Deciphering the biochemical crosstalk of histone sumoylation in human chromatin

Calvin Jon Antolin Leonen

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Champak Chatterjee, Chair

Pradip Rathod

Dustin J. Maly

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Deciphering the biochemical crosstalk of histone sumoylation in human chromatin

Calvin Jon Antolin Leonen

Chair of the Supervisory Committee:

Dr. Champak Chatterjee

Chemistry

The histone code hypothesis states that posttranslational modifications (PTMs) of amino acids at the histone protein termini can combinatorially regulate key DNA processes such as replication, repair, and transcription. The flexible histone termini extend beyond the globular octameric core of the nucleosome core particle, composed of a tetramer of histones H3 and H4, and two dimers of histone H2A and H2B. This octameric protein core tightly wraps about ~147 bp of DNA, and helps to package the vast eukaryotic genome in the small nuclear volume. Proteins that install, remove, or bind to histone PTMs, or marks, work together to modulate the histone PTM landscape and to regulate DNA-templated processes. The biochemical relationship between histone marks is also called *crosstalk*, and this may be positive or negative, depending on if a histone mark promotes the installation or removal of subsequent marks, respectively. The dysregulation of these proteins and the histone PTM landscape can have deleterious effects on human health, including cancer development and neurodegenerative diseases. Therefore, understanding the

biochemical mechanisms underlying the histone code will serve to improve our understanding of how genomic processes are regulated to maintain cellular homeostasis.

One histone mark that is poorly studied in cells is histone sumoylation. It was first reported in 2003 as a modification of H4 in human cells and associated with transcriptional repression. The Chatterjee lab has found that H4 Lys12 sumoylation (H4K12su) stimulates demethylase and deacetylase activities within the transcriptionally repressive LSD1-CoREST1-HDAC1 complex. Key to this stimulation is the non-canonical sumo-interacting motif (SIM) within CoREST1. I have characterized the SUMO-CoREST1 SIM interaction using two-dimensional NMR and found that despite differing in sequence from typical SIMs, the CoREST1 SIM binds to the same cleft on SUMO that all SIMs have been observed to bind. Importantly, mutation of the hydrophobic CoREST1 SIM residues to alanine resulted in significant inhibition of binding to SUMO in NMR. We are currently using NMR to investigate the effect of mutations within the CoREST1 SIM that have been observed in somatic cancers.

To directly address the function of H4K12su in transcription, I investigated SUMO's biochemical crosstalk with histone acetylation, a histone mark associated with and critical for active gene transcription. Using *in vitro* histone acetyltransferase assays and mass spectrometry I found that H4K12su inhibits p300 acetyltransferase activity on H4 in octamers, nucleosomes, chromatinized plasmids, and in cells. Chromatin incorporating H4K12su was also a poor template for in vitro transcription in human nuclear extracts. Additionally, I discovered a new negative biochemical crosstalk between H4K12su and transcriptionally relevant H3 Lys4 methylation by the COMPASS/Set1 methyltransferase *in vitro*. Cellular experiments confirmed this negative crosstalk between the transcription start and end sites of annotated protein-coding human genes. Collectively, my thesis work has revealed that H4K12su directly inhibits transcription, and may do so in part by the inhibition of histone acetyl- and methyltransferases.

Lastly, many mass spectrometric studies have identified sites of sumoylation by SUMO2/3 on human histones but none of them have been well biochemically validated. I have therefore sought to demonstrate histone sumoylation *in vivo* on both H4 and H2B, that latter which has not been previously shown in the literature. Incorporation of a trypsin cut site at the *C*-terminus of SUMO will allow us to discretely identify sites of sumoylation and validate them in cellular assays.

Acknowledgements

I have often described my time in graduate school as long days and fast weeks. Now that I am at the end of it, I can also say for certain that these last five years have flown by quicker than expected. And amidst the whirlwind of all the experimenting, learning, teaching, laughing, and crying, I now find myself as a turtle perched on a fencepost reflecting on all that has happened, as I did not get here all on my own. Here I hope I can express my gratitude for all the folks who have helped me along the way to find myself in this strange and wonderful new position of holding a research PhD and writing an acknowledgements section for the work I have done.

To my parents Celso and Emerie Leonen, and all my siblings and family, you have always supported me in everything I do. I can never thank you all enough nor show all the gratitude owed to you all for your love, acceptance, and care of me. I love you all so much and I hope I can continue to make you all proud. You all constantly inspire me to work hard and love harder every day. You all have had a hand in making this PhD, so we share in this achievement together.

I would like to especially thank my graduate research advisor Champak. You have pushed and challenged me to think more deeply about the questions I ask and seek to answer and about the experiments I conduct to address them. I appreciate immensely your understanding of my learning and research style and giving me the space to fail on my own with the grace of continuous support and direction when I cannot figure something out. Your belief in me as a scientist has encouraged me and given me the confidence that has kept me going through these years and formed into the researcher I am today.

Thank you to my lab mates past and present. You all have been such a joy and privilege to work

with and learn from. Our discussions of science have been invaluable in research, and the camaraderie we have built is sincerely motivating. I have also had the benefit of working with amazing collaborators here at UW and abroad. It was such a treat to engage in so many diverse projects and learn from each other during the process.

To my friends that I have made all along the way, I also owe a debt of thanks. We have celebrated and commiserated together during the highs and lows of life and graduate school. I could not have been successful in this whole process without your company and all the joyous experiences we had. I hope to continue to create more memories with you all and keep connected as we take the next steps in our lives and careers. You have all truly made me a better person today.

I would like to thank all my past and present teachers and advisors, formal and informal, from elementary school to now. I have been fortunate to have had an amazing and well-rounded education, fostered by teachers that have believed in me and my potential, even when I did not know it myself. You all have instilled in me a love for learning and the confidence to advance and be successful in my education from high school to college to graduate school.

Lastly, I would like to acknowledge that all my education and research here in Washington has been done on the traditional land of the first people of Seattle, the Duwamish People, past and present, and of the greater Coast Salish Peoples. I thank and honor with gratitude the land itself and the Duwamish Tribe.

I am disoriented sometimes thinking about how far I have come: from my family and community in Guam to the home away from home I have built here in Seattle. I am consoled and grounded by your love and support. Only then can I know my trajectory, by first looking back and remembering the initial forces that have put me in motion.

In dedication to my parents Celso and Emerie Leonen

Today is Monday
Monday, string beans

—Eric Carle

Sed omnia praeclara tam difficilia quam rara sunt

(But everything great is just as difficult to realize as it is rare to find)

—Viktor E. Frankl

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction to epigenetics and biochemical crosstalk in chromatin

1.1 Epigenetics

Understanding the molecular underpinnings of genome maintenance and regulation has been an extensive area of research over several decades, due to the intricate biochemical mechanisms that have evolved to both package and enforce an organism's genetic information. Classical genetics has significantly aided our understanding of how the instructions for life are encoded within double-stranded DNA, how this DNA is replicated and faithfully inherited from a cell to its progeny, how DNA is transcribed into RNA, and how RNA can ultimately be translated into protein. Genetics clearly explains the encoding and decoding of the information contained within genomic DNA, which remains largely fixed during an organism's lifetime. However, purely genetics alone have not been sufficient in explaining variations in expression from an organism's genotype and differences in the observed phenotype. Moreover, how a cell regulates which genes are expressed, repressed, or permanently silenced in development; in response to external stimuli or environmental factors; or in cellular transformation is incompletely explained by genetics alone. This suggested the existence of an additional, complex layer of genetic regulation, which is now its own field of study called epigenetics. Conrad Waddington in 1942 fist coined the term epigenetics, appending the prefix *epi-*, meaning "above" or "in addition to," onto genetics in order to categorize phenomena that could not be fully explained by genetics alone.[1,2] Examples of such phenomena include the observable differences in identical twins, who have identical genotypes;

position-effect variegation in *Drosophila;* and in the coat patterning in calico cats. Epigenetics is now a burgeoning field of research that seeks to uncover cellular mechanisms that regulate dynamic genome functions, including differentiation, development, responses to external stimuli and the environment, and maintenance of homeostasis from the largely static DNA sequence. One area of epigenetics involves the regulation of DNA-templated processes through the reversible post-translational modifications (PTMs) that occur on the histone proteins, which intimately bind and package DNA. In this thesis, I will present my research on uncovering how the small ubiquitin-like modifier (SUMO) protein and other PTMs of lysine residues in the histone *N*-termini, or *tails*, engage in biochemical *crosstalk* —the functional relationship between histone PTMs —to collectively regulate gene expression.

1.2 Chromatin

Chromatin is a massive and dynamic nucleoprotein complex that stores vast amounts of genetic material, about 3 billion base pairs of DNA in humans, within the minute cell nucleus. In 1974, Roger Kornberg was one of the first to report on the fundamental building block of chromatin as the nucleosome core particle (NCP).[3] Each NCP consists of ~147 bp of double-stranded DNA wrapped about 1.6 turns around a cylindrical protein spool consisting of equimolar amounts of the core histones H2A, H2B, H3, and H4.[3,4] Soon thereafter in 1995 one of the first X-ray crystal structures of the histone octamer was published, revealing an H3-H4 tetramer core bound by two dimers of H2A-H2B in a disk.[5] Initial micrococcal nuclease digestions of chromatin further corroborated the NCP as the core unit, generating an approximately 146 bp strand of dsDNA associated with two copies of all four histones.[6] Karolin Luger and co-workers in 1997 obtained the first high-resolution crystal structure of the entire NCP, which was a seminal finding toward our understanding of the histone-histone and histone-DNA interactions within chromatin.[7] The linker DNA between each NCP can vary from 20-90 bp, depending on the regulated spacing of

nucleosomes by various motor proteins, and may be specifically positioned to occlude, or alternately, make accessible the underlying DNA sequence to DNA-binding proteins such as transcription factors.[8,9] Some of the first electron micrographs of chromatin showed extended structures of “particles on a string” corresponding to nucleosomes more or less evenly spaced along the DNA.[10] This structural form of chromatin is now known as “open” or euchromatin (**Figure 1.1**). On the other hand, other regions of chromatin were found to be highly condensed and retained dye-staining throughout the cell cycle. This condensed region of chromatin was termed heterochromatin by Emil Heitz, and is largely associated with regions of DNA that are poorly expressed or silenced.[11,12]

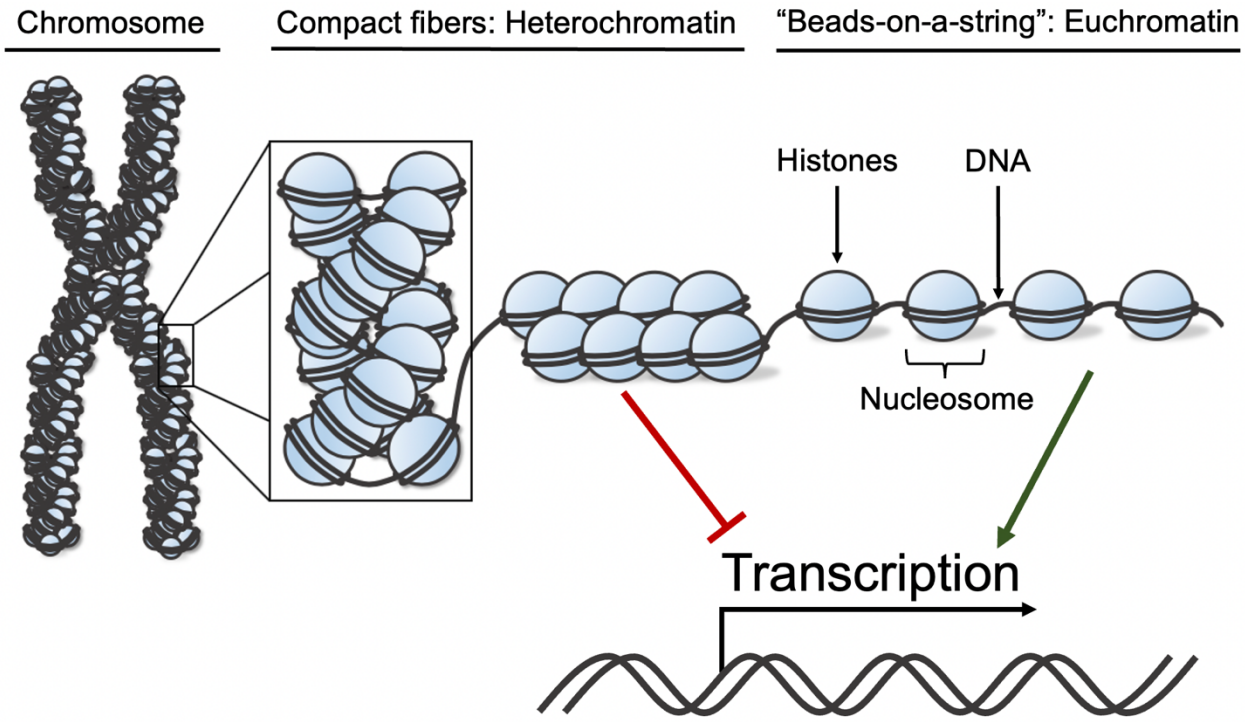


Figure 1.1. Cartoon diagram of chromatin. Chromatin is composed of DNA and histone proteins that form discrete nucleosome core particles. Condensed and transcriptionally silent chromatin is referred to as heterochromatin, whereas “open,” transcriptionally active chromatin is referred to as euchromatin.

While Emil’s initial hypothesis was that heterochromatin contained little to no protein-coding genes, we now know that heterochromatin is largely transcriptionally inactive, but, it does contain

genes; typically corresponding to developmental genes, and other transcribed sequences that produce regulatory RNAs like long non-coding RNAs and piwi-interacting RNAs.[13] Indeed, heterochromatin can be categorized as constitutive or facultative heterochromatin. Constitutive heterochromatin is comprised of regions of chromatin that are consistently permanently silenced, whereas facultative heterochromatin is temporarily repressed and can be activated into the euchromatin form. The observation of such genome organization and advances in chromatin research technology have led to the identification of topologically associated domains, or TADs, that delimit chromosome territories, compartments of gene-rich regions, and regulatory elements, among others.[14–16]

Early biochemical studies demonstrated that the histones were repressive towards transcription by RNA polymerases.[17] Studies in yeast where histones were depleted resulted in the activation and increased transcription of previously inactive genes.[18] Histones bound to DNA were found to induce transcriptional conflicts and prevented transcription elongation.[17,19,20] This presented an interesting conundrum in the histones' role in chromatin regulation and enabling DNA-templated processes like transcription. Toward this, Vincent Allfrey's early work with chemical histone acetylation suggested a possible role for histone acetylation in facilitating RNA synthesis in cellular extracts.[17] This dynamic interplay of histone modifications was hypothesized by Allfrey and colleagues as a potential molecular switch to turn gene expression on and off.[21]

An interesting phenotypic observation in Drosophila called position-effected variegation (PEV) highlighted chromatin and the histones' key role in transcription and gene regulation.[22,23] The protein Su(var)3-9 (suppressor of variegation, chromosome 3, 9th gene) was found to be responsible for PEV in the eye color of *D. melanogaster.* When the *white* gene that encodes an ABC transporter responsible for carrying red/brown pigments to *Drosophila* eyes was placed near heterochromatic regions, it resulted in a phenotype of white patches, or variegation, within the

typically red-colored eyes of the fruit fly. Su(var)3-9 was later identified to be a histone H3 Lys9 methyltransferase, a histone modifying enzyme that results in lysine di- and tri-methylation PTMs that recruit heterochromatin protein 1 (HP1) to silence genes and facilitate heterochromatin formation.[24] This effect was found to be part of the underlying mechanism of heterochromatin spreading, resulting in gene silencing. Thus, modification of the histones by small chemical PTMs was found to influence the expression of the underlying bound DNA sequence without any changes to the gene itself.

1.3 Histone post-translational modifications

Extending outwards from the NCP, and hence exposed to the nucleosol, are the functional group rich histone *N*-termini or *tails*. Histone tails are the sites of a wide range of chemically distinct and reversible post-translational modifications. These include varying degrees of arginine and lysine side-chain methylation (mono-, di-, and trimethylation), serine, threonine, or tyrosine phosphorylation, and various forms of lysine acylation. The latter ranges from modification by smaller acetyl to longer crotonyl and butyryl groups, and even entire proteins such as ubiquitin (Ub) and SUMO (small ubiquitin-like modifier) (**Figure 1.2**).[25–27] Histone PTMs or *marks* seldom exist in isolation and the diversity of modified histones revealed in numerous proteomic studies is truly staggering.[28] Despite the complex landscape of chromatin modifications, specific sets of marks may be correlated with distinct transcriptional states of their associated genes. This led to the hypothesis by David Allis and colleagues that combinatorial patterns of histone marks may constitute a *histone code* that regulates key DNA-templated processes such as transcription, replication, and damage repair.[29]

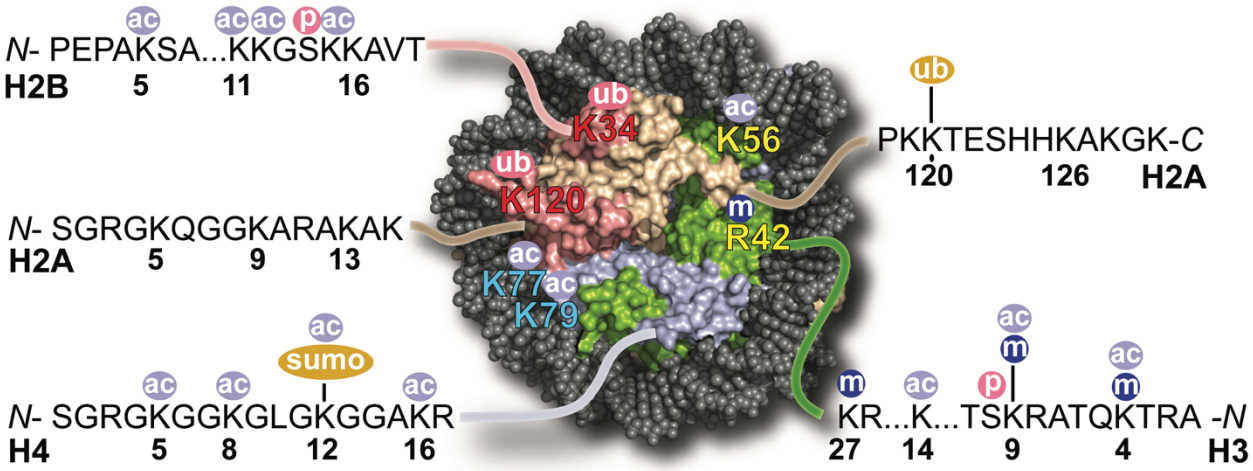


Figure 1.2. Histone tail marks. Schematic representation of the diversity of marks that have been accessed by histone semisynthesis. Chemical groups are indicated as ac= acetyl, m= methyl, p= phosphoryl, sumo= small ubiquitin-like modifier, ub= ubiquitin. The globular core of the nucleosome core particle is shown with histones colored as H2A (gold), H2B (red), H3 (green) and H4 (blue) and double-stranded DNA (gray). PDB code 1KX5.

1.4 Mechanisms of chromatin regulation by histone marks

Histone marks may influence chromatin structure and function by two mutually non-exclusive mechanisms. In some instances, a mark may directly change local chromatin structure, facilitating or denying access to numerous chromatin-modifying enzymes. Secondly, a mark may serve to recruit chromatin-associated proteins that deposit (*writers*), remove (*erasers*), or bind (*readers*) specific sets of marks.[30] Indeed, a body of literature exists for certain privileged histone marks, such as H4 Lys16 acetylation (H4K16ac), which is associated with actively transcribed genes[31] and the open euchromatin structure.[32] The absence of H4K16ac in chromatin is strongly associated with transcriptional repression, heterochromatin formation, and the appearance of repressive marks such as H3K9me3[33,34] and H3K27me3.[35,36] Further complexity in the writing and execution of the histone code arises from the crosstalk between marks, whereby one mark may lead to the addition or removal of others on the same histone or on a different histone.[37]

Beyond histone-centric crosstalk alone are the relationships between marks and DNA modifying enzymes[38] or long non-coding RNAs[39], which are also important regulators of cellular outcomes.

In addition to various histone PTMs, the histones themselves are expressed in various isoforms that are specifically recruited to and deposited in chromatin. One example is the centromeric histone H3, CENP-A, which is found within the centromeric region during cell division.[9] CENP-A demarcates the centromere and assists in binding with microtubules within the kinetochore during cell division to facilitate the equal distribution of chromosomes between the parent and offspring cells.[40] Thus, the combination of histone variants and their PTMs substantially increases the combinatorial power of the histone code to specifically regulate gene functions.

1.5 Histone semisynthesis to unravel a complex code

The empirical association between histone marks and gene function at different loci is a critical and important step toward understanding chromatin regulation, but it provides little clarity at the molecular level. Cellular signaling is inherently complex and immediate changes in chromatin prior to or following the installation of specific marks are not readily captured in cell-based experiments. Elucidating the precise biochemical crosstalk between sets of marks in chromatin is particularly challenging due to their complex pattern, varying abundance and dynamic nature in cells. Indeed, a molecular understanding of the roles for specific marks requires uniformly and site-specifically modified NCPs, in order to establish a direct line of causality between a mark and its functions. Therefore, the last decade has seen the development of a large number of chemical biology techniques to access homogeneously and site-specifically modified histones.[41] Protein semisynthesis is one early and key technique that has enabled detailed mechanistic studies of crosstalk in well-defined NCP substrates.[42]

The enzymatic modification of recombinant histones is a valuable strategy for generating substrates used in mechanistic studies, however, it requires knowledge of the site-specific writer and its isolation in an active form. Semisynthesis has several advantages including (1) the ease of scalability, (2) chemospecificity, and (3) the ability to install multiple different marks in close proximity to each other. The latter may be particularly challenging when the natural order of histone modification is unknown, leading to the inhibition of a desired enzymatic activity by partially modified histones.

Native chemical ligation (NCL) is a widely utilized semisynthetic strategy that overcomes inherently limited yields from the solid-phase peptide synthesis of proteins longer than ~50 amino acids, such as the histones.[43] NCL permits chemoselective amide bond formation between two polypeptide fragments, by the incorporation of a C-terminal α-thioester in one fragment and N-terminal Cys residue in the other (**Figure 1.3**). An initial reaction between the fragments joins them by a thioester linkage, which spontaneously rearranges to the native backbone amide bond at neutral-to-alkaline pH. The fragment with N-terminal Cys may be synthetic or recombinant in nature depending upon its desired PTM state. Expressed protein ligation (EPL) further extends the scope of NCL by employing an intein, a single-turnover enzyme that undertakes protein splicing in bacteria, to generate the C-terminal α-thioester.[44] Thus, either N- or C-terminally modified histone tails are readily accessible by protein semisynthesis employing two asymmetrical fragments, with the smaller synthetically accessible fragment containing any desired marks (**Figure 1.3**). The total chemical synthesis of histones by successive NCL steps has also rendered marks in the interior of histones, such as acetylation at H3K56[45] and phosphorylation at H2AY57, H3Y41 and H3T45, accessible in good quantities for mechanistic studies.[46]

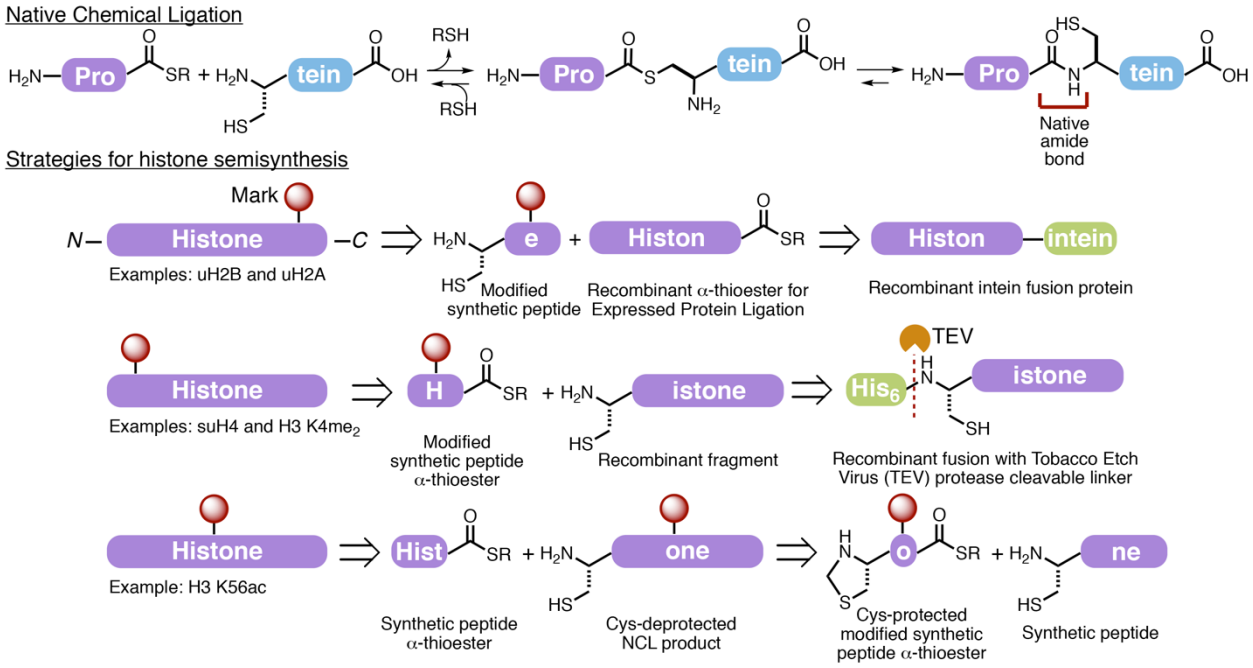


Figure 1.3. Native chemical ligation and strategies for histone semisynthesis. Native chemical ligation leads to the formation of a native amide bond between two polypeptide fragments (top). Different retrosynthetic strategies for the semisynthesis of a modified histone based on the location of the mark near the termini or histone core domain (bottom).

1.6 Understanding ubiquitin-signaling in chromatin

Ubiquitin (Ub) is best known as a molecular zip-code that destines proteins for degradation by the 26*S* proteasome.[47] However, when conjugated with histones it is associated with several non-proteolytic roles including gene silencing, DNA-repair and transcription elongation.[48] All four core histones as well as histone variants and the linker histone H1 are ubiquitylated, but the precise role of Ub in these varying contexts is unclear. This presents an excellent opportunity for the application of chemical techniques to obtain site-specifically ubiquitylated histones for mechanistic studies. Unlike smaller histone marks such as acetyl and methyl groups, however, the installation of Ub on a Lys side-chain during SPPS is extremely challenging. This has inspired the development of several NCL auxiliaries – Cys mimetic amino acid derivatives that first enable NCL of Ub-α-thioesters to the target Lys side-chain and are subsequently removed by strong

acids, photolysis or reduction, to yield the native isopeptide-linked ubiquitylated histone (**Figure 1.4**).[49–54] More recently, chemical analogs of the isopeptide linkage including disulfide[55], thioether[56], or triazole[57] linkages have proven to be useful mimics in biochemical studies of Ub signaling. The obvious limitation when using such analogs is that they must first be compared with the native ubiquitylated histone to ensure that no significant differences arise from the non-natural linkage. Collectively, these semisynthetic efforts have enabled studies of biochemical crosstalk involving histones modified by Ub and its family-members, and shed light on key elements of chromatin regulation.

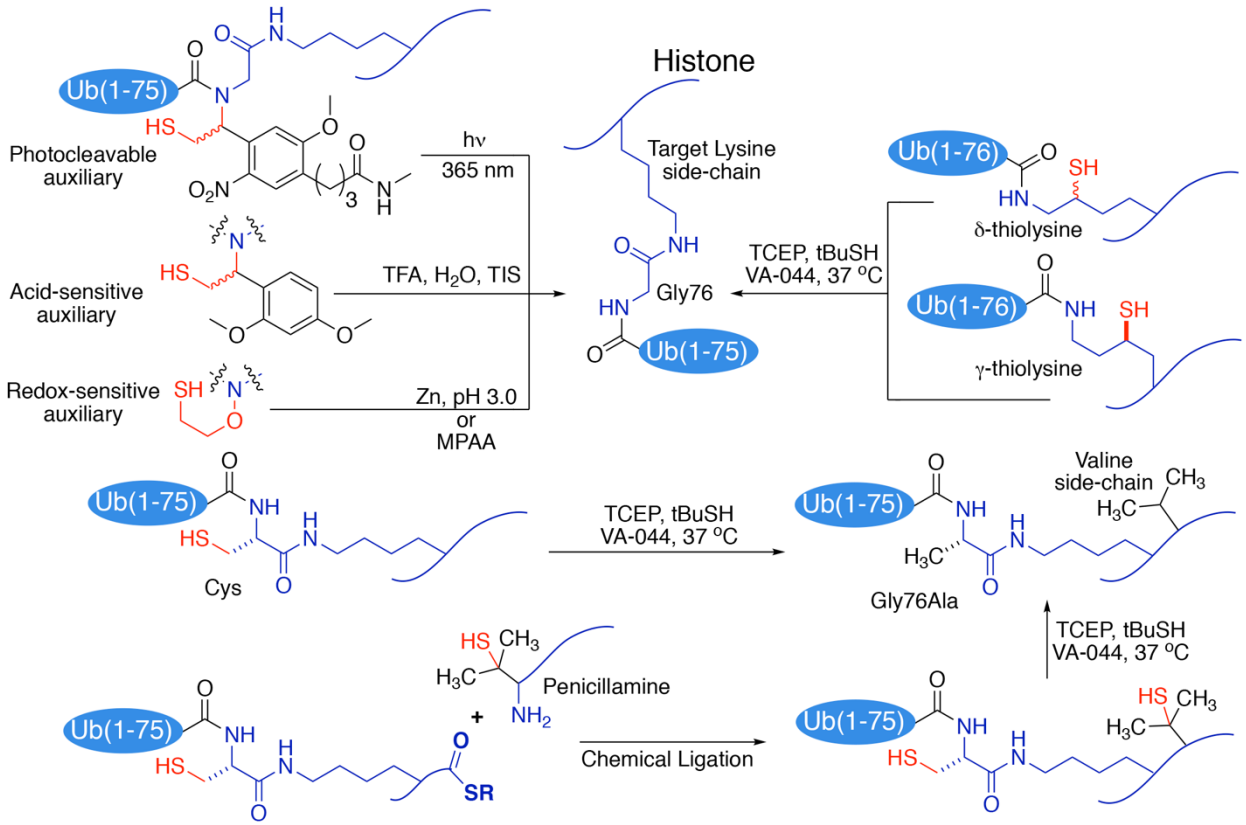


Figure 1.4. Cysteine and cysteine-mimetic amino acid derivatives used in histone semisynthesis. Cysteine and its mimetics employed in native chemical ligation are related by the 1,2- or 1,3-position of the thiol and amine groups, shown by red bonds. Each derivative is converted to the indicated amino acid in a single step to yield the wild-type modified histone. TCEP= tris(2-carboxyethyl)phosphine, tBuSH= *tert*-butanethiol, VA-044= 2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride, TFA= trifluoroacetic acid, TIS= triisopropylsilane, MPAA= 4-mercaptophenylacetic acid, Ub= ubiquitin.

1.7 Biochemical crosstalk between uH2B and H3 K79me2

One of the earliest known targets of modification by Ub are H2B[58] and H2A[59]. Unlike the multimeric K48-linked Ub chains associated with proteasomal targeting of protein, Ub conjugated with histones exists mostly in the monomeric form. Interestingly, ubiquitylation is associated with at least two opposing roles in chromatin. In the form of uH2B (at K120), it marks active genes as the RNA polymerase II transcription complex passes through[60], and uH2A (at K119) marks inactive gene promoters[61]. In the first example of biochemical *in trans* crosstalk between two modifications, Allis and Sun demonstrated that ubiquitylation of K123 in yeast H2B (K120 in humans) is a prerequisite for the installation of trimethylation at H3K4 by the Set1 (SET domain containing protein 1) methyltransferase.[62] Later, uH2B was also demonstrated to be required and sufficient for efficient dimethylation at H3K79 by the non-SET domain histone methyltransferase Dot1L (Disruptor of telomeric silencing 1-like).[63] However, the precise mechanistic role of uH2B in Dot1L function remained unknown, due to the low abundance of uH2B in cells (< 5%) and the inability to purify nucleosomes uniformly modified by this mark. Therefore, the first report of an NCL strategy to site-specifically ubiquitylate peptides at the Lys side-chain ε-amine enabled mechanistic studies of the crosstalk between uH2B and H3K79me2.[49] By employing semisynthetic uH2B reconstituted in NCPs, McGinty et al. demonstrated that the presence of Ub attached to the nucleosome at H2BK120 was necessary and sufficient to stimulate Dot1L activity (**Figure 1.5**).[64] Moreover, Dot1L stimulation was strictly intranucleosomal in nature and the extent of H3K79me2 corresponded 1:1 with the amount of uH2B in a nucleosome. Subsequent studies with a disulfide-linked analog of uH2B ($uH2B_{ss}$) revealed a degree of plasticity in the precise site of Ub attachment on H2B while retaining Dot1L stimulation, and established that spatial positioning of Ub rather than the precise histone target is key to its role.[55] Ala-mutagenesis studies of Ub surface residues revealed the dependence of Dot1L stimulation on an unexpected epitope centered on Leu71/Leu73 near the C-terminus of Ub.[65] This epitope differs significantly from the

canonical hydrophobic patch consisting of Leu8/Ile44/Val70 associated with most known functions of Ub. Further studies by Zhou et al. have suggested that Ub binding to the H2A tail may corral Dot1L to a more productive binding mode on nucleosomes.[66] This was demonstrated by the incorporation of the photocrosslinking amino acid photoleucine in place of Leu71 in Ub, and its subsequent crosslinking to H2A. The mechanistic implications of the Ub-H2A interaction and precisely how it accomplishes the proposed corralling effect await further interrogation.

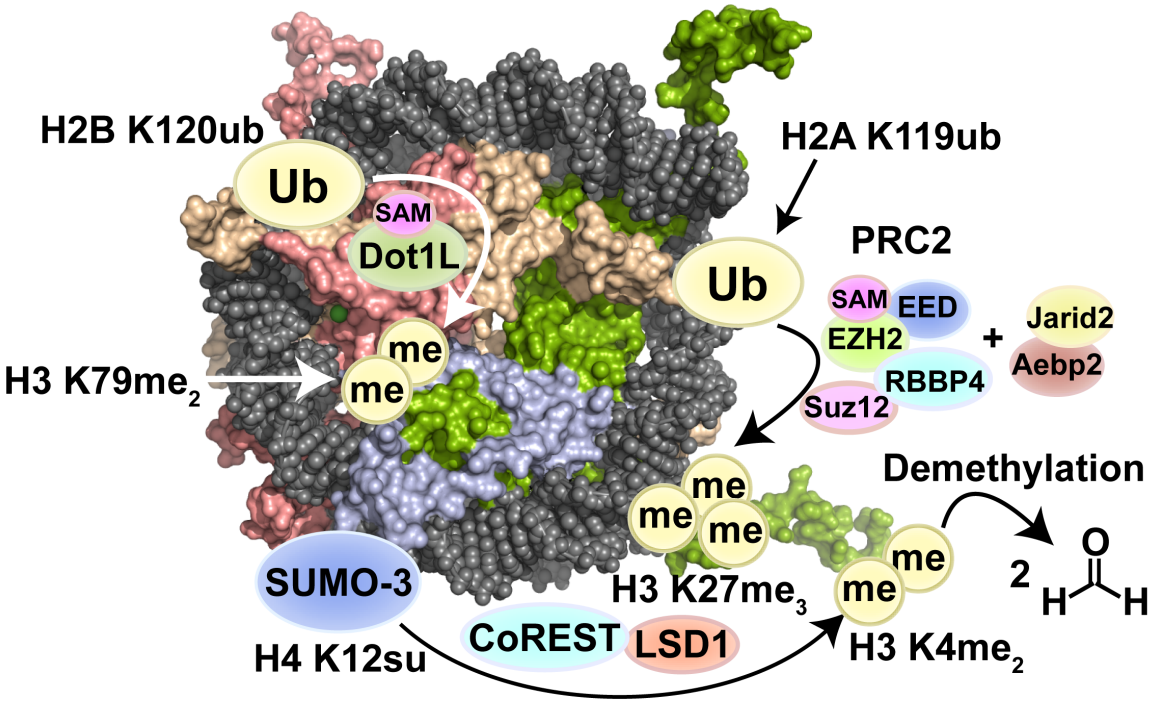


Figure 1.5. Biochemical crosstalk between histone marks in the nucleosome core particle. Arrows indicate the known direction of biochemical crosstalk between Ubiquitin (Ub), or the small ubiquitin-like modifier protein (SUMO) and methyl groups (me) at the indicated histone residues. Proteins mediating crosstalk are the Disruptor of telomeric silencing 1-like (Dot1L), the Lysine-Specific Demethylase 1-Corepressor of REST (LSD1-CoREST) complex and core subunits of the Polycomb Repressor Complex 2 (PRC2). These are the Suppressor of Zeste 12 protein homolog (Suz12), Enhancer of Zeste Homolog 2 protein (EZH2), Retinoblastoma-Binding Protein 4 (RBBP4) and Embryonic Ectoderm Development protein (EED). Jumonji And AT-Rich Interaction Domain Containing 2 (Jarid2) and Adipocyte Enhancer-Binding Protein 2 (Aebp2) are two key proteins that associate with PRC2 to mediate positive crosstalk between uH2A and H3K27me3. SAM= S-Adenosylmethionine.

1.8 Biochemical crosstalk between uH2A and H3K27me3

Ubiquitylation at H2AK119 (uH2A) is found in ~10% of all eukaryotic H2A[67], and it is a much more abundant modification than uH2B[62]. The higher abundance of uH2A led to its successful isolation in early efforts. Consistent with its occurrence in silenced genes, biophysical assays with linear nucleosome arrays reconstituted with cellular uH2A revealed the formation of compact structures associated with heterochromatin.[68] This contrasts with uH2B, which inhibits the compaction and higher order oligomerization of 12-mer nucleosome arrays.[69]

H2A is ubiquitylated by the E3 ligases Ring1 and Ring1B (or RNF2) that are components of the Polycomb repressive complex 1 (PRC1).[61] PRC1 plays critical roles in Polycomb silencing of developmentally critical *HOX* genes and X-chromosome inactivation. It is also implicated in crosstalk with the H3K27 methylating Polycomb repressive complex 2 (PRC2) whereby the two PRCs reinforce each other's activities.[70,71] With the established crosstalk between uH2B and H3K79me2 in mind, Whitcomb et al. employed semisynthetic wild-type uH2A to test crosstalk with H3K27me3 mediated by core components of the PRC2 complex.[72] Surprisingly, they discovered a small but significant inhibitory effect of uH2A on H3K27 trimethylation by PRC2 *in vitro*, which could not be readily explained. In a subsequent study, two additional protein components Jarid2–Aebp2 (Jumonji And AT-Rich Interaction Domain Containing 2-Adipocyte Enhancer-Binding Protein 2) were found to bind PRC2 and stimulate its activity ~25-fold on nucleosomes containing uH2A (**Figure 1.5**).[73] This confirmed the initially proposed crosstalk between PRC1 and PRC2, and indicated that its associated proteins modulate PRC2 function.

Recently, Liu and coworkers also prepared semisynthetic uH2A by the selective conjugation of a truncated Ub(1-75)-α-thioester to a Gly with N(α)-auxiliary conjugated at K119 in H2A.[54] In a stepwise ligation strategy similar to the genetically encoded orthogonal protection and activated

ligation approach pioneered by Virdee et al.[74], the authors prepared a fully amine-protected form of H2A and then generated Lys119 by the reduction of azidonorleucine incorporated at position 119 in response to a Met codon. The high yield of azidonorleucine incorporation in H2A with a mutant methionyl tRNA synthetase was a highlight of this approach. The uH2A was subsequently incorporated in nucleosomes and tested for crosstalk with various methylated states of H3K36, which mark active genes. Consistent with its role as a repressive mark, uH2A strongly inhibited the H3K36 di- and trimethylases, NSD2 (Nuclear SET Domain-Containing Protein 2) and SETD2 (SET Domain Containing 2), respectively. This confirmed the negative crosstalk between uH2A and H3K36me2/3 that was first proposed by Zhu and coworkers using oligonucleosomes assembled with histones isolated from HeLa cells.[75]

1.9 Chasing the high-hanging fruit: biochemical effects of histone sumoylation

SUMO is the small ubiquitin-like modifier and is part of the family of ubiquitin and ubiquitin-like (Ubl) proteins. There are three major isoforms of SUMO in humans, SUMO-1, SUMO-2, and SUMO-3. SUMO-2 and SUMO-3 share approximately 97 % sequence identity, have significant substrate overlap, and are typically referred to as SUMO-2/3. SUMO-2/3 are only about 50% identical to SUMO-1 in sequence, and their effects on substrates may differ widely. Sumoylation adds significant mass, surface area, and overall steric bulk to the proteins it modifies, forming an isopeptide bond at the epsilon-amine of lysine residues. Like ubiquitin, SUMO-2/3 can also form chains, however SUMO-2/3 chain structures and functions are not as well characterized as ubiquitin chains.

SUMO is evolutionarily conserved from parasites, yeast, plants, to humans, however there is only a single yeast SUMO ortholog, Smt3. SUMO only shares ~18% sequence identity with ubiquitin.[76]

Like ubiquitin and other Ubls, SUMO is initially translated as a pro-peptide with a *C*-terminal extension of several amino acids. SUMO proteases, or SUMO endopeptidases (SENPs, also sentrin proteases), must first cleave the propeptide after a di-glycine motif that is characteristic to Ub/Ubl proteins. There are at least six SENPs in humans whereas yeast only have two SUMO proteases, Ulp1 and Ulp2. In addition to the di-glycine motif, all Ub/Ubls share a distinctive β-grasp fold structural motif that features a five-stranded, mixed β-sheet that wraps around a cental α-helix.[77] After SUMO is cleaved to its mature form, it must undergo an E1, E2, and E3 enzymatic cascade, typical of the Ub/Ubl family, to be appended to its substrates.[78] First the heterodimeric SUMO-activating enzyme subunits 1 and 2 (SAE1/SAE2) complex (Uba2/Aos1 in yeast) utilizes ATP as a co-factor to convert the SUMO carboxy terminus into a highly reactive adenylated intermediate and capture as a *C*-terminal thioester adduct to the E1 via a cysteine resiude.[79,80] There is only one known SUMO E2, or SUMO-conjugating enzyme in humans, UBC9, which significantly differs from the multitude of E2's for ubiquitin. SUMO is transferred to UBC9 via a transthioesterification reaction with the SUMO-E1 adduct.[81] UBC9 can then transfer SUMO onto a substrate lysine residue, sometimes with the added specificity from an E3 ligase.[78] An E3 ligase recruits the UBC9~SUMO adduct to catalyze the transfer SUMO onto specific substrates. However, an E3 is not critical for SUMO conjugation to a substrate lysine. Many sites of sumoylation have also been identified to follow a consensus sequence of a ΨKXE amino acid motif, where Ψ is a large hydrophobic residue, X is any amino acid, and K is the lysine that is sumoylated.[82] This sumoylation consensus motif, however, has also been observed to not be absolutely necessary for substrate sumoylation. This is exemplified in the case of histone H4 sumoylation, where none of the observed sites of sumoylation are within a typical ΨKXE consensus motif.

Several decades after the discovery of histone ubiquitylation, Eisenman and Shiio reported the modification of the histone H4 tail by two distinct isoforms of the small ubiquitin-related modifier protein, SUMO-1 and SUMO-3, in human HEK293T and P493-6B cells.[83] More than a decade's efforts in understanding the roles of histone sumoylation in chromatin have produced varying hypotheses – from transcription repression to even proposed roles in transcription activation.[84] In the absence of a known histone-specific SUMO ligase, Shiio and Eisenman demonstrated that recruitment of the ΨKXE target-motif-specific E2 ligase, Ubc9, to the promoter region of a luciferase reporter plasmid sufficed to repress transcription and correlated with reduced histone acetylation. In a subsequent study, Berger and coworkers demonstrated similar gene-repressive effects of Smt3 in Baker's yeast on gene transcription.[85] Despite these pioneering inroads, little remained known regarding the specific role of SUMO site-specifically appended to H4, or indeed, its crosstalk with marks of active transcription, such as histone acetylation and methylation. Challenges toward investigating histone sumoylation parallel those of ubiquitylation and are additionally complicated by the presence of multiple SUMO isoforms, multiple sites of modification, and the lack of sumoylated H4 (suH4) specific antibodies.

Based on early mass spectrometric evidence of K12 in histone H4 being a site of modification by SUMO-3 in human cells[86,87], Dhall et al. interrogated the direct effect of histone sumoylation on chromatin structure. By employing a disulfide-linked analog of suH4 ($suH4_{ss}$) incorporated in 12-mer nucleosome arrays, they observed that sumoylation inhibits the formation of compact structures associated with heterochromatin.[88] The biophysical effect of suH4 paralleled what is known for uH2B and H4K16ac, which are associated with active genes, and was opposite to the observed effect of uH2A in chromatin decompaction. Furthermore, single molecule FRET experiments with fluorophore-labeled mononucleosomes revealed that the mechanism of inhibition by suH4 differs significantly from that of H4K16ac, in that suH4 reduces the rate at which two nucleosomes associate to form a compact dinucleosome. This surprising effect suggested

that suH4 might engage in biochemical crosstalk with other histone modifications to repress transcription, rather than directly favoring heterochromatin formation.

Many chromatin-modifying enzymes and their protein interaction partners are either modified by SUMO or have SUMO-interacting motifs (SIMs) through which they bind SUMO. Our current understanding of a SIM is a sequence with ~10 amino acids and a core of 3-4 hydrophobic amino acids such as Val or Ile, surrounded by acidic residues such as Glu or Asp, or sometimes phosphorylated Ser/Thr residues.[89] An alignment of several SIMs, however, shows little to no sequence conservation besides the hydrophobic core. The hydrophobic core can be generalized as V/I-V/I-X-V/I, where X is typically a Glu or Asp residue.[90] One critical gene repressive complex that contains a SIM is the LSD1-CoREST-HDAC1 (LCH) complex.[91] LSD1 or KDM1A, is a lysine specific histone demethylase that undertakes the demethylation of H3K4me1/2 and regulates multiple cellular pathways implicated in cellular development, differentiation, and embryonic pluripotency.[92] As a part of the LCH complex, LSD1 represses a subset of neuronal genes in non-neuronal cells in association with the RE1-silencing transcription factor (REST). Recently, it was proposed that LCH also represses REST-independent genes in a SUMO-2/3 dependent manner that requires the SIM in CoREST (the Corepressor of REST).[91] As both LCH and suH4 are associated with transcriptional repression, Dhall et al. tested their potential crosstalk by employing semisynthetic nucleosomes containing both wild-type suH4 and H3K4me2.[93] They discovered an ~2-fold stimulation of LSD1 activity by suH4, which was dependent upon the SIM in CoREST (**Figure 1.5**). This led them to suggest that suH4-mediated recruitment of the LCH complex near the H3 tail underlies enhanced demethylation. Consistent with the recruitment model, asymmetric dinucleosomes where suH4 and H3K4me2 were on adjacent nucleosomes, were not better substrates for LSD1 than methylated mononucleosomes alone. Thus, the semisynthesis of suH4 led to the discovery of the first histone modification that stimulates LSD1 activity. Further *in vivo* studies are needed to confirm the details of this exciting discovery, which highlights the power of

semisynthesis when applied toward elucidating the functional roles for histone marks that have long eluded investigation in cells.

1.10 Outlook and conclusion

The last decade witnessed an explosion in the number of biochemical studies with well-defined nucleosomal substrates that were aimed at testing proposed mechanisms of crosstalk between marks. Two key aspects of histone proteins- their facile reconstitution in nucleosomes and the presence of only a single and dispensable Cys residue (H3C110) between the four core histones- enabled numerous applications of NCL/EPL to study otherwise challenging histone marks. In conjunction with additional techniques such as amber codon suppression[94], methyllysine analogs[95], and sortase-mediated ligationp[96], semisynthesis has in principle rendered every atom in the ~210 kDa NCP accessible to chemical modification. The concurrent development of efficient strategies to tracelessly and site-specifically modify proteins by members of the Ub family has led to detailed mechanistic elucidation of the roles for these large protein modifiers in various chromatin contexts. In some cases, such as for histone sumoylation, a semisynthetic approach enabled both the generation and testing of novel hypotheses in the absence of effective molecular biological tools. We envision that as semisynthesis no longer remains the rate-limiting step toward obtaining modified histones, hypothesis-driven investigations of crosstalk between marks and unbiased screens for novel mediators of crosstalk will become routine. Instead of serving as a tool for the mechanistic elucidation of proposed crosstalk, *designer* semisynthetic nucleosomes with complex patterns of marks will be the launching pad for discovering new paradigms of biochemical crosstalk in chromatin.

Studying the roles of histone marks is crucial for our comprehension of how eukaryotic cells organize and regulate their genomes. Histone sumoylation is one such mark which has not been

extensively studied but appears consistently in proteomic studies. In this research thesis, I will describe how I utilized many of the tools described in this chapter to decipher the mechanistic role for human histone sumoylation. Our lab has demonstrated a role for H4 Lys12 sumoylation (H4K12su) in the stimulation of demethylation and deacetylation (unpublished results) of histone H3 in nucleosomes by the LSD1-CoREST1-HDAC1 complex.[93] Importantly, this stimulation relied on the SUMO interacting motif within CoREST1. Intriguingly, this CoREST SIM differs from the consensus SIM and has not been well characterized. In Chapter 2, I will explore the SUMO2/3-CoREST1 interaction with various techniques, including NMR. Based on our CoREST mutational data, I hypothesized that this CoREST1 SIM underlies the mechanism of SUMO crosstalk with the LSD1-CoREST1-HDAC1 complex. Furthermore, H4 sumoylation has been associated with transcription repression, but no study has mechanistically defined how this histone mark inhibits transcription. In Chapter 3 I interrogate the role of H4K12su in the inhibition of transcription by using well-defined reconstituted biochemical systems established by the Roeder lab at Rockefeller University. Through novel biochemical assays with lysine acetyl- and methyltransferases, I have now shown that SUMO engages in negative biochemical crosstalk with the writer enzymes that install histone acetylation and H3 Lys4 methylation. I also demonstrate for the first time, using the *in vitro* transcription system developed by the Roeder lab, that H4K12su directly inhibits gene transcription in nuclear extracts. Critical to these studies was the semisynthesis of H4K12su, which led to the reconstitution of sumoylated mononucleosomes and chromatinized plasmids for these biochemical assays. Finally, in Chapter 4 I describe my efforts to thoroughly explore the occurrence of histone sumoylation in human cells, which has not been extensively investigated before. Several reports have acknowledged histone sumoylation on all four core histones in human cells but did not provide complete data such as western blots with SUMO and histone-specific antibodies.[83,97] Moreover, some reports are contradictory in reports of preferential histone modification by SUMO-2/3 or by SUMO-1.[83,97] In this last chapter I began my experiments to elucidate histone sumoylation in human cells by utilizing the coupled

methodology of biochemical immunoprecipitation and mass spectrometry. These will enable us to identify both sumoylated histones and the specific sites of sumoylation. Such studies are critical for firmly establishing histone sumoylation as a permanent entry in the lexicon of histone PTMs.

