J. Org. Chem.* **2018**, *83* (17), 10525–10534.

(50) Eissler, S.; Kley, M.; Bächle, D.; Loidl, G.; Meier, T.; Samson, D. *J. Pept. Sci.* **2017**, *23* (10), 757–762.

(51) Yongxin Han, Fernando Albericio, and G. B. *J. Org. Chem.* **1997**, *62*, 4307–4312.

(52) Coin, I.; Beyermann, M.; Bienert, M. *Nat. Protoc.* **2007**, *2* (12), 3247–3256.

(53) Carpino, L. A.; Han, G. Y. H. *J. Org. Chem.* **1972**, *37* (22), 3404–3409.

Chapter 3

Biochemical investigation of the effect of histone H4 sumoylation on histone H3 deacetylation by the HDAC1-CoREST1 sub-complex

3.1 Introduction

Post-translational acetylation of histone N-termini is an established regulatory mechanism for many important cellular processes including gene transcription, DNA damage repair and cellular metabolism pathways.[1–5] Histone acetylation is a reversible modification that is controlled by the opposing enzymatic activities of histone acetyltransferases (HATs) and histone deacetylases (HDACs).[6,7] All four core histones are acetylated in vivo; however, acetylation of H3 and H4 N-termini in particular are strongly associated with transcription activation and gene activity.[1,2] The action of HDAC enzymes is therefore considered to be generally repressive for gene activity. Biochemically, histone acetylation acts to neutralize positively charged lysine ε-amines, disrupting electrostatic histone-DNA contacts to unfold compacted nucleosomes and facilitate transcription machinery binding to DNA[8,9] Furthermore, bromodomain-containing proteins recognize acetylated lysines adding further complexity to the regulation of chromatin structure and function through histone acetylation.[10–12]

A large body of evidence suggests that disruption of the spatial and temporal distribution of histone acetylation, due to misregulated activity of their corresponding HATs and HDAC enzymes, can lead to a variety of disease states including cancer and debilitating neurodegenerative diseases.[13–15] For example, loss or mutation of HDAC2 has been implicated

in colorectal and colon cancers by imparting increased cellular resistance to anti-proliferative and pro-apoptotic drugs.[16] In addition, the impairment of HDAC6 has been linked to Parkinson's disease due to aberrant aggresome formation and clearance of misfolded proteins in cells lacking this enzyme.[17–19] Decoding the crosstalk between histone acetylation, their effector proteins and other histone post-translational modifications (PTMs) is critical to understanding differences in gene transcription arising from subtle changes in the chromatin PTM landscape..

Due to the diversity within the HDAC family and their association with cancer proliferation, HDAC enzyme- and class-specific anticancer therapeutic development has garnered much attention over the last two decades.[20–22] HDACs are divided into four distinct classes (I, II, III and IV) based on sequence similarity to their yeast orthologues reduced potassium dependency-3 (Rpd3) and histone deacetylase 1 (Hda1), and their mechanisms of action (**Table 3.1, Figure 3.1**).[7] Classes I, II and IV HDACs are metal-dependent enzymes with large active site cavities imparting broad substrate scope. A zinc ion is coordinated by conserved active site His and Asp residues, and activates the carbonyl moiety of acetylated substrates for subsequent hydrolysis by water (**Scheme 3.1A**).[23,24] Class III HDACs (also known as sirtuins) use a nicotinamide adenine dinucleotide (NAD^+) cofactor and general acid-base catalysis to deacetylate their substrates, producing nicotinamide and acetylated ADP-ribose as by-products (**Scheme 3.1B**).

X-ray crystal structures for several of the metal-dependent HDACs have been reported.[25–28] Members of Class I, II and IV HDACs share a high degree of sequence similarity and tertiary structure within the ~250-400 residue deacetylase domain. Thus, residue-specific variances and additional structural elements lead to the observed differences in active site geometry, substrate recognition and protein-protein interactions among the zinc-dependent HDACs (**Figure 3.1**). Class I, comprised of HDAC1, 2, 3 and 8, contain large globular N-terminal deacetylase (DAC)

Table 3.1. Classification of metal-dependent HDAC isoforms

	Isoform	No. of residues	Deacetylase (DAC) domain residues[a]	DAC sequence identity (%) to[b]:
				Rpd3 (*S. cerevisiae*)
Class I	HDAC1	482	9-321	64
	HDAC2	488	9-322	65
	HDAC3	428	3-316	57
	HDAC8	377	14-324	43
				Hda1 (*S. cerevisiae*)
Class II	HDAC4	1,084	655-1,084	36
	HDAC5	1,112	684-1,028	34
	HDAC6	1,215	87-404, 482-800	38, 44
	HDAC7	912	518-685	37
	HDAC9	1,069	631-978	35
	HDAC10	669	1-323	41
				Rpd3 (Hda1)
Class IV	HDAC11	347	14-326	31 (24)

[a]Derived from UniProt accession numbers Q13547 (HDAC1), Q92769 (HDAC2), O15379 (HDAC3), P56524 (HDAC4), Q9UQL6 (HDAC5), Q9UBN7 (HDAC6), Q8WUI4 (HDAC7), Q9BY41 (HDAC8), Q9UKV0 (HDAC9), Q9UKV0 (HDAC10), Q96DB2 (HDAC11). [b]From reference 7.

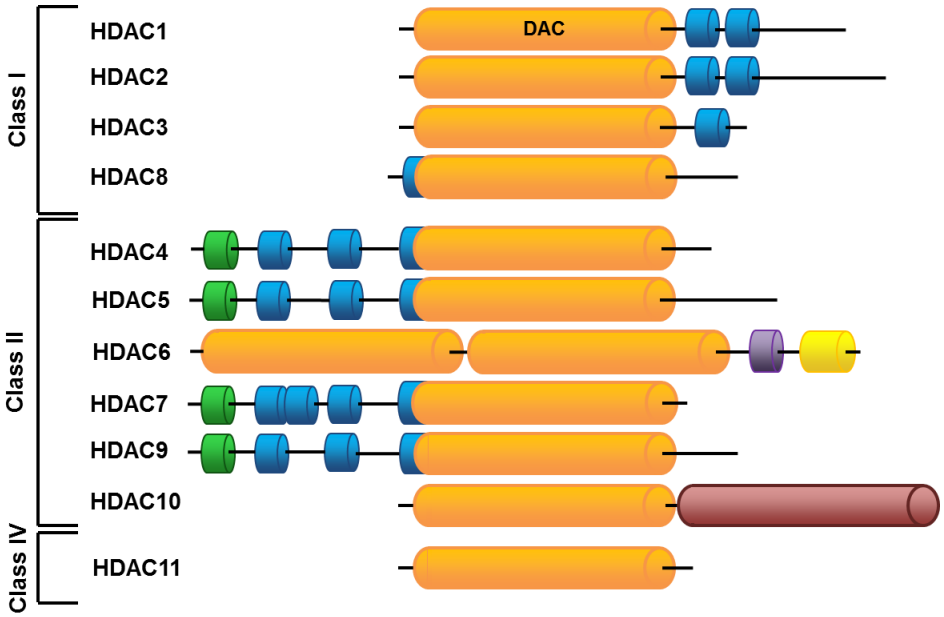
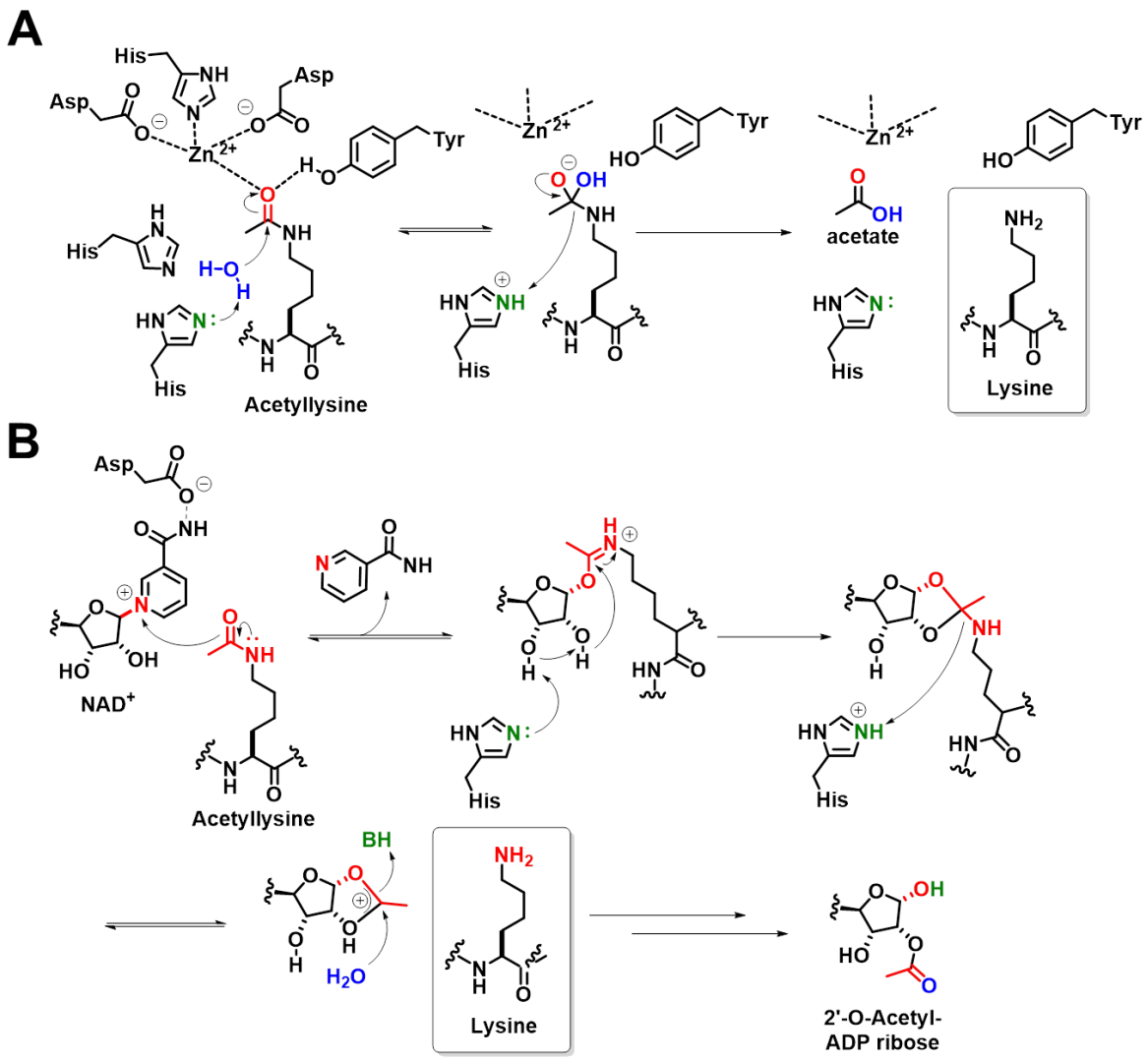


Figure 3.1. Sequence maps of metal-dependent HDAC isoforms. HDAC constructs are grouped by class (I, II, IV). Orange cylinders = DAC domains. Blue cylinders = sites of serine phosphorylation and protein 14-3-3 binding site. Green cylinders = sites of protein MEF binding. Purple cylinder = Ser/Glu repeat motif. Yellow cylinder = Zinc-finger motif/ubiquitin binding site. Red cylinder = leucine zipper motif. Adapted from reference 7.



Scheme 3.1. Deacetylation mechanisms of Zn-dependent (A) and NAD$^+$-dependent (B) HDACs. Adapted from reference 21.

domains and short, unstructured C-terminal domains. HDAC1 and 2, the most widely studied HDAC isoforms, share nearly identical sequences and largely redundant substrate specificities Thus, these proteins are often described together as HDAC1/2. Class II HDACs, comprised of HDAC4-7, 9 and 10 possess additional C- and/or N-terminal domains responsible for specific protein-protein interactions (e.g. with MEF2, 14-3-3). Little is currently known about the sole member of Class IV, HDAC11, but despite sequence homology to both Class I and II HDACs,

HDAC11 is highly conserved from yeast to humans and evolutionary analyses suggest it is the only member of its class.[29]

Like histone deacetylation activity, histone modification by the small ubiquitin-like modifier (SUMO) 2/3 proteins has been reported to downregulate transcription via mutational analysis and quantitative PCR experiments in yeast, and luciferase reporter assays in human cells (Chapter 1.4).[30,31] Despite its general association with gene repression, the Chatterjee lab has demonstrated that site-specifically sumoylated nucleosome arrays favor an open, euchromatic architecture in several biophysical characterization experiments.[32] The incongruity between the transcriptionally repressive nature of histone sumoylation in cells, and the preference for sumoylated arrays to adopt an open euchromatic structure has warranted further investigation into the precise mechanism(s) by which SUMO2/3 exerts its influence on gene regulation. If SUMO itself does not physically promote chromatin condensation, it is conceivable that recruitment of gene repressive histone-modifying enzymes (e.g. HDACs) or repressive protein complexes underlies its association with transcription repression.

HDAC enzymes can operate within one or more protein complexes with diverse biological functions. For example, the nucleosome remodeling and deacetylase (NuRD) and the lysine specific demethylase 1 (LSD1), co-repressor for element 1 silencing transcription factor (CoREST1), and HDAC1/2 complexe (LCH) finely tune histone methylation and acetylation states, which can stimulate or preclude the binding of cellular transcriptional machinery.[33–36] HDAC1/2 are central components of the LCH complex responsible for removing histone acetyl groups at target gene loci (**Figure 3.2**).[37,38] LSD1, an H3K4me1/2 specific histone demethylase, complements the repressive activity of HDAC1/2 in the LCH complex by removing gene-

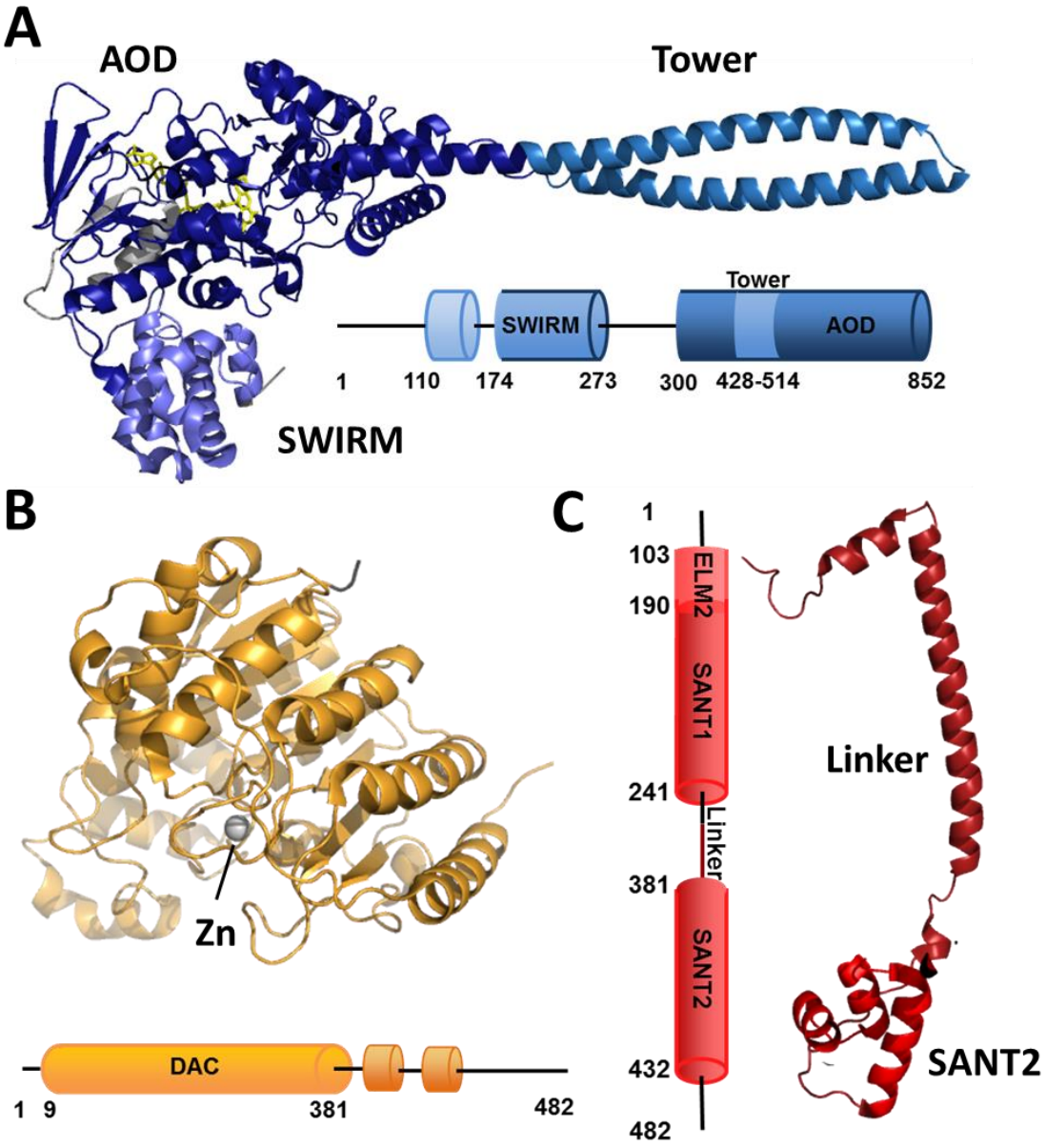


Figure 3.2. LCH complex components. (A) X-ray crystal structure of LSD1(171-835) and sequence map describing LSD1 domain organization. LSD1 FAD cofactor is depicted in yellow. PDB: 2V1D. (B) Crystal structure of HDAC1 (8-376) and sequence map describing HDAC1 domain organization. Zinc cofactor is depicted in gray. PDB: 4BKX. (C) Crystal structure of CoREST1(308-440) and sequence map describing CoREST1 domain organization. PDB: 2V1D.

activating histone methyl marks (Chapter 4.1).[39] LSD1 and HDAC enzymes are associated in the LCH complex through the scaffolding protein, CoREST1, by interacting with its α-helical

linker domain and N-terminal SANT1 domain, respectively (**Figure 3.2**). While CoREST1 is believed to be necessary for efficient nucleosome demethylation by LSD1 in vivo, and stimulates LSD1 activity toward mononucleosome substrates in vitro, to our knowledge there have been no studies describing the effect of CoREST1 on HDAC1/2 activity.[40] However, the sumoylation of nuclear proteins has been implicated in HDAC-mediated gene repression.[41] The LCH complex was previously found to repress the expression of the sodium channel α-subunit genes, *SCN1A* and *SCN3A*.[38] This gene repression was dependent on the binding interaction between CoREST1 and an unidentified sumoylated nuclear protein. The CoREST1-SUMO interaction was proposed to depend on a non-consensus SUMO interacting motif (ncSIM), I-D-I-E-V, identified within CoREST1(255-275) during co-precipitation assays. However, the target(s) of sumoylation which led to the observed repressive activity of the LCH complex remained unknown. Recently, the Chatterjee lab has demonstrated a stimulatory effect of the LSD1-CoREST1 (LC) sub-complex activity toward H3K4me2 in the context of chemically sumoylated H4 (suH4) nucleosomes.[42] Similar to the in vivo repression of *SCN* genes, the stimulatory effect on LSD1 activity was dependent on the interaction between CoREST1 and suH4, as point mutations in the SIM of CoREST1 (I270A, I272A, V274A, collectively called CoREST1-3A) failed to stimulate LSD1 demethylation activity.

In this chapter, I report that the deacetylation activity of the HDAC-CoREST1 (HC) sub-complex towards site-specifically acetylated nucleosomal substrates is stimulated by chemically sumoylated H4. We found that the degree of histone tail acetylation (mono- or polyacetylation) does not negatively influence deacetylation by the HC sub-complex, nor the stimulatory affect that suH4 imparts. Given the clear importance of the CoREST1-SUMO3 interaction to the activity enhancement of LC and HC sub-complexes, we also performed a quantitative assessment of their binding properties using heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) spectroscopy titration experiments. These results, in

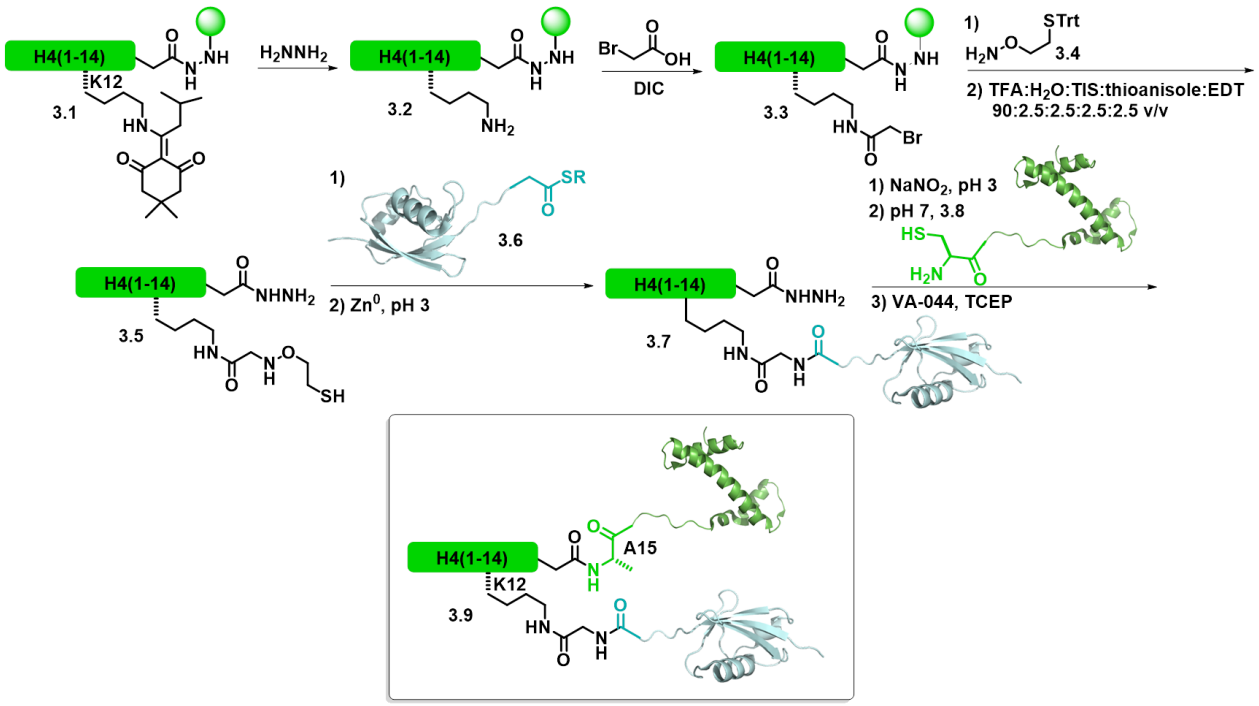
conjunction with previous cellular studies, clearly indicate that histone sumoylation plays a key role in the epigenetic hierarchy of dynamic chromatin remodeling for transcriptional repression.

3.2 Results and Discussion

3.2.1 Semisynthesis of sumoylated H4

In order to probe the effect of suH4 on HC activity, we required site-specifically modified H3 and H4 proteins. Natively-linked and site-specifically modified proteins are challenging substrates to obtain in quantities necessary for biochemical evaluation. The Chatterjee lab has previously reported an EPL-based method for the semi-synthesis of milligram quantities of suH4 (Chapter 1.6, **Scheme 3.2**).[42,43] The semisynthesis of suH4 hinges on a key chemical ligation step between recombinant H4(15-102)A15C truncate (**3.8**) and chemically sumoylated H4(1-14) C-terminal peptide hydrazide (**3.7**). First, the histone H4 truncate was expressed in *E. coli* as a fusion to tobacco etch virus (TEV) protease recognition sequence (ENLYFQ/S).[44] After isolation of the His_6-TEV-H4 truncate fusion construct via immobilized metal affinity chromatography (IMAC), treatment with recombinant TEV protease cleaved the affinity purification handle to release the free N-terminal cysteine, which was further purified by RP-HPLC (**Figures 3.S1-3.S3**). The low efficiency of TEV cleavage from this construct necessitated multiple rounds of both protease treatment and Ni^{2+}-affinity purifications to isolate sufficient quantities of the truncated H4(15-102)A15C construct.

To prepare the sumoylated H4 peptide hydrazide, recombinant His_6-SUMO3(2-91)C47S-MES thioester (**3.6**) was first generated by expression in *E. coli* as the C-terminal AVA-intein fusion (**Scheme 3.2**).[45,46] Following IMAC purification, dialysis of the fusion construct in a MESNa-containing buffer stimulated intein-mediated *N*-to-*S* acyl shift and transthioesterification to give the stable His_6-SUMO3(2-91)C47S-MES thioester, which was further purified by RP-HPLC



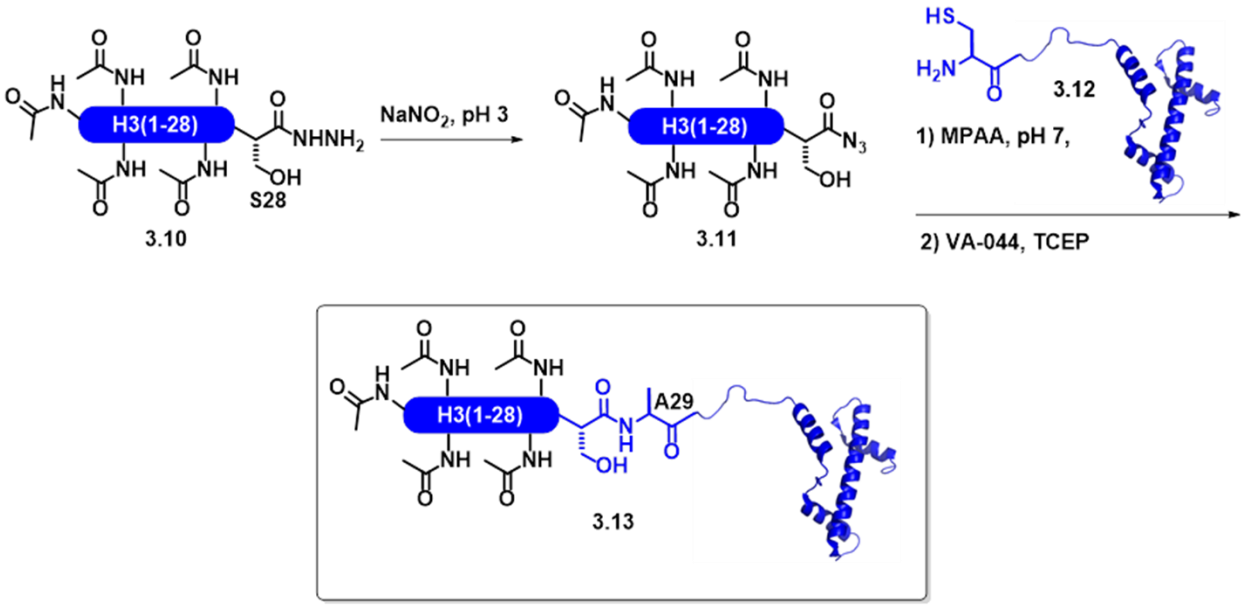
Scheme 3.2. Semisynthesis of suH4

(**Figure 3.S4**). Concurrently, H4(1-14) C-terminal peptide hydrazide containing orthogonally protected K12ivDde (**3.1**) was prepared via standard 9-fluorenylmethoxycarbonyl (Fmoc)-SPPS. Solid-phase removal of the ivDde group using hydrazine allowed installation of a 2-(aminooxy)ethanethiol ligation handle (**3.4**) on the K12 side chain over two steps (**3.5**) (**Scheme 3.2**, **Figure 3.S5**). NCL between the newly formed K12 ligation auxiliary and His_6-SUMO3(2-91)C47S-MES thioester, and subsequent zinc-mediated cleavage of the ligation handle N-O bond provided the sumoylated H4 peptide (**3.7**) (**Figures 3.S6, 3.S7**). The sumoylated H4 peptide was further ligated with recombinant **3.8** via diazotization of the C-terminal hydrazide (**Figure 3.S8**).[47,48] Radical-mediated desulfurization of Cys15 to the native Ala residue gave the full-length H4 substrate containing the native isopeptide linkage between SUMO3 and H4K12 (**3.9**) (**Figure 3.S9**)[49]

3.2.2 Semisynthesis of acetylated H3 for in vitro deacetylation assays

We are interested in studying the effect of suH4 on HDAC1 activity not only when a single H3 acetyllysine substrate is available at H3 K14, but, also in the presence of additional transcription activating PTMs — specifically additional acetylation marks. We hypothesized that these experiments could reveal any contextual differences in the ability of HDAC1 to initiate transcription repression, possibly through a hierarchy of specifity toward distinct acetyllysines in the H3 tail. My first goal was to ascertain whether several acetylated lysine resides would affect the overall activity of HDAC1, or its ability to bind a heavily modified histone substrate.

It was reported that the preferred site of lysine acetylation in histone H3 tails is at H3 K14 by acetyl transferases p300/CREB-binding protein (CBP) and p300/CBP-associated factor (PCAF), although all lysines in the H3 tail are observed to be acetylated in vivo.[50,51] In order to probe deacetylation of H3 by HDAC1, we designed an EPL approach for chemical access to all lysines in the H3 tail (**Scheme 3.3**). First, H3(1-28) C-terminal peptide hydrazides bearing either a single acetylation mark at lysine 14 (H3K14ac) or poly acetylation at lysines 9, 14, 18, 23 and 27 (H3polyac) were synthesized via Fmoc-SPPS (**3.10**) (**Figure 3.S10)**. Preparation of H3K14ac was performed by Caroline Weller (Caroline Weller, Ph.D thesis). Next, an H3Δ28 construct was cloned and transformed into *E. coli* as the His_{10}-SUMO3-H3(29-135)A29C,C110A fusion, bearing removable affinity purification and SUMO3 solubility tags. Following expression and IMAC purification of the H3 truncate fusion, cleavage of the His_{10}-SUMO3 tag was performed by incubation with the purified catalytic domain of the SUMO-isopeptidase, sentrin-specific protease 2 (SENP2, **Figure 3.S11**). Optimized cleavage conditions gave near quantitative cleavage in 8 h at 25 °C with 0.1 molar equivalent of SENP2 (**Figure 3.S12A**), thus eliminating the need for additional IMAC steps to remove residual H3 truncant fusion. The truncated H3(29-135)A29C protein was finally purified by RP-HPLC to remove remaining H3 truncate fusion. The H3 truncate was finally purified by RP-HPLC to yield tens of milligram quantities of the



Scheme 3.3. Semisynthesis of acetylated H3

H3Δ28 truncate (**3.12**)(**Figure 3.S12B,C**). The purified peptide hydrazide was converted to the corresponding azides (**3.11**) by oxidation with sodium nitrite, and then ligated to H3(29-135)A29C bearing an N-terminal cysteine to form the native amide bond between residues 28 and 29 (**Figure 3.S13**). Following desulfurization of C29 to the native Ala residue, the singly and polyacetylated H3 proteins were incorporated into histone octamers by refolding through dialysis with the remaining core histones (**Figure 3.S14, 3.S15**). The purified octamers were then assembled into mononucleosomes by mixing with high affinity 147 bp Widom 601 double stranded DNA and undertaking step-wise salt dilutions (**Figure 3.S16**). [52,53]

3.2.3 Deacetylation of nucleosomal substrates by the HDAC-CoREST1 sub-complex

With semisynthetic acetylated nucleosomes in hand, C-terminally FLAG-tagged HDAC1 was expressed in HeLa cells and purified by immunoprecipitation on M2 agarose resin. Accurate enzyme concentration was determined by SDS-PAGE and coomassie staining in comparison with bovine serum albumin (BSA) standards of known concentrations (**Figure 3.S17**). HDAC1

activity was confirmed with the Fluor de Lys (Enzo Life Sciences, Farmingdale, NY) HDAC fluorometric activity assay (**Figure 3.S18**). Isopeptidase activity from co-purified SUMO protease(s) was overcome through treatment with N-ethylmaleimide (Caroline Weller, Ph.D. thesis). Full-length wild-type His_6-CoREST1 and His_6-CoREST1-3A were purified from *E. coli* by IMAC using previously described protocols with minor modifications, and its final concentration determined by gel electrophoresis in comparison with BSA standards (**Figure 3.S19**).[38]

HDAC1 activity was monitored by first pre-forming the HC complex on ice for 30 minutes in an equimolar ratio of both proteins, followed by incubation with acetylated mononucleosomes containing either wild-type H4 or suH4 (**Figure 3.3**). Due to the additional HDAC1 substrate available in polyacetylated vs. singly acetylated nucleosomes (five acetyllysines compared to one acetyllysine), five-fold less of the polyacetylated substrates was used for a direct comparison of HDAC1 activity at each acetyllysine. We note that using the same concentration of H3polyac nucleosomes as that used in H3K14ac nucleosome deacetylation assays resulted in slower overall H3 deacetylation due to the 5-fold greater substrate present (**Figure 3.S20**), indicating no allosteric stimulation of HDAC1 activity by the acetyl groups themselves. Deacetylation assays using H3K14ac mononucleosomes were performed by Weller (Caroline Weller, Ph.D. thesis).

