Investigating the Effect of Histone H4 Sumoylation on Chromatin Structure and Function

Abhinav Dhall

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Champak Chatterjee, Chair

Michael H. Gelb

Dustin J. Maly

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Investigating the Effect of Histone H4 Sumoylation on Chromatin Structure and Function

Abhinav Dhall

Chair of the Supervisory Committee:
Champak Chatterjee
Chemistry

The packaging of DNA within the nucleus of a eukaryotic organism is facilitated by the formation of a nucleoprotein complex known as chromatin. There are various DNA binding proteins that play an important part in this process, but the four core histone proteins namely, H2A, H2B, H3 and H4 are essential for forming the scaffolding for DNA packaging. The basic repeating unit of chromatin is called a nucleosome and is formed when 147 bp of DNA wrap around an octamer of histone proteins. The N-terminal regions of all four histones, known as histone tails, protrude out from the nucleosome and are accessible to various chromatin binding complexes. These complexes are capable of modifying histone tails with post-translational modifications (PTMs) that range from small chemical groups such as methyl, acetyl and phosphoryl to small proteins including ubiquitin and different isoforms of the small ubiquitin-like modifier (SUMO). Dysregulation of histone PTMs has been observed in many disease states, which makes the study of these PTMs essential for understanding a variety of pathologies. We focused on the poorly understood modification of histones by SUMO protein, because several components of the sumoylation and desumoylation pathway have been implicated in cancers of the breast, colon, ovaries and prostate. H4 sumoylation has been indirectly linked to transcriptional

repression, but the precise effect of histone sumoylation on chromatin structure and function is still unknown.

We utilized our expertise in chemical biology to synthesize homogeneous sumoylated histone H4 and incorporated it into chromatin arrays. Through analytical ultracentrifugation experiments we discovered that sumoylation inhibits chromatin compaction and aggregation. To our knowledge, this is the first reported repressive mark that favors a euchromatic state of chromatin. We further corroborated these results with single-molecule FRET studies and uncovered an intrinsic effect of SUMO in preventing inter- and intra-array interactions. Mechanistic studies into the effect of sumoylation revealed that SUMO's steric bulk plays an important role in preventing the compaction of chromatin arrays. This unexpected result, namely, that a repressive mark favors the euchromatic state led us to further examine the biochemical pathways by which sumoylation might mediate transcriptional repression. We explored the effect of sumoylation of H4 on the activity of the LSD1/CoREST/HDAC complex, which has been known to mediate gene repression through its deacetylation and demethylation activities. We tested the hypothesis that SUMO can mediate repression by localizing the LSD1/CoREST complex by binding to the SUMO interaction motif found in CoREST. We focused on the demethylation of methylated Lys4 of H3 (H3 K4me2), which is a mark associated with transcriptionally active promoter regions. To this end, we performed the first reported synthesis of the native full-length H3 K4me2, which was incorporated into nucleosomes along with sumoylated H4. On comparing the rate of demethylation of H3 K4me2 by the LSD1/CoREST complex, in sumoylated and non-sumoylated nucleosomes, we discovered that sumoylation can facilitate the removal of this activating mark. This novel cross talk between sumoylation and methylation presents one possible pathway by which histone sumoylation can mediate transcriptional repression.

Acknowledgements

This Ph.D. has been first and foremost made possible by my mentor and advisor Dr. Champak Chatterjee whose constant motivation and optimism made it easy to handle experimental setbacks which are part and parcel of scientific research. He not only helped me become experimentally sound, but also taught me the importance of timeliness in research, discussing ideas with lab mates and reading scientific literature broadly, all of which are essential qualities for becoming a well-rounded scientist.

Needless to mention, none of this would have been possible without the help and support of my family. I have to thank my parents for giving me the freedom to choose my career path and always being supportive of my decisions. I have to thank my sister and brother-in-law for looking after me and making sure my life outside lab is as comfortable as possible. The amount of time and energy they have spent on me makes me think that this Ph.D. belongs as much to them as to me.

I have also been very lucky to be surrounded by amazing lab mates who have also played a big part in my learning process. From stimulated discussions on science to politics and current affairs, they have made me aware of the diversity in opinions and the importance of accepting that diversity. They have definitely helped make graduate school fun and enriching.

Finally, I also want to thank all my teachers over the years who have significantly contributed in my growth and learning as a student. They have instilled a sense of curiosity in me and also taught me the importance of critical thinking. Without the contributions of everyone mentioned here, the road to a doctorate degree would have been much more perilous.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO CHROMATIN AND POST-TRANSLATIONAL MODIFICATIONS

1.1 The Chromatin: History

The nucleus of a eukaryotic cell stores all the necessary information needed to successfully perform basic cellular processes such as protein synthesis, replication, metabolism and signaling. This vast amount of information, present in the form of approximately three billion base pairs of DNA, is packaged in the tiny volume of ~500 μm^3 of a cell's nucleus as a nucleoprotein complex known as chromatin. In the early 1880s the German scientist Walther Flemming coined the term 'chromatin' based on the nature of this nucleoprotein complex to get easily stained by dyes due to its high affinity toward them[1]. For many decades the protein component of chromatin did not garner much attention as the limelight stayed on DNA for being the 'genetic material' as shown by Avery and co-workers in 1944 through bacterial transformation experiments[2]. All the excitement about DNA finally culminated in the proposal of a double-helical structure of DNA by James D. Watson and Francis Crick in 1953[3]. By this time, advances in electron-microscopy, nuclease digestion and the development of new protein fractionation and purification techniques were making it possible to see rudimentary images of chromatin showing a 'beads on a string' structure[4,5]. Seminal work by C. L. Woodcock, R. Kornberg and T. J. Richmond resulted in the establishment of the basic repeating unit of chromatin, known as a nucleosome[6,7].

Since the discovery of the nucleosome in the 1970s, we have come a long way to perceive chromatin as a dynamic signal integration platform that is precisely organized inside the nucleus of a eukaryotic cell with the help of many proteins. The complex structure of a nucleosome and

the fact that different phenotypes can arise due to small changes in the DNA component of the nucleosome (e.g. DNA methylation), and protein components of the nucleosome (e.g. post-translational modification of histone proteins) has given birth to a new field of study known as epigenetics. After decades of discussion and debate over the precise definition of epigenetics, we have now come to know this field as the study of heritable changes in gene function, brought about by developmental and environmental cues, which are independent of the DNA sequence.

1.2 The Chromatin: Organization

The packaging of nuclear DNA starts with its wrapping around an octameric protein complex formed by two copies each of the four core histone proteins- H2A, H2B, H3 and H4. A nucleosome core particle (NCP) is formed when 147 base pairs of DNA wrap around the histone octamer making about 1.65 helical turns[8]. The NCP is 11 nm in diameter and 5.5 nm high particle forms the 'quantum' of chromatin and is repeated over the entire length of DNA with 10 to 90 bp of linker DNA present between each of them. This repeating pattern of NCPs forms the lowest level of organization for chromatin known as the 'beads on a string' structure (Figure 1.1). The DNA entry and exit site on a NCP is bound by linker histones H1 or H5 to further condense chromatin into a 30 nm fiber, which forms the next level of packaging. Even though the beads on a string structure of chromatin has been known and accepted for a long time, the precise structure of the 30 nm fiber has been a matter of contention for the past decade[9]. Based on EM, Cryo-EM and NMR studies, either the 2-start zig-zag model or the one start solenoid model of nucleosome stacking has been proposed[10]. No conclusive evidence has been provided to support one model over the other as the linker length of the reconstituted arrays employed in these studies, or the source of the extracted chromatin is known to affect the type of 30 nm fiber stacking observed. It is also possible that multiple states of packaging are

permissible and simultaneously exist *in vivo* for facilitating the transformation of chromatin from a highly condensed heterochromatic state to a less compact euchromatic state and vice versa.

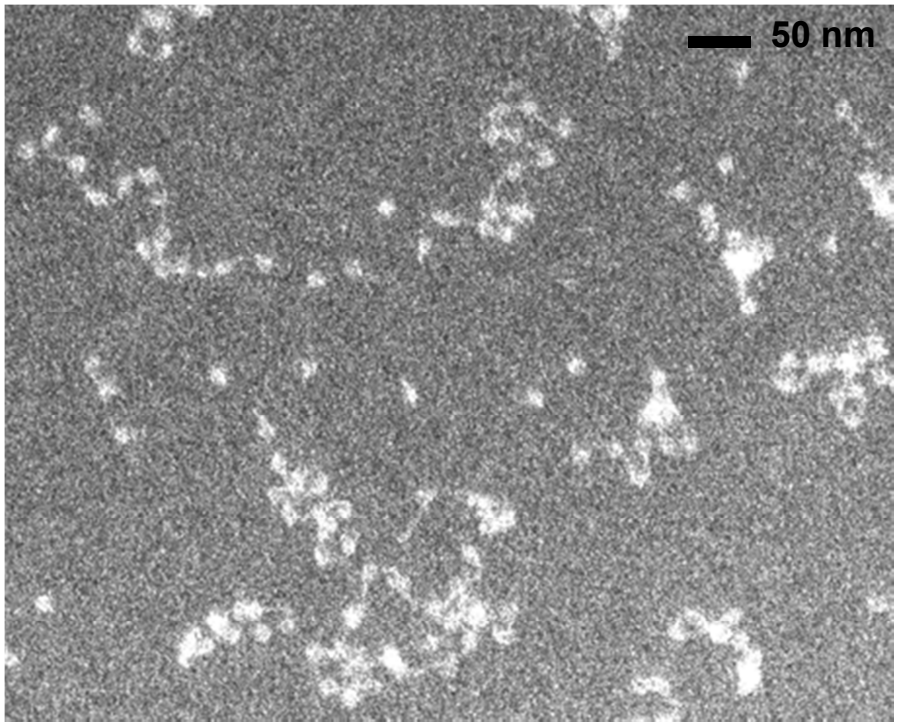


Figure 1.1. Negative stained electron microscopy image of reconstituted chromatin showing the lowest level of chromatin packaging, the 'beads on a string' structure. Naked DNA can be seen as the faint thread connecting individual mononucleosomes, which themselves appear as bright white dots.

To achieve further packaging, the 30 nm fibers form loops and interact with various scaffolding proteins such as CTCF and cohesin to form, on average, 800 kbp sized local chromatin interaction domains, termed as topologically associating domains (TADs)[11]. With recent advances in the chromatin conformation capture (3C) techniques we have now got a glimpse into the invariant nature of TADs and discovered vital regulatory clues such as interactions

between enhancers and their target promoters that occur only within a TAD but not across domains[11]. As the highest level of organization, the heterochromatic regions of the genome get sequestered near the nuclear membrane and the more actively transcribed regions are found in the interior of the nucleus[12]. The probabilistic nature of nuclear organization is evident by the multiple contacts observed for a single gene in a TAD through techniques like 4C and Hi-C. On the other hand, domains and compartments appear to be conserved across species and cell-types, forcing us to further examine the precise mechanisms by which three dimensional genome organization takes place. Since the three dimensional topology of chromatin strongly influences gene expression, it is vital to understand higher orders of chromatin organization.

1.3 The Nucleosome: Structure

The octameric core of a mononucleosome (MN) is approximately 100 kDa in size and is formed by the canonical histones H2A, H2B, H3 and H4, which are highly conserved across eukaryotes[8]. The core of the histone proteins consists of three α-helices which form the 'histone fold domain' (HFD), responsible for a majority of the protein-protein interactions that lead to the dimerization of H2A with H2B and H3 with H4 through a 'handshake' like interaction[13]. During nucleosome assembly the H3/H4 dimer interacts with another H3/H4 dimer via the H3 proteins to form a tetramer around which the wrapping of DNA is initiated. The tetrasome is capable of wrapping DNA in either a left-handed helix or a right handed helix[14]. Further interaction with an H2A/H2B dimer leads to the formation of a hexasome with the DNA partially wrapped around it. The dynamic wrapping and unwrapping of the DNA around the hexasome allows for further interaction with a second H2A/H2B dimer and leads to the formation of a stable nucleosome which locks the DNA in a left-handed helical conformation. Various histone chaperones are known to assist in the assembly and disassembly of a nucleosome[15,16]. The high resolution

crystal structure of a NCP by the Luger lab gave us further insight into the positioning of the DNA and the histones in this nucleo-protein complex (Figure 1.2)[17]. The symmetric nature of the histone octamer gives rise to a pseudo 2-fold axis of symmetry. The dyad axis of the nucleosome passes through a single base pair of DNA where the octamers are known to make the strongest interaction with the DNA. 73 bp of DNA is found on either side of this dyad, resulting in 147 bp of DNA in a NCP[17].

In addition to the structured core of the histones, there is a substantial unstructured N-terminal region, consisting of the first 20 to 30 residues observed in all the histones and is known as the 'histone tail'. The susceptibility of histone tails to proteolytic degradation was the first indication that they are solvent exposed and not as strongly associated with DNA as the core of the histone proteins. The high density of basic residues such as lysine and arginine makes the histone tails highly positively charged under physiological conditions and enables them to engage in electrostatic interaction with a negatively charged patch on the H2A/H2B dimer of a neighboring nucleosome. These tails play an essential role in regulating the structure and function of chromatin, as further discussed in the following sections.

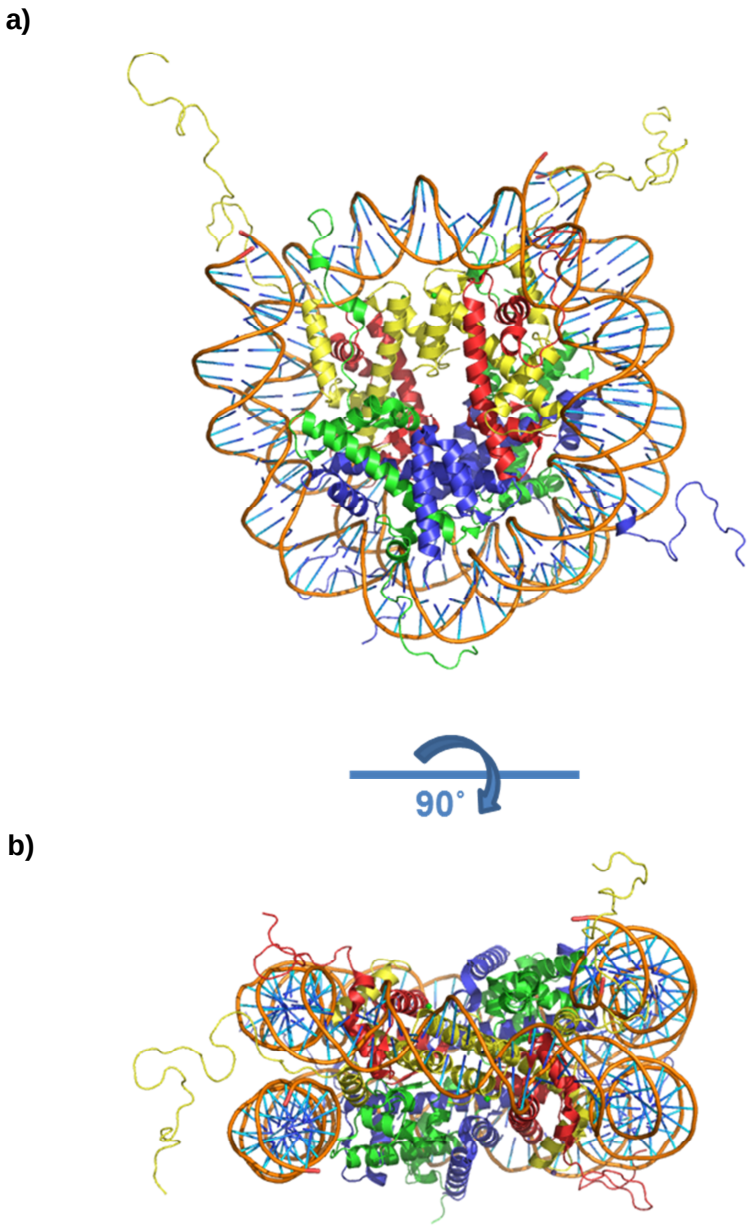


Figure 1.2. Structure of a mononucleosome. a) The structure of a mononucleosome at 1.9 Å resolution (PDB code 1KX5) showing individual histones and 147 bp of double-stranded DNA. The histone N-termini (tails) protrude from the nucleosome and are sites for extensive post-translational modifications. Green = H2A, Blue = H2B, Yellow = H3, Red = H4. b) Side-view of the mononucleosome showing the entry and the exit points of the dsDNA.

1.4 The Histones: Isoforms

Even though the primary function of histone proteins is to ensure efficient packaging of DNA, there are myriad other chromatin templated processes such as DNA repair, replication, meiotic recombination and chromosome segregation, that must be successfully performed to maintain cell viability. Various histone isoforms have evolved to perform specialized functions inside the eukaryotic cell. For example, the rapidly evolving histone H3 variant, known as CenH3, is essential for the formation of centromeres which ensures proper kinetochore assembly and equal segregation of chromosomes during mitosis and meiosis. The longer histone fold domain (HFD) of CenH3 helps in the formation of a more compact and rigid CenH3/H4 tetramer as compared to the canonical H3/H4 tetramers. This facilitates the formation of a heterochromatic centomeric region which is interspersed with canonical H3 containing nucleosomes and gives rise to a unique arrangement in which the CenH3 containing heterochromatin faces outward where kinetochore assembly takes place. Unlike regular histones whose transcription is tightly coupled with the replication of DNA, histone variants are constitutively expressed and incorporated into chromatin by specialized chaperones like the Holiday junction recognizing protein (HJURP) for CenH3[18].

A largely unexplored area of chromatin biology is the ability of histone variants to provide directionality to chromatin. An H2A variant, H2A.Z readily reconstitutes into heterogeneous nucleosomes containing one copy each of H2A and H2A.Z. Such nucleosome show greater salt-dependent stability than the canonical nucleosomes and can be readily observed in HeLa cells. The preferential interaction of H2A.Z containing face of the heterogeneous nucleosome with certain complexes like the polycomb repressive complex (PRC) or complexes containing homologues of SWI/SNF-related 1 (Swr1) can then lead to the propagation of these complexes in a particular direction. Studies done on H2A.Z containing nucleosomes have shown that the

larger acidic patch on H2A.Z leads to an enhanced binding of HP1α to nucleosomes, facilitating the establishment and spread of heterochromatin[19].

1.5 The Histones: Post-translational Modifications

The presence of histone variants provides some measure of diversity to the histone pool, but the numerous post-translational modifications (PTMs), also known as 'marks', found on the histone tails result in the generation of approximately 700 different isoforms of canonical histones as identified by proteomic studies employing HeLa cells[20]. This enormous complexity is a result of the presence of multiple modifiable residues on the histone tails such as histidine, serine, proline and many others. Histone tails are surprisingly rich in basic residues such as lysine and arginine which get extensively methylated and/or acetylated. Some modifications that may occur in multiples bring further complexity. For example, lysine methylation can occurring in varying degrees such as mono-, di- and trimethylation, has been shown to signal distinct functional outcomes. Arginine methylation is further complicated by the fact that there can be symmetric or asymmetric methylation which can be distinctly recognized by some interacting proteins. Lysine residues are also capable of getting acetylated which results in the neutralization of the positive charge on the lysine side chain, thereby altering the electrostatic environment around the histone tail. Proteomic studies have shown that histone PTMs are not only limited to small chemical groups like phosphoryl, glycosyl or ADP ribosyl, but small proteins including ubiquitin and various isoforms of the small ubiquitin-like modifier (SUMO) can also readily modify the histone tails (Figure 1.3).

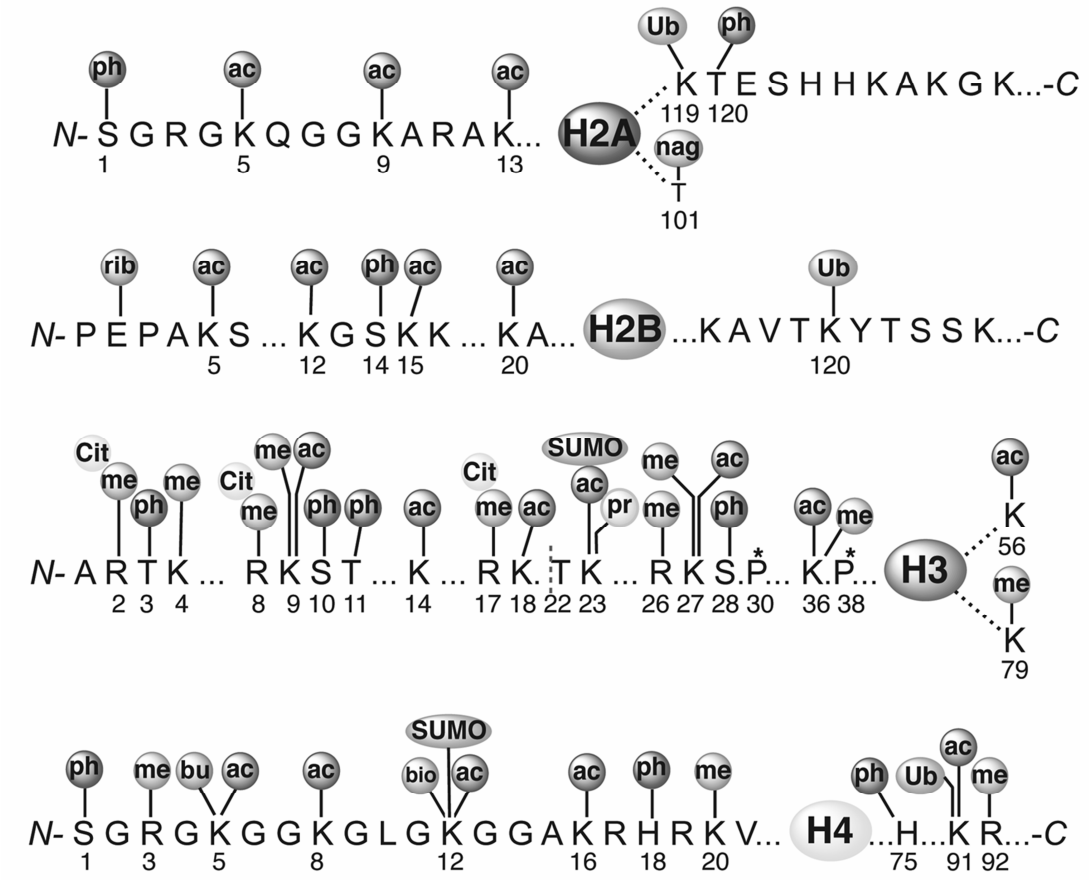


Figure 1.3. Schematic representation of modifications found at the N-termini, C-termini and the globular domains of the four core histones. Chemical groups are indicated as follows: ac, acetyl; bio, biotinyl; but, butyryl; cit, citrullyl; me, methyl; nag, N-acetylglucosaminyl; ph, phosphoryl; pr, propionyl; rib, ADP-ribosyl; SUMO, SUMOyl; and Ub, ubiquityl. Asterisks above Pro30 and Pro38 in the H3 tail indicate modification by *cis-trans* isomerization.