1.11 References

Portions of this chapter have been published as:

Leonen, C.J.A.; Upadhyay, E.; Chatterjee, C. Studies of biochemical crosstalk in chromatin with semisynthetic histones. *Curr. Opin. Chem. Biol.* **2018**, *45*, 27-34. https://doi.org/10.1016/j.cbpa.2018.02.005.

(1) Waddington, C. H. The Epigenotype. *Int. J. Epidemiol.* **2012**, *41* (1), 10–13. https://doi.org/10.1093/IJE/DYR184.

(2) Goldberg, A. D.; Allis, C. D.; Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **2007**, *128* (4), 635–638. https://doi.org/10.1016/j.cell.2007.02.006.

(3) Kornberg, R. D. Chromatin Structure : A Repeating Unit of Histones and DNA. *Science (80-.).* **1974**, *184* (4139), 868–871.

(4) Finch, J. T.; Lutter, L. C.; Rhodes, D.; Brown, R. S.; Rushton, B.; Levitt, M.; Klug, A. Structure of Nucleosome Core Particles of Chromatin. *Nat. 1977 2695623* **1977**, *269* (5623), 29–36. https://doi.org/10.1038/269029a0.

(5) Arents, G.; Burlingame, R. W.; Wang, B. I. C.; Love, W. E.; Moudrianakis, E. N. The Nucleosomal Core Histone Octamer at 3.1 Å Resolution: A Tripartite Protein Assembly and a Left-Handed Superhelix. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (22), 10148–10152. https://doi.org/10.1073/pnas.88.22.10148.

(6) Noll, M.; Kornberg, R. D. Action of Micrococcal Nuclease on Chromatin and the Location of Histone H1. *J. Mol. Biol.* **1977**, *109* (3), 393–404. https://doi.org/10.1016/S0022-2836(77)80019-3.

(7) Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. *Nature* **1997**, *389* (6648), 251–260. https://doi.org/10.1038/38444.

(8) Spadafora, C.; Bellard, M.; Compton, J. L.; Chambon, P. The DNA Repeat Lengths in Chromatins from Sea Urchin Sperm and Gastrula Cells Are Markedly Different. *FEBS Lett.* **1976**, *69* (1), 281–285.

(9) Chen, P.; Li, W.; Li, G. Structures and Functions of Chromatin Fibers. *Annu. Rev. Biophys.* **2021**, *50*, 95–116. https://doi.org/10.1146/annurev-biophys-062920-063639.

(10) Olins, A. L.; Olins, D. E. Spheroid Chromatin Units (v Bodies). *Science (80-.).* **1974**, *183* (4122), 330–332.

(11) Emil, H. Das Heterochromatin Der Moose. *Jahrbücher für Wissenschaftliche Bot.* **1928**, *69*, 762–818.

(12) Berger, F. Emil Heitz, a True Epigenetics Pioneer. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (10), 572. https://doi.org/10.1038/s41580-019-0170-y.

(13) Penagos-Puig, A.; Furlan-Magaril, M. Heterochromatin as an Important Driver of Genome Organization. *Front. Cell Dev. Biol.* **2020**, *8* (September), 1–10. https://doi.org/10.3389/fcell.2020.579137.

(14) Dixon, J. R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J. S.; Ren, B. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* **2012**, *485* (7398), 376–380. https://doi.org/10.1038/nature11082.

(15) Dekker, J.; Heard, E. Structural and Functional Diversity of Topologically Associating Domains. *FEBS Lett.* **2015**, *589* (20), 2877–2884. https://doi.org/10.1016/j.febslet.2015.08.044.

(16) Lieberman-Aiden, E.; van Berkum, N. L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B. R.; Sabo, P. J.; Dorschner, M. O.; Sandstrom, R.; Bernstein, B.; Bender, M. A.; Groudine, M.; Gnirke, A.; Stamatoyannopoulos, J.; Mirny, L. A.; Lander, E. S.; Dekker, J. Comprehensive Mapping of Long-Range Interactions Revels Folding Principles of the Human Genome. *Science (80-.).* **2009**, *326*, 289–294.

(17) Allfrey, V. G.; Littau, V. C.; Mirsky, A. E. On the Role of of Histones in Regulation Ribonucleic Acid Synthesis in the Cell Nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **1963**, *49* (1962), 414–421. https://doi.org/10.1073/pnas.49.3.414.

(18) Han, M.; Grunstein, M. Nucleosome Loss Activates Yeast Downstream Promoters in Vivo. *Cell* **1988**, *55* (6), 1137–1145. https://doi.org/10.1016/0092-8674(88)90258-9.

(19) Knezetic, J. A.; Luse, D. S. The Presence of Nucleosomes on a DNA Template Prevents Initiation by RNA Polymerase II in Vitro. *Cell* **1986**, *45* (1), 95–104. https://doi.org/10.1016/0092-8674(86)90541-6.

(20) Lorch, Y.; LaPointe, J. W.; Kornberg, R. D. Nucleosomes Inhibit the Initiation of Transcription but Allow Chain Elongation with the Displacement of Histones. *Cell* **1987**, *49* (2), 203–210. https://doi.org/10.1016/0092-8674(87)90561-7.

(21) Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. Acetylation and Methylation of Histones and Their Possible Role in The. *Proc. Natl. Acad. Sci. United States* **1964**, *51* (1938), 786–794. https://doi.org/10.1073/pnas.51.5.786.

(22) Muller, H. J.; Altenburg, E. The Frequency of Translocations Produced by X-Rays in Drosophila. *Genetics* **1930**, *15* (4), 283.

(23) Elgin, S. C. R.; Reuter, G. Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in Drosophila. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (8), 1–26. https://doi.org/10.1101/cshperspect.a017780.

(24) 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. Regulation of Chromatin Structure by Site-Specific Histone H3 Methyltransferases. *Nature* **2000**, *406* (6796), 593–599. https://doi.org/10.1038/35020506.

(25) Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **2007**, *128* (4), 693–705. https://doi.org/10.1016/j.cell.2007.02.005.

(26) Chen, Y.; Sprung, R.; Tang, Y.; Ball, H.; Sangras, B.; Kim, S. C.; Falck, J. R.; Peng, J.; Gu, W.; Zhao, Y. Lysine Propionylation and Butyrylation Are Novel Post-Translational Modifications in Histones. *Mol. Cell. Proteomics* **2007**, *6* (5), 812. https://doi.org/10.1074/MCP.M700021-MCP200.

(27) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; Lu, Z.; Ye, Z.; Zhu, Q.; Wysocka, J.; Ye, Y.; Khochbin, S.; Ren, B.; Zhao, Y. Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* **2011**, *146* (6), 1016–1028. https://doi.org/10.1016/J.CELL.2011.08.008.

(28) Huang, H.; Lin, S.; Garcia, B. A.; Zhao, Y. Quantitative Proteomic Analysis of Histone Modifications. *Chem. Rev.* **2015**, *115* (6), 2376–2418. https://doi.org/10.1021/cr500491u.

(29) Strahl, B. D.; Allis, C. D. The Language of Covalent Histone Modifications. *Nature* **2000**, *403* (6765), 41–45. https://doi.org/10.1038/47412.

(30) Ruthenburg, A. J.; Allis, C. D.; Wysocka, J. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Mol. Cell* **2007**, *25* (1), 15–30. https://doi.org/10.1016/J.MOLCEL.2006.12.014.

(31) Hebbes, T. R.; Thorne, A. W.; Crane-Robinson, C. A Direct Link between Core Histone Acetylation and Transcriptionally Active Chromatin. *EMBO J.* **1988**, *7* (5), 1395.

(32) Shogren-Knaak, M.; Ishii, H.; Sun, J.-M.; Pazin, M. J.; James R. Davie; Peterson, C. L. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. **2006**, *311* (5762), 844–847. https://doi.org/https://doi.org/10.1126/science.1124000.

(33) Nakayama, J.; Rice, J. C.; Strahl, B. D.; Allis, C. D.; Grewal, S. I. S. Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science (80-.).* **2001**, *292* (5514), 110–113. https://doi.org/10.1126/SCIENCE.1060118.

(34) Boggs, B. A.; Cheung, P.; Heard, E.; Spector, D. L.; Chinault, A. C.; Allis, C. D. Differentially Methylated Forms of Histone H3 Show Unique Association Patterns with Inactive Human X Chromosomes. *Nat. Genet. 2001 301* **2001**, *30* (1), 73–76. https://doi.org/10.1038/ng787.

(35) Kuzmichev, A.; Nishioka, K.; Erdjument-Bromage, H.; Tempst, P.; Reinberg, D. Histone Methyltransferase Activity Associated with a Human Multiprotein Complex Containing the Enhancer of Zeste Protein. *Genes Dev.* **2002**, *16* (22), 2893–2905. https://doi.org/10.1101/GAD.1035902.

(36) Filion, G. J.; Bemmel, J. G. van; Braunschweig, U.; Talhout, W.; Kind, J.; Ward, L. D.; Brugman, W.; Castro, I. J. de; Kerkhoven, R. M.; Bussemaker, H. J.; van Steensel, B. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **2010**, *143* (2), 212–224. https://doi.org/10.1016/J.CELL.2010.09.009.

(37) Fischle, W. Talk Is Cheap—Cross-Talk in Establishment, Maintenance, and Readout of Chromatin Modifications. *Genes Dev.* **2008**, *22* (24), 3375–3382. https://doi.org/10.1101/GAD.1759708.

(38) Cedar, H.; Bergman, Y. Linking DNA Methylation and Histone Modification: Patterns and Paradigms. *Nat. Rev. Genet. 2009 105* **2009**, *10* (5), 295–304. https://doi.org/10.1038/nrg2540.

(39) Tsai, M.-C.; Manor, O.; Wan, Y.; Mosammaparast, N.; Wang, J. K.; Lan, F.; Shi, Y.; Segal, E.; Chang, H. Y. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science (80-.).* **2010**, *329* (5992), 689–693. https://doi.org/10.1126/SCIENCE.1192002.

(40) Roulland, Y.; Ouararhni, K.; Naidenov, M.; Ramos, L.; Shuaib, M.; Syed, S. H.; Lone, I. N.; Boopathi, R.; Fontaine, E.; Papai, G.; Tachiwana, H.; Gautier, T.; Skoufias, D.; Padmanabhan, K.; Bednar, J.; Kurumizaka, H.; Schultz, P.; Angelov, D.; Hamiche, A.; Dimitrov, S. The Flexible Ends of CENP-A Nucleosome Are Required for Mitotic Fidelity. *Mol. Cell* **2016**, *63* (4), 674–685. https://doi.org/10.1016/j.molcel.2016.06.023.

(41) Dhall, A.; Chatterjee, C. Chemical Approaches To Understand the Language of Histone Modifications. *ACS Chem. Biol.* **2011**, *6* (10), 987–999. https://doi.org/10.1021/CB200142C.

(42) He, S.; Bauman, D.; Davis, J. S.; Loyola, A.; Nishioka, K.; Gronlund, J. L.; Reinberg, D.; Meng, F.; Kelleher, N.; McCafferty, D. G. Facile Synthesis of Site-Specifically Acetylated and Methylated Histone Proteins: Reagents for Evaluation of the Histone Code Hypothesis. *Proc. Natl. Acad. Sci.* **2003**, *100* (21), 12033–12038. https://doi.org/10.1073/PNAS.2035256100.

(43) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. Synthesis of Proteins by Native Chemical Ligation. *Science (80-.).* **1994**, *266* (5186), 776–779. https://doi.org/10.1126/SCIENCE.7973629.

(44) Muir, T. W.; Sondhi, D.; Cole, P. A. Expressed Protein Ligation: A General Method for Protein Engineering. *Proc. Natl. Acad. Sci.* **1998**, *95* (12), 6705–6710. https://doi.org/10.1073/PNAS.95.12.6705.

(45) Shimko, J. C.; North, J. A.; Bruns, A. N.; Poirier, M. G.; Ottesen, J. J. Preparation of Fully Synthetic Histone H3 Reveals That Acetyl-Lysine 56 Facilitates Protein Binding within Nucleosomes. *J. Mol. Biol.* **2011**, *408* (2), 187–204.

(46) Maity, S. K.; Jbara, M.; Mann, G.; Kamnesky, G.; Brik, A. Total Chemical Synthesis of Histones and Their Analogs, Assisted by Native Chemical Ligation and Palladium Complexes. *Nat. Protoc.* **2017**, *12* (11), 2293–2322. https://doi.org/10.1038/NPROT.2017.049.

(47) Varshavsky, A. The Ubiquitin System, an Immense Realm. *Annu. Rev. Biochem.* **2012**, *81*, 167–176. https://doi.org/10.1146/ANNUREV-BIOCHEM-051910-094049.

(48) Weake, V. M.; Workman, J. L. Histone Ubiquitination: Triggering Gene Activity. *Mol. Cell* **2008**, *29* (6), 653–663.

(49) Chatterjee, C.; McGinty, R. K.; Pellois, J. P.; Muir, T. W. Auxiliary-Mediated Site-Specific Peptide Ubiquitylation. *Angew. Chemie - Int. Ed.* **2007**, *46* (16), 2814–2818. https://doi.org/10.1002/ANIE.200605155.

(50) Ajish Kumar, K. S.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A. Highly Efficient and Chemoselective Peptide Ubiquitylation. *Angew. Chemie - Int. Ed.* **2009**, *48* (43), 8090–8094. https://doi.org/10.1002/ANIE.200902936.

(51) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. Dual Native Chemical Ligation at Lysine. *J. Am. Chem. Soc.* **2009**, *131* (38), 13592–13593. https://doi.org/10.1021/JA905491P.

(52) Fierz, B.; Kilic, S.; Hieb, A. R.; Luger, K.; Muir, T. W. Stability of Nucleosomes Containing Homogenously Ubiquitylated H2A and H2B Prepared Using Semisynthesis. *J. Am. Chem. Soc.* **2012**, *134* (48), 19548–19551. https://doi.org/10.1021/JA308908P.

(53) Weller, C. E.; Huang, W.; Chatterjee, C. Facile Synthesis of Native and Protease-Resistant Ubiquitylated Peptides. *ChemBioChem* **2014**, *15* (9), 1263–1267. https://doi.org/10.1002/CBIC.201402135.

(54) Bi, X.; Yang, R.; Feng, X.; Rhodes, D.; Liu, C. F. Semisynthetic UbH2A Reveals Different Activities of Deubiquitinases and Inhibitory Effects of H2A K119 Ubiquitination on H3K36 Methylation in Mononucleosomes. *Org. Biomol. Chem.* **2016**, *14* (3), 835–839. https://doi.org/10.1039/C5OB02323H.

(55) Chatterjee, C.; McGinty, R. K.; Fierz, B.; Muir, T. W. Disulfide-Directed Histone Ubiquitylation Reveals Plasticity in HDot1L Activation. *Nat. Chem. Biol.* **2010**, *6* (4), 267–269. https://doi.org/10.1038/NCHEMBIO.315.

(56) Valkevich, E. M.; Guenette, R. G.; Sanchez, N. A.; Chen, Y. C.; Ge, Y.; Strieter, E. R. Forging Isopeptide Bonds Using Thiol-Ene Chemistry: Site-Specific Coupling of Ubiquitin Molecules for Studying the Activity of Isopeptidases. *J. Am. Chem. Soc.* **2012**, *134* (16), 6916–6919. https://doi.org/10.1021/JA300500A.

(57) Weikart, N. D.; Sommer, S.; Mootz, H. D. Click Synthesis of Ubiquitin Dimer Analogs to Interrogate Linkage-Specific UBA Domain Binding. *Chem. Commun.* **2012**, *48* (2), 296–298. https://doi.org/10.1039/C1CC15834A.

(58) West, M. H. P.; Bonner, W. M. Histone 2B Can Be Modified by the Attachment of Ubiquitin. *Nucleic Acids Res.* **1980**, *8* (20), 4671–4680. https://doi.org/10.1093/NAR/8.20.4671.

(59) Goldknopf, I. L.; French, M. F.; Daskal, Y.; Busch, H. A Reciprocal Relationship between Contents of Free Ubiquitin and Protein A24, Its Conjugate with Histone 2A, in Chromatin Fractions Obtained by the DNase II, Mg++ Procedure. *Biochem. Biophys. Res. Commun.* **1978**, *84* (3), 786–793.

(60) Kim, J.; Guermah, M.; McGinty, R. K.; Lee, J. S.; Tang, Z.; Milne, T. A.; Shilatifard, A.; Muir, T. W.; Roeder, R. G. RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. *Cell* **2009**, *137* (3), 459–471. https://doi.org/10.1016/j.cell.2009.02.027.

(61) Wang, H.; Wang, L.; Erdjument-Bromage, H.; Vidal, M.; Tempst, P.; Jones, R. S.; Zhang, Y. Role of Histone H2A Ubiquitination in Polycomb Silencing. *Nat. 2004 4317010* **2004**, *431* (7010), 873–878. https://doi.org/10.1038/nature02985.

(62) Sun, Z.-W.; Allis, C. D. Ubiquitination of Histone H2B Regulates H3 Methylation and Gene Silencing in Yeast. *Nat. 2002 4186893* **2002**, *418* (6893), 104–108. https://doi.org/10.1038/nature00883.

(63) Ng, H. H.; Xu, R. M.; Zhang, Y.; Struhl, K. Ubiquitination of Histone H2B by Rad6 Is Required for Efficient Dot1-Mediated Methylation of Histone H3 Lysine 79. *J. Biol. Chem.* **2002**, *277* (38), 34655–34657.

(64) McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W. Chemically Ubiquitylated Histone H2B Stimulates HDot1L-Mediated Intranucleosomal Methylation. *Nature* **2008**, *453* (7196), 812–816. https://doi.org/10.1038/NATURE06906.

(65) Holt, M. T.; David, Y.; Pollock, S.; Tang, Z.; Jeon, J.; Kim, J.; Roeder, R. G.; Muir, T. W. Identification of a Functional Hotspot on Ubiquitin Required for Stimulation of Methyltransferase Activity on Chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (33), 10365–10370. https://doi.org/10.1073/PNAS.1504483112.

(66) L, Z.; MT, H.; N, O.; A, Z.; MM, M.; B, W.; TW, M. Evidence That Ubiquitylated H2B Corrals HDot1L on the Nucleosomal Surface to Induce H3K79 Methylation. *Nat. Commun.* **2016**, *7*. https://doi.org/10.1038/NCOMMS10589.

(67) Goldknopf, I. L.; French, M. F.; Musso, R.; Busch, H. Presence of Protein A24 in Rat Liver Nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (12), 5492–5495.

https://doi.org/10.1073/PNAS.74.12.5492.

(68) Jason, L. J. M.; Moore, S. C.; Ausió, J.; Lindsey, G. Magnesium-Dependent Association and Folding of Oligonucleosomes Reconstituted with Ubiquitinated H2A. *J. Biol. Chem.* **2001**, *276* (18), 14597–14601. https://doi.org/10.1074/jbc.M011153200.

(69) Fierz, B.; Chatterjee, C.; Mcginty, R. K.; Bar-dagan, M.; Raleigh, D. P.; Muir, T. W. Histone H2B Ubiquitylation Disrupts Local and Higher Order Chromatin Compaction. *Nat. Chem. Biol.* **2011**, *7* (2), 113–119. https://doi.org/10.1038/nchembio.501.Histone.

(70) Margueron, R.; Reinberg, D. The Polycomb Complex PRC2 and Its Mark in Life. *Nature* **2011**, *469* (7330), 343–349. https://doi.org/10.1038/NATURE09784.

(71) Blackledge, N. P.; Farcas, A. M.; Kondo, T.; King, H. W.; McGouran, J. F.; Hanssen, L. L. P.; Ito, S.; Cooper, S.; Kondo, K.; Koseki, Y.; Ishikura, T.; Long, H. K.; Sheahan, T. W.; Brockdorff, N.; Kessler, B. M.; Koseki, H.; Klose, R. J. Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* **2014**, *157* (6), 1445–1459.

(72) Whitcomb, S. J.; Fierzs, B.; McGinty, R. K.; Holt, M.; Ito, T.; Muir, T. W.; Allis, C. D. Histone Monoubiquitylation Position Determines Specificity and Direction of Enzymatic Cross-Talk with Histone Methyltransferases Dot1L and PRC2. *J. Biol. Chem.* **2012**, *287* (28), 23718–23725.

(73) Cooper, S.; Grijzenhout, A.; Underwood, E.; Ancelin, K.; Zhang, T.; Nesterova, T. B.; Anil-Kirmizitas, B.; Bassett, A.; Kooistra, S. M.; Agger, K.; Helin, K.; Heard, E.; Brockdorff, N. Jarid2 Binds Mono-Ubiquitylated H2A Lysine 119 to Mediate Crosstalk between Polycomb Complexes PRC1 and PRC2. *Nat. Commun. 2016 71* **2016**, *7* (1), 1–8. https://doi.org/10.1038/ncomms13661.

(74) Virdee, S.; Ye, Y.; Nguyen, D. P.; Komander, D.; Chin, J. W. Engineered Diubiquitin Synthesis Reveals Lys29-Isopeptide Specificity of an OTU Deubiquitinase. *Nat. Chem. Biol.* **2010**, *6* (10), 750–757. https://doi.org/10.1038/NCHEMBIO.426.

(75) Yuan, G.; Ma, B.; Yuan, W.; Zhang, Z.; Chen, P.; Ding, X.; Feng, L.; Shen, X.; Chen, S.; Li, G.; Zhu, B. Histone H2A Ubiquitination Inhibits the Enzymatic Activity of H3 Lysine 36 Methyltransferases. *J. Biol. Chem.* **2013**, *288* (43), 30832–30842.

(76) Su, H. L.; Li, S. S. L. Molecular Features of Human Ubiquitin-like SUMO Genes and Their Encoded Proteins. *Gene* **2002**, *296* (1–2), 65–73. https://doi.org/10.1016/S0378-1119(02)00843-0.

(77) Bayer, P.; Arndt, A.; Metzger, S.; Mahajan, R.; Melchior, F.; Jaenicke, R.; Becker, J. Structure Determination of the Small Ubiquitin-Related Modifier SUMO-1. *J. Mol. Biol.* **1998**, *280* (2), 275–286. https://doi.org/10.1006/JMBI.1998.1839.

(78) Cappadocia, L.; Lima, C. D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem. Rev.* **2017**, *118* (3), 889–918. https://doi.org/10.1021/ACS.CHEMREV.6B00737.

(79) Lois, L. M.; Lima, C. D. Structures of the SUMO E1 Provide Mechanistic Insights into SUMO Activation and E2 Recruitment to E1. *EMBO J.* **2005**, *24* (3), 439. https://doi.org/10.1038/SJ.EMBOJ.7600552.

(80) Desterro, J. M. P.; Rodriguez, M. S.; Kemp, G. D.; Hay, R. T. Identification of the Enzyme Required for Activation of the Small Ubiquitin-like Protein SUMO-1 *. *J. Biol. Chem.* **1999**, *274* (15), 10618–10624. https://doi.org/10.1074/JBC.274.15.10618.

(81) Desterro, J. M. .; Thomson, J.; Hay, R. T. Ubch9 Conjugates SUMO but Not Ubiquitin. *FEBS Lett.* **1997**, *417* (3), 297–300. https://doi.org/10.1016/S0014-5793(97)01305-7.

(82) Yang, S.-H.; Galanis, A.; Witty, J.; Sharrocks, A. D. An Extended Consensus Motif Enhances the Specificity of Substrate Modification by SUMO. *EMBO J.* **2006**, *25* (21), 5083. https://doi.org/10.1038/SJ.EMBOJ.7601383.

(83) Shiio, Y.; Eisenman, R. N. Histone Sumoylation Is Associated with Transcriptional Repression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230. https://doi.org/10.1073/pnas.1735528100.

(84) Ryu, H.-Y.; Zhao, D.; Li, J.; Su, D.; Hochstrasser, M. Histone Sumoylation Promotes Set3 Histone-Deacetylase Complex-Mediated Transcriptional Regulation. *Nucleic Acids Res.* **2020**, *48* (21), 12151–12168. https://doi.org/10.1093/NAR/GKAA1093.

(85) 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. Histone Sumoylation Is a Negative Regulator in Saccharomyces Cervisiae and Shows Dynamic Interplay with Positive-Acting Histone Modifications. *Genes Dev.* **2006**, *20*, 966–976. https://doi.org/10.1101/gad.1404206.6.

(86) Galisson, F.; Mahrouche, L.; Courcelles, M.; Bonneil, E.; Meloche, S.; Chelbi-Alix, M. K.; Thibault, P. A Novel Proteomics Approach to Identify SUMOylated Proteins and Their Modification Sites in Human Cells. *Mol. Cell. Proteomics* **2011**, *10* (2), M110.004796. https://doi.org/10.1074/mcp.M110.004796.

(87) Hendriks, I. A.; D'Souza, R. C. J.; Yang, B.; Verlaan-De Vries, M.; Mann, M.; Vertegaal, A. C. O. Uncovering Global SUMOylation Signaling Networks in a Site-Specific Manner. *Nat. Struct. Mol. Biol.* **2014**, *21* (10), 927–936. https://doi.org/10.1038/nsmb.2890.

(88) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long-Range Internucleosomal Interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837. https://doi.org/10.1074/jbc.M114.591644.

(89) Hecker, C. M.; Rabiller, M.; Haglund, K.; Bayer, P.; Dikic, I. Specification of SUMO1- and SUMO2-Interacting Motifs. *J. Biol. Chem.* **2006**, *281* (23), 16117–16127. https://doi.org/10.1074/jbc.M512757200.

(90) Kerscher, O. SUMO Junction—What's Your Function? New Insights through SUMO-Interacting Motifs. *EMBO Rep.* **2007**, *8* (6), 550. https://doi.org/10.1038/SJ.EMBOR.7400980.

(91) Ouyang, J.; Shi, Y.; Valin, A.; Xuan, Y.; Gill, G. Direct Binding of CoREST1 to SUMO-2/3 Contributes to Gene-Specific Repression by the LSD1/CoREST1/HDAC Complex. *Mol. Cell* **2009**, *34* (2), 145–154. https://doi.org/10.1016/j.molcel.2009.03.013.

(92) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **2004**, *119* (7), 941–953. https://doi.org/10.1016/j.cell.2004.12.012.

(93) Dhall, A.; Weller, C. E.; Chu, A.; Shelton, P. M. M.; Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-CoREST Complex. *ACS Chem. Biol.* **2017**, *12* (9), 2275–2280. https://doi.org/10.1021/acschembio.7b00716.

(94) Elsässer, S. J.; Ernst, R. J.; Walker, O. S.; Chin, J. W. Genetic Code Expansion in Stable Cell Lines Enables Encoded Chromatin Modification. *Nat. Methods* **2016**, *13* (2), 158–164. https://doi.org/10.1038/nmeth.3701.

(95) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **2007**, *128* (5), 1003–1012. https://doi.org/10.1016/j.cell.2006.12.041.

(96) Piotukh, K.; Geltinger, B.; Heinrich, N.; Gerth, F.; Beyermann, M.; Freund, C.; Schwarzer, D. Directed Evolution of Sortase A Mutants with Altered Substrate Selectivity Profiles. *J. Am. Chem. Soc.* **2011**, *133* (44), 17536–17539. https://doi.org/10.1021/JA205630G.

(97) Chen, W.-T.; Alpert, A.; Leiter, C.; Gong, F.; Jackson, S. P.; Miller, K. M. Systematic Identification of Functional Residues in Mammalian Histone H2AX. *Mol. Cell. Biol.* **2013**, *33* (1), 111. https://doi.org/10.1128/MCB.01024-12.

Chapter 2

Mechanism of SUMO crosstalk with the LSD1-CoREST1-HDAC1 complex

2.1 Introduction

The transcriptional repressor LCH complex consists of the lysine-specific demethylase 1 (LSD1), the Co-repressor of REST 1 (CoREST1) protein and Histone deacetylase 1/2 (HDAC1) (**Figure 2.1**).[1,2] LCH plays a critical role in regulating human gene expression by post-translationally modifying histone lysine side-chains through the enzymatic activities of LSD1 and HDAC1. Due to their direct activities on chromatin, and the resulting effects on gene transcription, LCH is a valuable therapeutic target. LCH has roles in various malignancies associated with the activities of each of its subunits. LSD1 is the lysine-specific demethylase and was the first demethylase discovered to reverse lysine side-chain methylation.[3] It is overexpressed in lung, breast, prostate,

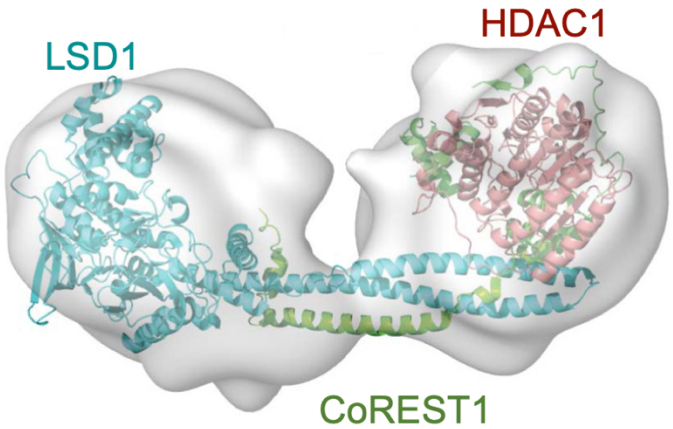


Figure 2.1 The LSD1:CoREST1:HDAC1 complex. Crystal structures of LSD1:CoREST1 (PDB: 2V1D) and HDAC1:MTA1 (PDB: 4BKX) fitted within the EM model of the ternary LSD1:CoREST1:LSD1 complex. Adapted from reference 46.

and blood cancers.[4–9] HDAC1/2 are responsible for removing acetyl marks associated with gene activation and active transcription.[10] The knockdown or inhibition of the histone deacetylase, HDAC1, has been observed to have antitumor activity and the misregulation of HDAC1 is associated with a plethora of human cancers.[11,12] CoREST1 serves as the central scaffolding protein that couples the enzymatic activities of LSD1 and HDAC1 to chromatin, and it is recruited by various protein-protein interactions to specific gene promoters.[13,14]

Inhibitors of both LSD1 and HDAC1 are valuable cancer therapeutics. Trichostatin-A (TSA) and Vorinostat (suberoylanilide hydroxamic acid, SAHA) were early developed HDAC inhibitors, and while TSA exhibited high cellular toxicity, Vorinostat is an FDA-approved drug to treat cutaneous T-cell lymphoma.[15,16] Several other HDAC inhibitors have now also been approved by the FDA for the treatment of various cancers.[15] Inhibitors of LSD1 are also being developed as therapeutics.[17] A dual inhibitor against both HDAC1 and LSD1 has also been developed to specifically target the LCH complex by linking together established HDAC1 and LSD1 inhibitors.[18] The dual inhibitor, named Corin, was demonstrated to inhibit both HDAC1 and LSD1 activities, as well as attenuate tumor growth in melanoma mouse xenografts.[18] Furthermore, the targeting of the LCH complex with a dual inhibitor may decrease off-target inhibition of HDAC1 and LSD1 family members, and not inhibit complexes containing only one or the other. Thus, the LCH complex represents an important biological target, and a careful understanding of its roles in the epigenetic regulation of gene expression will provide insights on the precise mechanism of inhibitor activity, and allow for the refinement and development of future therapeutic strategies.

LSD1 sequentially removes H3K4 mono- and dimethylation (H3K4me1/2) by utilizing the co-factor flavin adenine dinucleotide (FAD) as the oxidant. LSD1 is the founding member of the FAD-dependent family of demethylases (**Figure 2.2**). LSD1 contains three domains: the SWIRM (Swi3/Rcs8/Moira) domain, Tower domain, and the catalytic amine oxidase-like (AOL)

domain.[19,20] SWIRM plays roles in the nuclear localization of LSD1 and protein stability. The Tower domain interrupts and splits the AOL domain into two fragments and forms an elongated helix-turn-helix motif that forms a three-helix bundle with the SANT2 (SWI3/ADA2/N-CoR/TFIIB) domain of CoREST1.[21] The two AOL fragments come together at the start and end of the Tower domain and bind FAD within the active site to catalyze lysine demethylation (**Figure 2.2**). The FAD co-factor is reduced upon oxidation of the α-carbon in a methyl-group on H3K4me1/2 to generate the corresponding imine intermediate species.[22] The overall reaction is a two-electron transfer to FAD in the form of a hydride shift from the methyl group and proton shift from the lysine ε-amine. This requirement constrains LSD1 to mono- and di-methylated substrates as trimethyllysine lacks the necessary amine hydrogen to facilitate imine formation. FADH2 is reoxidized to FAD by molecular oxygen, thus producing a molecule of hydrogen peroxide per demethylated group. Water hydrolyzes the imine intermediate, releasing the former methyl α-carbon as a molecule of formaldehyde, and the protonated demethylated lysine. LSD1 has one homolog, LSD2, which lacks the Tower domain, and therefore does not appear to interact with CoREST proteins in cells.[19]

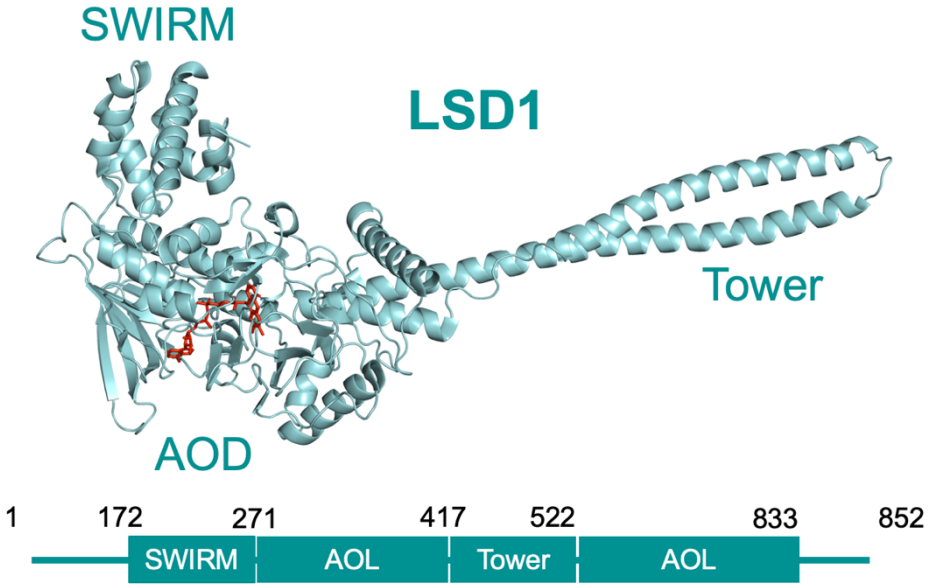


Figure 2.2 Structure of LSD1. The crystal structure of LSD1(171-836) alone is shown (PDB: 2V1D), with FAD (red) bound in its AOL domain active site. Below is a full-length domain map showing the relative positions of each of the domains in LSD1.

HDAC1 is a predominantly single domain protein and has been identified as part of the LCH complex through extensive interactions with the ELM2-SANT1 domain in CoREST1 (**Figure 2.3**).[1,23] There have been 18 HDACs identified till date, and these are categorized into class I, class IIa, class IIb, class III (the sirtuins), and class IV.[10] HDAC1 and its homolog HDAC2 are part of the class I HDACs, and along with class IIa/b and class IV, their activity depends on an active-site Zn^{2+}. The Zn^{2+} is utilized to promote nucleophilic attack by water on the acetamide carbonyl of acetyllysine. The final products released are the protonated deacetylated lysine and acetate anion. The sirtuin family of deacetylases, on the other hand, require the co-factor NAD^{+}.[24] While acetylated histone lysines have been observed to be a major substrate for HDAC1, it is sometimes referred to more generally as a lysine deacetylase based on its wide-ranging activity on non-histone substrates.

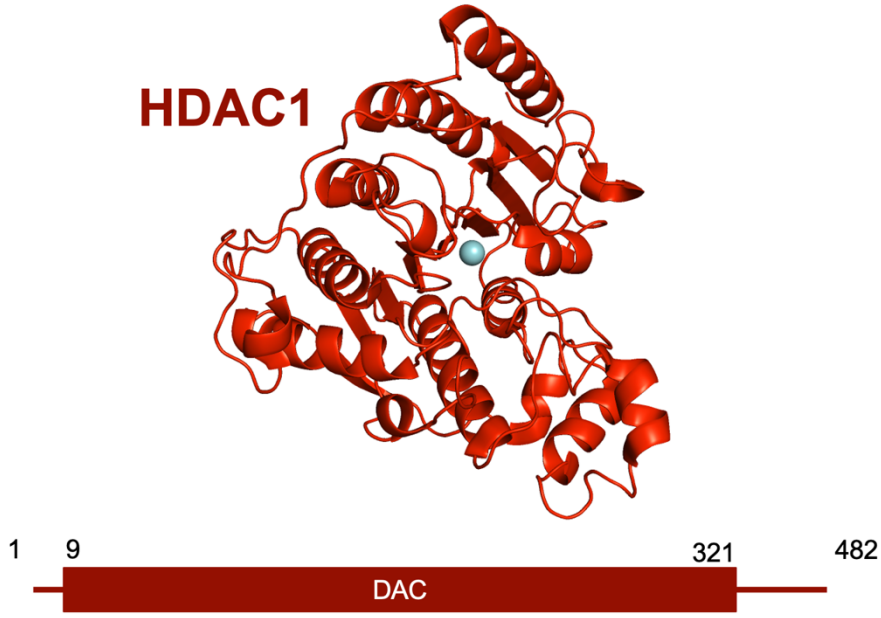


Figure 2.3. Structure of HDAC1. Crystal structure of HDAC1(8-376) only with a Zn^{2+} ion bound at its active site (PDB: 4BKX). Below is a full-length domain map showing the position of the structured deacetylase domain (DAC).

CoREST was first discovered as a binding partner of RE1 silencing transcription factor (REST), also known as neural-restrictive silencing factor (NRSF). CoREST bound to REST suppresses neuronal genes in non-neuronal cells to aid in the differentiation and regulation of non-neuronal

cells. There are three homologs of CoREST, encoded by the *RCOR1-3* genes, and each homolog forms distinct chromatin-modifying complexes. Structurally, CoREST1 contains a ELM2 (Egl-27 and MTA homology 2) and SANT1 domain at its *N*-terminus (**Figure 2.4**). A linker domain separates a second SANT domain, SANT2, near the *C*-terminus. The linker domain interacts with the Tower domain of LSD1, whereas the ELM2-SANT1 domains make contacts with HDAC1. The SANT2 domain of CoREST1 has been hypothesized to make contacts with nucleosomal DNA.

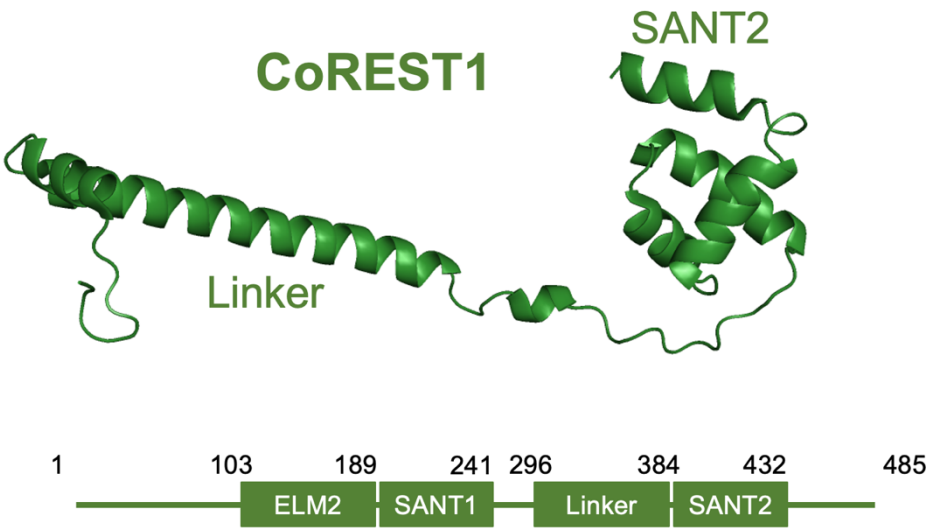


Figure 2.4. Structure of CoREST1. The crystal structure of CoREST1(308-440) (PDB: 2V1D) only is shown. Below is a full-length domain map showing relative positions of domains.

The small ubiquitin-like modifier protein 3 (SUMO3) is a histone PTM found at Lys12 in histone H4 (H4K12su).[25] Histone sumoylation is associated with transcriptional repression and recruitment of the heterochromatin protein 1 (HP1).[26] However, the precise mechanism of gene repression by SUMO is unknown. Surprisingly, previous biophysical studies in the Chatterjee lab that employed analytical ultracentrifugation and single-molecule FRET techniques revealed that H4K12su inhibits compaction to a similar degree as H4K16 acetylation, a transcription activating mark.[27] In contrast with acetylation, SUMO may provide numerous surfaces for binding chromatin modifiers and remodelers that facilitate gene repression. Ouyang and colleagues identified a non-canonical SUMO-interaction motif (ncSIM) within CoREST1 that bound SUMO2/3 but not

SUMO1.[13] We recently discovered that H4K12su stimulates demethylation and deacetylation by the LSD1-CoREST1[28,29] and HDAC1-CoREST1 (unpublished) sub-complexes, respectively, by engaging the SIM within CoREST1. The SUMO3-CoREST1 interaction via the ncSIM is central to the stimulation of LCH, as assays on nucleosomes containing unmodified H4 no longer showed stimulated activity. SUMO-interacting motifs (SIMs) have been characterized predominantly by a central hydrophobic core that has the consensus sequence of V/I-V/I-X-V/I, where X is typically a Glu or Asp residue (**Figure 2.5A**).[30] Compared with the diversity of 21 identified ubiquitin-binding domains, there has still only been a single SIM motif identified to bind SUMO, with few well-defined characteristics.[31] In addition to the hydrophobic core, an adjacent Ser-X-Ser was first thought to be a part of the SIM, however, this was later shown to be both insufficient and unnecessary.[32] An acidic patch consisting of glutamates (E) and aspartates (D) either *N*- or *C*-terminal to the hydrophobic core has also been observed, but this patch is not present in all SIMs due to the possibility of Ser phosphorylation imparting negative charge instead of E/D. Phosphorylation of serine residues proximal to the hydrophobic core has also been observed to shift SIM binding preference from SUMO2/3 to SUMO1.[33] Mutational analysis of the CoREST1 SIM between amino acids (255-275) showed that individual deletion of the acidic patch, hydrophobic core, or the small stretch of amino acids between the hydrophobic core and acidic patch significantly reduced binding.[13] When all three regions were deleted in CoREST1 its binding to SUMO2/3 was completely abolished.

Our understanding of SIMs is tenuous due to their large diversity, and even less is known about the non-consensus SIM within CoREST1, which diverges from better studied SIMs in proteins such as Promyelocytic Leukemia (PML) and the SUMO-ligase PIAS1. Currently, studies of the non-canonical SIM within CoREST have employed qualitative GST-pulldowns, and a quantitative characterization of non-canonical SIM binding to SUMO is entirely absent. The fact that H4K12su inhibits chromatin compaction yet strongly associates with gene repression and stimulates LSD1

function led to my proposed hypothesis that sumoylated H4 stimulates the LCH complex by binding the non-canonical SIM in CoREST1. I hypothesized that SUMO3 may position the LCH complex in a catalytically competent binding orientation, proximal to the methylated and acetylated H3 tails in mononucleosomes, in order to stimulate gene repression by the removal of gene activating acetyl and methyl marks. In order to test my hypothesis, I undertook a quantitative study of SUMO3 binding to the non-canonical SIM in CoREST1.

2.2 Results and discussion

2.2.1 Analysis of SUMO-interacting motifs

In order to visualize the large diversity of SUMO-interacting motifs in proteins, I aligned several canonical and non-canonical SIM sequences from various proteins that have been demonstrated to interact with SUMO1 or SUMO2/3 as a part of their mechanism of action **(Figure 2.5B)**.[32,34] The results of this alignment clearly revealed that the only well-conserved feature of these canonical SIMs is the previously reported hydrophobic core of V/I-V/I-X-V/I, or the reverse orientation of this sequence V/I-X- V/I -V/I. My alignment is consistent with a SIM analysis performed by Hecker and colleagues.[35] Interestingly, the SIM in the DNA-repair enzyme thymine DNA glycosylase (TDG) also appears to be a non-canonical SIM, containing two acidic residues at positions 2 and 3, sandwiched between two valines. Despite this difference, the observed sequence has been shown to be a *bona fide* SIM and interacts with the α-helix and β-strand 2 of SUMO1.[36] Upon closer inspection, it would appear that other residues, but predominantly negatively charged amino acids, may be accommodated at either position 2 or 3.

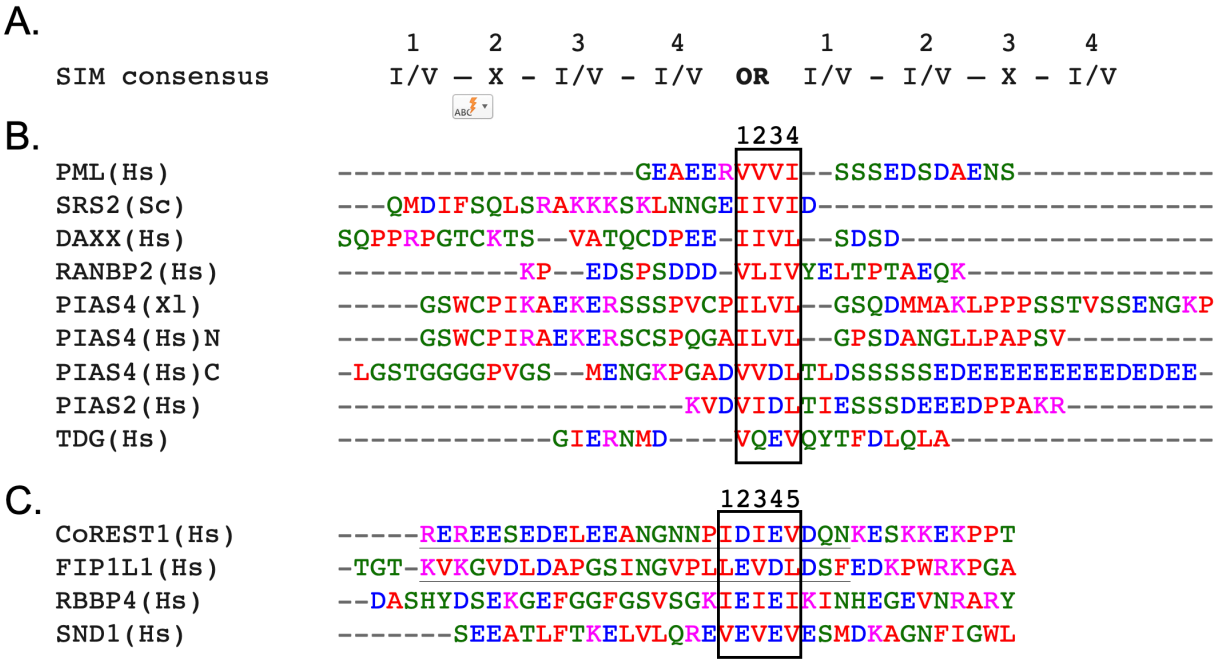


Figure 2.5. Sequence alignments of typical consensus SIMs and non-canonical SIMs. (A) Hydrophobic core consensus sequence SIM. (B) Alignment centered around the hydrophobic core SIM of various proteins known to functionally interact with SUMO. (C) Alignment of proteins identified to contain a non-canonical SIM hydrophobic core that was first identified in CoREST1. Hydrophobic core residues of the SIM are boxed, and positions indicated. Underlined sequences were used in NMR studies. Letter color indicates basic (pink); acidic (blue); small, hydrophobic, aromatic -Y (red); hydroxyl, sulfhydryl, amine, G (green) amino acids.