Time-points for the deacetylation assays were gathered and analyzed by quantitative Western blot analysis using either a H3K14ac-specific antibody or a pan-acetyllysine antibody that recognizes any acetylated lysine residue in H3 **(Figure 3.3**). The western blot membranes were also incubated with a second primary antibody towards one of the other assay components (i.e. H2A, H4 or HDAC-FLAG) as an internal loading control. Plotting the percentage of H3 acetylation relative to an initial zero time-point allowed us to monitor the activity of the HC complex over time, which revealed robust H3 deacetylation by the HC complex on both the

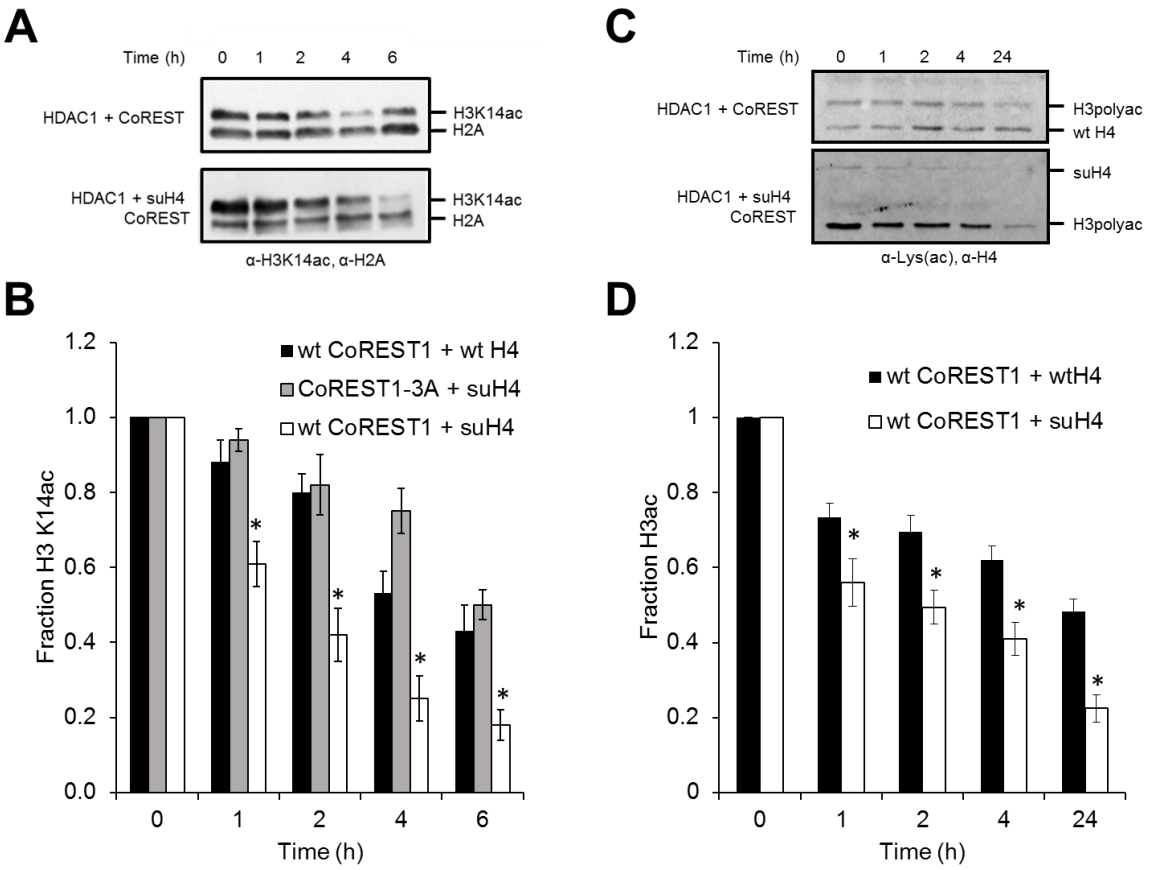


Figure 3.3. suH4 stimulates HC sub-complex deacetylation toward acetylated H3. (A) Representative Western blot time-course of H3K14ac deacetylation by HDAC1 using various histone H4 and CoREST1 constructs. (B) Histogram quantification of H3K14ac deacetylation. (C) Representative Western blot time-course of H3polyac deacetylation by HC sub-complex using either wild-type H4 or suH4. (D) Histogram quantification of H3polyac deacetylation. * $P < 0.05$. Error bars are standard error of the mean.

singly and polyacetylated substrates (**Figure 3.3**). Strikingly, the HC complex deacetylated H3 to >2-fold in the presence of suH4 compared to wild-type H4 over the same time period. The increased activity in the presence of suH4 was dependent on the interaction between SUMO3 and the SIM in CoREST1, because an HC sub-complex containing the CoREST1-3A mutant did not show enhanced deacetylation activity. Together, these results indicate that suH4 exerts a significant stimulatory effect on the HC sub-complex, similar to that observed for the LC sub-complex. Like the LC subcomplex, the enhanced activity was due to the interaction of sumoylated H4 and the intact SIM of CoREST1.

It is noteworthy that deacetylation of H3polyac mononucleosomes occurred at a slower overall rate to reach the same level of H3 deacetylation when compared to the singly acetylated substrate (24 hours versus 6 hours). However, the extent of H3 deacetylation in the early time-points of the experiment indicated a slightly increased rate of initial deacetylation. These results may indicate a preference of the HC complex to deacetylate specific lysine residues first and/or at a faster rate compared to other acetylated lysines within the H3 tail. To test for acetyllysine specificity we repeated the H3polyac nucleosome deacetylation assay and blotted time-points with either an H3K14ac-specific antibody and with the pan-acetyllysine antibody (**Figure 3.S21**). Results from this experiment demonstrated a slight, but statistically significant, difference in total H3 deacetylation at early time-points compared with H3K14ac-specific deacetylation at early time-points. We found that, not surprisingly, H3K14ac is deacetylated more slowly than the combination of all acetylated lysine residues in the H3 tail indicating that it is not a preferred substrate of HDAC1 However, this did not hold true at later time-points, and H3K14ac levels were less than global H3acetylation levels by the end of the time-course. Although more experimentation is necessary to characterize deacetylation rates at other lysine residues in the context of H3polyac nucleosomes, we suspect that K14ac may be a poorer HDAC1 substrate at the onset, due to the presence of acetylation at adjacent sites that may hinder binding to the active site of HDAC1. However, after deacetylation at adjacent sites has occurred H3 K14 appears to be a good substrate for deacetylation by HDAC1.

3.2.4 Quantitative characterization of the CoREST1-SUMO3 binding interface

Consensus SUMO interacting motifs are 3-4 hydrophobic amino acid stretches with the sequence (V/I)-X-(V/I)-(V/I) or (V/I)-(V/I)-X-(V/I/L), where X is generally another bulky aliphatic residue or Asp, Glu or Ser.[54,55] SIMs are often flanked by acidic patches of amino acids that aid in non-covalent interactions with the various SUMO isoforms. Thus far, SIMs have only been described for SUMO1-3 binding, while SIMs have yet to be demonstrated for SUMO 4 and 5,

even though SUMO4 shares ~82% sequence identity to SUMO2/3 and has nearly identical SIM binding cleft residues.[56–59] SUMO5, however, shares only ~46% sequences identity to SUMO2/3 (although ~87% with SUMO1) and is substantially variant in the SIM binding cleft (Chapter 1.4). Unsurprisingly, most known SUMO ligases contain at least one SIM in their sequence and have been shown to be critical for ligase activity.[60,61] Within the last several years, SIMs have been described in a plethora of other proteins, leading to an increased sense of the importance of protein sumoylation in cellular processes.[62]

The clear importance of the interaction between SUMO2/3 and the ncSIM of CoREST1 in our studies led us to further investigate this association. Until now, only qualitative methods have established a relationship between the two proteins. Previous studies have linked CoREST1 and SUMO through co-immunoprecipitation assays, and the Chatterjee lab has demonstrated enzymatic stimulation dependent on the SUMO3-CoREST1 interaction, and so we sought to quantitatively assess this important binding event for the first time.[26,42]

In order to generate the desired binding data we designed an HSQC NMR spectroscopy experiment in collaboration with the lab of Dr. Rachel Klevit (Department of Biochemistry, UW) and Dr. Rajan K. Paranji (Department of Chemistry, UW). By titrating increasing amounts of CoREST1 peptide, centered on the **I**-D-**I**-E-**V** hydrophobic core, into a solution containing isotopically labeled SUMO2/3, we expected to observe various backbone chemical shift perturbations (CSPs) of SUMO2/3 residues that are either directly or indirectly affected by CoREST1 binding. We anticipated that this experiment could elucidate the equilibrium dissociation constant (K_d) of the SUMO2/3-CoREST1 binding interaction as well as map the CoREST1 binding surface on SUMO. In these experiments, CSPs should continually increase until the SUMO binding interface is saturated with CoREST1 peptide. Binding curves reflect this saturation and can be fit with non-linear regression to derive the binding K_d. Unfortunately,

previous attempts to arrive at binding data between full-length CoREST1 or CoREST1(1-300) and SUMO3 using fluorescence polarization (FP) or Förster resonance energy transfer (FRET) experiments were unsuccessful due to the challenge of obtaining CoREST1 at high enough concentrations to saturate SUMO3 binding.

For NMR experiments, we expressed and purified ^{15}N-labeled His_6-SUMO3 from *E. coli.* In addition, we prepared an extended CoREST1 ncSIM peptide (252-277) as the C-terminal amide with acetylated N-terminus, as well as the corresponding 3A mutant via Fmoc-SPPS (**Figures 3.S22, 3.S23**). Upon dissolving labeled SUMO3 in a 10% D_2O-potassium phosphate buffer, the ^{1}H-^{15}N-HSQC NMR spectrum was collected. CoREST1 peptides were then titrated into the solution and subsequent NMR spectra were collected from a range of 1 to 24 molar equivalents of added CoREST peptide (**Figure 3.4A**). The resulting spectra displayed clear CSPs for several SUMO3 residues. The most perturbed SUMO3 residues were all found in, or adjacent to, the well-characterized consensus SIM binding-cleft between strand β2 and helix α1 of SUMO2/3, indicating a specific binding site for the peptide (**Figure 3.4B,C**). Plotting the magnitude of residue-specific CSPs against concentration of titrated CoREST1 peptide generated binding saturation curves. An average K_d of 1.98 ± 0.13 mM was extracted from ten binding saturation curves using GraphPad Prism software (**Table 3.2, Figure 3.S24**). This is in contrast to previously determined K_d values of 1-180 μM for proteins containing consensus SIMs but may speak to the fact that the peptide is entirely out of its protein context..[54,63,64] Importantly, when the CoREST1-3A peptide was used in identical NMR titration experiments, the CSPs did not saturate over the same concentration range for added peptide and we estimated a K_d > 20 mM (**Figure 3.S25**). This result confirms the importance of specific hydrophobic residues in the ncSIM of CoREST1 for SUMO binding and stimulation of HC and LC sub-complex activities.

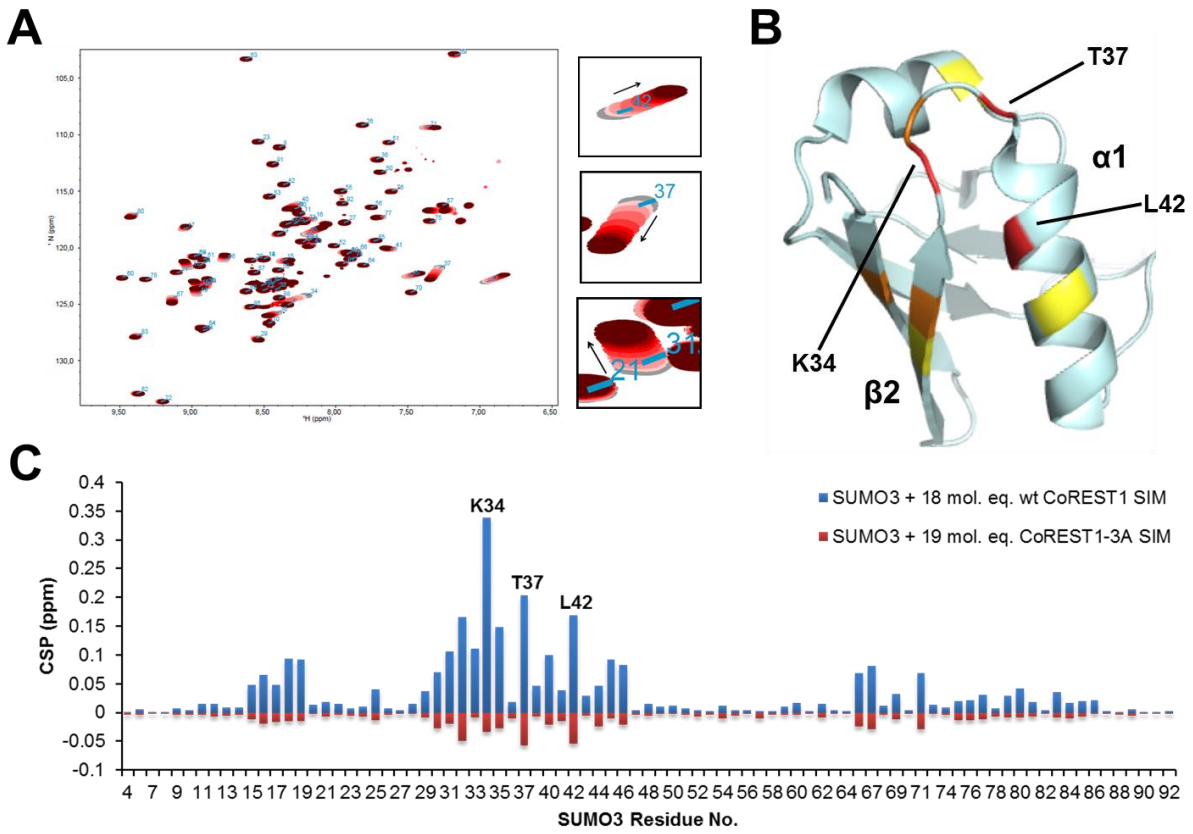


Figure 3.4. The CoREST1 SIM peptide binds SUMO3. (A) Overlaid ^{1}H-^{15}N-HSQC NMR spectra of free SUMO3 and 1-24 molar equivalents of titrated CoREST1(252-277). Magnified regions depict prominent chemical shift perturbations (CSPs) for L42, T37 and F31 upon increasing equivalents of CoREST1-SIM. (B) Crystal structure of SUMO3 with highlighted residue effected by CoREST1(252-277) binding. The three residues in red (K34, T37 and L42) indicate residues with the three largest CSPs. The three residues in orange (L19, F31 and R35) indicate the next three largest CSPs. The four residues in yellow (Q30, A45, N67 and T71) indicate the next four largest CSPs. PBD: 1U4A (C) Histogram depicting the absolute magnitude of CSP for each SUMO3 residue when either CoREST1(252-277) or CoREST1(252-277)-3A are added.

Table 3.2. Measured dissociation constants for ncSIM-SUMO3 binding interactions

ncSIM peptide	Dissociation constant (K_d, mM)
CoREST1(252-277)	1.98 ± 0.13
CoREST1(252-277)-3A	>20
FIP1L1(120-145)	3.57 ± 0.18
FIP1L1(120-145)-3A	>12

The relatively weak observed association (large K_d value) between the CoREST1 ncSIM peptide and SUMO3 raised additional questions about the physiological importance of this interaction. We surmise that either additional structural elements of full-length CoREST1 are needed to strengthen the binding interaction with SUMO3, or, that the weak SUMO-ncSIM associations are part of broader multivalent interactions between protein complexes and chromatin. For example, CoREST1 already binds nucleosomal DNA by means of its SANT2 domain and the interaction with SUMO3 may tailor its precise positioning on nucleosomes. To assess whether this weak affinity is specific to the CoREST1–SUMO interaction, or whether other proteins containing ncSIMs bind with similar affinity, we generated the Factor interacting with PAPOLA and CPSF1 (FIP1L1) (120-145) peptide and the corresponding 3A SIM variant (L138A, V140A and L142A, collectively called FIP1L1-3A) (**Figure 3.S26**). FIP1L1 (a pre-mRNA processing complex subunit) was previously reported to contain an ncSIM and bind SUMO2.[38] The wild-type FIP1L1 peptide was used in NMR titration experiments which revealed a K_d = 3.57 ± 0.18 mM, while the FIP1L1-3A bound SUMO3 with >3-fold weaker affinity, thus altering but not completely preventing peptide binding (**Table 3.2**). Again, the most perturbed residues were located in the SIM binding cleft of SUMO3 (**Figures 3.S27, 3.S28**). The weaker observed K_d for the FIP1L1 peptide is in agreement with slightly weaker association for full-length FIP1L1 with SUMO2 (relative to full-length CoREST1) via previously reported co-precipitation studies[38] The similar observed SUMO3 binding affinities for the CoREST1 and FIP1L1 peptides suggest that weak interactions may be a common property among proteins containing ncSIMs. This is an exciting discovery that will be further investigated in the Chatterjee labs, for example by searching for ncSIMs in other chromatin-associated proteins.

3.3 Conclusions and outlook

In conclusion, we have delineated a role for histone sumoylation in stimulating the histone deacetylation activity of the HC sub-complex. The enhanced HDAC activity toward acetylated

H3 was observed with both singly and polyacetylated substrates. In addition, the enzymatic stimulation of HDAC1 imparted by suH4 was dependent on the interaction between CoREST1 and SUMO3, as mutations in the CoREST1 SUMO interacting motif abrogated the enhanced activity. This phenomenon was confirmed through quantification of the CoREST1-SUMO3 binding interaction via HSQC NMR experiments, which revealed clear chemical shift perturbation in the SIM binding cleft of SUMO3. While the observed association between SUMO3 and CoREST was relatively weak, we surmise that additional structural components of CoREST1 are needed to further promote the interaction, or that weak SUMO-SIM associations are part of broader multivalent interactions in the context of chromatin. The results from this study raise interesting questions regarding HDAC1 specificity for the H3 tail residues, and the importance of multi-valent protein binding events in gene regulation. Future studies will aim to identify preferred or hindered H3 acetyllysine sites as HDAC substrates and whether histone sumoylation directly influences HDAC1 specificity. The activity of HDAC1 and LSD1 will also be explored in the context of the complete LCH complex in order to identify the effect of different PTMs on gene repression by LCH.

3.4 Experimental procedures

3.4.1 General methods

Fmoc-Lys(Boc)-Wang resin was purchased from Sigma-Aldrich Chemical Company (St. Louis, MO), and 2-chlorotrityl chloride and Rink amide resins were 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). 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 (RPHPLC) was performed on a Varian (Palo Alto, CA) ProStar HPLC or Agilent (Santa Clara, CA)1260 Infinity II 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 or Bruker Autoflex II MALDI-MS. 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).

3.4.2 Solid-phase peptide synthesis

Synthesis of H3(1-28)K14ac-C(O)NHNH$_2$ and H3(1-28)polyac-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 Fmoc-SPPS. 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. 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 *tert*-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. Fmoc-Lys(ac)-OH was used at indicated positions. Peptides were cleaved and deprotected by reaction of the resin at 20 µL/mg with standard cleavage cocktail (TFA:H_2O:triisopropylsilane:anisole 95:2.5:2.5 v/v) 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 preparative RP-HPLC with a gradient of 0-

50% B. This yielded 11% of the peptide H3(1-28)K14ac-C(O)$NHNH_2$ and 9% of H3(1-28)polyac-C(O)$NHNH_2$ based on initial resin loading. ESI-MS of H3(1-28)K14ac-C(O)$NHNH_2$. Calculated *m/z* $[M+H]^+$ 3,052.5 Da, observed 3,054.3 ± 1.5 Da (Caroline Weller, Ph.D. Thesis). ESI-MS of H3(1-28)polyac-C(O)NHNH2. Calculated *m/z* $[M+H]^+$ 3,221.7 Da, observed 3,221.5 ± 0.5 Da (**Figure 3.S10**).

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 Nα-deprotection chemistry. The 2-chlorotrityl 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), oxyma (1.31 mmol) and DIEA (1.31 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 N-terminus with the Boc group by reaction with ditert-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 mol. eq.) with *N,N′*-diisopropylcarbodiimide (DIC, 8

mol. eq.) 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.[43] Complete displacement was judged by test cleavage and subsequent ESI-MS analysis. Peptide was cleaved and deprotected by reaction of resin at 100 µL/mg with TFA:thioanisole:H_2O:TIS:1,2-ethanedithiol (90:2.5:2.5:2.5:2.5 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 9% of the peptide-auxiliary conjugate H4(1-14)K12aux-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 3.S5**).

Synthesis of CoREST1(252-277), CoREST1(252-277)-3A, FIP1L1(120-145) and FIP1L1(120-145)-3A

The peptides H_2N-REREESEDELEEANGNNPIDIEVDQN-C(O)NH_2 corresponding to CoREST1(252-277), the CoREST1-3A variant, H_2N-KVKGVDLDAPGSINGVPLLEVDLDSF-C(O)NH_2 corresponding to FIP1L1(120-145) and the FIPL1-3A variant were synthesized by microwave-assisted SPPS on a 0.05 mmol scale employing standard Fmoc-SPPS (CoREST1 peptides were synthesized and purified by C.J.A. Leonen). Briefly, each amino acid was coupled on Rink Amide resin (0.33 mmol/g) in 5 molar excess based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 5% piperazine/0.1 M HOBt 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 (0.25 mmol), oxyma (0.25 mmol) and DIC (0.25 mmol) in DMF. For Arg, an additional coupling reaction was performed for 25 min at 75 °C. After removal of the final Fmoc-protecting group, N-termini were acetylated by incubating the resin in 50:50 DMF:DCM

containing DIEA (40 mol. eq.) and acetic anhydride (20 mol. eq.) for 2x15 min. Peptides were cleaved and deprotected by reaction of the resin at 100 µL/mg with standard cleavage cocktail (TFA:H_2O:triisopropylsilane, 95:2.5:2.5 v/v) for 3 h 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 preparative RP-HPLC with a gradient of 20-80% B. ESI-MS of CoREST1(252-277). Calculated *m/z* $[M+H]^+$ 3,084.1 Da, observed 3,083.4.3 Da (**Figure 3.S23**). ESI-MS of CoREST1(252-277)-3A. Calculated *m/z* $[M+H]^+$ 2.971.9 Da, observed 2,971.2 Da (**Figure 3.S23**). ESI-MS of FIP1L1(120-145). Calculated *m/z* $[M+H]^+$ 2,738.0 Da, observed 2738.2 Da (**Figure 3.S26**). ESI-MS of FIP1L1(120-145)-3A. Calculated *m/z* $[M+H]^+$ 2,627.9 Da, observed 2,626.0 Da . (**Figure 3.S26**).

3.4.3 Overexpression and purification of SENP2 protease

E. coli BL21(DE3) cells containing the plasmid pRK793-His_6-SENP2 were grown in 6 L 2xYT (yeast extract, tryptone) supplemented with 30 µg/mL of Kanamycin 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 3 h at 37 °C. The cells were harvested by centrifugation at 7,000xg for 15 min. The cell pellet was resuspended in 15 mL lysis buffer: 20 mM tris, 350 mM NaCl, pH 7.4. Cells were lysed by sonication then centrifuged at 20,000xg for 15 min. The lysate supernatant was passed through a 0.45 µm filter to remove insoluble particles, then applied to ~10 mL Ni-NTA column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 1.5 h at 4 °C. The column was then washed with 2 column volumes (CV) of column buffer: 50 mM HEPES, 100 mM NaCl, containing 20 mM imidazole, 2 CV column buffer containing 75 mM imidazole, and 2 CV column buffer containing 100 mM imidazole. His_6-SENP2 was eluted with 10 CV column buffer containing 400 mM imidazole, and dialyzed into 3.5 L of column (no imidazole) for 3 h at 4 °C. Protein precipitation occurs during dialysis. Concentration of the suspension was determined by comparing the band intensity of His_6-

SENP2 protease on a coomassie-stained 15% SDS-PAGE gel relative to BSA standards of known concentration (**3.S11**).

3.4.4 Overexpression and purification of TEV protease

E. coli BL21(DE3) cells containing the plasmid pRK793-His_6-TEV were grown in 6 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 7,000xg 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 20,000xg for 15 min. The lysate supernatant was passed through a 0.45 µm filter then applied to ~10 mL Ni-NTA column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 1.5 h at 4 °C. The column was then washed thoroughly with lysis buffer containing 20 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 dithiothreitol (DTT), pH 7.5 for 18 h at 4 °C. Protein precipitation occurred during dialysis. Concentration of the suspension was determined by comparing the band intensity of His_6-TEV protease on a coomassie-stained 15% SDS-PAGE gel relative to BSA standards of known concentration (**Figure 3.S2**).

3.4.5 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 6 L 2xYT 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 7,000xg for 15 min. H4(15-102)A15C was purified using a previously established protocol.[65] Cells were resuspended in wash buffer (20 mM tris, 200 mM NaCl, 1mM 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 20,000xg 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 ~10 mL Ni-NTA resin. Column binding proceeded overnight at 4 °C, after which the resin was washed with 10 CV (column volumes) extraction buffer containing 25 mM imidazole. The protein was eluted with 2 x 2 CV extraction buffer containing 400 mM imidazole, then dialyzed into water containing 1 mM DTT (**Figure 3.S1**). 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 a concentration of 3.8 μM (0.1 mg/mL, 0.1 mol. eq.), 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 H4 N-terminal His_6-tag, and the column flow-through containing H4(15-102)A15C was 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* $[M+H]^+$ 10,070.9 Da, observed 10,071.9 ± 1.5 Da (**Figure 3.S3**).

3.4.6 Overexpression and purification of H3(29-135)A29C,C110A

E. coli BL21(DE3) cells containing pET15b-His_{10}-SUMO3-H3(29-134)A29C,C110A were grown in 6 L 2xYT 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 IPTG and cells were grown for an additional 2 h at 37 °C. The cells were harvested by centrifugation at 7,000xg for 15 min. Cells were resuspended in wash buffer (20 mM tris, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 7.4, 0.1% triton X-100) and lysed by sonication on ice. Inclusion bodies were pelleted by centrifugation at 20,000xg for 20 min and washed twice with wash buffer. Inclusion bodies were then dissolved in extraction buffer (6 M Gn-HCl, 20 mM tris, 100 mM NaCl,1 mM 2-mercaptoethanol, pH 7.4) and applied to ~10 mL Ni-NTA resin. Column

binding proceeded for 1.5 h at 4 °C, after which the resin was washed with 5 CV extraction buffer containing 20 mM imidazole. The protein was eluted with 5 CV extraction buffer containing 400 mM imidazole, then dialyzed into water containing 1 mM DTT. After dialysis, precipitated histone fusion was lyophilized to dryness. 30-40 mg of the dried histone was taken up in cleavage buffer consisting of 25 mM tris, pH 8, 150 mM NaCl, 25 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 5 mM DTT, 0.15% tween-20 to a concentration of 1 mg/mL. Purified SENP2 protease was added to a concentration of 4.2 μM (0.1 mg/mL, 0.1 mol. eq.). $Ar_{(g)}$ was bubbled through the reaction and then proceeded overnight at 37 °C. Histone proteins were pelleted by centrifugation was with 2x30 mL H_2O and lyophilized to dryness. H3(29-1035)A29C,C110A was purified by C4 preparative RP-HPLC employing a gradient of 40-70% B over 60 min. Typical yields were 5-6 mg/L of cell culture. ESI-MS of H3(29-135)A29C,C110A. Calculated *m/z* $[M+H]^+$ 12,274.4 Da, observed 12,274.6 ± 0.6 Da (**Figure 3.S12**).

3.4.7 Overexpression and purification of His_6-SUMO3(2-91)C47S-MES

E. coli BL21(DE3) cells were transformed with the plasmid pTXB1-His_6-SUMO3(2-91)C47S with C-terminal AVA intein fusion. 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 7,000xg for 15 min. The cell pellet was resuspended in lysis buffer: 50 mM NaH_2PO_4, pH 8, 300 mM NaCl, containing 5 mM imidazole. Cells were lysed by sonication then centrifuged at 20,000xg for 15 min. The lysate supernatant was passed through a 0.45 μm filter then applied to a ~10 mL Ni-NTA resin pre-equilibrated with lysis buffer. Proteins were bound to the column for 1 h at 4 °C. The column was then washed with 5 CV of lysis buffer, followed by 5 CV of lysis buffer containing 20 mM imidazole and 2.5 CV lysis buffer containing 50 mM imidazole. SUMO-3(2-91)C47S-AVA intein fusion was eluted

from the column with 5 CV lysis buffer containing 250 mM imidazole. Eluted protein was dialyzed against 2 L 100 mM NaH_2PO_4, pH 7.2, 150 mM NaCl containing 1 mM MESNa for 2x1 h. Intein thiolysis was induced through addition of additional MESNa to a final concentration of 100 mM and incubated at 30 °C for 18 h. Thiolysis reaction were then lyophilized to dryness and purified by C18 preparative RP-HPLC employing a gradient of 30-60% B (H4) over 60 min. Typical yield is 5-6 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.S4**).

3.4.8 Overexpression and purification of ^{15}N-labeled His_6-SUMO3

Bacterial expression plasmid pET28a containing human His_6-SUMO2 was used to clone the RING domain of Hrd1 C-terminally of SUMO3 via BamHI and XhoI. The protein was expressed in BL21 Escherichia coli (Invitrogen) in minimal MOPS media supplemented with $^{15}NH_4Cl$, ^{13}C glucose and D_2O (Cambridge Isotopes). SUMO3 was purified by Ni-NTA pull-down from bacterial lysate, followed by sequential on-column cleavage with human SENP1 and then with thrombin (Sigma).[66] After size exclusion chromatography on SDX75, SUMO3 was buffer exchanged into 10 mM potassium phosphate pH 6.5, 100 mM potassium chloride, 2 mM DTT, 0.1 mM EDTA, H_2O (90%), D_2O (10%) (NMR buffer). Labeled product was confirmed by C18 analytical RP-HPLC, 0-73% B over 30 min, and ESI-MS. Observed *m/z* $[M+H]^+$ 11,875.5 ± 5.7 (**Figure 3.S22**).