The solvent exposed nature of histone tails and the presence of multiple modifiable side chains allows for the simultaneous existence of various PTMs on histones. This has led to the proposal of a 'histone code' where specific combinations of PTMs or the sequential effect of certain PTMs, whose co-occurrence has been identified through proteomic or genomic methods, may result in a unique biological outcome[21]. The histone code attempts to provide some measure of

order to the complex, and seemingly chaotic world of histone post-translational modifications by associating groups of modifications with a desired phenotype. However, the significant changes in chromatin structure and function that may be brought about by a single modification cannot be ignored. As part of my thesis, I will provide evidence that sumoylation (modification by the protein SUMO) alone is sufficient to affect chromatin structure and function in the absence of any other modifications.

1.6 Post-translational Modifications: Readers, Writers and Erasers

Improved mass spectrometry techniques have resulted in the discovery of a dozen new PTMs in the last couple of decades[20]. With the discovery of new PTMs, there has been a constant effort to discover new enzymes responsible for installing these PTMs, commonly referred to as *writers*, enzymes that are responsible for reversing these PTMs, referred to as *erasers*, and proteins that bind specific PTMs, referred to as *readers*. With the exception of arginine methylation, all the other PTMs are known to be reversible, though the site-specificity of most of the writers and erasers still requires further investigation.

Decades before the discovery of the first histone acetyl transferase (HAT), Allfrey and colleagues had predicted the importance of histone acetylation and methylation in regulating RNA synthesis[22]. The discovery of the first HAT, general control of amino acid synthesis protein 5 (GCN5), opened up the doors for the search of other HATs and enzymes that would reverse PTMs, such as histone deacetylases (HDACs) which are responsible for converting acetylated lysine residues back to lysine. Just like many other chromatin modifying enzymes, HATs are associated with larger protein complexes that are responsible for modulating substrate and site-specificity. For example, GCN5 preferentially acetylates K9 and K14 of histone H3 in the context of free H3 but as part of the Spt-Ada-Gcn5-acetyltransferase (SAGA) complex, it acetylates

additional sites in H3 such as K56 and also H2B[23]. Acetylation is regarded to be generally associated with active transcription and another family of HATs, the CREB-binding protein/protein 300 (CBP/p300) is known to play a pivotal role in transcription. Other than histones, p300 is known to interact with hundreds of proteins and 47% of androgen-regulated genes are thought to be p300 dependent[23].

More than 1700 proteins in humans have been found to be acetylated[24]; hence the human cells have naturally evolved protein domains to read acetylation and also enzymes that reverse this modification. The most widely employed reader of acetylation is the highly conserved bromodomain which is observed in proteins functioning as transcriptional activators, chromatin remodelers and histone methyltransferases amongst others. The newest wave of epigenetics focused therapeutic drugs have been targeted against the bromodomain and extraterminal domain (BET) family of proteins whose aberrant activity is implicated in several human cancers[25,26]. Although bromodomain targeting inhibitors such as JQ1 and iBET are yet to receive FDA approval, HDAC inhibitors such as vorinostat (Zolinza; Merck) and romidepsin (Istodax; Celgene) have long been approved by the FDA for the treatment of cutaneous T cell lymphoma (CTCL) and peripheral T cell lymphoma (PTCL), respectively[27].

A modification which encodes multiple layers of complexity is the methylation of lysine and arginine residues by specific histone methyl transferases (HMTs). Methylation has been observed in various degrees (mono-, di- or trimethylation of lysines) and can also be symmetric or asymmetric in nature when observed on arginine residues. Unlike acetylation of histones, which is a prominent indicator of active gene transcription, methylation can be associated with discrete cellular processes based on the site and the degree of methylation. The trimethylation of lysine 4 of histone H3 (H3 K4me3), catalyzed by the Set1 (su(var)3-9, enhancer-of-zeste and trithorax) HMT in yeast, is found localized to the 5’ end of active genes where it is associated with active RNA Pol II. On the other hand, a less methylated form of the same protein, H3

K4me1, is associated with enhancer regions of various genes[28]. Exactly opposite to these roles is association of H3 K9me3 with constitutive and facultative heterochromatin regions in yeast and higher eukaryotes. The first ever identified HMT, Su(var)3-9, is responsible for installing this mark and establishing heterochromatic regions with the help of heterochromatin protein 1 (HP1) which specifically associates with this modification via its chromodomain[29].

Owing to the long *in vivo* half-life of methylated histones, this modification was thought to be irreversible, until in 2004 Yang Shi and colleagues described the discovery of the first histone demethylase called lysine specific demethylase 1 (LSD1 or KDM1A)[30]. LSD1 was shown to demethylate H3 K4me1/2 and is known to change its substrate specificity based on its binding partners. Since then another, much bigger family of demethylating enzymes, the Jumonji-C (JmjC) domain containing proteins, has been identified and is capable of demethylating trimethylated lysine residues too. This family of enzymes shows a very different mechanism of demethylation than LSD1 and makes use of Fe^{2+} and α-ketoglutarate to hydroxylate the methyl group present on the lysine side chain. The hydroxymethyl group is then released as formaldehyde to afford a demethylated lysine[31]. The Jumonji domain protein 2 (JMJD2) was the first identified demethylating enzyme in this family and was shown to specifically target H3 K9 and K36. The presence of a tandem tudor domain in JMJD2 targets this enzyme to active regions of the chromatin by recognizing methylation on H3 K4, which further facilitates gene activation by removing the repressive H3 K9me3 mark[32].

1.7 Post-translational Modifications: *Cis* Effects

In order to orchestrate quick responses to various stimuli, the cell uses phosphorylation cascades that relay information to the nucleus. These result in the activation of specific transcription factors and chromatin modifying enzymes to modulate the transcription level of

genes needed to achieve homeostasis. Unfortunately, our understanding of the precise changes in histone PTMs of a cell upon its receiving various external stimuli is still rudimentary, but the changes in chromatin landscape that happen due to cell cycle progression or internal stimuli like DNA damage, mitosis or meiosis and replication, are well characterized. Histone PTMs can have a transient or a long lasting effect on gene transcription by two mutually non-exclusive mechanisms known as the *cis* or the *trans* effect.

Modifications can affect chromatin structure in *cis* by having a direct effect on the stability of a nucleosome or the higher order packaging of chromatin. Modifications in the core domain of histones, such as acetylation of H3 K56 by the regulator of Ty1 transposition 109 (Rtt109) acetyltransferase, can weaken the critical electrostatic interactions made by the positive lysine side chain with the negatively charged phosphate backbone of DNA. The addition of an acetyl group can also destabilize the protein-DNA interactions by presenting steric clashes with the major groove of the DNA that passes over this residue[33]. Indeed, *in vitro* experiments employing optical tweezers showed increased unwrapping of DNA when H3 K56 is acetylated as it is present in the vicinity of DNA entry and exit points. The increased 'breathing' of DNA results in increased transcription factor binding, hence affecting the level of gene transcription[34]. Other than being a direct result of weakened interaction between the core of the histone octamer and the DNA, *cis* effects can also arise from the destabilization of the octameric core itself. Both acetylation and ubiquitylation of Lys91 in H4 have been observed in eukaryotes and shown to be important for successful response to DNA damage. This residue lies at the interface of the interaction between the H3/H4 tetramer with the H2A/H2B dimers, hence modification at this position leads to poor octamer formation and greater DNA unwrapping, as observed during micrococcal nuclease digestion assays on chromatin recovered from a strain of yeast stably expressing H4 K91A[35,36].

In a ground breaking study on the acetylation of the N-terminal region of H4, Shogren-Knaak et. al showed that the *cis* effects on chromatin structure are not only limited to modifications in the core of the octamer[37]. The importance of histone tails in mediating higher order chromatin compaction became evident when *in vitro* experiments showed that acetylation of Lys16 in H4 greatly inhibited inter-nucleosomal interactions by disrupting the electrostatic interaction between the positively charged H4 tail with an acidic patch on the H2A/H2B dimer of a neighboring nucleosome[37]. Since then, not many other modifications in the histone tails have been shown to affect chromatin structure, but as discussed in Chapter 2, I will present evidence of the *cis* effect of sumoylation of the H4 tail, which results in disruption of the higher order chromatin structures, albeit through a different mechanism than acetylation.

1.8 Post-translational Modifications: *Trans* Effects

In the complex landscape of chromatin, which can harbor hundreds of site-specific histone tail modifications at any given time, the accurate readout, deposition or removal of PTMs is carried out by chromatin-modifying enzymes that are generally occur within large multisubunit complexes. These subunits contain histone PTM recognition domains that bind specific PTMs and may even show site-specificity in their binding by recognizing the neighboring amino acids. The first histone reader domain was characterized in 1999 when an NMR based solution structure of the bromodomain of the HAT, p300/CBP-associated factor, was reported[38]. To date, three domains capable of recognizing acetylated lysine have been identified, namely the bromodomain, the plant homeodomain (PHD) and the double pleckstrin homology (PH) domain. Of these, bromodomains are the most commonly found and the most extensively studied. On the other hand, multiple readers of methyllysine group have been identified and extensively characterized. These include ATRX-DNMT3-DNMT3L (ADD), ankyrin, bromo-adjacent

homology (BAH), chromobarrel, chromodomain, double chromodomain (DCD), malignant brain tumor (MBT), plant homeodomain (PHD), PWWP (Pro-Trp-Trp-Pro), tandem Tudor domain (TTD), Tudor, WD40 and the zinc finger CW (zf-CW) domain.

A well characterized example of a *trans* effect of a PTM was reported by Narlikar and co-workers when they interrogated the mechanism of HP1 mediated heterochromatin spreading. Suv39 (known as Clr4 in yeast) preferentially dimethylates H3 K9 and also binds this mark with its chromodomain. HP1 then preferentially binds H3 K9me2 and recruits silencing factors to generate a heterochromatic state. The binding of HP1 displaces Suv39 which then binds and methylates neighboring MNs, thus resulting in the spreading of the heterochromatic region[39,40]. A cascade of trans effects has also been observed upon phosphorylation of S10 in H3. This mark is identified by the 14-3-3 protein which forms a complex with the histone acetyltransferase MOF. This results in the acetylation of H4 K16, which is further recognized by the bromodomain of BRD4. This releases the promoter-proximal pausing of RNA polymerase II and stimulates transcription elongation[41].

This crosstalk is not only limited to small chemical PTMs but has also been observed in the case of modification by ubiquitin. In 2008 Muir and co-workers showed that ubiquitylation at Lys120 of H2B stimulates methylation of H3 K79 by the HMT hDot1[42]. They further showed that the stimulation of hDot1 is not site-specifically linked to ubiquitylation at Lys120 but the presence of ubiquitylation in the spatial vicinity of this site is sufficient to stimulate methylation[43]. In Chapter 3, I will provide evidence for the crosstalk between sumoylation and methylation in the context of a nucleosome, and reveal how sumoylation can affect chromatin in *trans*.

1.9 Chemical Biology Tools: History

In the past few decades there has been a rapid expansion in the types and sites of modifications observed on histones. Currently there are more than 60 unique positions on the four core histones where modifications which include acetylation, ADP-ribosylation, butyrylation, biotinylation, citrullination, glycosylation, methylation, phosphorylation, propionylation, proline isomerization, sumoylation and ubiquitylation have been observed (Figure 1.4)[44]. Initial reports focused on elucidating the roles of histone modifications made use of solid-phase peptide synthesis (SPPS) to synthesize short peptides corresponding to histone tails and bearing appropriate modified amino acids such as an acetylated lysine. Using this approach Allfrey and co-workers determined that HDACs purified from calf thymus cells are capable of deacetylating Lys12 and Lys16 from the uniformly acetylated H4 tail, but require a long portion of the H4 tail for robust activity[45,46]. Peptide based approaches are relevant till date and are used to identify novel readers of modifications or the effect of neighboring PTMs in a high-throughput manner. In 2010, Garske and co-workers employed a 5000 member peptide library mimicking the H3 tail containing methylation at Lys4, and a combinatorial pattern of other PTMs on the neighboring residues. They used this one bead one compound (OBOC) approach to interrogate the effect of threonine phosphorylation and arginine methylation on the capacity of various PHD finger domains to recognize methylation of Lys4 in H3. This revealed an important aspect of chromatin binding proteins because they noted that while some modifications may act like a binary switch to completely prevent methyl readers from binding methylated Lys4, most of the PTMs have a protein-specific and graduated effect[47]. Hence, some PTMs can act as master switches, while many of them play a role in fine tuning a biological response.

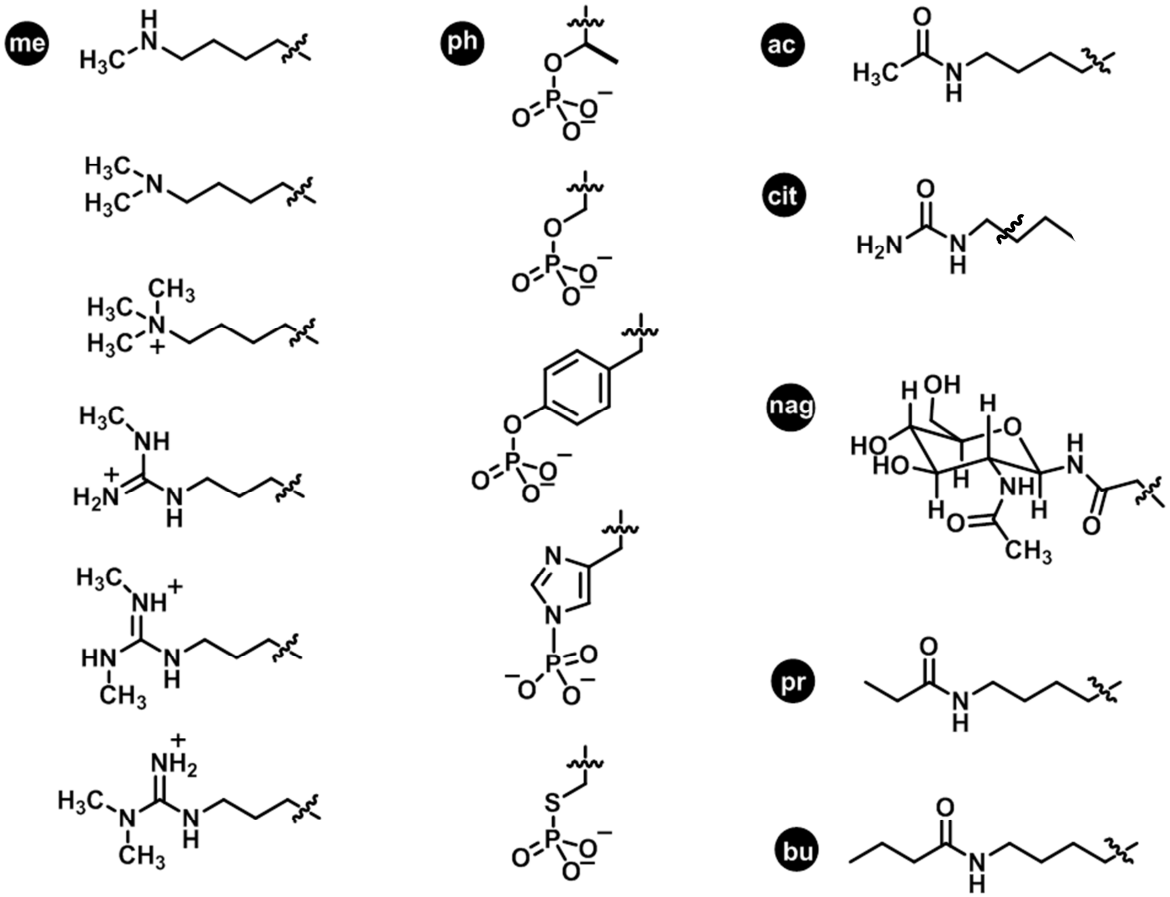


Figure 1.4. Chemical structures of a few post-translational modifications that have been observed *in vivo*. The left-most column denotes the various states of methylation observed for lysine and arginine residues. The middle column depicts phosphorylation of threonine, serine, tyrosine, histidine and cysteine residues. The right-most column shows acetylation of lysine, citrunillation of arginine, n-acetylglucosaminylation of asparagine, propionylation of lysine and butyrylation of lysine. Chemical groups are indicated as follows: ac, acetyl; but, butyryl; cit, citrullyl; me, methyl; nag, N-acetylglucosaminyl; ph, phosphoryl; pr, propionyl.

Even though peptide based approaches have been instrumental in the identification and characterization of histone PTM recognizing and modifying enzymes, the precise crosstalk of modifications and the roles of these effectors in the context of chromatin cannot be probed using peptides. Chromatin effectors are generally present in multisubunit complexes that contain multiple binding modules which can bind multiple histone tails, DNA or even neighboring nucleosome simultaneously. In order to probe such interactions, considerable effort has been put into synthesizing homogeneously modified full-length histones which can be incorporated into relevant chromatin substrates.

1.10 Chemical Biology Tools: Native Chemical Ligation

In 1994, Kent and co-workers were the first to report a technique known as native chemical ligation (NCL) which enabled the stitching of two peptides via a native peptide bond. Two peptide fragments which can be synthesized by solid-phase peptide synthesis employing either a tert-butyloxycarbonyl (Boc) or N-(9-fluorenyl)methoxycarbonyl (Fmoc) protecting group strategy could be ligated if the N-terminal peptide contained a C-terminal thioester and the C-terminal peptide contained a 1,2-amino thiol, such as a cysteine, at the N-terminus[48]. Due to the limitation of SPPS to only generate peptides approximately 50 amino acids long, NCL was further improved to make possible the synthesis of longer proteins. The new technique, known as expressed protein ligation (EPL), makes use of either a recombinantly produced C-terminal α-thioester or a recombinantly generated C-terminal protein fragment containing cysteine at the start site (Figure 1.5). The ligation between the two fragments in both NCL and EPL proceeds with an initial trans-thioesterification reaction between the α-thioester and the free thiol group present at the N-terminus of the C-terminal protein/peptide fragment.

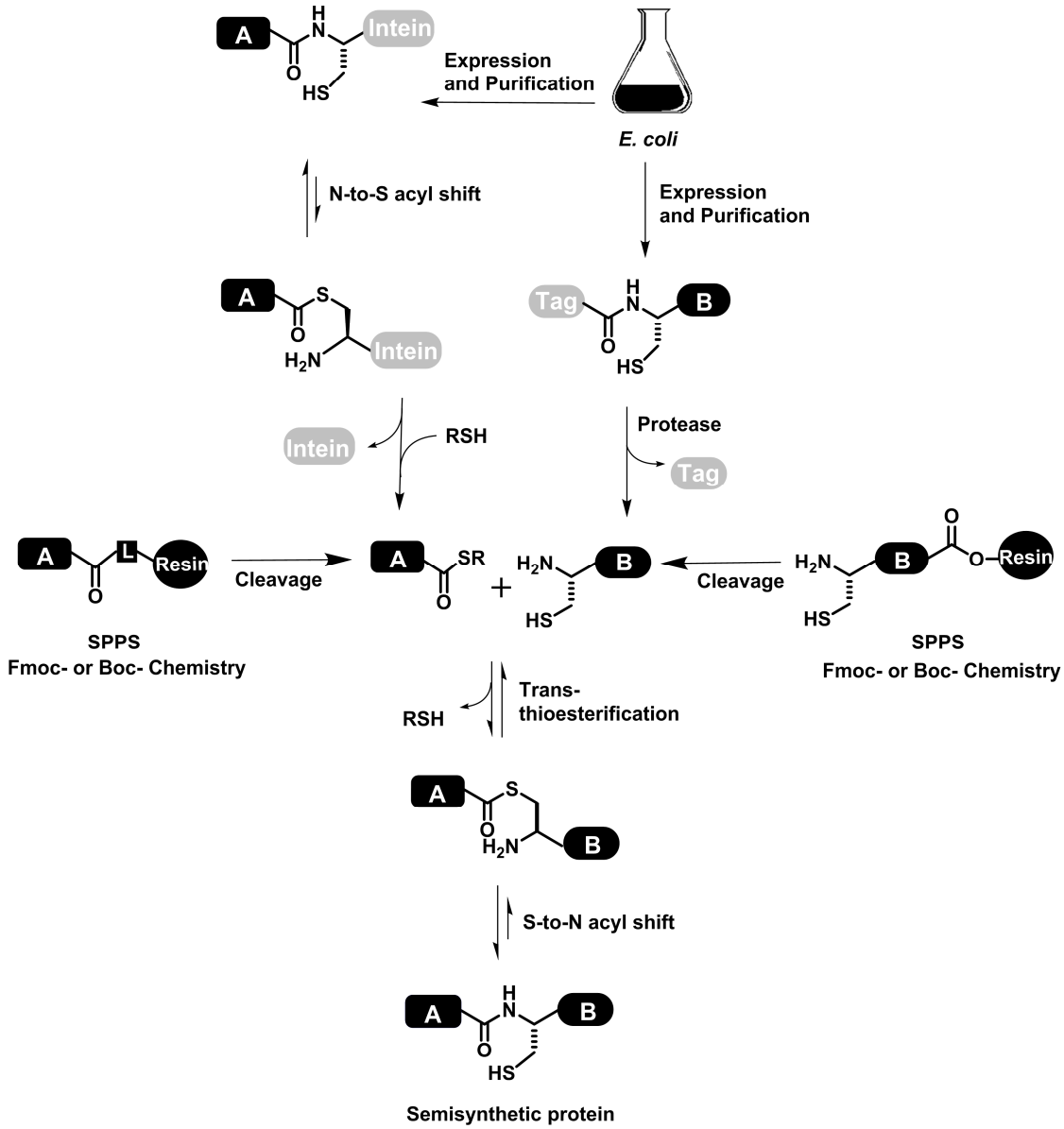


Figure 1.5. Native Chemical Ligation and Expressed Protein Ligation. Scheme depicting the semisynthesis of proteins by ligating fragments generated by solid-phase peptide synthesis or expressed and purified from *E. coli*. The presence of a C-terminal α-thioester in fragment A and a cysteine at the N-terminus of fragment B results in a reversible trans-thioesterification reaction when these two fragments are mixed together. The fast S-to-N acyl shift gradually converts the thioester bond between fragments A and B to a more stable amide bond which is less amenable to a N-to-S acyl shift, hence resulting in a stable semisynthetic protein.

A thermodynamically favored S-to-N-acyl shift then results in a native amide linkage between the two fragments to generate full-length target protein[49].