The acidic patch can also be seen to occur upstream or downstream of the hydrophobic core (**Figure 2.5B**). The negatively charged residues in these stretches have been observed to play a role in preferential binding to SUMO1 over SUMO2/3.[35] SIMs that do not contain an acidic patch, however can still bind SUMO1.[36,37] Whether the acidic patch is N-terminal or C-terminal from the hydrophobic core was observed to influence whether the SIM binds in a parallel or antiparallel orientation, as exemplified respectively in the SUMO E3 ligase, PIAS, and RanBP2 (Ran binding protein 2) SIMs (**Figure 2.6**).[35,38,39] Proximal serine or threonine residues can also introduce negative charge to the SIM upon their phosphorylation. The death-domain associate protein 6, DAXX, contains a SIM at its *C*-terminus with two serine residues following the hydrophobic core. Chang and colleagues showed that phosphorylation of these two serines by the kinase CK2

resulted in increased binding affinity to SUMO1 over SUMO2/3.[33] The DAXX SIM was also observed to bind parallel to strand β2 in SUMO1, consistent with the negative charge-position effect on binding orientation of SIMs. The lysine residues K37 and K39 in SUMO1 and K34 and K36 in SUMO2 have been implicated in forming salt bridges with phosphorylated and negatively charged residues within the acidic patch.[35]

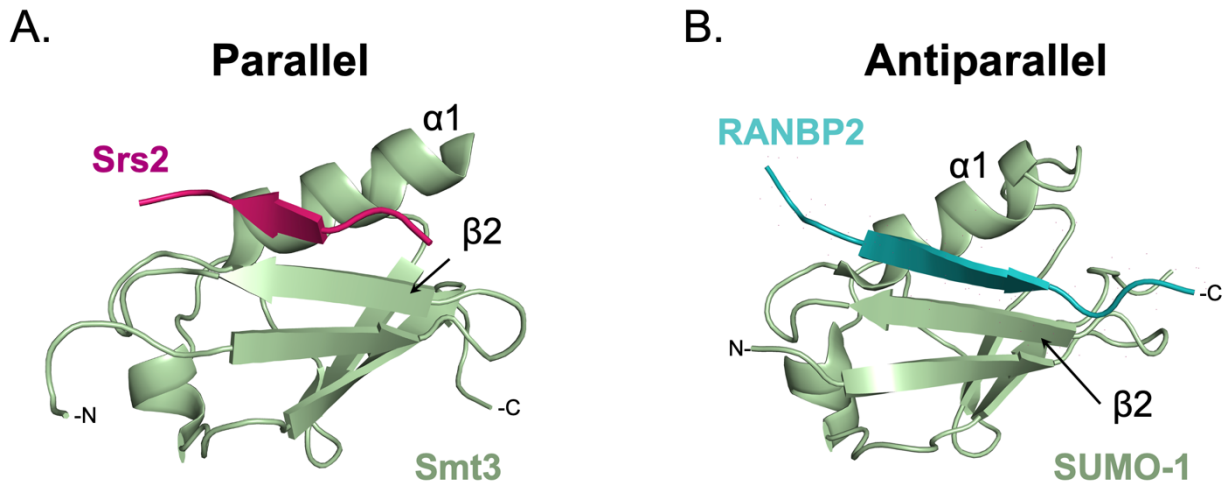


Figure 2.6. SIM binding to SUMO can result in a parallel or antiparallel β-strand addition to the β-sheet of SUMO. (A) The Srs2 SIM in yeast binds to the yeast SUMO ortholog Smt3 in a parallel orientation. (B) The RanBP2 SIM binds to SUMO1 in an antiparallel orientation.

In contrast with the consensus SIM, CoREST1's non-canonical SIM is "diluted" at its hydrophobic core, broken up by an additional amino acid at position 4 that is typically an I/V in canonical SIMs (**Figure 2.5C**). Despite its deviation from a traditional SIM sequence, Ouyang and colleagues observed that the noncanonical-SIM in CoREST is important for binding SUMO2/3 over SUMO1, and plays a functional role in transcriptional repression.[13] Several other proteins have this type of SIM including FIP1L1, RBBP4, and SND1, and these were also demonstrated to bind SUMO2/3 over SUMO1. Interestingly, mutation of the three hydrophobic residues within the non-canonical-SIM of CoREST1, FIP1L1, and RBBP4 resulted in the abrogation of binding to SUMO2/3, but not in SND1. This observation may be due to the potential existence of additional canonical or non-canonical SIMs within the full-length SND1 protein. Counterintuitively, the presence of an acidic

patch N-terminal to the CoREST1 SIM would have suggested that this SIM would preferentially bind to SUMO1, but this pattern of binding may not be preserved in non-canonical SIMs. It remains to be seen, however, whether this acidic patch dictates the CoREST1 SIM binding to SUMO3 in an antiparallel orientation with respect to β2 in SUMO3. There are no other apparent features adjacent to the hydrophobic core that would group these non-canonical SIMs together, including the absence of Ser-X-Ser motifs. In fact, post-translational modifications of the non-canonical SIM sequences have also yet to be explored. Thus, a detailed characterization of the non-canonical SIMs is first necessary to understand this subgroup of SUMO binding proteins where distal sequences from the non-canonical SIM may also be important for efficient SUMO binding.[13]

2.2.2 Investigations of SUMO-SIM binding by fluorescence anisotropy and TR-FRET

In order to quantify binding of the CoREST1 noncanonical SIM to SUMO3, we first attempted fluorescence anisotropy experiments.[40,41] This technique utilizes binding assays in a reflective black 96-well plate to measures the degree of dynamism of a protein complex via a fluorophore tag (**Figure 2.7A**). The degree of dynamism (or protein tumbling) is measured as the amount of anisotropic emission (r) from the protein upon excitation with plane-polarized light ($r = I_{\parallel}-I_{\perp}/I_{\parallel}+I_{\perp}$).[42]

Individually, a fluorescently-tagged protein tumbles in solution at an average rate related to its size and rigidity. Upon the addition of a significantly large binding partner, the tagged protein tumbles slower due to the increased overall mass of the protein complex. The slower tumbling of this complex leads to an increase in the observed fluorescence anisotropy and greater detection of plane polarized light relative to total light intensity. The change in polarization of the incident plane polarized light can be plotted against increasing amounts of the binding partner in order to calculate a K_d, or binding affinity for the protein complex.

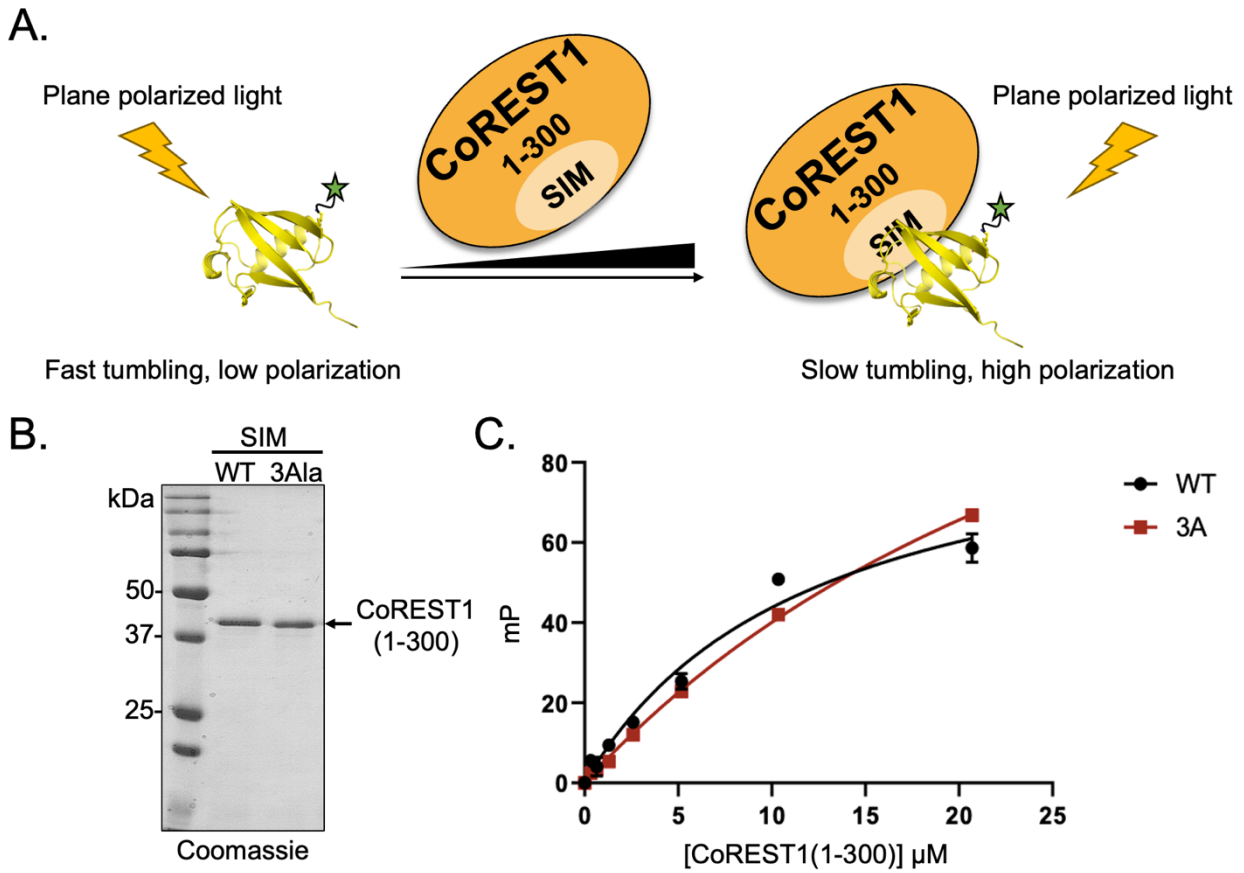


Figure 2.7. CoREST1(1-300) fails to saturate SUMO binding in fluorescence anisotropy assays. (A) Schematic of fluorescence anisotropy assay. Plane-polarized light excites the Alexa Fluor 488 attached to free SUMO in solution. Increasing addition of CoREST1(1-300) wild type (WT) or 3Ala mutant (3A) forms a larger, slower tumbling complex with SUMO, resulting in increased fluorescence anisotropy. (B) Coomassie gel of purified CoREST1(1-300) and CoREST1(1-300)-3Ala mutant. (C) Plot of fluorescence anisotropy assays using Su3-AF488 and increasing amounts of CoREST1(1-300) or the 3Ala mutant.

We chose to first investigate non-canonical SIM binding by free unconjugated SUMO3 before pursuing binding to sumoylated H4 within a nucleosome. In this way, we could assess SUMO-SIM interactions in the absence of other binding events that occur between CoREST1 and the nucleosome; such as DNA binding by the SANT2 domain. The SUMO3(K11C, C47S) mutant was expressed as an *N*-terminal fusion protein to AVA intein containing a His_6-tag at its *C*-terminus. The fusion protein was purified by immobilized metal-ion affinity chromatography (IMAC) and the intein was hydrolyzed with DTT to yield a tag-less, full length SUMO3 protein with a carboxylic

acid at the *C*-terminus. The K11C mutation, first reported by the Lima group, was engineered into the construct to allow for incorporation of an Alexa Fluor 488 dye (Invitrogen) via a maleamide linkage to give Su3-AF488. This specific site was chosen as it is located away from the SIM-binding cleft on the N-terminus of SUMO and this site was previously used in similar fluorescence anisotropy assays with canonical SIM sequences.[41] Analysis of Su3-AF488 after HPLC purification by ESI-MS (**Figure 2.S1**) and reconstitution in buffer for circular dichroism confirmed that the incorporation of the fluorescent dye and purification by HPLC did not significantly alter SUMO structure (**Figure 2.S2**). Titration of the full-length His_6-CoREST1 against Su3-AF488, however, failed to saturate SUMO binding. CoREST1 is a very difficult protein to concentrate to greater than 10 μM and tends to precipitate out of solution at higher concentrations. This is likely due to instability from lacking its corepressor binding partners, REST1, HDAC1, and/or LSD1. Additionally, I observed that lysing CoREST1 via French press method avoided the fragmentation of the protein observed previously upon extensive sonication, and increased the efficiency of cell lysis than our prior protocol of a limited physical sonication with the micro-tip sonicator.

In order to overcome the solubility and concentration issues of full-length CoREST1, I also tried a smaller protein construct, CoREST1(1-300), that still contains the SIM and had previously demonstrated reasonable binding to SUMO3 in pulldown experiments that was comparable to the full-length construct.[13] The CoREST(1-300) protein was slightly better behaved during protein concentration by spin concentrators and was employed for fluorescence anisotropy experiments. However, even this construct at the maximum 20 μM concentrations used was unable to saturate binding to Su3-AF488 (**Figure 2.7C**). Additionally, CoREST1(1-300)-3A, which contains the three hydrophobic residues within the SIM mutated to Ala, also showed some binding to SUMO, albeit with a slightly decreased degree of binding. Attempts to conduct TR-FRET experiments using a Terbium (Tb) cryptate-labeled anti-His_6 antibody (Cisbio) to bind His_6-CoREST1(1-300) and

titration of Su3-AF488 also were unsuccessful due to background FRET, leading us to choose an alternate strategy to measure the SUMO3 non-canonical SIM interaction.

2.2.3 Synthesis of CoREST1-SIM peptides

As I was unable to obtain SUMO-CoREST binding curves that reached binding saturation in order to measure a K_d, I decided to utilize NMR spectroscopy with SIM peptides and purified SUMO3. As an advantage to this technique, the $^{1}H/^{15}N$ HSQC-spectra for SUMO3 has been published and is readily available at the Biological Magnetic Resonance Data Bank (https://bmrb.io/). My binding partner of choice for NMR was a short peptide containing the CoREST1 noncanonical SIM that could be titrated at much higher concentrations in binding assays than the CoREST1(1-300) protein.

I synthesized the peptide, CoREST1-SIM, consisting of residues 255-269 (**Figure 2.5C**) using automated microwave-assisted solid-phase peptide synthesis. This peptide was chosen to include an acidic patch N-terminal to the hydrophobic core of the noncanonical SIM, which has been previously demonstrated to participate in SUMO-SIM interactions. The *N*- and *C*- termini of the peptide were also acetylated and amidated, respectively, in order to mimic its internal position in CoREST. The resulting peptide was cleaved from resin and purified by HPLC and analyzed by ESI-MS (**Figure 2.8A/B**). I also prepared the same peptide but with the 3 Ala mutation in the hydrophobic core as CoREST1-SIM(3A) (**Figure 2.8C/D**).

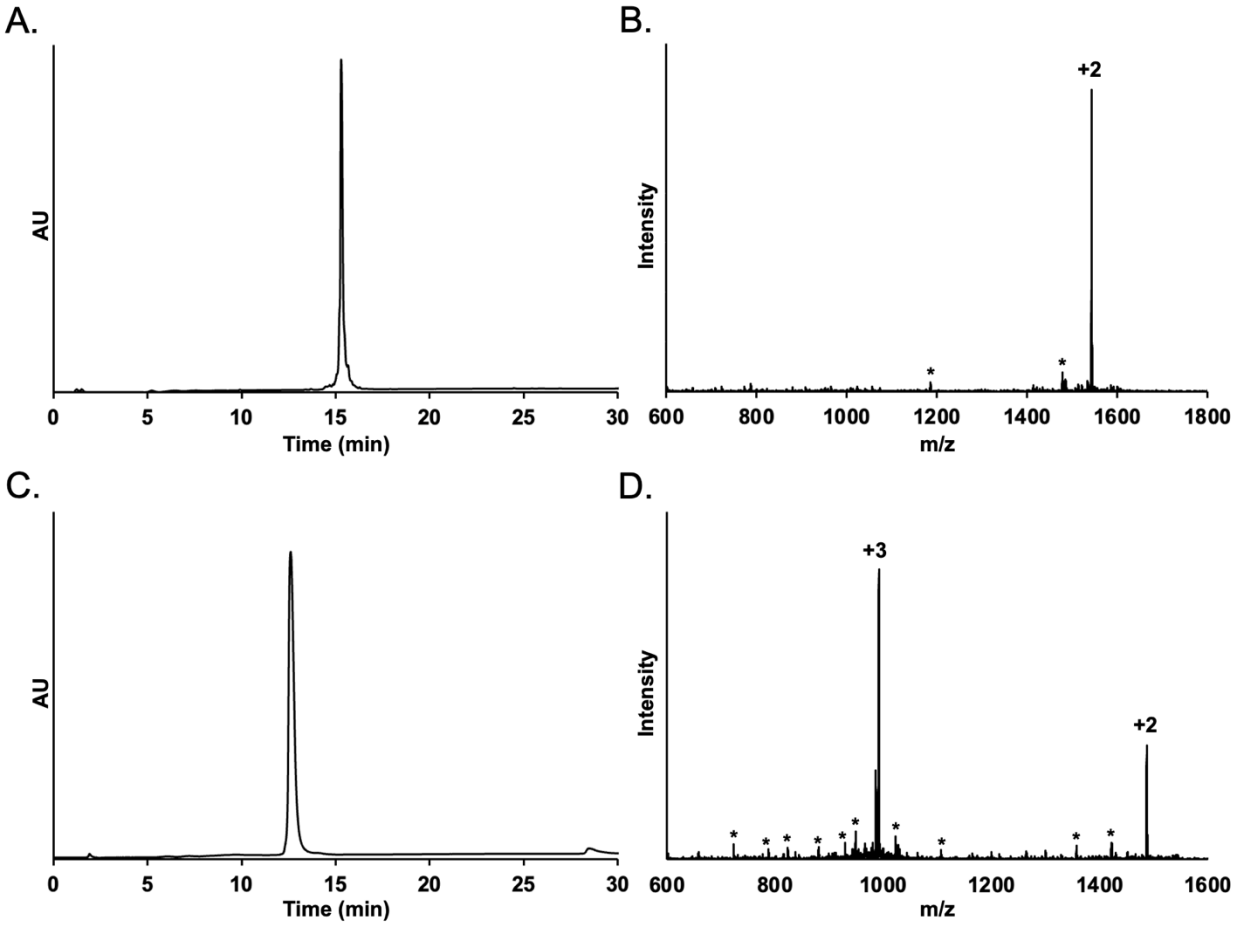


Figure 2.8. Purification of CoREST1-SIM and CoREST1-SIM(3A) peptides. (A) C18 analytical RP-HPLC chromatogram of purified CoREST1-SIM peptide, gradient 0-73% B, 30 min. (B) ESI-MS of purified CoREST1-SIM. Calculated m/z $[M+H]^+$ 3084.3 Da, observed 3084.4 Da. (C) C18 analytical RP-HPLC chromatogram of purified CoREST1-SIM(3A) peptide, gradient 0-73% B, 30 min. (D) ESI-MS of purified CoREST1-SIM(3A) peptide. Calculated m/z $[M+H]^+$ 2,972.2 Da, observed 2,972.4 Da. * = Fragmentation during ionization.

2.2.4 ^{15}N-heteronuclear single quantum correlation (HSQC) NMR titration of CoREST1-SIM peptides to characterize SUMO binding

NMR measures the chemical shifts of heteronuclear atoms such as ^{15}N, giving a characteristic chemical shift based on the electronic environment experienced by the nucleus. NMR is a powerful tool for measuring weak-to-medium protein-protein interactions due to fast exchange processes leading to distinct chemical environments. Typically, a protein of interest is expressed in minimal media containing isotopically enriched ^{15}N,^{13}C and ^{2}H atoms. First, a ligand-free spectrum of the labeled protein is collected and resonances are assigned in the 2-dimensional

HSQC spectrum to each amino acid within the protein. Upon the addition and binding of an unlabeled ligand to a sample, the chemical shifts of specific residues that participate in binding are perturbed. Changes in the 2D-HSQC spectrum are calculated as Chemical Shift Perturbations (CSPs).[43] These perturbations can be of several types; the HSQC cross-peak may grow or shrink in intensity; disappear completely; or shift to a new position in the spectrum. New cross-peaks may also appear that were missing in the ligand-free spectra. The type and degree of CSP informs on the nature of ligand-binding, and helps identify key interacting residues and surfaces on the receptor protein.

I (along with P. M. M. Shelton, Chatterjee Lab) collaborated with Dr. Tobias Ritterhoff from the lab of Dr. Rachel Klevit (Department of Biochemistry, University of Washington) and Dr. Rajan Paranji (NMR Facility, Department of Chemistry, UW) to conduct HSQC experiments. SUMO3 was expressed and purified by Dr. Ritterhoff as an $^{15}N/^{13}C/^{2}H$-labeled SUMO3. The CoREST1-SIM peptide was serially titrated against the $^{15}N/^{13}C/^{2}H$-SUMO3 from 1 to 24 molar equivalents, and spectra were collected at various points (**Figure 2.9**). Resonances were assigned to the SUMO3 spectrum based on the solved NMR solution spectrum of SUMO2 from the Biological Magnetic Resonance Data Bank (Entry 25577). This spectrum contained well-dispersed resonances across both dimensions, characteristic of a structured protein in solution. Upon the addition of the CoREST1-SIM peptide, some but not all resonances, experienced CSP suggesting that only specific residues within SUMO interact with the non-canonical SIM. We identified the top eleven residues that experienced the strongest CSP and plotted the change in CSP against the concentration of CoREST1-SIM peptide to generate binding isotherms (**Figure 2.S3**). A K_d of 1.98 ± 0.13 mM was calculated across all residues (**Table 2.1**). This affinity between SUMO3 and CoREST1-SIM peptide is relatively weak and provides an explanation for our inability to reach saturated binding in the fluorescence anisotropy experiments. This binding affinity is significantly weaker than previously observed with canonical SIM peptides, which are typically in the 1-200

μM range for either SUMO1 or SUMO2/3 binding to SIMs.[35,44] This significant difference in binding affinity may be due to the non-consensus sequence of the hydrophobic core, or residues downstream of the core, which are missing in the peptide I used and may be important toward tighter binding of SUMO to CoREST.

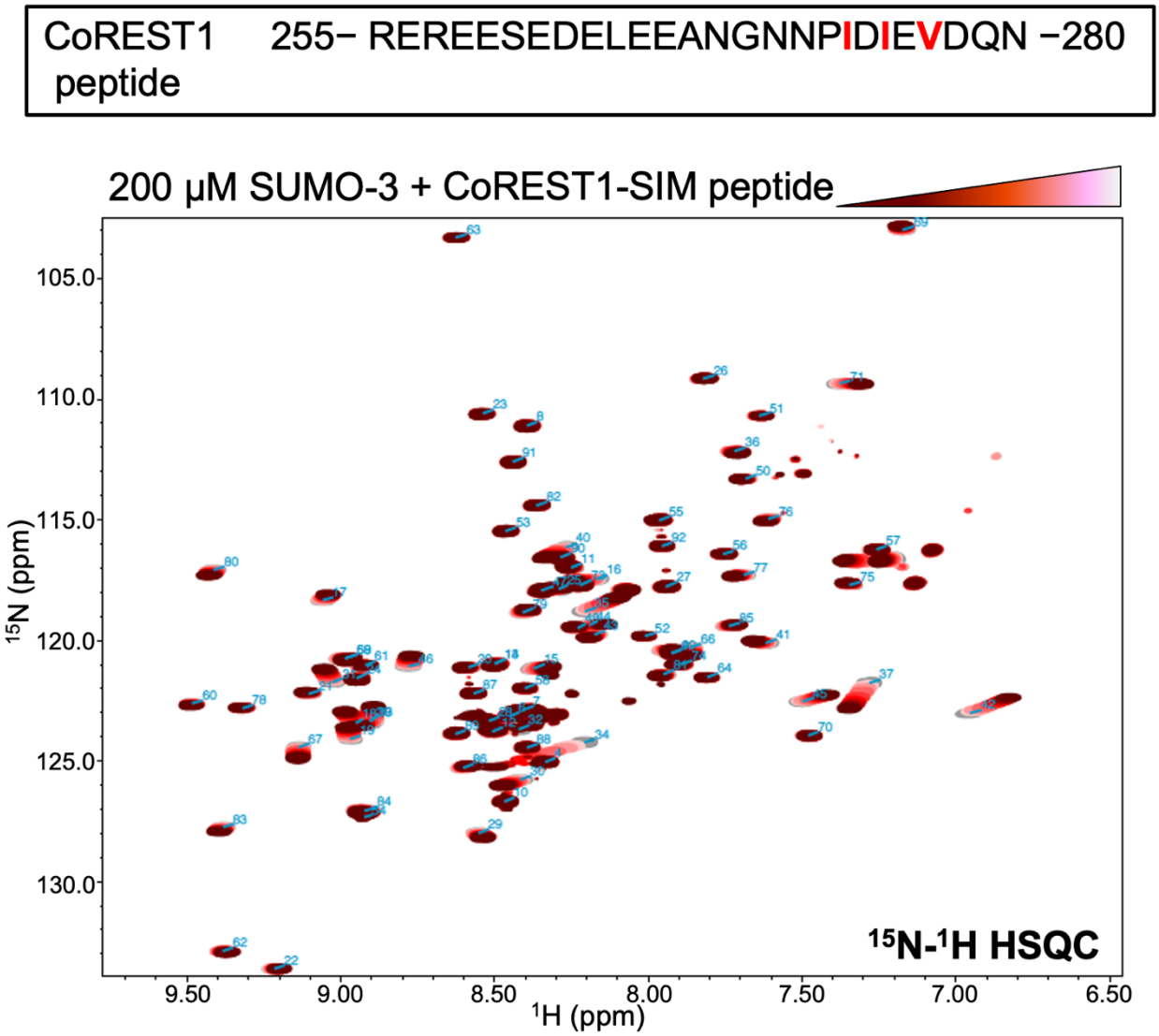


Figure 2.9. CoREST1-SIM peptide binds SUMO3 specifically. *Above* Sequence of CoREST1-SIM peptide. The hydrophobic core residues are in red. *Below* Overlaid HSQC NMR spectra of $^{15}N/^{2}H/^{13}C$-SUMO3 titrated with increasing amounts of CoREST1-SIM peptide.

Crucially, however, the CoREST1-SIM(3A) peptide also showed CSPs of the same SUMO residues but had an estimated K_d >20 mM (**Figure 2.10A, 2.S4, Table 2.1**). This strongly suggests that the two peptides bind to the same site on SUMO3 and that mutation of the hydrophobic core within the CoREST1 SIM significantly diminishes binding to SUMO3. Mutational analysis of the acidic patch in CoREST1-SIM remains to be performed, and will reveal the contribution of several glutamates in this patch toward binding SUMO3. We additionally conducted counter-titration experiments with unlabeled SUMO in the presence of $^{15}N/^{13}C/^{2}H$-SUMO3 and CoREST1-SIM peptide to demonstrate specific binding of the peptide to SUMO (**Figure 2.S5**). While we observed CSPs move in the opposite direction back towards the chemical shifts of $^{15}N/^{13}C/^{2}H$-SUMO3 alone, previously unperturbed residues also experienced some CSP. We presently attribute these new residue CSPs to the high concentration of unlabeled free SUMO3 required to compete with the weak $^{15}N/^{13}C/^{2}H$-SUMO3-non-canonical SIM interaction, which likely leads to non-physiological SUMO3-SUMO3 interactions (**Figure 2.S6**).

Table 2.1. Dissociation constants for SIM peptides against SUMO3 in HSQC NMR experiments.

Peptide	K_d (mM)
CoREST1-SIM	1.98 ± 0.13
CoREST1-SIM(3A)	>20
FIP1L1-SIM	3.57 ± 0.18
FIP1L1-SIM(3A)	>12

To further investigate if noncanonical SIMs typically bind to SUMO2/3 with such weak affinities, peptides of the FIP1L1 protein, which was also identified to contain a noncanonical SIM as COREST1 does, were synthesized by P. M. M. Shelton.[13] We used this peptide in similar HSQC-based titration experiments and observed that, indeed, the FIP1L1-non-canonical SIM peptide also bound SUMO with a K_d in the low millimolar range, and the FIP1L1-3Ala mutation significantly hindered its binding to SUMO3 (**Table 2.1**). This suggests that noncanonical SIM peptides, taken

out of the protein context, bind SUMO quite weakly yet highly specifically under our assay conditions. Interestingly, FIP1L1 does not contain an extended acidic patch (only 2 residues, ED, C-terminal to hydrophobic patch), which further suggests that sequences flanking the SIM may also be critical for SUMO-noncanonical SIM interactions.

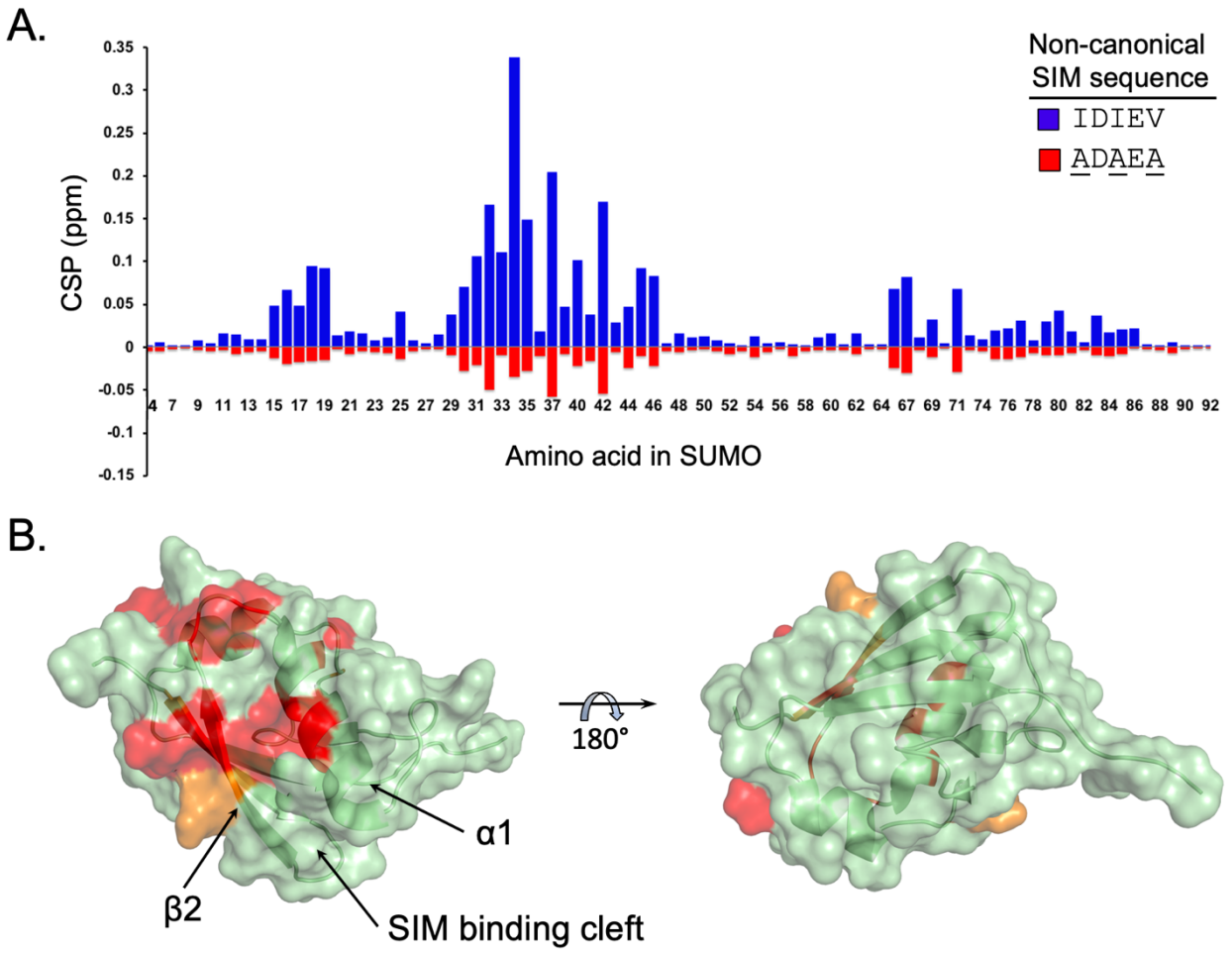


Figure 2.10. CoREST1-SIM binds to the same cleft on SUMO as consensus SIMs. (A) Plot of the degree of CSP versus amino acid residue within SUMO3 for both CoREST1-SIM in blue and CoREST1-SIM(3A) in red. The CSP values for the 3A peptide are plotted on the negative axis to mirror WT and are not negative value CSPs. (B) Transparent surface representation of SUMO3 (PDB: 1U4A) over the cartoon representation to indicate secondary structures. Residues experiencing the highest CSP are colored on the surface in red and intermediate CSPs are in orange.

Lastly, we mapped the highest CSPs onto the structure of SUMO3(C47S) (**Figure 2.10B**). Consistent with the binding site for typical consensus-SIMs between β2 and α1, the non-canonical

CoREST1-SIM also interacted with SUMO3 residues in this cleft. At the least, no residues were strongly perturbed by potential CoREST1-SIM binding on other surfaces of SUMO3. Thus, the CoREST1-SIM peptide likely binds the same highly conserved cleft in SUMO3 and may form a transient β-strand upon SUMO-binding. Our results suggest that the non-canonical hydrophobic core motif in the CoREST SIM is tolerated at the canonical SIM-binding groove in SUMO, and that there is significant leeway in potential non-canonical SIM sequences. Further experiments will be necessary to determine the orientation (parallel or antiparallel) of the CoREST1 and FIP1L1-SIM peptides, although CoREST1-SIM may be predicted to bind antiparallel based on the N-terminal acidic patch.

2.3 Conclusion and outlook

The LCH complex is an important chromatin-modifying complex and repressor of transcription through the activities of its HDAC1 and LSD1 subunits. The Chatterjee group has applied protein semisynthesis to generate site-specifically modified histones that were reconstituted into nucleosomes for mechanistic studies of the LCH complex components. We found that LSD1 alone, or in complex with CoREST1, is capable of demethylating H3K4me2 in short peptides, the full-length H3 protein, and nucleosome substrates.[29] Additionally, nucleosomes sumoylated at H4Lys12 and containing H3K4me2 are also efficient substrates for the LSD1-CoREST1 subcomplex due to the stimulation of demethylase activity by H4K12su.[28] I hypothesized that the efficient positioning of the LSD1 active site on nucleosomes by the SUMO-CoREST SIM interaction likely underlies the observed stimulation of LSD1 activity. Recent structural investigations into the LCH complex have revealed X-ray crystallographic[45] and cryo-EM structures[46] of the LSD1-CoREST1 sub-complex alone and when bound to a mononucleosome, respectively. One interesting observation is the positioning of the AOD domain of LSD1 ~100 Å away from the nucleosome core, bound to linker DNA.[45] Despite its distal positioning from the H3

tail, LSD1 demethylates H3K4me1/2, and its activity is inhibited by the mutation of important DNA-contacting residues in LSD1. One caveat for these studies is the absence of HDAC1 and extensive portions of CoREST1 which may provide additional contacts that alter LSD1 binding to the nucleosome.

Schwabe and co-workers showed that LSD1 and HDAC1 in the LCH complex are coupled in their activities by using NMR assays with H3 peptide substrates.[46] Future studies of the kinetic activities of the LCH complex and the coupling of HDAC1 and LSD1 activities on nucleosomal substrates will be necessary, as the nucleosome provides increased contact surfaces and potential constraints on LCH-binding. The observed coupling of HDAC1 and LSD1 activities does corroborate observations that H3K9ac inhibits demethylation of H3K4me by LSD1.[47] Importantly, none of these structural studies thus far have elucidated the CoREST1-SIM within their structures. However, this missing information is consistent with the observation that SIMs are often unstructured and only their interaction with SUMO results in the induced adoption of a β-sheet structure. Thus, continued interrogation of the SUMO2/3 isoform-specific binding of the noncanonical SIM in CoREST1 as well as structural characterization of LCH on a sumoylated nucleosome will provide new insights on the regulation of LCH activity in chromatin. Although we found that the noncanonical SIM in CoREST1 binds SUMO3 weakly in our NMR experiments with the CoREST1-SIM peptide, this interaction adds to the already established CoREST SANT2 domain-DNA interaction that anchors LSD1 to nucleosomes and likely plays a fine-tuning role. Furthermore, the hydrophobic core of the noncanonical SIM is clearly important for SUMO3 binding within the same cleft on SUMO where other SIMs are known to bind. Interestingly, several mutations have been observed within and proximal to the CoREST1 SIM associated with various cancers of the stomach, pancreas, and gall bladder, in the Catalogue of Somatic Mutations in Cancer (**Figure 2.11**). Such mutations present new targets for investigating the role of histone sumoylation in the etiology of human cancers.

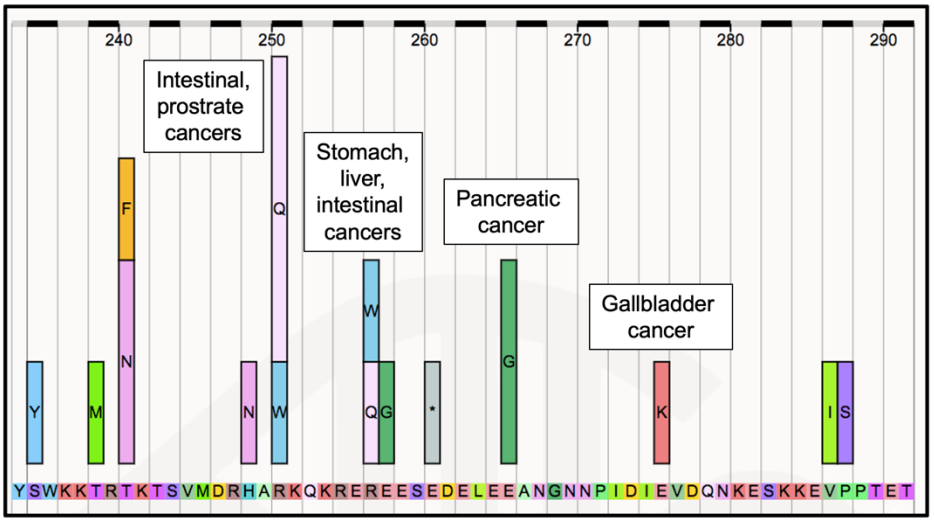


Figure 2.11. Somatic mutations within and around the CoREST1 SIM have been observed in various cancers. Chart of the CoREST1-SIM and flanking sequences. Amino acid mutations observed in caners and incidences (size of bar) within the sequence are shown above, including their associated origin. *Source: COSMIC*

2.4 Experimental procedures

2.4.1 General methods

Proteins and peptides were analyzed (4.6 x 150 mm, 5 µm) and purified (22 x 250 mm, 15 - 20 µm) with C4 and C18 reverse-phase HPLC columns from Vydac (Deerfiled, IL) on either a Varian Prostar (Palo Alto, CA) or Agilent (Santa Clara, CA) 1260 Infinity II LC system. The mobile phase consisted of Buffer A (0.1% trifluoracetic acid in water), and Buffer B (90% acetonitrile in water, 0.1% trifluoroacetic acid). UV-vis profiles of eluting peptides/proteins were monitored at 214 and 280 nm. Rink amide resin was purchased from Chem-Impex (Wood Dale, IL). Standard FMOC-L-amino acids were purchased from AGTC Bioproducts (Willmington, MA) or AnaSpec (Fremont, CA). All other chemicals were purchased form Sigma-Aldrich Chemical Company (St. Loius, MO) or Fisher Scientific (Pittsburgh, PA). DNA synthesis and gene sequencing were performed by Integrated DNA Technologies (Coralvill, IA) and Eurofins Genomics (Louisville, KY), respectively. Plasmid miniprep, PCR purification, and gel extraction kits were purchased from Qiagen

(Valencia, CA). HisPur Ni-NTA resin was purchased from Thermo Scientific (Waltham, MA). Routine peptide/protein mass spectrometry was performed by direct infusion on a Bruker (Billerica, MA) Esquire ion-trap mass spectrometer operating in positive mode. Circular dichroism measurements were performed on a JASCO (Easton, MD) J-720 spectropolarimeter. Fluorescence plate-based experiments were scanned on an EnVision (PerkinElmer, Waltham, MA) 2105 multimode plate reader.

2.4.2 Molecular cloning of SUMO3ΔG(K11C, C47S)-OH, CoREST1(1-300), and CoREST1(1-300)-3A

The plasmid pTXB1-SUMO3ΔG(C47S)-AvaDNAE-AAFN-His_6 containing the human SUMO3 gene, *SUMO3*, with a C47S mutation and lacking the terminal Gly, was used to generate the plasmid pTXB1-SUMO3ΔG(K11C, C47S)-AvaDNAE-AAFN-His_6, which mutates Lys11 to Cys. The modified pTXB1 plasmid was prepared by site-directed mutagenesis (Q5 Site-directed mutagenesis kit, NEB, Ipswitch, MA) using the original plasmid as the template in PCR using the plasmid below. The plasmid pET28b-His_6-CoREST1 containing the human CoREST1 gene, *RCOR1*, was used to generate pET28b-His6-CoREST1(1-300), which truncates the CoREST1 sequence after residue 300. The plasmid pET28b-His_6-CoREST1(3A) which contains Ala mutations at positions I273, I275, and V277 was used to generate pET28b-His_6-CoREST1(1-300)-3A. PCR attempts using pET28b-CoREST1 failed over several iterations of primer design and screening of conditions. The CoREST1 gene sequence may be too difficult for entire plasmid PCR amplification. Thus, we opted to use restriction enzymes to cut the plasmid at unique sites near the encoded amino acid residue 300 and insert a small dsDNA oligo containing a 5' sticky end corresponding to the first restriction enzyme, the missing CoREST1 codons up to residue 300, a stop codon, and the corresponding 3' sticky end for the second enzyme. This method avoids PCR amplification and allows the amplification of the plasmid solely in *E. coli*. The restriction enzymes Bsu36I and HindIII were used. The plasmids and ssDNA oligos (**Table 2.2**)

were ordered from IDT (Coralville, IA). The ssDNA oligos were annealed and gel extracted before use in T4 DNA ligase (NEB) reactions with the double cut pET28b-His_6-CoREST1 plasmid and transformation of *E. coli* XL10Gold ultracompetent cells (Agilent Technologies, Santa Clara, CA). The same oligos were used to generate pET28b-His_6-CoREST1(1-300)-3A using double cut pET28b-His_6-CoREST1(3A) as the vector. The desired sequences were confirmed by DNA sequencing using T7 and T7-reverse primers.

Table 2.2. List of ssDNA primers and oligos used for molecular cloning.

Primer	**DNA Sequence (5'- to -3')**
SU3K11C_F	GGA AGG AGT GTG CAC AGA GAA TGA CC
SU3K11C_R	TTG GGC TTT TCT TCG GAC
Oligo	**DNA Sequence (5'- to -3')**
CoREST1-frag-A	TCA GGT CAA GAA AGA GAA GCA TAG CAC CTA A
CoREST1-frag-a	AGC TTT AGG TGC TAT GCT TCT CTT TCT TGA CC
CoREST1-frag-B	TCA GGT CAA AAA AGA AAA ACA TAG CAC ATA A
CoREST1-frag-b	AGC TTT ATG TGC TAT GTT TTT CTT TTT TGA CC

2.4.3 Expression and purification of SUMO3ΔG(K11C, C47S)-OH

Six liters of LB medium supplemented with 100 µg/mL ampicillin were inoculated with *E. coli* BL21(DE3) cells transformed with the plasmid pTXB1-SUMO3ΔG(K11C, C47S)-Ava-AAFN-His_6. The cells were grown at 37 °C, shaking at 250 rpm until an OD600 ~0.6-0.8. Protein overexpression was then induced with 0.5 mM IPTG and cells were grown at 16 °C overnight (~18 h). Cells were collected by centrifugation (7 kRCF at 4 °C) the next morning and the cell pellet was either flash frozen in liquid nitrogen (LN2) or lysed in buffer containing 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 5 mM imidazole. Cells were pulse sonicated and clarified by sonication at 20 kRCF for 15 min at 4 °C. The supernatant was filtered through a 0.45 µm PES filter and the applied to 10 mL of Ni-NTA resin. Protein was bound to the column for 1 h at 4 °C,

nutating. The flow through was collected and the resin was washed with lysis buffer containing increasing amounts of imidazole from 5 mM (5 column volumes, CV), 20 mM (2 CV), 50 mM (2 CV), and 250 mM (2.5 CV). Fractions were analyzed by SDS-PAGE and fractions containing protein were pooled and dialyzed twice into 2 L 100 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, 1 mM sodium mercaptoethanesulfonate, pH 7.2, 4 °C, for 1 h, then overnight. The dialyzed protein was then supplemented with DTT to 100 mM and nutated at 30 °C for 18 h to thiolyze and hydrolyze off the AVA-intein. The hydrolysis reaction was lyophilized and purified by C18 preparative RP-HPLC employing a gradient of 5-75 % B over 60 min. Fractions containing pure Su3ΔG(K11C, C47S)-OH, identified by ESI-MS and pooled together. Typical yields were ~5 mg/L of cell culture. ESI-MS for Su3ΔG(K11C, C47S)-OH. Calculated *m/z* $[M+H]^+$ 10295.5 Da, observed 10295 ± 1.2 Da (**Figure 2.S1A,B**).

2.4.4 Expression and purification of His6-CoREST1(1-300) and 3A mutant

Two liters of 2xYT media were supplemented with 50 µg/mL kanamycin and inoculated with *E. coli* BL21(DE3) transformed with the pET28b-His_6-CoREST1(1-300) or 3Ala mutant and grown at 37 °C until OD600 ~0.6-0.8. Protein expression was induced with 0.3 mM IPTG and continued to grow overnight at 16 °C (~18 h). The next morning, cells were collected by centrifugation at 7000 kRCF and resuspended in lysis buffer (50 mM tris, pH 8, 500 mM NaCl, 10 % glycerol by French press. The lysate was clarified by centrifugation at 20 kRCF and passed through a 0.45 µm PES filter before application to 4 mL Ni-NTA resin. Protein was bound to the resin for 1 h, nutating at 4 °C. The resin was washed with increasing amounts of imidazole in lysis buffer: 25 mM (10 CV), 50 mM (10 CV), 100 mM (5 CV), 250 mM (2 CV), and 300 mM (6 CV). Fractions were analyzed by SDS-PAGE and fractions containing protein were pooled and dialyzed twice against 2 L 50 mM HEPES, pH 8, 250 mM NaCl, 10 % glycerol, for 1 h and then overnight at 4 °C. Protein was

concentrated and aliquoted into single use volumes. Samples were flash frozen in LN2 and kept at -80 °C until use.

2.4.5 Fluorescence anisotropy binding assay with SUMO3ΔG(K11C, C47S)-AlexaFluor488 and CoREST1(1-300)/3A

Dried Su3ΔG(K11C, C47S) was conjugated to Alex Fluor 488 C_5 maleimide (Invitrogen) according to manufacturer's protocol. The conjugated Su3ΔG-AF488 product was purified by C18 analytical RP-HPLC and analyzed by ESI-MS. Calculated *m/z* $[M+H]^+$ 10993.2 Da, observed 10993.9 ± 1.6 Da (**Figure 2.S1C/D**). Lyophilized Su3ΔG-AF488 was resuspended in assay buffer 50 mM HEPES (pH 8), 150 mM NaCl, 0.1 % Tween-20, and 0.6 mM dithiothreitol. Su3ΔG-AF488 was compared with Su3ΔG(C47S)-OH by circular dichroism to ensure labeling did not disrupt SUMO structure. Purified CoREST1(1-300) or CoREST1(1-300)-3A was dialyzed into assay buffer. Samples of fixed Su3ΔG-AF488 concentration at 50 nM and serially diluted CoREST1 were prepared in a black 384-well microplate and scanned on an EnVision microplate reader after incubation for 30 min, covered on ice.

2.4.6 Solid-phase peptide synthesis of CoREST1-SIM and CoREST1-SIM(3A) peptides

The peptides AcHN-REREESEDELEEANGNNPIDIEVDQN-C(O)NH_2 (CoREST1-SIM) and AcHN-REREESEDELEEANGNNPADAEADQN-C(O)NH_2 (CoREST1-SIM(3A)) were synthesized by microwave-assisted SPPS on a 0.05 mmol scale employing a 5 % piperazine, 0.1 M HOBt in DMF Fmoc-based N^{α}-deprotection chemistry for 2 min at 50 °C. Rink-amide resin (0.33 mmol/g) each amino acid was coupled in 5 molar excess based on resin loading. Coupling reactions were for 10 min at 50 °C with a mixture of Fmoc amino acid (0.25 mmol), oxyma (0.25 mmol), and DIC (0.25 mmol) in DMF. Arg, Val, and Ile residues were double coupled. Removal of the final Fmoc-protecting group proceeded on the machine. *N*-termini were acetylated by incubating the resin

with 40 mol eq DIEA and 20 mol eq acetic anhydride in 50:50 DMF:DCM, twice for 15 min. Peptide cleavage from the resin and global deprotection was carried out at 100 µL/mg resin with cleavage cocktail (92.5 TFA: 2.5 H_2O: 2.5 TIS: 2.5 thioanisole) for 3 h, shaking at RT. Peptide was precipitated and washed twice in cold diethyl ether. Dried peptides were purified by preparative RP-HPLC employing a 10-50% B gradient over 60 min. Fractions containing pure CoREST1-SIM peptide were identified by ESI-MS, pooled together, and lyophilized to dryness. FIP1L1 peptides were synthesized by P.M.M. Shelton.