3.4.9 Expressed protein ligation of H4(1-14)K12aux-C(O)$NHNH_2$ and SUMO3(2-91)C47S-MES

Purified H4(1-14)K12aux-C(O)$NHNH_2$ (17.6 mg, 12.9 µmol) and SUMO3(2-91)C47S-MES (22.5 mg, 2.2 µmol, 6 mol. eq.) were dissolved in 7.5 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)K12su(aux)(C47S)-C(O)$NHNH_2$. Calculated *m/z* $[M+H]^+$ 11,666.1 Da, observed 11,667.4 ± 3.4 Da (**Figure 3.S6**).

3.4.10 Zn-mediated auxiliary removal from H4(1-14)K12su(aux)(C47S)-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 14.5 mg purified H4(1-14)K12su(aux)(C47S)-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 the reduced product 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 preparative RP-HPLC with a gradient of 25-50% B over 45 min to give 9 mg of reduced product (62%). ESI-MS of H4(1-14)K12su(aux)(C47S)-C(O)$NHNH_2$. Calculated *m/z* $[M+H]^+$ 11,593 Da, observed 11,593.3.4 ± 6.3 Da (**Figure 3.S7**).

3.4.11 Expressed protein ligation of H4(1-14)K12su(C47S)-C(O)$NHNH_2$ and H4(15-102)A15C

Ligation was accomplished by first converting the C-terminal hydrazide of H4(1-14)K12su(C47S)-C(O)$NHNH_2$ to an acyl azide with $NaNO_2$ via the diazotization reaction, as described previously.[47] 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)K12su(C47S)-C(O)$NHNH_2$ (2.5 mg, 0.216 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 4.3 μL of 1 M 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 (5mg, 0.539

μmol, 2 mol. eq.) was 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 analytical RP-HPLC employing a gradient of 40-65% B over 30 min to give 1.5 mg (32%). ESI-MS of H4(A15C)K12su(C47S). Calculated *m/z* $[M+H]^+$ 21,628.7 Da, observed 21,625.0 ± 6.0 Da.

3.4.12 Expressed protein ligation of H3(1-28)K14ac-C(O)NHNH$_2$ or H3(1-28)polyac-C(O)NHNH2 and H3(29-135)A29C,C110A

Ligation was accomplished by first converting the C-terminal hydrazide of acetylated H3(1-28)-C(O)NHNH$_2$ constructs to an acyl azide with $NaNO_2$ via the diazotization reaction, as described previously.[47] Purified acetylated H3(1-28)-C(O)NHNH$_2$ was dissolved at 5 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 23 μL of a 1 M solution of $NaNO_2$ in water. The reaction was briefly mixed, then allowed to proceed 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, were added to the reaction. Immediately following this addition, 200 μL of a solution containing H3(29-135)A29C,C110A (4 mg, 0.33 μmol, 7 mol. eq.) dissolved at 0.5 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl was added. 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 analytical RP-HPLC employing a gradient of 35-70% B over 30 min. 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. ESI-MS of H3(1-135)polyac(A29C,C110A). Calculated m/z [M+H]+ 15,466.8 Da, observed 15,466.1 ± 0.8 Da (**Figure 3.S13**).

3.4.13 Desulfurization of H3K14ac(A29C,C110A), H3polyac(A29C,C110A) and H4K12su(A15C,C47S)

Purified H3(K14ac(A29C,C110A), H3polyac(A29C,C110A) or H4K12su(A15C) were 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) (H3K14ac). Calculated *m/z* $[M+H]^+$ 15,266.8 Da, observed 15,272.1 ± 7.0 Da. MALDI-TOF MS of H3(1-135)polyac(C110A) (H3polyac). Calculated *m/z* $[M+H]^+$ 15,434 Da, observed 15,439.9 Da (**Figure 3.S14**). ESI-MS of H4(1-102)K12su(C47S) (suH4). Calculated *m/z* $[M+H]^+$ 21,596.7 Da, observed 21,602.9 ± 5.5 Da (**3.S9**).

3.4.14 Overexpression and purification of H2A, H2B 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*). *E. coli* BL21(DE3) cells containing a pET3a-histone plasmid were grown in 6 L of 2xYT medium at 37 °C until OD600 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 'wash buffer' containing 20 mM tris, 200 mM NaCl, pH 7.5, 1 mM EDTA, 0.1% triton X-100 and lysed by sonication on ice. The lysate was centrifuged at 20,000xg to separate insoluble inclusion bodies containing histone protein. Inclusion bodies were washed again with wash buffer and centrifuged Histones were extracted with buffer containing 6 M Gn-HCl, 20 mM tris, pH 7.5. The solubilized histones were then dialyzed against 3.5 L H_2O, 1 mM DTT for 12 h at 4 °C to

precipitate crude proteins. Proteins were further purified to homogeneity by RP-HPLC (**Figure 3.S14**).

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

3.4.15 Overexpression and purification of His_6-CoREST1 and His_6-CoREST1-3A

The pET28b-His_6-CoREST1 and pET28b-His_6-CoREST1-3A plasmids were a kind gift from Dr. Grace Gill. Plasmid was freshly transformed into Rosetta (DE3) competent cells prior to each expression. For protein overexpression, *E. coli* cells containing the Rosetta (DE3) plasmid were grown at 37 °C in 2 L 2xYT medium containing 30 µg/mL kanamycin to an OD_{600} ~0.5, then cooled at 16 °C for 1 h without shaking. Protein expression was induced by the addition of 0.2 mM IPTG to the growth media and the cells were allowed to grow at 16 °C for 18 h. At the end of the induction period, the cells were harvested by centrifugation at 5,000 rpm, 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 (pulsed sonication 30, power level 6, 2 x 1 min.). The lysate was then centrifuged at 15,000 rpm and the supernatant passed through a 0.45 µm filter and applied to 5 mL Ni-NTA resin. Lysate was incubated with the column at 4 °C for 45 min, after which the column was washed with 10 CV buffer containing 50 mM tris, 250 mM NaCl, 10% glycerol, pH 8 (4 °C) (column buffer), 25 mM imidazole followed by 10 CV column buffer containing 100 mM imidazole and 5 CV column buffer containing 150 mM imidazole. His_6-CoREST1 or His_6-CoREST1-3A was eluted with 5 CV buffer containing 300 mM imidazole. Elution fractions were analyzed on a 12% SDS-PAGE gel, and pure fractions dialyzed against 50 mM HEPES, 250 mM NaCl, 10% glycerol, pH 8, for 3 h at 4 °C. After dialysis, samples were concentrated with 3,000 MWCO centrifugal concentrators at 4 °C (4-5 h). Protein concentration was determined by comparing the band intensity of His_6-CoREST1 or

His_6-CoREST1-3A on a coomassie-stained 12% SDS-PAGE gel relative to BSA standards of known concentration (**Figure 3.S19**).

3.4.16 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.[67] 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. 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 (**Figure 3.S17**).

Notes on HDAC1 purification (Caroline Weller, Ph.D. hesis)

In deacetylation assays with mononucleosomes (MNs), western blotting for histone H4 revealed desumoylation activity over the course of the assay. We suspected that trace amounts of a

desumoylating enzyme may be present, as HDAC1 is known to interact with SENP1.[68] Reaction of 0.5 μM purified HDAC1 with 1 μM of a SUMO3. C-terminal 7-amido-4-methylcoumarin derivative (SUMO3 AMC, Boston Biochem) in MN assay buffer indded revealed a moderate amount of SENP activity. The SENP2 catalytic domain was included as a positive control in this assay. Therefore, to inhibit co-purifying SENP activity, the 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, and the reduced SENP activity confirmed by reaction with SUMO3 AMC and suH4 containing MNs. HDAC1 deacetylation activity was re-confirmed by Fluor de Lys assay (**Figure 3.S18**), and by reaction with suH4 containing MNs.

3.4.17 Histone octamer formation

Histone octamers were assembled as previously reported, excluding reducing agents (e.g. DTT) to prevent downstream disruption of structurally important disulfides or metal-binding cysteine residues in enzymatic assays.[53] 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. Concentrations were determined using the 280 nm extinction coefficients: H2A, ε = 4,470 $M^{-1}cm^{-1}$; H2B, ε = 7,450 $M^{-1}cm^{-1}$; H3 C110A, ε = 4,470 $M^{-1}cm^{-1}$; H3 K14Ac, ε = 4,470 $M^{-1}cm^{-1}$; H3 K14polyac, ε = 4,470 $M^{-1}cm^{-1}$;H4, ε = 5,960 $M^{-1}cm^{-1}$; suH4, ε = 7,450 $M^{-1}cm^{-1}$. 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, ε = 44,700 $M^{-1}cm^{-1}$; H3polyac, ε = 44,700 $M^{-1}cm^{-1}$; H3K14Ac/suH4, ε = 47,680 $M^{-1}cm^{-1}$. H3polyac/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.

3.4.18 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% polyacrylamide/TBE gel run at 140 V for 30 min and stained with ethidium bromide.

3.4.19 Mononucleosome deacetylation assays

HDAC1 and CoREST1 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/CoREST1 solution was then combined with H3K14Ac or H3polyac MNs for final concentrations of 200 µM H3K14ac MN or 40 nM H3polyac MN, and 800 nM each of HDAC1 and CoREST1, in reaction buffer composed of 50 mM HEPES, 100 mM KCl, 63 mM NaCl, 0.15 mMDTT, 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) or α-H4 (Abcam 10158, lot #GR3186358-2) at 1:2500, and primary α-H3K14ac antibody (Abcam 52946, lot #GR149741-17) or α-panLys(ac) (Millipore AB3879, lot #2928563) at 1:1000 dilutions. Membranes were subsequently washed and incubated with goat α-rabbit-IR Dye 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. H3K14ac or panLys(ac) signal was normalized to H2A or H4 loading control signal, and quantified using NIH ImageJ software.

3.4.20 HSQC NMR titration experiments with ^{15}N-labeled SUMO3 ad CoREST1 or FIP1L1 ncSIM peptides

Sample Preparation: ^{15}N-labeled SUMO3 was diluted to 200 μM in 250-500 μL NMR buffer and the ^{1}H-^{15}N-HSQC spectrum was collected. Subsequently, 1-24 molar equivalents of ncSIM peptide were added from a stock solution (16-47 mM) dissolved in NMR buffer.

NMR data collection: Comparative titration studies of SUMO3-ncSIM peptides were prepared in a 3mm high field NMR tube (Norell Inc., Morganton, NC). ^{1}H-^{15}N TROSY-HSQC spectra were recorded on a Bruker (Billerica, MA) Avance III 800 MHz spectrometer equipped with a triple resonance cryoprobe (TCI-HCN-Z gradient) at 290K. A total of 200 slices along F1 (^{15}N) dimension with 1024 points along F2 dimension (^{1}H) were recorded with 8 scans per slice using the TROSY-HSQC pulse sequence that includes 3-9-19 watergate for solvent suppression.[69] Raw data were fourier transformed using Bruker Topspin 3.6 software with 2048 and 512 points in F2 and F1 dimensions, respectively, using a shifted Squared Sine Bell window function along both dimensions.

Data analysis: Processed spectra were imported into NMRViewJ software (v.9.2.0, One Moon Scientific, Inc) for titration fitting according to the following procedure. The effective value of

chemical shift perturbation (CSP) of the NH cross peaks of SUMO3 protein upon binding to a given concentration of the peptide is computed as follows, with the relevant definitions of following the parameters:

$$DCSP^{X} = ((\Delta CS^{1H})^2 + (0.2 * \Delta CS^{15N})^2)^{\frac{1}{2}}$$

where, ΔCS^{1H} is the change in Chemical Shift along 1H axis with respect to the control spectrum; ΔCS^{15N} is change in Chemical Shift along 15N axis with respect to the control spectrum; $DCSP^{X}$ is the effective change in chemical shift at a concentration of the binding peptide given by X μM. The measured $DCSP^{X}$ is related to the dissociation constant K_d by the following equation:

$$DCSP^{X} = DCSP^{0} + (DCSP^{\infty} - DCSP^{0}) * \frac{[(P_t + X + 10^{logKd}) - \{(P_t + X + 10^{logKd})^2 - 4 * P_t * X\}^{\frac{1}{2}}]}{2 * P_t}$$

Where, $DCSP^{0}$ is the canonical CSP value at 0 μM of peptide; $DCSP^{\infty}$ is the canonical CSP value at ∞ μM of peptide; K_d is equilibrium dissociation constant; Pt is the absolute concentration of the protein substrate in μM

NMRViewJ software automatically measures the effective chemical shift perturbation values for the plausible residues and fits the above equation to the measured shifts as a function of the peptide concentration. The final result of the analysis provides the dissociation constant in μM on a per residue basis, which helps us to obtain a protein wide profile of the binding regions.

3.5 Product characterization and supplemental data

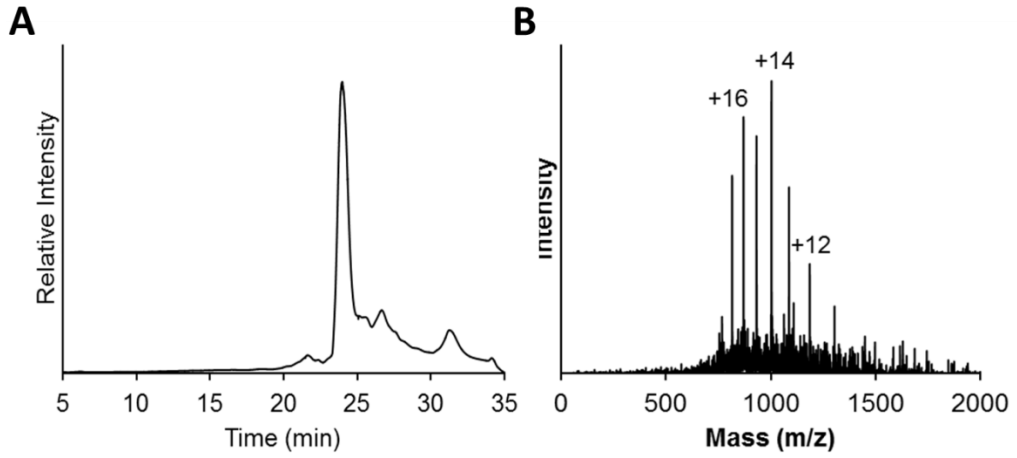


Figure 3.S1. Characterization of His_6-TEV-H4(15-102)A29C,C110A. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS. Calculated *m/z* $[M+H]^+$ 13,029.3 Da, observed 13,026 ± 0.2 Da.

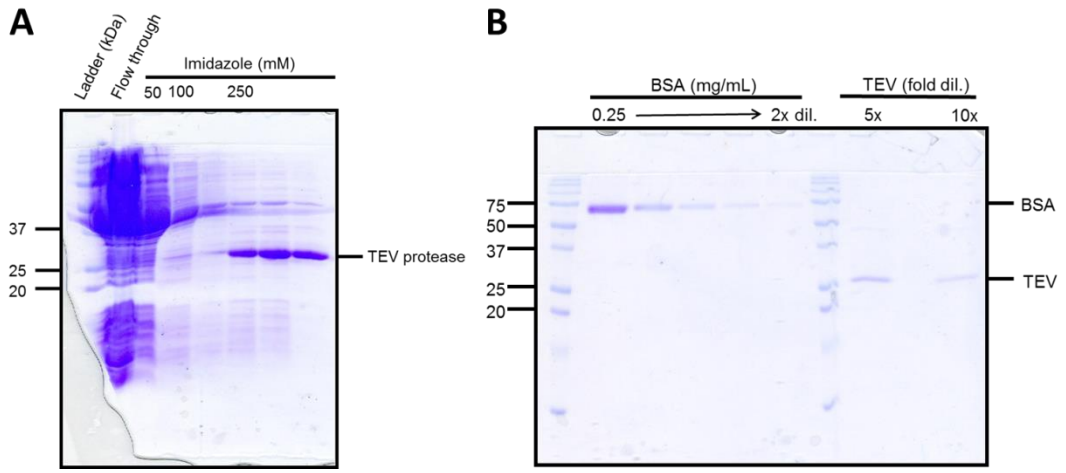


Figure 3.S2. Purification of His_6-TEV protease. (A) 15% SDS-PAGE of Ni-NTA IMAC purification of His_6-TEV protease. (B) Quantification of dialyzed TEV protease protein concentration through comparing densities to BSA standards of known concentration.

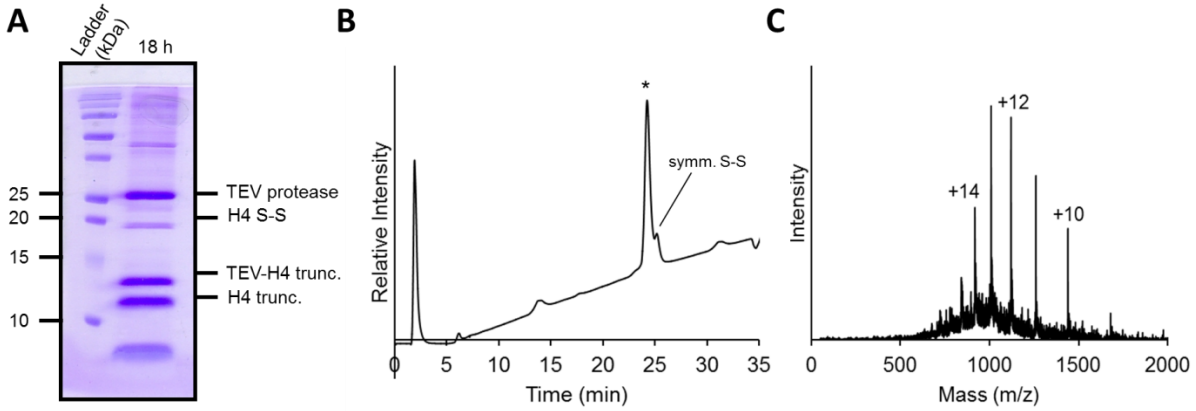


Figure 3.S3. Cleavage of His_6-TEV-H4(15-102)A29C,C110A. (A) 15% SDS-PAGE of TEV cleavage reaction after 18 h. (B) C18 analytical RP-HPLC of purified H4(15-102)A29C,C110A, 0-73% CH_3CN in H_2O, 30 min. gradient. (C) ESI-MS of purified H4(15-102)A29C,C110A.

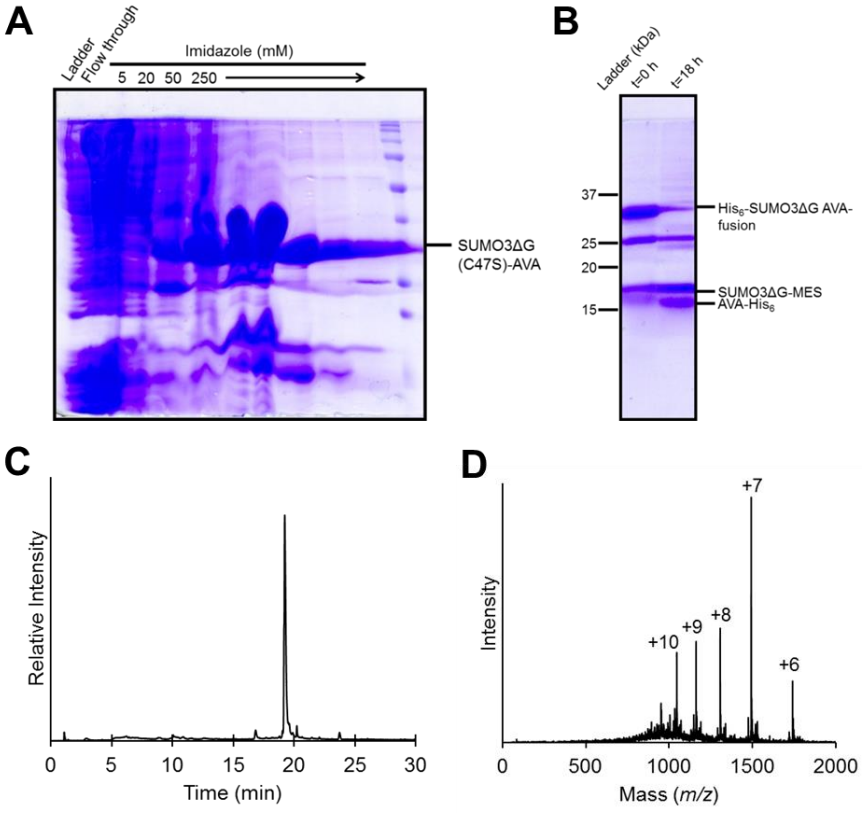


Figure 3.S4. Purification of His_6-SUMO3(2-91)C47S-MES. (A) 15% SDS-PAGE of Ni-NTA IMAC purification of His_6-SUMO3(2-91)C47S AVA-intein fusion. (B) 15% SDS-PAGE of intein-mediated thiolysis after 18 h (C) RP-HPLC of purified His_6-SUMO3(2-91)C47S-MES, 0-73% CH_3CN in H_2O, 30 min. gradient. (D) ESI-MS of purified His_6-SUMO3(2-

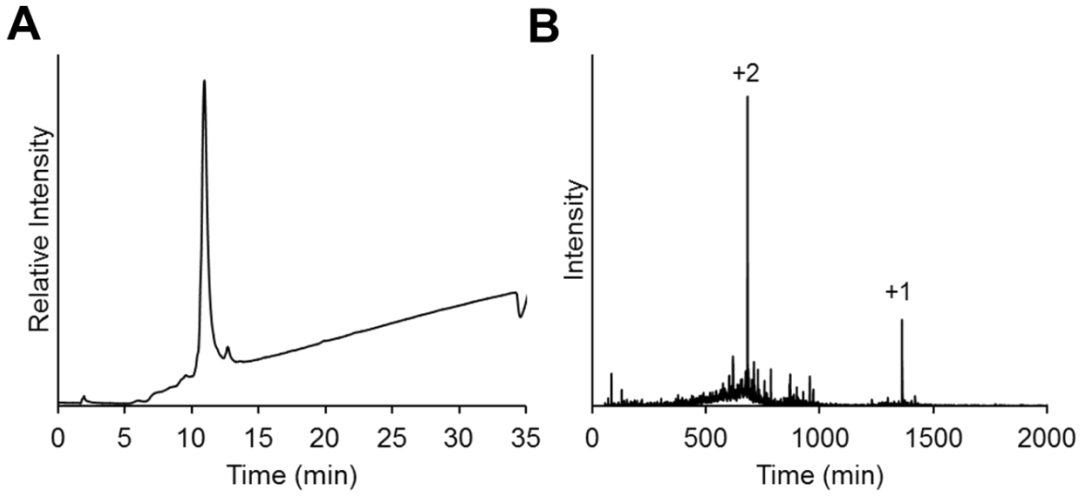


Figure 3.S5. Characterization of purified H4(1-14)K12aux-NHNH$_2$. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS.

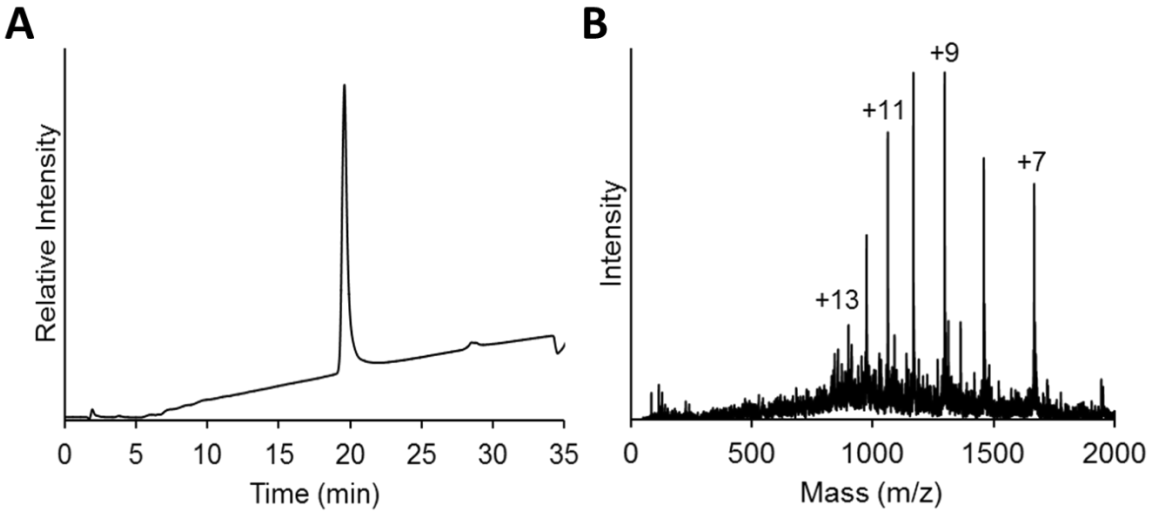


Figure 3.S6. Characterization of purified H4(1-14)K12su(aux)-NHNH$_2$. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS.

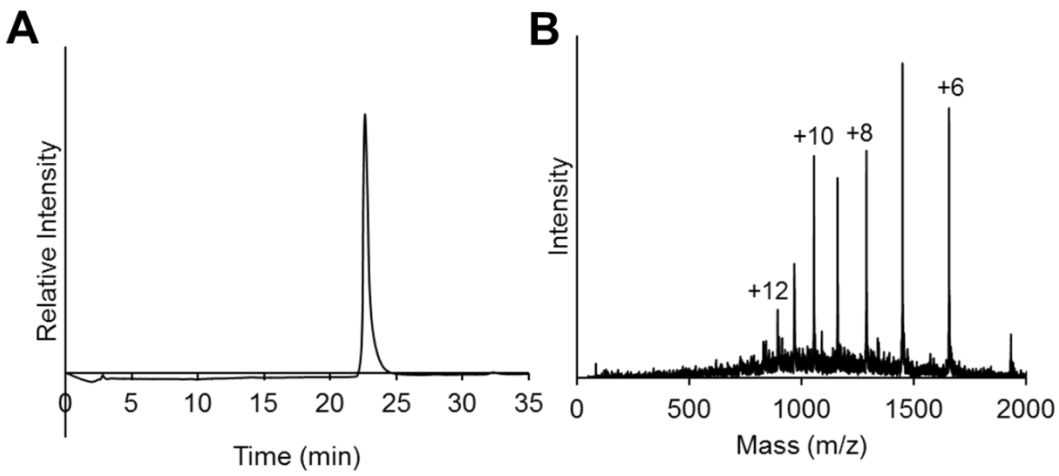


Figure 3.S7. Characterization of purified H4(1-14)K12su-NHNH$_2$. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS.

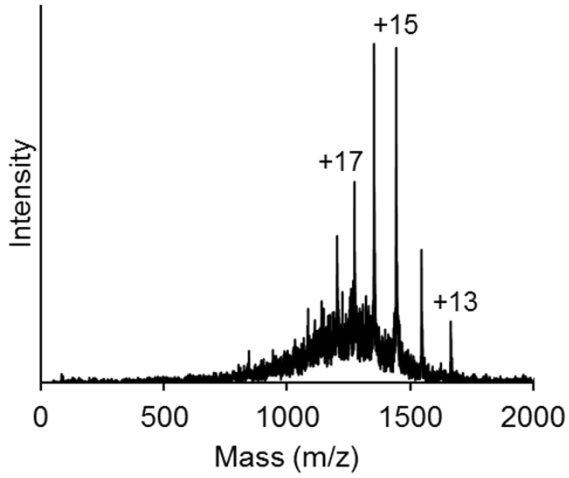


Figure 3.S8. ESI-MS characterization of purified H4(1-102)K12su(A15C).

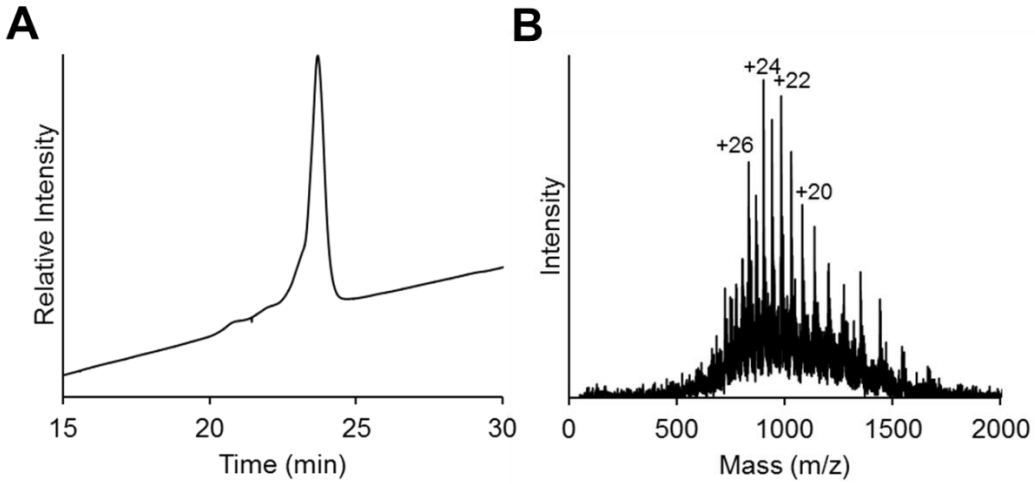


Figure 3.S9. Characterization of purified H4(1-102)K12su. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of H4(1-102)K12su.

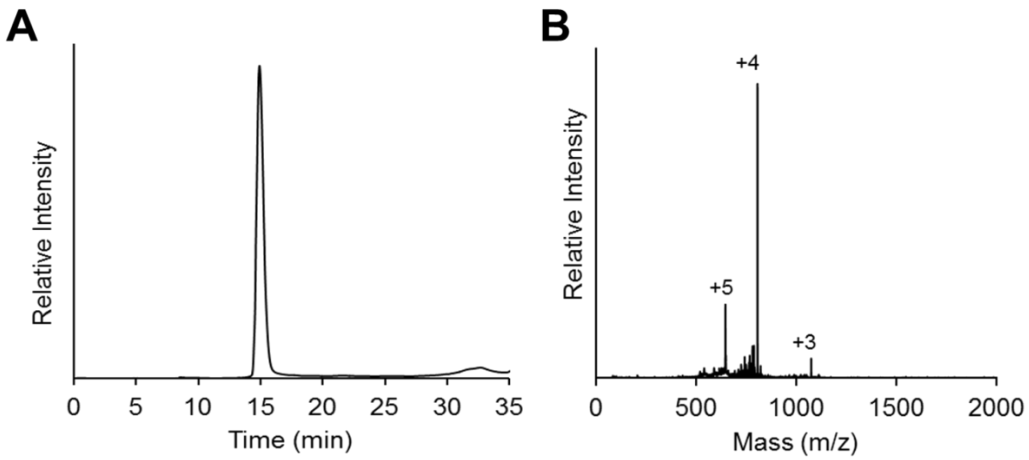


Figure 3.S10. Characterization of purified H3(1-28)polyac-$NHNH_2$. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of H3(1-28)polyac-$NHNH_2$.