The use of SPPS allows for the incorporation of multiple, discrete modified residues in the peptide fragment simultaneously. Several linkers such as diaminobenzoic acid or 2-hydroxy-3-mercaptopropionic acid have been developed to afford a C-terminal thioester at the end of peptide synthesis, but one linker that is quickly becoming the method of choice and is also employed by us in our studies, is the hydrazine linker that can be easily utilized to generate an α-thioester in high yield[50,51]. The generation of an N-terminal cysteine containing protein fragment by recombinant protein expression in *E. coli* requires more effort due to the fact that it is absolutely necessary to have methionine as the starting codon for protein synthesis. Poor processing of this methionine can result in an unexposed cysteine that can be strategically introduced right after the first methionine residue. In order to circumvent this problem, protein fusions can be employed, in which a tobacco etch virus (TEV) protease cleavage site can be introduced between the starting methionine and the required cysteine[52]. After purification of the fusion protein, the cysteine can be exposed by cleaving the extra N-terminal part by incubation with TEV protease. It is also possible to generate C-terminal thioesters by means of recombinant protein expression. Again, a protein fusion strategy is utilized for this task, in which the target protein is fused to the N-terminus of an engineered protein domain known as an intein. Inteins are autosplicing proteins that cut themselves out from the recombinantly obtained target-protein-intein fusion when exposed to a small-molecule thiol. Most of the EPL studies have made use of the slower *Mycobacterium xenopi* (*Mxe*)-GyrA intein to generate C-terminal thioesters of target proteins, but the recent discovery of faster self-splicing inteins, which have shorter splicing half-lives, has made the process much faster and also improved yields due to the reduced hydrolysis of the generated thioester[52,53].

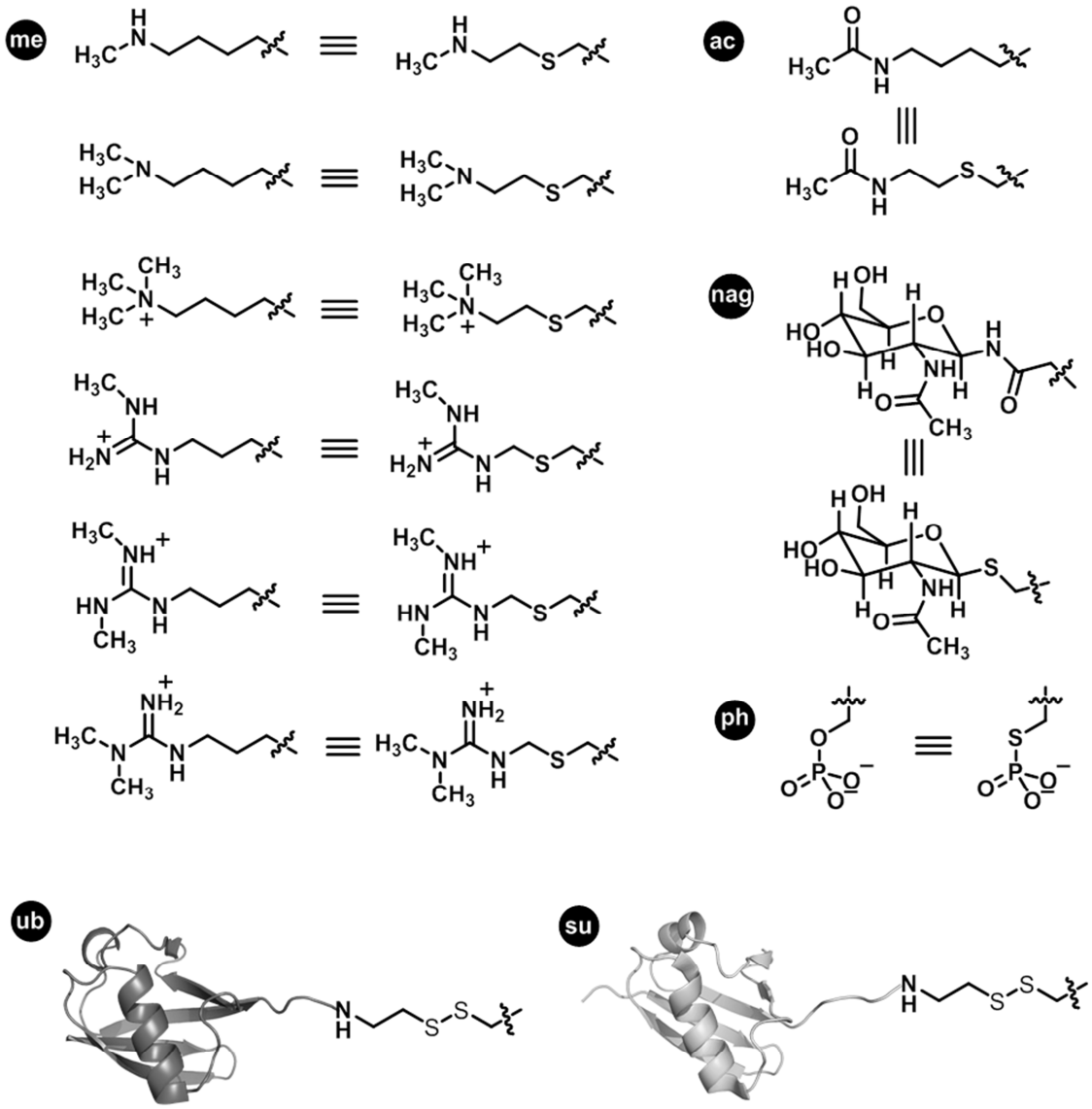


Figure 1.6. Cysteine based analogs of a few histone post-translational modifications that have been observed *in vivo*. Small chemical groups can be introduced by taking advantage of the nucleophilic nature of the sulfhydryl functional group in the side-chain of cysteine. This strategy has been used to generate analogs of lysine methylation. Sulfhydryl radical mediated click chemistry has been employed to generate mimics of acetylation and arginine methylation. Modification by small proteins can also be mimicked by means of disulfide-directed protein attachment. Chemical groups are indicated as follows: ac, acetyl; me, methyl; nag, N-acetylglucosaminyl; ph, phosphoryl; su, SUMOyl; ub, ubiquityl.

1.11 Chemical biology tools: Cysteine based modifications

The low occurrence of cysteine in intracellular proteins (approximately 1.5 % of residues are cysteines), and its reactive sulfhydryl side chain, makes it a valuable tool for chemical biology. Fortunately, histones are devoid of cysteines except for the presence of C110 in H3, which can be mutated to an alanine without any loss in structure or function[43]. Thus, cysteine based modification strategies have been developed to generate acetylated, methylated, sumoylated and ubiquitylated histone mimics (Figure 1.6). The most prominent example of this strategy is the generation of various methyl lysine analogs (MLAs) upon nucleophilic attack by the thiol of a cysteine on to an electrophilic ethylamine[54]. MLAs have been used to probe the effect of methylation on chromatin structure[55], mechanism of HP1 mediated gene silencing[39] and also the binding of various remodelers such as Isw1 and CHD1 to chromatin[56]. Another straightforward use of the cysteine residue has been to generate thialysine analogues that mimic acetylation at various Lys residues in the H4 tail. As further discussed in Chapter 2, we use one such strategy to generate acetylated H4 which was instrumental in revealing the site-specific effect of acetylation on chromatin structure. The versatility of cysteine based modifications is reflected in the fact that it has not only made possible to mimic small chemical modifications, but has also been utilized to site-specifically attach two proteins together. In Chapter 2, we present a site-specific disulfide formation strategy to generate sumoylated H4, which enabled unprecedented biophysical characterization of sumoylated chromatin.

1.12 REFERENCES

(1) Olins, D. E., and Olins, A. L. (2003) Chromatin history: our view from the bridge. *Nat. Rev. Mol. Cell Biol. 4*, 809–14.

(2) Avery, O. T., MacLeod, C. M., Mccarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of Pneuumococcal types. *J. Expt. Med. 79*, 137–58.

(3) Watson, J. D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids. *Nature 171*, 737-738.

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

(5) Olins, A. L., Olins, D. E. (1974) Spheroid Chromatin Units (v Bodies) Spheroid Chromatin Units (v Bodies). *Science 183*, 330–332.

(6) Kornberg, R. D. (1974) Chromatin Structure : A Repeating Unit of Histones and DNA. *Science184*, 868–871.

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

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

(9) Belmont, A. S. (2014) Large-scale chromatin organization: The good, the surprising, and the still perplexing. *Curr. Opin. Cell Biol. 26*, 69–78.

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

(11) Dekker, J., and Heard, E. (2015) Structural and functional diversity of Topologically Associating Domains. *FEBS Lett. 589*, 2877–2884.

(12) de Graaf, C. A., and van Steensel, B. (2013) Chromatin organization: form to function. *Curr. Opin. Genet. Dev. 23*, 185–90.

(13) Cutter, A. R., and Hayes, J. J. (2015) A brief review of nucleosome structure. *FEBS Lett. 589*, 2914–2922.

(14) Talbert, P. B., and Henikoff, S. (2010) Histone variants--ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol. 11*, 264–75.

(15) Gurard-Levin, Z. A., Quivy, J.-P., and Almouzni, G. (2014) Histone Chaperones: Assisting Histone Traffic and Nucleosome Dynamics. *Annu. Rev. Biochem. 83*, 487–517.

(16) Burgess, R. J., and Zhang, Z. (2014) Histone chaperones in nucleosome assembly and human disease. *Nat Struct Mol Biol. 20*, 14–22.

(17) McGinty, R. K., and Tan, S. (2015) Nucleosome structure and function. *Chem. Rev. 115*, 2255–2273.

(18) Torras-Llort, M., Moreno-Moreno, O., and Azorín, F. (2009) Focus on the centre: the role of chromatin on the regulation of centromere identity and function. *EMBO J. 28*, 2337–2348.

(19) Chakravarthy, S., and Luger, K. (2006) The histone variant macro-H2A preferentially forms “hybrid nucleosomes.” *J. Biol. Chem. 281*, 25522–25531.

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

(21) Strahl, B. D., and Allis, C. D. (2000) The language of covalent histone modifications. *Nature*

403, 41–45.

(22) Allfrey, V., Faulkner, R., and Mirsky, A. (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. 315*, 786–794.

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

(24) Farria, A., Li, W., and Dent, S. Y. R. (2015) KATs in cancer: functions and therapies. *Oncogene* 1–13.

(25) Brand, M., Measures, A. M., Wilson, B. G., Cortopassi, W. A., Alexander, R., Hess, M., Hewings, D. S., Rooney, T. P. C., Paton, R. S., and Conway, S. J. (2015) Small molecule inhibitors of bromodomain - Acetyl-lysine interactions. *ACS Chem. Biol. 10*, 22–39.

(26) Filippakopoulos, P., and Knapp, S. (2014) Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discov. 13*, 337–356.

(27) Bojang, P., and Ramos, K. S. (2014) The promise and failures of epigenetic therapies for cancer treatment. *Cancer Treat. Rev. 40*, 153–169.

(28) Black, J. C., Van Rechem, C., and Whetstine, J. R. (2012) Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell 48*, 491–507.

(29) Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature 406*, 593–599.

(30) Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004) Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell 119*, 941–953.

(31) Culhane, J. C., and Cole, P. A. (2007) LSD1 and the chemistry of histone demethylation. *Curr. Opin. Chem. Biol. 11*, 561–8.

(32) Mosammaparast, N., and Shi, Y. (2010) Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu. Rev. Biochem. 79*, 155–179.

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

(34) Manohar, M., Mooney, A. M., North, J. A, Nakkula, R. J., Picking, J. W., Edon, A., Fishel, R., Poirier, M. G., and Ottesen, J. J. (2009) Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding. *J. Biol. Chem. 284*, 23312–21.

(35) Yan, Q., Dutt, S., Xu, R., Graves, K., Juszczynski, P., Manis, J. P., and Shipp, M. A. (2009) BBAP Monoubiquitylates Histone H4 at Lysine 91 and Selectively Modulates the DNA Damage Response. *Mol. Cell 36*, 110–120.

(36) Ye, J., Ai, X., Eugeni, E. E., Zhang, L., Carpenter, L. R., Jelinek, M. A., Freitas, M. A., and Parthun, M. R. (2005) Histone H4 lysine 91 acetylation: A core domain modificationassociated with chromatin assembly. *Mol. Cell 18*, 123–130.

(37) Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., and Peterson, C. L. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science 311*, 844–7.

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

(39) Canzio, D., Chang, E. Y., Shankar, S., Kuchenbecker, K. M., Simon, M. D., Madhani, H. D., Narlikar, G. J., and Al-Sady, B. (2011) Chromodomain-mediated oligomerization of HP1

suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol. Cell 41*, 67–81.

(40) Al-Sady, B., Madhani, H. D., and Narlikar, G. J. (2013) Division of labor between the chromodomains of HP1 and Suv39 methylase enables coordination of heterochromatin spread. *Mol. Cell 51*, 80–91.

(41) Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009) Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation. *Cell 138*, 1122–1136.

(42) McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G., and Muir, T. W. (2008) Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature 453*, 812–6.

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

(44) Dhall, A., and Chatterjee, C. (2011) Chemical Approaches To Understand the Language of Histone Modifications. *ACS Chem. Biol. 6,* 987–999.

(45) Vidali, G., Gershey, E. L., and Allfrey, V. G. (1968) Chemical Studies of Histone Acetylation. *J. Biol. Chem. 243*, 6361–6366.

(46) Kiueger, D. E., and Vidali, G. (2003) The Synthesis of Diacetylated Histone H4-(1-37) for Studies on the Mechanism of Histone Deacetylation. *Bioorg. Chem. 1979*, *427*, 1–19.

(47) Garske, A. L., Oliver, S. S., Wagner, E. K., Musselman, C. A, LeRoy, G., Garcia, B. A, Kutateladze, T. G., and Denu, J. M. (2010) Combinatorial profiling of chromatin binding modules reveals multisite discrimination. *Nat. Chem. Biol. 6*, 283–290.

(48) Dawson, P. E., and Kent, S. B. H. (2000) Synthesis of Native Proteins by Chemical Ligation. *Ann. Rev. Biochem 69*, 923-960.

(49) Muir, T. W. (2003) Semisynthesis of Proteins by Expressed Protein Ligation. *Annu. Rev. Biochem. 72*, 249–289.

(50) Zheng, J.-S., Tang, S., Qi, Y.-K., Wang, Z.-P., and Liu, L. (2013) Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc. 8*, 2483–95.

(51) Fang, G.-M., Li, Y.-M., Shen, F., Huang, Y.-C., Li, J.-B., Lin, Y., Cui, H.-K., and Liu, L. (2011) Protein chemical synthesis by ligation of peptide hydrazides. *Angew. Chem. Int. Ed. Engl. 50*, 7645–9.

(52) Holt, M., and Muir, T. (2015) Application of the protein semisynthesis strategy to the generation of modified chromatin. *Annu. Rev. Biochem. 84*, 265–90.

(53) Fierz, B., and Muir, T. W. (2012) Chromatin as an expansive canvas for chemical biology. *Nat. Chem. Biol. 8*, 417–27.

(54) Simon, M. D., Chu, F., Racki, L. R., de la Cruz, C. C., Burlingame, A. L., Panning, B., Narlikar, G. J., and Shokat, K. M. (2007) The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell 128*, 1003–12.

(55) Lu, X., Simon, M. D., Chodaparambil, J. V, Hansen, J. C., Shokat, K. M., and Luger, K. (2008) The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure *15*, 1122–1124.

(56) Leonard, J. D., and Narlikar, G. J. (2015) A Nucleotide-Driven Switch Regulates Flanking DNA Length Sensing by a Dimeric Chromatin Remodeler. *Mol. Cell 57*, 850–859.

CHAPTER 2

BIOPHYSICAL STUDIES OF SUMOYLATED HISTONE H4

2.1 INTRODUCTION

In humans, the SUMO family of proteins consists of three functional paralogs, SUMO-1, SUMO-2 and SUMO-3, which are conjugated to the ε-amine of Lys side-chains in their protein targets by a family of E1-E3 ligases. The mature forms of SUMO-2 and SUMO-3 are 97% identical in sequence, but share only 47% sequence identity with SUMO-1. Another paralog of SUMO, called SUMO-4, which shares 86% sequence identity with SUMO2/3 has been observed in kidney, spleen and lymph nodes, but has yet to be identified as a covalent modification of other proteins[1]. On the other hand, SUMO-1 and SUMO-2/3 have been observed as modifications of multiple proteins in higher eukaryotes such as parasites, plants and humans. SUMOs are conjugated to their substrates by an enzyme cascade very similar to the one involved in ubiquitylation. The SUMO E1 enzyme is a heterodimer comprised of SUMO-activating enzyme subunit 1 (SAE1; also known as Aos1) and SUMO activating enzyme subunit 2 (SAE2; also known as Uba2). The SAE1/SAE2 E1 activates SUMO's C terminus in an ATP-dependent reaction that involves the formation of a SUMO adenylate species followed by a SAE2-SUMO thioester bond formation. Thus far, only one E2 conjugating enzyme, Ubc9, has been discovered for the sumoylation pathway. Ubc9 is capable of binding a consensus motif, ψKXE (where ψ is a bulky aliphatic residue, typically L/I/V and X can be any amino acid), in target proteins and sumoylating them even in the absence of an E3 ligase, which is typically involved in providing substrate specificity. The largest family of SUMO E3 ligases is a group of SP-RING-

containing proteins known as protein inhibitor of activated STAT (PIAS) family[2]. Although knockout studies in mice suggest some redundancy in function between the various PIAS splice variants, their cellular levels of expression and localization can provide unique function to each one of them. For example, PIASy$^{-/-}$ mice develop and age normally, but primary fibroblasts lacking PIASy show a highly reduced propensity to undergo senescence[3].

All four core histones, along with a few histone variants such as H2A.Z, H2AX and many other chromatin associated proteins including the linker histone H1, HDAC1 and HP1α are modified by SUMO[4]. However, the modification is transient in nature unlike methylation or acetylation which can persist throughout the cell cycle. Initial studies by Shiio and Eisenman identified histone H4 as the primary histone target for modification by SUMO in human 293T cells and B-lymphocytes[5]. The site of sumoylation was proposed to be in the flexible N-terminal tail of H4 as a synthetic 26-mer tail peptide was observed to be sumoylated by Ubc9 *in vitro*. Subsequent proteomic studies by Thibault and co-workers identified K12 in H4 as the major site of conjugation by SUMO-2/3 in HeLa cells, although *in vivo* modification to varying degrees at other sites could not be ruled out[6]. A recent report of the sumoylome in HeLa cells, which employed immunoprecipitation of SUMO targets from cell lysates with monoclonal antibodies directed toward SUMO-1 or SUMO-2/3, revealed that H4 is modified to similar amounts by these paralogs[7]. Collectively, these studies established H4 as a target for sumoylation in mammalian cells, and opened up the interrogation of the direct effects of SUMO on chromatin structure and function.

Recently, a double affinity chromatin immunoprecipitation strategy, termed ChDIP that employed FLAG-tagged histones and hexahistidine (His_6)-tagged SUMO was used to examine the genome-wide prevalence of histone-associated SUMO in baker's yeast *Saccharomyces cerevisiae*. This revealed a two-fold increase in SUMO near the ends of chromosomes, close to the telomeres, although SUMO was also found to a lesser degree throughout genes[8].

Furthermore, a dynamic and inverse correlation was observed between sumoylation and histone acetylation at the Gal promoter site when *S. cerevisiae* was challenged with different carbon sources. These results were interpreted to suggest that histone sumoylation may repress gene transcription by opposing histone acetylation. In support of this theory, FLAG-epitope-tagged linear SUMO-1/3-H4 fusion proteins pulled down HDAC1 which is generally associated with gene repression. Thus, several lines of indirect evidence point to a potential repressive role for histone sumoylation. However, the specific effects of chromatin sumoylation have remained a mystery for over a decade.

For the first part of my thesis project, I investigated the direct effect of H4 sumoylation on chromatin structure. One mechanism by which histone PTMs may influence the chromatin landscape is by modulating the degree of chromatin compaction. Electrostatic interaction between the positively charged H4 tail and an acidic patch at the H2A-H2B interface on an adjacent nucleosome is known to be critical for inter-nucleosomal interactions. Modifications such as acetylation at K16 in the H4 tail (H4 K16ac) are known to abrogate this interaction and prevent chromatin compaction by neutralizing the positive charge of the lysine side chain[9]. On the other hand, H4 K20me3, a repressive modification found in mammalian cells, is the only known modification to facilitate chromatin compaction *in vitro,* as suggested by sedimentation velocity analysis of chromatin arrays reconstituted with a thialysine analog of H4 K20me3[10]. Another large modification, ubiquitylation of Lys119 of H2A, which is associated with gene repression and lies away from the interface of a dinucleosome, allows for the compaction of chromatin arrays and can mediate gene repression[11,12]. We wondered if sumoylation at Lys12 of H4 may lead to gene repression by facilitating the compaction of chromatin arrays, especially when the crystal structure of a tetranucleosome suggests that this region of the H4 tail is far removed from the interface of interaction between neighboring nucleosomes.

In order to test this hypothesis, we required homogeneous sumoylated H4. Thus far, mechanistic studies of the precise roles for histone PTMs in dictating chromatin structure have been limited by the unavailability of homogeneously modified histones from biological sources. This is due to the fact that histone modifications are heterogeneous and dynamic in nature. Furthermore, many of the enzymes that install site-specific histone modifications remain unknown, which precludes enzymatic generation of uniformly modified histones *in vitro*. Entirely synthetic and semisynthetic methods that yield uniformly modified histones can overcome the limitations imposed by the heterogeneity of biological samples. These approaches have previously enabled detailed characterization of methylated, acetylated and ubiquitylated histones and have provided valuable insight regarding modification-specific effects on chromatin structure. In order to generate homogeneously sumoylated human H4 protein, we employed a disulfide-directed protein modification strategy. This permitted the unprecedented biophysical analysis of homogenously sumoylated mononucleosomes (MNs) and nucleosome arrays, which revealed the inhibitory effect of sumoylation on intra- and inter-array interactions. We further corroborated the sedimentation coefficient measurements with a single-molecule Försters resonance energy transfer (FRET) based technique to measure the intrinsic effect of SUMO on inter-nucleosomal interactions independent of the nucleosome repeat length of the linker DNA between adjacent nucleosomes. Furthermore, our mechanistic studies revealed that the effect of SUMO is likely a consequence of its steric bulk and not just due to charge neutralization of K12 in H4. Thus, my results have revealed that sumoylation in the H4 tail enforces a *cis* effect on the structure of chromatin by preventing compaction and opened avenues for various biochemical investigations to understand the proposed repressive nature of sumoylation.