2.4.7 HSQC-NMR titration experiments with CoREST1-SIM/3A or FIP1L1-SIM/3A and 15N/13C/2H-SUMO3

15N/13C/2H-SUMO3 was expressed and purified by Dr. Tobias Ritterhoff. NMR experiments were carried out by T. Ritterhoff and Dr. Rajan Paranji. Peptide synthesis and sample preparations were carried out by C.J.A. Leonen and P.M.M. Shelton. In brief, lyophilized peptide and SUMO were dissolved in 10 mM potassium phosphate (pH 6.5), 100 mM KCl, 2 mM DTT (added fresh), 0.1 mM EDTA, 10 % D_2O. SUMO was held at 200 µM. 1-24 equivalents of SIM peptide were added to the SUMO sample, along with additional SUMO from stock to maintain 200 µM. Counter titration experiments were conducted with the addition of unlabeled-SUMO. Samples prepared in a 3 mm high field NMR tube (Norell Inc., Morganton, NC) were run on a Bruker (Billerica, MA) Avance III 800 MHz spectrometer equipped with a triple resonance cryoprobe (TCI-HCN-Z gradient) at 290 K. Data were processed and visualized in NMRviewJ.[48]

2.5 Product characterization and supplemental data

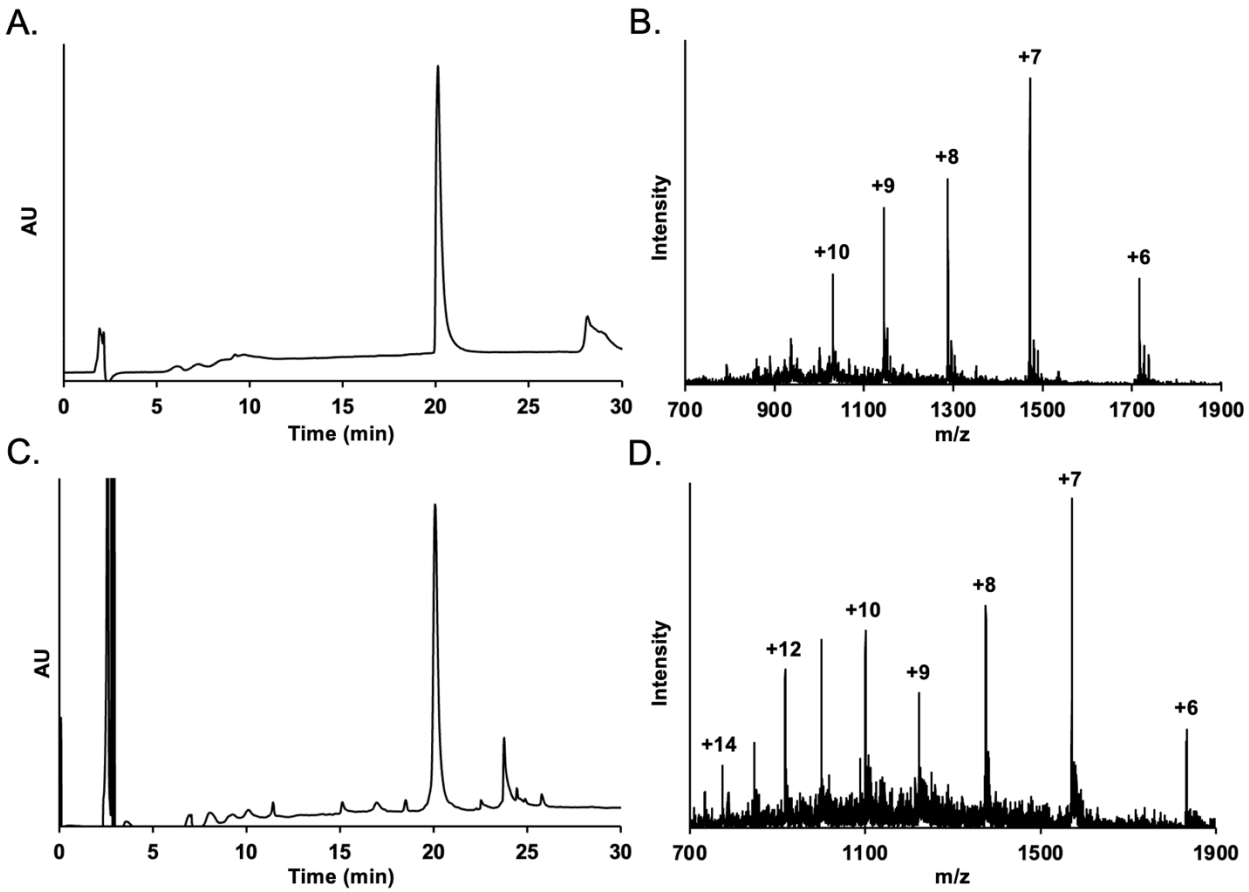


Figure 2.S1. Purification of SUMO-3ΔG(K11C, C47S)-AlexaFluor488. (A) C18 analytical RP-HPLC chromatogram of purified SUMO-3ΔG(K11C, C47S), gradient 0-73% B, 30 min. (B) ESI-MS of purified SUMO-3ΔG(K11C, C47S). (C) C18 analytical RP-HPLC chromatogram of purified SUMO-3ΔG(K11C, C47S)-AlexaFluor488, gradient 0-73% B, 30 min. (D) ESI-MS of purified SUMO-3ΔG(K11C, C47S)-AlexaFluor488.

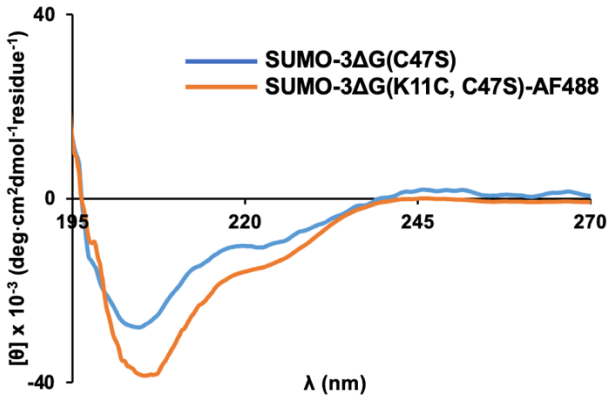


Figure 2.S2. SUMO-3ΔG(K11C, C47S)-AlexaFluor488 structure is similar to SUMO-3ΔG(C47S) by circular dichroism. Overlaid CD spectra of SUMO-3ΔG(K11C, C47S)-AlexaFluor488 and SUMO-3ΔG(C47S).

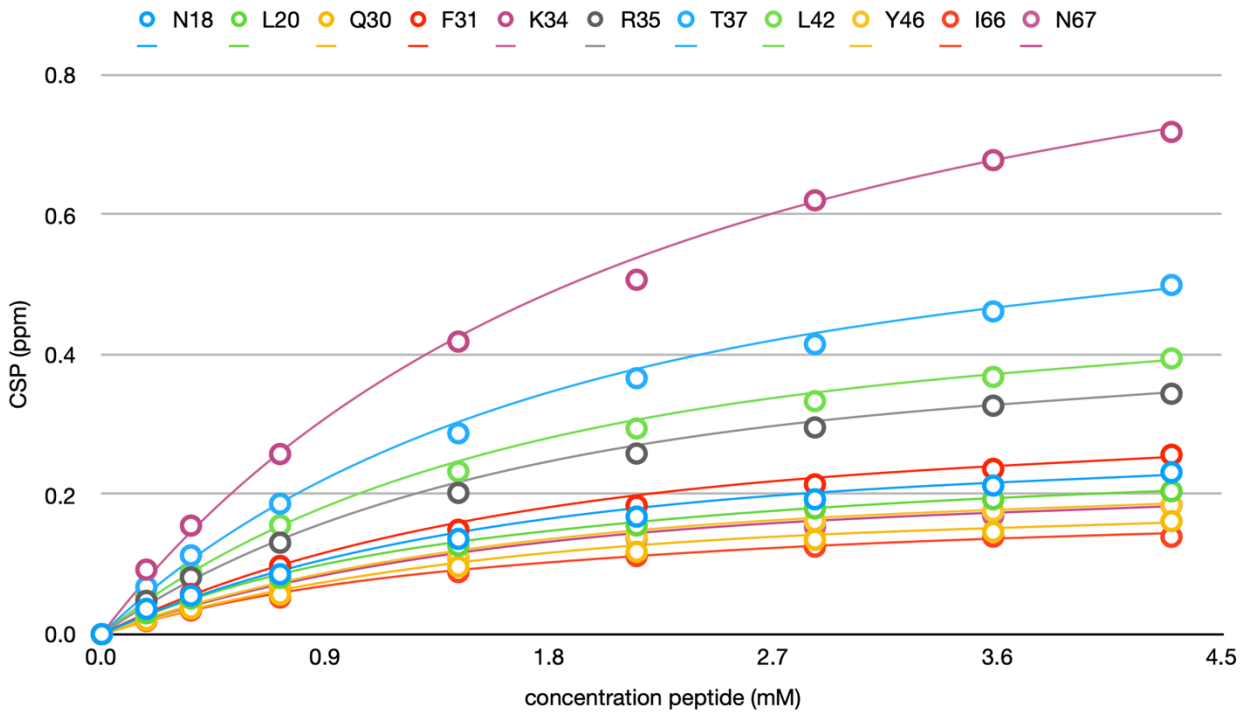
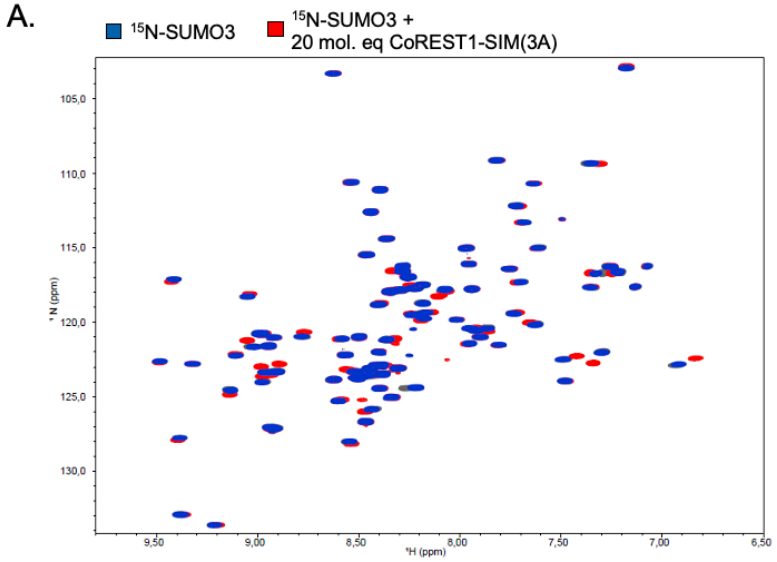
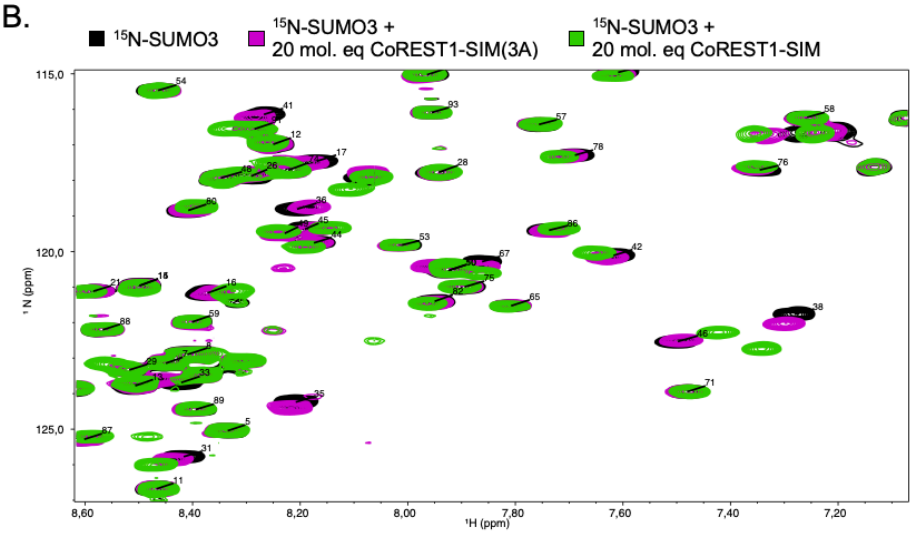


Figure 2.S3. Plot of CSP versus CoREST1-SIM peptide. Chemical shift perturbations (CSPs) were plotted against CoREST1-SIM peptide concentration used in HSQC NMR experiments for the eleven most shifted residues in SUMO-3.

A.


15N-SUMO3
15N-SUMO3 +
20 mol. eq CoREST1-SIM(3A)
105,0
110,0
115,0
120,0
125,0
130,0
15N (ppm)
9,50
9,00
8,50
8,00
7,50
7,00
6,50
1H (ppm)

B.


15N-SUMO3
15N-SUMO3 +
20 mol. eq CoREST1-SIM(3A)
15N-SUMO3 +
20 mol. eq CoREST1-SIM
115,0
120,0
125,0
15N (ppm)
8,60
8,40
8,20
8,00
7,80
7,60
7,40
7,20
1H (ppm)

C.

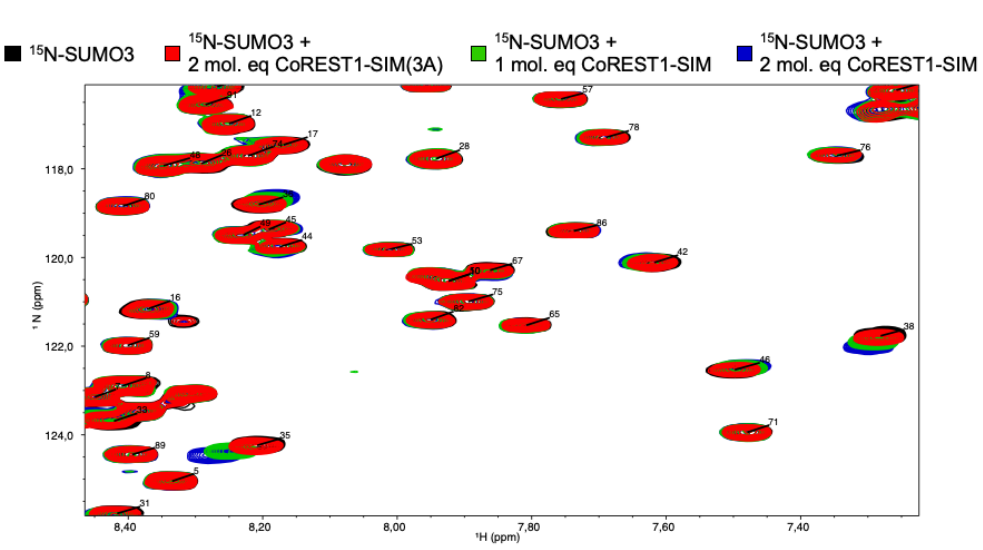
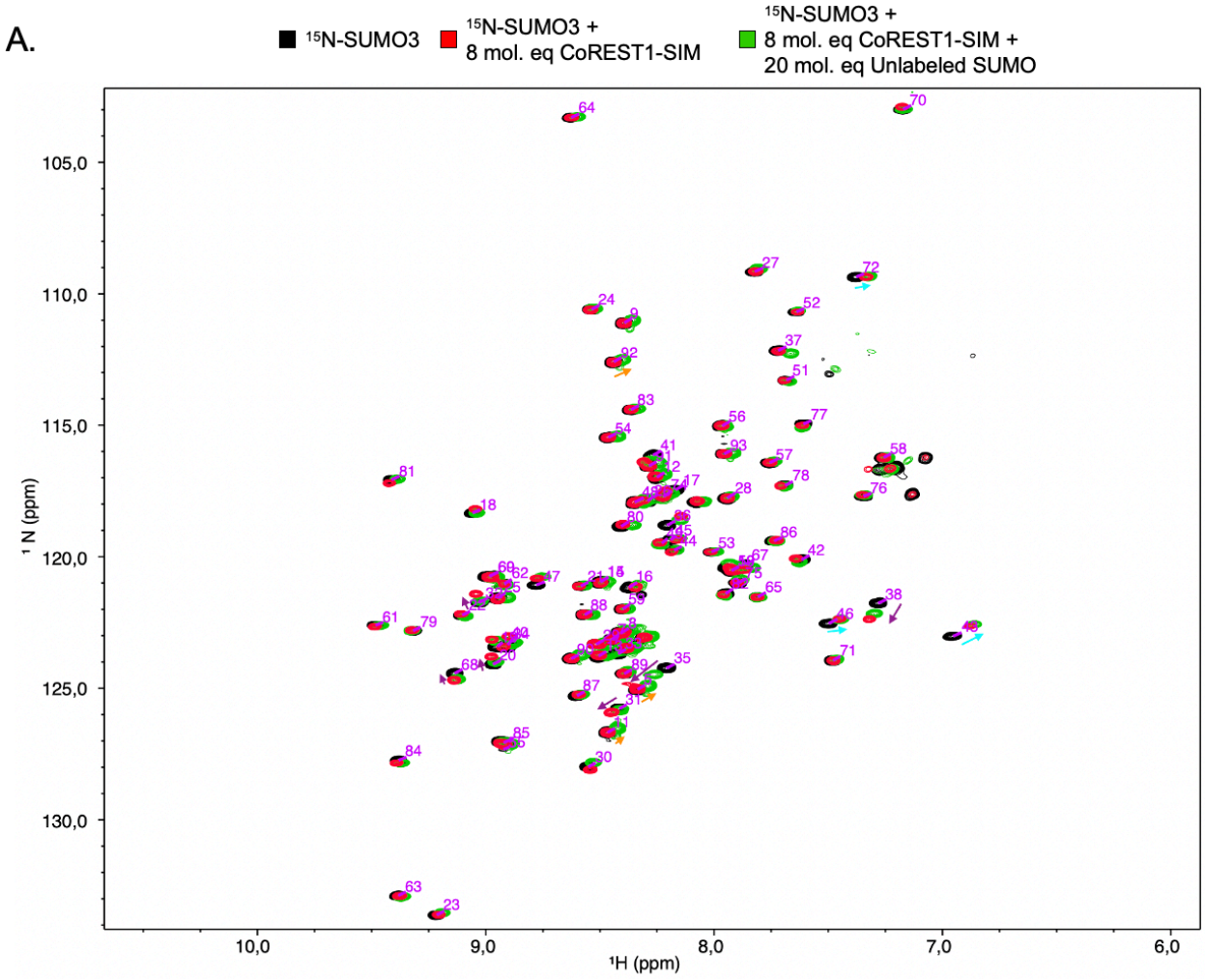

15N-SUMO3
15N-SUMO3 +
2 mol. eq CoREST1-SIM(3A)
15N-SUMO3 +
1 mol. eq CoREST1-SIM
15N-SUMO3 +
2 mol. eq CoREST1-SIM
118,0
120,0
122,0
124,0
15N (ppm)
8,40
8,20
8,00
7,80
7,60
7,40
1H (ppm)

Figure legend for figure shown above (p. 60).

Figure 2.S4. CoREST1-SIM3A binds weakly to SUMO-3. (A) Overlaid spectra of 20 molar equivalents of CoREST1-SIM(3A) peptide incubated with ^{15}N-SUMO3, or ^{15}N-SUMO3 only. (B) Zoomed-in overlaid spectra of ^{15}N-SUMO3 ± 20 molar equivalents of CoREST1-SIM or CoREST1-SIM(3A) peptide for comparison of CSPs. (C) Zoomed-in overlaid spectra of ^{15}N-SUMO3 ± CoREST1 peptide at 1 and 2 equivalents or CoREST1-SIM(3A) peptide at 2 equivalents to demonstrate weaker CSPs from the 3A mutant.


A.
15N-SUMO3
15N-SUMO3 +
8 mol. eq CoREST1-SIM
15N-SUMO3 +
8 mol. eq CoREST1-SIM +
20 mol. eq Unlabeled SUMO
105,0
110,0
115,0
120,0
125,0
130,0
¹ N (ppm)
10,0
9,0
8,0
7,0
6,0
¹H (ppm)

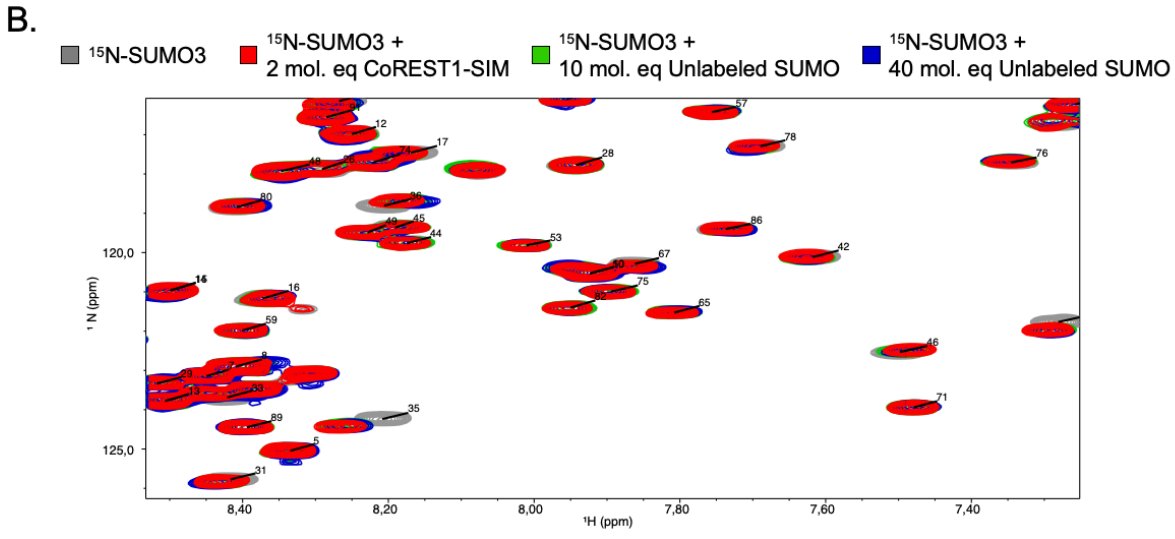

B.
15N-SUMO3
15N-SUMO3 +
2 mol. eq CoREST1-SIM
15N-SUMO3 +
10 mol. eq Unlabeled SUMO
15N-SUMO3 +
40 mol. eq Unlabeled SUMO
120,0
125,0
¹ N (ppm)
8,40
8,20
8,00
7,80
7,60
7,40
¹H (ppm)

Figure legend for figure shown above (p. 62).

Figure 2.S5. Counter titration experiment with unlabeled SUMO. (A) Overlaid spectra of ^{15}N-SUMO3, ^{15}N-SUMO3 + 8 equivalents of CoREST1-SIM, and ^{15}N-SUMO3 + 8 eq. CoREST1-SIM + 20 eq of unlabeled SUMO. Peptide and unlabeled SUMO were pre-incubated before addition of ^{15}N-SUMO3. The signature peaks that we see shifting upon peptide titration, shift back upon SUMO counter-titration (see purple arrows). Peaks that don't shift upon peptide titration but do shift upon additional SUMO titration (orange arrows). Peaks that shift upon peptide titration, but don't shift back upon SUMO counter-titration (cyan arrows). (B) Zoomed-in overlaid spectra of increasing amounts of unlabeled SUMO used to compete for CoREST1-SIM binding at 2 equivalents.

2.6 References

(1) You, A.; Tong, J. K.; Grozinger, C. M.; Schreiber, S. L. CoREST Is an Integral Component of the CoREST- Human Histone Deacetylase Complex. *Proc. Natl. Acad. Sci.* **2001**, *98* (4), 1454–1458. https://doi.org/10.1073/PNAS.98.4.1454.

(2) Lee, M. G.; Wynder, C.; Cooch, N.; Shiekhattar, R. An Essential Role for CoREST in Nucleosomal Histone 3 Lysine 4 Demethylation. *Nat. 2005 4377057* **2005**, *437* (7057), 432–435. https://doi.org/10.1038/nature04021.

(3) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **2004**, *119* (7), 941–953. https://doi.org/10.1016/j.cell.2004.12.012.

(4) Lv, T.; Yuan, D.; Miao, X.; Lv, Y.; Zhan, P.; Shen, X.; Song, Y. Over-Expression of LSD1 Promotes Proliferation, Migration and Invasion in Non-Small Cell Lung Cancer. *PLoS One* **2012**, *7* (4), e35065. https://doi.org/10.1371/JOURNAL.PONE.0035065.

(5) N, S.; A, G.; S, S.; H, L.; J, K.; R, B. Elevated Expression of LSD1 (Lysine-Specific Demethylase 1) during Tumour Progression from Pre-Invasive to Invasive Ductal Carcinoma of the Breast. *BMC Clin. Pathol.* **2012**, *12*. https://doi.org/10.1186/1472-6890-12-13.

(6) Kahl, P.; Gullotti, L.; Heukamp, L. C.; Wolf, S.; Friedrichs, N.; Vorreuther, R.; Solleder, G.; Bastian, P. J.; Ellinger, J.; Metzger, E.; Schüle, R.; Buettner, R. Androgen Receptor Coactivators Lysine-Specific Histone Demethylase 1 and Four and a Half LIM Domain Protein 2 Predict Risk of Prostate Cancer Recurrence. *Cancer Res.* **2006**, *66* (23), 11341–11347. https://doi.org/10.1158/0008-5472.CAN-06-1570.

(7) Schenk, T.; Chen, W. C.; Göllner, S.; Howell, L.; Jin, L.; Hebestreit, K.; Klein, H.-U.; Popescu, A. C.; Burnett, A.; Mills, K.; Casero, R. A.; Marton, L.; Woster, P.; Minden, M. D.; Dugas, M.; Wang, J. C. Y.; Dick, J. E.; Müller-Tidow, C.; Petrie, K.; Zelent, A.

Inhibition of the LSD1 (KDM1A) Demethylase Reactivates the All-Trans-Retinoic Acid Differentiation Pathway in Acute Myeloid Leukemia. *Nat. Med. 2012 184* **2012**, *18* (4), 605–611. https://doi.org/10.1038/nm.2661.

(8) Harris, W. J.; Huang, X.; Lynch, J. T.; Spencer, G. J.; Hitchin, J. R.; Li, Y.; Ciceri, F.; Blaser, J. G.; Greystoke, B. F.; Jordan, A. M.; Miller, C. J.; Ogilvie, D. J.; Somervaille, T. C. P. The Histone Demethylase KDM1A Sustains the Oncogenic Potential of MLL-AF9 Leukemia Stem Cells. *Cancer Cell* **2012**, *21* (4), 473–487. https://doi.org/10.1016/J.CCR.2012.03.014.

(9) Mohammad, H. P.; Smitheman, K. N.; Kamat, C. D.; Soong, D.; Federowicz, K. E.; Van Aller, G. S.; Schneck, J. L.; Carson, J. D.; Liu, Y.; Butticello, M.; Bonnette, W. G.; Gorman, S. A.; Degenhardt, Y.; Bai, Y.; McCabe, M. T.; Pappalardi, M. B.; Kasparec, J.; Tian, X.; McNulty, K. C.; Rouse, M.; McDevitt, P.; Ho, T.; Crouthamel, M.; Hart, T. K.; Concha, N. O.; McHugh, C. F.; Miller, W. H.; Dhanak, D.; Tummino, P. J.; Carpenter, C. L.; Johnson, N. W.; Hann, C. L.; Kruger, R. G. A DNA Hypomethylation Signature Predicts Antitumor Activity of LSD1 Inhibitors in SCLC. *Cancer Cell* **2015**, *28* (1), 57–69. https://doi.org/10.1016/J.CCELL.2015.06.002.

(10) Milazzo, G.; Mercatelli, D.; Muzio, G. Di; Triboli, L.; Rosa, P. De; Perini, G.; Giorgi, F. M. Histone Deacetylases (HDACs): Evolution, Specificity, Role in Transcriptional Complexes, and Pharmacological Actionability. *Genes (Basel).* **2020**, *11* (5). https://doi.org/10.3390/GENES11050556.

(11) Li, Y.; Seto, E. HDACs and HDAC Inhibitors in Cancer Development and Therapy. *Cold Spring Harb. Perspect. Med.* **2016**, *6* (10). https://doi.org/10.1101/CSHPERSPECT.A026831.

(12) KJ, F.; RW, J. Histone Deacetylases and Their Inhibitors in Cancer, Neurological Diseases and Immune Disorders. *Nat. Rev. Drug Discov.* **2014**, *13* (9), 673–691. https://doi.org/10.1038/NRD4360.

(13) Ouyang, J.; Shi, Y.; Valin, A.; Xuan, Y.; Gill, G. Direct Binding of CoREST1 to SUMO-2/3 Contributes to Gene-Specific Repression by the LSD1/CoREST1/HDAC Complex. *Mol. Cell* **2009**, *34* (2), 145–154. https://doi.org/10.1016/j.molcel.2009.03.013.

(14) Shi, Y. J.; Matson, C.; Lan, F.; Iwase, S.; Baba, T.; Shi, Y. Regulation of LSD1 Histone Demethylase Activity by Its Associated Factors. *Mol. Cell* **2005**, *19* (6), 857–864. https://doi.org/10.1016/j.molcel.2005.08.027.

(15) Eckschlager, T.; Plch, J.; Stiborova, M.; Hrabeta, J. Histone Deacetylase Inhibitors as Anticancer Drugs. *Int. J. Mol. Sci. 2017, Vol. 18, Page 1414* **2017**, *18* (7), 1414. https://doi.org/10.3390/IJMS18071414.

(16) Bubna, A. K. Vorinostat—An Overview. *Indian J. Dermatol.* **2015**, *60* (4), 419. https://doi.org/10.4103/0019-5154.160511.

(17) Fang, Y.; Liao, G.; Yu, B. LSD1/KDM1A Inhibitors in Clinical Trials: Advances and Prospects. *J. Hematol. Oncol. 2019 121* **2019**, *12* (1), 1–14. https://doi.org/10.1186/S13045-019-0811-9.

(18) Kalin, J. H.; Wu, M.; Gomez, A. V.; Song, Y.; Das, J.; Hayward, D.; Adejola, N.; Wu, M.; Panova, I.; Chung, H. J.; Kim, E.; Roberts, H. J.; Roberts, J. M.; Prusevich, P.; Jeliazkov, J. R.; Burman, S. S. R.; Fairall, L.; Milano, C.; Eroglu, A.; Proby, C. M.; Dinkova-Kostova, A. T.; Hancock, W. W.; Gray, J. J.; Bradner, J. E.; Valente, S.; Mai, A.; Anders, N. M.; Rudek, M. A.; Hu, Y.; Ryu, B.; Schwabe, J. W. R.; Mattevi, A.; Alani, R. M.; Cole, P. A. Targeting the CoREST Complex with Dual Histone Deacetylase and Demethylase Inhibitors. *Nat. Commun.* **2018**, *9* (1). https://doi.org/10.1038/S41467-017-02242-4.

(19) Perillo, B.; Tramontano, A.; Pezone, A.; Migliaccio, A. LSD1: More than Demethylation of Histone Lysine Residues. *Exp. Mol. Med. 2020 5212* **2020**, *52* (12), 1936–1947. https://doi.org/10.1038/s12276-020-00542-2.

(20) Chen, Y.; Yang, Y.; Wang, F.; Wan, K.; Yamane, K.; Zhang, Y.; Lei, M. Crystal Structure of Human Histone Lysine-Specific Demethylase 1 (LSD1). *Proc. Natl. Acad. Sci.* **2006**, *103* (38), 13956–13961. https://doi.org/10.1073/PNAS.0606381103.

(21) Forneris, F.; Binda, C.; Adamo, A.; Battaglioli, E.; Mattevi, A. Structural Basis of LSD1-CoREST Selectivity in Histone H3 Recognition. *J. Biol. Chem.* **2007**, *282* (28), 20070–20074. https://doi.org/10.1074/JBC.C700100200.

(22) Anand, R.; Marmorstein, R. Structure and Mechanism of Lysine-Specific Demethylase Enzymes *. *J. Biol. Chem.* **2007**, *282* (49), 35425–35429. https://doi.org/10.1074/JBC.R700027200.

(23) GW, H.; Y, W.; VR, R.; T, H.; J, Q.; Y, N.; BH, H. Stable Histone Deacetylase Complexes Distinguished by the Presence of SANT Domain Proteins CoREST/Kiaa0071 and Mta-L1. *J. Biol. Chem.* **2001**, *276* (9), 6817–6824. https://doi.org/10.1074/JBC.M007372200.

(24) P, B.; H, J.; C, W.; H, L. The Substrate Specificity of Sirtuins. *Annu. Rev. Biochem.* **2016**, *85*, 405–429. https://doi.org/10.1146/ANNUREV-BIOCHEM-060815-014537.

(25) Galisson, F.; Mahrouche, L.; Courcelles, M.; Bonneil, E.; Meloche, S.; Chelbi-Alix, M. K.; Thibault, P. A Novel Proteomics Approach to Identify SUMOylated Proteins and Their Modification Sites in Human Cells. *Mol. Cell. Proteomics* **2011**, *10* (2), M110.004796. https://doi.org/10.1074/mcp.M110.004796.

(26) Shiio, Y.; Eisenman, R. N. Histone Sumoylation Is Associated with Transcriptional Repression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230. https://doi.org/10.1073/pnas.1735528100.

(27) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long-Range Internucleosomal Interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837. https://doi.org/10.1074/jbc.M114.591644.

(28) Dhall, A.; Weller, C. E.; Chu, A.; Shelton, P. M. M.; Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-CoREST Complex. *ACS Chem. Biol.* **2017**, *12* (9), 2275–2280. https://doi.org/10.1021/acschembio.7b00716.

(29) Dhall, A.; Shelton, P. M. M.; Delachat, A. M.-F.; Leonen, C. J. A.; Fierz, B.; Chatterjee, C. Nucleosome Binding by the Lysine Specific Demethylase 1 (LSD1) Enzyme Enables Histone H3 Demethylation. *Biochemistry* **2020**, *59* (27), 2479–2483. https://doi.org/10.1021/ACS.BIOCHEM.0C00412.

(30) Kerscher, O. SUMO Junction—What's Your Function? New Insights through SUMO-Interacting Motifs. *EMBO Rep.* **2007**, *8* (6), 550. https://doi.org/10.1038/SJ.EMBOR.7400980.

(31) Kliza, K.; Husnjak, K. Resolving the Complexity of Ubiquitin Networks. *Front. Mol. Biosci.* **2020**, *0*, 21. https://doi.org/10.3389/FMOLB.2020.00021.

(32) Song, J.; Durrin, L. K.; Wilkinson, T. A.; Krontiris, T. G.; Chen, Y. Identification of a SUMO-Binding Motif That Recognizes SUMO-Modified Proteins. *Proc. Natl. Acad. Sci.* **2004**, *101* (40), 14373–14378. https://doi.org/10.1073/PNAS.0403498101.

(33) 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. Structural and Functional Roles of Daxx SIM Phosphorylation in SUMO Paralog-Selective Binding and Apoptosis Modulation. *Mol. Cell* **2011**, *42* (1), 62–74. https://doi.org/10.1016/J.MOLCEL.2011.02.022.

(34) Armstrong, A. A.; Mohideen, F.; Lima, C. D. Recognition of SUMO-Modified PCNA Requires Tandem Receptor Motifs in Srs2. *Nat. 2012 4837387* **2012**, *483* (7387), 59–63. https://doi.org/10.1038/nature10883.

(35) Hecker, C. M.; Rabiller, M.; Haglund, K.; Bayer, P.; Dikic, I. Specification of SUMO1- and SUMO2-Interacting Motifs. *J. Biol. Chem.* **2006**, *281* (23), 16117–16127. https://doi.org/10.1074/jbc.M512757200.

(36) Baba, D.; Maita, N.; Jee, J.-G.; Uchimura, Y.; Saitoh, H.; Sugasawa, K.; Hanaoka, F.; Tochio, H.; Hiroaki, H.; Shirakawa, M. Crystal Structure of Thymine DNA Glycosylase Conjugated to SUMO-1. *Nature* **2005**, *435* (7044), 979–982. https://doi.org/10.1038/nature03634.

(37) Lussier-Price, M.; Mascle, X. H.; Cappadocia, L.; Kamada, R.; Sakaguchi, K.; Wahba, H. M.; Omichinski, J. G. Characterization of a C-Terminal SUMO-Interacting Motif Present in Select PIAS-Family Proteins. *Structure* **2020**, *28* (5), 573-585.e5. https://doi.org/10.1016/J.STR.2020.04.002.

(38) Reverter, D.; Lima, C. D. Insights into E3 Ligase Activity Revealed by a SUMO-RanGAP1-Ubc9-Nup358 Complex. *Nature* **2005**, *435* (7042), 687–692. https://doi.org/10.1038/nature03588.

(39) Jing, S.; Ziming, Z.; Weidong, H.; Yuan, C. Small Ubiquitin-like Modifier (SUMO) Recognition of a SUMO Binding Motif: A Reversal of the Bound Orientation. *J. Biol. Chem.* **2005**, *280* (48), 40122–40129. https://doi.org/10.1074/JBC.M507059200.

(40) Rossi, A. M.; Taylor, C. W. Analysis of Protein-Ligand Interactions by Fluorescence Polarization. *Nat. Protoc.* **2011**, *6* (3), 365. https://doi.org/10.1038/NPROT.2011.305.

(41) Streich, F. C.; Lima, C. D. Capturing a Substrate in an Activated RING E3/E2-SUMO Complex. *Nature* **2016**, *536* (7616), 304–308. https://doi.org/10.1038/nature19071.

(42) Lea, W. A.; Simeonov, A. Fluorescence Polarization Assays in Small Molecule Screening. *Expert Opin. Drug Discov.* **2011**, *6* (1), 17. https://doi.org/10.1517/17460441.2011.537322.

(43) MP, W. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16. https://doi.org/10.1016/J.PNMRS.2013.02.001.

(44) Song, J.; Durrin, L. K.; Wilkinson, T. A.; Krontiris, T. G.; Chen, Y. Identification of a SUMO-Binding Motif That Recognizes SUMO-Modified Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (40), 14373–14378. https://doi.org/10.1073/pnas.0403498101.

(45) Kim, S. A.; Zhu, J.; Yennawar, N.; Eek, P.; Tan, S. Crystal Structure of the LSD1/CoREST Histone Demethylase Bound to Its Nucleosome Substrate. *Mol. Cell* **2020**, *78* (5), 903-914.e4. https://doi.org/10.1016/J.MOLCEL.2020.04.019.

(46) Song, Y.; Dagil, L.; Fairall, L.; Robertson, N.; Wu, M.; Ragan, T. J.; Savva, C. G.; Saleh, A.; Morone, N.; Kunze, M. B. A.; Jamieson, A. G.; Cole, P. A.; Hansen, D. F.; Schwabe, J. W. R. Mechanism of Crosstalk between the LSD1 Demethylase and HDAC1 Deacetylase in the CoREST Complex. *Cell Rep.* **2020**, *30* (8), 2699-2711.e8.

https://doi.org/10.1016/J.CELREP.2020.01.091.

(47) Forneris, F.; Binda, C.; Vanoni, M. A.; Battaglioli, E.; Mattevi, A. Human Histone Demethylase LSD1 Reads the Histone Code. *J. Biol. Chem.* **2005**, *280* (50), 41360–41365. https://doi.org/10.1074/JBC.M509549200.

(48) BA, J. Using NMRView to Visualize and Analyze the NMR Spectra of Macromolecules. *Methods Mol. Biol.* **2004**, *278*, 313–352. https://doi.org/10.1385/1-59259-809-9:313.

Chapter 3

Sumoylation of the human histone H4 tail inhibits p300-mediated transcription by RNA polymerase II in cellular extracts

3.1 Introduction

Chromatin is Nature's elegant architectural solution to the challenge of packing approximately 3 billion base-pairs of human genomic DNA in an average nuclear volume of only about 500 cubic microns. Histones constitute the main protein component of chromatin and their reversible post-translational modifications (PTMs), or *marks*, regulate chromatin structure and function by a range of direct and indirect mechanisms.[1] Based upon their association with either transcriptionally active or silenced regions of chromatin, histone marks were proposed to constitute an epigenetic *code* for gene function.[2] As a consequence of their early discovery and the development of modification-specific chemical and molecular biological tools, marks such as methylation,[3] acetylation[4] and ubiquitylation[5] have been extensively investigated *in vitro* and in cell culture. In contrast, histone modification by the small ubiquitin-like modifier (SUMO) protein is a poorly understood mark due both to its very low abundance in cells, which prevents the isolation of sumoylated histones in quantities required for biochemical analysis, and to a lack of sumoylated histone-specific antibodies for cellular studies. First reported in human HEK293T and P493-6 B cells by Shiio and Eisenman,[6] histone sumoylation also occurs in yeast,[7] parasitic protozoans,[8] and plants.[9] Similar to histone ubiquitylation, sumoylation occurs on all core histones, the linker histone H1, the histone variants H2A.Z and H2A.X and the centromeric histone variant Cse4 in yeast.[10,11] Myriad roles have been proposed for histone sumoylation in different organisms,

including transcriptional regulation, kinetochore assembly, the regulation of chromatin structure, and double-strand break repair.[12] Pioneering efforts to identify specific lysine sites of sumoylation identified K12 in histone H4 as a major recurring site of sumoylation by SUMO-2/3 (H4K12su),[13,14] although multiple proximal lysines in the H4 N-terminal tail may also be enzymatically sumoylated *in vitro*.[10] Genetic studies in yeast and human cells have typically associated H4 sumoylation with the repression of gene transcription, although mechanistic studies of the direct roles for histone sumoylation in human cells have remained intractable due to the dynamic nature and low abundance of sumoylation.[6,15]

In an effort to understand the direct effects of H4K12su in chromatin, we previously applied a disulfide-directed chemical sumoylation strategy to generate uniformly and site-specifically sumoylated nucleosome arrays.[16] Biophysical studies of chromatin-array compaction remarkably showed that H4K12su is incompatible with the compact chromatin structures seen in transcriptionally silent heterochromatin. Subsequent biochemical studies revealed that H4K12su stimulates intranucleosomal activity of the H3K4me2-specific histone demethylase LSD1.[17] These studies suggested that sumoylated H4 does not directly enable heterochromatin formation and may instead act by recruiting LSD1 to genes. However, a potentially direct effect of histone H4 sumoylation on promoter-driven transcription by RNA polymerase II (RNAPII) and associated initiation factors that are key for efficient eukaryotic gene transcription has remained unknown.

Pioneering studies of the reconstitution of class II promoter-driven accurate eukaryotic transcription in both nuclear extracts and purified systems has led to insights into roles for histone modifications in gene function.[18,19] The ability to reconstitute chromatinized plasmid templates using chemically modified histones enables studies of the roles of specific histone modifications in transcription and investigations of their crosstalk with key enzymes associated with transcription initiation and elongation.[20] Multiple proteins involved in gene transcription bind to and modify

histone tails, which enables the remodeling of chromatin prior to and during transcription. One such modification, acetylation of lysine sidechains on H3 and H4 by the acetyltransferase p300, is necessary for efficient activator-driven transcription of both 11-nm chromatin[21] and 30-nm linker histone H1-containing heterochromatin, likely through mechanisms that include direct decompaction of chromatin upon H4K16 acetylation and octamer eviction by the chromatin remodeler NAP1.[22,23]

Due to their proposed opposing roles in gene transcription, we investigated the precise nature of biochemical crosstalk between histone sumoylation and histone acetylation by p300. Histone H4 site-specifically sumoylated at Lys12 (H4K12su) was synthesized with the aid of a traceless ligation auxiliary, 2-aminooxyethanethiol,[24] and then incorporated into histone octamers for subsequent reconstitution of cognate mononucleosomes and chromatinized plasmids. Each sumoylated substrate was subjected to acetyltransferase assays with the full-length p300 enzyme, which revealed a consistent inhibition of acetylation in the H4K12su tail. Consistent with this observation and requirements for both H3 and H4 acetylation for *in vitro* transcription of chromatin,[21] replacing wild-type (wt) H4 with H4K12su in chromatinized plasmid templates dramatically inhibited p300-dependent, RNAPII-mediated transcription *in vitro*. Bottom-up mass spectrometry on chromatinized histones, following a novel in-gel desumoylation protocol, revealed decreased acetylation in H4K12su by p300 when compared to wt H4. Consistent with a role in gene repression, H4K12su also inhibited H3K4 methylation by the extended catalytic module of the Set1/COMPASS methyltransferase complex.[25] To confirm the negative crosstalk with SUMO in human cells, linear non-hydrolyzable genetic fusions of SUMO-H4 were introduced in HEK293T cells and analyzed by Western blotting and ChIP-seq.

Collectively, our observations provide the first unambiguous biochemical demonstration that sumoylated histone H4 directly inhibits RNAPII-mediated transcription from chromatin templates

and reveal its direct negative crosstalk with histone acetylation by p300 and methylation by Set1/COMPASS that are strongly associated with active gene transcription.

3.2 Results and discussion

3.2.1 Reconstitution of site-specifically sumoylated octamers and nucleosomes

Site-specifically sumoylated human histone H4 at Lys12, H4K12su, was obtained by a semisynthetic strategy using the ligation auxiliary 2-aminooxyethanethiol (**Figure 3.1A and B**).[17] The semisynthetic sumoylated H4 was incorporated into octamers with purified recombinant human histones H2A, H2B and H3 (**Figure 3.1C**). Nucleosomes were reconstituted from sumoylated octamers using the 147 bp Widom 601 double-stranded DNA (**Figure 3.1D**).

3.2.2 Histone octamer acetylation by p300

We previously showed that H4K12su stimulates activity of the H3K4me1/2 demethylase, LSD1, in the context of a LSD1-CoREST sub-complex.[17] The stimulation of histone deacetylase (HDAC) activity of the Set3c complex in yeast was also recently proposed for sumoylated histone H2B.[26] Although the erasure of specific methyl and acetyl marks in the H3 and H4 tails may facilitate the transcriptionally repressed state of chromatin, there remains no information regarding the re-installation of these marks by the corresponding writer enzymes in the presence of H4K12su. Key among the histone acetyltransferases is the enzyme p300 that is recruited to chromatin by transcriptional activators for histone tail acetylation prior to transcription initiation.[21,27,28] Given its essential role in transcription, we investigated the effect of H4K12su on histone acetylation by p300 prior to and during transcription by RNAPII.

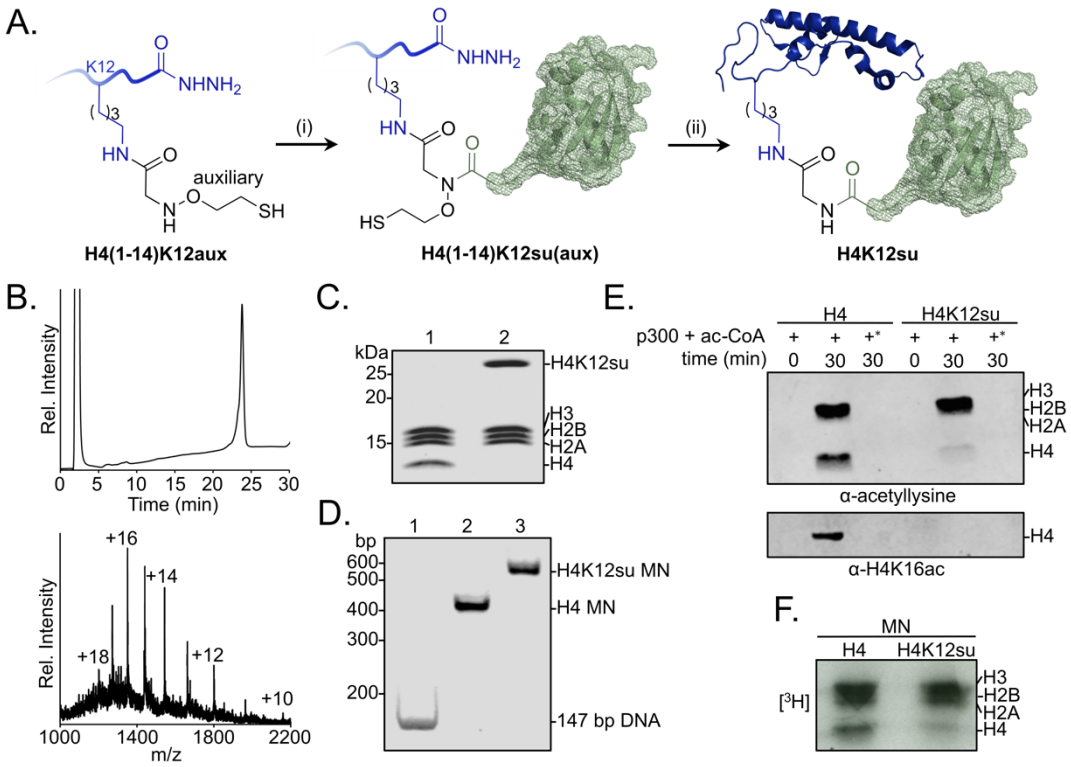


Figure 3.1. Sumoylation inhibits p300-mediated H4 acetylation in octamer and mononucleosome substrates. (A) Synthetic scheme for H4K12su. (i) An H4(1-14)K12aux peptide was ligated with a SUMO-3(2-91)C47S α-thioester. (ii) The sumoylated H4(1-14) peptidyl hydrazide containing the auxiliary was converted to a C-terminal α-thioester and ligated with H4(15-102)A15C. The auxiliary was then reductively cleaved from the ligation product. Cys15 in the final ligation product was desulfurized to the native Ala15 to yield site-specifically sumoylated H4K12su. (B) C4 analytical RP-HPLC trace of purified H4K12su (top). ESI-MS of purified H4K12su (bottom). Calculated mass 21,596.7 Da. Observed, 21,594.2 ± 3.4 Da. (C) Coomassie-stained 15% SDS-PAGE of reconstituted octamers containing wt H4 or H4K12su. (D) Ethidium bromide stained 5% TBE gel of mononucleosomes containing wt H4 or H4K12su. (E) Western blots of p300 assay products with octamer substrates containing wt H4 or H4K12su, probed with a site-independent pan-acetyllysine antibody (top) and an H4K16ac-specific antibody (bottom). An asterisk indicates assays with heat-inactivated p300 to exclude non-enzymatic acetylation. (F) Fluorogram of p300 assay products with [^{3}H]-acetyl-CoA as the co-factor and mononucleosome substrates containing wt H4 or H4K12su.

To investigate the direct biochemical crosstalk between H4K12su and acetylation, a Western-blot-based histone acetyltransferase (HAT) assay was developed using a sequence-independent pan-acetyllysine antibody to detect lysine acetylation in all four histones (**Figure 3.S1A**). In order to effectively compare acetylation of wt H4 with H4K12su and to strictly exclude any acetylation of the surface-exposed lysines in SUMO-3 attached to H4 (**Figure 3.S1B**), it was proteolyzed

from H4K12su prior to analysis. To this end, after acetylation by p300, the assay product was heat inactivated at 65 °C for 10 min, followed by addition of the purified catalytic domain of human SENP2 containing residues 365-590 (**Figure 3.S1C**).[29] Heat inactivation precluded p300 activity during desumoylation, and enabled the direct comparison of acetylation status in H4 and H4K12su.

Histone octamers containing wt H4 were first acetylated with full-length p300 immuno-affinity purified from HEK293T cells with an N-terminal FLAG epitope-tag (**Figure 3.S1D**).[30] Western blot analysis showed the robust acetylation of all four histones (**Figure 3.1E** top panel and **3.S1E**). This is consistent with previous *in vitro* assays that revealed acetylation of all four histones by p300.[31] Strikingly, H4 from octamers containing H4K12su was devoid of acetylation, including H4K16ac (**Figure 3.1E** bottom panel), which is strongly associated with chromatin decompaction and active gene transcription.[22,32,33] Importantly, the inhibition of H4 tail acetylation does not arise from allosteric inactivation of p300, based on the observation that the other histones in H4K12su octamers were acetylated to the same extent as in wt H4 octamers. Additionally, Western blots confirmed that SUMO-3 did not inhibit p300 autoacetylation, which is associated with robust acetyltransferase activity (**Figure 3.S1F**). Hence the inhibition of H4 acetylation in H4K12su is likely due to lysine acetylation site-occlusion by proximal SUMO-3 in the H4 tail.