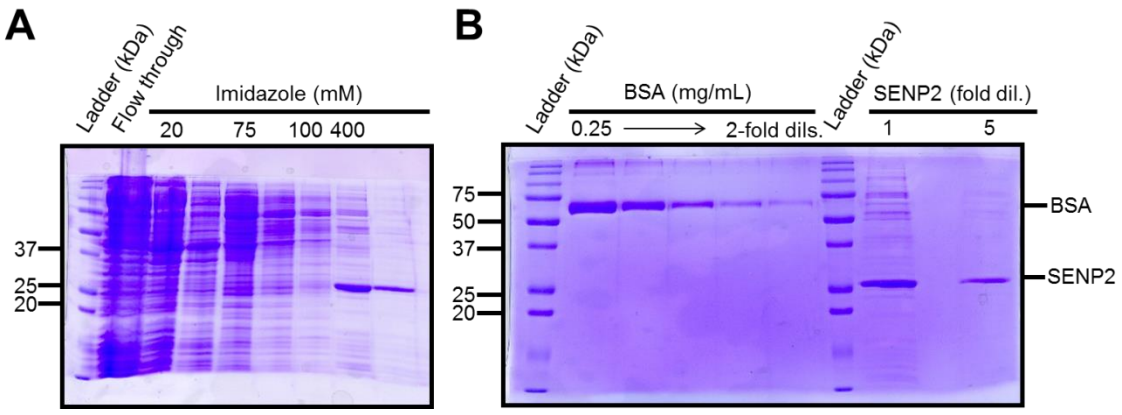


Figure 3.S11. Purification of His_6-SENP2 protease. (A) 15% SDS-PAGE of Ni-NTA IMAC purification of His_6-SENP2 protease. (B) Quantification of dialyzed SENP2 protease protein concentration through comparing densities to BSA standards of known concentration.

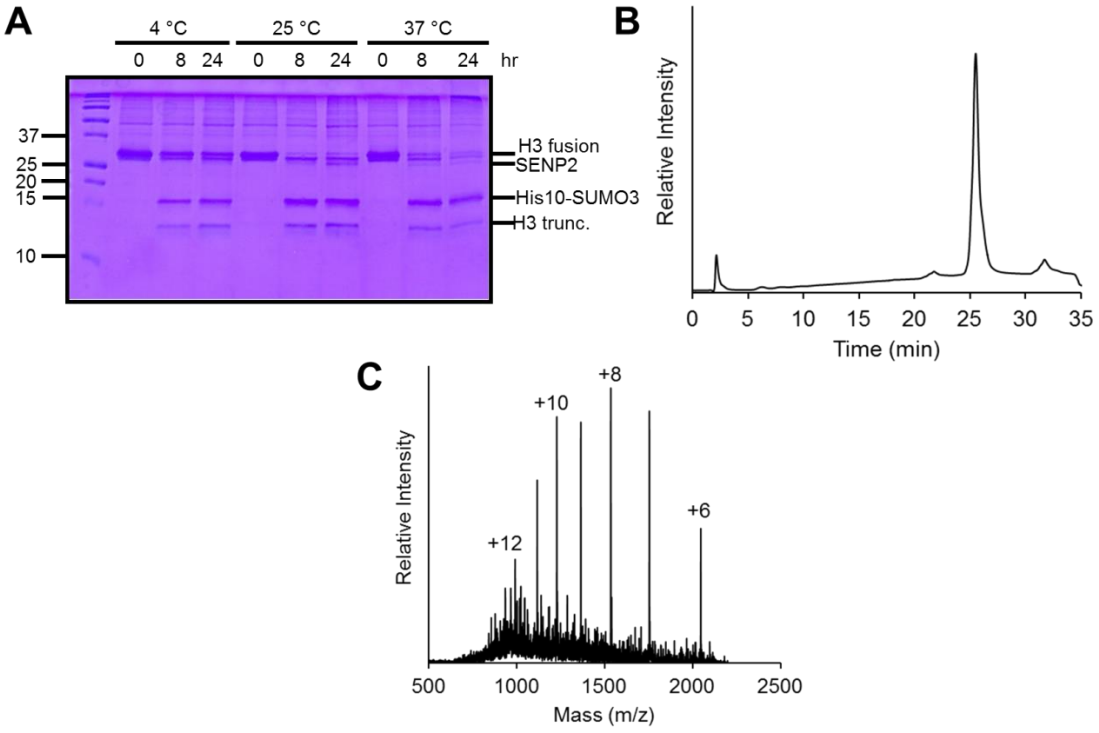
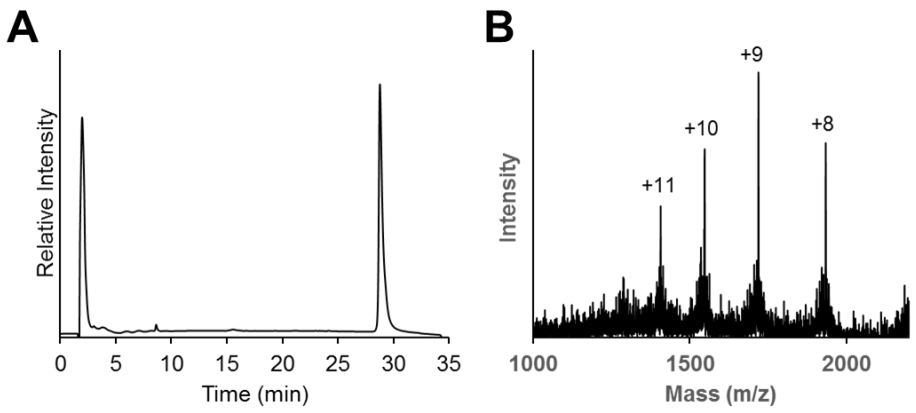


Figure 3.S12. Cleavage of His_{10}-SUMO3-H3(29-135)A29C,C110A. (A) 15% SDS-PAGE of SENP2 cleavage reaction optimization. (B) C18 analytical RP-HPLC of purified H3(29-135)A29C,C110A, 0-73% CH_3CN in H_2O, 30 min. gradient. (C) ESI-MS of H3(29-135)A29C,C110A.



Characterization of purified H3(1-135)polyac(A29C,C110A). (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of H3polyac(A29C, C110A).

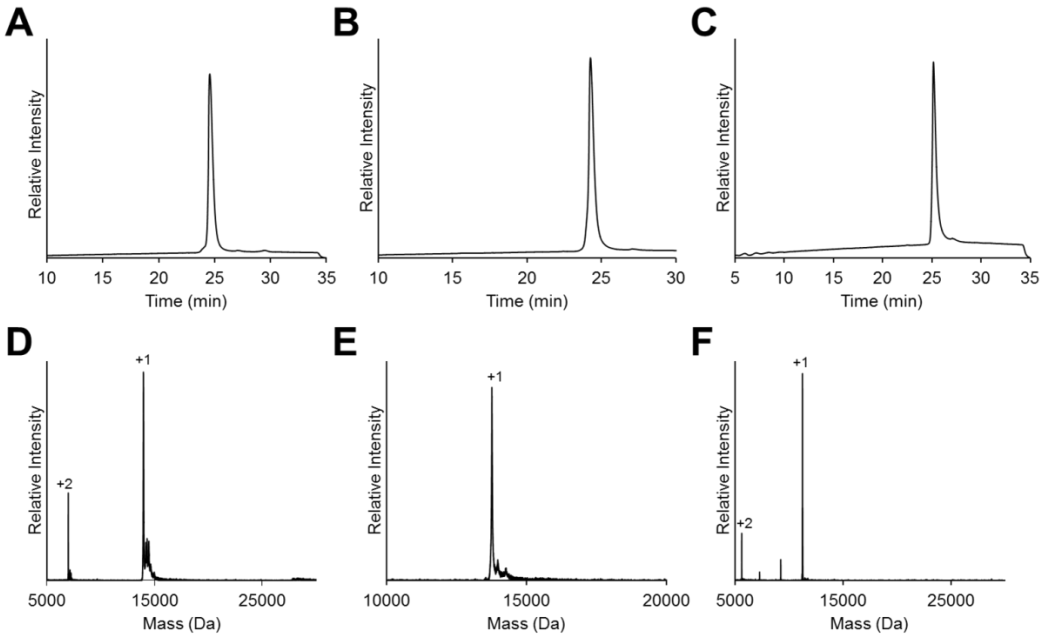


Figure 3.S14. Characterization of purified core histones. (A-C) C18 analytical RP-HPLC of H2A (A), H2B, (B) and H4 (C), 0-73% CH_3CN in H_2O, 30 min. gradient. (D-F) MALDI-TOF MS of purified H2A (D), H2B (E) and H4 (F).

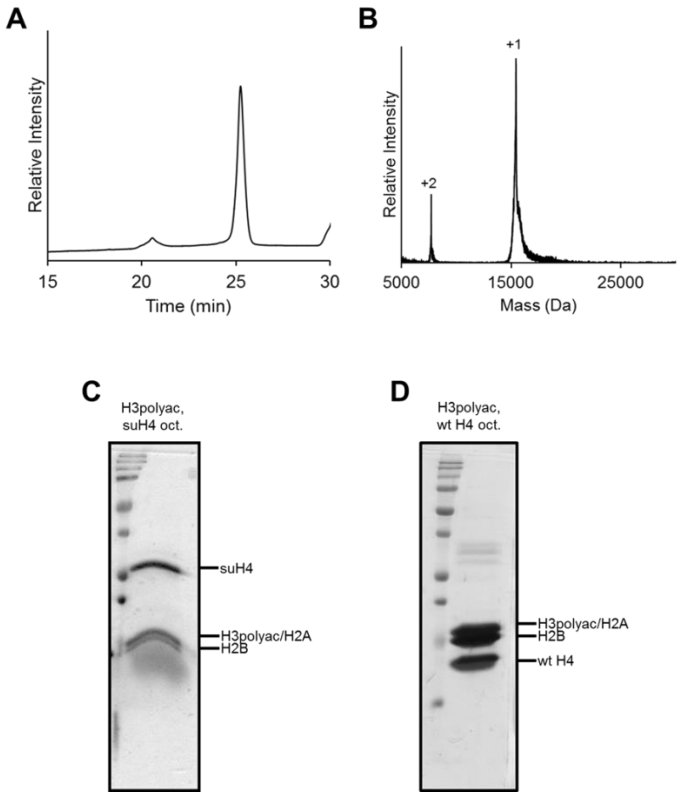


Figure 3.S15. Characterization of purified H3(1-135)polyac(C110A) and H3polyac octamers. (A) C18 analytical RP-HPLC of H3(1-135)polyac(C110A), 0-73% CH_3CN in H_2O, 30 min. gradient. (B) MALDI-TOF MS of purified H3(1-135)polyac(C110A). (C) 15% SDS-PAGE of H3polyac, suH4 mononucleosomes. (D) 15% SDS-PAGE of H3polyac, wild-type H4 mononucleosomes.

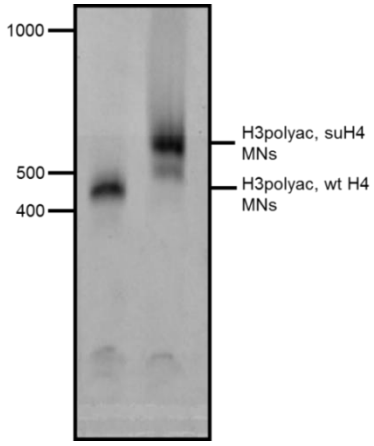


Figure 3.S16. Characterization of H3polyac mononucleosomes. 5% polyacrylamide/TBE gel of H3polyac, suH4 and H3polyac, wild-type H4 mononucleosomes.

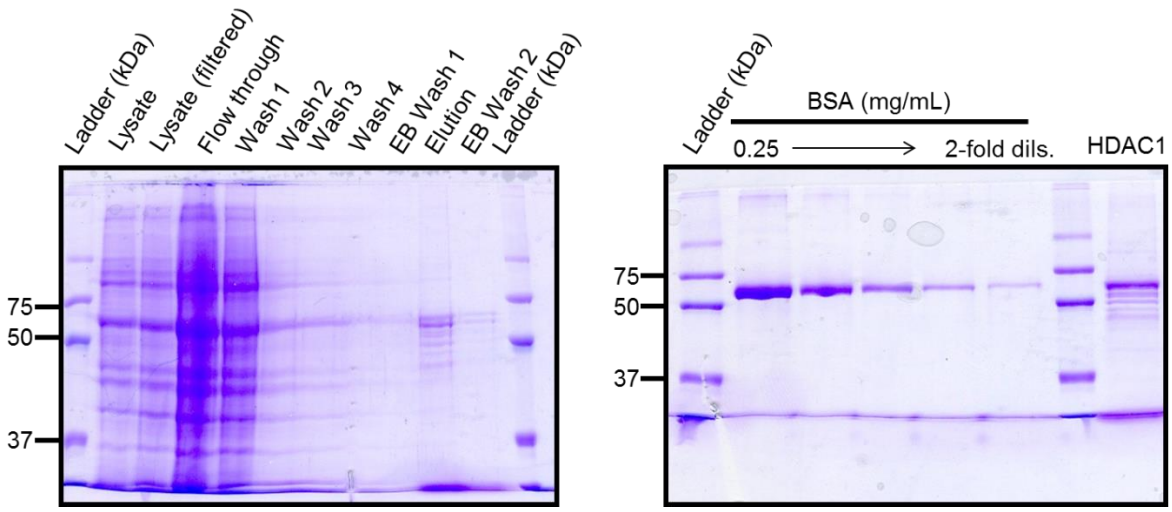


Figure 3.S17. Purification of HDAC1-FLAG. (A) 10% SDS-PAGE of FLAG immunopurification of HDAC1-FLAG. EB = elution buffer. (B) 10% SDS-PAGE quantification of dialyzed HDAC1-FLAG protein concentration through comparing densities against BSA standards of known concentration.

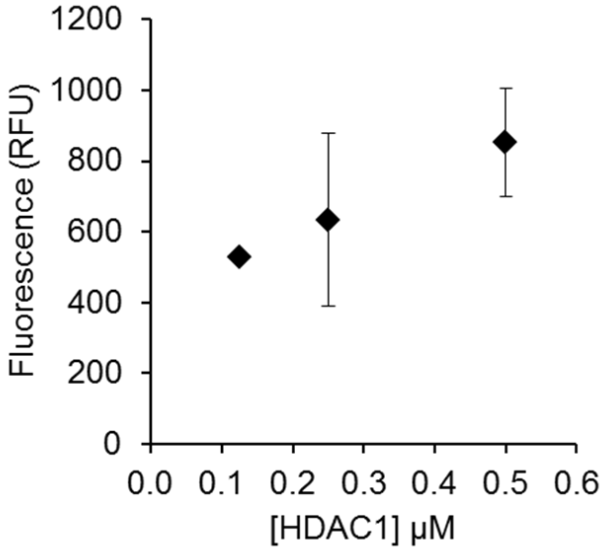


Figure 3.S18. Fleur de Lys assay using HDAC1-FLAG.

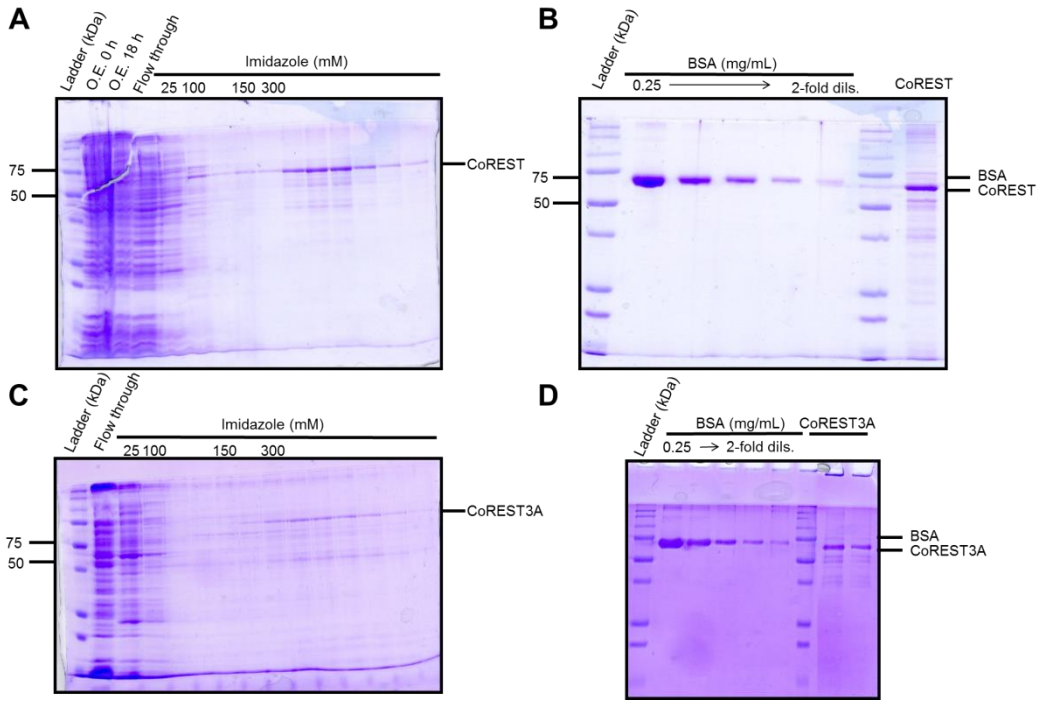


Figure 3.S19. Purification of His_6-CoREST and His_6-CoREST-3A (A) 12% SDS-PAGE of Ni-NTA IMAC purification of His_6-CoREST. (B) Quantification of dialyzed His_6-CoREST protein concentration through comparing densities to BSA standards of known concentration. (C) 12% SDS-PAGE of Ni-NTA IMAC purification of His_6-CoREST-3A. (D) Quantification of dialyzed His_6-CoREST-3A protein concentration through comparing densities to BSA standards of known concentration.

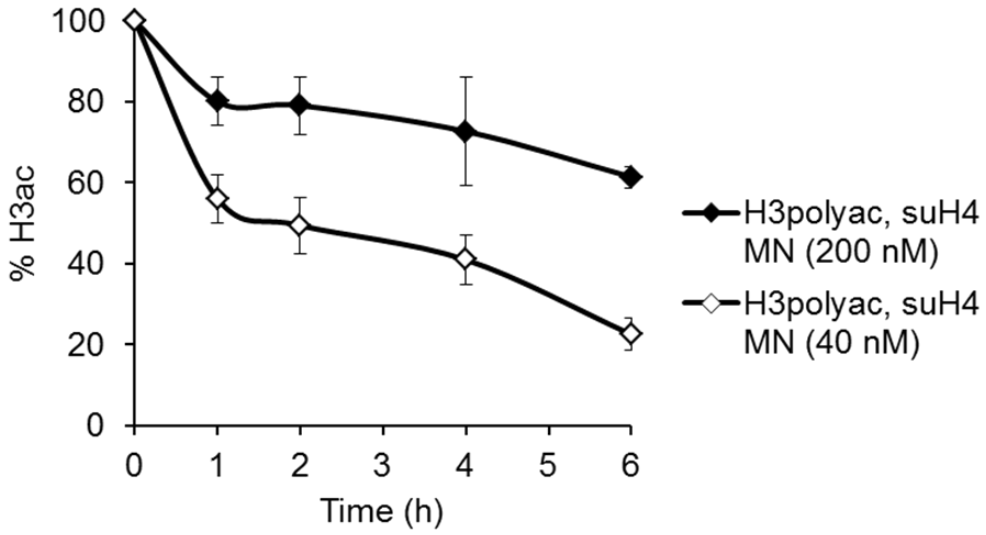


Figure 3.S20. H3polyac, suH4 MN deacetylation assay. Time-course of H3polyac, suH4 MNs (200 nM and 40 nM) deacetylation by HDAC1.

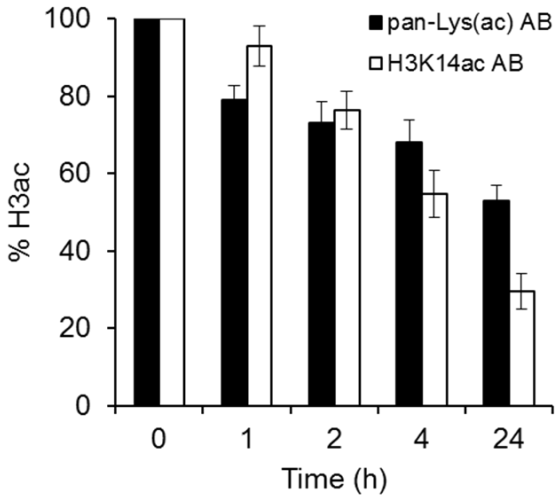


Figure 3.S21. H3K14ac-specific deacetylation in H3polyac MNs. H3polyac, wild-type H4 MN HDAC1 deacetylation assays performed using both α-H3K14ac and α-pan-Lys(ac) antibodies.

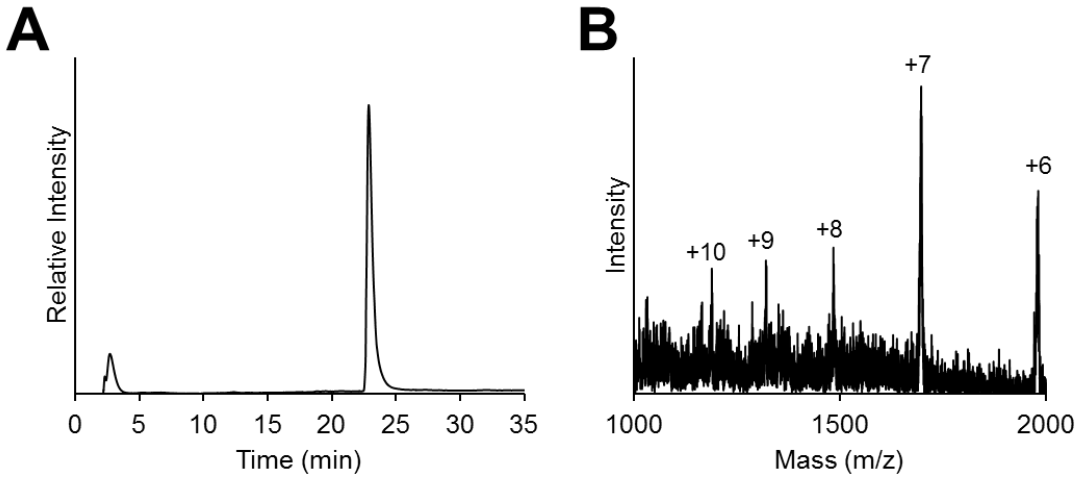


Figure 3.S22. Characterization of $^{2}H/^{13}C/^{15}N$-labeled SUMO3. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of $^{2}H/^{13}C/^{15}N$-labeled SUMO3.

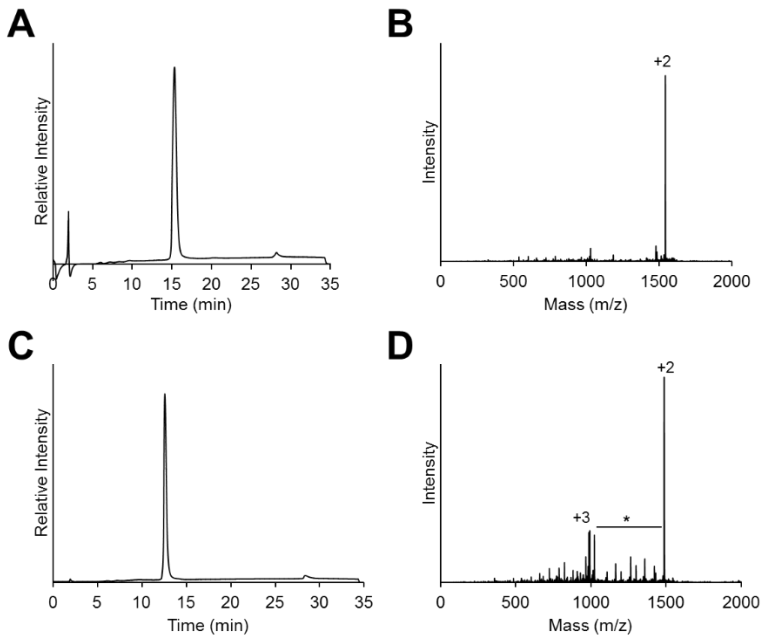


Figure 3.S23. Characterization of purified CoREST1(252-277) and CoREST1(252-277)-3A. (A) CoREST1(252-277) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of CoREST1(252-277). (C) CoREST1(252-277)-3A C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of CoREST1(252-277)-3A. * = on-instrument peptide fragmentation.

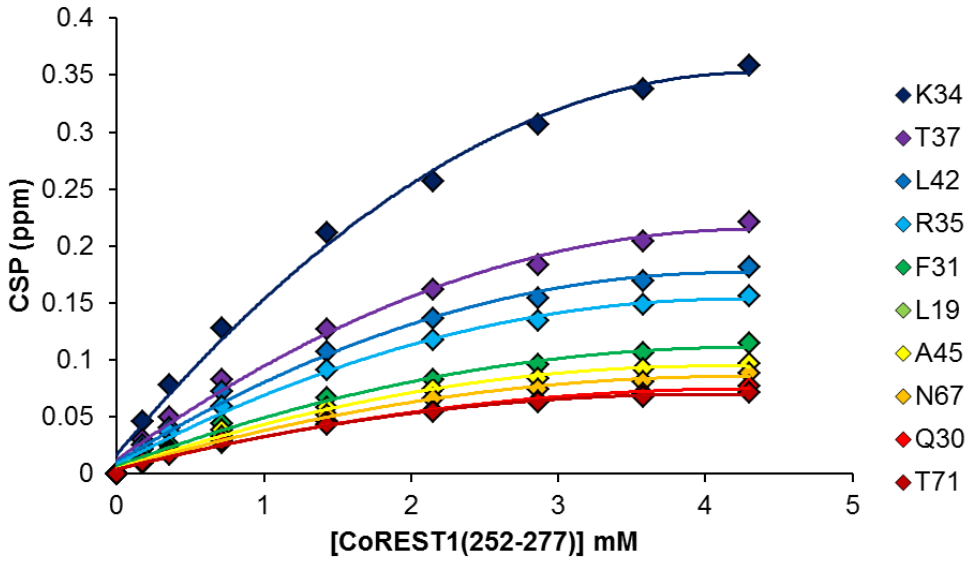


Figure 3.S24. CoREST1(252-277)-SUMO3 binding. Isothermal binding saturation curves for the ten most perturbed SUMO3 residues

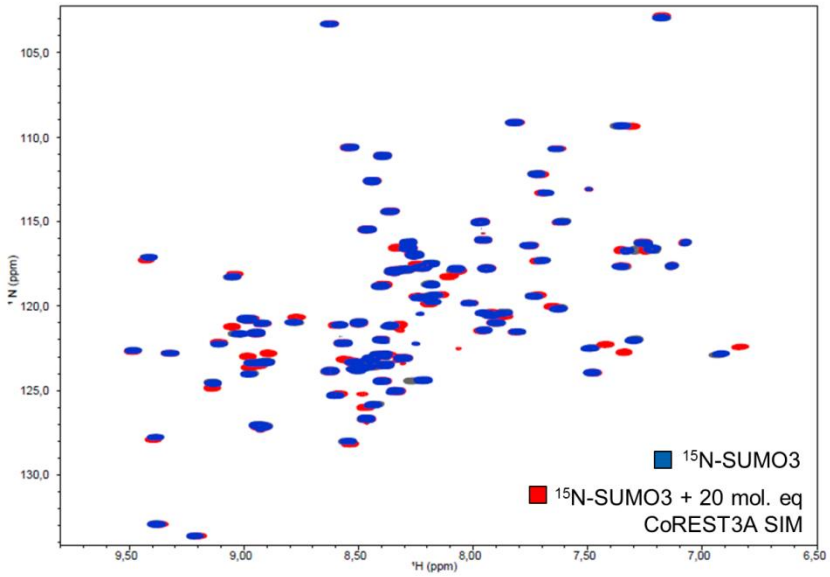


Figure 3.S25. SUMO3 ^{1}H-^{15}N-HSQC NMR spectra with titrated CoREST(252-277)-3A peptide. SUMO residues when 20 mol. eq. CoREST(252-277)-3A was titrated into solution displayed similar CSPs to those observed when 2 mol. eq. of wild-type CoREST1(252-277) was used.

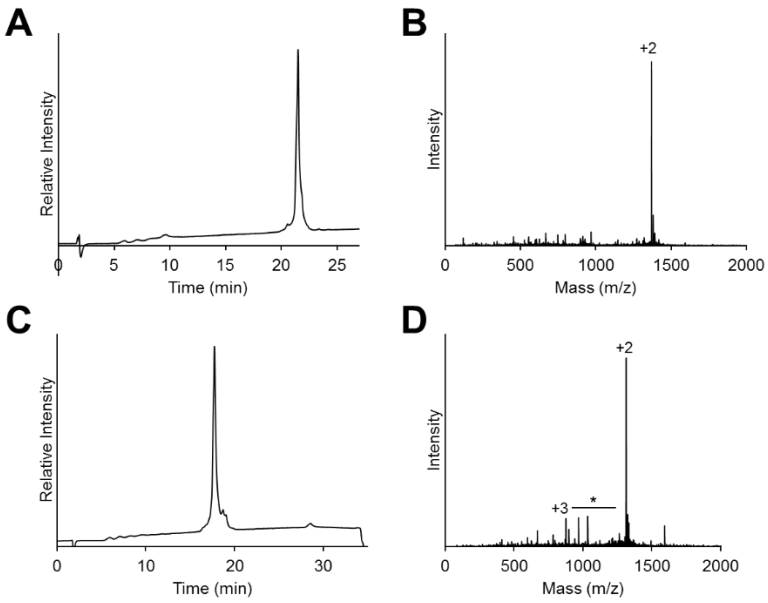


Figure 3.S26. Characterization of purified FIP1L1(120-145) and FIP1L1(120-145)-3A peptides. (A) FIP1L1(120-145) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of FIP1L1(120-145). (C) FIP1L1(120-145)-3A C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of FIP1L1(120-145)-3A. * = on-instrument peptide fragmentation.

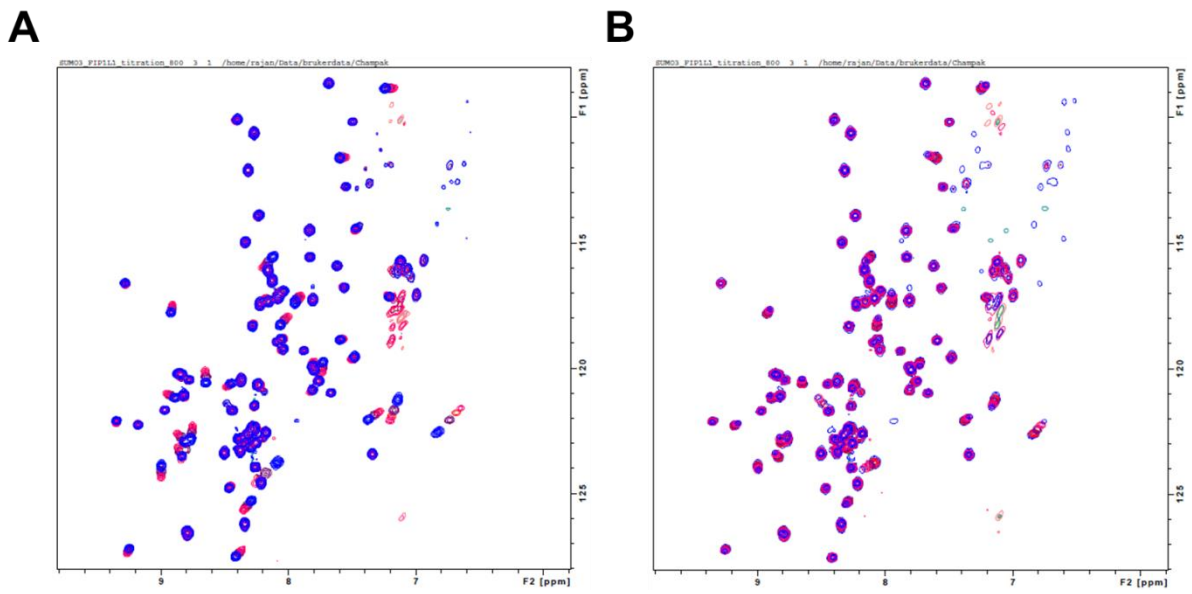


Figure 3.S27. SUMO3 ^{1}H-^{15}N-HSQC NMR spectra with titrated FIP1L1 SIM peptides. (A) Overlaid ^{1}H-^{15}N-HSQC NMR spectra of free SUMO3 and 1-24 molar equivalents of titrated FIP1L1(120-145). (B) Overlaid ^{1}H-^{15}N-HSQC NMR spectra of free SUMO3 and 1-24 molar equivalents of titrated FIP1L11(120-145)-3A. SUMO residues when 24 mol. eq. FIP1L1(120-145-3A was titrated into solution displayed similar CSPs to those observed when 8 mol. eq. of wild-type FIP1L1(120-145) was used.