2.2 RESULTS AND DISCUSSION

2.2.1 Synthesis of Homogeneous Sumoylated H4

To study the effect of sumoylation on chromatin structure, we needed access to large quantities of H4 modified by SUMO-3 at Lys12 in the histone tail. The first ever reported synthesis of a peptide modified by a large modification such as ubiquitin was presented by Chatterjee et al, where they made use of a photocleavable ligation auxiliary which was coupled to the side-chain of a target lysine in the peptide[13]. EPL between the peptide and a ubiquitin thioester followed by UV irradiation resulted in a site-specifically ubiquitylated peptide. Even though the ligation auxiliary resulted in a native isopeptide linkage, the complex synthetic route rendered this method intractable for widespread use. Later, the same group reported a facile and high yielding disulfide-directed ubiquitylation strategy which used a strategically introduced cysteine as the site of attachment. The disulfide analog of ubiquitylation of H2B at Lys120 ($uH2B_{ss}$) was shown to be a good mimic of native uH2B as it successfully stimulated methylation activity of hDot1 toward H3 in the context of a nucleosome[14]. We adopted a similar strategy to afford H4 protein homogeneously sumoylated at Lys12. Using site-directed mutagenesis we generated the H4 K12C mutant, which contained a unique cysteine residue at the site of SUMO attachment. The thiol group of the cysteine side chain was activated by reaction with 2,2'-dithiobis(5-nitropyridine) to generate **2** (Figure 2.1a). The full-length human SUMO-3 C47S mutant was expressed as a fusion with the GyrA intein and chitin-binding domain. Intein-mediated thiolysis was performed in the presence of a small molecule, 2-aminoethanethiol, to yield SUMO-3 C47S with a free C-terminal thiol **3** (Figure 2.1a). Finally, reaction of **2** and **3** at pH 7.0 led to the formation of the disulfide-linked analogue $suH4_{ss}$. Furthermore, the SUMO-3 C47S mutation, which was necessitated to ensure selective disulfide formation at the SUMO C-terminus, does not lead to observable structural changes in SUMO. Our semisynthetic strategy readily yielded milligram quantities of homogeneous $suH4_{ss}$.

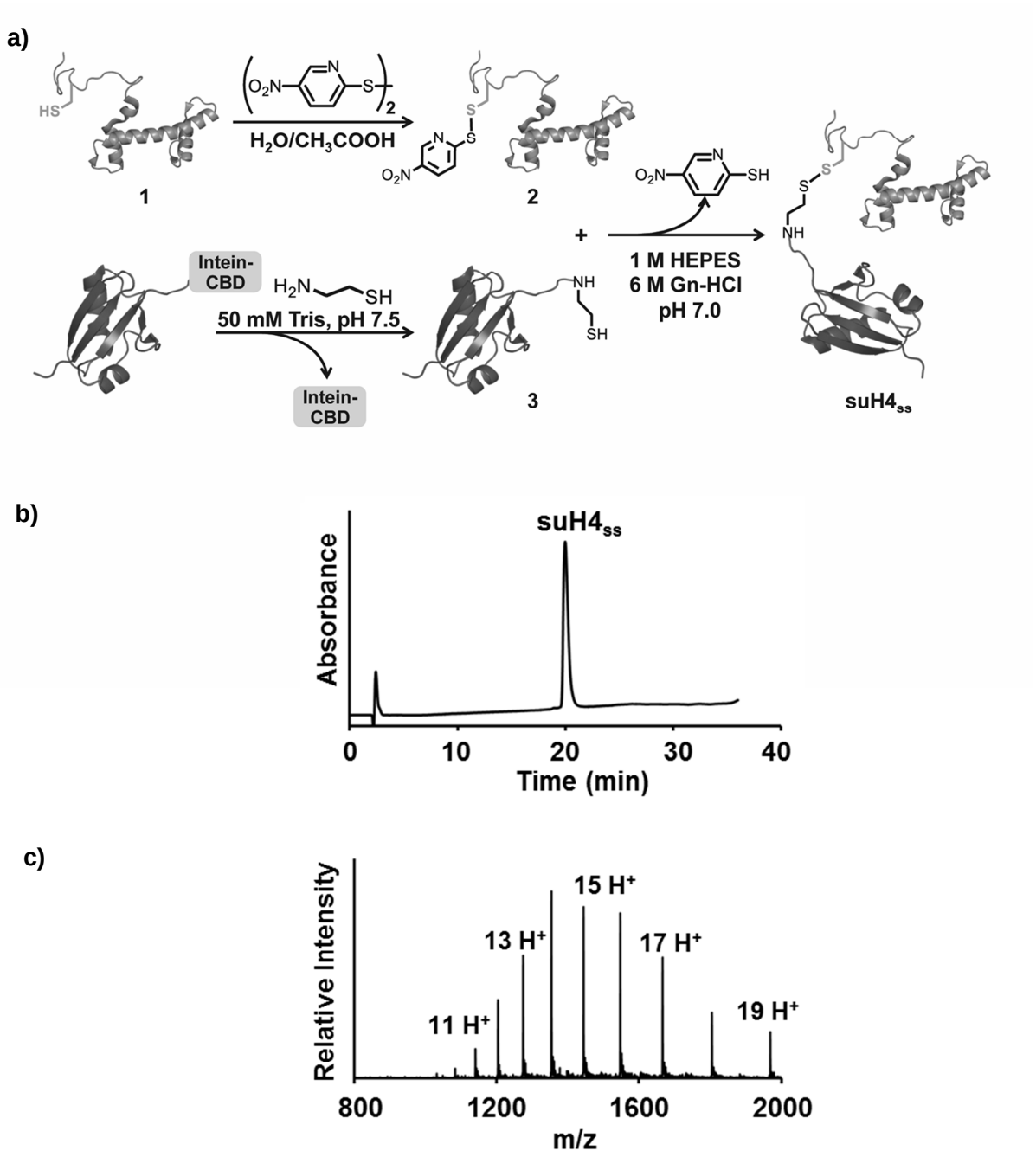


Figure 2.1. Disulfide directed site-specific modification of H4 by SUMO-3. a) Synthetic scheme depicting the activation of the H4 K12C mutant, **1**, with a small molecule to form a mixed disulfide, **2**. SUMO-3 C47S-aminoethanethiol, **3**, is generated using intein-mediated thiolysis. Subsequent reaction of **2** and **3** leads to the formation of $suH4_{ss}$. b) C4 analytical RP-HPLC chromatogram of pure $suH4_{ss}$. c) ESI-MS of pure $suH4_{ss}$. Calculated m/z for $[M+H]^+$ 21,646 Da, observed 21,646 ± 3 Da.

2.2.2 Synthesis of Thialysine Analog of Acetylated H4

In a seminal study by Shogren-Knaak et. al, nucleosomal arrays containing acetylated Lys16 in H4 were shown to inhibit chromatin compact and aggregation. The authors made use of the EPL methodology to generated a H4 K16ac by ligating an N-terminal H4 peptide (residues 1 to 22) containing Lys16ac with a recombinantly expressed H4Δ1-22 R23C mutant[9]. Even though the final product, containing a ligation scar at 23rd residue, was shown to be a good substitute for wild-type H4 K16ac, we decided to use a much simpler, scar free approach to generate H4 K16ac. By employing radical-initiated thiol-ene coupling chemistry between a H4 K16C mutant and N-vinyl acetamide, we generated a thialysine analog of H4 K16ac, referred to as H4 K_s16ac (Figure 2.2a)[15]. The thia-analog is indistinguishable from wild-type H4 K16ac in its ability to inhibit chromatin compaction mediated by Mg^{2+} and is successfully deacetylated by HDACs[15,16]. A major advance over previous uses of this methodology was our utilization of a thermal activation strategy, rather than UV light, to generate the thiyl radical. This prevented unwanted side-reactions resulting from extended UV exposure of protein and led to clean reaction mixtures that were purified to homogeneity by HPLC and characterized by ESI-MS (Figure 2.2b and 2.2c).

2.2.3 Reconstitution of Nucleosomes with suH4$_{ss}$, H4 K_s16ac or wt-H4

Wt-H4, H4 K_s16ac or suH4$_{ss}$ was incorporated into histone octamers by mixing with equimolar quantities of human histones H2A, H2B and H3 C110A. The H3 C110A mutant is known to have no observable effect on nucleosome structure or function, and was used to preclude undesired disulfide exchange of Cys110 with suH4$_{ss}$. The octamers were purified away from histone dimers and tetramers by size exclusion chromatography and their composition was confirmed by 15% SDS-PAGE and RP-HPLC (Figure 2.3a and 2.3b). Using equimolar ratio of the octamer and a 147 bp 601 strongly positioning double-stranded DNA (dsDNA) sequence reported by the

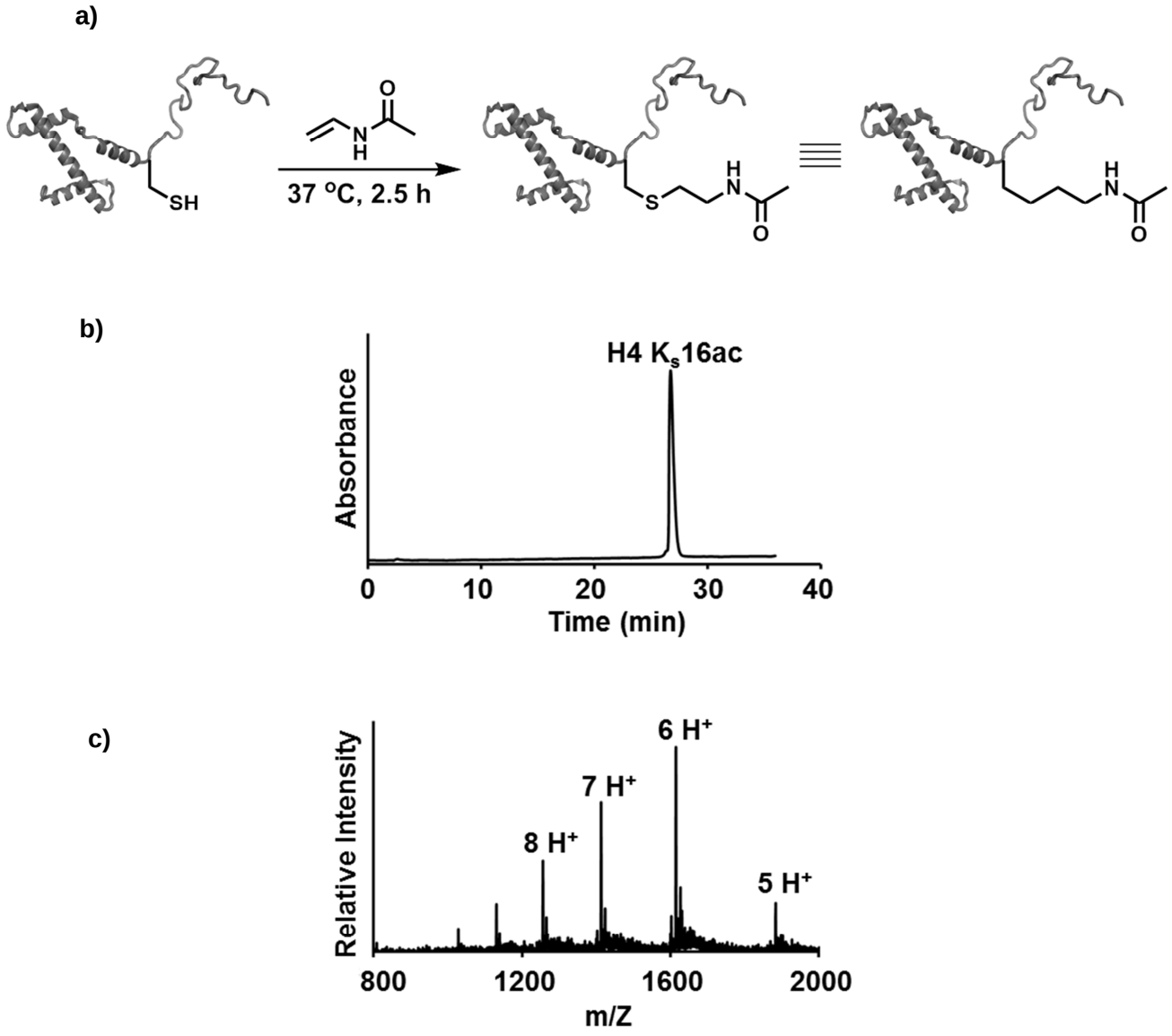


Figure 2.2. Synthesis of the thialysine analog of H4 K16ac. a) Synthetic scheme depicting the site-specific reaction of N-vinylacetamide with the cysteine side chain to generate H4 K_s16ac, a thia-analog of lysine acetylation. b) C4 analytical RP-HPLC chromatogram of pure H4 K_s16ac. c) ESI-MS of pure H4 K_s16ac. Calculated m/z for $[M+H]^+$ 11,296.1 Da, observed 11,295.6 ± 1 Da.

Widom and co-workers[17], we reconstituted wild-type, acetylated or sumoylated nucleosome core particles (NCPs) by means of step-wise salt dialysis. To our surprise, the sumoylated octamers formed a single predominant suMN species, indicating that octamer formation with pre-sumoylated H4 is possible *in vivo*. In the absence of reducing agents the disulfide linkage was very stable. The addition of 1 mM DTT led to identical migration of the sumoylated and wtMNs, thus confirming the presence of equal amounts all four histones in suMN (Figure 2.3c).

2.2.4 Assembly of Nucleosome Arrays with suH4$_{ss}$, H4 K$_s$16ac or wt-H4

To study the direct effects of sumoylation on chromatin compaction and aggregation, we assembled sumoylated nucleosomes on dsDNA containing 12 repeats of the strongly binding 601 sequence with a 30 bp linker between each nucleosome (12_177_601 sequence). Similar to our strategy for nucleosome formation, the arrays were assembled by mixing equimolar ratios of the 601 nucleosome forming sites in the 12_177_601 dsDNA with sumoylated octamers. Upon slow dilution of the initial 2 M NaCl buffer to 10 mM NaCl, array formation was observed with a single predominant species observed by 1% APAGE for each type of octamer (Figure 2.4a). The complete occupancy of all 12 octamer binding sites was confirmed by partial digestion with the non-specific endonuclease micrococcal nuclease (MNase). The observation of 12 bands upon brief MNase digestion suggested that all 12 601 DNA sites were occupied by histones which prevents the digestion of DNA at those places (Figure 2.4b). If the digestion is allowed to proceed for long, then as expected for a saturated array of nucleosomes, only ~150 bp DNA is observed. The complete occupancy of our arrays was further confirmed by negative stain electron microscopy in collaboration with Dr. Christopher Woodcock at the University of Massachusetts, Amherst (Figure 2.4c). To also confirm that each of the 12 positions contained all four histones, we performed specific digestion with the endonuclease ScaI which cuts within a unique AGTACT sequence incorporated in each linker region.

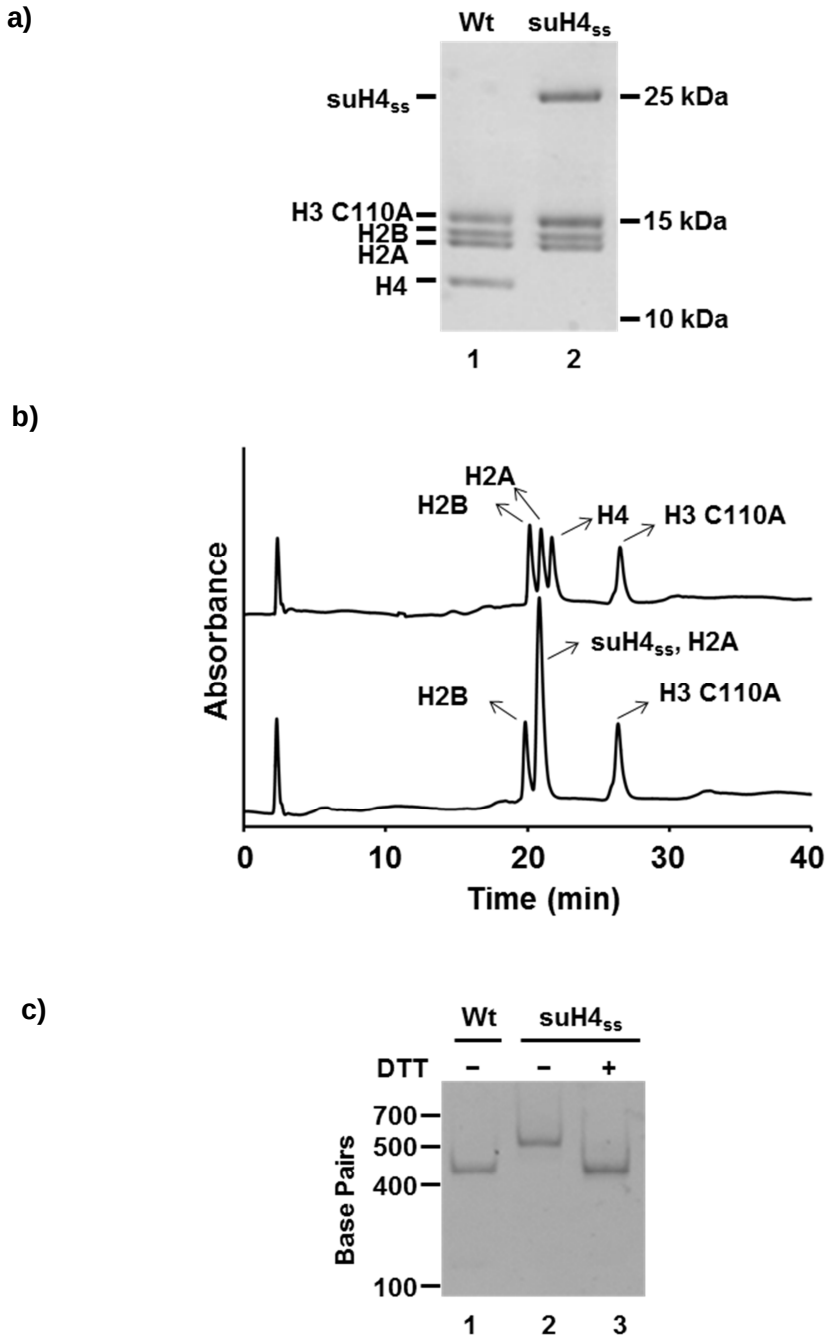


Figure 2.3. Characterization of unmodified and sumoylated octamers and nucleosomes. a) Coomassie-stained 15% SDS-PAGE of octamers containing wild-type (Wt) or sumoylated histone H4 ($suH4_{ss}$). b) C4 analytical RP-HPLC of wild-type (top) and sumoylated (bottom) octamers. c) Ethidium bromide stained 5% TBE gel of wild-type and $suH4_{ss}$ containing MNs. Lane 3 shows the result of reducing the disulfide-linkage in $suH4_{ss}$.

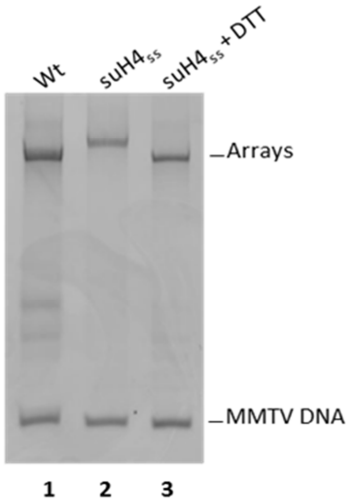
The mononucleosomes resulting from ScaI digestion were visualized by 5% TBE gel and were found to migrate identically to the corresponding NCP reconstituted as described in 2.2.3. Due to the presence of reducing agents in the ScaI digestion buffer, a band corresponding to partial reduction of the SUMO-3-H4 disulfide was also observed (Figure 2.4d). 12-mer arrays bearing either wild-type H4 or H4 K_s16ac that have been shown to compact maximally and minimally, respectively, at 1 mM Mg^{2+} concentrations, were also assembled as benchmarks for array compaction experiments. The quality of these arrays was assessed by subjecting them to MNase and ScaI digestion.

2.2.5 Sumoylation Inhibits Chromatin Compaction

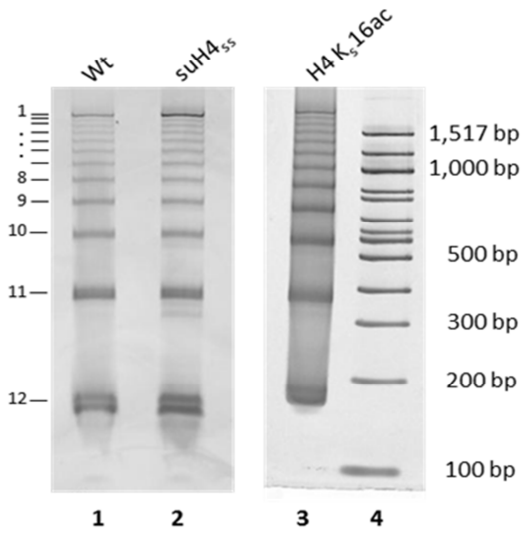
The majority of chromatin in the nucleus of mammalian cells is present in a compact and transcriptionally less active heterochromatic state. The presence of divalent cations such as Mg^{2+}, which neutralize the negative charge of the DNA backbone and the presence of chromatin effectors such as the linker histone H1, is crucial for maintaining the compacted state inside the nucleus. Compaction can be recapitulated *in vitro* in the absence of chromatin binders by the addition of 1 mM $MgCl_2$. The Mg^{2+} cations neutralize the electrostatic repulsion between the negatively charged phosphate groups of the DNA backbone. This allows the positively charged histone tails, especially the H4 tail, to interact with negatively charged regions of the octameric core of the neighboring nucleosome and brings adjacent nucleosomes closer to facilitate compaction. This transformation of the decompacted arrays to the compact form upon addition of 1 mM $MgCl_2$ can be assessed by subjecting the 12-mer nucleosome arrays to analytical ultracentrifugation (AUC).