3.2.3 Mononucleosome acetylation by p300

The histone acetylation assay was next undertaken with mononucleosomes containing either wt H4 or H4K12su. We failed to see significant nucleosome acetylation with pan-acetyllysine antibodies with or without pre-incubation of p300 with acetyl-CoA prior to the addition of nucleosomes (**Figure 3.S1G**). Due to the significantly decreased activity of p300 with nucleosomal substrates, a [^{3}H]-acetyl-CoA co-factor was employed and the transfer of acetyl

groups to histones was observed by fluorography. H4 acetylation was also suppressed in mononucleosomes containing H4K12su, but not in unmodified H4 mononucleosomes (**Figure 3.1F**). Our results unequivocally suggested that SUMO-3 in the H4 tail is inhibitory toward p300-mediated acetylation of chromatin, a process that is necessary for active gene transcription.

Based on the lower acetyltransferase activity observed with mononucleosomes than with octamers, we wondered if dsDNA may inhibit p300 activity. To test this, an equimolar amount of free 147 bp Widom 601 dsDNA was included in the octamer acetylation assay. The presence of DNA was sufficient to inhibit p300 activity to a similar extent as observed with mononucleosomes (**Figure 3.S1H**). This unexpected inhibition of p300 activity by free DNA suggests that additional factors, such as transcription factor and RNAPII binding, enable robust p300 activity on histones during transcription initiation and elongation.

3.2.4 The effect of H4K12su on cell-free transcription from chromatinized templates

Based on our observation that H4K12su inhibits the acetylation of key H4 tail residues that are associated with p300-dependent active transcription, including H4K16 that is acetylated in euchromatin, we sought to investigate the direct effect of H4K12su on transcription in our reconstituted cell-free system. These assays employed a p300/Gal4-VP16-dependent transcription system using chromatinized plasmids assembled with reconstituted octamers containing either wt H4 or H4K12su (**Figure 3.2A**). The plasmid DNA template consisted of five *gal4* binding sites and a ~400 bp G-less cassette.[28] Due to the absence of any engineered strong nucleosome positioning sequences in the template, chromatinization was undertaken with the histone chaperone NAP1 and the chromatin remodelers Acf1 and ISWI (**Figure 3.S2**). Limited micrococcal nuclease digestion of the transcription templates revealed the periodic spacing of nucleosomes in chromatin assembled with either wt H4 or H4K12su octamers, clearly indicating

that H4K12su does not inhibit the formation of recombinant chromatin (**Figure 3.2B**). In this background, addition of the transcription activator Gal4-VP16, p300, acetyl-CoA, [α-^{32}P]-CTP, rNTPs, and transcriptional machinery from a HeLa nuclear extract resulted in the transcription of a 365 base RNA from the chromatin template assembled with wt H4 histones (**Figure 3.2C**). Surprisingly, and in contrast to the direct structural decompaction of chromatin by H4K12su, transcription from templates assembled with H4K12su was drastically inhibited when compared with templates assembled with wt H4. The addition of Trichostatin A (TSA), a nanomolar inhibitor of class I and II HDACs, did not lead to significant changes in transcription, indicating that the repressive effect of H4 K12su is not significantly mediated through HDAC1 in chromatinized templates assembled with non-acetylated histones.[34] Importantly, our results unambiguously demonstrated transcriptional repression when site-specifically sumoylated H4 was present in chromatin.

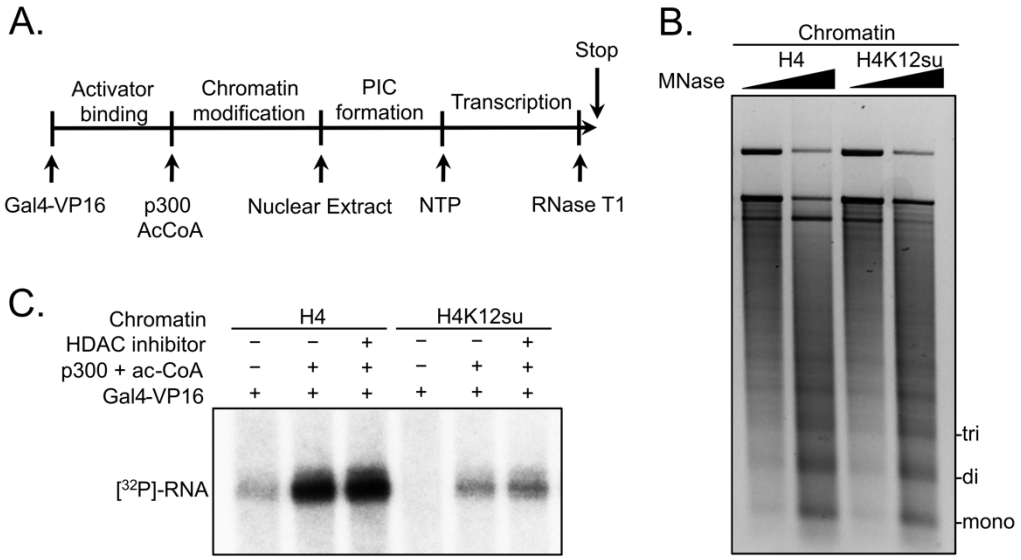


Figure 3.2. Histone H4 sumoylation inhibits *in vitro* transcription from chromatinized plasmid templates. (A) Scheme outlining steps during the *in vitro* transcription assay with chromatinized plasmids, nuclear extracts, activator Gal4-VP16 and p300. (B) Micrococcal nuclease digestion analysis of plasmids chromatinized with wt H4 or H4K12su indicating the similar occupancy and spacing of mononucleosomes. (C) Autoradiogram of ^{32}P-labeled 365-base RNA transcript generated from p300-mediated transcription from chromatinized templates containing wt H4 or H4K12su in the presence or absence of the HDAC inhibitor, Trichostatin A.

3.2.5 H4 acetylation is inhibited prior to gene transcription in chromatin containing H4K12su

Based on our observations with sumoylated octamer and nucleosome substrates, we wondered if the inhibition of transcription by H4K12su also correlated with diminished H4 tail acetylation by p300. Previous *in vitro* transcription studies with chromatinized plasmids containing either K-to-R mutations in the H4 tail or truncated H4 missing tail residues 1-19 revealed an ~80 % reduced transcriptional output relative to transcription from chromatin containing wt H4.[21] Chromatinized plasmids containing either wt H4 or H4K12su were incubated with Gal4-VP16, p300 and acetyl-CoA for 30 min to enable steps preceding transcription; and the histones were subsequently resolved by SDS-PAGE and analyzed by tandem mass spectrometry after chemical propionylation, trypsination, and separation by capillary-liquid chromatography (**Figure 3.S3**).[35] A critical innovation in the bottom-up analysis workflow was our use of the SENP2 catalytic domain to desumoylate H4K12su within the polyacrylamide gel matrix after SDS-PAGE. This procedural step was important to generate the same H4(4-17) tryptic peptide from wt H4 and H4K12su after p300-mediated acetylation. The H4(4-17) peptide contains K5,8,12 and 16 that are known to be acetylated by p300 *in vitro* and *in vivo*.[31]

Analysis of the H4(4-17) tryptic peptides arising from wt H4 revealed a remarkable degree of hyperacetylation within 30 min. The most abundant peptide corresponded to the K5,8,12,16 tetra-acetylated form with some tri-acetylated species also present (**Figure 3.3A**, **Table 3.1**, and **Figures 3.S4-S5**). This is consistent with the fact that p300 acetylates histones to facilitate transcription.[21–23,28] No significant degree of monoacetylation was observed, and a low abundance of diacetylated peptide was detected after manually searching the MS-MS spectra over the expected elution time (**Figure 3.S6**). In comparison, chromatin assembled with H4K12su generated significantly fewer hyperacetylated peptides, with approximately equal amounts of tri- and di-acetylated H4(4-17) peptides (**Figure 3.3B** and **Figures 3.S7-S8**). Small amounts of

unmodified H4(4-17) peptides were also observed (**Figure 3.S9**). This clearly indicated that H4K12su directly inhibits p300-mediated H4 tail acetylation in the steps prior to transcription. Given the importance of H4 tail acetylation for efficient transcription, H4K12su may inhibit transcription, in-part, by directly inhibiting p300 activity on H4.

Since all four acetylated states of H4(4-17) were observed during tandem-MS-based analysis of p300 assay products with octameric and chromatin substrates, we interrogated the site-specificity of p300 in the H4 tail. Consistent with previous reports, we observed that K5 and K8 are preferred sites in the double acetylated H4 tail, over acetylation at K12 and K16 (**Table 3.2**).[36] Additionally, K12 was preferentially acetylated over K16 in the triply acetylated H4 tail peptide (**Table 3.3**). These observations were consistent between H4 or H4K12su containing substrates, indicating that the intrinsic substrate preference of p300 is unchanged in the presence of SUMO-3 (**Table 3.4**).

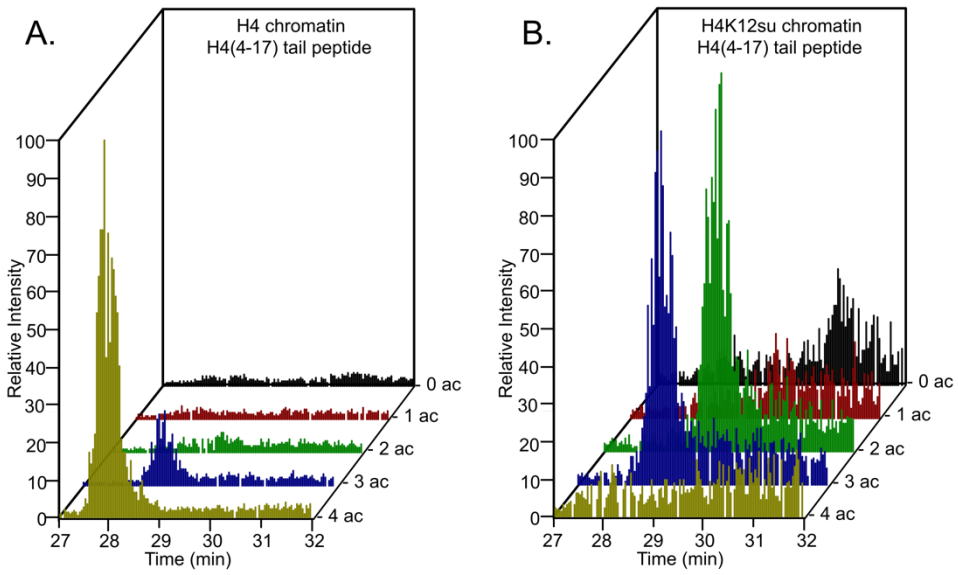


Figure 3.3. Comparison of H4 tail acetylation by p300 in chromatinized plasmid templates with activator Gal4-VP16. (A) Extracted ion chromatograms of all H4(4-17) tryptic peptides obtained after SDS-PAGE resolution and in-gel trypsination of acetylated chromatin containing wt H4. (B) Extracted ion chromatograms of all H4(4-17) tryptic peptides obtained after SDS-PAGE resolution, in-gel desumoylation and trypsination of acetylated chromatin containing H4K12su. The extracted m/z of each spectrum is centered on the $[M+2H]^{2+}$ precursor ion.

Table 3.1. H4(4-17) tail peptides acetylated by p300 in chromatinized plasmid templates with activator Gal4-VP16.[a,b]

H4(4-17) peptide	[M]	$[M+2H]^{2+}$	PSM H4	PSM H4K12su
prGKprGGKprGLGKprGGAKprR	1549.89	775.95	n.d.	1
prGKprGGKacGLGKprGGAKprR	1535.88	768.95	n.d.	n.d.
prGKacGGKacGLGKprGGAKprR	1521.86	761.94	2	3
prGKacGGKacGLGKacGGAKprR	1507.85	754.93	5	n.d.[c]
prGKacGGKacGLGKprGGAKacR	1507.85	754.93	n.d.	3[d]
prGKacGGKacGLGKacGGAKacR	1493.83	747.92	7	n.d.[c]

[a] Peptides were chemically propionylated before and after trypsinization to cap unmodified lysine sidechains and newly generated *N*-termini. [b] MS-MS spectra observed contained major fragments for the shown modification pattern over other potential patterns, however, no singly acetylated peptides were observed for wt H4. [c] Acetylation at K12 is not possible for H4K12su. [d] The triply acetylated peptide from H4K12su is blocked from acetylation at K12, but is propionylated after in-gel desumoylation. PSM= peptide spectral match. n.d. = not detected.

Table 3.2. Comparisons of relative ion-intensities of characteristic fragment-ions from an enzymatically di-acetylated and chemically propionylated H4(4-17) peptide, $[M+2H]^{2+}$ = 762 Da, after activator and p300-mediated acetylation of chromatinized plasmids containing wt H4.[a]

Species	Ion	% of Total ion intensity		Avg. ratio
K5ac, K8ac	b_5^+	1.316	1.654	12.4
K5pr, K8pr		0.072	0.253	
K12ac, K16ac	y_9^+	0.042	0.306	0.2
K12pr, K16pr		3.176	0.971	

[a] Only two unique spectra were observed and analyzed for the doubly acetylated and propionylated H4(4-17) tail peptide from wt H4 chromatin.

Table 3.3. Comparisons of relative ion-intensities of characteristic fragment-ions from an enzymatically tri-acetylated and chemically propionylated H4(4-17) peptide, $[M+2H]^{2+}$ = 755 Da, after activator and p300-mediated acetylation of chromatinized plasmids containing wt H4.[a]

Species	Ion	% of Total ion intensity			Avg. ratio
K5ac, K8ac, K12ac	b_9^+	3.032	4.938	3.981	24.1 ± 13.3
K5ac, K8ac, K12pr		0.339	0.096	0.147	
K5ac, K8ac, K12ac	b_{10}^+	0.875	0.862	1.109	10.3 ± 5.1
K5ac, K8ac, K12pr		0.112	0.152	0.064	
K12pr, K16ac	y_8^+	0.271	n.d.	0.012	0.2
K12ac, K16pr		0.885	1.005	1.263	
K12pr, K16ac	y_9^+	n.d.	0.123	0.054	0.02
K12ac, K16pr		2.587	4.442	6.296	

[a] Three unique spectra corresponding to the tri-acetylated and propionylated H4(4-17) tail peptide from wt H4 chromatin were analyzed. Error reported is standard deviation of the mean. n.d. = not detected.

Table 3.4. Comparisons of relative ion-intensities of characteristic fragment-ions from an enzymatically di-acetylated, desumoylated and chemically propionylated H4(4-17) peptide, $[M+2H]^{2+}$ = 762 Da, after activator and p300-mediated acetylation of chromatinized plasmids containing H4K12su.[a]

Species	Ion	% of Total ion intensity			Avg. ratio
K5ac, K8ac	b_5^+	2.096	2.308	1.770	71.6
K5pr, K8pr		n.d.	0.071	0.016	
K5ac, K8ac	b_6^+	2.603	2.835	2.900	33.1
K5pr, K8pr		0.053	n.d.	0.170	
K12pr, K16ac	y_6^+	0.038	0.080	n.d.	0.04
K12pr, K16pr		1.227	1.378	1.375	
K12pr, K16ac	y_8^+	0.027	0.013	0.138	0.1 ± 0.09
K12pr, K16pr		0.815	0.732	0.677	

[a] Three unique spectra corresponding to the di-acetylated and propionylated H4(4-17) tail peptide from H4K12su chromatin were analyzed. Error reported is standard deviation of the mean. n.d. = not detected.

3.2.6 Biochemical crosstalk between H4 sumoylation and acetylation in human cells

In order to further probe the biochemical relationship between histone H4 sumoylation and acetylation in human cells, HEK293T cells were transfected with a plasmid encoding an N-terminally HA-tagged *SUMO-3-H4* gene fusion. This fusion positioned SUMO at the N-terminus proximal to K12 in the tail while maintaining intact epitopes for antibodies toward H4K12ac or H4K16ac. To preclude desumoylation by intracellular SENPs, the two C-terminal Gly residues in SUMO-3 were omitted to generate the fusion protein HA-Su3(ΔGG)-H4. After 48 h of transfection, nuclear chromatin was isolated and treated with *Micrococcal nuclease* (MNase) to produce mononucleosomes (**Figure 3.4A**). Only nucleosomes containing either one or two copies of HA-Su3(ΔGG)-H4 were immunoprecipitated using anti-HA agarose beads, and the acetylation state of H4K12 and H4K16 investigated in Western blots using site-specific antibodies. This analysis revealed that the attachment of SUMO-3 to the N-terminus of H4 significantly inhibits H4K16ac and has a smaller inhibitory effect on H4K12ac (**Figure 3.4B**). If the K12 sidechain were sumoylated, as it is in semisynthetic H4K12su, this site also be blocked from acetylation. The

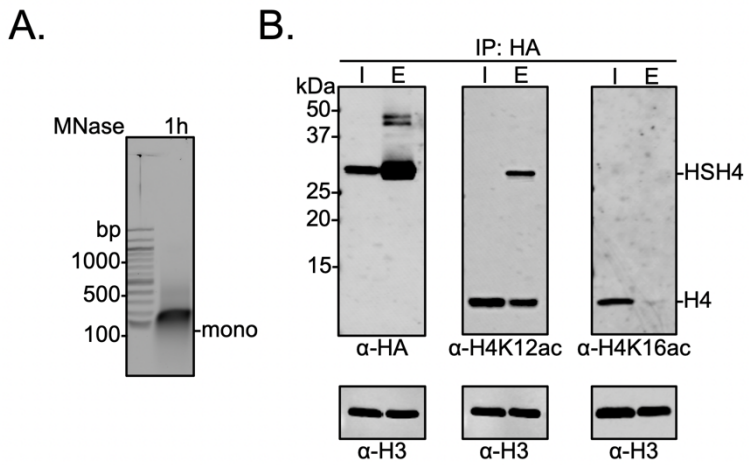


Figure 3.4. H4 acetylation is inhibited by SUMO in human cells. (A) Extended micrococcal nuclease digestion of chromatin generates mononucleosome-sized DNA fragments. (B) Lysates from HEK293T cells transfected with a plasmid containing HA-SUMO3ΔGG-H4 (HSH4) were immunoprecipitated against HA and blotted against the indicated antibodies. Input (I) and elution (E) were run with H3 as a loading control to compare total bulk chromatin to HSH4 associated chromatin.

presence of all the other histones in the pulldown also demonstrated that the ectopically expressed HA-Su3(ΔGG)-H4 fusion was globally incorporated into chromatin within human cells (**Figure 3.S10**). Thus, the biochemical relationship observed in HEK293T cells is consistent with our observation that H4K12su inhibits H4K16ac *in vitro*, and further indicates a degree of plasticity in the precise site of sumoylation that may exhibit negative biochemical crosstalk with H4K16ac. The cellular studies also revealed that H4 tail sumoylation not only inhibits p300 activity, but also likely inhibits other known H4K16 acetyltransferases including hMOF[37] and CBP.[38]

3.2.7 H3K4 methylation by COMPASS is inhibited in nucleosomes containing H4K2su

Along with H4K16ac, trimethylation at Lys 4 in histone H3 (H3K4me3) is a key mark associated with active gene transcription.[39] In humans, H3K4me3 is installed by the SETD1A/B and MLL-1/2 enzyme complexes, while the corresponding yeast enzyme complex is COMPASS (Complex of proteins associated with Set1).[40] Importantly, the catalytic module that imparts enzymatic activity

and product specificity is evolutionarily conserved in animals and yeast, and consists of Set1 and the subunits Swd1, Bre2, Swd3 and Sdc1 in COMPASS. In an effort to understand the mechanism of auto-regulation in SET1/MLL enzymatic complexes, we recently reported the reconstitution and structural characterization of an extended catalytic module (eCM) of COMPASS that contains both the nSET domain of Set1 and the Spp1 subunit (**Figure 3.5A**).[25] Although ubiquitylation at H2BK120 stimulates the methyltransferase activity of the eCM, it is not absolutely critical for nucleosome methylation by SET1/COMPASS complexes *in vitro*.[41] Based on our previous observation that H4K12su biochemically opposes the presence of H3K4me2 in nucleosomes by stimulating the activity of the H3K4me1/2 demethylase LSD1, we asked if H4K12su also directly opposes the installation of H3K4me3 in nucleosomes. Recent cryo-EM structures of the COMPASS eCM bound to the nucleosome core particle show significant spatial separation between the disordered H4 tail and the eCM (**Figure 3.5A**), thereby making it hard to predict any biochemical crosstalk between sumoylation and methylation.[25,42] In order to shed light on this problem, methylation assays were undertaken with mononucleosome substrates containing either wt H4 or H4K12su and the six-subunit eCM (**Figure 3.S11**). The degree of H3K4me1/2/3 was measured by Western blot with antibodies specific for the different H3 methylation states (**Figure 3.5B**). These experiments clearly showed that H4K12su inhibits the installation of H3K4me1/2/3 on nucleosomes. Moreover, the negative biochemical crosstalk arises from the presence of the Spp1 subunit in the eCM because the core 5 protein catalytic module (Set1, Swd1, Bre2, Swd3, Sdc1) remained active on nucleosomes with or without the presence of H4K12su (**Figure 3.5C**). Thus, we conclude that H4K12su in nucleosomes engages in negative biochemical crosstalk with p300-mediated acetylation within the same H4 tail, in *cis*, and may engage in negative biochemical crosstalk with the COMPASS-mediated methylation in the H3 tail, in *trans*.

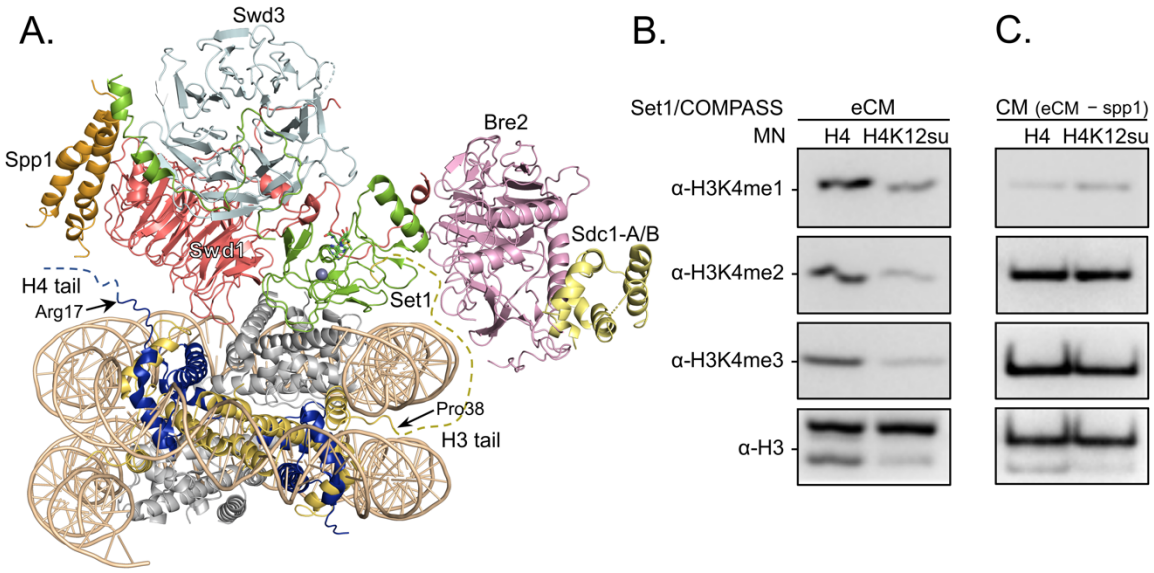


Figure 3.5. H3K4 methylation by the extended catalytic module of the COMPASS methyltransferase complex is inhibited by H4 sumoylation. (A) Structure of the COMPASS extended catalytic module (eCM) bound to a mononucleosome (PDB code 6UGM). The disordered H3 and H4 tails are shown in gold and blue, respectively, with the last observable amino acid indicated. Dotted lines indicate missing N-terminal amino acids. Spp1 (orange) suppressor of PRP protein 1. Swd3 (light blue), Set1 complex WD40 repeat protein 3. Swd1 (red), Set1 complex WD40 repeat protein 1. Set1 (green), SET domain protein 1. Bre2 (pink), brefeldin-A sensitivity protein 2. Sdc1-A/B (yellow), suppressor of CDC25 protein 1. (B) Western blots of the products from methylation assays with mononucleosome substrates containing wt H4 or H4K12su and the COMPASS eCM complex. Mono-, di- and trimethylated states of H3K4 were detected by the indicated modification-specific antibodies. (C) Western blots of the products from methylation assays with mononucleosome substrates containing wt H4 or H4K12su and the COMPASS catalytic module. Mono-, di- and trimethylated states of H3K4 were detected by the indicated modification-specific antibodies.

3.2.8 Biochemical crosstalk between H4 sumoylation and methylation in human cells

In order to probe the biochemical relationship between H4 methylation and sumoylation in HEK293T cells, the degree of H3K4 methylation was investigated in Western blots using site- and methylation-state specific antibodies with immunoprecipitated mononucleosomes containing HA-Su3(ΔGG)-H4. Interestingly we observed no significant effect on the bulk levels of H3K4 methylation associated with HA-Su3(ΔGG)-H4-incorporated nucleosomes relative to total HEK293T chromatin input (**Figure 3.6A**). This suggested that sumoylated H4 may not generally inhibit all the SET1/MLL family of methyltransferases found in humans on a genome-wide scale.

Although Set1/COMPASS is the only H3K4 methyltransferase in yeast, humans have the SETD1A/B and MLL1-4 family of enzymes. The MLL-family methyltransferases do not contain analogs of the COMPASS Spp1 subunit that may enforce negative crosstalk with H4 (**Figure 3.5A**).[43] However, the human SETD1A/B complex has an analog of the Spp1 subunit, namely CFP1, which targets SET1 to actively transcribed gene promoters.[44] Therefore we wondered if the biochemical crosstalk between H4 sumoylation and H3 methylation may be seen in the chromatin occupancy of sumoylated H4 at transcription start sites (TSSs). To address this specific crosstalk, we employed the Flp-In™ T-REx 293 cell line (Invitrogen) and Flp-recombinase mediated recombination between FRT sites to generate a stable HEK293T cell line containing a single copy of the FLAG-HA-SUMO-3(ΔGG)-H4(Δ1-11) gene under a doxycycline-inducible CMV promoter. In this system, fusion to the N-terminus of H4(Δ1-11) places SUMO-3 at the $^{\alpha}$N of H4K12 and closely mimics enzymatic sumoylation at the $^{\varepsilon}$N of H4K12. The addition of doxycycline led to the expression and incorporation of FLAG-HA-SUMO-3(ΔGG)-H4(Δ1-11) in chromatin as early as 4 h post-induction (**Figure 3.6B**). Cells were harvested after 24 h of doxycycline treatment and regions of genomic DNA-associated with FLAG-HA-SUMO-3(ΔGG)-H4(Δ1-11), or with H3K4me3, were identified by ChIP with anti-HA or anti-H3K4me3 antibodies, respectively, followed by high-throughput sequencing of the associated DNA. Our analysis of chromatinized DNA associated with FLAG-HA-SUMO-3(ΔGG)-H4(Δ1-11) and H3K4me3 revealed a striking inverse correlation in their genomic localization (**Figure 3.6C**). Although H3K4me3 was strongly accumulated at TSSs, the fusion protein was diminished from the TSSs and enriched at the transcription end sites (TESs) (**Figure 3.6C**). Thus, ChIP-seq analysis of gene occupancy reiterated the biochemical crosstalk between H4 sumoylation and H3K4me3.

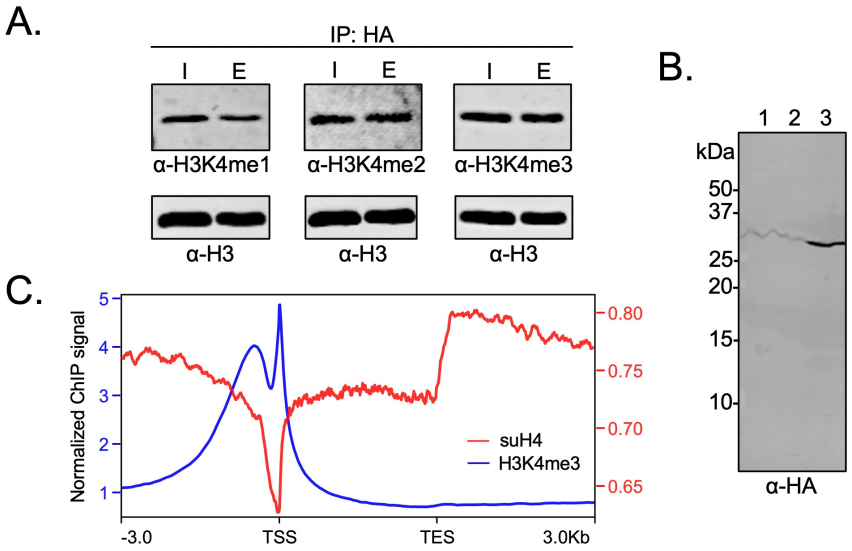


Figure 3.6. H3K4 methylation is regulated in part by H4 sumoylation. (A) Lysates from HEK293T cells transfected with a plasmid containing HA-SUMO3ΔGG-H4 (HSH4) were immunoprecipitated against HA and blotted against the indicated antibodies. Input (I) and elution (E) were run with H3 as a loading control to compare total bulk chromatin with HSH4 associated chromatin. (B) FLAG-HA-SUMO-3(ΔGG)-H4(Δ1-11) (suH4) expression was induced with doxycycline and cells collected at 4 h and fractionated. Lane 1 cytoplasm fraction; lane 2 nucleoplasm fraction; lane 3 chromatin fraction. (C) RPMK normalized ChIP-seq signals for SUMO-H4 (red, right y-axis) and H3K4me3 (blue, left y-axis) are plotted across length normalized gene bodies for 19,531 UniProt annotated protein coding genes plus 3kb upstream and downstream of the TSS and TES, respectively.

3.3 Conclusion and outlook

Histone marks in eukaryotic chromatin represent a range of biological pathways that modulate chromatin structure and function.[45] Marks may directly alter chromatin structure either through their steric bulk or by changing the charge of amino acid sidechains. Additionally, marks may recruit chromatin-modifying enzymes that change the modification state of other histones within a nucleosome. The biochemical crosstalk between marks is considered positive when one mark directs installation of another, and negative when one mark opposes the installation of another. The specific sumoylation of histones in chromatin was associated with the repression of gene

transcription through early studies in yeast and human cells.[6,15] We previously discovered the negative biochemical crosstalk between H4 sumoylation and H3K4me2 in nucleosomes mediated by the CoREST-LSD1 sub-complex, which suggested that sumoylation of actively transcribed regions enriched in H3K4me2 may lead to histone demethylation and silencing.[17] However, the direct effect of H4K12su in chromatin on p300-mediated gene transcription by RNAPII remained entirely unknown. Our semisynthesis of H4K12su using the ligation auxiliary, 2-aminooxyethanethiol, enabled the first interrogation of biochemical crosstalk between histone sumoylation and p300-catalyzed histone acetylation in reconstituted octamers, nucleosomes, and chromatinized plasmids. Interestingly, we found that although H4K12su has no significant impact on the acetylation of other histones, in either histone octamer or nucleosomal contexts, it significantly impedes acetylation in the unstructured H4 N-terminal tail. While the simplest explanation is that steric bulk of the 93-amino acid SUMO-3 at K12 prevents acetylation at proximal lysines, this should not be taken for granted as ubiquitin-family modifications at lysines do not always occlude enzymatic activity at proximal sites. For example, the ubiquitylation at K119 in the H2A C-terminus by the polycomb repressive complex 1 (PRC1) E3 ligases Ring1B/Bmi1 does not inhibit ubiquitylation at K124/K127/K129 in the H2A tail by the BRCA1-BARD1 heterodimeric E3 ligase.[46] Therefore, reduced acetylation may equally arise from a specific spatial orientation of the H4 tail upon its sumoylation that limits access to the p300 active site.

Given the essential role of acetylated lysines in the H4 tail on chromatin structure and gene transcription *in vitro* and in cells,[21] we surmised that diminished H4 acetylation may adversely influence gene transcription. Surprisingly, replacing wt H4 with H4K12su had no significant effect on the efficiency of plasmid chromatinization by the histone chaperone Nap1 and remodelers ACF1 and ISWI, resulting in regularly positioned nucleosomes for both wt H4 and H4K12su. Notably, however, an analysis of these chromatinized plasmid substrates in activator-and p300-dependent transcription assays with nuclear extracts revealed a strong repression of transcription

by H4K12su. Toward a further understanding of the H4 acetylation events whose loss may lead to this repression, bottom-up mass spectral analysis with data-dependent acquisition of tryptic H4 peptides after SDS-PAGE resolution of the transcription assay components demonstrated reduced acetylation in the H4(4-17) peptide when chromatin was reconstituted with H4K12su. To the best of our knowledge, this is the first mass-spectral analysis of acetylation in the tail of chromatin-associated H4 following *in vitro* transcription, and also the first demonstration that the catalytic domain of the SUMO-specific protease SENP2 can desumoylate histones in SDS-PAGE gels. Given the inherent challenges of detecting SUMO target sites in substrates, due to the lack of a convenient trypsin cleavage site at the C-terminus of SUMO, the ability to selectively remove SUMO using SENP2 may be particularly useful for analyzing sumoylated proteins in complex mixtures that require some degree of separation by SDS-PAGE. The fact that chromatinized plasmids containing H2B ubiquitylated at K120 show similar levels of transcription to unmodified chromatin suggests that the inhibition of transcription we observed may not strictly be due to the steric bulk of SUMO in chromatin.[20] Consistent with this notion, SUMO-3 fused to the N-terminus of H4 cells still prevented acetylation at K16 in HEK293T cells from 15 residues away. Importantly, HA-Su3(ΔGG)-H4 fusion protein was readily incorporated into chromatin within 24 h, as evidenced by the co-immunoprecipitation of all four endogenous histones and ~150 bp DNA after MNase digestion. The presence of endogenous wt H4 in HA-tagged mononucleosomes indicates that HA-Su3(ΔGG)-H4 forms both symmetric and asymmetric nucleosomes in cells. The observation of significantly inhibited acetylation at H4K16 within the HA-Su3(ΔGG)-H4 fusion protein in cells was consistent with out *in vitro* biochemical observations. Interestingly, endogenous wt H4 in HA-tagged asymmetric mononucleosomes was also deficient in H4K16ac, when compared with H4K16ac in bulk chromatin. This indicates an additional trans-mechanism of acetyltransferase inhibition by SUMO, in addition to its cis-inhibitory effect in the H4 tail

In addition to histone acetylation, another key histone mark associated with active transcription and enriched at promoter regions is H3K4me3. Installed by the Set1 containing COMPASS complex in yeast and the SET1/MLL1-2 family of methyltransferases in humans, H3K4me3 activates transcription in p53- and p300-dependent transcription from chromatinized plasmids.[41] And although H2BK120ub stimulates the methylation of H3K4 on these templates, it is not absolutely critical for SET1 complex activity.[41] Structures of the 5-protein core catalytic module of COMPASS[47] and the 6-protein extended catalytic module were recently reported.[24] While the CM complex does not change methyltransferase activity in the presence of H2BK120ub, the eCM demonstrates some activity on nucleosomes that is further enhanced by the presence of H2BK120ub.[25] Consistent with this observation, the human SET1 complex also retains some *in vitro* activity on chromatinized templates lacking H2BK120ub.[41] From the cryo-EM structure of the eCM complex bound to nucleosomes, and the relative position of the unstructured H4 tail (**Figure 3.4A**), we wondered if H4K12su would have an effect on H3K4me3 methylation by COMPASS. We discovered that although the CM is not hindered by the presence of H4K12su in nucleosomes, the eCM is significantly hindered by SUMO. From the differences in composition of the two subcomplexes, we propose that SUMO may sterically interact with the Spp1 subunit in the eCM and may reduce nucleosome binding and/or productive catalysis by the Set1 protein. Interestingly, our results from methyltransferase assays, in conjunction with previous observations that H4 tail sumoylation inhibits chromatin compaction, appear to indicate that SUMO attached to the H4 tail does not extend away from the nucleosome, but instead may occupy a fixed space that prevents the close apposition of both adjacent nucleosomes and histone-modifying enzymes in chromatin. Future structural studies will aim to identify the precise placement of SUMO in sumoylated nucleosomes. As histone acetylation by p300 also stimulates SETD1 activity on chromatin, the direct effect of H4K12su on p300-stimulated SETD1 methylation at H3K4 remains an interesting question.[41]

Our cellular studies further revealed the complexity of crosstalk between H4K12su and H3K4me3 in humans. Although immunoprecipitated mononucleosomes containing HA-Su3(ΔGG)-H4 did not show a strong negative correlation between SUMO and H3K4me3 across all regions of chromatin, this is consistent with our hypothesis that negative crosstalk is mediated by the Spp1 compnent of COMPASS (Figure 3.5B-C). The mammalian ortholog of Spp1, namely the CXXC-type zinc finger protein 1 (CFP1), is entirely absent in the MLL family of H3K4 methylatransferases and is only present in the SETD1A/B family.[48,49] SETD1A/B and MLL1-2 methyltransferase complexes are known to act on promoters, whereas MLL3-4 complexes modify enhancers. The knockdown and modulation of CFP1 levels has shown disruption of H3K4 methylation across the genome and is known to be an important regulator of a subset of actively transcribed genes.[44,48,50] Therefore, given the strong association of SETD1A/B with active gene promoters, we undertook ChIP-seq experiments to probe the occupancy of genes in HEK239 cells by H3K4me3 and FLAG-HA-Su3(ΔGG)-H4(Δ1-11). Gratifyingly, we observed a stark negative correlation between gene occupancy by SUMO and H3K4me3 between the TSS and TES of genes in HEK239 cells. Thus, our extensive investigations of biochemical crosstalk revealed that H4K12su does not inhibited H3K4 methylation genome-wide, but, instead may specifically inhibit the. Methylation of transcriptionally active genes by the SETD1A/B complexes.

Collectively, the disruption of H4 tail acetylation and H3 tail methylation by the presence of H4K12su along with the inhibition of p300-mediated transcription from chromatinized templates have revealed multiple, albeit not exhaustive, biochemical pathways by which histone sumoylation may inhibit gene transcription (**Figure 3.7**). These results have shed light on important aspects of chromatin regulation by histone H4 sumoylation and provide a strong mechanistic basis for the proposed roles for SUMO from studies in yeast and cultured human cells.

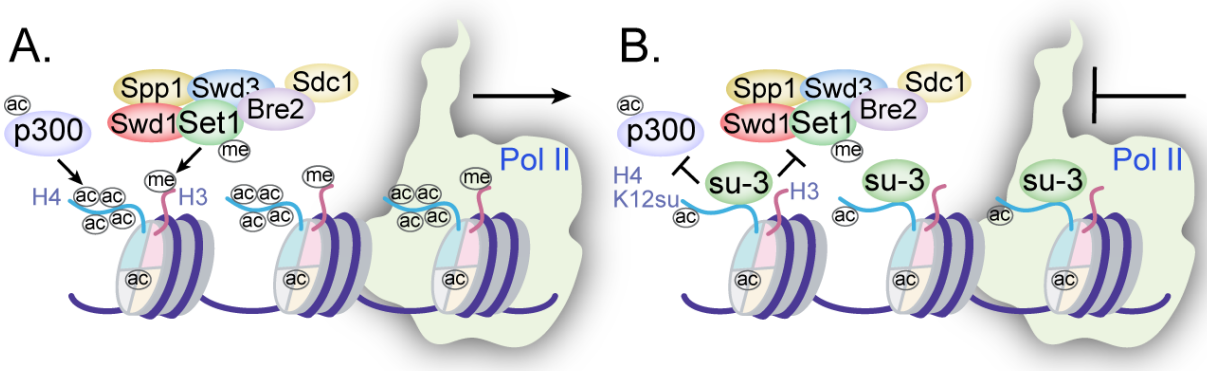


Figure 3.7. Mechanisms of chromatin regulation by H4K12su. (A) Transcription from chromatinized templates containing wt H4 is accompanied by acetylation of all four histones by p300, and with the methylation of the H3 tail by the COMPASS/SET1 complexes. (B) The inhibition of PolII-mediated transcription from chromatinized templates containing H4K12su is accompanied by reduced p300-mediated H4 tail acetylation. H4K12su also inhibits H3 tail methylation by the extended catalytic module of the COMPASS complex. For clarity, only one of two histone tails, each, is shown for H3 and H4.

3.4 Experimental procedures

3.4.1 Key resources

Table 3.5. Key resources and reagents used in this study are listed.

Resource or Reagent	Source	Identifier
acetyl-CoA	Roche	10101893001
Calcium Phosphate Transfection Kit	Thermo	K278001
Lipofectamine 3000	Invitrogen	L3000001
Anti-DYKDDDDK G1 Affinity Resin	GenScript	L00432
HisPur Ni-NTA resin	Thermo	88221
Anti-HA magnetic beads	Pierce	88836
Micrococcal nuclease solution	Thermo	88216
cOmplete, Mini, EDTA-free protease inhibitor cocktail	Roche	11836170001
[^{3}H]-acetyl-CoA	American Radiolabeled Chemicals	ART0213B
Pierce Trypsin Protease, MS Grade	Thermo	90057
Propionic anhydride	Sigma Aldrich	P51478
DMEM	Gibco	11956118
DPBS	Gibco	14190250

Fetal bovine serum	Gibco	16000044
Amplify fluorographic reagent	GE Amersham	NAMP100
Kodak GBX developer and fixer	Carestream Health	1900943
Trifluoroacetic acid	Alfa Aesar	AA31771-36
Formic Acid	Acros Organics	AC147932500
Acetonitrile (ACN)	Fisher	A996
C18 Zip tip	Millipore	ZTC18S096
Glacial acetic acid	Fisher	A38C-212
Bacterial Strains	**Source**	**Identifier**
E. coli BL21(DE3) competent cells	Thermo	FEREC0114
E. coli DH5α competent cells	NEB	C2987HVIAL
Human cell lines	**Source**	**Identifier**
HEK 293T	ATCC	CRL-3216
Flp-In T-Rex 293 cell line	Invitrogen	R78007
Plasmids	**Source**	**Identifier**
pST100-20xNCP*601*a	Gift from Dr. Robert K. McGinty	-
pcDNA3.1-p300-His_6	Addgene	23252
pET28a-His_6-SENP2(365-590)	Addgene	16357
pcDNA3.1-HA-SUMO3(ΔGG)-H4	GenScript	-
Vector-FLAG-HA-SUMO-3(ΔGG)H4(Δ1-11)	This study	-
Primers	**Source**	**Identifier**
5'-ATC CTT GTA ATC GTG TAT GTC TAG TGT ACT C-3'	IDT	p300_Ctrm_FLAG_R
5'-GAT GAC GAT AAA TAG TGA TAC TAA GCT TA AGT TTA AAC-3'	IDT	p300_Ctrm_FLAG_F
Antibodies	**Source**	**Identifier**
Polyclonal anti-acetyllysine antibody	Millipore	AB3879
Polyclonal anti-H4K16ac antibody	Active Motif	39167
Polyclonal anti-H4K12ac antibody	Active Motif	39066
Monoclonal anti-H3K4me1	Cell Signaling Technology	5326
Polyclonal anti-H3K4me2	Abcam	ab7766
Polyclonal anti-H3K4me3	Abcam	ab8580
Polyclonal anti-Histone H3	Abcam	ab1791
Monoclonal anti-Histone H3	Abcam	ab24834
Monoclonal anti-HA	Cell Signaling Technology	3724
Monoclonal anti-FLAG	Sigma-Aldrich	F1804
anti-rabbit, HRP conjugated	GE Healthcare	NA934
IRDye 680RD Goat anti-Rabbit IgG	Li-COR Biosciences	926-68071
IRDye 800CW Goat anti-Rabbit IgG	Li-COR Biosciences	926-32211
IRDye 800CW Goat anti-Mouse IgG	Li-COR Biosciences	926-32210

3.4.2 HPLC purification

Proteins and peptides were analyzed (4.6 x 150 mm, 5 µm) and purified (22 x 250 mm, 15 - 20 µm) with C4 and C18 reverse-phase HPLC columns from Vydac (Deerfiled, IL) on either a Varian Prostar (Palo Alto, CA) or Agilent (Santa Clara, CA) 1260 Infinity II LC system. The mobile phase consisted of Buffer A (0.1% trifluoracetic acid in water), and Buffer B (90% acetonitrile in water, 0.1% trifluoroacetic acid). UV-vis profiles of eluting peptides/proteins were monitored at 214 and 280 nm.

3.4.3 Electrospray ionization mass spectrometry

Routine peptide/protein mass spectrometry was performed by direct infusion on a Bruker (Billerica, MA) Esquire ion-trap mass spectrometer operating in positive mode.

3.4.4 Recombinant human histone purification

Human histones H2A 2-A, H2B 1-K, H3.2, and H4 were expressed from pET3a plasmids in *E. coli* BL21(DE3) cells.[16] The insoluble histones were extracted from exclusion bodies with 6 M Gn-HCl, 10 mM Tris, pH 7.5. Histones were precipitated by dialysis against Millipore water in Spectra/Por 6 3.5 kDa molecular weight cut-off dialysis tubing and lyophilized to dryness. Crude histones were dissolved in 6 M Gn-HCl and purified by preparative C4 RP-HPLC.

3.4.5 Semisynthesis of H4K12su

The peptide H4(1-14)K12ivDde-C(O)NHNH$_2$ was synthesized on 2-chlorotrityl chloride resin by standard 9-fluorenylmethoxycarbonyl (Fmoc)-based solid phase peptide synthesis on a CEM Liberty Blue Automated Microwave Peptide Synthesizer (Matthews, NC).[17] The protected ligation auxiliary, *O*-(2-(tritylthio)ethyl)hydroxylamine, was incorporated at K12 by coupling bromoacetic acid at deprotected K12 on the resin followed by displacement of the bromide with 0.25 M auxiliary in dry DMSO. After acidolytic cleavage from the resin and C18 RP-HPLC purification, the H4(1-

14)K12aux-C(O)NHNH$_2$ peptide (6 equivalents) was ligated with purified SUMO-3(2-91)C47S-MESNa thioester (1 equivalent) in ligation buffer consisting of 6 M Gn-HCl, 100 mM Na_2HPO_4, pH 7.3 to generate H4(1-14)K12su(aux)-C(O)NHNH$_2$ after 24 h at 25 °C. After C18 RP-HPLC purification, the sumoylated H4(1-14) peptidyl hydrazide was converted to the C-terminal acyl azide by diazotization with 15 equivalents of $NaNO_2$ in 200 mM Na_2HPO_4, 6 M Gn-HCl, pH 3.0 at -20 °C for 20 min. The acyl azide was converted to the 4-mercaptophenylacetic acid (MPAA) C-terminal α-thioester in situ and ligated with purified H4(15-102)A15C truncant protein (2 equivalents) in a ligation buffer consisting of 200 mM Na_2HPO_4, 6 M Gn-HCl, 200 mM MPAA, pH 6.5. The pH was adjusted to 6.8-7.0 and ligation allowed to proceed at 25 ^{o}C for 24 h. The full length H4(A15C)K12su(aux) ligation product was purified by C4 RP-HPLC and the ligation auxiliary subsequently cleaved by activated Zn in 6 M Gn-HCl, pH 3.0, at 37 °C under argon over 24 h. This yielded pure H4(A15C)K12su after C4 RP-HPLC purification. Finally, radical mediated desulfurization of Cys15 to the native Ala15 in histone H4 was accomplished by first dissolving H4(A15C)K12su 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 the radical initiator 2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) to a concentration of 10 mM. Desulfurization was allowed to proceed at 37 °C for 24 h and the final sumoylated histone product, H4K12su, was purified by C4 RP-HPLC.

3.4.6 147 bp Widom 601 DNA preparation

A plasmid containing 20 repeats of the 147 bp Widom 601 sequence, pST100-20xNCP601a was a kind gift from Professor Robert K. McGinty (UNC, Chapel Hill).[51] The plasmid was used to transform *E. coli* DH5α and propagated in LB media. The plasmid was extracted from cells by alkaline cell lysis and precipitated with isopropanol. The precipitate was collected by centrifugation and washed with 70% ethanol, followed by resuspension in 20 mM Tris, pH 8.0. The plasmid was purified by SOURCE 15Q 4.6/100 PE strong anion exchange column and analyzed by 1%

agarose gel electrophoresis. Fractions pure from RNA contamination were pooled and dialyzed against 10 mM tris, pH 8, overnight. The plasmid was cleaved by EcoRV to liberate the 147 bp 601 Widom sequence overnight. The plasmid backbone was separated by PEG precipitation followed by SOURCE 15Q 4.6/100 PE anion exchange. Pure fractions of 147 bp dsDNA, as seen by 1.5% agarose gel electrophoresis, were pooled and dialyzed against 10 mM Tris, pH 7.0, overnight and stored frozen at -20 °C.

3.4.7 Octamer and mononucleosome formation

All four histones were combined in equimolar amounts in 7 M Gn-HCl, 20 mM tris, pH 7.5 at a final concentration of 1 mg/mL and dialyzed against 10 mM Tris, 2 M NaCl, 1 mM EDTA, pH 7.5 overnight at 4 °C. The self-assembled crude octamers were purified by size-exclusion chromatography on a Superdex-200 10/300 GL column attached to an AKTA FPLC (GE Healthcare, Chicago, IL). Pure fractions were identified by 15% SDS-PAGE after coomassie staining. Pure octamer fractions were combined and concentrated, diluted with 10% glycerol and flash frozen for long-term storage. Mononucleosomes were formed by mixing equivalent amounts of histone octamers and 147 bp Widom *601* DNA in a buffer consisting of 2 M NaCl and 10 mM Tris, pH 7.0, followed by dialysis against 10 mM Tris to a final NaCl concentration of 200 mM. Small-scale mononucleosome formation tests were first undertaken to fine-tune the molar ratio of octamer to 147 bp DNA before large-scale preparations. This avoided the presence of unbound 147 bp DNA in the mononucleosome preparations as seen by 5% TBE gels stained with ethidium bromide. Freshly prepared mononucleosomes containing wt H4 or H4K12su were stored on ice at 4 °C for a maximum period of 3 weeks.[16]

3.4.8 Purification of the SENP2 catalytic domain

A pET28a plasmid containing the SENP2 catalytic domain (cat.SENP2) was obtained from Addgene (Catalog number 16357).[52] The protein was expressed and purified from *E. coli* (DE3)

cells. Briefly, cells were grown in 2xYT medium at 37 °C until OD_{600} 0.6-0.8 and protein expression was induced with 0.5 mM isopropyl-*β*-D-galactopyranoside for 3 h at 25 °C. Cells were collected by centrifugation and lysed by sonication in 20% (w/v) sucrose, 20 mM Tris, 350 mM NaCl, 20 mM imidazole, 1 mM *β*-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF), pH 8.0. The lysate was clarified by centrifugation and bound to Ni^{2+}-NTA resin over 1 h with continuous nutation at 4 °C. The resin was then washed with a buffer containing 20 mM Tris, 350 mM NaCl, pH 8.0, followed by successive buffers containing increasing amounts of imidazole from 20 mM to 100 mM to elute non-specifically bound proteins. Finally, the cat.SENP2 was eluted with a buffer containing 20 mM Tris, 350 mM NaCl, 400 mM imidazole, pH 8.0. Fractions containing cat.SENP2 were identified by 15% SDS-PAGE, pooled and dialyzed in Spectra/Por 6 dialysis tubing with 15 kDa molecular-weight cut-off against 2 L of dialysis buffer containing 20 mM Tris, 100 mM NaCl, 5% (v/v) glycerol, pH 8.0 at 4 °C for two hours, twice. The dialyzed protein was stored at -80 °C in aliquots flash frozen in liquid nitrogen.