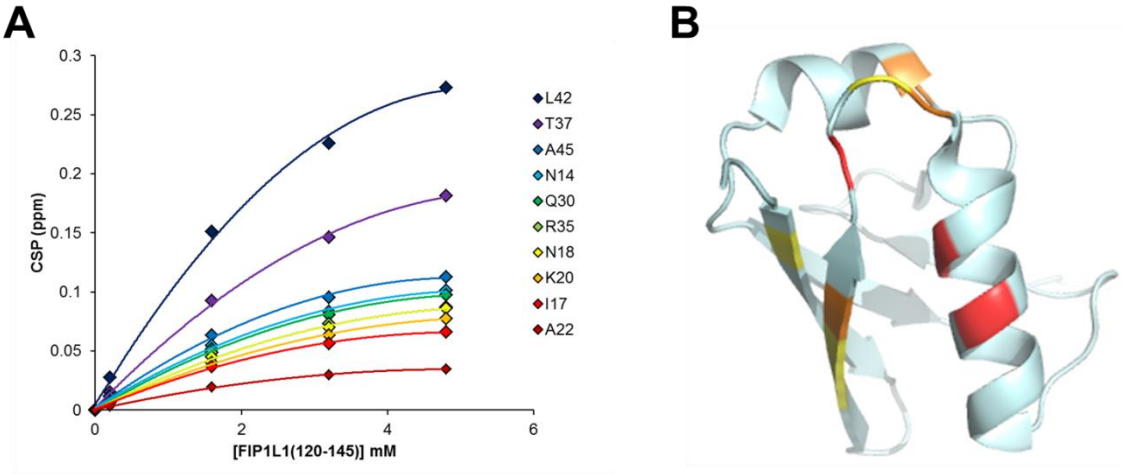


Figure 3.S28. FIP1L1(120-145)-SUMO3 binding. (A) Isothermal binding saturation curves for ten SUMO3 residues that are perturbed upon FIP1L1(120-145) peptide binding (B) Crystal structure of SUMO3 with highlighted residue effected by FIP1L1(120-145) binding. The three residues in red (L42, K34 and A45) indicate residues with the three largest CSPs. The three residues in orange (T71, T37 and F31) indicate the next three largest CSPs. The four residues in yellow (Q30, N18, V29 and H36) indicate the next four largest CSPs. PBD: 1U4A

3.6 References

(1) Grunstein, M. *Nature* **1997**, *389* (6649), 349–352.

(2) Struhl, K. *Genes Dev.* **1998**, *12* (5), 599–606.

(3) Sterner, D. E.; Berger, S. L. *Microbiol. Mol. Biol. Rev.* **2000**, *64* (2), 435–459.

(4) Guan, K.-L. et al. *Science (80-.).* **2010**, *327*, 1000–1007.

(5) Miller, K. M.; Tjeertes, J. V.; Coates, J.; Legube, G.; Polo, S. E.; Britton, S.; Jackson, S. P. *Nat. Struct. Mol. Biol.* **2010**, *17* (9), 1144–1151.

(6) Sharon Y. Roth, John M. Denu, C. D. A. *Annu. Rev. Biochem.* **2001**, *70*, 81–120.

(7) Yang, X. J.; Seto, E. *Nat. Rev. Mol. Cell Biol.* **2008**, *9* (3), 206–218.

(8) Robinson, P. J. J.; An, W.; Routh, A.; Martino, F.; Chapman, L.; Roeder, R. G.; Rhodes, D. *J. Mol. Biol.* **2008**, *381* (4), 816–825.

(9) Devaiah, B. N.; Case-Borden, C.; Gegonne, A.; Hsu, C. H.; Chen, Q.; Meerzaman, D.; Dey, A.; Ozato, K.; Singer, D. S. *Nat. Struct. Mol. Biol.* **2016**, *23* (6), 540–548.

(10) Jacobson, R. H.; Ladurner, A. G.; King, D. S.; Tjian, R. *Science (80-.).* **2000**, *288* (5470), 1422–1425.

(11) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Müller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S. *Cell* **2012**, *149* (1), 214–231.

(12) Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M. M. *Nature* **1999**, *399* (6735), 491–496.

(13) Seligson, D. B.; Horvath, S.; Shi, T.; Yu, H.; Tze, S.; Grunstein, M.; Kurdistani, S. K. *Nature* **2005**, *435* (7046), 1262–1266.

(14) Hockly, E.; Richon, V. M.; Woodman, B.; Smith, D. L.; Zhou, X.; Rosa, E.; Sathasivam, K.; Ghazi-Noori, S.; Mahal, A.; Lowden, P. A. S.; Steffan, J. S.; Marsh, J. L.; Thompson, L. M.; Lewis, C. M.; Marks, P. A.; Bates, G. P. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (4), 2041–2046.

(15) Yang, X. J.; Seto, E. *Oncogene* **2007**, *26* (37), 5310–5318.

(16) Ropero, S.; Fraga, M. F.; Ballestar, E.; Hamelin, R.; Yamamoto, H.; Boix-Chornet, M.; Caballero, R.; Alaminos, M.; Setien, F.; Paz, M. F.; Herranz, M.; Palacios, J.; Arango, D.; Orntoft, T. F.; Aaltonen, L. A.; Schwartz, S.; Esteller, M. *Nat. Genet.* **2006**, *38* (5), 566–569.

(17) Kawaguchi, Y.; Kovacs, J. J.; McLaurin, A.; Vance, J. M.; Ito, A.; Yao, T. P. *Cell* **2003**, *115* (6), 727–738.

(18) Pandey, U. B.; Nie, Z.; Batlevi, Y.; McCray, B. A.; Ritson, G. P.; Nedelsky, N. B.; Schwartz, S. L.; Diprospero, N. A.; Knight, M. A.; Schuldiner, O.; Padmanabhan, R.; Hild, M.; Berry, D. L.; Garza, D.; Hubbert, C. C.; Yao, T. P.; Baehrecke, E. H.; Taylor, J. P.

Nature **2007**, *447* (7146), 859–863.

(19) Ouyang, H.; Ali, Y. O.; Ravichandran, M.; Dong, A.; Qiu, W.; MacKenzie, F.; Dhe-Paganon, S.; Arrowsmith, C. H.; Zhai, R. G. *J. Biol. Chem.* **2012**, *287* (4), 2317–2327.

(20) West, R. W. J. A. C. *J. Clin. Investig.* **2014**, *124*, 30–39.

(21) Bürli, R. W.; Luckhurst, C. A.; Aziz, O.; Matthews, K. L.; Yates, D.; Lyons, K. A.; Beconi, M.; McAllister, G.; Breccia, P.; Stott, A. J.; Penrose, S. D.; Wall, M.; Lamers, M.; Leonard, P.; Müller, I.; Richardson, C. M.; Jarvis, R.; Stones, L.; Hughes, S.; Wishart, G.; Haughan, A. F.; O'Connell, C.; Mead, T.; McNeil, H.; Vann, J.; Mangette, J.; Maillard, M.; Beaumont, V.; Munoz-Sanjuan, I.; Dominguez, C. *J. Med. Chem.* **2013**, *56* (24), 9934–9954.

(22) Weerasinghe, S. V. W.; Estiu, G.; Wiest, O.; Pflum, M. K. H. *J. Med. Chem.* **2008**, *51* (18), 5542–5551.

(23) Smith, B. C.; Denu, J. M. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2009**, *1789* (1), 45–57.

(24) Gantt, S. M. L.; Decroos, C.; Lee, M. S.; Gullett, L. E.; Bowman, C. M.; Christianson, D. W.; Fierke, C. A. *Biochemistry* **2016**, *55* (5), 820–832.

(25) Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. *Nature* **1999**, *401* (6749), 188–193.

(26) Vannini, A.; Volpari, C.; Filocamo, G.; Casavola, E. C.; Brunetti, M.; Renzoni, D.; Chakravarty, P.; Paolini, C.; De Francesco, R.; Gallinari, P.; Steinkühler, C.; Di Marco, S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (42), 15064–15069.

(27) Watson, P. J.; Fairall, L.; Santos, G. M.; Schwabe, J. W. R. *Nature* **2012**, *481* (7381), 335–340.

(28) Millard, C. J.; Watson, P. J.; Celardo, I.; Gordiyenko, Y.; Cowley, S. M.; Robinson, C. V.; Fairall, L.; Schwabe, J. W. R. *Mol. Cell* **2013**, *51* (1), 57–67.

(29) Gao, L.; Cueto, M. A.; Asselbergs, F.; Atadja, P. *J. Biol. Chem.* **2002**, *277* (28), 25748–25755.

(30) Shiio, Y.; Eisenman, R. N. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230.

(31) 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.; Johnson, E. S.; Berger, S. L. *Genes Dev.* **2006**, *20* (8), 966–976.

(32) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837.

(33) You, a; Tong, J. K.; Grozinger, C. M.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 1454–1458.

(34) Humphrey, G. W.; Wang, Y.; Russanova, V. R.; Hirai, T.; Qin, J.; Nakatani, Y.; Howard, B. H. *J. Biol. Chem.* **2001**, *276* (9), 6817–6824.

(35) Xue, Y.; Wong, J.; Moreno, G. T.; Young, M. K.; Côté, J.; Wang, W. *Mol. Cell* **1998**, *2* (6), 851–861.

(36) Zhang, Y.; Ng, H. H.; Erdjument-Bromage, H.; Tempst, P.; Bird, A.; Reinberg, D. *Genes Dev.* **1999**, *13* (15), 1924–1935.

(37) Wu, M.; Hayward, D.; Kalin, J. H.; Song, Y.; Schwabe, J. W. R.; Cole, P. A. *Elife* **2018**, *7*, 1–19.

(38) Ouyang, J.; Shi, Y.; Valin, A.; Xuan, Y.; Gill, G. *Mol. Cell* **2009**, *34* (2), 145–154.

(39) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. *Cell* **2004**, *119* (7), 941–953.

(40) Lee, M. G.; Wynder, C.; Cooch, N.; Shiekhattar, R. *Nature* **2005**, *437* (7057), 432–435.

(41) Yang, S. H.; Sharrocks, A. D. *Mol. Cell* **2004**, *13* (4), 611–617.

(42) Dhall, A.; Weller, C. E.; Chu, A.; Shelton, P. M. M.; Chatterjee, C. *ACS Chem. Biol.* **2017**, *12* (9), 2275–2280.

(43) Weller, C. E.; Huang, W.; Chatterjee, C. *Chembiochem* **2014**, *15* (9), 1263–1267.

(44) William G. Dougherty, Stephen A. Johnston, Eric D. Howard, Kerstin K. Leuther, T. D. P. *Anal. Biochem.* **1994**, *216*, 413–417.

(45) Stevens, A. J.; Brown, Z. Z.; Shah, N. H.; Sekar, G.; Cowburn, D.; Muir, T. W. *J. Am. Chem. Soc.* **2016**, *138* (7), 2162–2165.

(46) Shah, N. H.; Dann, G. P.; Vila-Perelló, M.; Liu, Z.; Muir, T. W. *J. Am. Chem. Soc.* **2012**, *134* (28), 11338–11341.

(47) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. *Nat. Protoc.* **2013**, *8* (12), 2483–2495.

(48) Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. Bin; Lin, Y.; Cui, H. K.; Liu, L. *Angew. Chemie - Int. Ed.* **2011**, *50* (33), 7645–7649.

(49) Wan, Q.; Danishefsky, S. J. *Angew. Chemie - Int. Ed.* **2007**, *46* (48), 9248–9252.

(50) Ogryzko, V. V.; Schiltz, R. L.; Russanova, V.; Howard, B. H.; Nakatani, Y. *Cell* **1996**, *87* (5), 953–959.

(51) Schiltz, R. L.; Mizzen, C. A.; Vassilev, A.; Cook, R. G.; Allis, C. D.; Nakatani, Y. *J. Biol. Chem.* **1999**, *274* (3), 1189–1192.

(52) Lowary, P. T.; Widom, J. *J. Mol. Biol.* **1998**, *276*, 19–42.

(53) Luger, K.; Rechsteiner, T. J.; Flaus, A. J.; Waye, M. M.; Richmond, T. J. *J. Mol. Biol.* **1997**, *272* (3), 301–311.

(54) Hecker, C. M.; Rabiller, M.; Haglund, K.; Bayer, P.; Dikic, I. *J. Biol. Chem.* **2006**, *281* (23), 16117–16127.

(55) Song, J.; Zhang, Z.; Hu, W.; Chen, Y. *J. Biol. Chem.* **2005**, *280* (48), 40122–40129.

(56) Song, J.; Durrin, L. K.; Wilkinson, T. a; Krontiris, T. G.; Chen, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (40), 14373–14378.

(57) Rodriguez, M. S.; Dargemont, C.; Hay, R. T. *J. Biol. Chem.* **2001**, *276* (16), 12654–12659.

(58) Guo, D.; Li, M.; Zhang, Y.; Yang, P.; Eckenrode, S.; Hopkins, D.; Zheng, W.; Purohit, S.; Podolsky, R. H.; Muir, A.; Wang, J.; Dong, Z.; Brusko, T.; Atkinson, M.; Pozzilli, P.; Zeidler, A.; Raffel, L. J.; Jacob, C. O.; Park, Y.; Serrano-Rios, M.; Martinez Larrad, M. T.; Zhang, Z.; Garchon, H. J.; Bach, J. F.; Rotter, J. I.; She, J. X.; Wang, C. Y. *Nat. Genet.* **2004**, *36* (8), 837–841.

(59) Bohren, K. M.; Nadkarni, V.; Song, J. H.; Gabbay, K. H.; Owerbach, D. *J. Biol. Chem.* **2004**, *279* (26), 27233–27238.

(60) Cappadocia, L.; Pichler, A.; Lima, C. D. *Nat. Struct. Mol. Biol.* **2015**, *22* (12), 968–975.

(61) Kaur, K.; Park, H.; Pandey, N.; Azuma, Y.; De Guzman, R. N. *J. Biol. Chem.* **2017**, *292* (24), 10230–10238.

(62) Taupitz, K. F.; Dörner, W.; Mootz, H. D. *Chem. - A Eur. J.* **2017**, *23* (25), 5978–5982.

(63) Chang, C. C.; Naik, M. T.; Huang, Y. S.; Jeng, J. C.; Liao, P. H.; Kuo, H. Y.; Ho, C. C.; Hsieh, Y. L.; Lin, C. H.; Huang, N. J.; Naik, N. M.; Kung, C. C. H.; Lin, S. Y.; Chen, R. H.; Chang, K. S.; Huang, T. H.; Shih, H. M. *Mol. Cell* **2011**, *42* (1), 62–74.

(64) Namanja, A. T.; Li, Y. J.; Su, Y.; Wong, S.; Lu, J.; Colson, L. T.; Wu, C.; Li, S. S. C.; Chen, Y. *J. Biol. Chem.* **2012**, *287* (5), 3231–3240.

(65) 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.; Chatterjee, C. *Nat. Commun.* **2016**, *7*, 1–10.

(66) Bossis, G.; Melchior, F. *Mol. Cell* **2006**, *21* (3), 349–357.

(67) Qiu, Y.; Zhao, Y.; Becker, M.; John, S.; Parekh, B. S.; Huang, S.; Hendarwanto, A.; Martinez, E. D.; Chen, Y.; Lu, H.; Adkins, N. L.; Stavreva, D. A.; Wiench, M.; Georgel, P. T.; Schiltz, R. L.; Hager, G. L. *Mol. Cell* **2006**, *22* (5), 669–679.

(68) Cheng, J.; Wang, D.; Wang, Z.; Yeh, E. T. H. *Mol. Cell. Biol.* **2004**, *24* (13), 6021–6028.

(69) Pervushin, K. V.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **1998**, *12* (2), 345–348.

Chapter 4

Biochemical investigation of the role of histone H4 sumoylation and H3 acetylation on H3 demethylation by the LSD1-CoREST1 sub-complex

4.1 Introduction

Histone methylation is an abundant post-translational modification (PTM) with critical roles in gene transcription, cell cycle progression, differentiation and DNA damage response pathways.[1–4] Lysine and arginine residues within histones are the most prominent targets of methylation, although recent reports suggest H2AQ105 can also be methylated on the amide nitrogen while histidine residues may be methylated in rare cases.[5–7] Arginine residues are either symmetrically or asymmetrically dimethylated on either terminal amine within H2A (R3), H3 (R2, R8, R17, R26, R42) and H4 (R3).[8] Lysines within H3 (K4, K9, K23, K27, K36, K79) and the H4 (K20) tail are most prominently modified and can be mono-, di- or trimethylated on their ε-amine.[9] To date, no accounts have uncovered a biological role for lysine methylation in H2B, although chromatin immunoprecipitation and sequencing (ChIP-seq) studies have suggested potential monomethylation at H2BK5 within promoter regions.[10] Additionally, the DNA damage response variant of H2A (H2A.X) is reportedly methylated at K134.[11] The vast diversity that is possible with regards to the site and degree of histone methylation imparts exquisite epigenetic regulation for chromatin-templated processes that dictate many downstream cellular events. Methyllysine-binding protein domains [e.g. plant homeo domain (PHD), tudor and chromo domains] have been characterized in numerous chromatin-modifying and chromatin-associated proteins, adding further complexity to the interpretation of the histone methylome.[12]

When considering its roles in gene transcription, histone lysine methylation is associated with both gene activation and repression.[1] The effect and regulation of many lysine methylation events, however, is poorly understood and even more so when considered as part of a complex biochemical environment such as chromatin. Di- and trimethylation of H3K4 (H3K4me2/3) is an example of an established gene activating PTM that is significantly upregulated at transcription start sites of transcribed genes in ChIP studies, which either blocks or recruits various effector proteins for transcription initiation (**Figure 4.1**).[13–17] Conversely, di- and trimethylation at H3K9 and H3K27 have been consistently associated with gene down regulation.[18–20] Heterochromatin protein 1α (HP1α) chromo domain-mediated recognition of H3K9me3 stimulates chromatin condensation through dimerization with adjacent HP1α-bound nucleosomes.[19,21] Interestingly, monomethylation of H3K9, H3K27 and H4K20 are associated with actively transcribed genes, while di- and trimethylation of the same residues associate with repression, followed by heterochromatin formation (**Figure 4.1**).[10] Clearly, cells can create elaborate and dynamic epigenetic landscapes to introduce structural and chemical variation into chromatin. Therefore, delineating the code embedded in histone methylation is key to understanding many diverse cellular processes, and in particular understanding the regulation of gene transcription.

Histone methylation is controlled by the opposing activities of histone methyltransferase (HMT) and histone demethylase (HDM) enzymes. The first human HMT (SUV39H1) was described in 2000 by Jenuwein et al, and shown to specifically methylate H3K9.[22] Since then, at least 37 human genes encoding for proteins with HMT activity have been identified.[2] Lysine methyltransferases (KMTs) are grouped into two classes based on their active site geometry: MLL/SET [su(var)3-9, enhancer-of-zeste and trithorax] domain containing enzymes, and the disruptor of telomeric silencing 1 (DOT1) family of KMTs.[23] The active sites of both families contain the methyl group donor, *S*-adenosyl-(L)-methionine (SAM), and share overall similar

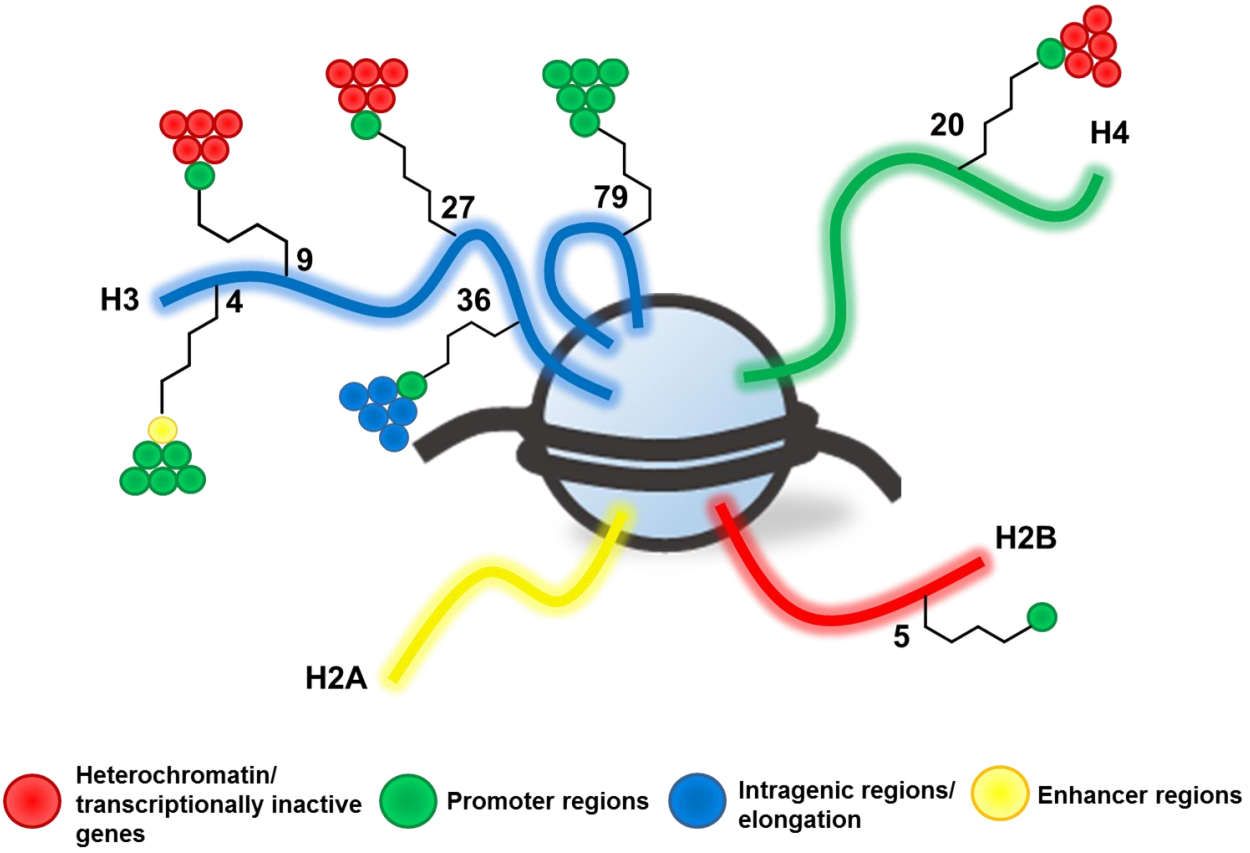


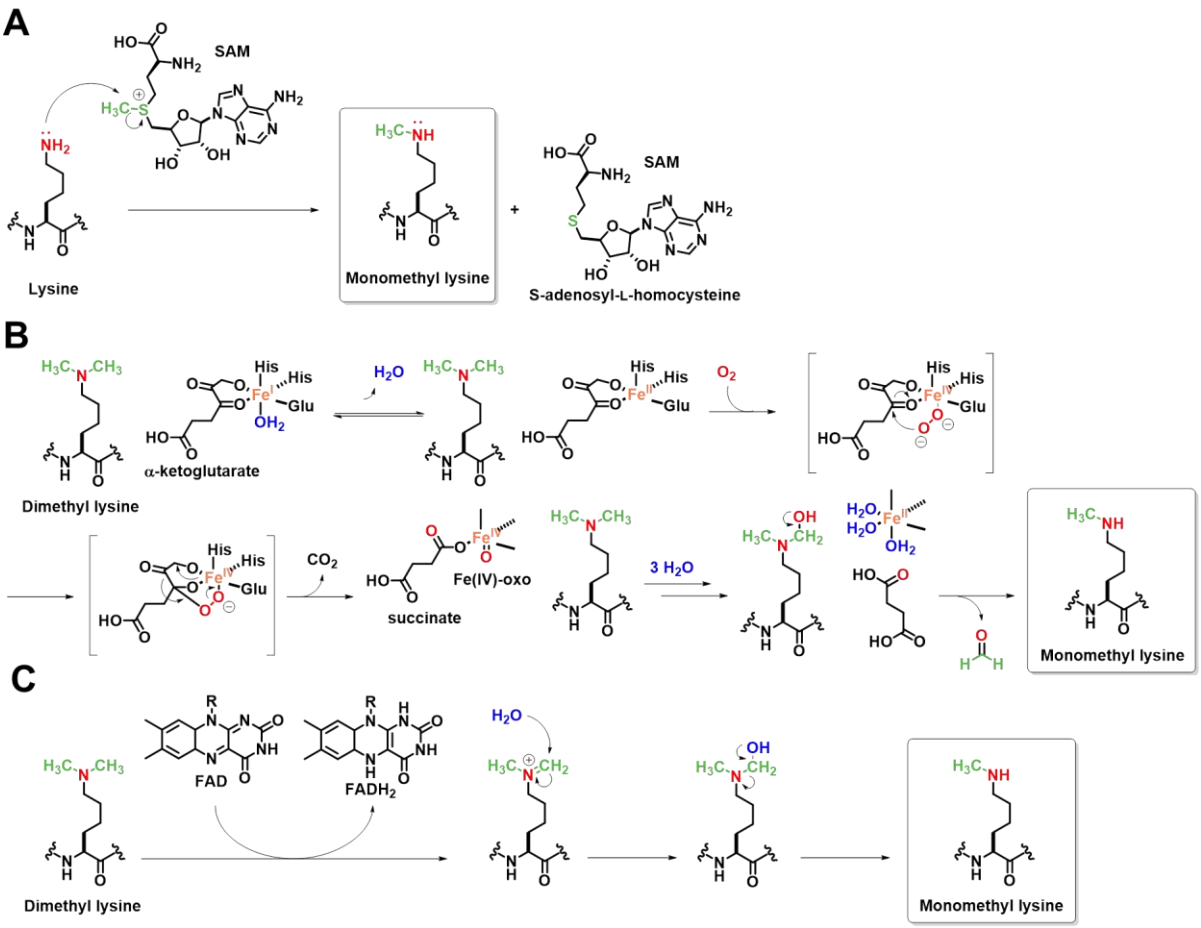
Figure 4.1. Histone lysine methylation. Cartoon representation of known histone mono-, di- and trimethylation sites and their genomic distributions.

mechanisms of methyl transfer (**Scheme 4.1A**). Due to the drastically different active site pH within SET and DOT1 KMTs, it is speculated that they differ their methylation mechanism in the mode of final methyllysine deprotonation and substrate release.[23]

For years, histone methylation was thought to be an irreversible modification. Then in 2004, Yang Shi and coworkers demonstrated the histone demethylase activity of a previously unannotated gene product (KIAA0601) with homology to flavin-containing amine oxidases, which they named lysine specific demethylase 1 (LSD1/KDM1A).[24] To date, at least 20 bona fide lysine demethylases (KDMs) have been reported and are now broadly grouped into two classes.[25–27] The majority of histone demethylases belong to the Jumonji domain (JmjC) family

that house a catalytic Fe^{2+} cofactor, and are divided further into sub-classes (KDM2-7) based on additional protein-protein interaction domains. In tandem with active site α-ketoglutarate and molecular oxygen, JmjC demethylases oxidize their methylated substrates to an unstable carbinolamine intermediate through a highly reactive Fe(IV) species (**Scheme 4.1B**).[23] Demethylation is achieved through a terminal hydrolysis step, releasing formaldehyde as a by-product. The flavin adenine dinucleotide (FAD)-dependent demethylase family consists of only two members, LSD1 and LSD2 (KDM1B). Unlike JmjC demethylases, FAD-dependent enzymes are chemically limited to processing mono- and dimethylated lysine substrates. Mechanistically similar to homologous small molecule monoamine oxidases, the active site FAD cofactor of LSD1/2 accepts a hydride from methylated lysine residue to produce demethylated lysine and formaldehyde (**Scheme 4.1C**).[23] Re-oxidation of the inactive $FADH_2$ by molecular oxygen to regenerate FAD allows further rounds catalysis and produces hydrogen peroxide as a result.