AUC is a biophysical technique that reports on the size and shape of a macromolecule. Nucleosome arrays were suspended in an aqueous buffer and subject to ultrahigh centrifugal

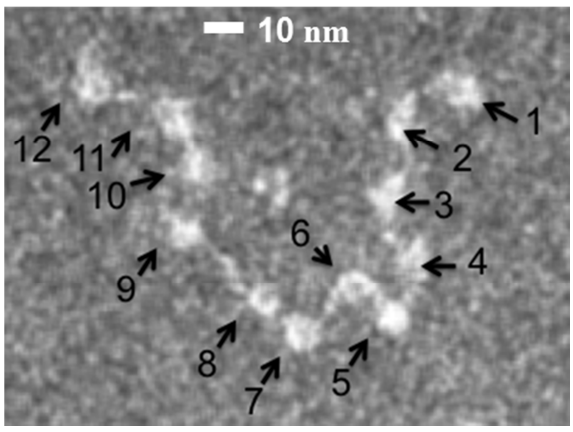
a)


Wt
suH4ss
suH4ss +DTT
Arrays
MMTV DNA
1
2
3

b)


Wt
suH4ss
H4 Ks16ac
1
8
9
10
11
12
1,517 bp
1,000 bp
500 bp
300 bp
200 bp
100 bp
1
2
3
4

c)


10 nm
1
2
3
4
5
6
7
8
9
10
11
12

d)

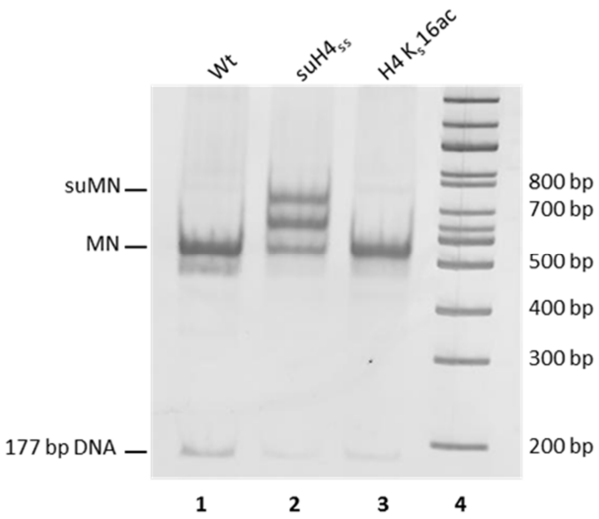


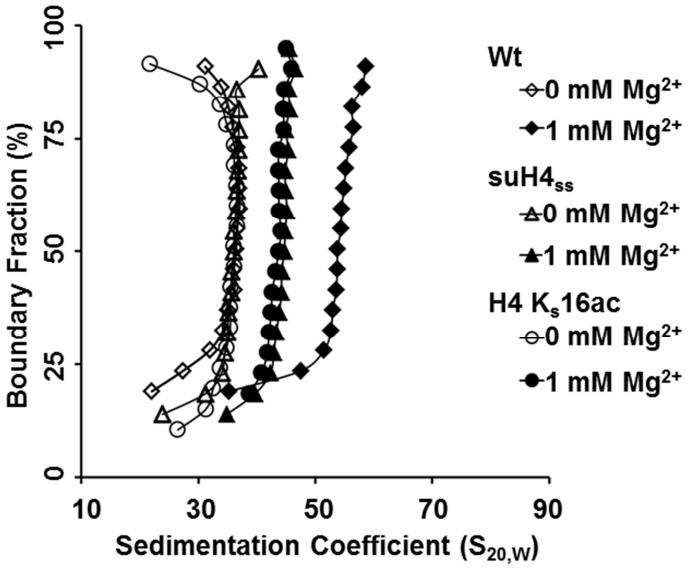
Figure 2.4. Characterization of unmodified, sumoylated and acetylated nucleosomal arrays. a) Ethidium bromide stained 1% APAGE gel showing the reduction of sumoylated arrays in the presence of 1 mM DTT (Lane 3). b) Ethidium bromide stained 5% TBE gel showing micrococcal nuclease digestion of wild-type, sumoylated and acetylated arrays. c) Electron micrograph showing full occupancy of sumoylated arrays assembled on a 12_177_601 DNA substrate (Mean occupancy= 11.7 ± 0.56, n=17, and the error is standard deviation). d) Ethidium bromide stained 5% TBE gel showing ScaI digestion products of wild-type and modified 12-mer arrays. Both partial and complete reduction of the disulfide-linkage in $suH4_{ss}$ can be observed in lane 2 due to the presence of thiols in the digestion buffer.

force in an AUC experiment. Particles moving through a viscous fluid experience a frictional drag that is proportional to the velocity and the hydrodynamic radius of the macromolecule. Bulky or elongated particles experience more frictional drag than compact, smooth spherical ones. If two particles have the same effective molar weight, the particle that sediments faster has a lower hydrodynamic radius, hence is more compact. Thus the sedimentation coefficient (s), which has dimensions of seconds, is inversely proportional to the frictional coefficient of a particle. For many substances, the value of s lies between 1 and 100 × 10^{-13} seconds. The Svedberg unit (abbreviation S or Sv) is defined as 10^{-13} seconds. Sedimentation coefficient corrected for the viscosity and density of the solvent, relative to that of water at 20 °C is represented as $S_{20,w}$, this allows for the uniform comparison of sedimentation coefficients when different buffers are used.

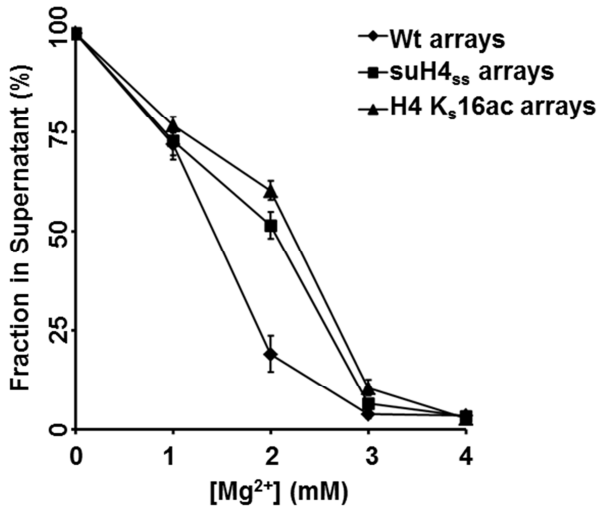
Sumoylated (containing $suH4_{ss}$), acetylated (containing H4 K_s16ac) and unmodified (containing wt-H4) arrays were subject to AUC to assess the effect of sumoylation on chromatin compaction. In the absence of any divalent cations, all three types of arrays showed similar sedimentation coefficients, suggesting that they have similar structures before compaction has been initiated. Upon addition of 1 mM $MgCl_2$, which is known to induce compaction *in vitro*, we observed maximal compaction for unmodified arrays ($\Delta S_{20,w}$ = 16.9 ± 0.9 S). Surprisingly, sumoylation inhibited chromatin compaction similar to acetylation of the H4 tail. Sedimentation velocity analysis showed that both sumoylation ($\Delta S_{20,w}$ = 9.5 ± 1.0 S) and acetylation ($\Delta S_{20,w}$ = 10.0 ± 1.0 S) result in an open form for chromatin even in the presence of divalent cations (Figure 2.5a).

To confirm that the reduction in compaction observed for sumoylated arrays is due to the presence of SUMO-3 on the H4 tail, we made use of the reversible nature of our disulfide-directed sumoylation strategy. We reduced the disulfide bond by adding 1 mM DTT to the sumoylated samples, and performed AUC experiment with these arrays. We observed maximal

a)


Boundary Fraction (%)
0
25
50
75
100
10
30
50
70
90
Sedimentation Coefficient ($S_{20,W}$)
Wt
0 mM Mg^{2+}
1 mM Mg^{2+}
$suH4_{ss}$
0 mM Mg^{2+}
1 mM Mg^{2+}
H4 K_s16ac
0 mM Mg^{2+}
1 mM Mg^{2+}

b)


Fraction in Supernatant (%)
0
25
50
75
100
0
1
2
3
4
[Mg^{2+}] (mM)
Wt arrays
$suH4_{ss}$ arrays
H4 K_s16ac arrays

c)

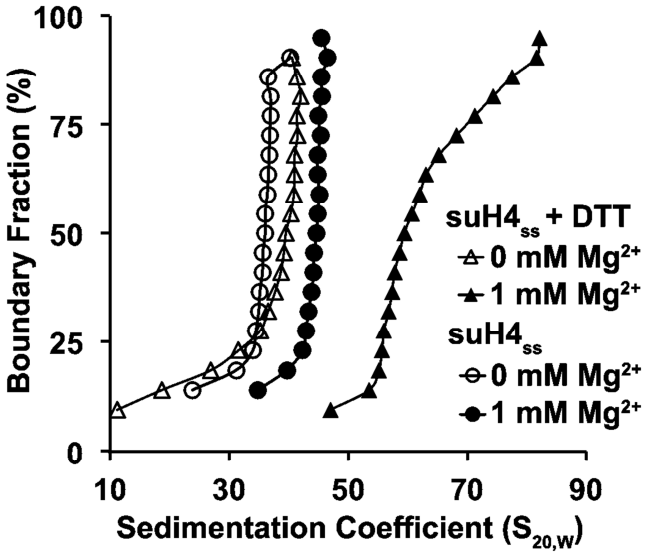


Figure 2.5. Determination of chromatin compaction by analytical ultracentrifugation experiments with sumoylated ($suH4_{ss}$) and non-sumoylated arrays. a) Sedimentation coefficient distributions of nucleosomal arrays containing unmodified histone H4, H4 K_s16ac, or suH4ss were calculated by van Holde-Weischet analysis at 0 and 1 mM Mg^{2+} (open and solid symbols, respectively). b) Oligomerization behavior of nucleosomal arrays containing unmodified histone H4, H4 K_s16ac, or $suH4_{ss}$ was assessed by determining Mg^{2+}-dependent solubility. n = 14–16 and error bars show s.e.m. c) Sedimentation coefficient distributions of arrays containing $suH4_{ss}$ or where SUMO-3 was removed by reduction with dithiothreitol (DTT) were calculated by van Holde-Weischet analysis at 0 and 1 mM Mg^{2+} (open and solid symbols, respectively).

compaction by the reduced arrays similar to unmodified arrays, which confirmed that the inhibition of compaction is indeed due to the presence of SUMO-3 on the H4 tail (Figure 2.5c)

2.2.6 Sumoylation Inhibits Chromatin Aggregation

The compacted state of nucleosome arrays at 1 mM $MgCl_2$ is thought to mimic the 30 nm chromatin fibers proposed to occur *in vivo*, but the majority of the chromatin is present in a further compacted and aggregated form. This higher order of chromatin compaction can be

induced *in vitro* by increasing the concentration of divalent cations in the solution containing soluble nucleosome arrays. In the absence of NCP, no aggregation is observed for DNA even in the presence of high concentration of divalent cations. Thus, aggregation is promoted by the attraction between NCP and limited by the repulsion between linker DNA. Indeed trypsinizing the histone tails[18,19], which contain a majority of the positively charged amino acid residues of the NCP, lowers this attractive interaction and the aggregation of nucleosomal arrays is suppressed. Similarly, the presence of various modifications on the histone tails can affect the degree of aggregation and higher order chromatin structure formation. We assessed the ability of sumoylation to inhibit or allow chromatin aggregation by performing chromatin aggregation assays in increasing amounts of $MgCl_2$. Again, we made use of unmofdified and acetylated arrays as benchmarks of maximal and minimal aggregation, respectively. Once again, sumoylation behaved similar to H4 K16ac and prevented chromatin aggregation to a similar degree as acetylation. At 2 mM $MgCl_2$, where ~80% of the unmodified arrays had aggregated and hence precipitated out of the solution, more than 50% of sumoylated and acetylated arrays still remained soluble. Thus, sumoylation has an inhibitory effect on inter-array and intra-array interactions (Figure 2.5b).

2.2.6 Degree of Sumoylation can Modulate Chromatin Compaction

Having discovered the *cis* effects of uniform sumoylation on chromatin structure, we wondered if cells may use this modification to modulate chromatin compaction by varying the degree of sumoylation on nucleosomes. As asymmetrically modified mononucleosomes have been observed *in vivo*[20], it is also possible that non-uniform and sub-stoichiometric sumoylation exists inside the nucleus to maintain local chromatin structure. We interrogated the effect of asymmetric sumoylation by assembling nucleosomal arrays containing a 1:1 ratio of $suH4_{ss}$ to wt-H4. These 50% sumoylated arrays, containing a stochastic distribution of suH4ss, were

subject to the same biophysical analyses as the uniformly sumoylated arrays (Figure 2.6a). Contrary to uniform 100% sumoylation, 50% sumoylation ($\Delta S_{20,w}$ = 18.4 ± 0.6 S) allowed chromatin compaction identical to unmodified arrays ($\Delta S_{20,w}$ = 16.9 ± 0.9 S). Surprisingly, 50% sumoylation did not allow the aggregation of chromatin to the same degree as unmodified chromatin arrays. The aggregation observed for 50% sumoylated arrays was in-between that observed for 100% sumoylated arrays and unmodified arrays. We hypothesize that in order to transition from a heterochromatic to a euchromatic state the cell may employ sumoylating enzymes to install SUMO on the few exposed H4 tails to incrementally open up the aggregated state of chromatin. As aggregation is impaired upon partial sumoylation, more histone tails are exposed and the sumoylating enzymes can install additional SUMO on nucleosomes. Upon reaching a state of uniform 100% sumoylation, the formation of 30 nm fibers is disrupted, as observed through our biophysical studies, and a euchromatic state is established.

2.2.7 Single Molecule Studies of Sumoylated Nucleosomes

In mammalian cells, the length of the linker DNA varies anywhere from 10 bp to 90 bp, and this variation in length has been shown to affect the structure of 30 nm fibers formed upon compaction[21]. It is possible that the inhibitory effect of uniform sumoylation discovered in our biophysical analyses is specific to the 30 bp linker length present in our nucleosome arrays. Both, AUC and chromatin aggregation analysis report on an ensemble of nucleosomal arrays and do not provide information about the biophysical property of each individual sumoylated array. We sought to address whether the observed effect of $suH4_{ss}$ on inter-nucleosomal interactions is independent of linker length by employing a single-molecule Förster resonance energy transfer (FRET) assay. In collaboration with Dr. Tae-Hee Lee's lab at the Pennsylvania State University, we reconstituted two sets of mononucleosomes, one with a FRET donor and another with an acceptor, by attaching appropriate fluorophores to the 147 bp 601 Widom DNA.

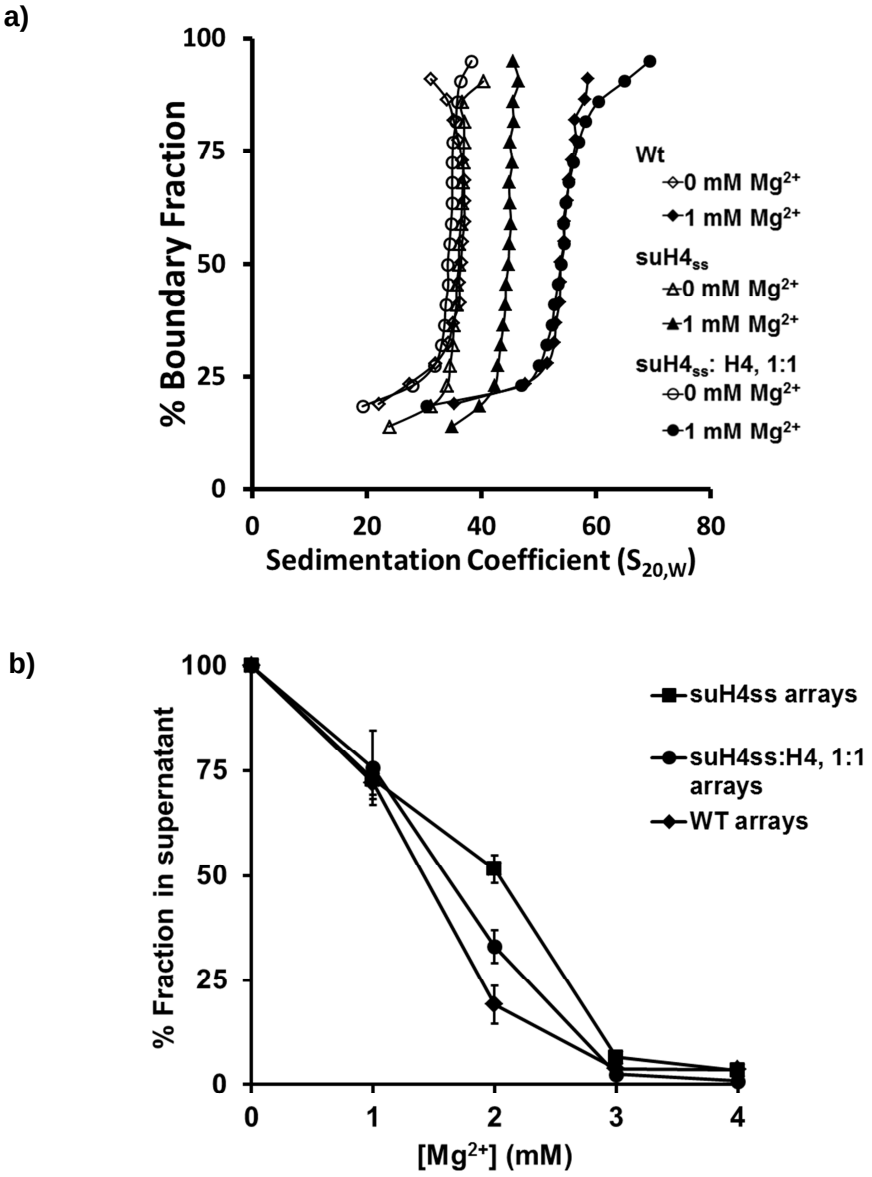
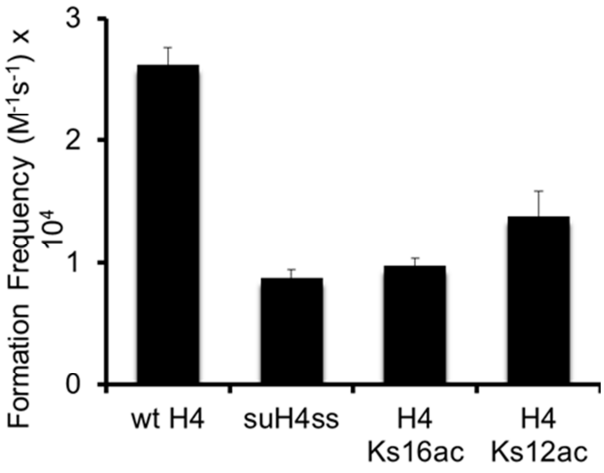


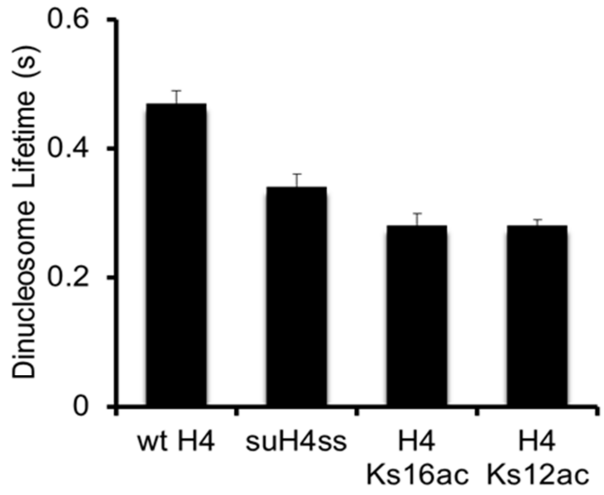
Figure 2.6. Determination of chromatin aggregation and compaction by analytical ultracentrifugation experiments with unmodified (Wt), sumoylated ($suH4_{ss}$) and 50% sumoylated ($suH4_{ss}$:H4, 1:1) arrays. a) Sedimentation coefficient distributions of nucleosomal arrays containing unmodified histone H4, suH4ss or 1:1 $suH4_{ss}$:H4 were calculated by van Holde-Weischet analysis at 0 and 1 mM Mg^{2+} (open and solid symbols, respectively). b) Oligomerization behavior of nucleosomal arrays containing unmodified histone H4, $suH4_{ss}$ or 1:1 $suH4_{ss}$:H4 was assessed by determining Mg^{2+}-dependent solubility. n = 10–16 and error bars show s.e.m.

One 147 bp sequence was labeled with a Cy3 dye at the 129th base (601A) on both strands and the other was labeled with a Cy5 dye at the 60th base (601B) on both strands. The Cy3-labeled MNs were then immobilized on a streptavidin-coated microscope slide and a dilute solution of Cy5-labeled MNs was introduced onto the slide. For each type of modification, we measured the frequency and the lifetime of FRET events between the two sets of fluorescently labeled MNs in the presence of 3 mM Mg^{2+}. The frequency of dinucleosome formation was converted to a rate constant to reflect the effect of a particular modification on the propensity of two NCPs to form a dinucleosome (depicted as $k_{forward}$). We observed that wild-type, suH4$_{ss}$, H4 K$_s$16ac, H4 K$_s$12ac MNs formed compact dinucleosomes 26,200, 8,740, 9,710, and 1,3800 times per second per mole, respectively (Figure 2.7a). Sumoylation at Lys12 of H4, which is a proposed repressive mark, resulted in lowering the frequency of dinucleosome formation to the same extent as acetylation at Lys16 of H4, which is an activating mark. Whereas unmodified nucleosomes formed dinucleosome two times more frequently. Next, we measured the lifetime of FRET observed once the dinucleosomes were formed to measure the stability of this state arising from these different modifications (Figure 2.7b). The longer the lifetime of FRET observed for a particular modification, the less is the propensity of the dinucleosome to fall apart once it is formed (depicted as $1/k_{reverse}$). By combining these two measurements we calculated the free energy change during dinucleosome stack formation ($\Delta G = -RT ln(k_{forward}/k_{reverse})$. The free energy changes of stacking for wild-type, suH4$_{ss}$, H4 K$_s$16ac, H4 K$_s$12ac MNs are -23.3 ± 0.2, -19.8 ± 0.2, -19.5 ± 0.2, and -20.5 ± 0.4 kJ/mol, respectively (Figure 2.7c). Again, just like the AUC based ensemble measurements, ΔG values based on single-molecule FRET analysis showed that sumoylation at Lys12 of H4 inhibits inter-nucleosome interaction identical to acetylation at Lys16 of H4. To understand how two modifications that are chemically dissimilar can have identical inhibitory effect on chromatin compaction, we further investigated the precise mechanism by which SUMO prevents compaction.

a)


Formation Frequency ($M^{-1}s^{-1}$) x 10^4
3
2
1
0
wt H4
suH4ss
H4 Ks16ac
H4 Ks12ac

b)


Dinucleosome Lifetime (s)
0.6
0.4
0.2
0
wt H4
suH4ss
H4 Ks16ac
H4 Ks12ac

c)

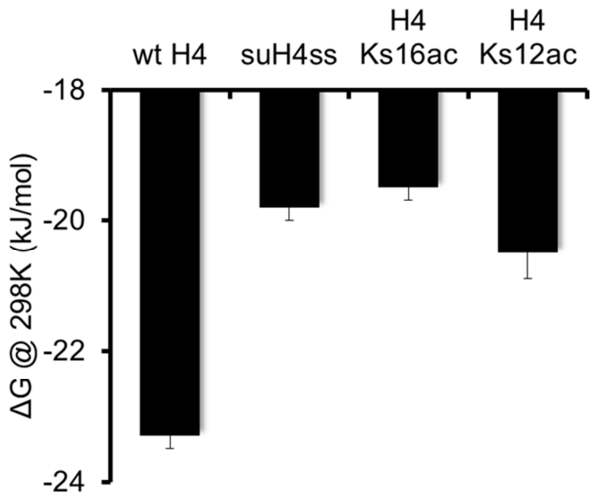


Figure 2.7. Single molecule FRET measurements of dinucleosome formation with unmodified (wt H4), sumoylated ($suH4_{ss}$) and two different acetylated (H4 K_s16ac, H4 K_s12ac) nucleosomes. a) Frequency of dinucleosome formation in the presence of various modifications on the H4 tail. b) Lifetime of a dinucleosome in the presence of various modifications on the H4 tail. c) Change in Gibbs free energy at standard state upon dinucleosome formation in the presence of various modifications on the H4 tail.