3.4.9 DNA cloning and sequencing

A Q5 site-directed mutagenesis kit (NEB, Ipswitch, MA) was used to generate pcDNA3.1-p300-FLAG from the plasmid pcDNA3.1-p300-His_6 obtained from Addgene (Catalog number 23252) by following the manufacturer's protocols. Oligos for molecular cloning were purchased from Integrated DNA Technologies (Coralville, IA) and DNA sequencing was carried out by Eurofins Genomics (Louisville, KY). pcDNA3.1-HA-SUMO3(ΔGG)-H4 was prepared by GenScript.

3.4.10 Purification of pcDNA for transfection

DNA for transient transfection was prepared by Miraprep of *E. coli* DH5α cells as previously described using a Qiagen (Valencia, CA) DNA miniprep kit.[53] Briefly, transformed *E. coli* DH5α cells were grown in 50 mL LB media supplemented with ampicillin (50 μg/mL) overnight at 37 °C. Cells were collected by centrifugation and resuspended in P1 buffer supplemented with fresh

RNase. After alkaline lysis and neutralization, the supernatant was cleared by centrifugation. The supernatant was diluted with an equal volume of 96% (v/v) ethanol prior to loading onto five Qiagen miniprep spin columns. At this point, the DNA was washed and eluted according to the Qiagen protocol. Purity of the eluted DNA was checked by measuring the $A_{280\ nm}/A_{260\ nm}$ ratio on a NanoDrop 2000c spectrophotometer and by agarose gel electrophoresis. The correct gene sequence for p300 was also confirmed by sequencing prior to transfection in human cells.

3.4.11 HEK293T cell culture

HEK293T cells were cultured in T75 flasks using Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and incubated at 37 °C in a 5% CO_2 atmosphere.

3.4.12 Transient transfection of HEK293T cells

HEK293T cells were cultured to ~60% confluency before transient transfection with pcDNA3.1-p300-FLAG. The cell growth medium was changed at least 1 h before transfection. Cells were transfected with the Calcium Phosphate Transfection Kit (Thermo Fisher Scientific, Waltham, MA) using 15 µg DNA per T75 flask for FLAG-p300. Lipofectamine 3000 (Invitrogen) was used to transfect cells with plasmid containing HA-SUMO3(ΔGG)-H4, and the media was changed 24 h later. Cells were then grown in transfection medium over 48 hours before detachment by trypsin and collection by centrifugation. The cell pellet was washed thrice with ice-cold Dulbecco's Phosphate-buffered Saline before either protein purification, immunoprecipitation, or storage at -80 °C.

3.4.13 Purification of the full-length histone acetyltransferase p300

The washed HEK293T cell pellet was resuspended in lysis buffer consisting of 20 mM Tris, pH 7.4 at 4 °C, 0.5 M KCl, 5 mM $MgCl_2$, 0.1% (v/v) IGEPAL CA-630, 1 mM PMSF, 1x EDTA-free protease inhibitor (Roche, Basel, CH), 10 % (v/v) glycerol then frozen in dry ice and allowed to

thaw at room temperature. The lysate was nutated for 30 min at 4 °C before clarifying by centrifugation at 17,620*rcf* for 30 min at 4 °C. The soluble fraction was bound to anti-DYKDDDDK resin (Genscript, Piscataway, NJ) for 1 h at 4 °C, then washed four times in lysis buffer and finally once in lysis buffer containing 150 mM KCl without protease inhibitors. The full-length p300-FLAG protein was eluted with 200 µg/mL 3xFLAG peptide in 150 mM KCl containing lysis buffer. The purity and concentration of p300-FLAG were assessed by 8% SDS-PAGE using BSA standards (ThermoFisher) and by western blot with an anti-FLAG antibody (Sigma). Aliquots of p300-FLAG in 10% (v/v) glycerol containing elution buffer were flash frozen in liquid nitrogen and stored at -80 °C.

3.4.14 Histone acetylation assays

Histone acetylation assays were conducted as outlined in Figure S1B. In brief, 30 µL volume assays were prepared by mixing histone octamers (300 nM), acetyl-CoA (25 µM), and p300-FLAG (150 pM) in histone acetyltransferase buffer consisting of 50 mM Tris, pH 8.0, 1 mM DTT, 1 mM PMSF, 0.1 mM EDTA and 10% (v/v) glycerol. Acetylation assays with octamer substrates were undertaken at 30 °C for 30 min and stopped by heating the assay mixture to 65 °C for 10 min which denatures p300. For fluorography, 389 nM mononucleosomes containing either H4 or H4K12su were incubated with 2 µL [^{3}H]-acetyl-CoA (15.95 Ci/mmol) and p300 (292 pM) overnight at 30 °C before heat denaturation at 65 °C for 10 min.[54] Heat denatured samples were cooled down to 30 °C before the addition of cat.SENP2 (3 equivalents relative to H4K12su). To ensure similar sample handling, cat.SENP2 was added to both H4 and H4K12su assay samples. After 2 h of desumoylation with cat.SENP2, the samples were denatured with 6xLaemmli and resolved by SDS-PAGE prior to analysis by western blotting or by fluorography.

3.4.15 Native chromatin immunoprecipitation of mononucleosomes

Nuclei were collected from pelleted HEK293T cells transfected with HA-SUMO3(ΔGG)-H4 by resuspension in a modified STM-N/D buffer (250 mM sucrose, 50 mM tris, pH 7.4 @ 4 °C, 5 mM $MgCl_2$, 10 mM iodoacetamide, 20 mM N-ethylmaleimide, 0.5 % IGEPAL CA-630, 1x cOmplete protease inhibitor cocktail, 0.5 % sodium deoxycholate). Suspensions were incubated on ice for 5 min and then spun down at 25 kRCF, 4 °C, 5 min to pellet nuclei. The cytoplasmic fraction was removed, and the nuclei were resuspended in p300-buffer(150) (20 mM Tris, pH 7.4 at 4 °C, 0.15 M NaCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$, 0.1% IGEPAL CA-630, 1 mM PMSF, 10 mM iodoacetamide, 20 mM N-ethylmaleimide , 1x cOmplete, 10 % glycerol). The nuclei were digested with micrococcal nuclease diluted in 50 mM tris, pH 8 @ 37 °C, 5 mM $CaCl_2$, 20 mM NEM, 10 mM iodoacetamide, 0.1 % triton x-100, 1x cOmplete. Digestion went for 1 h in a 37 °C water bath, and mixed every 30 min. The digestion was halted by addition of 0.5 M EGTA and then passed through a 25G needle five times. The resulting solution was diluted with one volume p300-buffer(500) (p300-buffer with 0.5 M NaCl) and supplemented with NaCl to 0.5 mM. The digested chromatin was clarified by centrifugation at 17.2 kRCF, 4 °C, 20 min, and the soluble fraction was collected. A sample of the soluble fraction was mixed with 1 % SDS and analyzed by 1.2 % agarose gel electrophoresis to check for mononucleosomes (~150 bp sized DNA). The soluble fraction containing chromatin digested to mononucleosomes was applied to anti-HA magnetic resin pre-equilibrated in p300-buffer(500) and nutated overnight at 4 °C. The unbound fraction was removed, and the resin was washed 5 times with p300-buffer(500). Bound proteins were eluted with 3xLaemmli (no DTT) and boiled for 3 min. The elution was supplemented with DTT to 50 mM after removal from the resin. Input and elution samples were run on 15 % SDS-PAGE and transferred to PVDF membranes for Western blotting.

3.4.16 Histone methyltransferase assays

Nucleosome (0.5 μM) and COMPASS CM or COMPASS eCM (1 μM) were incubated together in 20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM DTT, 0.2 mM SAM for 30 min at 30 °C. The reaction was quenched in SDS-PAGE loading buffer and analyzed by western blot.[25]

3.4.17 Western blotting

Western blots were performed in modified Towbin buffer consisting of 25 mM Tris, 192 mM glycine, 4 mg/L SDS, 10% (v/v) methanol, and proteins were transferred onto PVDF membrane at 35 V for 2 h, on ice. Membranes were blocked in 5% (w/v) non-fat milk powder in phosphate-buffered saline (PBS) for 1 h at 25 ^{o}C before incubating overnight in diluted primary antibody in 5% (w/v) non-fat milk powder in PBST (PBS containing 0.05% (v/v) Tween-20) at 4 °C. Overnight incubated membranes were washed in PBST before incubating with IR-dye conjugated secondary antibody or HRP-conjugated secondary antibody in 5% (w/v) non-fat milk powder containing PBST for 1 h at 25^{o}C. After incubation with secondary antibodies, the membranes were washed first in PBST and then PBS before scanning on a Li-COR Biosciences (Lincoln, NE) Odyssey IR scanner or developing with ECL reagents.

3.4.18 Fluorography

After SDS-PAGE separation of methylation assay components, gels were soaked in Amplify solution (GE Healthcare, Chicago, IL) for 30 min before drying on a vacuum air dryer. Dried gels were exposed to X-ray film for 1 week at -80 °C. Images were developed and fixed using Kodak (Rochester, NY) GBX solutions.

3.4.19 Chromatin assembly and MNase digestion

Chromatin assembly and micrococcal nuclease analysis proceeded essentially as described previously.[23] Briefly, octamer and hNAP-1 were incubated together on ice for 15 min before the

addition of hTopo1, dACF1/ISWI, relaxed circular plasmid DNA, and ATP-Mg mix (0.5 M creatine phosphate, 0.5 M ATP, 1 M $MgCl_2$, 5 mg/mL creatine kinase) and incubated at 30 °C for 2 h. Micrococcal nuclease (TaKaRa Bio Inc., Shiga, JP) was added to chromatin and digested at 25 °C for 10 min. DNA was purified from the digested chromatin by miniprep kit and analyzed by 1.25% agarose gel stained with ethidium bromide.

3.4.20 *In vitro* transcription assay

Assays were conducted as outlined in Figure 2A following previously described protocols.[23,55] Gal4-VP16 (30 ng), chromatin template (25 ng), p300 (10 ng), and acetyl-CoA (5 μM) were incubated in 20 mM HEPES-KOH, pH 7.9, 1 mM EDTA, 100 mM KCl, 10% (v/v) glycerol, 0.2 mg/mL BSA at 30 °C for 10 min. Preinitiation complex formation was initiated by the addition of HeLa extract (30 μg total protein) with continued incubation at 30 °C for 10 min. Transcription was initiated by the addition of ATP, UTP, CTP, 3'-O-methyl-GTP, and [α-^{32}P]-CTP followed by incubation at 30 °C for 1 h. Assay products were resolved by agarose gel and analyzed by autoradiography.

3.4.21 In-gel desumoylation and propionylation of lysine residues

Histone modification analysis by mass spectrometry was conducted as previously described with one key modification.[35] Histones were separated by SDS-PAGE and stained in coomassie stain consisting of 0.05% (w/v) Coomassie brilliant blue dissolved in 45:45:10 (v/v) methanol/H_2O/acetic acid. Bands were mostly destained in 45:45:10 (v/v) methanol/H_2O/acetic acid and specific histone bands were excised from the gel and diced into 1 mm cubes. Gel pieces were sequentially washed in 100 mM NH_4HCO_3, pH 8.0, and then dehydrated in sufficient acetonitrile to cover the gel pieces with gentle agitation for 15 min at 25 °C. Gel pieces were dried by lyophilization before rehydrating in the presence of cat.SENP2 (1 μg/μL) in histone acetyltransferase assay buffer on ice for 30 min, and further incubated overnight at 30 °C. After

desumoylation, the gel pieces were sequentially washed in 100 mM NH_4HCO_3, pH 8.0, and then dehydrated in sufficient acetonitrile to cover the gel pieces with gentle agitation for 15 min at 25 ^oC. Histones were propionylated in-gel with 1:2 (v/v) 100 mM NH_4HCO_3/propionic anhydride for 20 min with shaking at 25 ^oC. The propionylation reagent was aspirated away from gel pieces and they were sequentially washed in 100 mM NH_4HCO_3, pH 8.0, and then dehydrated in sufficient acetonitrile to cover the gel pieces with gentle agitation for 15 min at 25 ^oC. In-gel propionylation was undertaken thrice to ensure complete reaction.

3.4.22 In-gel tryptic digestion and peptide extraction

Histones were digested by rehydrating gel pieces with 12.5 ng/µL trypsin (Pierce) in 50 mM NH_4HCO_3 on ice for 30 min, and then incubation at 30 °C overnight. The bicarbonate solution after overnight digestion contained histone peptides that were transferred into a new empty tube. Peptides were further extracted from the trypsinized gel pieces by the sequential addition of 20 µL milliQ pure H_2O and shaking for 15 min, followed by incubation with 20 µL acetonitrile and shaking for 15 min. All three solutions – bicarbonate, water and acetonitrile – were pooled together. The peptide extraction process was repeated once, and the combined peptide solutions were lyophilized to dryness. The dry tryptic peptides were propionylated to cap newly formed N-termini by resuspending in 30 µL 100 mM NH_4HCO_3 and adjusting to pH 8.0 before the addition of 15 µL 1:3 (v/v) propionic anhydride:ACN. The pH was quickly readjusted to pH 8.0 with NH_4OH before shaking for 15 min at 25 ^oC. The reaction mixture was dried on the lyophilizer and the propionylation of N-termini was repeated once to ensure complete reaction.

3.4.23 C18 Zip-tip desalting of histone peptides

The dried histone peptides were dissolved in 50 µL 0.1 % (v/v) TFA in milliQ pure H_2O by vortexing and sonication in a water-bath, and the pH was adjusted to ~ 4.0 with TFA prior to desalting with C18 Zip-tips (EMD Millipore). The Zip-tip was first washed sequentially with 3x10 µL acetonitrile

and then washed with 3x10 µL 0.1% (v/v) aqueous TFA. Peptides were bound to the resin by pipetting the peptide solution through the Zip-tip resin in 10 µL aliquots until all the solution had passed through the resin at least once. The Zip-tip resin was then washed by stepwise transfer of 5x10 µL of 0.1 % (v/v) aqueous TFA into waste. The solid-phase bound peptides were then eluted into a clean empty tube by the stepwise transfer of 5x10 µL of 75% (v/v) acetonitrile in water containing 0.5% (v/v) acetic acid. The combined eluted peptide solutions were lyophilized to dryness.

3.4.24 Nano-LC and tandem mass spectrometry analysis of histone peptides

The lyophilized histone tryptic peptides were dissolved in 15 µL of 10% (v/v) acetonitrile and 1% (v/v) formic acid in milliQ pure water by vortexing and sonication in a water-bath. Samples were loaded onto a C18 trap column (LC Packing) and then into a capillary column using an Agilent 1100 Series capillary LC system (Santa Clara, CA). The capillary column was packed in house with Phenomenex C12 Jupiter resin. The mobile phase used consisted of A (0.6% (v/v) acetic acid in milliQ pure water) and B (0.6% (v/v) acetic acid in acetonitrile). Peptides were separated over a 70 min gradient from 2% to 98% buffer B, and directly eluted through the nano electrospray ionization source into a Thermo Finnigan LTQ XL mass spectrometer (San Jose, CA) for MS and MS-MS analysis. Peptide ions were identified and analyzed manually using the Thermo Xcalibur Qual Browser. Precursor fragments were identified over at least two peptide spectra, when possible, with an error of ± 0.8 Da. Mass spectrometer parameters were: Isolation width: 1 m/z; normalized collision energy: 35.0; scan range m/z: 300.00–2000.00. Dynamic exclusion was enabled.

3.4.25 Generation of stable cell lines

FLAG-HA-Su3(ΔGG)-H4($\Delta 1-11$) was sub-cloned by Gibson assembly from DNA fragments into the AflII site of the pcDNA5/FRT/TO vector (Invitrogen) to stably introduce these genes into the

genome of HEK293 Flp-In TREx cells under the control of a tetracycline (Tet)-inducible promoter. All plasmids were verified by DNA sequencing. Transient transfection and selection of stable transfectants were performed according to the manufacturer's recommendations. Verification of Tet-inducible expression of the stably incorporated gene was confirmed by immunoblotting for the HA-tagged protein in cell lysates.

3.4.26 Native chromatin immunoprecipitation for high-throughput sequencing

ChIP-seq experiments were performed using minor modifications of previously described methods.[56] Briefly, doxycycline-induced, or uninduced, Flp-In™ T-REx 293 cells were resuspended in 10 mM HEPES, pH 7.5, 100 mM NaCl, 300 mM Sucrose, 3 mM $MgCl_2$, 10 mM NEM, 0.5% Triton X-100, 1x cOmplete protease inhibitor cocktail. The suspension was incubated on ice for 2 min then centrifuged at 2,500 x g for 5 min at 4° C to pellet nuclei. Nuclei were taken up in a buffer consisting of 50 mM Tris, pH 8.0, 150 mM NaCl, 5 mM $CaCl_2$, 10 mM NEM, 1% Triton X-100 and 1x cOmplete protease inhibitor cocktail. Chromatin was digested to mononucleosomes with 10 units of MNase (ThermoFisher, Catalog 88216) for 15 min at 37 °C. Digestion was stopped by the addition of 1 mM EDTA and the reaction mixture clarified by passing through a 25 gauge needle ten times then centrifugation at 17,200 x g for 10 min at 4 °C in a microcentrifuge. 15 µg of chromatin was incubated overnight at 4 °C with α-HA antibody (Sigma, Catalog number H9658) or α-H3K4me3 antibody (Active Motif, Catalog number 39159) coated Dynabeads (ThermoFisher). Bound chromatin was washed using RIPA buffer with 500 mM LiCl for 5 times. Bound chromatin was eluted in TE buffer and RNA was removed by incubation with RNAseA for 30 min at 37 °C followed by Proteinase K treatment at 65 °C for 30 min. DNA was purified using Phenol:Chloroform:Isoamyl alcohol. Finally, precipitated DNA was further washed 3 times with 80% aqueous ethanol and eluted in nuclease-free water for use in library preparation.

3.4.27 Library preparation

Library preparation was performed using NEBNext® UltraTM II DNA Library Prep Kit from Illumina® according to the manufacturer's protocol. Final libraries were amplified using KAPA HiFi Hotstart ready mix then purified and size selected for 300 to 600 bp final library product using AMPURE beads (Beckman Coulter). Final library quantification was performed using Qubit dsDNA HS kit and average size was determined using high sensitivity BioAnalyzer DNA kit (Agilent). Libraries were sequenced as single-end 75 cycles on the Illumina Nextseq 550 platform. Experiments were performed in duplicate.

3.5 Product characterization and supplemental data

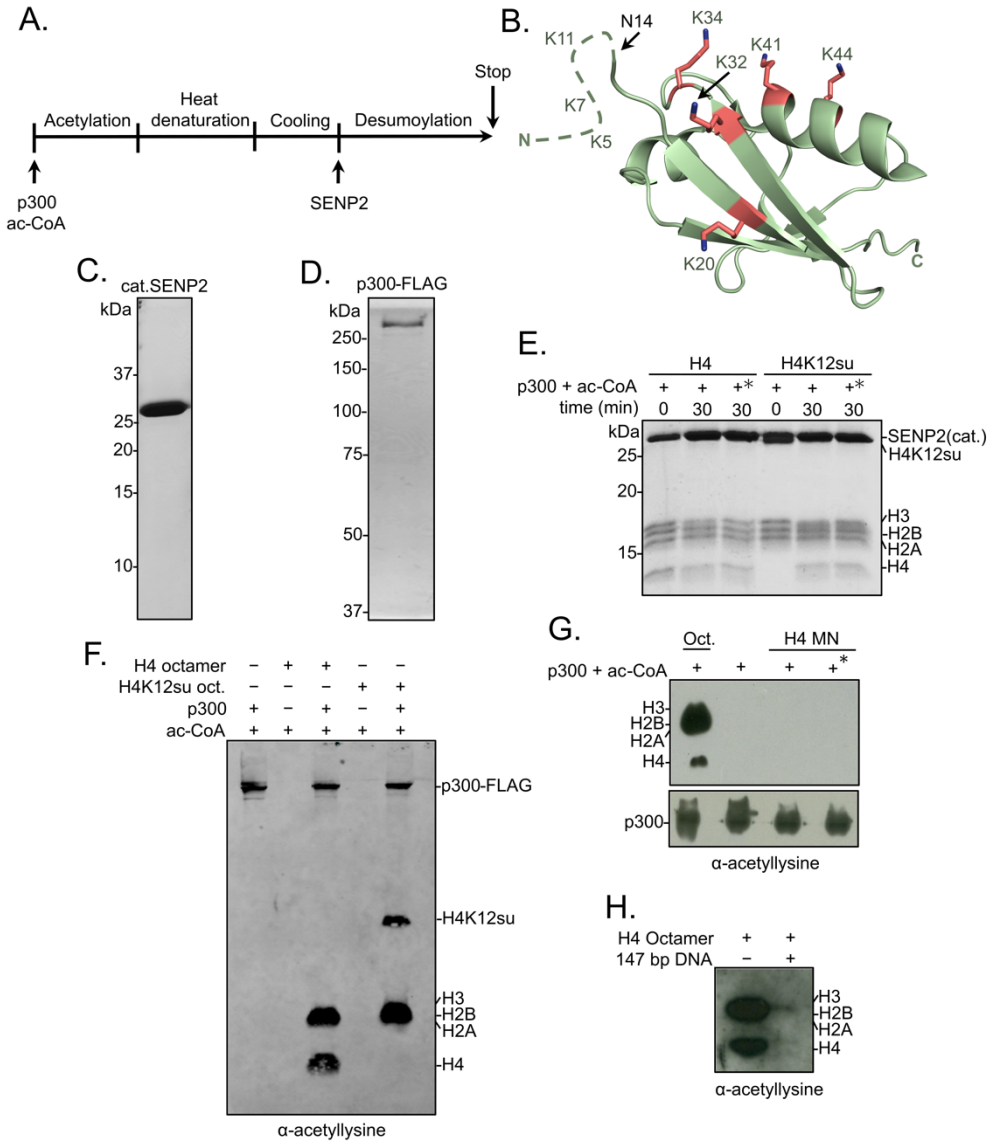


Figure 3.S1. Histone octamer and mononucleosome acetylation by p300. (A) Scheme outlining the p300 histone acetyltransferase assay with octamer and mononucleosome substrates. (B) Cartoon representation of SUMO-3 showing all surface-exposed lysine residues in stick representation. The dashed line represents N-terminal residues not observed in the structure. PDB code 1U4A. (C) Coomassie-stained SDS-PAGE of purified catalytic domain of SENP2, cat.SENP2, consisting of residues 365-590. (D) Coomassie-stained SDS-PAGE of purified p300-FLAG from HEK293T cells. (E) Coomassie-stained SDS-PAGE corresponding to the HAT assay shown in Figure 1E. Asterisk indicates heat-inactivated p300 was used. (F) Histone acetylation assay with octamers containing wt H4 or H4K12su. Autoacetylation of p300 was observed with a pan-acetyllysine antibody. (G) Histone acetylation assay with p300 and wt H4 containing octamers and mononucleosomes. No cat.SENP2 was used in this assay. The asterisk indicates pre-incubation of p300-FLAG with acetyl-CoA for 1 h to allow for the build-up of activating p300-autoacetylation prior to the addition of mononucleosome substrate.[11] (H) Histone acetylation assay with wt H4 octamer with or without equimolar amounts of 147 bp Widom *601* dsDNA.

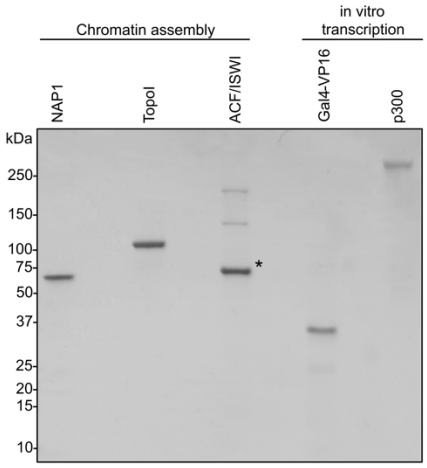
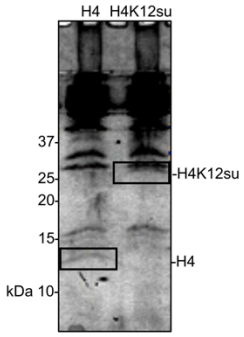
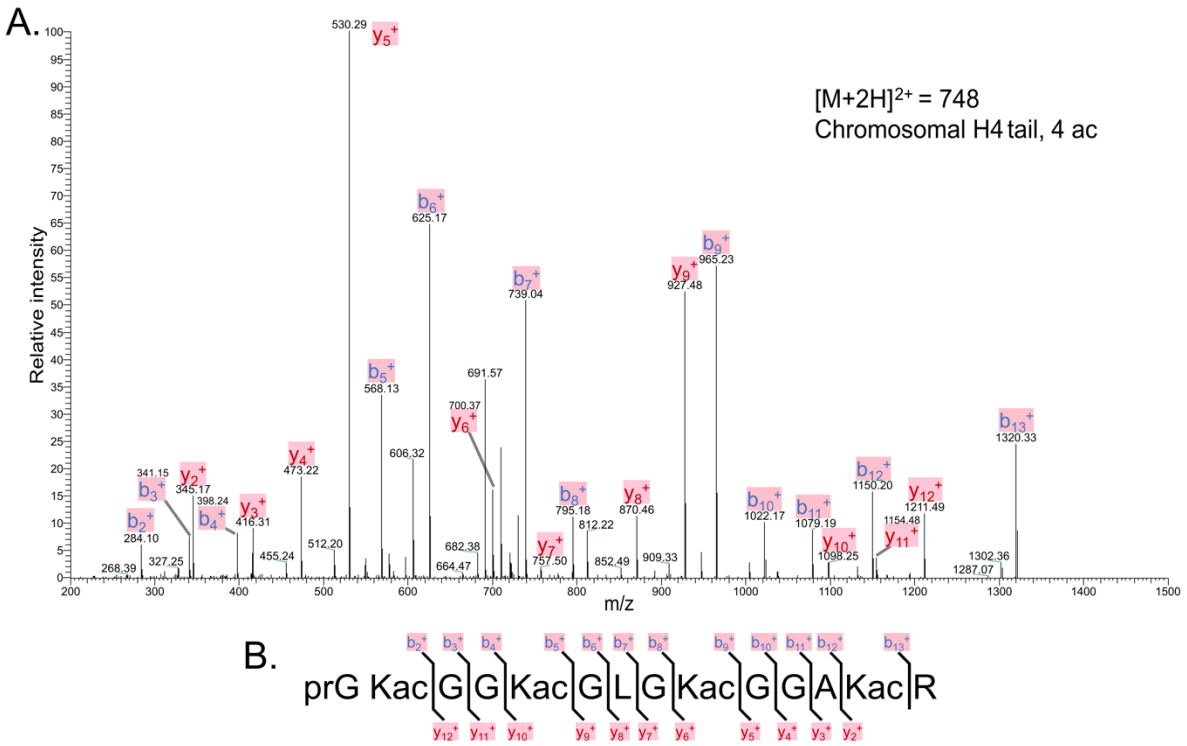


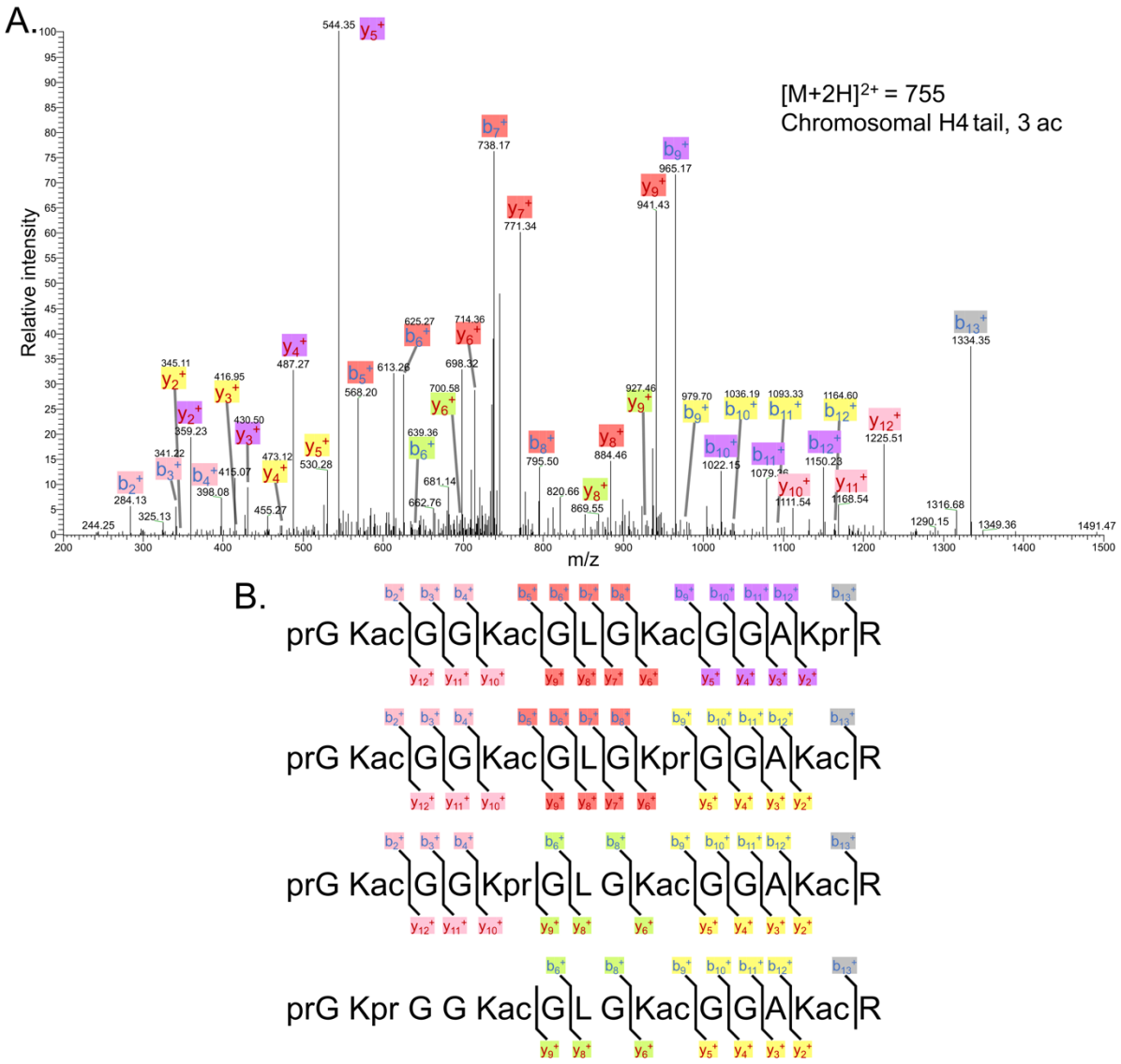
Figure 3.S2. Coomassie-stained SDS-PAGE of chromatin assembly proteins and *in vitro* transcription components. Asterisk indicates BSA used as a stabilizer.



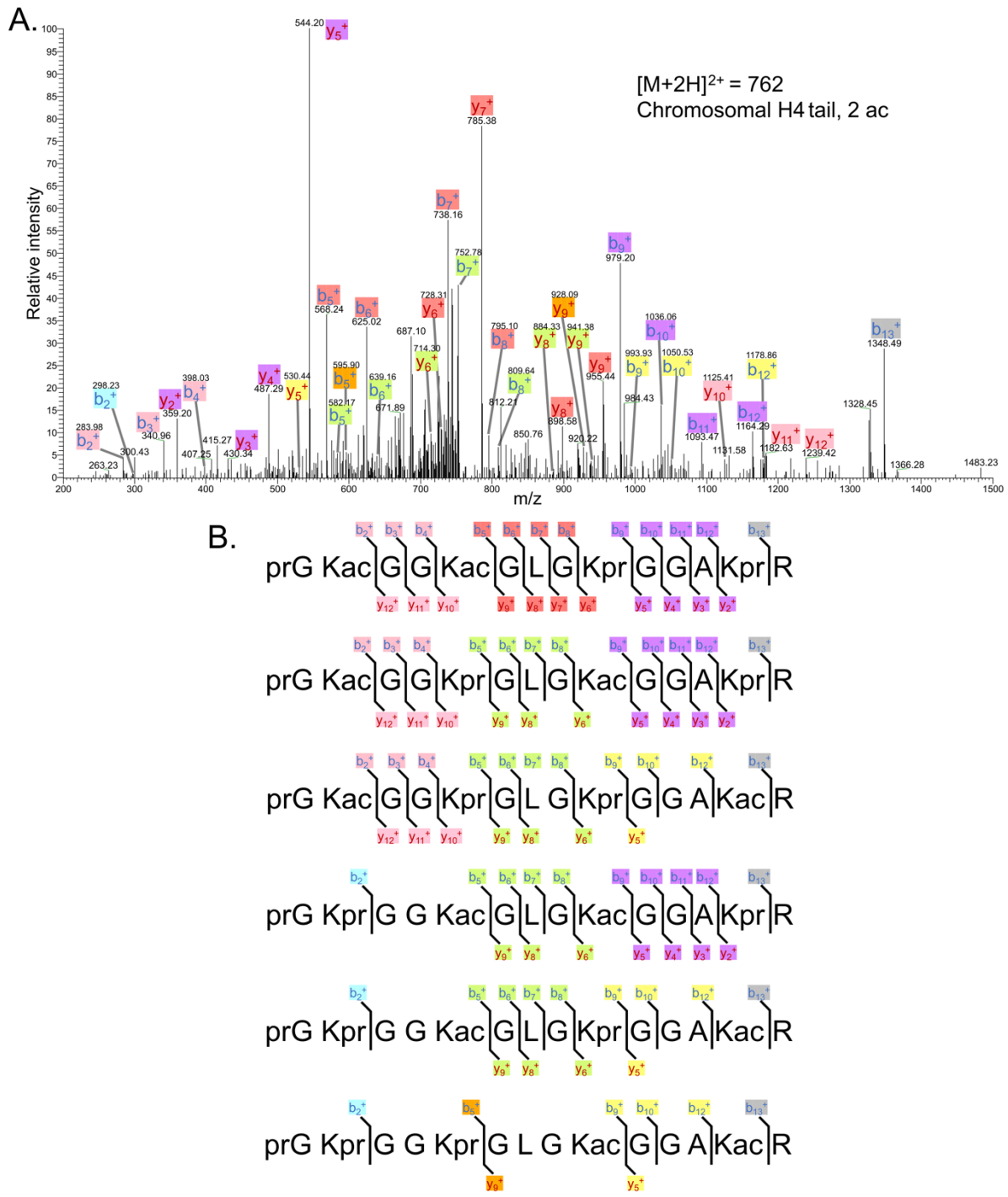
Supplementary Figure 3.S3. Coomassie-stained SDS-PAGE of histone acetylation assay on chromatinized plasmids containing wt H4 or H4K12su with p300 and activator Gal4-VP16. Gel bands excised for MS-MS analysis are indicated.



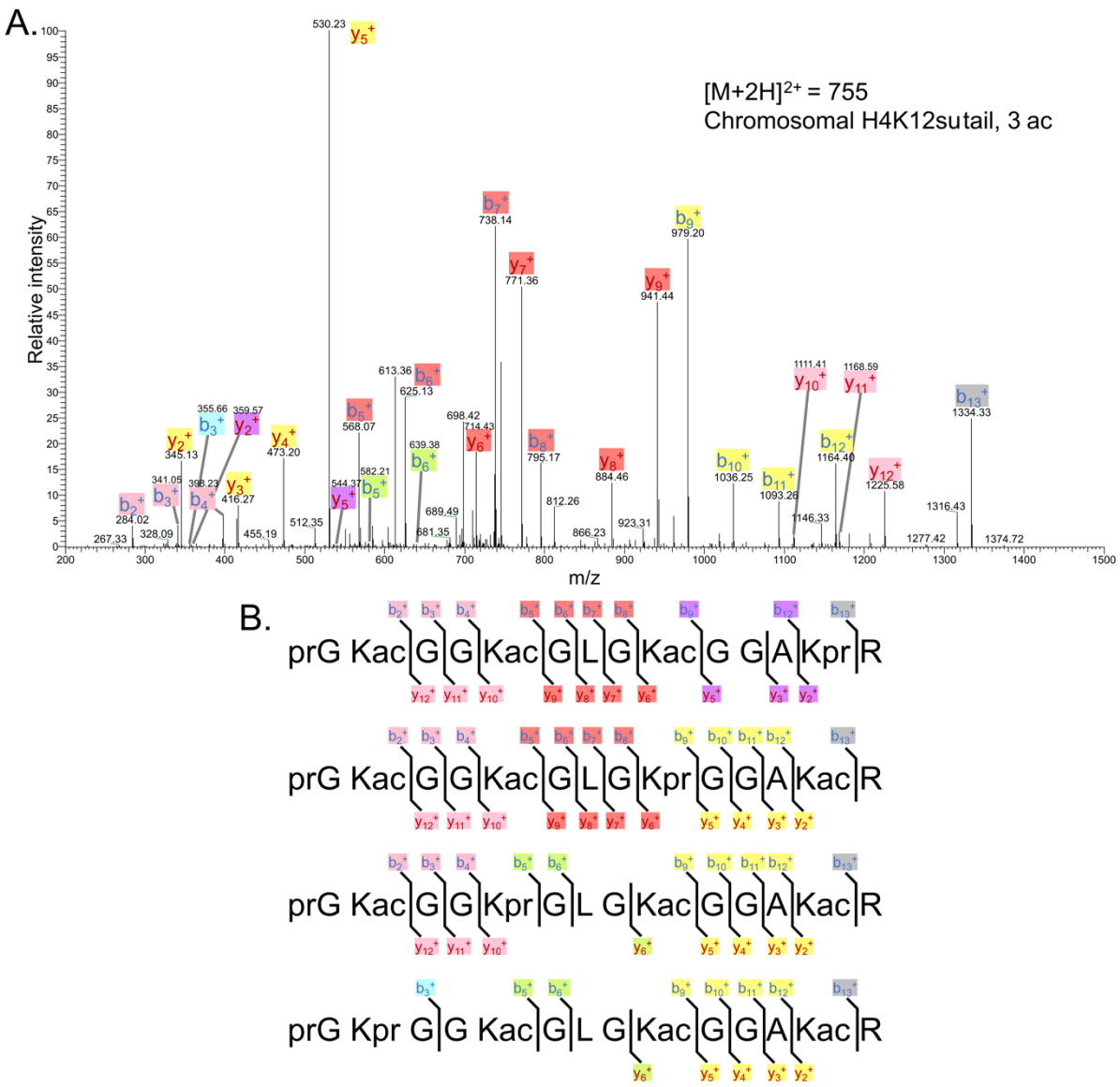
Supplementary Figure 3.S4. (A) Representative MS-MS spectrum of tetra-acetylated tryptic peptide H4(4-17) generated after *in vitro* acetylation of chromatinized plasmids containing wt H4 with p300 and activator Gal4-VP16. (B) Peptide fragment-ion map of the tetra-acetylated H4(4-17) peptide indicating all ions identified over three spectra.



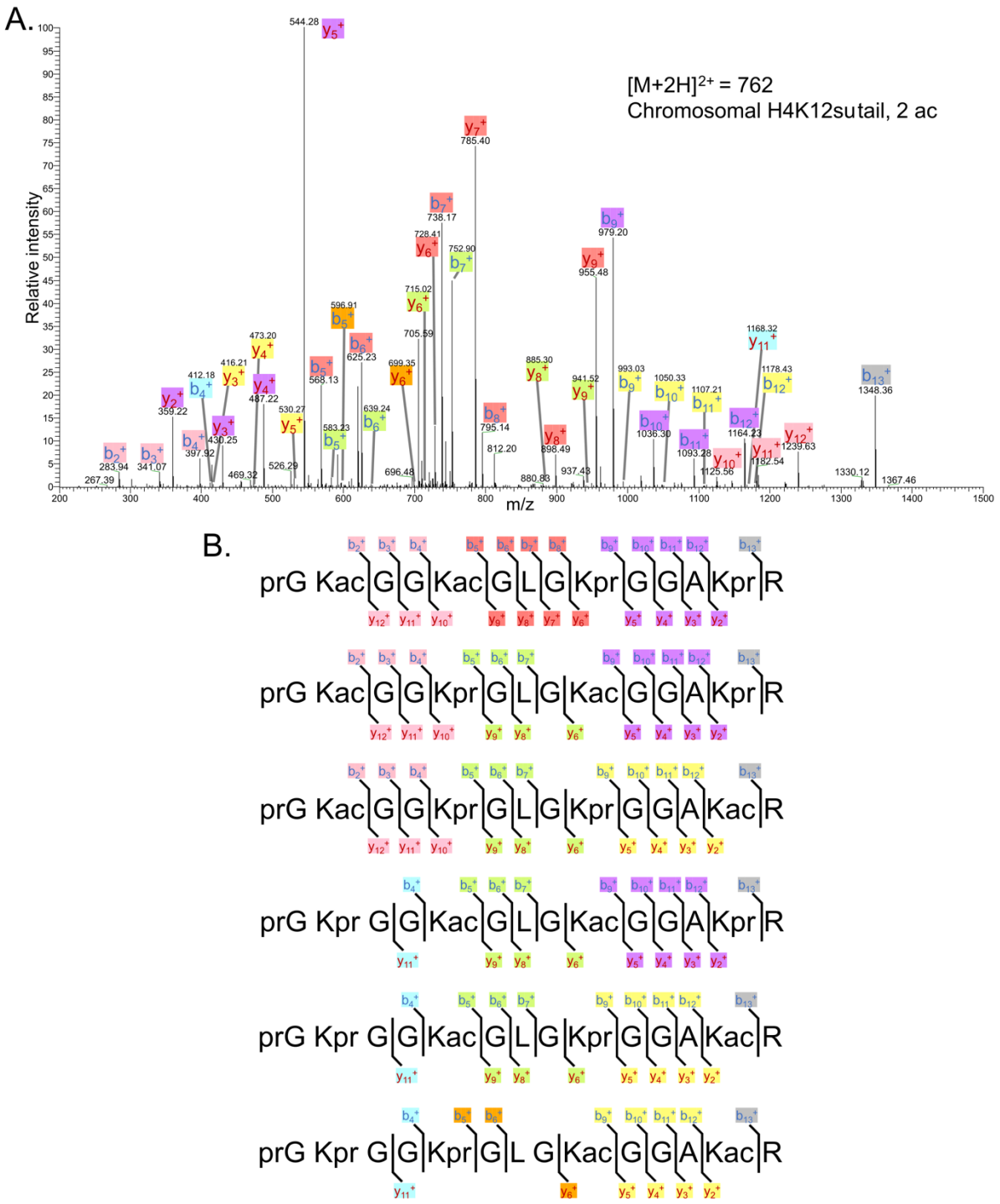
Supplementary Figure 3.S5. (A) Representative MS-MS spectrum of tri-acetylated tryptic peptide H4(4-17) generated after *in vitro* acetylation of chromatinized plasmids containing wt H4 with p300 and activator Gal4-VP16. (B) Peptide fragment-ion maps of the four possible tri-acetylated H4(4-17) peptide patterns, indicating all ions identified over three spectra.



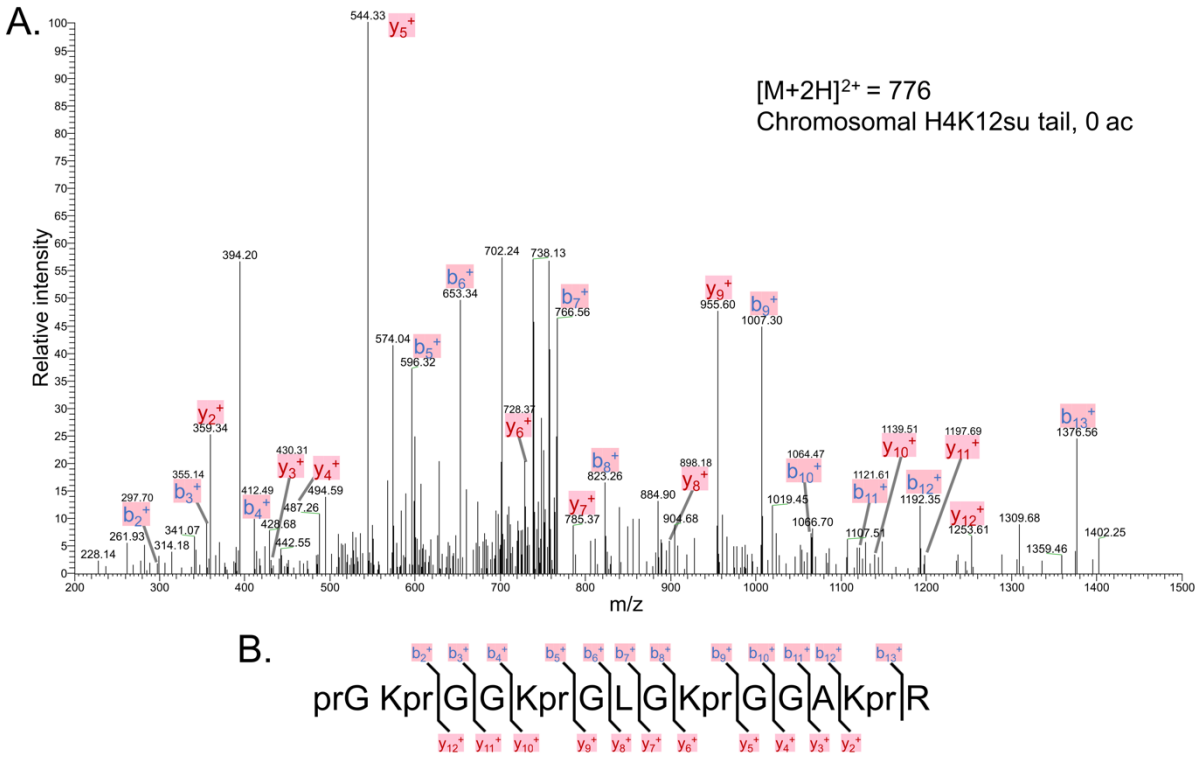
Supplementary Figure 3.S6. (A) Representative MS-MS spectrum of di-acetylated tryptic peptide H4(4-17) generated after *in vitro* acetylation of chromatinized plasmids containing wt H4 with p300 and activator Gal4-VP16. (B) Peptide fragment-ion maps of the six possible di-acetylated H4(4-17) peptide patterns, indicating all ions identified over two spectra.



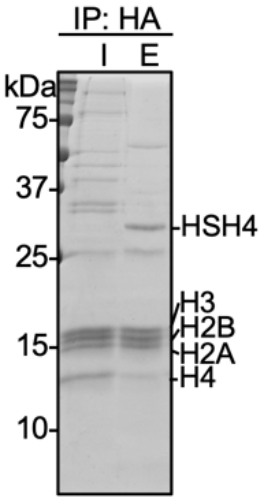
Supplementary Figure 3.S7. Representative MS-MS spectrum of tri-acetylated tryptic peptide H4(4-17) generated after *in vitro* acetylation of chromatinized plasmids containing H4K12su with p300 and activator Gal4-VP16 followed by in-gel desumoylation. (B) Peptide fragment-ion maps of the four possible tri-acetylated H4(4-17) peptide patterns, indicating all ions identified over three spectra. Acetylation on K12 in H4K12su is not possible due to presence of SUMO-3 at K12.



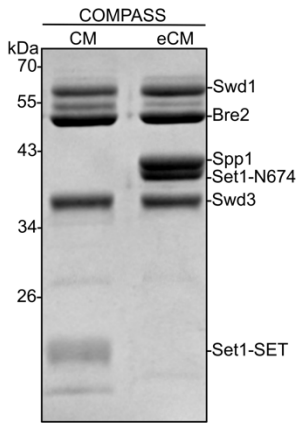
Supplementary Figure 3.S8. Representative MS-MS spectrum of di-acetylated tryptic peptide H4(4-17) generated after *in vitro* acetylation of chromatinized plasmids containing H4K12su with p300 and activator Gal4-VP16 followed by desumoylation. (B) Peptide fragment-ion maps of the six possible di-acetylated H4(4-17) peptide species, indicating all ions identified over three spectra.



Supplementary Figure 3.S9. The MS-MS spectrum of unacetylated tryptic peptide H4(4-17) after *in vitro* acetylation of chromatinized plasmids containing H4K12su with p300 and activator Gal4-VP16 followed by desumoylation. (B) Peptide fragment-ion map of the unmodified H4(4-17) peptide, indicating identified ions.



Supplementary Figure 3.S10. Coomassie-stained SDS-PAGE gel of input (I) and elution (E) samples from immunoprecipitation against HA of MNase digested nuclear extracts prepared from HEK293T cells transfected with HA-Su3(ΔGG)-H4.



Supplementary Figure 3.S11. Coomassie-stained SDS-PAGE gel of recombinant *K. lactis* COMPASS CM and COMPASS eCM sub-complexes used in H3K4 methylation assays. The Sdc1 subunit (10 kDa) is not observed on this gel due to its size. Set1-SET is the catalytic domain. Set1-N674 includes the nSET domain, beginning at residue 674.

3.6 References

Portions of this chapter have been published as:

Leonen, C.J.A.; Shimada, M.; Weller, C.E.; Nakadai, T.; Hsu, P.L.; Tyson, E.L.; Mishra, A.; Shelton, P.M.M.; Sadilek, M.; Hawkins, R.D.; Zheng, N.; Roeder, R.G.; Chatterjee, C. Sumoylation of the human histone H4 tail inhibits p300-mediated transcription by RNA polymerase II in cellular extracts. *eLife* **2021**, 10:e67952.