LSD1 is a particularly interesting enzyme due to its high specificity, associations with multiple protein complexes and strong implications in human health and disease. Initial characterization of LSD1 demonstrated exquisite specificity for H3K4me1/2, and a dependence on the interaction with co-repressor for element 1 silencing transcription factor (CoREST1) for the demethylation of nucleosomal substrates.[28–30] The first in vitro characterization of LSD1 activity on methylated peptides by Mattevi et al demonstrated that LSD1 is a relatively slow enzyme with a substrate turnover (k_{cat}) of 3.2 min^{-1} and Michaelis constant (K_M) of 3.0 μM, resulting in a catalytic efficiency (k_{cat}/K_M) of just 1.1 $\mu M^{-1}min^{-1}$ for a monomethylated substrate.[31,32] The authors also showed that acetylation of K9 resulted in a ~6-fold increase in the K_M value (17.5 μM), while S10 phosphorylation completely abolished LSD1 activity. This demonstrates the susceptibility of LSD1 to the specific chromatin PTM landscape. LSD1 consists of an unstructured N-terminal region proposed to contain a putative nuclear localization signal,



Scheme 4.1. Proposed mechanisms for methylation and demethylation of lysine residues. (A) *S*-adenosylmethionine mediated methylation of free lysine by histone methyltransferases. (B) Fe and α-ketoglutarate mediated demethylation of dimethyl lysine by JmjC-containing demethylases. (C) FAD-dependent demethylation of dimethyl lysine by LSD1/2 enzymes.

followed by a SWIRM domain.[28,33] Despite sequence and structural similarities between the SWIRM domains of LSD1 and the DNA binding protein, Swi3, NMR studies with short DNA oligonucleotides failed to detect binding with an isolated LSD1 SWIRM domain construct.[28,33,34] The C-terminal amine oxidase domain (AOD), which binds the catalytic FAD cofactor, is split by two antiparallel helices that form the tower domain responsible for binding CoREST1.[28,33] LSD1 is known to require the first 21 amino acids of the H3 N-terminus for efficient demethylation in vitro, with only the first 12 residues of the H3 tail binding within the active site according to

computationally modelled substrate-enzyme binding interactions.[35] Since its discovery, much has been uncovered regarding LSD1 structure and substrate specificity, and we now know that the large active site cavity supports demethylation of non-histone proteins such as the transcription factor p53 and DNA methyltransferase 1 (DNMT1).[36,37]

LSD1 participates in several protein complexes through hydrophobic and electrostatic interactions within the tower and SWIRM domains. LSD1 has been associated with repressive nucleosome remodeling and deacetylase (NuRD) and LCH (Chapter 3.1) complexes, aiding in their gene repressive activities by H3K4me1/2 demethylation.[38,39] LSD1 not only associates with other proteins, but binding of the LSD1-CoREST1-REST complex to long non-coding (lnc) RNA, HOTAIR, was shown to coordinate chromatin targeting for LSD1 activity.[40] Further studies showed an unexpected switch in LSD1 specificity toward the transcriptionally repressive H3K9me1/2 when it associates with the androgen receptor transcription factor, thereby implicating LSD1 in gene activation as well as repression activities.[41] It is no wonder that disruption of the many roles of LSD1 through genetic mutations of the enzyme are associated with a host of health-related issues ranging from developmental delays to cancers of the breast, prostate and lungs.[42–45]

The Chatterjee lab has been interested in the role of LSD1 as a corepressor due to its proposed connection to the small ubiquitin-like modifier (SUMO) family of proteins, which have also been associated with gene repression (Chapter 1.4). In 2009, Grace Gill and coworkers demonstrated down regulation of LSD1 target genes based on the interactions between the LCH complex and a sumoylated nuclear protein.[46] The dependence of this repression on the interaction between CoREST1 and SUMO2/3 prompted the Chatterjee lab to ascertain the effect of sumoylated H4 (suH4) on activity of the LSD1-CoREST1 (LC) sub-complex toward H3K4me2 in the context of chromatin. To this end, we generated mononucleosomes containing semisynthetic suH4 and

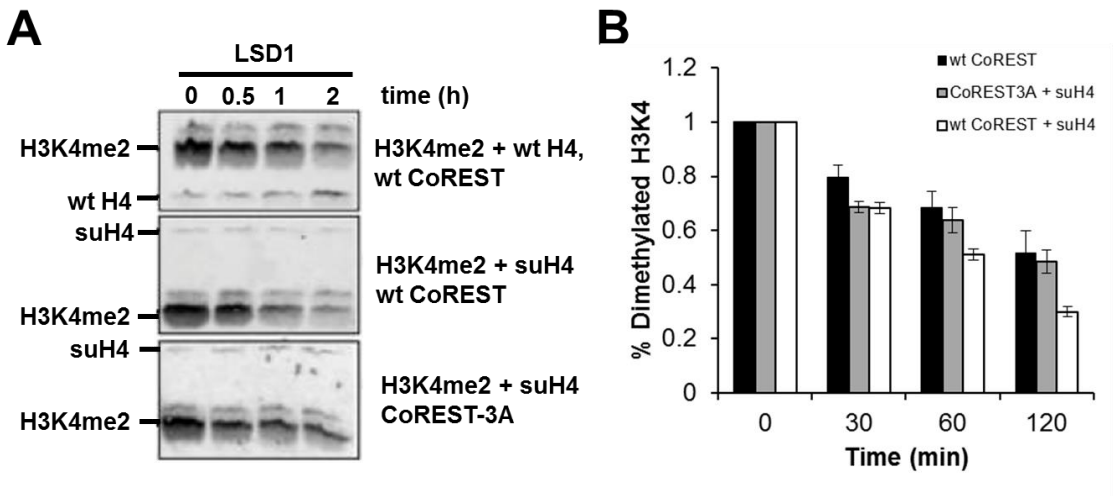


Figure 4.2. suH4 stimulates LC demethylation activity toward H3K4me2 mononucleosomes. (A) Representative Western blot analyses of LC sub-complex demethylation time-course of H3K4me2 mononucleosomes using varying H4 and CoREST1 constructs. α-H3K4me2 and α-H4 antibodies used. (B) Histogram quantification of H3K4me2 demethylation time courses. * $P < 0.05$.

H3K4me2, and incubated them with recombinant LC sub-complex (**Figure 4.2**). Quantitative Western blot analyses of the demethylation assay indicated a ~2-fold stimulation of LC activity when compared to nucleosomes containing wild-type H4.[47] Furthermore, this stimulation was

H3K4me2, and incubated them with recombinant LC sub-complex (**Figure 4.2**). Quantitative Western blot analyses of the demethylation assay indicated a ~2-fold stimulation of LC activity when compared to nucleosomes containing wild-type H4.[47] Furthermore, this stimulation was dependent on the intact SUMO interacting motif (SIM) of CoREST1. Upon demonstrating the effect of suH4 on LC activity, we have been further interested in elucidating the precise mechanistic details of LC stimulation by suH4, and asking whether histone sumoylation may help overcome previously reported LC inhibition by H3 acetylation.

In this chapter, we report a detailed analysis of PTM context-dependent demethylation activity by LSD1 and the LC sub-complex towards an array of methylated and acetylated H3 substrates.

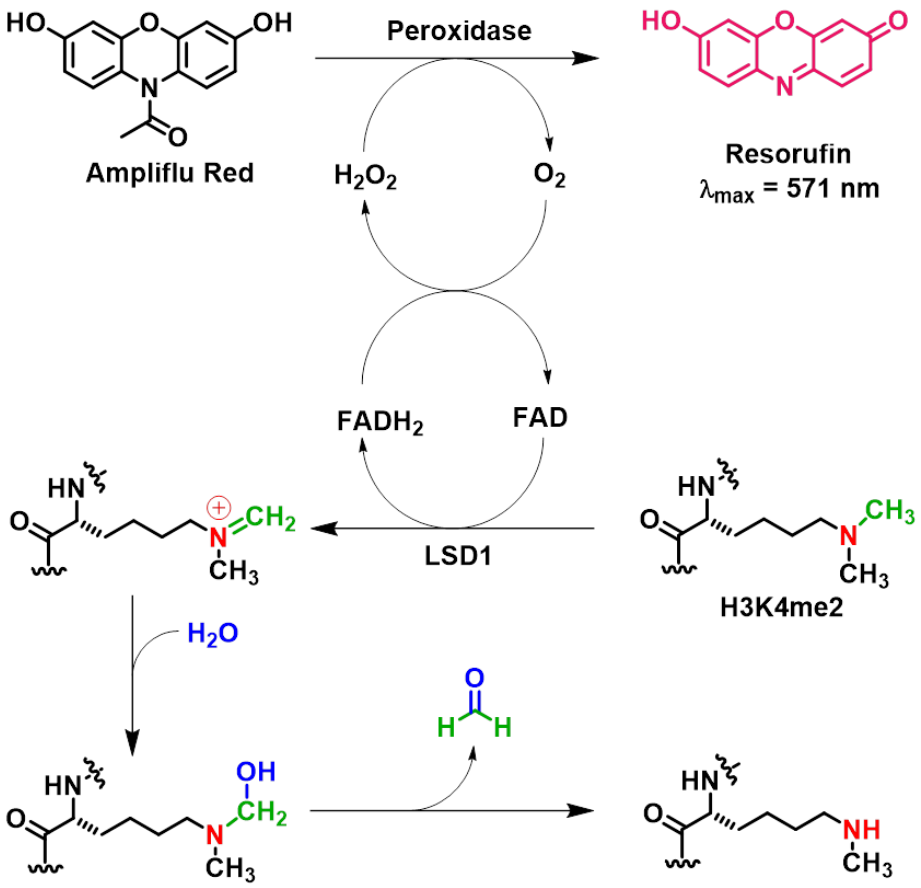
We further developed and optimized a high-throughput, microplate-based assay platform for monitoring LC demethylation kinetics. We show that differential monoacetylation of H3 peptides has no effect on the overall efficiency of LC K4me2 demethylation, whereas multiply acetylated H3 tails completely blocked LC activity. Toward ascertaining the effect of H3 acetylation on LC activity, both in the presence and absence of suH4, microplate assay development enabled steady-state LC kinetics on a suH4-containing mononucleosome for the first time. Importantly, previous studies involving recombinant LC sub-complex have employed truncated variants of these proteins for ease of purification. Here, we employed full-length LSD1 and CoREST1 which provides the first accurate interpretation of LC sub-complex activity.

4.2 Results and Discussion

4.2.1 H3(1-21) peptide synthesis and steady-state demethylation kinetics

Histone H3 acetylation is known to inhibit LSD1 demethylation toward a H3K4(1-21)me1 peptide substrates.[32,35] Previous reports have indicated that monoacetylation at K9 reduced the enzyme-substrate binding affinity by 6-fold, and hyperacetylation completely blocked LSD1 activity. Before attempting to delineate the effects of additional epigenetic marks on H3K4me2 demethylation in the context of a mononucleosome, we first sought to establish a complete picture of the effect of H3 acetylation on LSD1 using peptide substrates and to measure Michaelis-menten kinetics with full-length LSD1. We expected that the results from our study could both inform future experiments with regards to which full-length H3 constructs to pursue, and also fill in gaps in the literature regarding LSD1 specificity and the crosstalk between H3 methylation and acetylation

Towards this goal, we prepared six different modified H3(1-21)K4me2 peptides via 9-fluorenylmethoxycarbonyl solid-phase peptide synthesis (Fmoc-SPPS) and purified them by reverse-phase high-performance liquid chromatography (RP-HPLC) (**Figure 4.S1**). In addition



Scheme 4.2. Peroxidase-coupled demethylation assay using LSD1.

to an H3(1-21) peptide containing solely the K4me2 modification, peptides with monoacetylation at K9, K14, and K18, as well as di- and tri-acetylated variants were prepared to assess the site-specific contribution of lysine acetylation to LSD1 activity. With peptide substrates in hand, full-length His_6-LSD1(1-852) and His_6-CoREST1(1-482) were overexpressed and purified from *Escherichia coli* as described in Chapter 3 (**Figure 4.S2**, Chapter 3.4, **Figure 3.S19A,B**).

LSD1 demethylation assays were carried out using a cuvette-based horse radish peroxidase (HRP) coupled assay developed by Mattevi and coworkers (**Scheme 4.2**).[32] In this reaction, the small molecule, Ampliflu™ Red (AR), is stoichiometrically oxidized to resorufin by HRP in the presence of hydrogen peroxide. H_2O_2 is a stoichiometric by-product of the demethylation

reaction resulting from the re-oxidation of $FADH_2$ by molecular oxygen. The production of resorufin was monitored by its distinct maximum absorbance at 571 nm. LSD1 demethylation assays on peptide substrates were performed in at least triplicate and kinetic parameters were derived from non-linear regression analysis of initial velocity versus substrate concentration plots (**Table 4.1, Figure 4.S3**).

Kinetic experiments were first performed using full-length LSD1 alone, and results from peptide demethylation assays on H3(1-21)K4me2 were in good agreement with previously reported values for this substrate (**Table 4.1**). Demethylation assays with acetylated H3(1-21)K4me2 substrates indicated that a single acetylation mark on any of the three additional lysine residues in the H3(1-21) tail fragment did not drastically affect the catalytic efficiency of LSD1 (<2-fold effect). The substrate turnover in assays with H3K4me2K9ac and H3K4me2K14ac peptides was significantly reduced (~3.3-6.4-fold), but this was accompanied by an increased binding affinity (decreased K_M value) to LSD1 by nearly the same factor (~1.8-3.9-fold), resulting in similar k_{cat}/K_M values. Not surprisingly, acetylation of K18 (the most distal lysine residue from the site of demethylation) showed very little effect on LSD1 substrate binding or turnover. Strikingly, the addition of a second or third acetylation mark in the H3 tail completely abolished LSD1 activity, and no signal above background was observed for these substrates (**Table 4.1, Figure 4.S4**). As H3K4 methylation and H3 tail acetylation are marks of actively transcribed genes, this phenomenon suggests that in order to suppress gene transcription, HDAC activity may first be necessary, or at least concomitant, for LSD1-mediated histone demethylation. Consistent with this requirement, repressive complexes that contain LSD1 also contain HDAC1/2 activity. The effect of CoREST1 on the demethylation of peptide substrates was also assessed. Results from these experiments showed that CoREST1 has only a modest detrimental effect on LSD1 activity in the context of peptide substrates (**Table 4.1, Figure 4.S5**). In three previous reports describing the effect of truncated CoREST1 on LSD1 peptide demethylation, two studies found

Table 4.1. LSD1 kinetics on H3(1-21)K4me2 peptide substrates

LSD1

Substrate	K_M μM	k_{cat} min^{-1}	k_{cat}/K_M $\mu M^{-1}min^{-1}$
H3 (1-21) $K4me_2$	15.1 ± 2.4	8.9 ± 0.5	0.59 ± 0.1
H3 (1-21) $K4me_2$ K9ac	8.5 ± 2.6	2.7 ± 0.7	0.32 ± 0.05
H3 (1-21) $K4me_2$ K14ac	3.9 ± 0.9	1.4 ± 0.1	0.36 ± 0.1
H3 (1-21) $K4me_2$ K18ac	11.7 ± 3.0	6.4 ± 0.8	0.55 ± 0.2
H3 (1-21) $K4me_2$ K9, K14ac	nd	nd	nd
H3 (1-21) $K4me_2$ K9, K14, K18ac	nd	nd	nd

nd = no activity detected

LSD1+CoREST

Substrate	K_M μM	k_{cat} min^{-1}	k_{cat}/K_M $\mu M^{-1}min^{-1}$
H3 (1-21) K4me2	15.4 ± 2.9	5.3 ± 0.5	0.34 ± 0.1
H3 (1-21) K4me2 K9ac	7.1 ± 1.2	2.0 ± 1.2	0.28 ± 0.05
H3 (1-21) K4me2 K14ac	3.4 ± 0.1	1.0 ± 0.1	0.29 ± 0.1
H3 (1-21) K4me2 K18ac	9.5 ± 0.6	5.3 ± 0.5	0.56 ± 0.1
H3 (1-21) K4me2 K9, K14ac	nd	nd	nd
H3 (1-21) K4me2 K9, K14, K18ac	nd	nd	nd

nd = no activity detected

that CoREST1 had no influence, while one report described a >2-fold weaker substrate binding affinity.[23,37,48] Our findings argue that CoREST does not significantly influence LSD1 demethylation of peptide substrates.

4.2.2 Semisynthesis of H3K4me2 and H3K_c4me2 proteins

To study LSD1 in the more physiologically relevant context of mononucleosomes, access to homogenous, site-specifically modified pools of full-length histone proteins is crucial. To this end, established expressed protein ligation (EPL) strategies in our lab were employed to prepare full-length H3K4me2 and suH4 (Chapter 1.6, 3.1, **Scheme 3.3**).[47,49] First, the H3(1-28)K4me2 C-terminal peptide hydrazide was prepared via Fmoc-SPPS (**Figure 4.S6**). We note that in comparison to the synthesis of H3(1-21)K4me2 peptides, the H3(1-28) peptide was

significantly more challenging to synthesize due to several truncations in the sequence that arise from difficult couplings. In general, methylated peptides were produced by automated SPPS until the Q5 residue. After confirmation that this initial peptide was produced, Fmoc-Lys(me2)-OH was manually coupled to Q5 using milder microwave and base-free conditions and lower amino acid equivalents to conserve the expensive reagent and to suppress amino acid epimerization. While this strategy was effective for producing H3(1-21)K4me2 peptides, we observed very poor coupling of Fmoc-Lys(me2)-OH to the H3(5-28) resin-bound peptide, which ultimately resulted in an inseparable mixture of H3 peptide hydrazides. Attempts to optimize the coupling reaction by increasing amino acid equivalents, increasing the solution dielectric constant and introducing chaotropic agents benefitted the coupling yield, but we were still limited to <50% overall coupling efficiency. We speculate that the additional peptide length in the H3(1-28) peptide construct caused significant peptide aggregation within the resin polymer matrix, preventing access to the deprotected N-terminus for the incoming amino acid. Despite, increasing the Fmoc-Lys(me2)-OH coupling efficiency, the desired product was still inseparable from impurities by RP-HPLC. We overcame this challenge by capping the uncoupled N-terminal glutamine residue of H3(5-28) with benzoic anhydride, causing a significant shift in retention time by RP-HPLC that enabled purification of the desired H3(1-28)K4me2-$NHNH_2$ peptide (**Figure 4.S7**).

The purified peptide hydrazide was then used in a ligation reaction with recombinant H3(29-135)A29C,C110A that was overexpressed and purified from *E. coli* (Chapter 3.4**, Figures 3.S11-S12)**. Ligation products were purified by RP-HPLC. The natively-linked and site-specifically modified H3K4me2 protein was attained upon a terminal desulfurization of the Cys29 incorporated for NCL to the native Ala29 (**Figures 4.S8**).

In order to conserve the precious, natively-linked H3K4me2 protein and mononucleosome substrates, we first chose to optimize the LC-mediated demethylation reaction using a readily generated analog of H3K4me2, namely the thialysine analog H3K$_c$4me2. In 2007, Shokat and coworkers reported a convenient method for the synthesis of thialysine variants of methylated histones.[50] Briefly, Lys4 of H3 was mutated to cysteine, and the H3K4C construct was expressed in *E. coli* and further purified by RP-HPLC (**Figure 4.S9**). H3K4C was then used in an S_N2 reaction with 2-chloro-*N,N*-dimethylethanamine hydrochloride to obtain tens of milligram quantities of H3K$_c$4me2 (**Figure 4.S10**). The H3K$_c$4me2 protein was incorporated into histone octamers containing either wild-type H4 or semisynthetic suH4, along with wild-type H2A and H2B (**Figure 4.11A**, Chapter 3, **Figures 3.S9, 3.S14**). Given the large quantities of nucleosomes necessary for multiple demethylation assays, mass production of 147 bp Widom 601 DNA was undertaken using a 20x147 bp DNA repeat plasmid as previously described by Luger et al (**Figure 4.S12**).[51] Large-scale mononucleosome preparation was optimized following modified protocols originally reported by Luger and coworkers in order to obtain modified semisynthetic mononucleosomes (**Figure 4.S11B**).[51]

4.2.3 Development and optimization of a microplate-based demethylation assay

Upon generation of the methylated and sumoylated nucleosome substrates, we quickly realized several limitations in standard cuvette-based demethylation assays that we routinely used with peptidic substrates. First, mononucleosomes are intrinsically poorer substrates than peptides, which results in slower LSD1 substrate turnover and diminished absorbance signals at early time-points. In addition, the larger volumes required for cuvette-based assays presented a challenge with regard to the total amount of semisynthetic nucleosome substrates needed per assay. In fact, the challenge in assaying highly concentrated semisynthetic mononucleosomes necessitated a lower working concentration range in the demethylation assay, which led to

further reductions in absorbance signal. I demonstrate this with peptide substrates in **Figure 4.S13**.

I overcame the challenges associated with cuvette-based demethylation assays for mononucleosome substrates by developing a microplate assay platform and switching from measuring changes in absorbance to monitoring the fluorescence at 590 nm emission wavelength of the resorufin assay product. Fluorescence spectroscopy is inherently more sensitive than absorbance measurements due to the far lower background associated with light emission from molecules in an excited state. While Tan et al had previously described a 96-well plate fluorescence-based assay for LSD1 demethylation on $H3K_c4me2$-containing nucleosomes, few experimental details were reported and the 96-well format offered little in regards to conserving semisynthetic histones.[52] Therefore, I developed a 384-well plate fluorescence-based demethylation assay that I expected would greatly reduce the amount of required substrate and present a high-throughput method for measuring the kinetics of nucleosome demethylation.

Several factors were considered in optimizing the microplate-based assay. These included the assessment of background noise, plate reader-specific parameter adjustments, working concentration ranges and the overall quality of the enzyme complex. Fluorescence signal, unlike absorbance, is non-linear at high concentrations due to the re-absorbance of light emitted from the fluorophore before it reaches the detector. I demonstrated this non-linear fluorescence phenomenon, known as the inner-filter effect, using our resorufin detection platform (**Figure 4.S14**).[53] Due to this non-linearity in fluorescence signal, relative fluorescence units (RFU) cannot be simply converted into a workable unit of measurement (i.e. product concentration) in the same way that absorbance units can be converted using the Beer-Lambert law. With this in

mind, I reasoned that standard curves using known concentrations of resorufin must be included in all fluorescence-based assays.

Next, initial experiments with peptide and nucleosome substrates showed an unanticipated upward "drift" in fluorescence demethylation signal, indicating significant residual background fluorescence from one or more of the assay components. Performing fluorescence measurements on individual assay components alone indicated that various buffers were not a source of background noise, and neither were the LC complex nor the HRP enzymes (**Figure 4.S15A-C**). Upon incubating Ampliflu^{TM} Red (AR) in buffer alone I noticed a steady increase in fluorescence signal accounting for the previously observed demethylation assay background noise (**Figure 4.S15D**). Indeed, AR photooxidation is a well-characterized issue in light-based measurements using AR, especially at low working concentrations.[54–56] In order to suppress AR photooxidation, I limited the time between fluorescence "reads" for each hour-long assay to every 1.5 min. To account for any remaining AR photooxidation, assay blanks to which no methylated substrate is added were run in triplicate for every demethylation assay, and subtracted from the substrate demethylation signal during data processing. This subtraction turned out to be absolutely essential for the ability to measure demethylation kinetics.

4.2.4 Microplate-based steady-kinetics of LSD1 demethylation

After establishing reproducible microplate assay conditions, we were interested in assessing the influence of suH4 on LC activity when additional acetylation marks are present in the H3 tail. Thus far, I have prepared H3K$_c$4me2 mononucleosomes containing either wild-type H4 or suH4 and have successfully and reproducibly measured LC sub-complex kinetics under steady-state conditions **(Figure 4.3A**). Initial kinetics experiments demonstrated inconsistent and irreproducible kinetic outputs when especially low amounts of LC sub-complex (<200 nM) were used with low substrate concentration under steady-state parameters. In addition, we note that

inconsistent demethylation signal was observed at LC and substrate concentrations lower than 50 nM, and LSD1 activity is also known to significantly decrease at salt concentrations higher than 100 mM.[35,52] Therefore, we decided to measure apparent LC sub-complex kinetics using variable amounts of enzyme at salt concentrations lower than 100 mM.

To this end, we incubated 500 nM H3K$_c$4me2 and suH4 mononucleosomes with 50-1,000 nM LC in a 384-well plate. LC sub-complex quality was routinely checked by SDS-PAGE analysis and demethylation reactions with peptide substrates were first undertaken to check for robust activity. As previously mentioned, assay blanks were run in triplicates and resorufin standard curves (50-500 nM) were generated on the same plate (**Figure 4.3B-C**). The demethylation reaction was initiated by the addition of mononucleosomes to each well, and then resorufin production was measured over time. The raw fluorescence data was corrected by subtracting

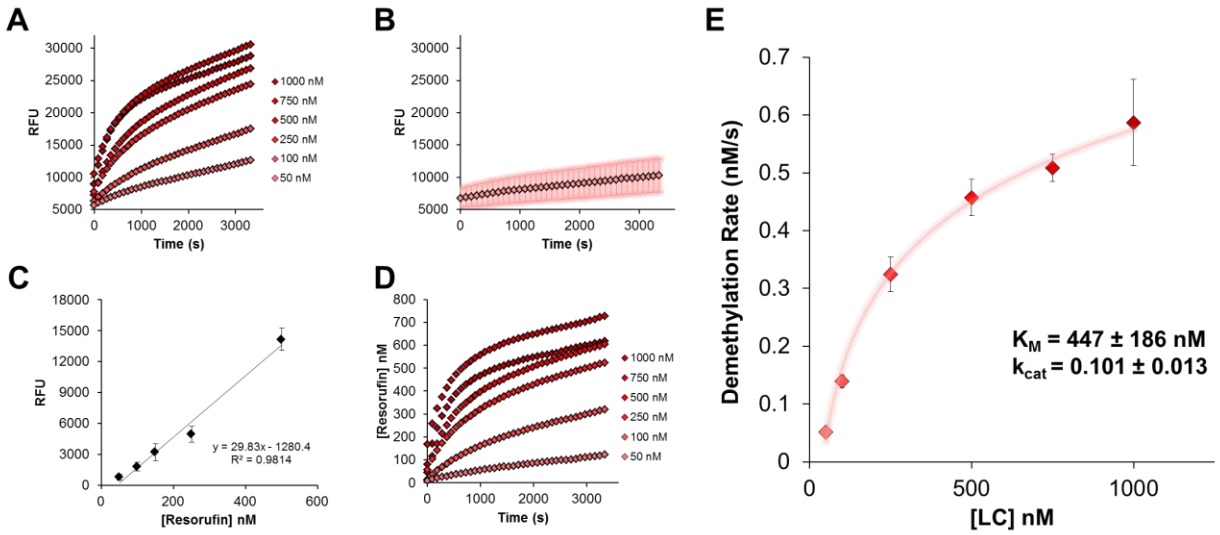


Figure 4.3. Steady-state LC sub-complex kinetics toward H3K$_c$4me2, suH4 mononucleosomes. (A) Overlaid fluorescence outputs from LC sub-complex demethylation assays H3K$_c$4me2, suH4 mononucleosomes, 50-1,000 nM. (B) Fluorescence output from averaged assay blanks. Error bars are standard error of the mean. (C) Averaged resorufin standard curve. Error bars are standard error of the mean (D) Correct LC sub-complex demethylation data, converted to units of nanomolar. (E) Steady-state demethylation rates versus LC sub-complex concentration plot. Error bars are standard error of the mean.

the resorufin blank and converted to the concentration of resorufin product using a resorufin standard curve (**Figure 4.3D**). The rate of demethylation during the first 500 seconds of the reaction was analyzed to derive the demethylation rates at early time points, which were then plotted against varying LC concentration (**Figure 4.3E**). Non-linear regression analysis of 2-D plots of rate versus enzyme complex concentration was used to derive a K_M of 447 ± 186 nM and a k_{cat} of 0.101 ± 0.013 min^{-1}. This result marks the first report of steady-state kinetic analysis using a 384-well microplate platform and semisynthetic mononucleosomes containing suH4, which will be followed by similar kinetic analysis with methylated and acetylated mononucleosomes.

4.3 Conclusions and outlook

In conclusion, we have harnessed the utility of protein semisynthesis and coupled-enzyme assays to monitor the kinetics of methylated and sumoylated mononucleosome substrates. Detailed kinetic parameters of the LC sub-complex were first established on peptidic H3(1-21)K4me2 substrates which demonstrated the susceptibility of LSD1 to H3 tail acetylation. We showed that monoacetylation at K9, K14 or K18 had little effect on LSD1 catalytic efficiency toward H3K4me2, despite changes in turnover rate and binding affinity of these substrates. However, the addition of one or two more acetylation marks in the H3 tails completely abolished LSD1 activity. Intrigued by these results, we are interested in assessing LSD1 kinetics in the context of mononucleosome substrates. Given our established LC sub-complex stimulation by suH4 seen in western blots, we are further interested in establishing what, if any, effect histone sumoylation has on LC sub-complex activity in the presence of H3 acetylation. To this end, we developed a high-throughput microplate-based demethylation assay in order to conserve precious semisynthetic mononucleosomal substrates and to facilitate data collection and interpretation. By using this platform, I have successfully established steady-state kinetic parameters for the LC-mediated demethylation of $H3K_c4me2$ and suH4 containing

mononucleosomes for the first time. Future experiments in this platform will aim to assess potential differences in demethylation kinetics of suH4 versus wild-type H4 containing mononucleosomes, and measure LC sub-complex activity towards methylated and acetylated mononucleosomes in the presence or absence of sumoylation.

4.4 Experimental procedures

4.4.1 General methods

2-chlorotrityl chloride and Rink amide resins were purchased from AnaSpec (Fremont, CA). Standard Fmoc-L-amino acids were purchased from MilliporeSigma (Burlington, MA), AGTC Bioproducts (Wilmington, MA), or AnaSpec. All other chemical reagents were purchased from MilliporeSigma 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). 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). 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 (RPHPLC) was performed on a Varian (Palo Alto, CA) ProStar HPLC or Agilent (Santa Clara, CA)1260 Infinity II 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 or Bruker Autoflex II MALDI-MS. 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).

4.4.2 Solid-phase peptide synthesis

Synthesis of H3(1-21)K4me2 peptides

The peptide H_2N-ARTK(me2)QTARKSTGGKAPRKSTGGQLA-OH corresponding to the first 21 N-terminal residues of the human histone H3 protein was synthesized by SPPS on a 0.05 mmol scale employing standard Fmoc-based Nα-deprotection chemistry and Lys(ac) residue(s) at the indicated positions. Briefly, Rink amide resin (0.3-0.6 mmol/g) was coupled with amino acids 5-21 in 5 mol. eq. based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 20% piperidine in DMF for 2 min at 90 °C. Coupling reactions were undertaken for 2 min at 90 °C with a mixture of Fmoc-amino acid HBTU (4.8 mol. eq.) and DIEA (5 mol. eq.) in DMF. For Arg, an additional coupling reaction was performed for 10 min at 90 °C. Fmoc-Lys(Ac)-OH were coupled at the indicated position in place of Fmoc-Lys(Boc)-OH. Synthesis of the peptide was confirmed by test cleavage of 5 mg resin and RP-HPLC and ESI-MS analysis. Fmoc-Lys(me2)-OH was coupled to resin-bound peptide using a mixture if amino acid (2 mol. eq.), ethyl cyano(hydroxyimino)acetate (Oxyma Pure, 2 mol. eq.) and diisopropylcarbodiimide (DIC, 2 mol eq) in DMF overnight. Deprotection of the Fmoc group was performed by treating resin with 20% piperidine in DMF for 25 min. Remaining coupling reactions were undertaken for 2-3 h at room temperature with a mixture of Fmoc-amino acid (8 mol eq.), Oxyma (8 mol. eq.) and DIC (8 mol. eq.) in DMF. Products were confirmed by test cleavage of 5 mg of resin. Peptide was cleaved and deprotected by reaction of resin at 100 µL/mg with a mixture of TFA:thioanisole:H_2O:triispropylsilane (TIS) (90:2.5:2.5:2.5:2.5 v/v) for 3 h 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 10-35% B (**Figure 4.S1**).

ESI-MS of H3(1-21)K4me2: Calculated *m/z* $[M+H]^+$ 2,282.7 Da, observed 2,282.1 ± 0.4 Da.

ESI-MS of H3(1-21)K4me2, K9ac: Calculated *m/z* $[M+H]^+$ 2,324.4 Da, observed 2,324.7 ± 1.3 Da.

ESI-MS of H3(1-21)K4me2, K14ac: Calculated *m/z* $[M+H]^+$ 2,324.4 Da, observed 2,235.7 ± 1.2 Da.

ESI-MS of H3(1-21)K4me2, K18ac: Calculated *m/z* $[M+H]^+$ 2,324.4 Da, observed 2,324.9 ± 1.0 Da.

ESI-MS of H3(1-21)K4me2, K9,14ac: Calculated *m/z* $[M+H]^+$ 2,366.7 Da, observed 2,366.5 ± 1.2 Da.

ESI-MS of H3(1-21)K4me2, K9,14,18ac: Calculated *m/z* $[M+H]^+$ 2,408.8Da, observed 2,408.1 ± 0.4 Da.

Synthesis of H3(1-28)K4me2-C(O)NHNH$_2$ peptides

The peptide H_2N-ARTK(me2)QTARKSTGGKAPRKQLATKAARKS-C(O)NHNH_2 corresponding to the first 28 N-terminal residues of the human histone H3 protein was synthesized by manual and microwave-assisted SPPS on a 0.1 mmol scale employing standard 9-fluorenylmethoxycarbonyl (Fmoc)-based Nα-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. 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 to the resin in a solution containing Fmoc-Ser(O^{t}Bu)-OH (0.4 mmol), *O*-(6-chlorobenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HCTU, 0.38 mmol) , and DIEA (0.8 mmol) for 60 min at 30 °C. From *tert*-butyl-serinyl 2-chlorotrityl hydrazine resin amino acids 5-27 were 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), Oxyma (5 mol. eq.) and DIC (5 mol. eq.) in DMF. For Arg, an additional coupling reaction was performed for 25 min at 75 °C. Deprotection of the Fmoc group was performed using 20% piperidine in DMF for 25 min. Fmoc-Lys(me2)-OH was coupled to the resin-bound peptide in a solution of amino acid (8 mol. eq.), Oxyma (8 mol. eq.), DIC (8 mol. eq.) and potassium thiocyanate (0.4 M) in 50:50 v/v DMOS and DMF overnight. Uncoupled Glu5 residues were capped by mixing the resin in a solution of benzoic anhydride (20 mol. eq.) and DIEA (40 mol. eq.) in 50:50 v/v DMF and DCM. The remaining amino acids were coupled in solutions of amino acid (8 mol. eq.), Oxyma (8 mol. eq.), DIC (8 mol. eq.) and potassium thiocyanate (0.4 M) in DMF. The peptide was cleaved and deprotected by reaction of resin at 100 µL/mg with TFA:H_2O:TIS:thioanisole (90:2.5:2.5:2.5:2.5 v/v) for 3 h 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 10-35% B (**Figure 4.S6**). ESI-MS of H3(1-28)K4me2-$NHNH_2$: Calculated *m/z* $[M+H]^+$ 3,039.6 Da, observed 3,040.2 ± 0.2 Da.