2.2.7 Mechanistic Studies of SUMO's Inhibitory Effect

The primary mechanism by which a small chemical modification such as an acetyl group exerts a *cis* effect on chromatin structure is by neutralizing the positive charge on the lysine side-chain at the site of acetylation. This impairs the electrostatic interaction between the positively charged histone tail with DNA and with negatively charged regions in the octameric core of the adjacent nucleosomes. In case of modification by a larger group, such as the ~11 kDa SUMO-3 protein, many additional pathways may be involved in inhibiting the compaction of chromatin. Attachment of SUMO to H4 neutralizes the positive charge of the lysine residue, but then additionally brings its own charged surfaces which can have electrostatic interactions with the adjacent nucleosomes. SUMO is a bulky modification and can also act as a steric wedge between neighboring MNs and hence prevent compaction. A similar mechanism was proposed when inhibition of chromatin compaction was observed upon ubiquitylation of Lys120 of H2B[22]. Even though ubiquitin and SUMO-3 share only ~18% sequence identity, they have similar steric bulk due to very similar tertiary structures. In order to assess whether SUMO also prevents compaction by being a steric wedge, we made use of site-specific disulfide formation strategy to attach ubiquitin (Ub) at Lys12 of H4 ($uH4_{ss}$). In order to understand whether the inhibition of chromatin compaction observed by us is a result of charge neutralization at SUMO's site of attachment, we made use of thiol-ene click chemistry to install acetylation at Lys12 of H4 (H4 K_s12ac), which results in charge neutralization at that site. We incorporated $uH4_{ss}$ and H4 K_s12ac into nucleosome arrays and performed chromatin compaction and aggregation analyses on them. We observed that charge neutralization plays an important part in inhibiting chromatin compaction but its effect varies with the position of the modified lysine. Acetylation at Lys12 ($\Delta S_{20,w}$ = 11.8 ± 0.4 S) prevented chromatin compaction, but to a lesser degree than sumoylation at Lys12 ($\Delta S_{20,w}$ = 9.5 ± 1.0 S) or acetylation at Lys16 of H4 ($\Delta S_{20,w}$ = 10.0 ± 1.0 S) (Figure 2.8a).

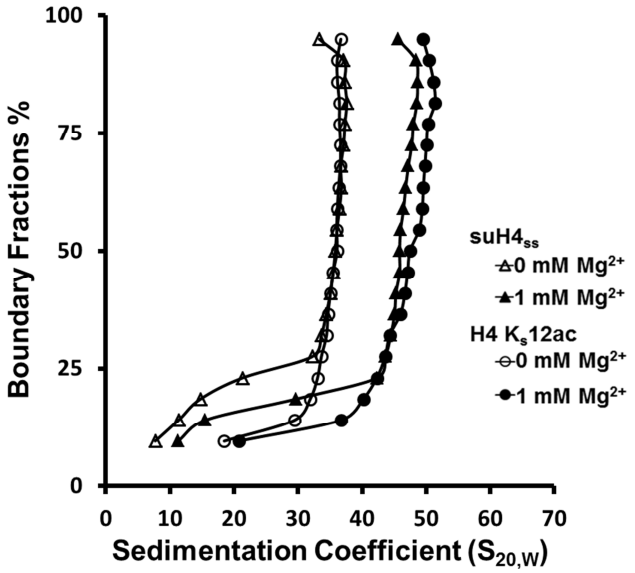
This suggests that array decompaction by sumoylation is not entirely due to charge neutralization at the site of attachment. The steric bulk and electrostatic interactions by SUMO-3 play additional roles. Ubiquitylation ($\Delta S_{20,w}$ = 8.7 ± 1.4 S for $uH4_{ss}$) inhibited chromatin compaction in an identical manner to sumoylation ($\Delta S_{20,w}$ = 9.5 ± 1.0 S for $suH4_{ss}$) indicating that SUMO may also act as a steric wedge between adjacent MNs and inhibit compaction of chromatin (Figure 2.8b). A similar trend was observed for chromatin aggregation where ubiquitylation and sumoylation showed identical aggregation behavior whereas H4 K_s12ac aggregated to a lesser degree (Figure 2.8c).

2.3 CONCLUSION AND OUTLOOK

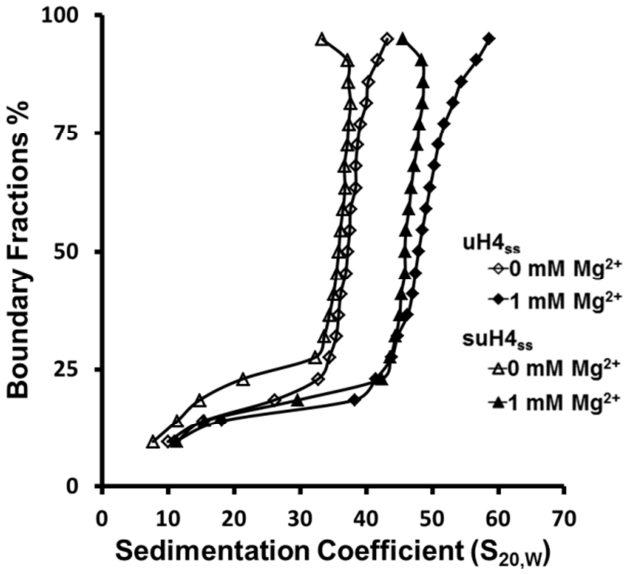
The rapidly expanding repertoire of chemical biology tools has vastly increased our accessibility to various analogs of modifications observed in histone proteins in cells[23]. We made use of few of these methodologies to generate homogeneous acetylated, ubiquitylated and sumoylated histone H4. This allowed us to understand the site-specific effect of these modifications on chromatin structure. We incorporated the modified H4 into chromatin arrays consisting of 12 nucleosomes on a long linear dsDNA, and performed various ensemble averaged and single-molecule biophysical measurements on them.

Acetylation of histones has long been associated with euchromatin and active transcription. Using our thermally initiated thiolene click-chemistry methodology, we generated site-specifically acetylated H4 and showed that the effect of acetylation on chromatin compaction can vary with the position of this modification on the H4 tail[24]. While acetylation at Lys16 and Lys12 both inhibit compaction, the degree of inhibition is greater when acetylation is present on Lys16.

a)


Boundary Fractions %
100
75
50
25
0
0
10
20
30
40
50
60
70
Sedimentation Coefficient ($S_{20,W}$)
suH4ss
0 mM Mg2+
1 mM Mg2+
H4 Ks12ac
0 mM Mg2+
1 mM Mg2+

b)


Boundary Fractions %
100
75
50
25
0
0
10
20
30
40
50
60
70
Sedimentation Coefficient ($S_{20,W}$)
uH4ss
0 mM Mg2+
1 mM Mg2+
suH4ss
0 mM Mg2+
1 mM Mg2+

c)

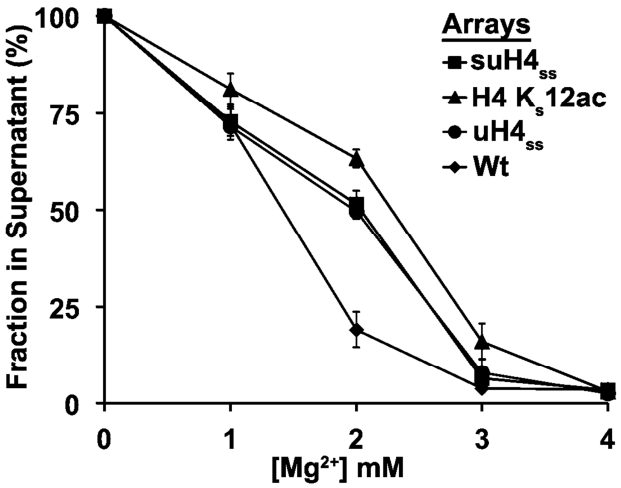


Figure 2.8. Mechanistic investigation into the inhibitory effect of H4 sumoylation on chromatin oligomerization by array precipitation experiments with unmodified (Wt), sumoylated ($suH4_{ss}$), ubiquitylated ($uH4_{ss}$) and acetylated (H4 K_s12ac) arrays. a) Sedimentation coefficient distributions of nucleosomal arrays containing suH4ss or H4 K_s12ac were calculated by van Holde-Weischet analysis at 0 and 1 mM Mg^{2+} (open and solid symbols, respectively). b) Sedimentation coefficient distributions of nucleosomal arrays containing suH4ss or $uH4_{ss}$ were calculated by van Holde-Weischet analysis at 0 and 1 mM Mg^{2+} (open and solid symbols, respectively). c) Oligomerization behavior of nucleosomal arrays containing unmodified histone H4, $suH4_{ss}$, $uH4_{ss}$ or H4 K_s12ac was assessed by determining Mg^{2+}-dependent solubility. n = 10–16 and error bars show s.e.m.

We also discovered that sumoylation, even though linked to gene repression, results in a decompacted form of chromatin similar to acetylation. Upon interrogating the mechanism of this effect, which involved mimicking the steric bulk of SUMO through the attachment of ubiquitin to the H4 tail, we discovered that the steric bulk of SUMO plays an important part in inhibiting compaction. We then corroborated these results through single-molecule FRET measurements and also discovered that it's the intrinsic property of sumoylation to reduce the rate of dinucleosome formation.

Further investigations have to be undertaken to understand the positional variance in the inhibitory effect of SUMO. The H4 tail contains neighboring lysine residues at positions 5 and 12 which may also be sumoylated *in vivo*. It would be interesting to investigate if sumoylation shows the same mechanistic plasticity as was observed for ubiquitylation of H2B[14]. Histones have also been found to be modified by other isoforms of SUMO such as SUMO-1, which has very low sequence identity with SUMO-3. It would be interesting to see if there are any isoform-specific effects of sumoylation on chromatin structure. These isoforms have been shown to have overlapping targets which does not always accompany overlapping function[25]. Other than histone H4, sumoylation has also been discovered on other canonical histones, such as H3. Through our disulfide-directed protein modification strategy, we can easily access sumoylated H3 and explore the roles of sumoylation in altering chromatin structure and function.

It is also important to reconcile our *in vitro* observations with the biological roles of site-specifically sumoylated H4. To this end we are also actively pursuing the generation of monoclonal antibodies in mice which can distinguish between the SUMO-1 and SUMO-2/3 and are specific for sumoylation of H4 at Lys12. It would be interesting to investigate the spatial and temporal aspects of this modification in cells and also to probe its binding partners. We anticipate these future studies to shed light on the various roles played by sumoylation in important cellular processes.

2.4 EXPERIMENTAL PROCEDURES

2.4.1 General Laboratory Methods

All commonly used chemical reagents and solvents were purchased from Sigma-Aldrich Chemical Company (Milwaukee, WI) or Fischer Scientific (Pittsburgh, PA). A Superdex S-200 10/300 GL size-exclusion column was obtained from GE Healthcare (Waukesha, WI). Chemically competent DH5α and BL21(DE3) cells were purchased from Novagen (Madison, WI). T4 DNA ligase and restriction enzymes were purchased from New England BioLabs (Ipswitch, MA). Primer synthesis and gene sequencing were performed by Integrated DNA Technologies (Coralville, IA) and Genewiz (South Plainfeld, NJ), respectively. Site-directed mutagenesis was performed with a QuikChange Site-Directed Mutagenesis kit from Agilent Technologies (Santa Clara, CA). Criterion 5% TBE gels were purchased from Bio-Rad (Hercules, CA). Centrifugal filtration units were from Sartorius (Goettingen, Germany) and Slide-A-Lyzer dialysis cassettes and MINI dialysis units were from Pierce (Rockford, IL). Qiafilter Plasmid Giga, PCR purification and gel extraction kits were purchased from Qiagen (Valencia, CA). Micrococcal nuclease was purchased from Sigma-Aldrich Chemical Company (Milwaukee, WI). Size-exclusion chromatography was performed on an AKTA FPLC system (GE Healthcare) equipped with a P-920 pump and UPC-900 monitor. Analytical reversed-phase HPLC (RP-HPLC) was performed on a Varian ProStar instrument with Vydac C18 or C4 columns (5 micron, 4 x 150 mm), employing 0.1% TFA in water (HPLC buffer A), and 90% acetonitrile, 0.1% TFA in water (HPLC buffer B) as the mobile phases. Typical analytical gradients were 30-70% buffer B over 30 min at a flow rate of 1 mL/min. Preparative scale purifications were conducted on Vydac C18 or C4 preparative columns (15-20 micron, 50 x 250 mm) over 60 min at a flow rate of 9 mL/min. Electrospray ionization mass spectrometry (ESI-MS) analysis was performed on a Bruker Esquire LC-ion trap spectrometer (Bruker Daltonics, Billerica,

Massachusetts). All protein starting materials and ligation products were analyzed by both C18 or C4 analytical RP-HPLC and ESI-MS. The list of primers used in this study is as follows:

DNA Sequence (5'-to-3')	Primer
AGCAAGCTGATGAAGGCCTACTCTGAGAGGCAGGGCTTGTCAATG	hS3-C47S-FP
CATTGACAAGCCCTGCCTCTCAGAGTAGGCCTTCATCAGCTTGCT	hS3-C47S-RP
GGGAATTC*CATATG*GAAAACCTGTACTTCCAGTCTGGTCGTGGTAAAGGTGG (*NdeI*)	hH4-FP
GCCCGC*GGATCC*TCAACCACCGAAACCGTACAGGGTACGACCC (*BamHI*)	hH4-RP
GGTAAAGGTGGTAAAGGTCTGGGTTGCGGTGGTGCTAAACGTCACCGTAAA	hH4 K12C-FP
TTTACGGTGACGTTTAGCACCACCGCAACCCAGACCTTTACCACCTTTACC	hH4 K12C-RP
GGTGGTAAAGGTCTGGGTAAAGGTGGTGCTTGCCGTCACCGTAAAGTTCTG	hH4 K16C-FP
CAGAACTTTACGGTGACGGCAAGCACCACCTTTACCCAGACCTTTACCACC	hH4 K16C-RP

2.4.2 Cloning, Overexpression and Purification of hH2A 2-A (gene *HIST2H2AA3*), hH2B (gene *HIST1H2BK*), hH3 C110A (gene *HIST2H3C*), hH4 (gene *HIST1H4c*)

Full-length human histone genes in the pET3a vector were a generous gift from Dr. Peter Moyle at the University of Queensland, Australia. Plasmids containing full-length human histone genes were used to transform *E. coli* BL21(DE3) cells. 1.5 mL Eppendorf tube containing 50 µL of competent *E. coli* BL21(DE3) cells were thawed on ice and 50 ng of circular plasmid was introduced into the tube. The mixture was incubated on ice for 30 min at the end of which the cells were subjected to heat shock by incubating in a water bath at 42 ˚C for 45 seconds. The tube was further incubated on ice for 2 minutes to prevent damage to the *E.coli* cells. After addition of 1 ml of autoclaved luria broth (LB), which is free from antibiotics and has been pre-incubated at 37 ˚C, the tube was further incubated at 37 ˚C and shaken at 250 rpm for 1 hour. 300 µL of the suspended *E. coli* were plated onto a 1.5% LB-Agar plate and incubated overnight at 37 ˚C. Colonies were picked from the plate and grown overnight in 5 mL LB containing 100 µg/mL ampicillin at 37 ˚C. Cell stocks were prepared in the morning by mixing 800 µL of the overnight solution with 300 µL of sterile 80% glycerol stock solution. The mixture was flash frozen using liquid nitrogen and stored at -80 ˚C. 25 mL LB cultures were started using the cell

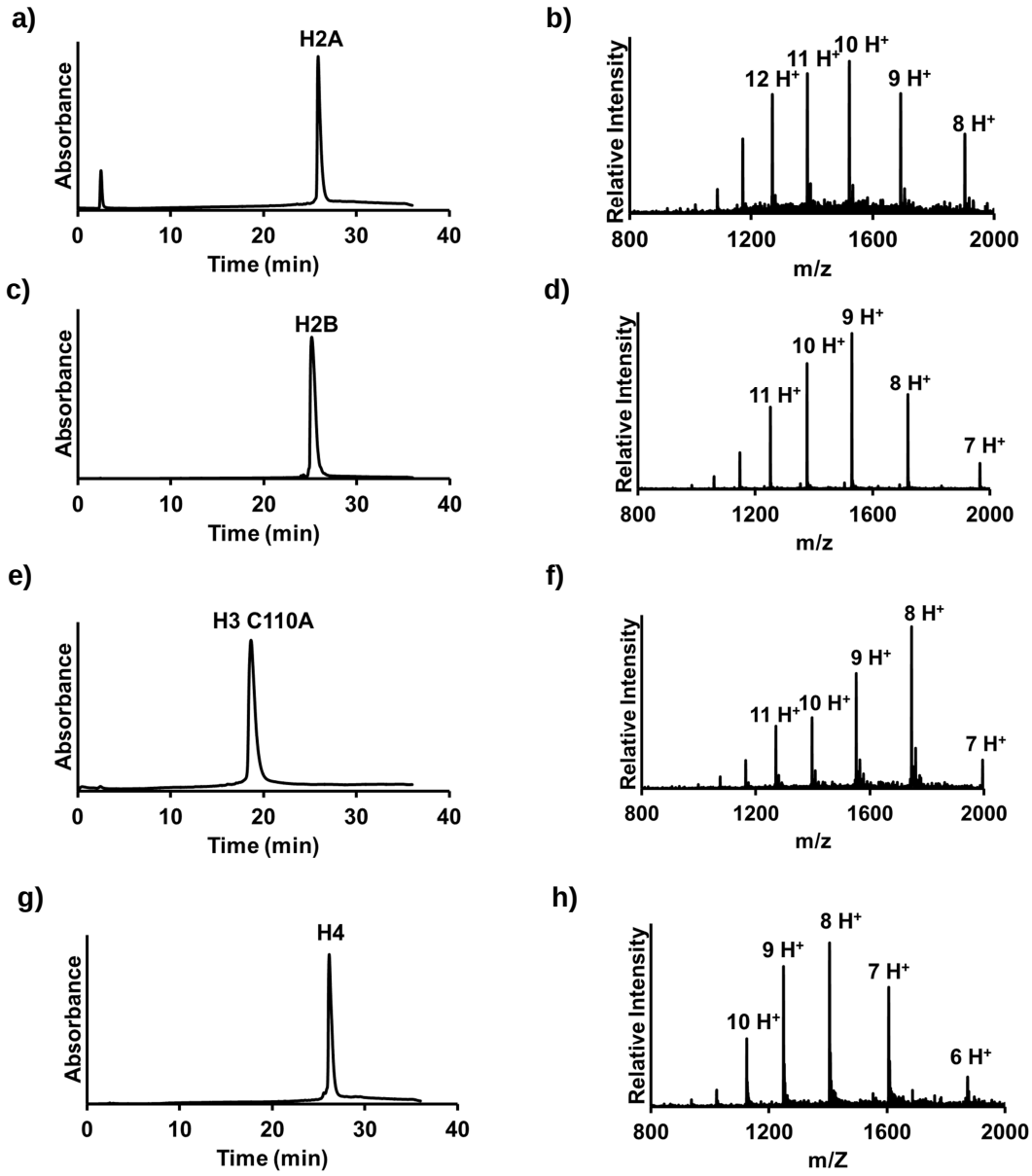


Figure 2.4.1. Characterization of recombinant human histones. a) RP-HPLC trace of histone H2A (isoform H2A 2-A) on a gradient of 0-73% buffer B. b) ESI-MS of H2A, calculated m/z $[M+H]^+$ 13,964.5 Da, observed 13,964.9±1.5 Da. c) RP-HPLC trace of histone H2B (isoform H2B 1-K) on a gradient of 0-73% buffer B. d) ESI-MS of H2B, calculated m/z $[M+H]^+$ 13,759.5 Da, observed 13,758.2±1.3 Da. e) RP-HPLC trace of histone H3 C110A (isoform H3.2) on a gradient of 30-90% buffer B. f) ESI-MS of H3, calculated m/z $[M+H]^+$ 15,225.1 Da, observed 15,224.7±2.9 Da. g) RP-HPLC trace of histone 4 (isoform H4 g) on a gradient of 0-73% buffer B. h) ESI-MS of histone H4, calculated m/z $[M+H]^+$ 11,237.0 Da, observed 11,237.4±1.8 Da.

Stocks and used 1% inoculation to grow the cells in 6 L of YT medium (16 g tryptone, 10 g yeast extract, 5 g NaCl per liter of media) at 37 ˚C until OD_{600} ~0.7. Protein expression was induced for 3 h at 37 ˚C by the addition of 0.3 mM IPTG. Cells were harvested by centrifugation at 7,000xg, resuspended in 150 mM NaCl, 50 mM Tris, pH 7.5 and lysed by sonication till the viscosity of the solution was reduced along with an observed change in color. The lysate was centrifuged at 20,000xg to separate insoluble inclusion bodies and histones were extracted from these bodies using a buffer consisting of 6 M guanidine hydrochloride (Gn-HCl), 10 mM Tris, pH 7.5. The re-solubilized histones were purified away from other proteins and DNA/RNA by size-exclusion chromatography on a Superdex S-200 column. Fractions containing the desired histones were identified by 15% SDS-PAGE and the proteins were further purified to homogeneity by RP-HPLC and characterized by ESI-MS (Figure 2.4.1).

2.4.3 Purification of TEV Protease

His_6-TEV protease was overexpressed and purified from *E. coli* BL21(DE3) cells using the expression plasmid pRK793. Cells were grown to an OD_{600} of ~0.5 at 37 ˚C and protein expression was induced by the addition of 0.3 mM IPTG followed by an additional growth for 4 h at 37 ˚C. Cells were harvested by centrifugation at 7,000xg, resuspended in 150 mM NaCl, 50 mM Tris, pH 7.5 and lysed by sonication till the viscosity of the solution was reduced along with an observed change in color. The lysate was centrifuged at 20,000xg to remove insoluble cellular debris and supernatant was incubated with Ni^{2+}-NTA resin for purification by affinity chromatography. After 30 min of incubation at 4 ˚C, the resin was washed with 5 column volumes (CV) of lysis buffer containing 50 mM imidazole, followed by a 100 mM imidazole containing buffer. His_6-TEV was eluted using lysis buffer containing 500 mM imidazole. Observation of white precipitate indicated the presence of protein but all the fractions were nonetheless visualized by 15% SDS-PAGE. Fractions containing 90% pure His_6-TEV were

pooled and dialyzed against 150 mM NaCl, 50 mM Tris, pH 7.5 buffer overnight. Aliquots were made, flash frozen with liquid nitrogen and stored at -80 ˚C for future use.