(1) Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **2007**, *128* (4), 693–705.

(2) Strahl, B. D.; Allis, C. D. The Language of Covalent Histone Modifications. *Nature* **2000**, *403* (6765), 41–45. https://doi.org/10.1038/47412.

(3) Greer, E. L.; Shi, Y. Histone Methylation: A Dynamic Mark in Health, Disease and Inheritance. *Nat. Rev. Genet.* **2012**, *13* (5), 343–357. https://doi.org/10.1038/nrg3173.

(4) MD, S.; M, G. Functions of Site-Specific Histone Acetylation and Deacetylation. *Annu. Rev. Biochem.* **2007**, *76*, 75–100. https://doi.org/10.1146/ANNUREV.BIOCHEM.76.052705.162114.

(5) VM, W.; JL, W. Histone Ubiquitination: Triggering Gene Activity. *Mol. Cell* **2008**, *29* (6), 653–663. https://doi.org/10.1016/J.MOLCEL.2008.02.014.

(6) Shiio, Y.; Eisenman, R. N. Histone Sumoylation Is Associated with Transcriptional Repression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230. https://doi.org/10.1073/pnas.1735528100.

(7) Ryu, H.-Y.; Su, D.; Wilson-Eisele, N. R.; Zhao, D.; López-Giráldez, F.; Hochstrasser, M. The Ulp2 SUMO Protease Promotes Transcription Elongation through Regulation of Histone Sumoylation. *EMBO J.* **2019**, *38* (16), e102003. https://doi.org/10.15252/EMBJ.2019102003.

(8) Issar, N.; Roux, E.; Mattei, D.; Scherf, A. Identification of a Novel Post-Translational Modification in Plasmodium Falciparum: Protein Sumoylation in Different Cellular Compartments. *Cell. Microbiol.* **2008**, *10* (10), 1999. https://doi.org/10.1111/J.1462-5822.2008.01183.X.

(9) Miller, M. J.; Barrett-Wilt, G. A.; Hua, Z.; Vierstra, R. D. Proteomic Analyses Identify a Diverse Array of Nuclear Processes Affected by Small Ubiquitin-like Modifier Conjugation in Arabidopsis. *Proc. Natl. Acad. Sci.* **2010**, *107* (38), 16512–16517. https://doi.org/10.1073/PNAS.1004181107.

(10) Hendriks, I. A.; Vertegaal, A. C. O. A Comprehensive Compilation of SUMO Proteomics. *Nat. Rev. Mol. Cell Biol. 2016 179* **2016**, *17* (9), 581–595. https://doi.org/10.1038/nrm.2016.81.

(11) Ohkuni, K.; Takahashi, Y.; Fulp, A.; Lawrimore, J.; Au, W.-C.; Pasupala, N.; Levy-Myers, R.; Warren, J.; Strunnikov, A.; Baker, R. E.; Kerscher, O.; Bloom, K.; Basrai, M. A. SUMO-Targeted Ubiquitin Ligase (STUbL) Slx5 Regulates Proteolysis of Centromeric Histone H3 Variant Cse4 and Prevents Its Mislocalization to Euchromatin. *Mol. Biol. Cell* **2016**, *27* (9), 1500–1510. https://doi.org/10.1091/MBC.E15-12-0827.

(12) Ryu, H.-Y.; Hochstrasser, M. Histone Sumoylation and Chromatin Dynamics. *Nucleic Acids Res.* **2021**, *49* (11), 6043–6052. https://doi.org/10.1093/NAR/GKAB280.

(13) Galisson, F.; Mahrouche, L.; Courcelles, M.; Bonneil, E.; Meloche, S.; Chelbi-Alix, M. K.; Thibault, P. A Novel Proteomics Approach to Identify SUMOylated Proteins and Their Modification Sites in Human Cells. *Mol. Cell. Proteomics* **2011**, *10* (2), M110.004796. https://doi.org/10.1074/mcp.M110.004796.

(14) Hendriks, I. A.; D'Souza, R. C. J.; Yang, B.; Verlaan-De Vries, M.; Mann, M.; Vertegaal, A. C. O. Uncovering Global SUMOylation Signaling Networks in a Site-Specific Manner. *Nat. Struct. Mol. Biol.* **2014**, *21* (10), 927–936. https://doi.org/10.1038/nsmb.2890.

(15) 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. Histone Sumoylation Is a Negative Regulator in Saccharomyces Cervisiae and Shows

Dynamic Interplay with Positive-Acting Histone Modifications. *Genes Dev.* **2006**, *20*, 966–976. https://doi.org/10.1101/gad.1404206.6.

(16) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long-Range Internucleosomal Interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837. https://doi.org/10.1074/jbc.M114.591644.

(17) Dhall, A.; Weller, C. E.; Chu, A.; Shelton, P. M. M.; Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-CoREST Complex. *ACS Chem. Biol.* **2017**, *12* (9), 2275–2280. https://doi.org/10.1021/acschembio.7b00716.

(18) Dignam, J. D.; Lebovitz, R. M.; Roeder, R. G. Accurate Transcription Initiation by RNA Polymerase II in a Soluble Extract from Isolated Mammalian Nuclei. *Nucleic Acids Res.* **1983**, *11* (5), 1475–1489. https://doi.org/10.1093/nar/11.5.1475.

(19) Roeder, R. G. 50+ Years of Eukaryotic Transcription: An Expanding Universe of Factors and Mechanisms. *Nat. Struct. Mol. Biol. 2019 269* **2019**, *26* (9), 783–791. https://doi.org/10.1038/s41594-019-0287-x.

(20) Kim, J.; Guermah, M.; McGinty, R. K.; Lee, J. S.; Tang, Z.; Milne, T. A.; Shilatifard, A.; Muir, T. W.; Roeder, R. G. RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. *Cell* **2009**, *137* (3), 459–471.

(21) An, W.; Palhan, V. B.; Karymov, M. A.; Leuba, S. H.; Roeder, R. G. Selective Requirements for Histone H3 and H4 N Termini in P300-Dependent Transcriptional Activation from Chromatin Woojin. *Mol. Cell* **2002**, *9*, 811–821.

(22) PJ, R.; W, A.; A, R.; F, M.; L, C.; RG, R.; D, R. 30 Nm Chromatin Fibre Decompaction Requires Both H4-K16 Acetylation and Linker Histone Eviction. *J. Mol. Biol.* **2008**, *381* (4), 816–825. https://doi.org/10.1016/J.JMB.2008.04.050.

(23) Shimada, M.; Chen, W. Y.; Nakadai, T.; Onikubo, T.; Guermah, M.; Rhodes, D.; Roeder, R. G. Gene-Specific H1 Eviction through a Transcriptional Activator→p300→NAP1→H1 Pathway. *Mol. Cell* **2019**, *74* (2), 268-283.e5. https://doi.org/10.1016/j.molcel.2019.02.016.

(24) Weller, C. E.; Huang, W.; Chatterjee, C. Facile Synthesis of Native and Protease-Resistant Ubiquitylated Peptides. *ChemBioChem* **2014**, *15* (9), 1263–1267. https://doi.org/10.1002/CBIC.201402135.

(25) Hsu, P. L.; Shi, H.; Leonen, C.; Kang, J.; Chatterjee, C.; Zheng, N. Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS. *Mol. Cell* **2019**, *76*, 712–723. https://doi.org/10.1016/J.MOLCEL.2019.10.013.

(26) Ryu, H.-Y.; Zhao, D.; Li, J.; Su, D.; Hochstrasser, M. Histone Sumoylation Promotes Set3 Histone-Deacetylase Complex-Mediated Transcriptional Regulation. *Nucleic Acids Res.* **2020**, *48* (21), 12151–12168. https://doi.org/10.1093/NAR/GKAA1093.

(27) Kraus, W. L.; Kadonaga, J. T. P300 and Estrogen Receptor Cooperatively Activate Transcription via Differential Enhancement of Initiation and Reinitiation. *Genes Dev.* **1998**, *12* (3), 331. https://doi.org/10.1101/GAD.12.3.331.

(28) Kundu, T. K.; Palhan, V. B.; Wang, Z.; An, W.; Cole, P. A.; Roeder, R. G. Activator-Dependent Transcription from Chromatin in Vitro Involving Targeted Histone Acetylation by P300. *Mol. Cell* **2000**, *6* (3), 551–561. https://doi.org/10.1016/S1097-2765(00)00054-X.

(29) Mikolajczyk, J.; Drag, M.; Békés, M.; Cao, J. T.; Ronai, Z.; Salvesen, G. S. Small Ubiquitin-Related Modifier (SUMO)-Specific Proteases: Profiling the Specificities and Activities of Human SENPs. *J. Biol. Chem.* **2007**, *282* (36), 26217–26224. https://doi.org/10.1074/jbc.M702444200.

(30) Chen, L.; Mu, Y.; Greene, W. C. Acetylation of RelA at Discrete Sites Regulates Distinct Nuclear Functions of NF-KB. *EMBO J.* **2002**, *21* (23), 6539. https://doi.org/10.1093/EMBOJ/CDF660.

(31) Ogryzko, V. V.; Schiltz, R. L.; Russanova, V.; Howard, B. H.; Nakatani, Y. The Transcriptional Coactivators P300 and CBP Are Histone Acetyltransferases. *Cell* **1996**, *87* (5), 953–959. https://doi.org/10.1016/S0092-8674(00)82001-2.

(32) A, A.; PB, B. Activation of Transcription through Histone H4 Acetylation by MOF, an Acetyltransferase Essential for Dosage Compensation in Drosophila. *Mol. Cell* **2000**, *5* (2), 367–375. https://doi.org/10.1016/S1097-2765(00)80431-1.

(33) Shogren-Knaak, M.; Ishii, H.; Sun, J.-M.; Pazin, M. J.; James R. Davie; Peterson, C. L. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. **2006**, *311* (5762), 844–847. https://doi.org/https://doi.org/10.1126/science.1124000.

(34) BE, S.; S, M.; J, W.; J, T.; MT, C.; R, T.; L, W. Kinetics and Comparative Reactivity of Human Class I and Class IIb Histone Deacetylases. *Biochemistry* **2004**, *43* (34), 11083–11091. https://doi.org/10.1021/BI0494471.

(35) Sidoli, S.; Garcia, B. A. Characterization of Individual Histone Post-Translational Modifications and Their Combinatorial Patterns by Mass Spectrometry-Based Proteomics Strategies. *Methods Mol. Biol.* **2017**, *1528*, 121–148. https://doi.org/10.1146/annurev.bi.48.070179.001111.

(36) Schiltz, R. L.; Mizzen, C. A.; Vassilev, A.; Cook, R. G.; Allis, C. D.; Nakatani, Y. Overlapping

but Distinct Patterns of Histone Acetylation by the Human Coactivators P300 and PCAF within Nucleosomal Substrates. *J. Biol. Chem.* **1999**, *274* (3), 1189–1192. https://doi.org/10.1074/jbc.274.3.1189.

(37) Taipale, M.; Rea, S.; Richter, K.; Vilar, A.; Lichter, P.; Imhof, A.; Akhtar, A. HMOF Histone Acetyltransferase Is Required for Histone H4 Lysine 16 Acetylation in Mammalian Cells. *Mol. Cell. Biol.* **2005**, *25* (15), 6798. https://doi.org/10.1128/MCB.25.15.6798-6810.2005.

(38) Bannister, A. J.; Kouzarides, T. The CBP Co-Activator Is a Histone Acetyltransferase. *Nat. 1996 3846610* **1996**, *384* (6610), 641–643. https://doi.org/10.1038/384641a0.

(39) Santos-Rosa, H.; Schneider, R.; Bannister, A. J.; Sherriff, J.; Bernstein, B. E.; Emre, N. C. T.; Schreiber, S. L.; Mellor, J.; Kouzarides, T. Active Genes Are Tri-Methylated at K4 of Histone H3. *Nat. 2002 4196905* **2002**, *419* (6905), 407–411. https://doi.org/10.1038/nature01080.

(40) A, S. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **2012**, *81*, 65–95. https://doi.org/10.1146/ANNUREV-BIOCHEM-051710-134100.

(41) Tang, Z.; Chen, W.-Y.; Shimada, M.; Nguyen, U. T. T.; Kim, J.; Sun, X.-J.; Sengoku, T.; McGinty, R. K.; Fernandez, J. P.; Muir, T. W.; Roeder, R. G. SET1 and P300 Act Synergistically, through Coupled Histone Modifications, in Transcriptional Activation by P53. *Cell* **2013**, *154* (2), 297–310. https://doi.org/10.1016/J.CELL.2013.06.027.

(42) Worden, E. J.; Zhang, X.; Wolberger, C. Structural Basis for COMPASS Recognition of an H2B-Ubiquitinated Nucleosome. *Elife* **2020**, *9*. https://doi.org/10.7554/ELIFE.53199.

(43) Xue, H.; Yao, T.; Cao, M.; Zhu, G.; Li, Y.; Yuan, G.; Chen, Y.; Lei, M.; Huang, J. Structural Basis of Nucleosome Recognition and Modification by MLL Methyltransferases. *Nat. 2019 5737774* **2019**, *573* (7774), 445–449. https://doi.org/10.1038/s41586-019-1528-1.

(44) Brown, D. A.; Cerbo, V. Di; Feldmann, A.; Ahn, J.; Ito, S.; Blackledge, N. P.; Nakayama, M.; McClellan, M.; Dimitrova, E.; Turberfield, A. H.; Long, H. K.; King, H. W.; Kriaucionis, S.; Schermelleh, L.; Kutateladze, T. G.; Koseki, H.; Klose, R. J. The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin. *Cell Rep.* **2017**, *20* (10), 2313. https://doi.org/10.1016/J.CELREP.2017.08.030.

(45) *Epigenetics*, 2nd ed.; Allis, C. D., Caparros, M.-L., Jenuwein, T., Reinberg, D., Eds.; Cold Spring Harbor Laboratory Press, 2015.

(46) Uckelmann, M.; Densham, R. M.; Baas, R.; Winterwerp, H. H. K.; Fish, A.; Sixma, T. K.; Morris, J. R. USP48 Restrains Resection by Site-Specific Cleavage of the BRCA1 Ubiquitin

Mark from H2A. *Nat. Commun. 2018 91* **2018**, *9* (1), 1–16. https://doi.org/10.1038/s41467-017-02653-3.

(47) Hsu, P. L.; Li, H.; Lau, H.-T.; Leonen, C.; Dhall, A.; Ong, S.; Chatterjee, C.; Zheng, N. Crystal Structure of the COMPASS H3K4 Methyltransferase Catalytic Module. *Cell* **2018**, *174*, 1106–1116. https://doi.org/10.1016/J.CELL.2018.06.038.

(48) Lee, J. H.; Skalnik, D. G. CpG-Binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. *J. Biol. Chem.* **2005**, *280* (50), 41725–41731. https://doi.org/10.1074/jbc.M508312200.

(49) Meeks, J. J.; Shilatifard, A. Multiple Roles for the Mll/Compass Family in the Epigenetic Regulation of Gene Expression and in Cancer. *Annu. Rev. Cancer Biol.* **2017**, *1*, 425–446. https://doi.org/10.1146/annurev-cancerbio-050216-034333.

(50) Clouaire, T.; Webb, S.; Skene, P.; Illingworth, R.; Kerr, A.; Andrews, R.; Lee, J. H.; Skalnik, D.; Bird, A. Cfp1 Integrates Both CpG Content and Gene Activity for Accurate H3K4me3 Deposition in Embryonic Stem Cells. *Genes Dev.* **2012**, *26* (15), 1714–1728. https://doi.org/10.1101/gad.194209.112.

(51) Dyer, P. N.; Edayathumangalam, R. S.; White, C. L.; Bao, Y.; Chakravarthy, S.; Muthurajan, U. M.; Luger, K. Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. *Methods Enzymol.* **2003**, *375*, 23–44. https://doi.org/10.1016/S0076-6879(03)75002-2.

(52) Reverter, D.; Lima, C. D. Preparation of SUMO Proteases and Kinetic Analysis Using Endogenous Substrates. *Methods Mol. Biol.* **2009**, *497* (4), 1–13. https://doi.org/10.1007/978-1-59745-566-4.

(53) Pronobis, M. I.; Deuitch, N.; Peifer, M. The Miraprep: A Protocol That Uses a Miniprep Kit and Provides Maxiprep Yields. *PLoS One* **2016**, *11* (8), 1–12. https://doi.org/10.1371/journal.pone.0160509.

(54) 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. HDAC1 Acetylation Is Linked to Progressive Modulation of Steroid Receptor-Induced Gene Transcription. *Mol. Cell* **2006**, *22* (5), 669–679. https://doi.org/10.1016/j.molcel.2006.04.019.

(55) An, W.; Roeder, R. G. Reconstitution and Transcriptional Analysis of Chromatin In Vitro. *Methods Enzymol.* **2004**, *377*, 460–474. https://doi.org/10.1016/S0076-6879(03)77030-X.

(56) Hawkins, R. D.; Hon, G. C.; Lee, L. K.; Ngo, Q.; Lister, R.; Pelizzola, M.; Edsall, L. E.; Kuan, S.; Luu, Y.; Klugman, S.; Antosiewicz-Bourget, J.; Ye, Z.; Espinoza, C.; Agarwahl, S.; Shen, L.; Ruotti, V.; Wang, W.; Stewart, R.; Thomson, J. A.; Ecker, J. R.; Ren, B. Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. *Cell Stem Cell* **2010**, *6* (5), 479–491. https://doi.org/10.1016/J.STEM.2010.03.018.

Chapter 4

Biochemical investigation of histone sumoylation in human cells

4.1 Introduction

Histone sumoylation is the addition of the small ubiquitin-like modifier onto the lysine ε-amine of substrate proteins. Within humans, there are 5 identified isoforms of SUMO, with SUMO1-3 being the major isoforms and the most studied thus far (**Figure 4.1A**). SUMO2 and SUMO3 are ~97% identical and are often grouped together as SUMO2/3, with SUMO1 sharing only 50% identity with SUMO2/3. Much less is known about the functions of SUMO4 and SUMO5. SUMO1-3 are typically conserved in higher eukaryotes. In yeast, there is only a single SUMO protein, Smt3 (generally referred to as SUMO and sumoylation in yeast studies). Sumoylation occurs through a process involving an E1-E2-E3 enzymatic cascade that is similar to, but distinct from, the conjugation of ubiquitin and other ubiquitin-like proteins to their targets (**Figure 4.1B**).[1] The ATP-dependent SAE1/SAE2 (Aos1/Uba2 in yeast) E1 first activates mature SUMO, which is then transferred to the sole E2 SUMO-ligase, UBC9. From its sumoylated thioester form, UBC9 can transfer SUMO directly onto a substrate lysine or engage with an E3 that helps Ubc9 sumoylate a specific substrate. Typically, SUMO is conjugated at a ψKXE consensus motif where ψ is a hydrophobic residue, K is the lysine targeted, and X is any amino acid.[2,3] Seven highly active SUMO endopeptidases (SENP1-7 in humans and Ulp1/2 in yeast) function to reversibly remove SUMO from lysine sidechains as well as to convert nascent polySUMO proteins and pre-SUMO peptides to the mature form by cleaving after the *C*-terminal Gly-Gly motif in these proteins after they are translated on ribosomes.[4]

A.

```
(hs)SUMO1      MSDQE----AKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQG  56
(hs)SUMO2      MADEK----P----KEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQG  52
(hs)SUMO3      MSEEK----P----KEGVKTE-NDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQG  51
(sc)Smt-3      MSDSEVNQEAKPEVK--PEVKPETHINLKVS-DGSSEIFFKIKKTTPLRRLMEAFAKRQG  57
(ce)Smo-1      MA----------DDAAQAGDNAEYIKIKVVGQDSNEVHFRVKYGTSMAKLKKSYADRTG  49

(hs)SUMO1      VPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGG 97
(hs)SUMO2      LSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG 93
(hs)SUMO3      LSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG 92
(sc)Smt-3      KEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG 98
(ce)Smo-1      VAVNSLRFLFDGRRINDDDTPKTLEMEDDDVIEVYQEQLGG 90
```

B.

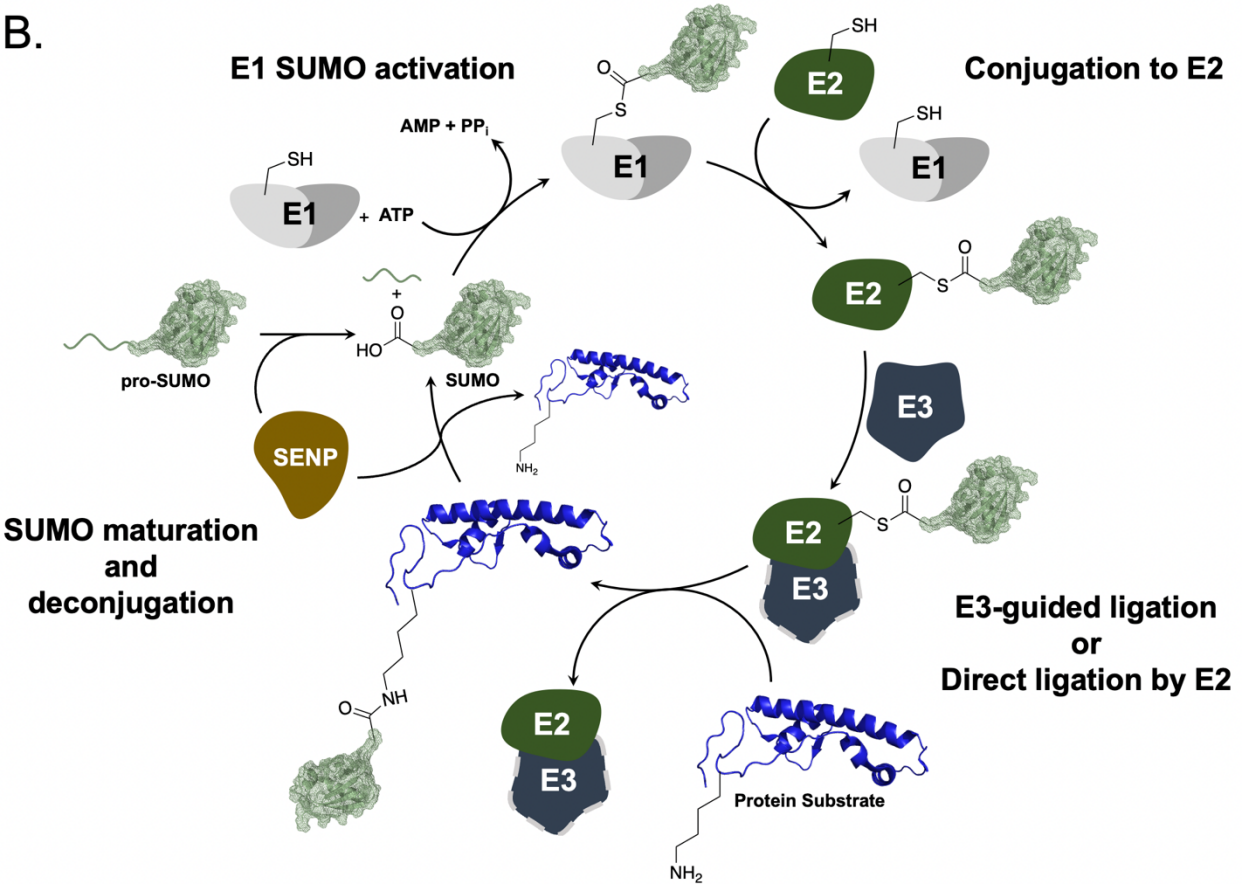


Figure 4.1 SUMO is well conserved and modifies protein lysines via an E1-E2-E3 cascade. (A.) Sequence alignment of human (hs) SUMO isoforms and SUMO orthologs in yeast *S. cerevisiae* (sc) and *C. elegans* (ce). (B.) Schematic enzymatic cascade for protein substrate sumoylation. Immature pro-SUMO is first cleaved by SENPs after a Gly-Gly motif (boxed in A.) to yield mature SUMO that is activated by the E1. Activated SUMO is transferred to the E2 conjugating enzyme. A SUMO E3 ligase may participate to direct sumoylation of a specific substrate that the E3 binds, or, the E2 can directly sumoylate substrate lysine residues. SENPs can also deconjugate SUMO from substrates at the isopeptide bond. Protein substrate PDB: 1KX5, SUMO PDB: 1U4A.

The protein sumoylome is incredibly large and diverse, covering proteins involved in nearly all aspects of nuclear function. This includes replication, transcription, DNA-damage response, and even cell division.[5–9] Interestingly, SUMO2/3 is >90% conjugated to protein substrates in the immortalized HEK293 cell line. However, larger pools of unconjugated SUMO2/3 are found across primary mouse tissues harvested from various organs (~52% conjugated) in large proteomics studies, suggesting stark differences in the "sumoylome" between immortalized cell lines and primary cells.[10] The majority of protein sumoylation occurs within the nucleus of cells, and particularly in promyelocytic leukemia (PML) bodies, on transcription factors (TFs), DNA-damage response proteins, and chromatin-modifying proteins.[3] Transcription factor sumoylation is typically characterized by increased binding specificity upon sumoylation, either by inducing conformational changes or through modulating interactions with other TFs and DNA-interacting proteins.[9]

Mass spectral (MS) proteomics studies have been critical in identifying sumoylated proteins and their specific lysine sites. However, biochemical confirmation is often necessary in order to verify mass-spectral data and to determine biological relevance. Among the proteins that are known to be sumoylated are the histones. All four core histones have been identified in humans to be sumoylated by SUMO2/3, including the histone variants H2A.X, H2A.Z, and H3.3 (**see Table 4.1**). SUMO1 was not initially identified to modify the core histones in humans, but mass-spectral studies have further identified all four core histones and the linker histone H1 as substrates for SUMO1 upon inhibition of the proteasome system with the inhibitor MG132.[11] Both H3 and H1 were further validated in *in vitro* sumoylation assays as SUMO1 substrates. A very recent study has linked PIAS1 as a SUMO1 E3 ligase for H3, H2A, and H2B sumoylation when ectopically co-expressed in HEK239T cells.[12] The authors also observed direct binding of PIAS1 and other PIAS family proteins to H3 and H2A.Z. Interestingly, H4 sumoylation was not stimulated by PIAS1 co-expression, suggesting that H4 sumoylation may not regulated by PIAS1 or may be preferentially

Table 4.1 Histones and sites identified to be sumoylated in human and yeast cells via mass spectrometry and/or *in vivo* biochemical analysis.[a]

Organism	Histone	Sites	Ref
H. sapiens	H2A	5, 75, 95, 97, 118, 119¨, 125, 127	10,11[b], **12**[b],13,14
	H2A.J	118	15
	H2A.V	120	14
	H2A.X	**5, 9, 13, 15, 118, 119¨, 127¨, 134‴**	10,15,16[c]
	H2A.Z	101, 115, 120	10,14
	macroH2A	72, 119, 122, 133, 166, 188, 238, 250, 294, 322	10
	H2B	4, 5¨, 11, 20¨, 43, 46¨, 57, 108¨, 116¨, 120¨,[d]	10–14
	H3	14¨, 18¨, 23‴, 27, 37, 56¨, 79, 115, 122	10,**11**,**12** 13,15,17
	H3.3	37, 122	10
	H3.1t	23, 27	10
	H4	5, 8, 12‴, 16, 20, 77, 91	**6**[e],10,11, 14,15,17
	H1	No sites identified	**11**
S. cerevisiae	H2A	**126**[f]	**18**
	H2A.Z	**126, 133**	**18,19**
	H2B	**6/7, 16/17**	**18,20**
	H3	No sites identified	**18**
	Cse4	**65, 215, 216**	**21–23**
	H4	**5, 8, 12, 16, 20**	**18,20**

[a]Bold-face indicates some degree of *in vivo* biochemical analysis. In humans, histone sumoylation by SUMO2/3 was identified unless otherwise noted. [b]Histone sumoylation observed by SUMO1. [c]H2A.X preferentially modified by SUMO1, residually by SUMO3; mentioned observation of all core histones modified by SUMO1 but no data was shown. [d]Possibly also on H2B.S. [e]Reported sumoylation observed on H3, H2A, and H2B but did not show data. [f]Mutation at this site did not result in a change in histone sumoylation levels and ID was by mass spec. ¨Two dots "‥", or more, indicates two or more MS identifications and/or *in vivo* biochemical validations of site modification in independent publications.

sumoylated by SUMO2/3 instead. All four core histones, H2A.Z, and the centromeric histone Cse4 are sumoylated in yeast (**Table 4.1**). Poly-sumoylation has only been biochemically characterized in yeast; however, mass spectral studies where SENP6, which depolymerizes poly-SUMO chains, was knocked down with shRNAs in human U2OS cells resulted in the enrichment of His10-SUMO2 conjugates for all four core histones.[13] Importantly, this only suggested that these

histones may be poly-sumoylated and implicated SENP6 in their regulation as these assays did not specifically enrich for poly-sumo conjugated proteins but for all His10-SUMO2 conjugates. In the nematode, *Caenorhabditis elegans*, histones H3, H4, and the linker histone H1 were also identified to be sumoylated under stress conditions in SILAC-based (stable isotopic labeling of amino acids in cell culture) mass-spectral experiments.[24] Interestingly, in mouse breast cancer model cells, histones H3, H4, and H2B were observed to be sumoylated in SILAC-based experiments looking at changes in SUMO2/3 protein conjugation arising from the progression to metastisis.[25] Histone sumoylation has also been observed in the parasite *Plasmodium falciparum* (on H4), which is the agent that causes malaria, and in the plant *Arabidopsis thaliana* (on H2B).[26,27] The observation of histone sumoylation across a diverse group of eukaryotes, including unicellular parasites and fungi, multicellular invertebrates, insects, plants, and chordate animals demonstrates that this histone modification is evolutionarily conserved and likely plays important roles in the organism. The functions of histone sumoylation are actively being investigated by several groups, with the primary observations being the regulation of transcription in both humans and yeast. Interestingly, the role histone sumoylation plays in transcription appears to depend on the specific histone it modifies; however, these observations have only been studied at different histones in human cells and yeast and are yet to be confirmed as general/identical mechanisms in both organisms. Additionally, the specific site of sumoylation within a histone may also have differential effects on transcription that awaits elucidation. Thus, the characterization of specific histone and lysine sumoylation is incomplete and warrants urgent investigation as a means to understand the histone code for gene function.

Histone sumoylation was first observed by Shiio and Eisenman in 2003 on all four canonical histones, although only H4 was predominantly sumoylated and further interrogated.[6] The authors found that SUMO-2/3 modifies H4 at its *N*-terminus, and through an artificial *in vivo* Gal4-reporter system, suggested that H4 sumoylation was associated with transcriptional repression. In

immunoprecipitation of sumoylated H4 (H4su) was also the concomitant appearance of heterochromatin protein 1 (HP1), the histone deacetylase 1 (HDAC1). The observation of reduced H3 acetylation upon chromatin sumoylation was also consistent with a role for H4su in gene repression. It was not until 2014 that the first sumoylation sites in histones were identified in H4 at Lys12 and H3.2 at Lys23. This spurred investigations by the Chatterjee lab, due to their expertise in histone semisynthesis, to uncover the site-specific effects of H4 Lys12 sumoylation (H4K12su) in chromatin and on transcription. Using a disulfide linked H4K12su ($suH4_{ss}$), we observed that chromatin arrays resisted compaction when $suH4_{ss}$ was homogenously incorporated in place of H4.[28] This observation was contrary to the "expected" compacted chromatin arrays that have been previously associated with transcriptionally repressive marks such as H3K9me3. For example, similar chromatin compaction assays with ubiquitylated H2A, which is associated with heterochromatin formation and transcriptional repression, revealed no inhibitory effects on chromatin compaction and fiber association.[29] Another study found that when ubiquitin was attached to H2B, a mark associated with active gene transcription, it actually opposed the compaction of nucleosome arrays that resemble chromatin.[30] These experiments are clear examples of the complexity of structural effects from ubiquitin and ubiquitin-like proteins in chromatin. Furthermore, these data suggest that despite being large protein modifications, steric bulk alone cannot be presumed to be the major factor in influencing the dynamic structure of chromatin. The use of chemically defined chromatin arrays also made these analyses possible and highlights the need for protein semisynthesis in the categorical investigation of histone PTMs. In our lab, Caroline Weller developed an auxiliary-mediated semisynthetic method for traceless synthesis of natively linked H4K12su for biochemical assays, which often require reducing conditions.[31] Using this semisynthetic H4K12su, our lab demonstrated the stimulation of the transcriptionally repressive LSD1-CoREST1 subcomplex on nucleosomes reconstituted with H3 dimethylated at Lys4 (H3K4me2).[32] The deacetylase activity of the CoREST1-HDAC1 subcomplex was also stimulated on reconstituted nucleosomes containing H4K12su and H3

acetylated at various lysines (Chatterjee Lab, unpublished results). Together these results strengthen our hypothesis that H4K12su recruits the LSD1-CoREST1-HDAC1 complex to remove histone PTMs associated with active transcription. In addition to these studies, I have recently demonstrated that H4K12su engages in negative biochemical crosstalk with H4 tail acetylation *in cis* via the acetyltransferase p300, and H3K4 methylation *in trans* via the yeast Set1/COMPASS complex. The inhibition of enzymatic activities by H4K12su likely also plays a part in transcriptional repression. Importantly, reconstituted chromatin templates containing H4K12su directly inhibited GAL4-VP16 activated and p300-dependent *in vitro* transcription in nuclear extracts. Analysis of the H4 tail modifications after GAL4-VP16 and p300 acetylation of chromatin templates, prior to transcription, revealed diminished acetylation of lysine residues in H4K12su but not in H4. This suggested that inhibition of H4 tail acetylation by H4K12su may inhibit *in vitro* transcription. Inhibition of *in vitro* transcription via the reduction of H4 acetylation is consistent with previous studies.[33–35]

Studies of histone sumoylation in yeast have benefitted from the genetic tractability of *Saccharomyces cerevisiae* (baker's yeast), *Schizosaccharomyces pombe* (budding yeast), and other yeast model organisms. In fact, sumoylation was the first histone PTM associated with transcription repression studied in *S. cerevisiae*.[18] Nathan et al. identified all four core histones in *S. cerevisiae* to be sumoylated, which still accounted for less than 5% of total cellular histones, and that sumoylated H2B and H4 were associated with transcriptional repression *in vivo*.[18] They also found that histone sumoylation opposed and antagonized histone PTMs associated with active transcription, like H2B ubiquitylation and histone acetylation. Surprisingly, chromatin immunoprecipitation (ChIP) experiments revealed that sumoylated H2B occurred throughout the genome with a slight increase in occupancy near telomeres.[18] Additionally, ChIP studies using an antibody against yeast Smt3 (SUMO) showed sumoylated protein occupancy at constitutively active genes in *S. cerevisiae*, in stark contrast with SUMO's putative role in transcription

repression.[36,37] Ryu et al. recently expanded our understanding of yeast histone sumoylation dynamics and crosstalk with H2B ubiquitylation, observing that the catalytic removal of poly-SUMO chains on H2B by the SUMO protease, Ulp2, is important for transcription elongation.[20] H2BK123ub (K120 in humans) was also observed to promote H2B and H4 poly-sumoylation, demonstrating a positive crosstalk between these two histone modifications in cells lacking Ulp2.[20] Persistent H2B ubiquitylation and poly-sumoylation inhibited the recruitment of the kinase Ctk1 (pTEFb/CDK9 in animals), which is responsible for phosphorylation of Ser2 in the *C*-terminal domain of RNA polymerase II subunit Rpo21/Rpb1. Removal of both H2B ubiquitylation and poly-sumoylation were necessary for transcription, and it is still unclear why the cell evolved this sequential addition of ubiquitin and SUMO followed by their removal for transcription to occur. H2B ubiquitylation is known to engage in crosstalk with H3K4me1/2/3 methylation by the SET1/COMPASS complex and H3K79me1/2 methylation by Dot1.[38] The methylation of both of these lysine residues in H3 are associated with active transcription.[39–41] Interestingly, Dot1 binding to the nucleosome is also stimulated by H4K16ac, coupling H4 acetylation to the orchestration of histone methylations implicated in transcription.[42] H2BK123 ubiquitylation itself is installed by the yeast E2-E3 ligases Bre1-Rad6 (UBE2A/B and RNF20/40, respectively, in humans), which associates with the critically important transcription elongation factor, Paf1 complex (Paf1C).[38,43] Furthermore, histone sumoylation repressed transcription of ncRNAs at internal cryptic promoter sites through the recruitment of the Set3 histone deacetylase complex (Set3C) to remove histone acetylation at the 5' ends of mRNA genes.[44] Nevertheless, the regulation of histone desumoylation by Ulp2 to facilitate active transcription and the maintenance of sumoylation to inhibit spurious transcription at cryptic promoter sites, while very interesting still remain unresolved with respect to the precise timing of sumoylation and the players involved.

Studies of histone sumoylation in humans and higher metazoans have been vastly limited by the complexity of SUMO isoforms and the lack of specific antibodies. Investigations into conservation

of the proposed roles for yeast H2B sumoylation in higher eukaryotes are urgently needed. Neither appreciable amounts of H2B sumoylation nor the poly-sumoylation of histones are currently known in human cells. In fact, histone sumoylation in human cells has only been biochemically observed after the transfection of plasmids containing epitope-tagged histone and SUMO, and most information comes from MS-based proteomic studies. Therefore, there is a need to investigate the signals that result in the deposition of SUMO onto histones and to identify specific E3 ligases involved in histone sumoylation. In this chapter I provide successful results from our initial attempts to identify histone sumoylation in human cells using transfected plasmids. The unprecedented observation of H2B sumoylation (H2Bsu) in human cells has set the stage for new investigations in our lab that will focus on the mechanistic roles for H2Bsu in comparison with H4su.

4.2 Results and discussion

4.2.1 Histones H4 and H2B are sumoylated in HEK293T cells

Currently there are only two publications that report and provide data for histone sumoylation in human cells using non-mass spectrometric methods (**Table 4.1**). However, no sites were identified for histone H4 or H2B sumoylation, although the latter has now been studied in yeast by several groups. A couple of studies have also reported, but did not show data for, sumoylation by SUMO1 or SUMO2/3 present on all four core histones. To address the small amount of western blot-based evidence of histone sumoylation in human cells and the lack of biochemical validation of the sites, we decided to replicate the first experiments showing histone H4 sumoylation in HEK293T cells, as well as to identify and biochemically validate sumoylation sites by bottom-up mass spectrometry. Therefore, HEK293T cells were cultured and transiently transfected with plasmids encoding FLAG-tagged H4 and HA-tagged SUMO3 that contains a Q87R mutation. This mutation has been previously used in proteomics studies in order to introduce a trypsin cleavage

site at the *C*-terminus of SUMO, allowing for ease of identification of SUMO sites via bottom-up MS analysis.[15] This cleavage site was necessary in order to generate a small SUMO peptide remnant consisting of H_2N-QQTGG- that would serve to identify the specific lysine site modified. This position was specifically chosen to mirror the yeast SUMO, Smt3, which has an Arg at this site instead of Glu (**Figure 4.1A**). A plasmid encoding HA-SUMO3(Q87R,Q89P) was also generated and used. The Q89P mutation was previously used in a study to prevent SUMO deconjugation by SUMO proteases.[45]

Cells co-transfected with these plasmids were grown in the presence of transfection medium for 24 h and then the media was changed. After an additional 24 h of growth, the cells were collected and lysed using trichloroacetic acid (TCA). This specific method of cell lysate preparation was necessary in order to protect protein sumoylation from the highly active cellular SENPs. FLAG-H4 was expressed in cells when probed with an anti-FLAG antibody in immunoblots (**Figure 4.2A input**). However, there were no higher molecular weight (MW) FLAG-tagged bands observed in the input or after enrichment by immunoprecipitation (**Figure 4.2A**). This suggested that under these specific conditions, FLAG-H4 was not detectably sumoylated in cells. Both HA-SUMO3s also expressed in the cotransfected HEK293T cells (**Figure 4.2B input**). Interestingly, HA-reactive higher MW bands were observed in the immunoprecipitates for both forms of SUMO used (Q87R and Q87R/Q89P) and these surprisingly corresponded to the expected size for the epitope-tagged sumoylated H4 (**Figure 4.2B IP**). This therefore suggested that H4 may be sumoylated in cells despite the absence of a cross-reactive FLAG-H4 band at the appropriate size where HA-SUMO was observed. We speculated that this may be due to (1) endogenous H4 being sumoylated by HA-SUMO3, and/or (2) due to poor detection of the small amounts of FLAG-tagged H4su in immunoblots with the anti-FLAG antibody. Importantly, HA-SUMO3 did not non-specifically bind to the FLAG-resin (**Figure 4.S2**), suggesting that the HA-reactive bands in co-transfected lysates are covalently bound to FLAG-H4. This also suggested that the FLAG

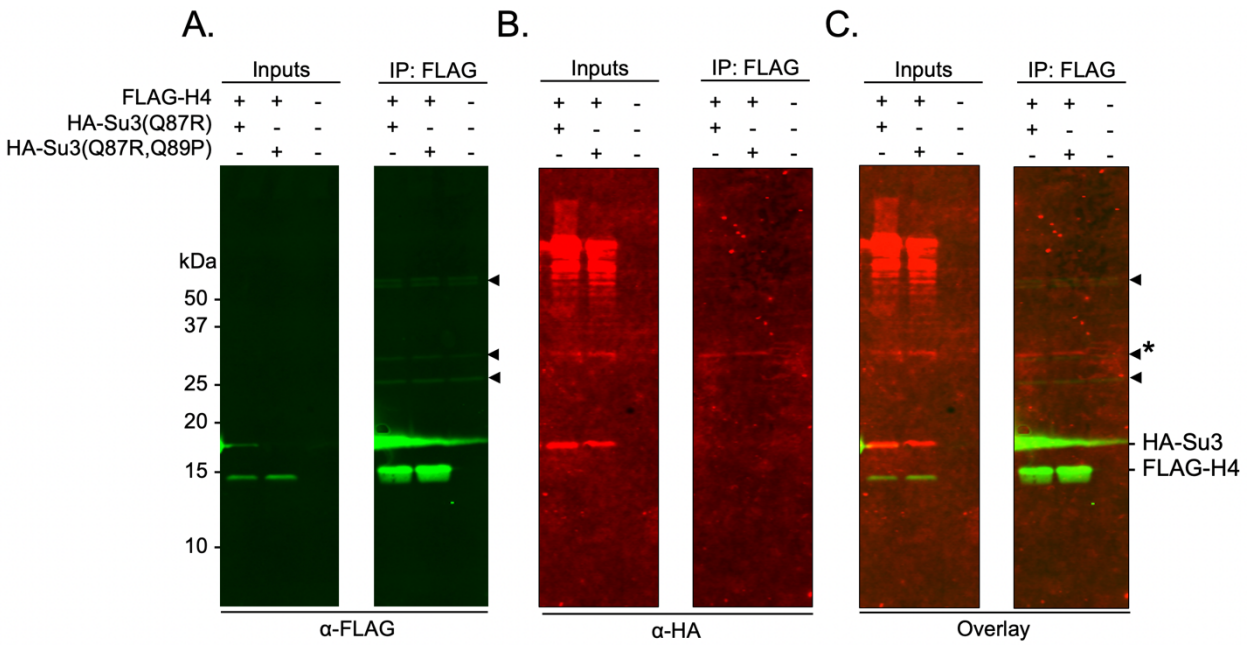


Figure 4.2 Immunoprecipitation from HEK293T cells suggests H4 sumoylation. HEK293T cells were transiently transfected with plasmids encoding the indicated epitope-tagged proteins. After 48 h, lysates were prepared by TCA method. Lysates were used in immunoprecipitation against FLAG, resolved by SDS-PAGE, and were immunoblotted against FLAG (A) and HA (B) on the same blot (C). Asterisk indicates expected migration size of sumoylated H4 with tags. Arrow heads indicate positions of immunoglobulin heavy and light chains leached from resin.

antibody used was not sensitive enough to pick up the FLAG-tag at the same levels as the HA antibody. Attempts to increase the amount of FLAG-H4 expression by increasing the amount of transfected plasmid did not produce increased amounts of FLAG-H4 expression under the conditions used (**Figure 4.S3**). These sets of experiments suggested that H4 is sumoylated in HEK293T cells, but additional experiments are needed to further corroborate this observation.

A similar set of experiments was conducted using FLAG-H2B, which expressed strongly in HEK293T cells after transient transfection with either HA-SUMO3(Q87R) or HA-SUMO3(Q87R,Q89P) (**Figure 4.3A input and 4.S4**). After FLAG immunoprecipitation, FLAG-H2B was strongly enriched, including the observation of several strong higher molecular weight bands. A few HA-reactive bands, but not bands corresponding to unconjugated HA-SUMO, specifically appeared in the FLAG immunoprecipitates, indicating HA-SUMO conjugation to

FLAG-H2B (**Figure 4.3B**). Gratifyingly, these same HA-reactive bands overlayed with a set of higher MW FLAG-H2B bands, strongly suggesting the modification of FLAG-H2B (**Figure 4.3C**). This data represents the first unambiguous data demonstrating H2B sumoylation by SUMO3 in human cells. Bottom-up MS coupled with site-directed mutagenesis of Lys residues to Arg would identify and validate the site of sumoylation and are being undertaken in the Chatterjee labs.

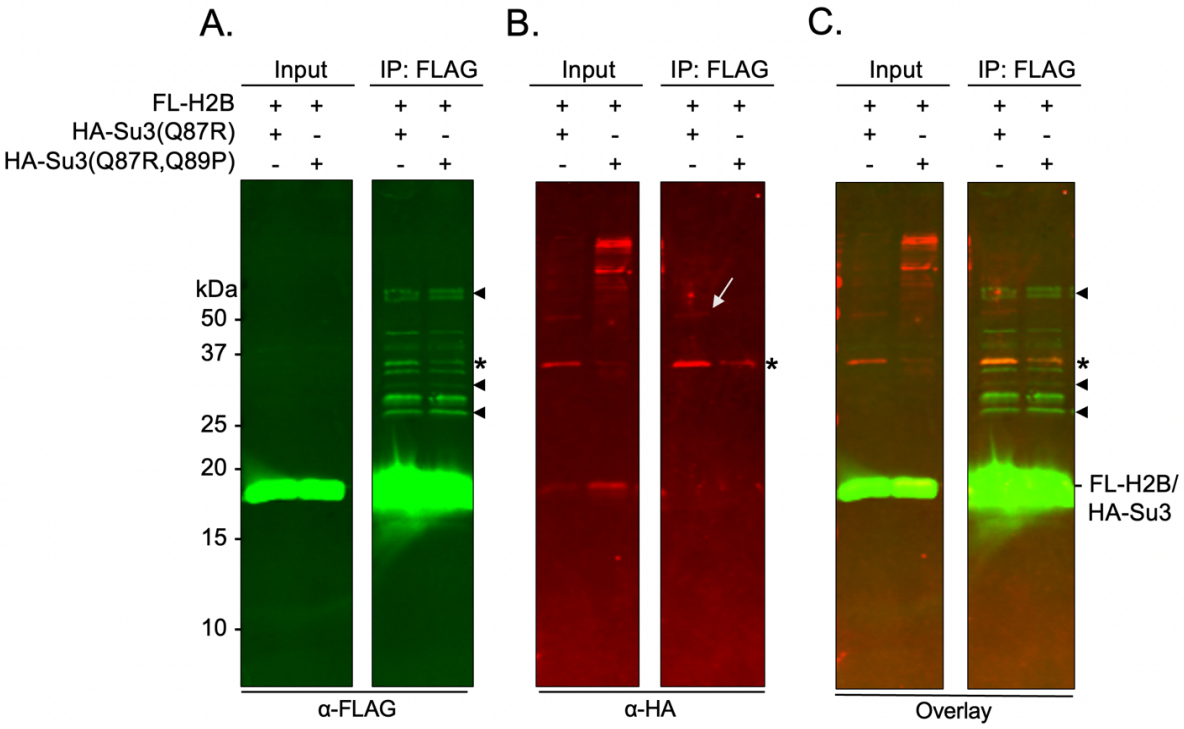


Figure 4.3. H2B is sumoylated in HEK293T cells. HEK293T cells were transiently transfected with plasmids encoding the indicated epitope-tagged proteins. After 48 h, lysates were prepared by TCA method. Lysates were used in immunoprecipitation against FLAG, resolved by SDS-PAGE, and were immunoblotted against FLAG (A) and HA (B) on the same blot (C). Asterisk indicates sumoylated H2B with tags. White arrow indicates a possible H2Bsu modified by another large protein modifier such as ubiquitin or SUMO. Arrow heads indicate positions of immunoglobulin heavy and light chains leached from resin.

Interestingly, despite its resistance to proteolysis, the Q87R/Q89P SUMO mutant resulted in less H2B sumoylation than the Q87R mutant SUMO at this site. An additional higher MW band is also observable in the HA blot that is just under 50 kDa in apparent MW, pointing to the possibility of a second SUMO or endogenous ubiquitin modification on H2B. No FLAG-reactive band overlays with this 50 kDa band, although this suggests and would be the first observation of either a multi-

mono, poly-SUMO, or co-SUMO and ubiquitin modified H2B in human cells. Both poly-sumoylation and potentially co-modification by SUMO and ubiquitin of H2B would be consistent with biochemical observations on the function of yeast H2B poly-sumoylation pathway to turn on transcription.[20] Further characterization of this band by immunoblotting with a ubiquitin-specific antibody and mass spectrometry will be necessary to confirm the identity of this higher molecular weight species.

4.2.2 Histones H4 and H2B are sumoylated in HeLa cells

To further validate H4 and H2B sumoylation in human cells, the same set of experiments were conducted in another human cell line. All four plasmids expressing FLAG-H2B, FLAG-H4, HA-SUMO3(Q87R) or HA-SUMO3(Q87R,Q89P) were transiently transfected into HeLa cells. Lysates were prepared by TCA precipitation and used in a FLAG immunoprecipitation assay (Figure 4.4 and 4.S5). Expression of FLAG-H2B and FLAG-H4 was clearly observed in the inputs and were enriched in the IP (Figure 4.4). Both HA-SUMO3s also expressed well in HeLa cells. FLAG immunoprecipitates clearly showed co-IP of HA-tagged proteins from cells transfected with either FLAG-H4 or FLAG-H2B (Figure 4.4B). Both HA-SUMO3(Q87R) and HA-SUMO3(Q87R,Q89P) modified FLAG-H2B, but only HA-SUMO3(Q87R) clearly modified FLAG-H4 (Figure 4.4C). Together with the transfections in HEK293T cells, these data strongly demonstrate that histones H4 and H2B are substrates for sumoylation in human cells.