4.4.3 Overexpression and purification of full-length His_6-LSD1 (KDM1A, isoform b)

The pET15b-LSD1 plasmid was a kind gift from Dr. Yi Zhang. *E. coli* BL21(DE3) cells were transformed with the plasmid. For overexpression of His_6-LSD1, cells were grown at 37 °C in 3 L 2xYT media containing 100 µg/mL ampicillin to an OD_{600} of ~0.8-1.0 and were cooled down to 25 °C before induction. Protein expression was induced by the addition of 0.3 mM IPTG to the growth media and the cells were further grown at 25 °C for 4-6 h. At the end of the induction period, the cells were harvested by centrifugation at 5,000 rpm, resuspended in PBS, 0.2 mM PMSF, 5% glycerol buffer and lysed by sonication for 3 x 3 min with 1 min. rest in between. The lysate was centrifuged at 15,000 and the supernatant was applied to a Ni-NTA resin (5 mL) at 4 °C for 1.5 h. Column was washed with 5 column volumes (CV) each of lysis buffer containing 25 and 50 mM imidazole. Full-length His_6-LSD1 eluted with lysis buffer containing 250 mM

imidazole. Pure fractions were dialyzed against 20 mM tris, pH7.5, 5% glycerol buffer for 3 h and further purified by anion exchange chromatography. The fractions were visualized on a 12% SDS-PAGE gel and the fractions containing the desired protein in 90% purity were combined and concentrated. The amount of active LSD1 was determined by using the FAD extinction coefficient at 458 nm as 10,790 $cm^{-1}M^{-1}$. The total protein concentration was determined by comparing the LSD1 intensity on a coomassie stained gel 10% SDS-PAGE gel relative to BSA standards of known concentration (**Figure 4.S2**). 50-60% of the total protein was typically purified in the active form, based on cofactor occupancy.

4.4.4 Overexpression and purification of full-length His_6-CoREST1

(Refer to Chapter 3.4.15, **Figure 3.S19**)

4.4.5 Overexpression and purification of H3(29-135)A29C,C110A

(Refer to Chapters 3.4.3 and 3.4.6, **Figures 3.S11, 3.S12**)

4.4.6 Expressed protein ligation of H3(1-28)K4me2-C(O)$NHNH_2$ and H3(29-135)A29C,C110A

(Refer to Chapter 3.4.12)

4.4.7 Desulfurization of H3K4me2(A29C,C110A)

(Refer to Chapter 3.4.13, **Figure 4.S8**) ESI-MS of H3K4me2(C110A): Calculated *m/z* $[M+H]^+$ 15,252.0 Da, observed 15,253.3 ± 3.5 Da.

4.4.8 Purification of 147 bp Widom 601 DNA

The protocol DNA preparation reported by Luger et al was followed with modifications.[51] The pET15b-LSD1 plasmid was a kind gift from Dr. Ning Zheng. *E. coli* DH5α cells were transformed

with the plasmid. Cells were grown at 37 °C in 200 mL 2LB media containing 100 μg/mL ampicillin overnight. Plasmid DNA isolated using the Miraprep protocol reported by Peifer et al.[57] Isolated plasmid was visualized on a 0.6% agarose/TBE gel. 147 bp DNA fragments were liberated from the plasmid backbone through incubation with EcoRV-HF restriction endonuclease according to the manufacturer's recommendations. 0 h and 6 h time-points were visualized on a 5% polyacrylamide/TBE gel to confirmation complete digestion. To the solution, 0.3375 vol. 40% polyethylene glycol (PEG)-6000 and 0.15 vol. 5 M NaCl were added and incubated on ice for 1 h to precipitate the plasmid backbone. 2.5 vol. absolute ethanol and incubated at -20 °C overnight. Solution was centrifuged at 15,000 rpm, 4 °C for 30 min. Pelleted DNA was washed with 70% ethanol. DNA pellet air dried and resuspended in sterile H_2O, pH 8. DNA concentration measured via absorbance at 260 nm wavelength and using the extinction coefficient for 147 bp DNA of 2,784,500 $M^{-1}{\cdot}cm^{-1}$ (**Figures 4.S12**)

4.4.9 Overexpression and purification of H3K4C

E. coli BL21(DE3) cells containing the pET15b-H3(K4C) were grown at 37 °C in 6 L YT medium until OD_{600} ~0.7. Protein expression was induced by the addition of 0.3 mM IPTG for 3 h at 37 °C. Cells were harvested by centrifugation at 5,000 rpm, resuspended in 50 mM tris, pH 7.5 150 mM NaCl, 1 mM ethylene diamine tetraacetic acid (EDTA), 0.5% Triton X-100, 1 mM dithiothreitol (DTT) (wash buffer) and lysed by sonication. The lysate was centrifuged at 15,000 rpm for 20 min and the pelleted was washed with washed buffer followed by centrifugation. Insoluble histones were recovered from inclusion bodies with an extraction buffer consisting of 6 M guanidinium chloride (Gn-HCl), 50 mM tris, pH 7.5, 100 mM NaCl. The re-solubilized histones were dialyzed into H_2O/1 mM DTT overnight at 4 °C. Precipitated histone was collected and dried via lyophilization before C4 preparative RP-HPLC purification with a gradient of 30-70% B. Typical yields were 5-6 mg/L. ESI-MS of H3(K4C): Calculated *m/z* $[M+H]^+$ 15,199.8 Da, observed 15,199.7 ± 3.9 Da (**Figure 4.S9**).

4.4.10 Alkylation of H3K4C to generate H3K$_c$4me2

The methyllysine analog was synthesized according to the protocol published by Shokat et. al.[50] Briefly, Lyophilized H3(K4C) (5 mg) was dissolved in 900 µL alkylation buffer, 1 M HEPES pH 7.8, 4 M Gn-HCl, 10 mM D/L-methionine; and DTT was added (20 µl of 1 M, dissolved immediately before use). The histones were reduced for 1 hr at 37 °C and (2-chloroethyl)-dimethylammonium chloride was added (50 µL of 1 M, dissolved immediately before use) to perform the alkylation. After the reaction proceeded for 2 hr at room temperature additional DTT was added (10 ml of 1 M) to reduce any oxidized histone H3 K4C. The reaction incubated at room temperature for 30 min and treated with additional alkylating agent (50 mL, 1M dissolved immediately before use) and allowed to proceed for an additional 2 h at room temperature. The reaction was quenched with β-mercaptoethanol (50 µL, 14.2 M) and then purified to homogeneity by C4 preparative RP-HPLC with a gradient of 30-70% B. ESI-MS of H3(K4C): Calculated *m/z* $[M+H]^+$ 15,271.7 Da, observed 15,271.54 ± 1.4 Da (**Figure 4.S10**).

4.4.11 Histone octamer formation

(Refer to Chapter 3.4.17, **Figure 4.S11A**)

4.4.12 Large-scale mononucleosomes assembly

Prior to large scale mononucleosome preparation, a histone to octamer ratio at which free DNA is saturated or nearly saturated was determined on a small scale (refer to Chapter 3.4.18). Pure histone octamers and 147 bp Widom 601 DNA were combined in 150-300 µL of a high-salt refolding buffer consisting of 2 M NaCl, 20 mM tris, pH 8, to a final concentration of 10 µM. Octamers were transferred to a 100-500 µL, 3,500 MWCO dialysis button and placed in 700 mL of 20 mM tris, pH 8, 2 M NaCl in a 3.5 L beaker. Mononucleosomes were dialyzed to 400 mM NaCl by slow addition of 2800 mL 20 mM tris, pH 8 over 24 h. Mononucleosomes were transferred to a 1L beaker containing 500 mL previous dialysis buffer and further dialyzed to 200

mM NaCl by addition of 500 mL 20 mM tris, pH 8 overnight. Histone precipitation occurs after first dialysis. Mononucleosomes were gathered and the supernatant was separated from precipitated histone proteins. Mononucleosomes were checked on a 5% TBE gel for free DNA, and concentration was determined by measuring the absorbance at 260 nm wavelength and using the extinction coefficient for 147 bp DNA of 2,784,500 $M^{-1,}cm^{-1}$ (**Figure 4.S11B**).

4.4.13 Cuvette-based LSD1 demethylation assay

Stock solutions of H3(1-21)K4me2 peptides (140-220 μM), Ampliflu™ Red (AR, 100 mM, Sigma-Aldrich, St. Louis, MO), HPOFF horse radish peroxidase (76%, 100 μM, Worthington Biochemicals, Lakewood, NJ) and NaCl (500 mM) in 50 mM HEPES, pH 7.9 (assay buffer) were freshly prepared prior to kinetics experiments. Kinetics buffer consisting of AR (125 μM), HPOFF (0.95 uM) and LSD1 (or LC) (0.625 μM) were prepared from the stock solutions. Assay buffer, kinetics buffer and NaCl solution were combined in a sub-micro quartz cuvette (Starna Scientific, Ltd., Atascadero, CA) so that final concentration were 100 μM AR, 0.1 μM HPOFF, 0.5 μM LSD1 (or LC) and 50 mM NaCl upon addition of peptide substrate. Bubbles in the cuvette were carefully avoided. Demethylation reactions were initiated by the addition of 1-120 μM peptide to the cuvette, and quickly placed a NanoDrop 2000c UV-Vis spectrophotometer (ThermoFisher Scientific, Waltham, MA). Demethylation assays were prepared in the dark, and protected from light at all times. The production of resorufin was monitored at 571 nm for 20 min. with reads every 6 sec. Resorufin absorbance units were converted to units of product concentration using a resorufin extinction coefficient of 54,000 $M^{-1}cm^{-1}$. Initial velocities were derived from the first 50-100 sec of each demethylation reaction and plotted versus substrate concentration. Michaelis constants (K_M) and catalytic constants (k_{cat}) were derived using non-linear regression analysis in GraphPad Prism software.

4.4.12 384-well microplate-based LSD1 demethylation assay

Assay preparation and execution: Prior to nucleosome demethylation assays, LSD1 and CoREST1 were mixed in an equimolar ratio on ice for 15-30 min. The quality of the individual proteins and the 1:1 complex were routinely checked on a 10% or 12% SDS-PAGE gel prior to use (**Figure 4.S16**). Semisynthetic mononucleosome quality was checked by 5% TBE gel prior to use (**Figure 4.S11B**). Kinetics buffer (KB) consisting of Ampliflu^{TM} Red (AR, 250 μM) and HPOFF (25 μM) in 50 mM HEPES, pH 7.9 (assay buffer, AB), were freshly prepared before each assay from AR (25 mM) and HPOFF (200 μM) in AB stock solutions. A 1 M NaCl in AB stock solution was prepared for adjusting the reaction salt concentration. Resorufin stock solutions were prepared by mixing 2.1 mg AR in 1 mL AB (~10 mM, 200x solution). This solution is an opaque, deep red mixture. A 1x homogenous resorufin solution was prepared from the 200x mixture and its concentration measured on a NanoDrop 2000c spectrophotometer using the resorufin extinction coefficient of 54,000 $M^{-1}cm^{-1}$ at 573 nm wavelength. A 2.5 μM resorufin solution was prepared from the 1x stock solution. In a black, flat-bottom 384-well microplate (Corning Inc., Corning, NY), LC, AB, KB and NaCl were combined in demethylation wells so that their final concentrations were 0.05-1 μM LC, 5 μM HPOFF, 50 μM AR and 50 mM NaCl upon final addition of mononucleosomes. Blank wells were prepared as described above with 0.05 μM LC. Resorufin standard wells were prepared as described for blank wells and including 50-500 nM resorufin. Microplate assays were prepared in the dark, and protected from light at all times. Reactions were initiated through quick addition of semisynthetic mononucleosomes (0.5 μM) to demethylation assay wells and placement in a BioTek (Winooski, VT) Synergy 2 microplate reader.

Plate reader and assay parameters: Resorufin production was measured for 1 h with fluorescence reads every 1.5 min. 528 nm wavelength excitation and 590 nm wavelength emission filters were used for each assay. The plate reader detector was set to read from the

top at 4 mm distance, and sensitivity was set to ‘65’. Wells F9-F14 were used for all demethylation assay measurements and wells G9-G16 were used for blank and resorufin standard wells.

Data analysis: Assay blank RFU values were averaged and subtracted from the raw, averaged demethylation signal for each set of LC complex concentration data. Corrected RFU values were converted to units of concentration (nM) using averaged slope and y-intercept values from all resorufin standard curves. The initial reaction rates were derived from the first 500 sec of product concentration versus time plots, and further plotted against LC sub-complex concentration. Non-linear regression analysis was performed with GraphPad Prism software by constraining substrate concentration to 0.5 μM to generate K_M and k_{cat} values.

4.5 Product characterization and supplemental data

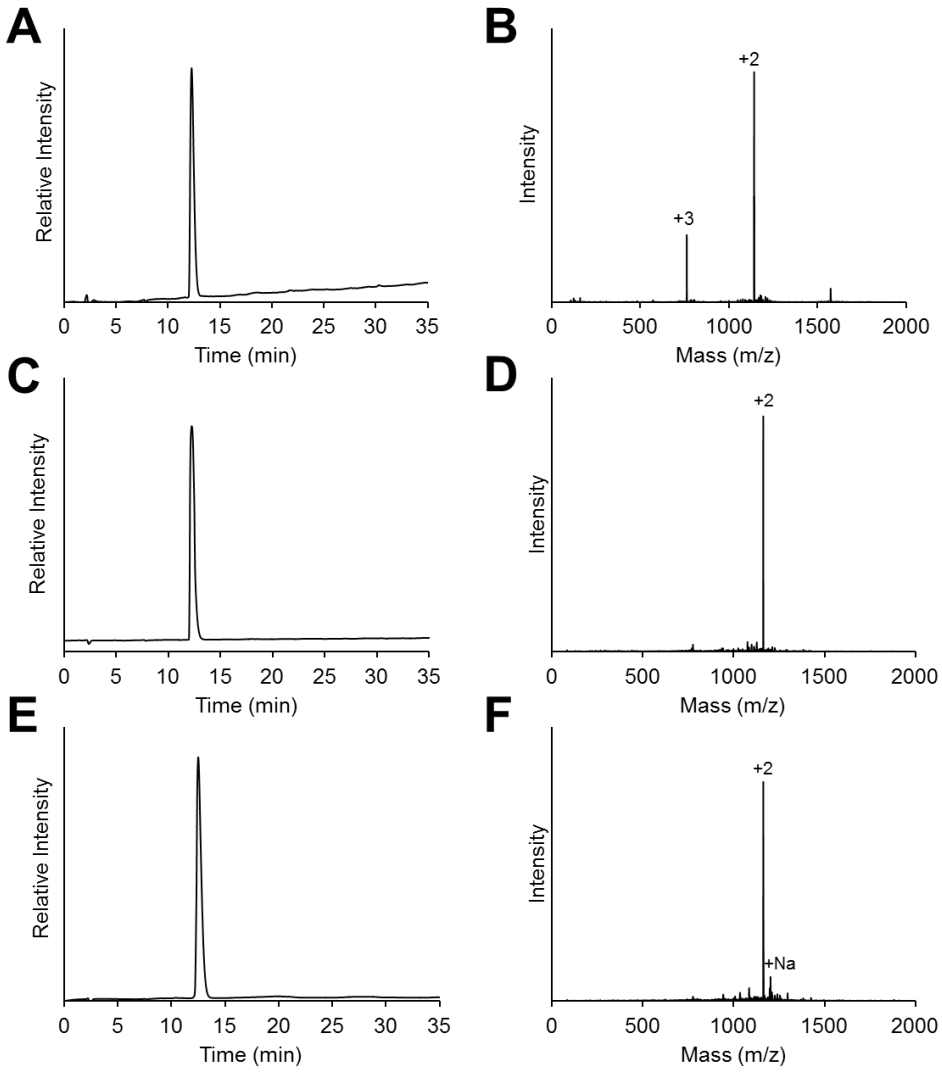


Figure 4.S1. Characterization of H3(1-21)K4me2 peptides. (A) C18 analytical RP-HPLC of H3(1-21)K4me2, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of purified H3(1-21)K4me2. (C) C18 analytical RP-HPLC of H3(1-21)K4me2, K9ac 0-73% CH_3CN in H_2O, 30 min. gradient. (D) ESI-MS of purified H3(1-21)K4me2, K9ac. (E) C18 analytical RP-HPLC of H3(1-21)K4me2, K14ac, 0-73% CH_3CN in H_2O, 30 min. gradient. (F) ESI-MS of purified

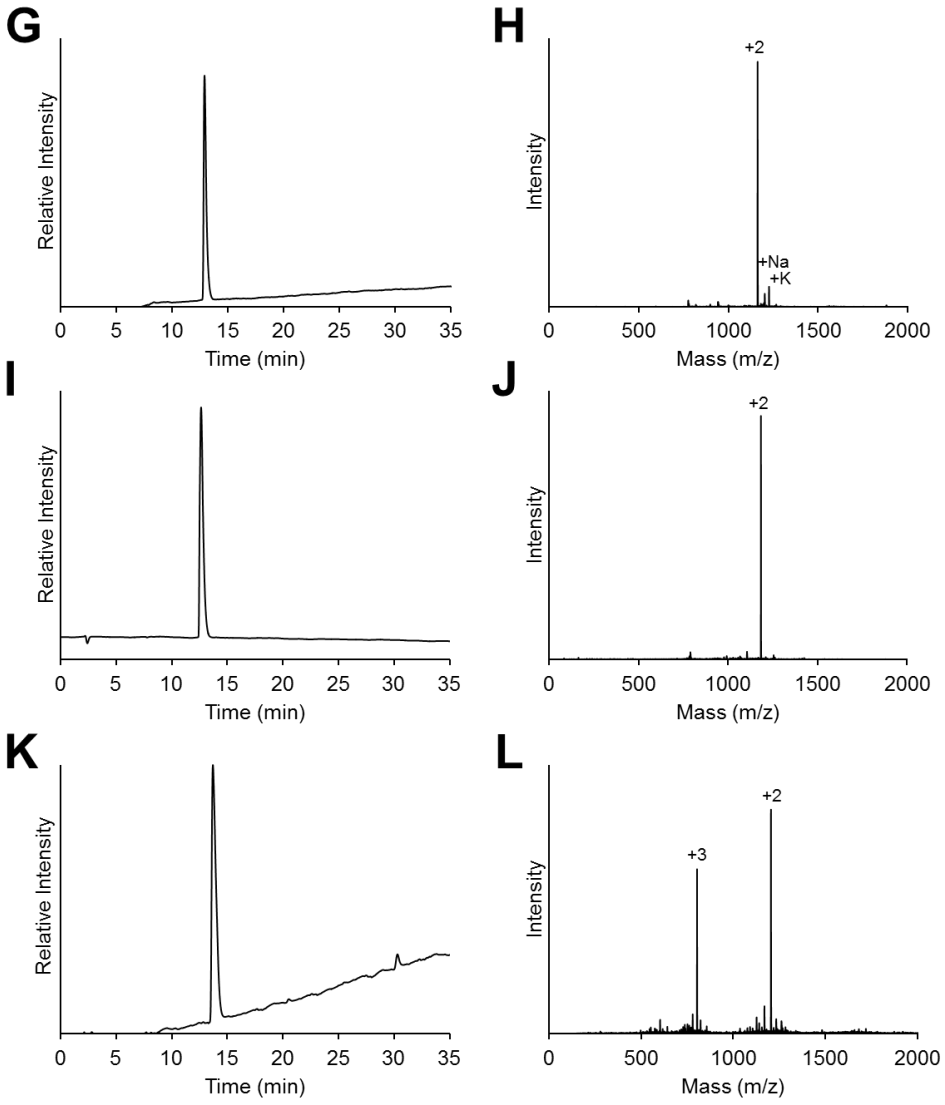


Figure 4.S1 continued. Characterization of H3(1-21)K4me2 peptides (G) C18 analytical RP-HPLC of H3(1-21)K4me2, K18ac 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS of purified H3(1-21)K4me2, K18ac. (C) C18 analytical RP-HPLC of H3(1-21)K4me2, K9,14ac, 0-73% CH_3CN in H_2O, 30 min. gradient. (D) ESI-MS of purified H3(1-21)K4me2, K9,14ac. (E) C18 analytical RP-HPLC of H3(1-21)K4me2, K9,14,18ac, 0-73% CH_3CN in H_2O, 30 min. gradient. (F) ESI-MS of purified H3(1-21)K4me2, K9,14,18ac.

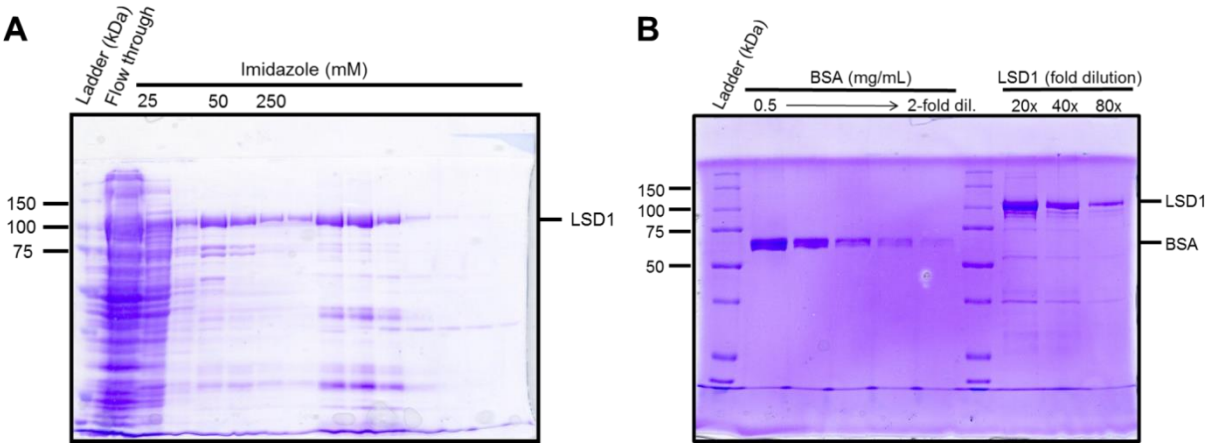


Figure 4.S2. Purification of His_6-LSD1. (A) 15% SDS-PAGE of Ni-NTA IMAC purification of His_6-LSD1. (B) Quantification of IEC-purified His_6-LSD1 protein concentration through comparing densities to BSA standards of known concentration.

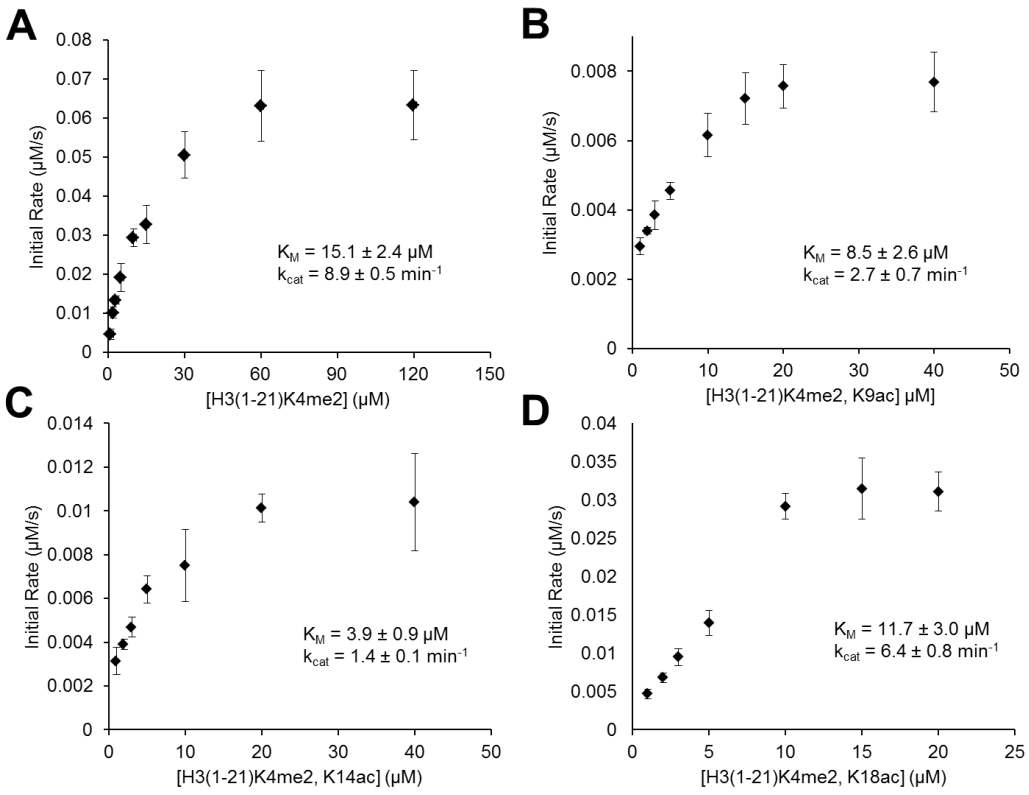


Figure 4.S3. LSD1 kinetics on peptide substrates. (A) Initial demethylation rate versus H4(1-21)K4me2 concentration plot, 1-120 μM (B) Initial demethylation rate versus H4(1-21)K4me2, K9ac concentration plot, 1-40 μM (C) Initial demethylation rate versus H4(1-21)K4me2, K14ac concentration plot, 1-40 μM (D) Initial demethylation rate versus H4(1-21)K4me2, K18ac concentration plot, 1-20 μM. n≥3

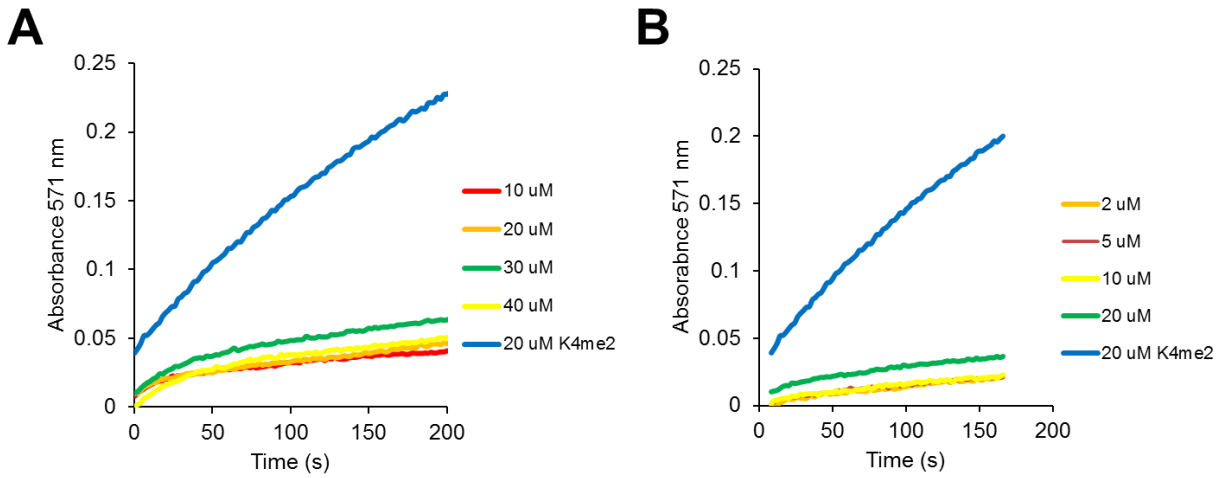


Figure 4.S4. LSD1 demethylation of H3(1-21)K4me2, K9,K14ac and H3(1-21)K4me2, K9,14,18ac peptides. (A) Overlaid absorbance outputs from LSD1 demethylation assays on the H3(1-21)K4me2, K9,14ac peptide, 10-40 μM, in comparison to demethylation of 20 μM H3(1-21)K4me2 peptide (B) Overlaid absorbance outputs from LSD1 demethylation assays on the H3(1-21)K4me2, K9,14,18ac peptide, 2-20 μM, in comparison to demethylation of 20 μM H3(1-21)K4me2 peptide

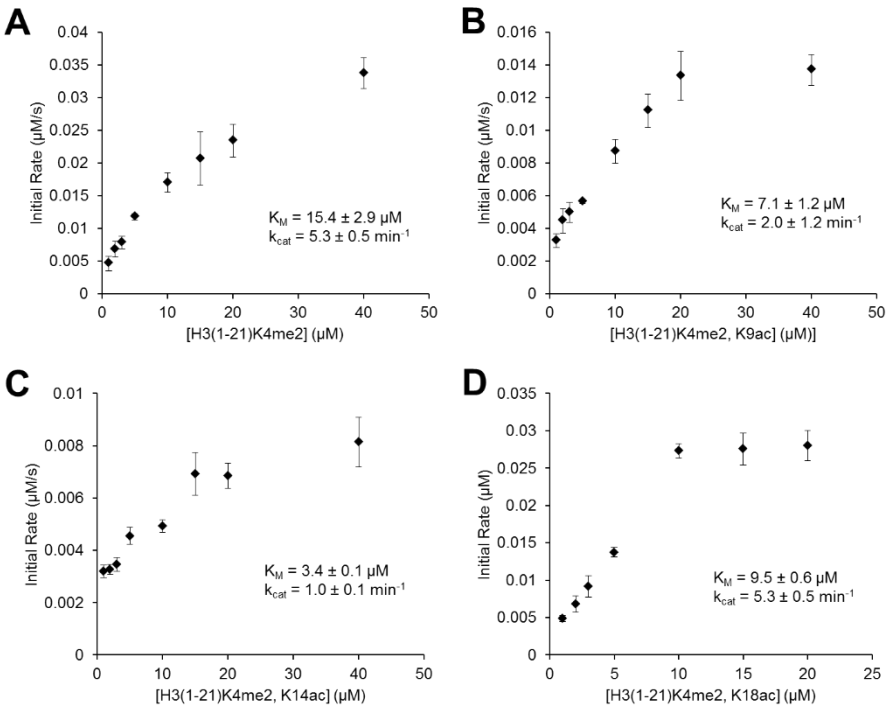


Figure 4.S5. LC sub-complex kinetics on peptide substrates. (A) Initial demethylation rate versus H4(1-21)K4me2 concentration plot, 1-40 μM (B) Initial demethylation rate versus H4(1-21)K4me2, K9ac concentration plot, 1-40 μM (C) Initial demethylation rate versus H4(1-21)K4me2, K14ac concentration plot, 1-40 μM (D) Initial demethylation rate versus H4(1-21)K4me2, K18ac concentration plot, 1-20 μM. n≥3

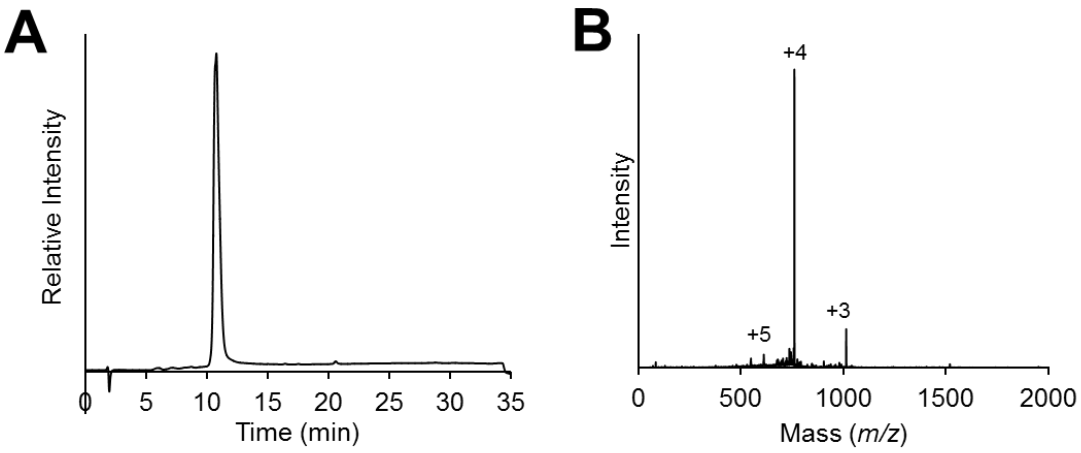


Figure 4.S6. Characterization of H3(1-28)K4me2-C(O)-NHNH$_2$ (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS.