2.4.4 Generation of human H4 K12C and H4 K16C.

Human histone H4 was cloned between the NdeI and BamHI restriction sites in the pET15b vector with a Tobacco Etch Virus (TEV) protease cleavage sequence between the His6-tag and the H4 N-terminus. Site-directed mutagenesis was then employed to generate the mutants H4 K12C and H4 K16C from this plasmid. The primers used to perform the mutagenesis are mentioned in Table 1. *E. coli* BL21(DE3) cells were transformed with the mutant plasmids according to protocol mentioned in 2.4.1 and grown at 37 ˚C in 6 L YT medium until OD_{600} ~0.7. Protein expression was induced by the addition of 0.3 mM IPTG for 3 h at 37 ˚C. Cells were harvested by centrifugation at 7000xg, resuspended in 150 mM NaCl, 50 mM Tris, pH 7.5 and lysed by sonication. The lysate was centrifuged at 20,000xg for 20 min and insoluble histones were recovered from inclusion bodies with an extraction buffer consisting of 6 M Gn-HCl, 100 mM NaCl, 50 mM Tris, pH 7.5. The re-solubilized histones were purified by Ni^{2+}-affinity chromatography. Briefly, proteins were bound to the Ni^{2+}-column overnight at 4 ˚C with gentle shaking and the column subsequently washed with 3 volumes of extraction buffer containing 25 mM Imidazole. Immobilized proteins were eluted using 500 mM imidazole in the extraction buffer. Histone-containing fractions were identified by 15% SDS-PAGE, combined and dialyzed into 1 mM DTT overnight at 4 ˚C. Purified His_6-TEV protease was then added in a 1:10 molar ratio (enzyme:substrate) and His_6-tag cleavage was performed overnight in a buffer containing 1 mM EDTA, 10 mM DTT, 10 mM Cys, 50 mM Tris, pH 6.9. The cleavage products were dialyzed into extraction buffer and applied to a Ni^{2+}-column. This led to retention of the cleaved His6-tag, His_6-TEV and any uncleaved histones on the column. The eluted tag-less H4 K12C and H4 K16C were further purified to homogeneity by RP-HPLC and characterized by ESI-MS (Figure 2.4.2).

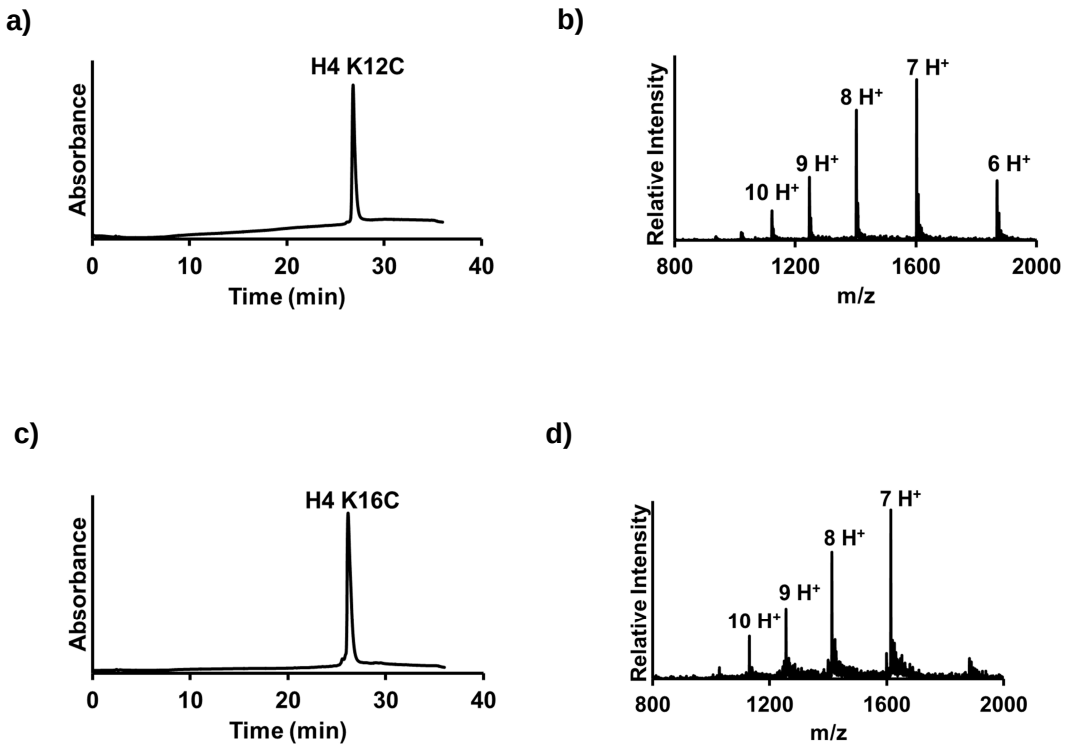


Figure 2.4.2. Characterization of mutants of histone H4. a) C4 analytical RP-HPLC trace of human H4 K12C on a 0-73% buffer B gradient over 30 minutes. b) ESI-MS of pure H4 K12C, calculated m/z $[M+H]^+$ 11,212.1 Da, observed 11,213.9±3.2 Da. c) C4 analytical RP-HPLC trace of human H4 K16C on a 0-73% buffer B gradient over 30 minutes. d) ESI-MS of pure H4 K16C, calculated m/z $[M+H]^+$ 11,212.1 Da, observed 11,213.9±3.2 Da.

2.4.5 Purification of human SUMO-3 C47S-aminoethanethiol and Ub-aminoethanethiol

The full-length human ubiquitin gene in pTXB1 was obtained from Dr. Tom Muir at Princeton University. The full-length human SUMO-3 gene was cloned in in pTXB1 between Nde1 and Sap1 restriction sites. Site-directed mutagenesis was employed to generate the SUMO-3 C47S mutant from human SUMO-3 cloned in the pTXB1 vector. *E. coli* BL21(DE3) cells transformed with SUMO-3 C47S and Ub plasmids were grown to an OD_{600} of ~0.5 at 37 ˚C followed by the addition of 0.3 mM IPTG. The cells were grown for another 4 h at 25 ˚C, harvested by centrifugation at 7,000xg, resuspended in lysis buffer containing 200 mM NaCl, 20 mM Tris, pH 7.5, and lysed by sonication. The lysate was centrifuged at 20,000xg for 20 min and the soluble supernatant containing SUMO-3 C47S- and Ub-intein-chitin binding domain fusion proteins was bound to chitin beads at 4 ˚C overnight. The column was subsequently washed with 3 column volumes of lysis buffer and intein-mediated production of SUMO-3 C47S-aminoethanethiol and Ub-aminoethanethiol was initiated by incubating the column with lysis buffer containing 100 mM cysteamine, pH 7.5, for 48 h at 4 ˚C. After this period, the desired adducts were eluted from the column with lysis buffer, further purified by RP-HPLC and characterized by ESI-MS (Figure 2.4.3).

2.4.6 Generation of H4 K12C-Npys (2)

RP-HPLC purified H4 K12C was dissolved in a 1:3 mixture of water:acetic acid (v/v) and reacted with an excess of 2,2'-dithiobis(5-nitropyridine) (DTNP) for 2 h at 25 ˚C with vigorous shaking. The desired product H4 K12C-Npys (2) (Figure 2.1a) was purified away from unreacted starting materials by RP-HPLC and characterized by ESI-MS (Figure 2.4.4).

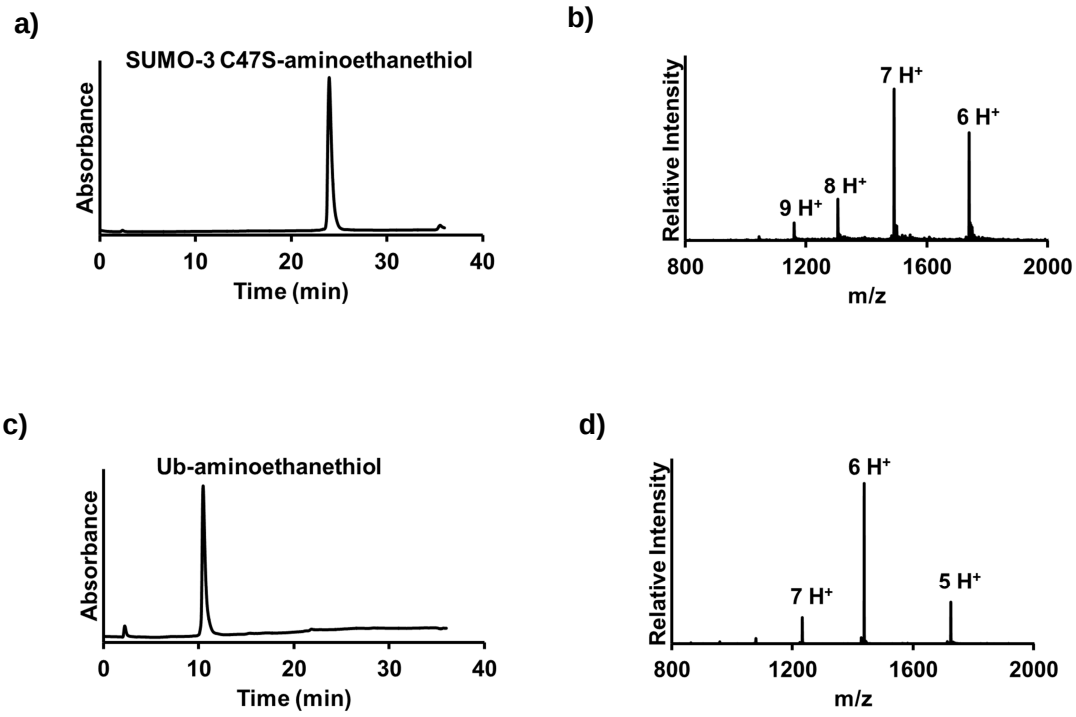


Figure 2.4.3. Characterization of SUMO-3 and Ubiquitin containing a reduced thiol functional group at the C-terminus. a) C18 analytical RP-HPLC trace of SUMO-3 C47S-aminoethanethiol (**3**) on a 15-50% buffer B gradient over 30 minutes. b) ESI-MS of pure **3**, calculated m/z $[M+H]^+$ 10,437.6 Da, observed 10,437.3±1.1 Da. c) C4 analytical RP-HPLC trace of $uH4_{ss}$ on 30-70% buffer B gradient over 30 minutes. d) ESI-MS of pure $uH4_{ss}$ calculated m/z $[M+H]^+$ 19,833.1 Da, observed 19,832.8±3.3 Da.

2.4.7 Synthesis of $suH4_{ss}$ and $uH4_{ss}$

One equivalent of H4 K12C-Npys (2) and two equivalents of SUMO-3 C47S-aminoethanethiol (3) (Figure 2.1a), or Ub-aminoethanethiol were dissolved in a reaction buffer consisting of 6 M Gn-HCl, 1 M HEPES, pH 6.9 and allowed to react for 1 h at 25 ˚C with continuous shaking. The desired products, $suH4_{ss}$ or $uH4_{ss}$ were purified away from starting materials by RP-HPLC and characterized by ESI-MS (Figure 2.4.4).

2.4.8 Synthesis of H4 K_s16ac and H4 K_s12ac

To a solution of 2.2 mg of H4 K16C or H4 K12C in 200 µL of 6 M Gn-HCl, 200 mM NaOAc, pH 6.0, L-glutathione was added to a final concentration of 15 mM to ensure complete reduction of the thiol groups. Following this, 50 mM N-vinylacetamide, 100 mM dimethyl sulfide, and 50 mM of the free-radical initiator VA-044 were added (2). The reaction vessel was covered with Aluminum foil to protect from light and incubated at 37 °C for 2 h. The desired products, H4 Ks16ac and H4 Ks12ac, were purified by RP-HPLC and characterized by ESI-MS (Figure 2.4.5).

2.4.9 Generation of 147 bp 601 DNA

The 147 bp 601 DNA was amplified from a plasmid containing the 1_147_601 fragment by PCR using the forward primer 5'-CTGGAGAATCCCGGTGCCGAGG-3' and reverse primer 5'-ACAGGATGTATATATCTGACACG-3'. The PCR product was purified using a QIAquick PCR purification kit, eluted in sterile water and lyophilized to concentrate the DNA. The dry DNA was then resuspended in sterile water to a final concentration of 20-25 µM.

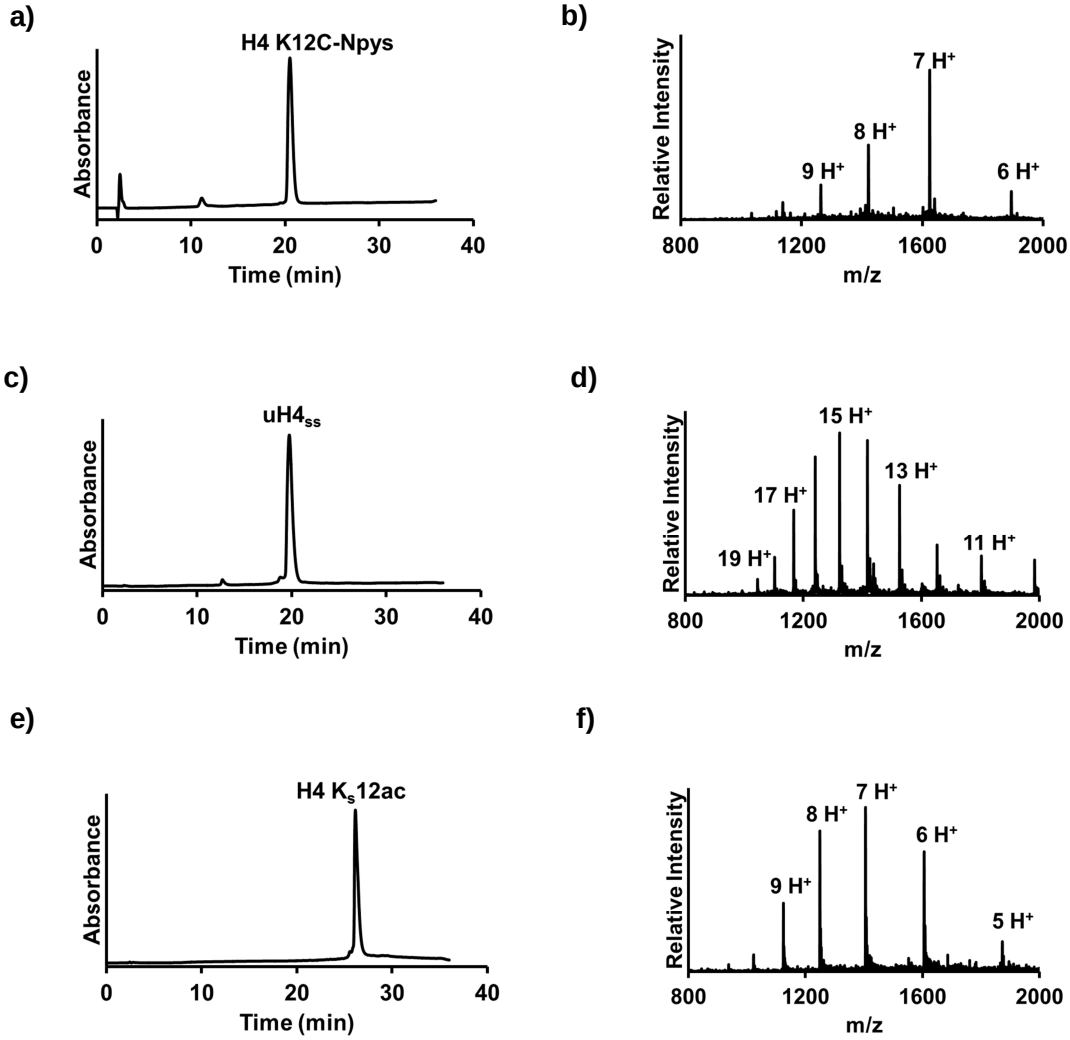


Figure 2.4.4. Characterization of ubiquitylated ($uH4_{ss}$) and acetylated (H4 K_s12ac) histone H4. a) C4 analytical RP-HPLC trace of human H4 K12C-Npys (2) on a 30-70% buffer B gradient over 30 minutes. b) ESI-MS of pure 2, calculated m/z $[M+H]^+$ 11366.5 Da, observed 11,364.1±2.2 Da. c) C4 analytical RP-HPLC trace of $uH4_{ss}$ on 30-70% buffer B gradient over 30 minutes. d) ESI-MS of pure $uH4_{ss}$ calculated m/z $[M+H]^+$ 19,833.1 Da, observed 19,832.8±3.3 Da. e) C4 analytical RP-HPLC trace of human H4 K_s12ac on a 0-73% buffer B gradient over 30 minutes. f) ESI-MS of pure H4 K_s12ac, calculated m/z $[M+H]^+$ 11,296.1 Da, observed 11,295.6 ± 1.0 Da.

2.4.10 Generation of Histone Octamers

Histone octamers were assembled as previously reported with the exclusion of reducing agents. Each of the four core histones was dissolved at ~4 mg/mL in an unfolding buffer containing 7 M Gn-HCl, 20 mM Tris, pH 7.5. The histones were mixed in equimolar amounts and the resulting mixture dialyzed into refolding buffer (3 x 1 L) consisting of 2 M NaCl, 1 mM EDTA, 10 mM Tris, pH 7.5. Crude octamers were concentrated and purified by size-exclusion on a Superdex S-200 column. Fractions containing pure histone octamers were identified by 15% SDS-PAGE, combined and concentrated prior to nucleosome assembly.

2.4.11 Generation of Mononucleosomes

Pure histone octamers and 147 bp 601 DNA were combined in 10 µL of a high-salt refolding buffer consisting of 2 M NaCl, 1 mM EDTA, 10 mM Tris, pH 7.5, to a final concentration of 2 µM in each. After incubation at 37 ˚C for 15 min, 3.3 µL of Dilution Buffer 1 containing 1 mM EDTA, 0.5 mM PMSF, 10 mM HEPES, pH 7.9, was added and the temperature dropped to 30 ˚C. Further dilutions of 6.7, 5, 3.6, 4.7, 6.7, 10, 30, and 20 µL, respectively, were performed every 15 min. A final dilution was undertaken with 100 µL of Dilution Buffer 2 containing 1 mM EDTA, 0.1% (v/v) NP-40, 0.5 mM PMSF, 20% (v/v) glycerol, 10 mM Tris, pH 7.5. After an additional 15 min at 30 ˚C, the mononucleosomes (MNs) were concentrated and analyzed by separation on a Criterion 5% TBE gel run in 0.5X TBE, followed by staining with ethidium bromide. For sumoylated MNs, reduction with 1 mM DTT for 30 min on ice led to the removal of SUMO and co-migration of the reduced MNs with unmodified wild-type MNs. This indicated the identical core histone composition of sumoylated and unmodified MNs.

2.4.12 Generation of 12_177_601 Array DNA

A plasmid containing 12 copies of a 177 bp repeat of the 601 nucleosome positioning sequence (12_177_601) flanked by EcoRV sites was purified from a 6 L culture of DH5α cells using a Qiafilter Plasmid Giga kit. The 12_177_601 sequence was obtained by preparative-scale digestion of the plasmid with EcoRV followed by selective precipitation of the 12_177_601 fragment with 6% polyethylene glycol (PEG)-6000 on ice followed by centrifugation at 26,000xg for 30 min at 4 ^{o}C. After phenol extraction and ethanol precipitation the DNA was redissolved in TE buffer (1 mM EDTA, 10 mM Tris, pH 7.5) and quantified by Abs_{260} prior to array formation.

2.4.13 Generation of 12-mer Nucleosomal Arrays

Pure histone octamers (2 μM) and 12_177_601 DNA (0.17 μM) were combined in 75 μL of reconstitution buffer consisting of 2 M KCl, 0.1 mM EDTA, 10 mM Tris pH 7.8. Additionally, 0.7 mM of a weaker-binding 147 bp fragment of the MMTV DNA was added to prevent overloading of the array with octamers. Stepwise dialysis was performed at 4 ˚C against reconstitution buffer containing 1.4 M NaCl, 1.2 M NaCl, 1 M NaCl, 0.8 M NaCl, 0.5 M NaCl and 10 mM NaCl for 90 min each, followed by a final dialysis step against reconstitution buffer containing 10 mM NaCl. DNA fragments were removed by selective precipitation of the arrays with $MgCl_2$. The crude array mixture was incubated with 4 mM $MgCl_2$ for 10 min on ice to allow the aggregation of the arrays. The mixture is spun at 15,000 g at 4 ˚C for 10 min and the supernatant is immediately removed. The final concentration of $MgCl_2$ used to precipitate arrays can vary based on the modifications present on the histone tails. For example, arrays containing acetylation at Lys 16 of H4 have to be precipitated at 6 mM $MgCl_2$ due to tendency of acetylation to prevent aggregation. After removal of the supernatant, the arrays were resuspended in TEN buffer, dialyzed against fresh TEN buffer followed by concentration determination by Abs_{260}. Wild-type

and modified arrays were visualized on 1% agarose-2% polyacrylamide gel electrophoresis (APAGE) by running gels at 180 V for 20 min at 4 °C followed by staining with ethidium bromide.

The saturation of octamer-binding sites in 12-mer arrays was confirmed by the digestion of 0.17 pmol of arrays with 10 U of ScaI restriction enzyme (New England Biolabs, Catalog number: R3122S) in NEB buffer 3 at 25 ^{o}C for 12 h, followed by separation on a Criterion 5% TBE gel run in 0.5 x TBE buffer and staining with ethidium bromide. The presence of an MN band as well as the absence of free DNA and higher-running species indicated full array occupancy. DTT in the digestion and restriction enzyme buffer led to partial loss of the SUMO moiety in sumoylated arrays resulting in three distinct MN bands corresponding to the doubly, singly and non-sumoylated species. The presence of 12 MNs per array was also confirmed by partial digestion with the enzyme micrococcal nuclease (MNase). In a typical assay, 0.17 pmol of 12-mer arrays were digested with 0.2 U of MNase for 60 s on ice. The reaction was stopped by the addition of 0.2% (w/v) SDS and 20 mM EDTA followed by DNA purification using a QIAquick PCR purification kit (Qiagen). The DNA fragments were separated on a Criterion 5% TBE gel run in 0.5x TBE buffer and visualized after ethidium bromide staining.