4.2.3 Use of a 3xFLAG-tag on H4 improves visualization of H4su.

One caveat of the set of experiments conducted so far with FLAG-H4 is that it is relatively weakly ectopically expressed in human cells and sumoylation on H4 occurs at a significantly low proportion relative to total FLAG-H4.[18] To attempt to increase the signal obtained in immunoblots and to potentially increase the amount of H4su immunoprecipitated for MS analysis, a 3xFLAG-

tag was introduced in place of the single FLAG-tag on H4. Due to the increased length of the FLAG epitope tag, it may interact more strongly with anti-FLAG antibodies to increase yields in purification and to lower the limit of detection of moderately expressed recombinant FLAG-tagged proteins.

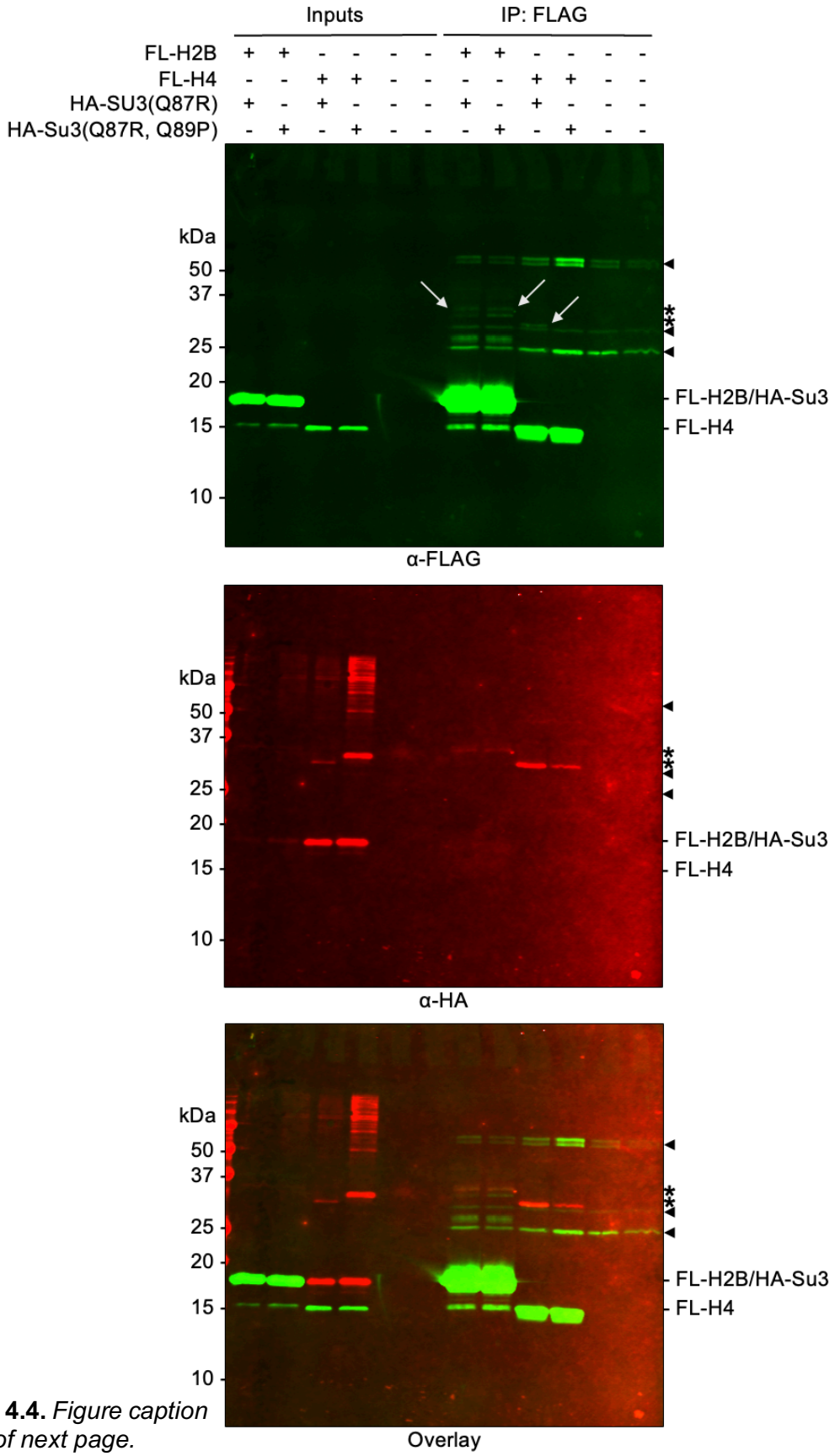


Figure 4.4. *Figure caption at top of next page.*

Figure 4.4 H2B and H4 are sumoylated in HeLa cells. HeLa and HEK293T cells were transiently transfected with plasmids encoding the indicated proteins. Lysates prepared by TCA method were immunoprecipitated against FLAG, resolved by SDS-PAGE, and immunoblotted against FLAG and HA on the same blot. Asterisk indicates sumoylated histones with tags. Arrow heads indicate positions of immunoglobulin heavy and light chains leached from resin. White arrows indicate FLAG reactive H2B or H4 bands that correspond with their sumoylation.

Transfection of HEK293T cells with a plasmid containing 3xFLAG-H4 with HA-SUMO(Q87R) showed a clear band corresponding to the molecular weight (MW) of a sumoylated H4 (**Figure 4.S6**), but not with 3xFLAG-H4 transfection alone. No higher MW bands were previously observed in similarly transfected lysates for FLAG-H4. Additionally, the signal intensity from 3xFLAG-H4-HA-SUMO3 is noticeably greater than previous transfections with FLAG-H4, demonstrating the increased signal obtained by the 3xFLAG-tag. Immunoprecipitation (IP) of 3xFLAG-H4 from TCA-prepared lysates specifically pulled out this higher MW band that is reactive with both FLAG and HA in immunoblots (**Figure 4.5 and S8**). This data strongly supports histone H4 sumoylation in human cells by the direct detection of the FLAG- and HA- epitope tags in the same overlapping protein band after IP. In addition to scaling up this system for MS analysis, the introduction of the 3xFLAG tag at the N-terminus of H2B and experiments with this construct are currently being conducted in our laboratory by a second year graduate student Madeline Currie.

4.3 Conclusion and outlook

Histone sumoylation is an emerging key histone modification that is implicated in regulating DNA-damage and gene transcription-related processes. Despite histone sumoylation being first discovered in human cells through biochemical pulldown and immunoblot experiments, most studies identifying histone sumoylation in human cells have been unbiased mass spectrometry-based proteomics efforts. These studies have been integral for establishing the mark and identifying the sites of sumoylation on all its protein substrates, but they lacked detailed characterization of the effect of sumoylation on chromatin function. Studies in yeast however have

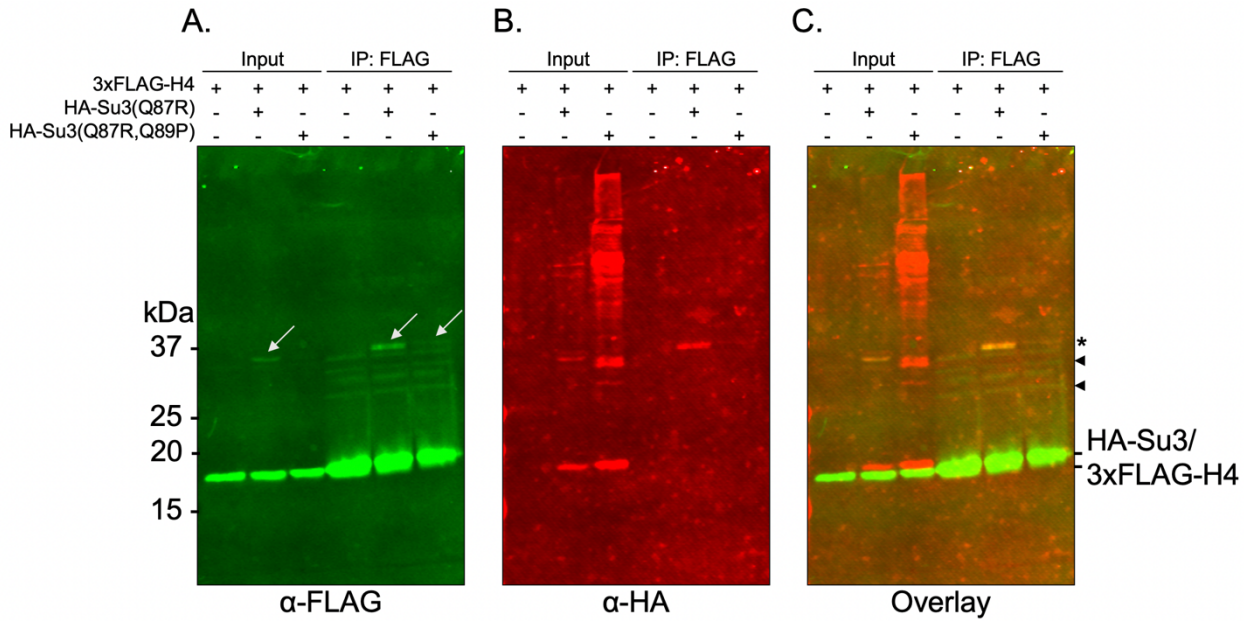


Figure 4.5 Immunoprecipitation using 3xFLAG-H4 clearly demonstrates H4 sumoylation in HEK293T cells. HEK293T cells were transiently transfected with plasmids encoding the indicated proteins. After 48 h post-transfection, cells were collected, and lysates were prepared by TCA method. Lysates immunoprecipitated against FLAG, were resolved by SDS-PAGE, and immunoblotted against FLAG and HA. Asterisk indicates sumoylated H4 with tags. Arrow heads indicate positions of immunoglobulin chains leached from resin. White arrows indicate FLAG reactive H4 bands that correspond with their sumoylation.

benefitted from a combination of its generally greater genetic tractability, and from there being only a single yeast SUMO isoform, Smt3. Indeed, SUMO1 and SUMO2/3 often have different substrate scopes in human cells, and therefore are functionally non-redundant in many instances. Histone sumoylation in yeast also appears to occur to a greater extent than in human cells, based on recent reports, allowing for biochemical interrogations into its functional roles. Histone H2B poly-sumoylation has been shown to regulate transcription in yeast, and H4 poly-sumoylation was also observed.[20,44] So far only three studies have biochemically investigated histone sumoylation, the first being the original report identifying H4 sumoylation by SUMO2/3 and the other two being studies on SUMO1 modification.[6,12,16] Two of these studies also reported sumoylation occurring at trace levels on other core histones, but did not include supporting evidence.[6,16] In fact, no site analysis by mutation has been done on any of the human core histones in human cells. There is therefore a significant gap in our understanding of the mechanistic roles for histone sumoylation

in regulating transcription in human cells, and to test if some of the mechanisms proposed in yeast are also conserved in humans and higher eukaryotes.

To understand the functions for sumoylated histones in cells, I sought to first biochemically validate histone sumoylation of H4 and H2B in human cell lines. I also designed my histone and SUMO protein constructs to be amenable to proteomic analysis by MS in order to identify the specific sites of histone sumoylation. This information will be used to inform which histone lysine residues to mutate and test for downstream functional effects in cells. Through transient transfections of epitope-tagged H4, H2B, and SUMO3, I have now demonstrated that H4 and H2B are sumoylated in both HEK293T and HeLa cell lines, using immunoprecipitation with epitope-specific antibodies,. Importantly, histone H2B sumoylation is a previously uncharacterized histone modification in human cells. Additionally, I have observed for the first time either a multi-mono, di-sumo chain, or co-modification by sumo and ubiquitin on H2B in human cells, which will be further investigated in the Chatterjee lab. Interestingly, in a previous study using a linear fusion of HA-SUMO3 to the N-terminus of H4 transfected into HEK293T cells, I also observed an additional higher MW band that lies above where the SUMO3-H4 fusion protein runs (**Figure 3.4**). This piece of data also suggests either a multi-mono, di-sumo chain, or co-modification by sumo and ubiquitin on H4 in humans. Further biochemical investigations and mass spectrometry will be needed to identify and validate these surprising observations of a higher band at a molecular weight greater than that of SUMO-H4 alone.

Now that the expression system for generating histone sumoylation is well-established in human cells, these transfections can be scaled up to generate quantities of sumoylated histones to be analyzed by mass spectrometry for site-identification. The identification of specific sites by mass spectrometry and their comparison with published SUMO proteomics studies will help inform site-directed mutagenesis of histones for further studies in cells. Additionally, these site-IDs and

expression systems can then be used in cellular studies using various stresses, such as DNA-damage, heat, and oxidation, to test if chromatin sumoylation levels are indeed modulated by stress as was proposed in some studies.[10,24,46] What is well understood thus far is that histone sumoylation appears to be generally repressive towards transcription through its anti-correlation with signals of active transcription, and its association with transcriptionally repressive chromatin complexes. There is still much to be uncovered in the realm of histone sumoylation and in ascertaining what roles it may play in different nuclear processes like DNA-damage (with respect to SUMO1 modification of H2A.X) and transcriptional control, as well as identifying the signals that lead to its deposition in and removal from chromatin.

4.4 Experimental procedures

4.4.1 General methods

DNA synthesis and gene sequencing were performed by Integrated DNA Technologies (Coralville, IA) and Eurofins Genomics (Louisville, KY), respectively. Plasmid miniprep, PCR purification, and gel extraction kits were purchased from Qiagen (Valencia, CA). Western blots were performed with PVDF membranes and visualized using IR-Dye secondary antibodies (LI-COR Biosciences, Lincoln, NE) scanned on an Odyssey infrared imaging system (LI-COR Biosciences). HEK293T and HeLa cells were purchased from ATCC and cultured in T75 flasks or 6-well plates using Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and incubated at 37 °C in a humidified, 5% CO_2 atmosphere.

4.4.2 Molecular cloning of pcDNA3 plasmids for transient transfections

The plasmid pcDNA3-HA-SUMO3 containing the human SUMO3 (sentrin 2) gene was purchased from Addgene (plasmid #17361). The modified pcDNA3 plasmid containing SUMO3 with a Q87R mutation or the additional Q89P mutation was prepared by site-directed mutagenesis (Q5 Site-

directed mutagenesis kit, NEB, Ipswitch, MA) using the original plasmid as the template in PCR using the primers below (**Table 4.2**). Both PCR's utilized the same reverse primer. The plasmid pcDNA3-FLAG-H4 containing the human H4 gene was generated by restriction digest cloning from a pME18s-FLAG-H4 plasmid. First, a HindIII restriction site was introduced in order to cut out the FLAG-H4 encoding sequence using HindIII and XbaI. The same enzymes were used to cut the pcDNA3-FLAG-H2B plasmid (a gift from Professor Moshe Oren at Weizmann Institute of Science) as the vector source for the ligation of the FLAG-H4 insert by T4 DNA ligase (NEB). Products were used to transform *E. coli* XL10Gold ultracompetent cells (Agilent Technologies, Santa Clara, CA). The pcDNA3-3xFLAG-H4 was generated by PCR using the primers below and the pcDNA3-FLAG-H4 plasmid as the template. The desired sequences were confirmed by agarose gel electrophoresis and DNA sequencing.

Table 4.2. List of ssDNA primers used for molecular cloning.

Primer	DNA Sequence (5'- to -3')
Q87R_GG_F	CGA CGT GTT CCG GCA GCA GAC GG
Q87R_GG_R	ATG GTG TCC TCG TCC TCC ATC TCC
Q87R_Q89P_F	CGA CGT GTT CCG GCA GCC GAC GG
FP-HindIII-FLAGH4	CTT ATG GAC TAC AAG GAC GAC GAC GAC
RP-HindIII-FLAGH4	CTT GGC AAT TCC GCA GCT TTT AGA G
3xFLAG-H4-Fwd	GAT TAT AAA GAT CAT GAC ATC GAC TAC AAG GAC GAC GAC
3xFLAG-H4-Rev	ACC GTC ATG GTC TTT GTA GTC CAT AAG CTT GGG TCT CCC

4.4.3 Preparation of pcDNA for transfection

DNA for transient transfection was prepared by Miraprep of *E. coli* DH5α cells as previously described using a Qiagen (Valencia, CA) DNA miniprep kit.[45] Briefly, transformed *E. coli* DH5α cells were grown in 50 mL LB media supplemented with ampicillin (50 μg/mL) overnight at 37 °C. Cells were collected by centrifugation and resuspended in P1 buffer supplemented with fresh RNase. After alkaline lysis and neutralization, the supernatant was cleared by centrifugation. The supernatant was diluted with an equal volume of 96% (v/v) ethanol prior to loading onto five

Qiagen miniprep spin columns. At this point, the DNA was washed and eluted according to the Qiagen protocol. Purity of the eluted DNA was checked by measuring the $A_{280\ nm}/A_{260\ nm}$ ratio on a NanoDrop 2000c spectrophotometer and by agarose gel electrophoresis. The correct gene sequence for p300 was also confirmed by sequencing prior to transfection in human cells.

4.4.4 Transient transfection of human cells

HEK293T or HeLa cells were cultured to ~60% confluency before transient transfection with pcDNA3.1-p300-FLAG. The cell growth medium was changed at least 1 h before transfection. Cells were transfected using Lipofectamine 3000 (Invitrogen) at 2.5 µg DNA per well in a 6-well format or 15 µg in a T75 flask, and the media was changed 24 h later. Cells were grown in transfection medium for an additional 24 hours before detachment using trypsin and collection by centrifugation. The cell pellet was washed thrice with ice-cold Dulbecco's Phosphate-buffered Saline before lysate preparation using trichloroacetic acid or storage at -80 °C after flash freezing in liquid nitrogen (LN2).

4.4.5 FLAG immunoprecipitation from TCA-prepared lysates

Lysates prepared using trichloroacetic acid (TCA) and immunoprecipitation using anti-DYKDDDK G1 affinity resin (Genscript, Piscataway, NJ) were carried out as previously described.[18,47] Two wells of confluent cells were combined from a 6-well plate and lysed in 500 µL of cold 20 % TCA, vortexed well, and kept on ice. The suspension was sonicated 4x 1 s using a microtip sonicator and then diluted with 900 µL of 5 % TCA. The precipitate was collected by centrifugation at 6,000 RCF for 10 min. The pellet was resuspended in 200 µL of 1x Laemmli without reducing agent or dye and neutralized with 50 µL of 2 M tris (unbuffered). Samples were boiled for 3 min, vortexed, and clarified by centrifugation at 6,000 RCF. For immunoprecipitations (IPs), ~ 200 µL of the clarified lysate was diluted with 800 µL IP buffer (50 mM tris, pH 7.4 at 4 °C, 150 mM NaCl, 0.5 % IGEPAL-CA630). The diluted lysate was applied to pre-equilibrated FLAG resin in IP buffer and

nutated for 2 h at 4 °C. The resin was washed 4x with IP buffer and proteins were eluted in 1x Laemmli (with dye but no reducing agent) and boiled for 5 min. Eluted samples were mixed with DTT before running SDS-PAGE and immunoblotting.

4.4.6 Immunoblotting (Western blot)

Western blots were performed in modified Towbin buffer consisting of 25 mM Tris, 192 mM glycine, 4 mg/L SDS, 10% (v/v) methanol, and proteins were transferred onto PVDF membrane at 35 V for 2 h, on ice. Membranes were blocked in 5% (w/v) non-fat milk powder in phosphate-buffered saline (PBS) for 1 h at 25 ^{o}C before incubating overnight in diluted primary antibody (**Table 4.3**) in 5% (w/v) non-fat milk powder in PBST (PBS containing 0.05% v/v Tween-20) at 4 °C. Overnight incubated membranes were washed in PBST before incubating with IR-dye conjugated secondary antibody or HRP-conjugated secondary antibody in 5% (w/v) non-fat milk powder containing PBST for 1 h at 25^{o}C. After incubation with secondary antibodies, the membranes were washed first in PBST and then PBS before scanning on a Li-COR Biosciences (Lincoln, NE) Odyssey IR scanner.

Table 4.3. List of antibodies used in this study.

Antibodies	Source	Identifier
Rabbit monoclonal anti-HA	Cell Signaling Technology	3724
Mouse monoclonal anti-FLAG	Sigma-Aldrich	F1804
IRDye 680RD Goat anti-Rabbit IgG	Li-COR Biosciences	926-68071
IRDye 800CW Goat anti-Mouse IgG	Li-COR Biosciences	926-32210

4.5 Product characterization and supplemental data

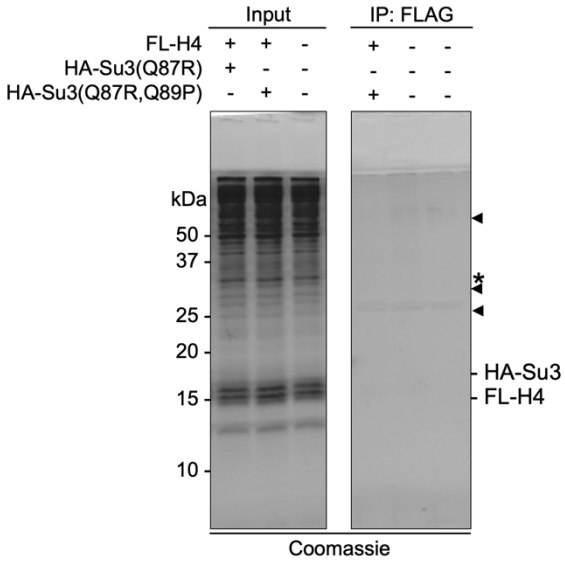


Figure 4.S1 Immunoprecipitation from HEK293T cells suggests H4 sumoylation. Corresponding coomassie-stained gel from **Figure 4.2** of the input lysates and immunoprecipitates. Asterisk indicates expected migration size of sumoylated H4 with tags. Arrow heads indicate positions of immunoglobulin heavy and light chains released from resin.

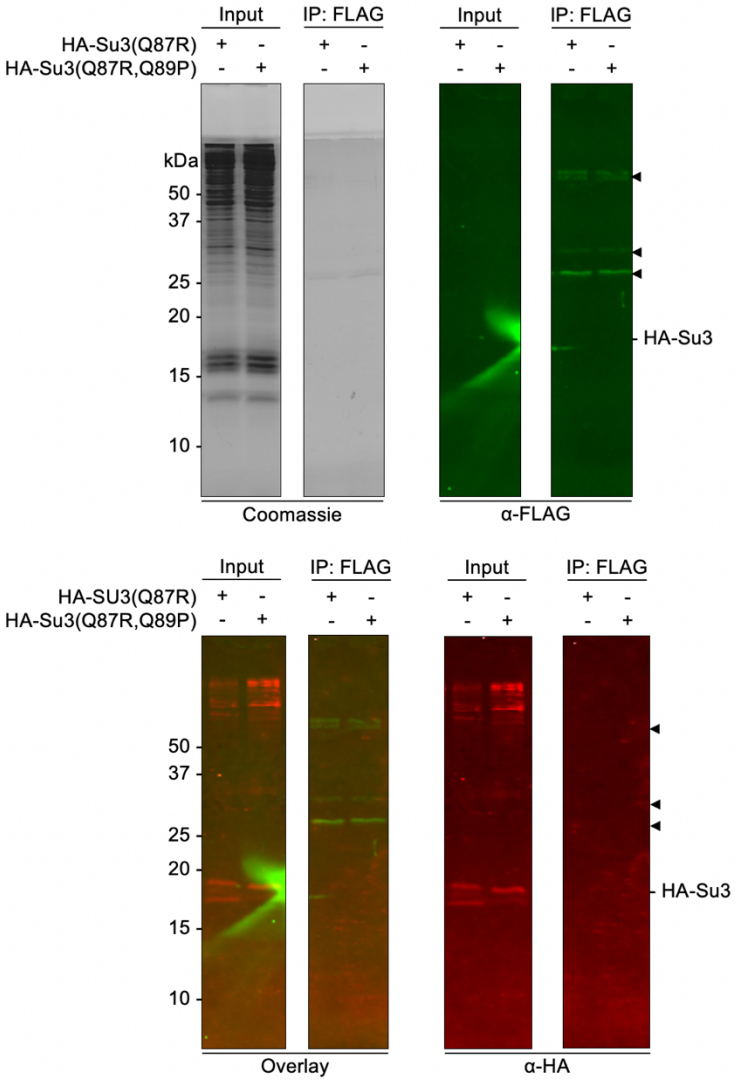


Figure 4.S2. HA-tagged SUMO does not cross react or bind non-specifically with FLAG antibody in immunoprecipitation or immunoblot. Transient transfection with plasmid containing HA-SUMO3(Q87R) or HA-SUMO3(Q87R,Q89P) in HEK293T cells. Cellular extracts were prepared by TCA method and immunoprecipitated with FLAG resin and blotted against HA and FLAG. Arrow heads indicate positions of immunoglobulin heavy and light chains released from resin.

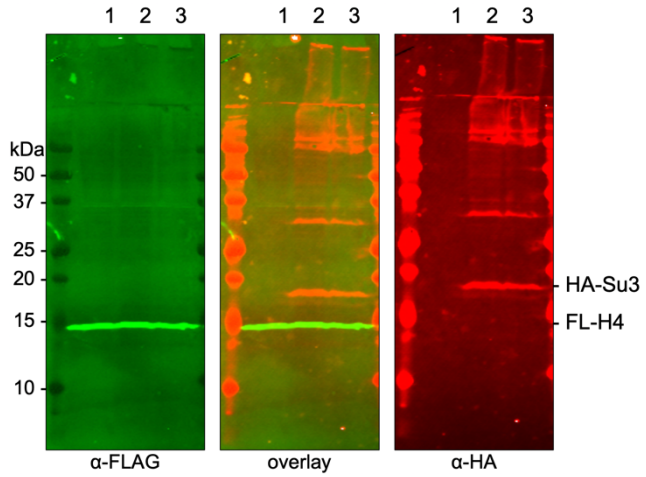


Figure 4.S3. Increasing amount of FLAG-H4 plasmid in transfection does not significantly increase its expression. Transient transfection with plasmid containing FLAG-H4 at 2.5 µg alone, lane 1; with HA-SUMO3(Q87R), lane 2; and at 5 µg with HA-SUMO3(Q87R), lane 3, in HEK293T cells. Cellular extracts were prepared by TCA method and immunoblotted against HA and FLAG.

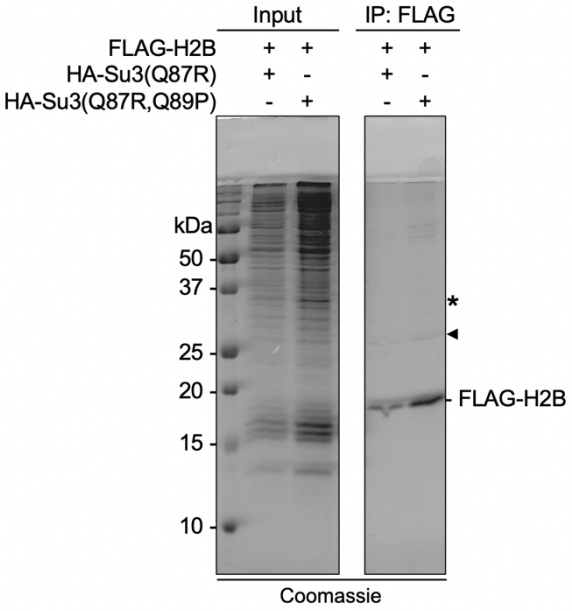


Figure 4.S4. H2B is sumoylated in HEK293T cells. Corresponding coomassie-stained gel from **Figure 4.3** of the input lysates and immunoprecipitates. Asterisk indicates expected migration size of sumoylated H4 with tags. Asterisk indicates expected migration size of sumoylated H2B with tags. Arrow heads indicate positions of immunoglobulin heavy and light chains released from resin.

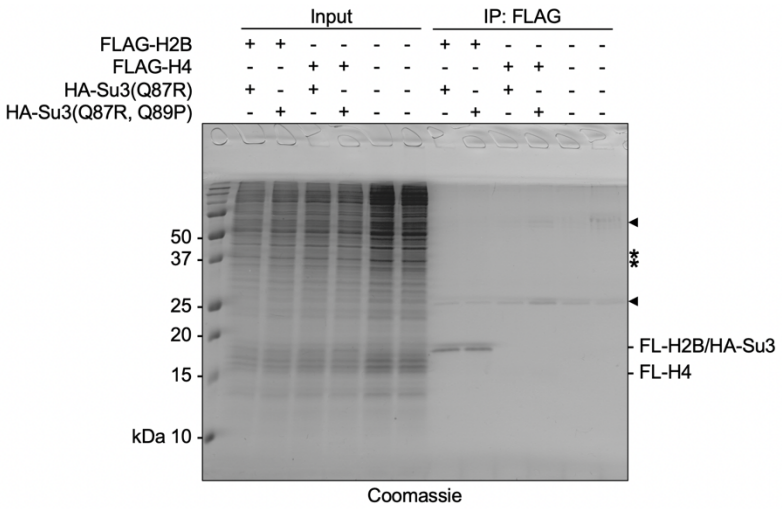


Figure 4.S5. H2B and H4 are sumoylated in HeLa cells. Corresponding coomassie-stained gel from **Figure 4.4** of the input lysates and immunoprecipitates. Asterisk indicates expected migration size of sumoylated histones with tags. Arrow heads indicate positions of immunoglobulin heavy and light chains released from resin.

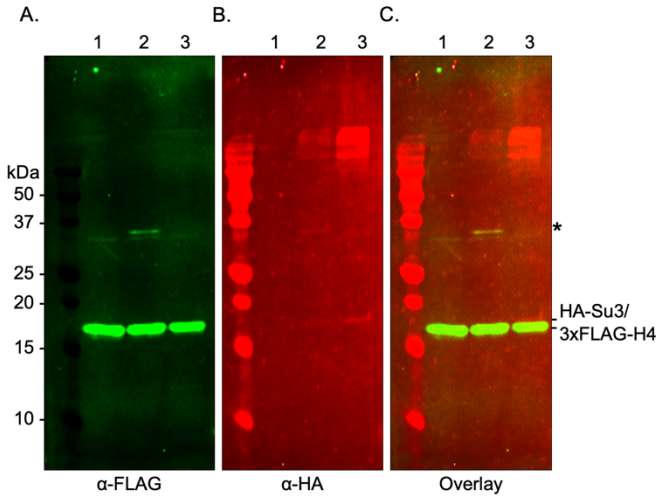


Figure 4.S6. 3xFLAG-H4 transfected with HA-Su3(Q87R) in HEK293T cells shows a clear FLAG band corresponding to a sumoylated H4. TCA lysates were prepared from transiently transfected HEK293T cells with 3xFLAG-H4 alone (lane 1) or with either HA-Su3(Q87R) (lane 2) or HA-Su3(Q87R,Q89P) (lane 3). Lysates were immunoblotted against FLAG and HA after SDS-PAGE. Asterisk indicates sumoylated H4 band.

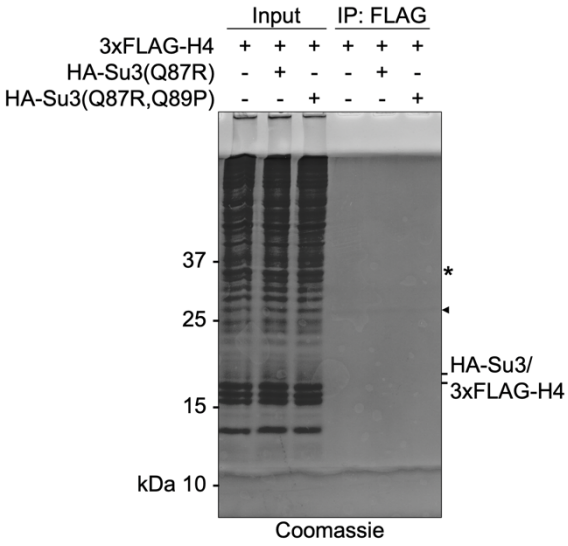


Figure 4.S7 Immunoprecipitation using 3xFLAG-H4 clearly demonstrates H4 sumoylation in HEK293T cells. Corresponding coomassie-stained gel from **Figure 4.5** of the input lysates and immunoprecipitates. Asterisk indicates expected migration size of sumoylated H4 with tags. Arrow heads indicate positions of immunoglobulin light chain released from resin.

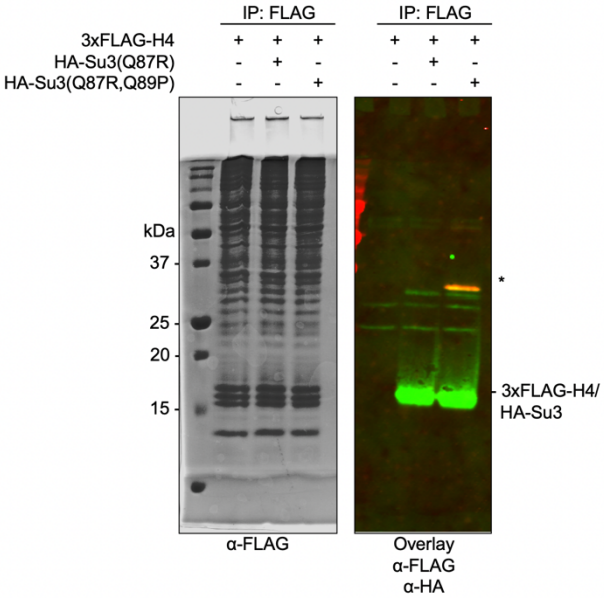


Figure 4.S8 Immunoprecipitation using 3xFLAG-H4 clearly demonstrates H4 sumoylation in HEK293T cells. Replicate FLAG IP with clear overlap of a band reactive to both FLAG and HA at the appropriate MW for a sumoylated H4.

4.6 References

(1) Cappadocia, L.; Lima, C. D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chem. Rev.* **2017**, *118* (3), 889–918. https://doi.org/10.1021/ACS.CHEMREV.6B00737.

(2) Yang, S.-H.; Galanis, A.; Witty, J.; Sharrocks, A. D. An Extended Consensus Motif Enhances the Specificity of Substrate Modification by SUMO. *EMBO J.* **2006**, *25* (21), 5083. https://doi.org/10.1038/SJ.EMBOJ.7601383.

(3) Hendriks, I. A.; Vertegaal, A. C. O. A Comprehensive Compilation of SUMO Proteomics. *Nat. Rev. Mol. Cell Biol. 2016 179* **2016**, *17* (9), 581–595. https://doi.org/10.1038/nrm.2016.81.

(4) Hickey, C. M.; Wilson, N. R.; Hochstrasser, M. Function and Regulation of SUMO Proteases. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (12), 755. https://doi.org/10.1038/NRM3478.

(5) Wei, L.; Zhao, X. Roles of SUMO in Replication Initiation, Progression, and Termination. *Adv. Exp. Med. Biol.* **2017**, *1042*, 371. https://doi.org/10.1007/978-981-10-6955-0_17.

(6) Shiio, Y.; Eisenman, R. N. Histone Sumoylation Is Associated with Transcriptional Repression. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (23), 13225–13230. https://doi.org/10.1073/pnas.1735528100.

(7) Jackson, S. P.; Durocher, D. Molecular Cell Review Regulation of DNA Damage Responses by Ubiquitin and SUMO. **2013**. https://doi.org/10.1016/j.molcel.2013.01.017.

(8) Eifler, K.; Cuijpers, S. A. G.; Willemstein, E.; Raaijmakers, J. A.; El Atmioui, D.; Ovaa, H.; Medema, R. H.; Vertegaal, A. C. O. SUMO Targets the APC/C to Regulate Transition from Metaphase to Anaphase. *Nat. Commun. 2018 91* **2018**, *9* (1), 1–15. https://doi.org/10.1038/s41467-018-03486-4.

(9) Rosonina, E. A Conserved Role for Transcription Factor Sumoylation in Binding-Site Selection. *Curr. Genet.* **2019**, *65*, 1307–1312. https://doi.org/10.1007/s00294-019-00992-w.

(10) Hendriks, I. A.; Lyon, D.; Su, D.; Skotte, N. H.; Daniel, J. A.; Jensen, L. J.; Nielsen, M. L. Site-Specific Characterization of Endogenous SUMOylation across Species and Organs. *Nat. Commun. 2018 91* **2018**, *9* (1), 1–17. https://doi.org/10.1038/s41467-018-04957-4.

(11) Matafora, V.; D'Amato, A.; Mori, S.; Blasi, F.; Bachi, A. Proteomics Analysis of Nucleolar SUMO-1 Target Proteins upon Proteasome Inhibition. *Mol. Cell. Proteomics* **2009**, *8* (10),

2243. https://doi.org/10.1074/MCP.M900079-MCP200.

(12) Chen, Z.; Zhang, Y.; Guan, Q.; Zhang, H.; Luo, J.; Li, J.; Wei, W.; Xu, X.; Liao, L.; Wong, J.; Li, J. Linking Nuclear Matrix-Localized PIAS1 to Chromatin SUMOylation via Direct Binding of Histones H3 and H2A.Z. *J. Biol. Chem.* **2021**, *0* (0), 101200. https://doi.org/10.1016/J.JBC.2021.101200.

(13) Liebelt, F.; Jansen, N. S.; Kumar, S.; Gracheva, E.; Claessens, L. A.; Verlaan-de Vries, M.; Willemstein, E.; Vertegaal, A. C. O. The Poly-SUMO2/3 Protease SENP6 Enables Assembly of the Constitutive Centromere-Associated Network by Group DeSUMOylation. *Nat. Commun. 2019 101* **2019**, *10* (1), 1–18. https://doi.org/10.1038/s41467-019-11773-x.

(14) Lumpkin, R. J.; Gu, H.; Zhu, Y.; Leonard, M.; Ahmad, A. S.; Clauser, K. R.; Meyer, J. G.; Bennett, E. J.; Komives, E. A. Site-Specific Identification and Quantitation of Endogenous SUMO Modifications under Native Conditions. *Nat. Commun. 2017 81* **2017**, *8* (1), 1–11. https://doi.org/10.1038/s41467-017-01271-3.

(15) Hendriks, I. A.; D'Souza, R. C. J.; Yang, B.; Verlaan-De Vries, M.; Mann, M.; Vertegaal, A. C. O. Uncovering Global SUMOylation Signaling Networks in a Site-Specific Manner. *Nat. Struct. Mol. Biol.* **2014**, *21* (10), 927–936. https://doi.org/10.1038/nsmb.2890.

(16) Chen, W.-T.; Alpert, A.; Leiter, C.; Gong, F.; Jackson, S. P.; Miller, K. M. Systematic Identification of Functional Residues in Mammalian Histone H2AX. *Mol. Cell. Biol.* **2013**, *33* (1), 111. https://doi.org/10.1128/MCB.01024-12.

(17) Galisson, F.; Mahrouche, L.; Courcelles, M.; Bonneil, E.; Meloche, S.; Chelbi-Alix, M. K.; Thibault, P. A Novel Proteomics Approach to Identify SUMOylated Proteins and Their Modification Sites in Human Cells. *Mol. Cell. Proteomics* **2011**, *10* (2), M110.004796. https://doi.org/10.1074/mcp.M110.004796.

(18) 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. Histone Sumoylation Is a Negative Regulator in Saccharomyces Cervisiae and Shows Dynamic Interplay with Positive-Acting Histone Modifications. *Genes Dev.* **2006**, *20*, 966–976. https://doi.org/10.1101/gad.1404206.6.

(19) Kalocsay, M.; Hiller, N. J.; Jentsch, S. Chromosome-Wide Rad51 Spreading and SUMO-H2A.Z-Dependent Chromosome Fixation in Response to a Persistent DNA Double-Strand Break. *Mol. Cell* **2009**, *33* (3), 335–343. https://doi.org/10.1016/J.MOLCEL.2009.01.016.

(20) Ryu, H.-Y.; Su, D.; Wilson-Eisele, N. R.; Zhao, D.; López-Giráldez, F.; Hochstrasser, M. The Ulp2 SUMO Protease Promotes Transcription Elongation through Regulation of Histone Sumoylation. *EMBO J.* **2019**, *38* (16), e102003. https://doi.org/10.15252/EMBJ.2019102003.

(21) Ohkuni, K.; Takahashi, Y.; Fulp, A.; Lawrimore, J.; Au, W.-C.; Pasupala, N.; Levy-Myers, R.; Warren, J.; Strunnikov, A.; Baker, R. E.; Kerscher, O.; Bloom, K.; Basrai, M. A. SUMO-Targeted Ubiquitin Ligase (STUbL) Slx5 Regulates Proteolysis of Centromeric Histone H3 Variant Cse4 and Prevents Its Mislocalization to Euchromatin. *Mol. Biol. Cell* **2016**, *27* (9), 1500–1510. https://doi.org/10.1091/MBC.E15-12-0827.

(22) Ohkuni, K.; Levy-Myers, R.; Warren, J.; Au, W.-C.; Takahashi, Y.; Baker, R. E.; Basrai, M. A. N-Terminal Sumoylation of Centromeric Histone H3 Variant Cse4 Regulates Its Proteolysis To Prevent Mislocalization to Non-Centromeric Chromatin. *G3 Genes|Genomes|Genetics* **2018**, *8* (4), 1215. https://doi.org/10.1534/G3.117.300419.

(23) Ohkuni, K.; Suva, E.; Au, W.-C.; Walker, R. L.; Levy-Myers, R.; Meltzer, P. S.; Baker, R. E.; Basrai, M. A. Deposition of Centromeric Histone H3 Variant CENP-A/Cse4 into Chromatin Is Facilitated by Its C-Terminal Sumoylation. *Genetics* **2020**, *214* (4), 839–854. https://doi.org/10.1534/GENETICS.120.303090.

(24) Drabikowski, K.; Ferralli, J.; Kistowski, M.; Oledzki, J.; Dadlez, M.; Chiquet-Ehrismann, R. Comprehensive List of SUMO Targets in Caenorhabditis Elegans and Its Implication for Evolutionary Conservation of SUMO Signaling. *Sci. Reports 2018 81* **2018**, *8* (1), 1–12. https://doi.org/10.1038/s41598-018-19424-9.

(25) Subramonian, D.; Raghunayakula, S.; Olsen, J. V.; Beningo, K. A.; Paschen, W.; Zhang, X.-D. Analysis of Changes in SUMO-2/3 Modification during Breast Cancer Progression and Metastasis. *J. Proteome Res.* **2014**, *13* (9), 3905–3918. https://doi.org/10.1021/PR500119A.

(26) Issar, N.; Roux, E.; Mattei, D.; Scherf, A. Identification of a Novel Post-Translational Modification in Plasmodium Falciparum: Protein Sumoylation in Different Cellular Compartments. *Cell. Microbiol.* **2008**, *10* (10), 1999. https://doi.org/10.1111/J.1462-5822.2008.01183.X.

(27) Miller, M. J.; Barrett-Wilt, G. A.; Hua, Z.; Vierstra, R. D. Proteomic Analyses Identify a Diverse Array of Nuclear Processes Affected by Small Ubiquitin-like Modifier Conjugation in Arabidopsis. *Proc. Natl. Acad. Sci.* **2010**, *107* (38), 16512–16517. https://doi.org/10.1073/PNAS.1004181107.

(28) Dhall, A.; Wei, S.; Fierz, B.; Woodcock, C. L.; Lee, T. H.; Chatterjee, C. Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long-Range Internucleosomal Interactions. *J. Biol. Chem.* **2014**, *289* (49), 33827–33837. https://doi.org/10.1074/jbc.M114.591644.

(29) Jason, L. J. M.; Moore, S. C.; Ausió, J.; Lindsey, G. Magnesium-Dependent Association and Folding of Oligonucleosomes Reconstituted with Ubiquitinated H2A. *J. Biol. Chem.* **2001**, *276* (18), 14597–14601.

(30) Fierz, B.; Chatterjee, C.; Mcginty, R. K.; Bar-dagan, M.; Raleigh, D. P.; Muir, T. W. Histone H2B Ubiquitylation Disrupts Local and Higher Order Chromatin Compaction. *Nat. Chem. Biol.* **2011**, *7* (2), 113–119. https://doi.org/10.1038/nchembio.501.Histone.

(31) 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. Aromatic Thiol-Mediated Cleavage of N-O Bonds Enables Chemical Ubiquitylation of Folded Proteins. *Nat. Commun.* **2016**, *7*, 1–10. https://doi.org/10.1038/ncomms12979.

(32) Dhall, A.; Weller, C. E.; Chu, A.; Shelton, P. M. M.; Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-CoREST Complex. *ACS Chem. Biol.* **2017**, *12* (9), 2275–2280. https://doi.org/10.1021/acschembio.7b00716.

(33) An, W.; Roeder, R. G. Direct Association of P300 with Unmodified H3 and H4 N Termini Modulates P300-Dependent Acetylation and Transcription of Nucleosomal Templates. *J. Biol. Chem.* **2003**, *278* (3), 1504–1510. https://doi.org/10.1074/jbc.M209355200.

(34) An, W.; Palhan, V. B.; Karymov, M. A.; Leuba, S. H.; Roeder, R. G. Selective Requirements for Histone H3 and H4 N Termini in P300-Dependent Transcriptional Activation from Chromatin Woojin. *Mol. Cell* **2002**, *9*, 811–821.

(35) Kundu, T. K.; Palhan, V. B.; Wang, Z.; An, W.; Cole, P. A.; Roeder, R. G. Activator-Dependent Transcription from Chromatin in Vitro Involving Targeted Histone Acetylation by P300. *Mol. Cell* **2000**, *6* (3), 551–561. https://doi.org/10.1016/S1097-2765(00)00054-X.

(36) Rosonina, E.; Duncan, S. M.; Manley, J. L. SUMO Functions in Constitutive Transcription and during Activation of Inducible Genes in Yeast. *Genes Dev.* **2010**, *24* (12), 1242. https://doi.org/10.1101/GAD.1917910.

(37) HW, L.; J, Z.; GF, H.; M, A.; H, G. O.; R, O.-S.; K, H.; JD, P. Chromatin Modification by SUMO-1 Stimulates the Promoters of Translation Machinery Genes. *Nucleic Acids Res.* **2012**, *40* (20), 10172–10186. https://doi.org/10.1093/NAR/GKS819.

(38) A, W.; J, S.; J, D.; M, J.; A, S. The Paf1 Complex Is Essential for Histone Monoubiquitination by the Rad6-Bre1 Complex, Which Signals for Histone Methylation by COMPASS and Dot1p. *J. Biol. Chem.* **2003**, *278* (37), 34739–34742. https://doi.org/10.1074/JBC.C300269200.

(39) Chong, S. Y.; Cutler, S.; Lin, J.-J.; Tsai, C.-H.; Tsai, H.-K.; Biggins, S.; Tsukiyama, T.; Lo, Y.-C.; Kao, C.-F. H3K4 Methylation at Active Genes Mitigates Transcription-Replication Conflicts during Replication Stress. *Nat. Commun. 2020 111* **2020**, *11* (1), 1–16. https://doi.org/10.1038/s41467-020-14595-4.

(40) Steger, D. J.; Lefterova, M. I.; Ying, L.; Stonestrom, A. J.; Schupp, M.; Zhuo, D.; Vakoc, A. L.; Kim, J.-E.; Chen, J.; Lazar, M. A.; Blobel, G. A.; Vakoc, C. R. DOT1L/KMT4 Recruitment and H3K79 Methylation Are Ubiquitously Coupled with Gene Transcription in Mammalian Cells. *Mol. Cell. Biol.* **2008**, *28* (8), 2825–2839. https://doi.org/10.1128/MCB.02076-07.

(41) Hyun, K.; Jeon, J.; Park, K.; Kim, J. Writing, Erasing and Reading Histone Lysine Methylations. *Exp. Mol. Med.* **2017**, *49* (4). https://doi.org/10.1038/emm.2017.11.

(42) Altaf, M.; Utley, R. T.; Lacoste, N.; Tan, S.; Briggs, S. D.; Côté, J. Interplay of Chromatin Modifiers on a Short Basic Patch of Histone H4 Tail Defines the Boundary of Telomeric Heterochromatin. *Mol. Cell* **2007**, *28* (6), 1002. https://doi.org/10.1016/J.MOLCEL.2007.12.002.

(43) Hou, L.; Wang, Y.; Liu, Y.; Zhang, N.; Shamovsky, I.; Nudler, E.; Tian, B.; Dynlacht, B. D. Paf1C Regulates RNA Polymerase II Progression by Modulating Elongation Rate. *Proc. Natl. Acad. Sci.* **2019**, *116* (29), 14583–14592. https://doi.org/10.1073/PNAS.1904324116.

(44) Ryu, H.-Y.; Zhao, D.; Li, J.; Su, D.; Hochstrasser, M. Histone Sumoylation Promotes Set3 Histone-Deacetylase Complex-Mediated Transcriptional Regulation. *Nucleic Acids Res.* **2020**, *48* (21), 12151–12168. https://doi.org/10.1093/NAR/GKAA1093.

(45) Mukherjee, S.; Thomas, M.; Dadgar, N.; Lieberman, A. P.; Iñiguez-Lluhi, J. A. Small Ubiquitin-like Modifier (SUMO) Modification of the Androgen Receptor Attenuates Polyglutamine-Mediated Aggregation. *J. Biol. Chem.* **2009**, *284* (32), 21296–21306. https://doi.org/10.1074/jbc.M109.011494.

(46) Miller, M. J.; Barrett-Wilt, G. A.; Hua, Z.; Vierstra, R. D. Proteomic Analyses Identify a Diverse Array of Nuclear Processes Affected by Small Ubiquitin-like Modifier Conjugation in Arabidopsis. *Proc. Natl. Acad. Sci.* **2010**, *107* (38), 16512–16517. https://doi.org/10.1073/PNAS.1004181107.

(47) K, R.; J, R.; MA, O. Rad6-Dependent Ubiquitination of Histone H2B in Yeast. *Science* **2000**, *287* (5452), 501–504. https://doi.org/10.1126/SCIENCE.287.5452.501.