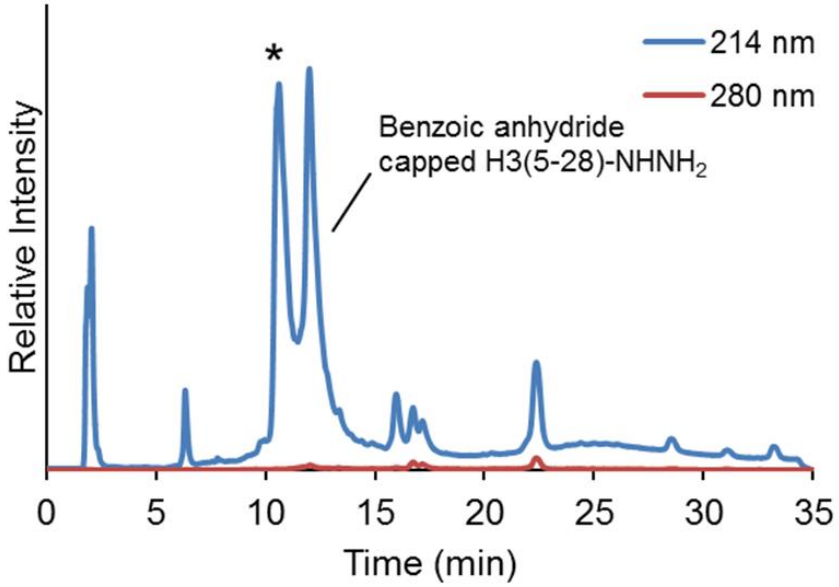


Figure 4.S7. Separation of H3(1-28)K4me2-NHNH2 from impurities. C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. * = Desired peptide product.

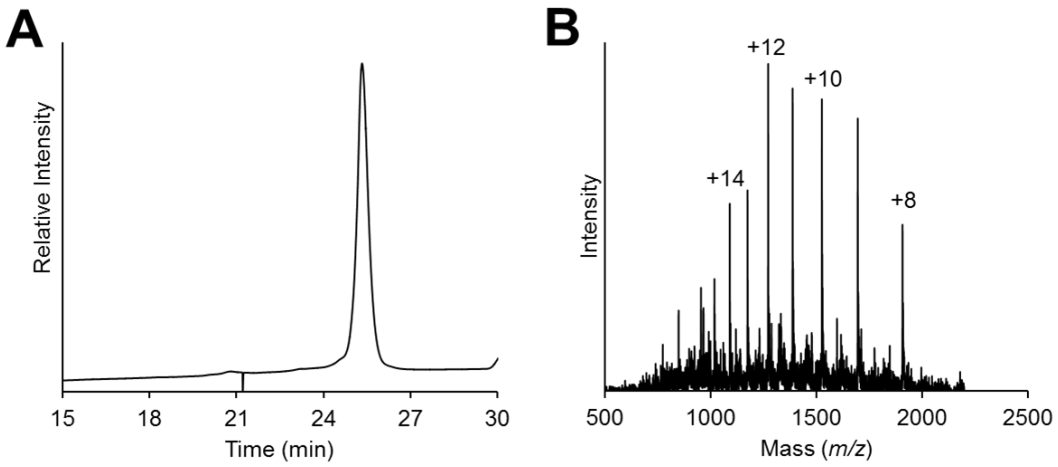


Figure 4.S8. Characterization of H3K4me2(C110A). (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS.

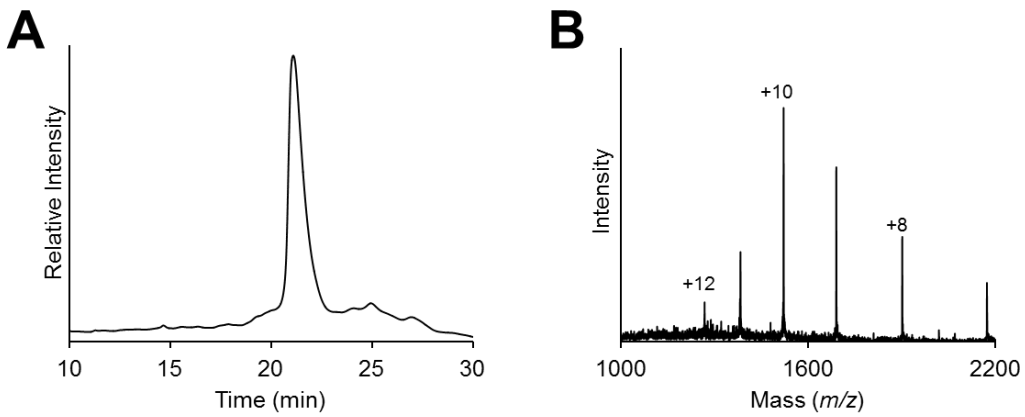


Figure 4.S9. Characterization of H3K4C. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS.

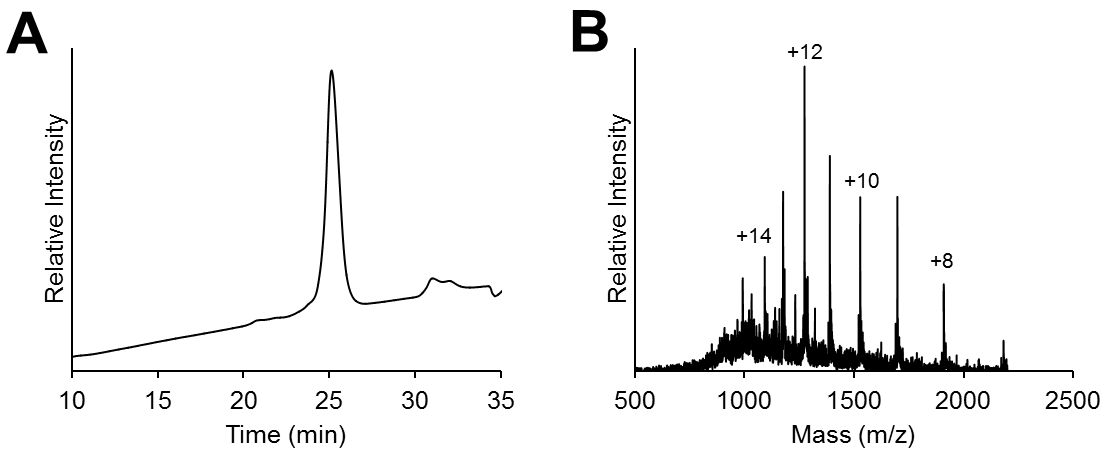


Figure 4.S10. Characterization of H3K$_c$4me2. (A) C18 analytical RP-HPLC, 0-73% CH_3CN in H_2O, 30 min. gradient. (B) ESI-MS.

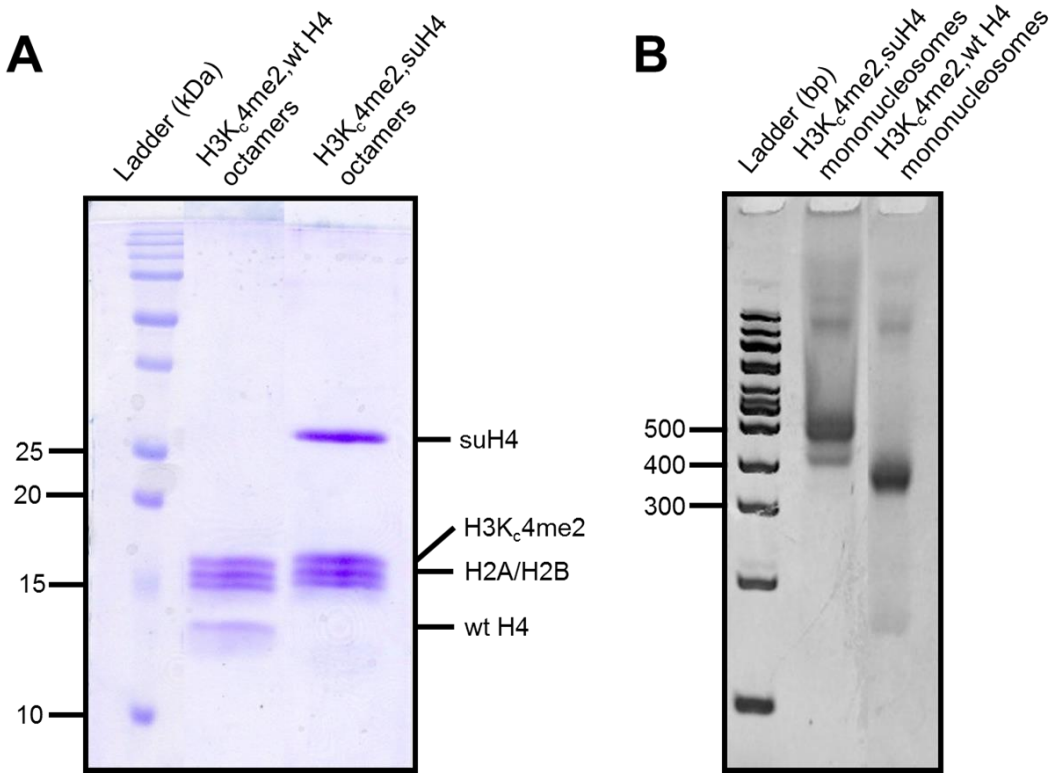


Figure 4.S11. Visualization of H3K$_c$4me2 octamers and mononucleosomes. (A) 15% SDS-PAGE gel of purified H3Kc4me2, wt H4 and H3Kc4me2, suH4 octamers. (B) 5% TBE gel of H3Kc4me2, wt H4 and H3Kc4me2, suH4 mononucleosomes.

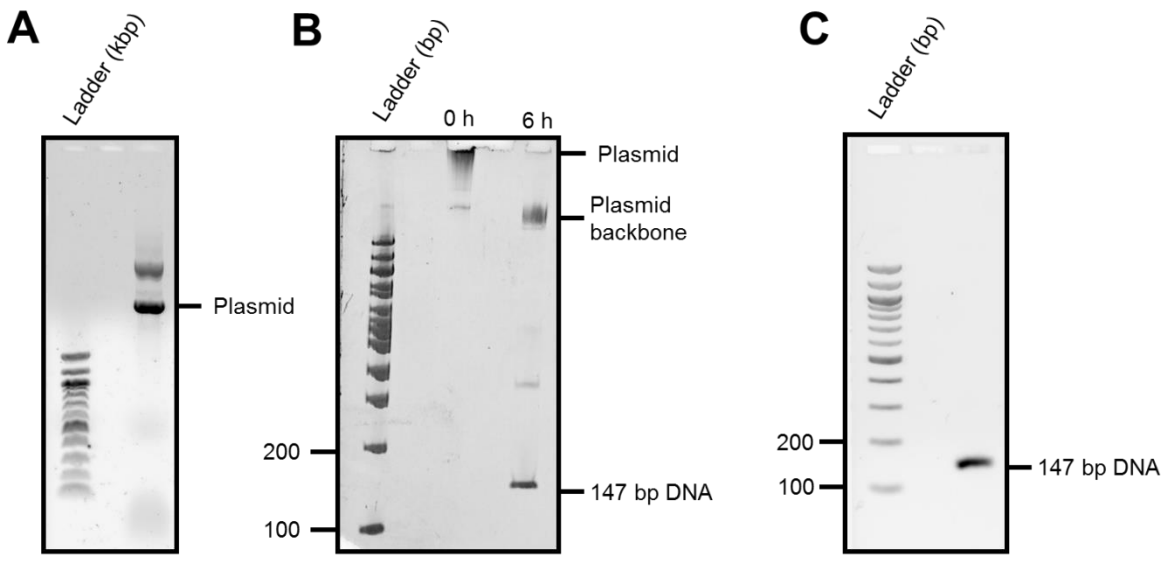


Figure 4.S12. Large-scale preparation of 147 bp Widom DNA. (A) 0.6% agarose/TBE gel of isolated 20x147bp repeat DNA plasmid (B) 5% polyacrylamide/TBE gel of EcoRV-HF restriction digest time-points. (C) 1.5% agarose/TBE gel of purified 147 bp Widom DNA.

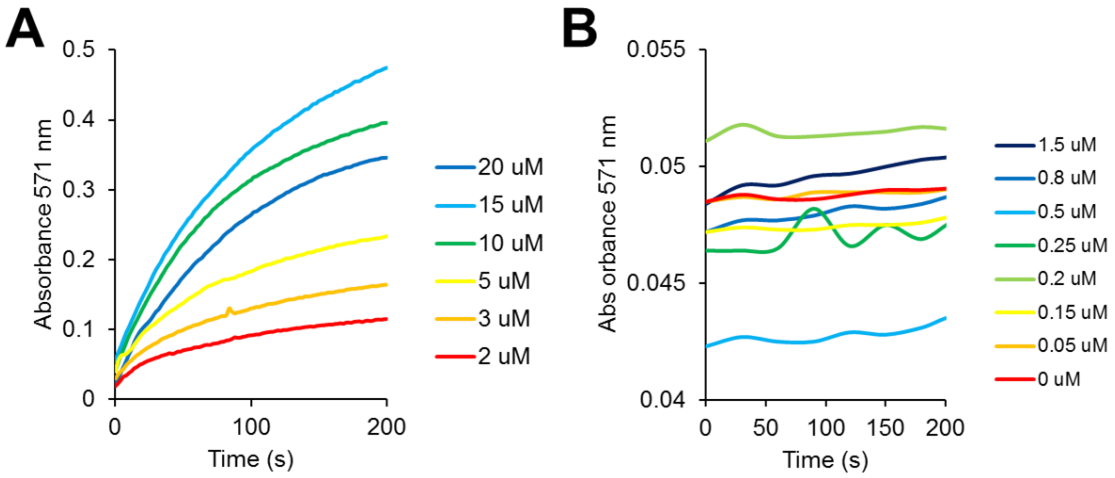


Figure 4.S13. Absorbance spectroscopy sensitivity is not sufficient at low methylated substrate concentrations. (A) Overlaid absorbance outputs from LSD1 demethylation assays on the H3(1-21)K4me2, K18ac peptide in a high concentration regime (2-20 μM) (B) Overlaid absorbance outputs from LSD1 demethylation assays on the H3(1-21)K4me2, K18ac peptide in a low concentration regime 0-1.5 μM).

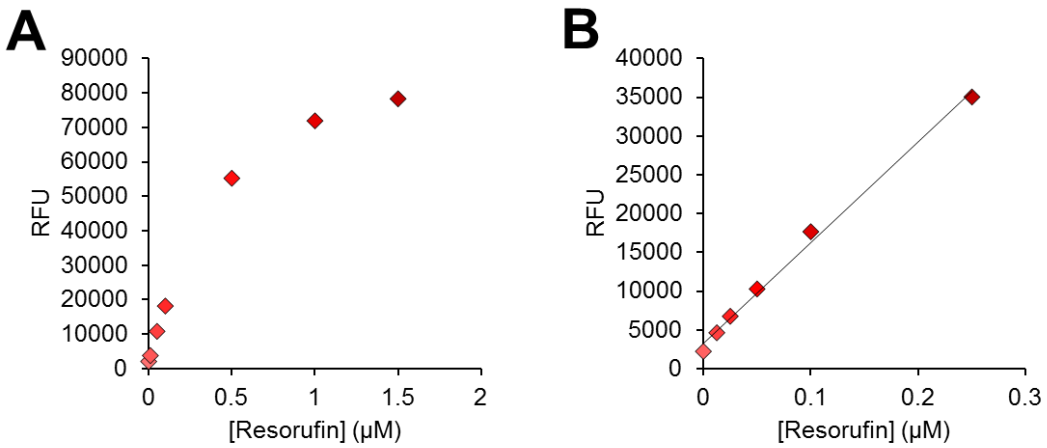


Figure 4.S14. Inner-filter effect in fluorescence spectroscopy. (A) Fluorescence signal from a high resorufin concentration regime (B) Fluorescence signal from a low resorufin concentration regime.

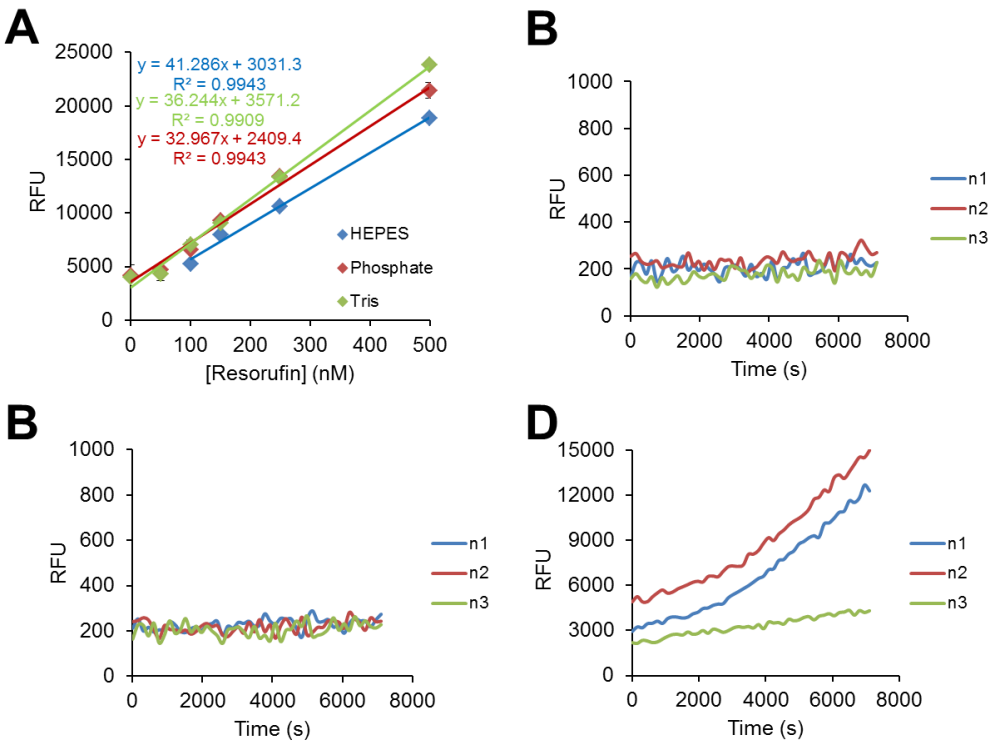


Figure 4.S15. Contribution of assay components to demethylation assay signal noise. (A) Resorufin standard curves generated in either HEPES buffer (50 mM, pH 7.9), tris buffer (20 mM, pH 7.9) or phosphate buffer (100 mM, pH 7.9). n=3 (B) Assay time course of LSD1 alone in 50 mM HEPES, pH 7.9 buffer. (C) Time course of HPOFF peroxidase alone in 50 mM HEPES, pH 7.9 buffer. (D) Time course of AR alone in 50 mM HEPES, pH 7.9 buffer.

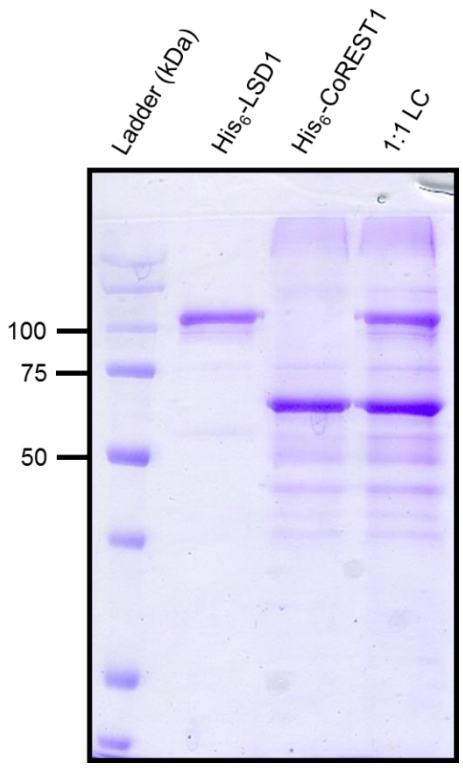


Figure 4.S16. Demethylation assay enzyme components. 10% SDS-PAGE gel of purified LSD1, CoREST1 and the preformed 1:1 LC subcomplex used in demethylation assays.

4.6 References

(1) Kouzarides, T. *Curr. Opin. Genet. Dev.* **2002**, *12* (2), 198–209.

(2) Greer, E. L.; Shi, Y. *Nat. Rev. Genet.* **2012**, *13* (5), 343–357.

(3) Fiszbein, A.; Kornblihtt, A. R. *Neurogenesis* **2016**, *3* (1), 1–5.

(4) Gong, F.; Miller, K. M. *Mutat. Res. - Rev. Mutat. Res.* **2017**, *780* (August 2017), 37–47.

(5) Allfrey, E. L. *J. Biol. Chem.* **1969**, *244* (18), 4871–4877.

(6) Borun, T. W.; Pearson, D.; Paik, W. K. *J. Biol. Chem.* **1972**, *247* (13), 4288–4298.

(7) Tessarz, P.; Santos-Rosa, H.; Robson, S. C.; Sylvestersen, K. B.; Nelson, C. J.; Nielsen, M. L.; Kouzarides, T. *Nature* **2014**, *505* (7484), 564–568.

(8) Bedford, M. T. *J. Cell Sci.* **2007**, *120* (24), 4243–4246.

(9) Lachner, M.; O'Sullivan, R. J.; Jenuwein, T. *J. Cell Sci.* **2003**, *116* (11), 2117–2124.

(10) Barski, A.; Cuddapah, S.; Cui, K.; Roh, T. Y.; Schones, D. E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. *Cell* **2007**, *129* (4), 823–837.

(11) Sone, K.; Piao, L.; Nakakido, M.; Ueda, K.; Jenuwein, T.; Nakamura, Y.; Hamamoto, R. *Nat. Commun.* **2014**, *5*, 1–12.

(12) Teske, K. A.; Hadden, M. K. *Eur. J. Med. Chem.* **2017**, *136*, 14–35.

(13) Liang, G.; Lin, J. C. Y.; Wei, V.; Yoo, C.; Cheng, J. C.; Nguyen, C. T.; Weisenberger, D. J.; Egger, G.; Takai, D.; Gonzales, F. A.; Jones, P. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (19), 7357–7362.

(14) Bernstein, B. E.; Humphrey, E. L.; Erlich, R. L.; Schneider, R.; Bouman, P.; Liu, J. S.; Kouzarides, T.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (13), 8695–8700.

(15) Zegerman, P.; Canas, B.; Pappin, D.; Kouzarides, T. *J. Biol. Chem.* **2002**, *277* (14), 11621–11624.

(16) Ruthenburg, A. J.; Allis, C. D.; Wysocka, J. *Mol. Cell* **2007**, *25* (1), 15–30.

(17) Kim, T.; Buratowski, S. *Cell* **2009**, *137* (2), 259–272.

(18) Kaimori, J.; Maehara, K.; Hayashi-takanaka, Y.; Harada, A. *Nat. Publ. Gr.* **2016**, No. April, 1–10.

(19) Carroll, Â. O.; Rea, S.; Mechtler, K.; Lachner, M. *Nature* **2001**, *410* (March), 116–120.

(20) Kathrin Plath, Jia Fang, Susanna K. Mlynarczyk-Evans, Ru Cao, Kathleen A. Worringer, H. W.; CecileC. dela Cruz, ArieP. Otte, Barbara Panning, Y. Z. *Science (80-.).* **2002**, *298* (5616), 1039–1043.

(21) Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T. *Nature* **2001**, *410* (6824), 120–124.

(22) Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z. W.; Schmid, M.; Opravil, S.; Mechtier, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T. *Nature* **2000**, *406* (6796), 593–599.

(23) Smith, B. C.; Denu, J. M. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2009**, *1789* (1), 45–57.

(24) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. *Cell* **2004**, *119* (7), 941–953.

(25) Kooistra, S. M.; Helin, K. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (5), 297–311.

(26) Mosammaparast, N.; Shi, Y. *Annu. Rev. Biochem.* **2010**, *79* (1), 155–179.

(27) Shi, Y.; Whetstine, J. R. *Mol. Cell* **2007**, *25* (1), 1–14.

(28) Yang, M.; Gocke, C. B.; Luo, X.; Borek, D.; Tomchick, D. R.; Machius, M.; Otwinowski, Z.; Yu, H. *Mol. Cell* **2006**, *23* (3), 377–387.

(29) Shi, Y.-J. Y.; Matson, C.; Lan, F.; Iwase, S.; Baba, T. *Mol. Cell* **2005**, *19* (6), 857–864.

(30) Lee, M. G.; Wynder, C.; Cooch, N.; Shiekhattar, R. *Nature* **2005**, *437* (7057), 432–435.

(31) Forneris, F.; Binda, C.; Vanoni, M. A.; Mattevi, A.; Battaglioli, E. *FEBS Lett.* **2005**, *579* (10), 2203–2207.

(32) Forneris, F.; Binda, C.; Vanoni, M. A.; Battaglioli, E.; Mattevi, A. *J. Biol. Chem.* **2005**, *280* (50), 41360–41365.

(33) Stavropoulos, P.; Blobel, G.; Hoelz, A. *Nat. Struct. Mol. Biol.* **2006**, *13* (7), 626–632.

(34) Da, G.; Lenkart, J.; Zhao, K.; Shiekhattar, R.; Cairns, B. R.; Marmorstein, R. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (7), 2057–2062.

(35) Forneris, F.; Binda, C.; Dall'Aglio, A.; Fraaije, M. W.; Battaglioli, E.; Mattevi, A. *J. Biol. Chem.* **2006**, *281* (46), 35289–35295.

(36) Wang, J.; Hevi, S.; Kurash, J. K.; Lei, H.; Gay, F.; Bajko, J.; Su, H.; Sun, W.; Chang, H.; Xu, G.; Gaudet, F.; Li, E.; Chen, T. *Nat. Genet.* **2009**, *41* (1), 125–129.

(37) Huang, J.; Sengupta, R.; Espejo, A. B.; Lee, M. G.; Dorsey, J. A.; Richter, M.; Opravil, S.; Shiekhattar, R.; Bedford, M. T.; Jenuwein, T.; Berger, S. L. *Nature* **2007**, *449* (7158), 105–108.

(38) Wang, Y.; Zhang, H.; Chen, Y.; Sun, Y.; Yang, F.; Yu, W.; Liang, J.; Sun, L.; Yang, X.; Shi, L.; Li, R.; Li, Y.; Zhang, Y.; Li, Q.; Yi, X.; Shang, Y. *Cell* **2009**, *138* (4), 660–672.

(39) You, a; Tong, J. K.; Grozinger, C. M.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 1454–1458.

(40) Tsai, M.; Manor, O.; Wan, Y.; Mosammaparast, N.; Wang, J. K.; Lan, F.; Shi, Y.; Segal, E.; Chang, H. Y. *Science (80-.).* **2010**, *329* (August), 689–693.

(41) Metzger, E.; Wissmann, M.; Yin, N.; Müller, J. M.; Schneider, R.; Peters, A. H. F. M.; Günther, T.; Buettner, R.; Schüle, R. *Nature* **2005**, *437* (September), 25–28.

(42) Augert, A.; Eastwood, E.; Ibrahim, A. H.; Wu, N.; Grunblatt, E.; Basom, R.; Liggitt, D.; Eaton, K. D.; Martins, R.; Poirier, J. T.; Rudin, C. M.; Milletti, F.; Cheng, W. Y.; Mack, F.; MacPherson, D. *Sci. Signal.* **2019**, *12* (567), 1–16.

(43) Sehrawat, A.; Gao, L.; Wang, Y.; Bankhead, A.; McWeeney, S. K.; King, C. J.; Schwartzman, J.; Urrutia, J.; Bisson, W. H.; Coleman, D. J.; Joshi, S. K.; Kim, D. H.; Sampson, D. A.; Weinmann, S.; Kallakury, B. V. S.; Berry, D. L.; Haque, R.; Van Den Eeden, S. K.; Sharma, S.; Bearss, J.; Beer, T. M.; Thomas, G. V.; Heiser, L. M.; Alumkal, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (18), E4179–E4188.

(44) Chen, J.; Ding, J.; Wang, Z.; Zhu, J.; Wang, X.; Du, J. *Oncotarget* **2017**, *8* (12), 19609–19630.

(45) Tunovic, S.; Barkovich, J.; Sherr, E. H.; Slavotinek, A. M. *Am. J. Med. Genet. Part A* **2014**, *164* (7), 1744–1749.

(46) Ouyang, J.; Shi, Y.; Valin, A.; Xuan, Y.; Gill, G. *Mol. Cell* **2009**, *34* (2), 145–154.

(47) Dhall, A.; Weller, C. E.; Chu, A.; Shelton, P. M. M.; Chatterjee, C. *ACS Chem. Biol.* **2017**, *12* (9), 2275–2280.

(48) Forneris, F.; Binda, C.; Adamo, A.; Battaglioli, E.; Mattevi, A. *J. Biol. Chem.* **2007**, *282* (28), 20070–20074.

(49) Weller, C. E.; Huang, W.; Chatterjee, C. *Chembiochem* **2014**, *15* (9), 1263–1267.

(50) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. *Cell* **2007**, *128* (5), 1003–1012.

(51) Pamela N. Dyer; Raji S. Edayathumangalam; Cindy L. White; Yunhe Bao; Srinivas Chakravarthy; Uma M. Muthurajan; Karolin Luger. In *Methods in Enzymology*; 2004; Vol. 375, pp 23–44.

(52) Kim, S. A.; Chatterjee, N.; Jennings, M. J.; Bartholomew, B.; Tan, S. *Nucleic Acids Res.* **2015**, *43* (10), 4868–4880.

(53) Doren, M. O. P. S. R. Van. *Anal. Biochem.* **2007**, *371* (1), 43–51.

(54) Lefrançois, P.; Vajrala, V. S. R.; Arredondo, I. B.; Goudeau, B.; Doneux, T.; Bouffier, L.; Arbault, S. *Phys. Chem. Chem. Phys.* **2016**, *18* (37), 25817–25822.

(55) Summers, F. A.; Zhao, B.; Ganini, D.; Mason, R. P. *Methods Enzymol.* **2013**, *526*, 1–17.

(56) Debski, D.; Smulik, R.; Zielonka, J.; Michałowski, B.; Jakubowska, M.; Debowska, K.; Adamus, J.; Marcinek, A.; Kalyanaraman, B.; Sikora, A. *Free Radic. Biol. Med.* **2016**, *95*, 323–332.

(57) Pronobis, M. I.; Deuitch, N.; Peifer, M. *PLoS One* **2016**, *11* (8), 1–12.