2.4.14 Analytical Ultracentrifugation

Following previous reports, 12-mer array samples for analytical ultracentrifugation were prepared in 400 µL of measurement buffer consisting of 10 mM NaCl and 10 mM Tris, pH 7.8 containing either 0.1 mM EDTA (referred to as 0 mM Mg^{2+} buffer), or 1 mM Mg^{2+}. Samples were equilibrated under vacuum for 2 h at 20 ˚C in a Beckman XL-1 Analytical Ultracentrifuge with absorbance optics (Beckman Coulter Inc., Brea, CA) before sedimentation at 12,000 r.p.m. in 12 mm double-sector cells in an 8-place An-50 Ti analytical rotor. Sample absorbance was measured at 260 nm in a continuous scan mode. The integrity of arrays and octamers was

assessed by 1% APAGE and 15% SDS-PAGE after completion of ultracentrifugation analysis. No reduction of the disulfide bond was observed in $suH4_{ss}$ and all core histones were maintained in the correct ratios after sedimentation.

2.4.15 Chromatin Precipitation Assays

Reconstituted 12-mer arrays were dialyzed into a measurement buffer consisting of 10 mM NaCl and 10 mM Tris, pH 7.8. The samples were subsequently mixed with stock $MgCl_2$ solutions at twice the final concentration, incubated for 10 min at 22 ˚C and then centrifuged at 16,000xg for 10 min at 22 ^{o}C to remove insoluble aggregates. The fraction of initial 12-mer arrays at 0 mM $MgCl_2$ that remained in solution at 1 mM, 2 mM, 3 mM and 4 mM of $MgCl_2$ was determined by measuring Abs_{260} of the supernatant immediately after centrifugation.

2.4.16 Electron Microscopy

Nucleosomal arrays were dialyzed into HEN buffer (0.25 mM EDTA, 2.5 mM NaCl, and 10 mM Hepes pH 7.8), fixed with 0.1% glutaraldehyde for 4 h at 4 ˚C and dialyzed overnight against fresh HEN buffer. Samples were diluted as needed, applied to glow-discharged carbon-coated grids, and stained with 0.1% aqueous uranyl acetate. Grids were examined in a Tecnai 12 TEM operated at 100 kV in the tilted darkfield mode, and digital images were recorded using a TVIPS 2024 × 2024 CCD camera.

2.4.17 Single Molecule FRET Measurements

Quartz microscope slides were silanized and coated with polyethyleneglycol (PEG 5,000 m.w., Laysan Bio, Arab, AL). One percent of the PEG molecules have a biotin at one end (Laysan

Bio). The slide surfaces were passivated with BSA before immobilizing nucleosomes through streptavidin-biotin conjugation. Nucleosomes were prepared as previously reported. To detect dinucleosome formation events, Cy3-labeled nucleosomes (601A) were immobilized attached on the surface, and Cy5-labeled nucleosomes (601B) were injected on the slide. The fluorescence signals were taken within 1 h after nucleosome immobilization to avoid nucleosome disassembly. A protocatechuate dioxygenase and protocatechuic acid mixture in addition to Trolox (2 mM) was used to elongate the dye photobleaching lifetime and stabilize the emission.

Fluorescence signals from single Cy3 and Cy5 fluorophores were collected with an electron multiplying CCD camera (EMCCD, iXon+897; Andor Technology, Belfast UK) on a prism coupled total internal reflection setup based on a commercial microscope (TE2000; Nikon, Tokyo, Japan). The excitation of the FRET donor (Cy3) was achieved with a laser at 532 nm (Laser Quantum). Fluorescence emission from FRET pairs (Cy3 and Cy5) were separated into two spectral regions (550–645 and 645–750 nm) with a dichroic mirror, and the two separate images of Cy3 and Cy5 regions were projected on a single EMCCD to collect fluorescence signals simultaneously from the two fluorophores in a time-resolved manner at a frame rate of 25 ms. The time series of fluorescence intensities of Cy3 and Cy5 were obtained from the series of fluorescence images and were plotted against the elapsed time. FRET efficiency at each time point was calculated with a formula, Icy5/(Icy3 + Icy5). To construct the lifetime histogram of dinucleosomes, the duration of every FRET event was measured and each event was counted as one event of the measured duration. These events with their measured durations were constructed as the lifetime histogram of a dinucleosomal state. For dinucleosome formation frequency calculation, the total number of FRET events was counted and divided by the total observation time and the acceptor labeled mononucleosome concentration.

2.5 REFERENCES

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

(2) Geiss-Friedlander, R., and Melchior, F. (2007) Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol. 8*, 947–56.

(3) Bischof, O., Schwamborn, K., Martin, N., Werner, A., Sustmann, C., Grosschedl, R., and Dejean, A. (2006) The E3 SUMO Ligase PIASy Is a Regulator of Cellular Senescence and Apoptosis. *Mol. Cell 22*, 783–794.

(4) Hendriks, I. A., D'Souza, R. C., Chang, J.-G., Mann, M., and Vertegaal, A. C. O. (2015) System-wide identification of wild-type SUMO-2 conjugation sites. *Nat. Commun. 6*, 7289.

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

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

(7) Becker, J., Barysch, S. V, Karaca, S., Dittner, C., Hsiao, H., Diaz, M. B., Herzig, S., Urlaub, H., and Melchior, F. (2013) Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat. Struc. Mol. Biol. 20,* 525-532.

(8) 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., and Berger, S. L. (2006) Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes Dev. 20*, 966–76.

(9) Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., and Peterson, C. L. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science 311*, 844–7.

(10) Lu, X., Simon, M. D., Chodaparambil, J. V, Hansen, J. C., Shokat, K. M., and Luger, K. (2008) The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat. Struc. Mol. Biol. 15*, 1122–1124.

(11) Margueron, R., and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature 469*, 343–9.

(12) Jason, L. J., Moore, S. C., Ausio, J., and Lindsey, G. (2001) Magnesium-dependent association and folding of oligonucleosomes reconstituted with ubiquitinated H2A. *J. Biol. Chem. 276*, 14597–601.

(13) Chatterjee, C., McGinty, R. K., Pellois, J.-P., and Muir, T. W. (2007) Auxiliary-mediated site-specific peptide ubiquitylation. *Angew. Chem. Int. Ed. Engl. 46*, 2814–8.

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

(15) Li, F., Allahverdi, A., Yang, R., Lua, G. B. J., Zhang, X., Cao, Y., Korolev, N., Nordenskiöld, L., and Liu, C.-F. (2011) A direct method for site-specific protein acetylation. *Angew. Chem. Int. Ed. Engl. 50*, 9611–4.

(16) Hoyle, C. E., and Bowman, C. N. (2010) Thiol-ene click chemistry. *Angew. Chem. Int. Ed. Engl. 49*, 1540–73.

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

(18) Schwarz, P. M., and Hansens, J. C. (1994) Formation and Stability of Higher Order Chromatin Structures *12*, 16284–16289.

(19) Schwarz, P. M., Felthauser, a, Fletcher, T. M., and Hansen, J. C. (1996) Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry 35*, 4009–15.

(20) Voigt, P., LeRoy, G., Drury, W. J., Zee, B. M., Son, J., Beck, D. B., Young, N. L., Garcia, B. A, and Reinberg, D. (2012) Asymmetrically modified nucleosomes. *Cell 151*, 181–93.

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

(22) Fierz, B., Chatterjee, C., McGinty, R. K., Bar-Dagan, M., Raleigh, D. P., and Muir, T. W. (2011) Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat. Chem. Biol. 7*, 113–9.

(23) Dhall, A., and Chatterjee, C. (2011) Chemical Approaches To Understand the Language of Histone Modifications. *ACS Chem. Biol. 6,* 987–999.

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

(25) Cubeñas-Potts, C., and Matunis, M. J. (2013) SUMO: A Multifaceted Modifier of Chromatin Structure and Function. *Dev. Cell 24*, 1–12.

CHAPTER 3

BIOCHEMICAL INVESTIGATION OF THE CROSSTALK BETWEEN H4 SUMOYLATION AND H3 METHYLATION

3.1 INTRODUCTION

The various functional roles of chromatin are dictated, in a large part, by the intricate crosstalk between various histone post-translational modifications. This crosstalk is mediated by numerous chromatin effector protein complexes. Many of these complexes are capable of binding specific PTMs and further modifying chromatin to modulate a downstream process. Our detailed biophysical studies ruled out the possibility that sumoylation of H4 might mediate transcriptional repression by facilitating the compaction of chromatin. We next probed the *trans* effects of sumoylation to discover the biochemical mechanisms by which H4 sumoylation may mediate gene repression. Several chromatin binding complexes are known to harbor writer and eraser modules that can install PTMs associated with heterochromatin and remove PTMs associated with euchromatin, respectively. We considered various scenarios by which sumoylation of H4 could either facilitate the installation of repressive marks on chromatin or facilitate the removal of activating marks from chromatin, hence helping establish a transcriptionally repressive state.

Acetylation of histones is primarily associated with active transcription and a member of the SAGA (Spt-Ada-Gcn5-acetyltransferase) complex, the histone acetyl transferase GCN5, is known to acetylate K9 and K14 of free H3, and also in nucleosomes[1]. Binding of methylated H3

K9 by HP1 is essential for the establishment of a heterochromatic state[2]. As GCN5 blocks the site of methylation by installing acetyl groups on the H3 tail, it prevents the formation of heterochromatin and facilitates the formation of the pre-initiation complex (PIC). It is possible that SUMO can antagonize the binding of the SAGA complex and hence block acetylation. An example of this is seen in case of ubiquitin where a module in the SAGA complex, USP22, is responsible for deubiquitinating H2B. This is surprising because uH2B is also associated with transcriptional activation just like the SAGA complex, but one potential reason for this surprising activity could be to get rid of the steric bulk associated with ubiquitin, and allow other subunits of the SAGA complex to bind. Since the SAGA complex doesn't contain any SUMO proteases, one can imagine that the presence of SUMO on the H4 tail in a nucleosome could prevent acetylation of the H3 tail by blocking the binding of the SAGA complex, and thus allow the formation of heterochromatin. Indirect evidence for this was observed by Shiio and Eisenman when they probed the histone acetylation status of GAL4 sites in HeLa cells using an anti-acetylated histone H3 antibody after the cells were co-transfected with 4XGAL14D luciferase reporter and GAL4-UBC9. A significant decrease in the acetylation of H3 was observed in the co-transfected cells but not in the mock or control cells[3].

Another pathway by which sumoylation could mediate gene repression is by facilitating the installation of PTMs associated with gene repression. An opposite example of this crosstalk has been well established in the case of ubiquitylation of Lys120 of H2B. Muir and co-workers demonstrated that the presence of ubiquitylation on H2B activates the methyltransferase hDot1L, which then installs methylation on Lys79 of H3 in the ubiquitylated MN[4]. H3 K79 methylation is observed in 90% of the yeast chromatin, a majority of which is actively transcribed[5]. Further studies in metazoan supported this observation when ChIP-seq using mouse 3T3 cells demonstrated that all H3 K79 methylation marks are localized within the body

of transcribed genes, and that the amount of enrichment of these marks correlates with expression level[6]. Unfortunately none of the known complexes that install repressive marks have known SUMO interaction motifs (SIMs). Instead we focused on another scenario in which sumoylation could mediate gene repression by recruiting chromatin binding complexes that contain subunits responsible for removing PTMs associated with transcriptional activation.

A large number of transcription factors have been found to be modified by SUMO. While a few of them activate transcription upon sumoylation, most of them exhibit decreased transactivation activity when modified by SUMO[7]. In case of the transcription factor Sp3, it was proposed that part of the mechanism by why sumoylation inhibits its transactivation activity is by localizing Spr3 to the nuclear periphery and nuclear dots[8]. The exact mechanisms of the inhibitory effect of SUMO did not become clear till the molecular basis of transcriptional repression upon sumoylation of transcription factors was extensively characterized for the transcription factor Elk-1. The transcriptional activity of Elk-1 is inhibited upon modification by SUMO at Lys230 and Lys249[9]. The authors were surprised to observe transcriptional repression of the luciferase gene under the control of a GAL4 promoter by a direct fusion of GAL4 with SUMO alone. ChIP analysis of the GAL-driven promoter-reporter revealed that sumoylation recruited HDAC2 to the promoters, resulting in histone deacetylation, and hence transcriptional repression[10]. Recently, Gill and co-workers described a role for SUMO-2/3 in gene-specific recruitment and activity of the lysine specific demethylase 1 (LSD1), co-repressor of RE1 silencing transcription factor (CoREST) and histone deacetylase1 (HDAC1) corepressor complex[11].

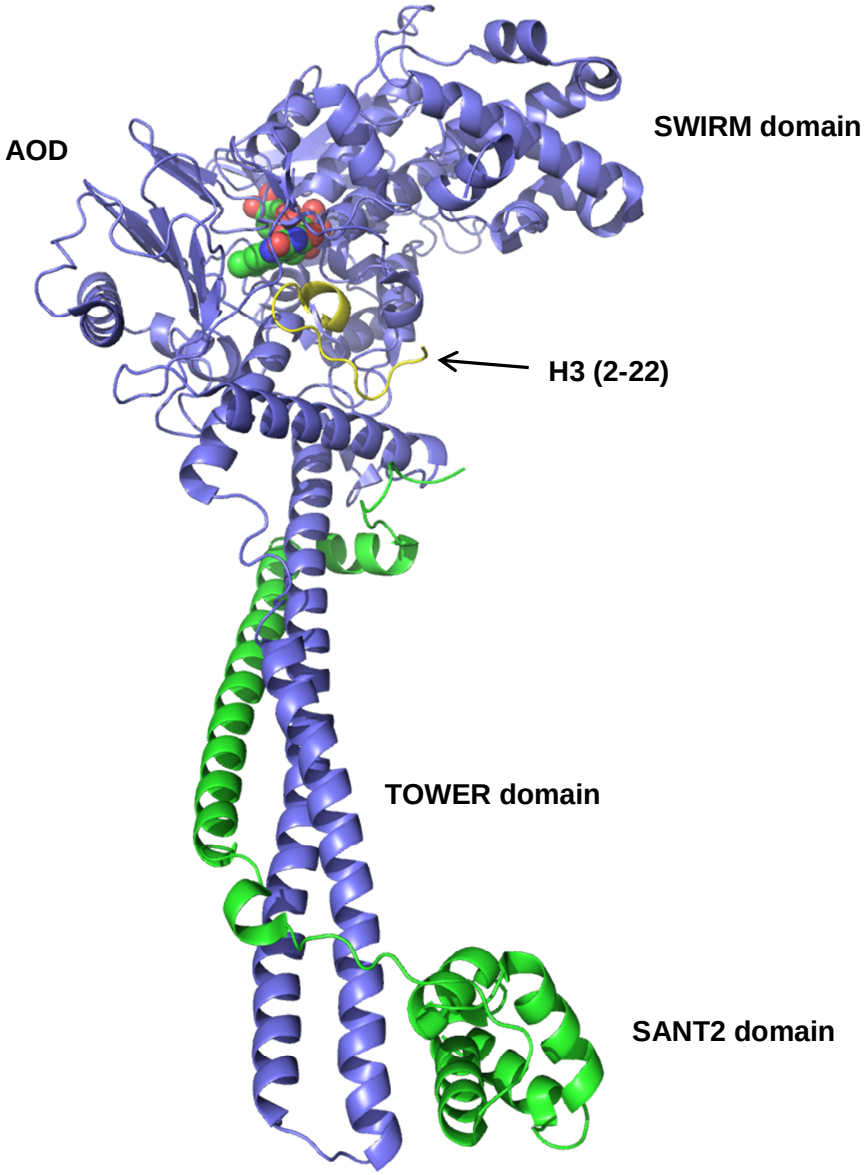


Figure 3.1. Crystal structure of LSD1 and CoREST bound to the H3 (2-22) peptide at 3.1 Å resolution. CoREST (305-482) is depicted in Green, LSD1 (123-852) in Blue and the H3 (2-22) tail in yellow. FAD bound to the amineoxidase domain (AOD) of LSD1 is depicted with spheres. PBD code 2V1D.

They showed that CoREST directly but non-covalently binds SUMO-2/3, but not SUMO-1. This interaction bridges LSD1 binding to SUMO-2/3. The deconjugation of SUMO-2/3 led to derepression of certain genes due to an increase in the level of H3 K4 methylation as a result of the loss of promoter occupancy by CoREST1 and LSD1. This led the authors to propose a role for sumoylation in regulation of histone modifications such as methylation, and in the transcriptional silencing of some neuronal-specific genes in non-neuronal cells. Since the LSD1/CoREST complex (Figure 3.1) has also been shown to bind nucleosomes[12], we investigated whether sumoylation of the H4 tail was responsible for facilitating the demethylation of H3 K4me1/2 by LSD1, and hence mediating gene repression.

For a long time methylation was thought to be an irreversible mark, but in 2004 Shi and co-workers demonstrated the demethylation activity of LSD1 on K4-methylated histone H3[13]. LSD1 is known to consist of an unstructured N-terminal region thought to contain a putative nuclear localization signal followed by a SWIRM domain. Recently the first 170 amino acids of LSD1 were shown to be important for binding the ubiquitin E3 ligase, RNF168 and mediating DNA damage response[14]. The C-terminal amine oxidase domain (AOD) of LSD1 is bifurcated by two antiparallel helices which form the TOWER domain responsible for binding CoREST[15]. LSD1 is known to require the first 21 amino acids of the H3 N-terminus for efficient demethylation *in vitro*[16], with only the first 12 residues of the H3 tail binding the active site according to a docking model[15]. Hyperacetylation in this region of the H3 tail has been shown to completely abrogate the catalytic activity of LSD1, while various other individual PTMs such as arginine methylation and serine phosphorylation also significantly impair the demethylation of H3K4me1/2[17]. Since its discovery, a lot has been uncovered about the structure and the substrate specificity of LSD1, and we now know that the large catalytic cavity of this enzyme even allows for the demethylation of non-histone proteins such as the transcription factor p53[18]. On the other hand,

the enzymatic activity of LSD1 on nucleosomes isolated from mammalian cells has been shown to be dependent on various binding factors such as CoREST in case of the LSD1/CoREST/HDAC1 complex[19] or the MTA proteins in case of the NuRD complex[20]. CoREST was first characterized as a corepressor to the REST (RE1 silencing transcription factor/neural restrictive silencing factor) transcription factor which represses neural genes in non-neuronal cells by recruiting various HDAC1/2 containing complexes[21,22]. CoREST consists of a C-terminal ELM2 (Egl-27 and MTA1 homology 2) domain whose function is not entirely known (Figure 3.2). It also has two SANT (SWI-SNF, ADA, N-CoR, and TFIIIB) domains which are involved in DNA binding and the SANT2 domain has been shown to be critical for LSD1-mediated demethylation of nucleosomes[12]. The SANT domains are connected by a linker region which interacts with the TOWER domain in LSD1 to form a stable LSD1/CoREST complex[15]. Interestingly, even though the SWIRM domains of human LSD1 and Swi3 have similar structure and sequence[23,24], NMR studies with short oligonucleotides failed to detect any binding with the isolated LSD1 SWIRM domain[15]. Hence, the DNA binding SANT2 domain of CoREST was proposed to be necessary and sufficient to stimulate LSD1 mediated demethylation of nucleosomes extracted from HeLa cells[12]. Also, similar to the trend observed with the H3 (1-21) peptide, hyperacetylation of nucleosomes impaired the demethylation activity of the LSD1/CoREST complex in vitro.

Previous studies involving recombinantly expressed and purified LSD1/CoREST complex have made use of the truncated form of these proteins for the ease of purification. LSD1Δ1-170 was shown to be sufficient for demethylating H3[25] and CoRESTΔ1-281 was sufficient to stimulate the activity of LSD1 toward nucleosomal substrate. In order to assess the effect of sumoylated H4 in facilitating the demethylation of H3K4me2 from nucleosomes, we employed full length LSD1 and CoREST1.

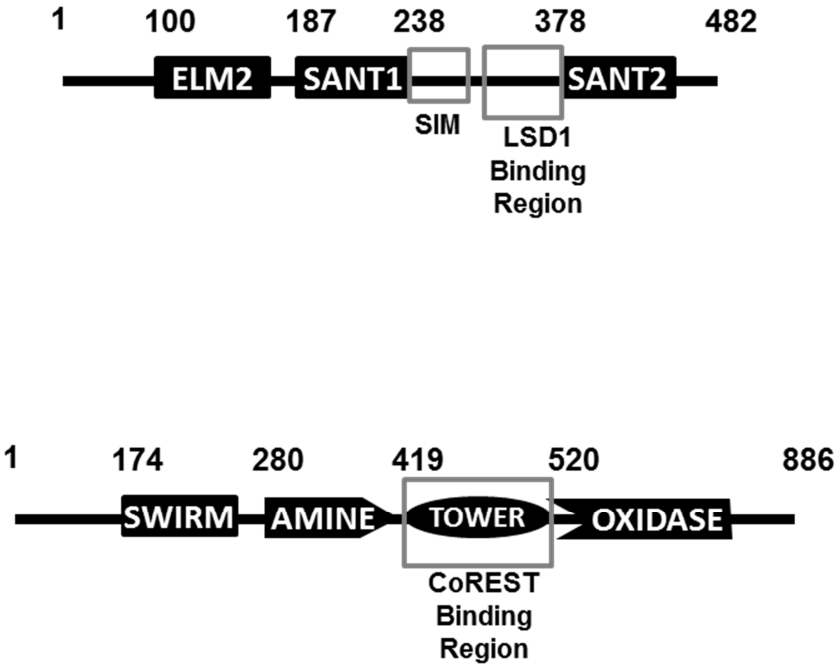


Figure 3.2. Diagram depicting the various DNA and protein binding domains and regions that have been identified in LSD1 and CoREST. ELM2 stands for the Egl-27 and MTA1 homology 2 domain. SANT1 stands for Swi3, Ada2, N-Cor, and TFIIIB domain. SIM stands for SUMO interaction motif, also known as SUMO binding motif and SWIRM stands for Swi3p, Rsc8p and Moira domain.

This was necessitated by the fact that the SUMO binding motif is localized in the N-terminal region of CoREST[11]. We first confirmed that the full-length complex shows similar kinetics with peptides and full-length H3, as previously reported for the truncated proteins. Next, we performed demethylation assays with sumoylated and unmodified MNs as substrates and discovered a novel crosstalk between sumoylation and demethylation. Sumoylated MNs showed faster demethylation with the LSD1/CoREST complex, which suggests that sumoylation can mediate gene repression by facilitating the removal of activating PTMs from chromatin.

3.2 RESULTS AND DISCUSSION

3.2.1 Effect of Neighboring Acetylation on Demethylation by LSD1

We sought to establish the kinetics of demethylation of recombinant full-length LSD1 in comparison with the previously reported kinetics for the N-terminal truncant of the enzyme, with H3 (1-21) peptide containing dimethylation at Lys4 as the substrate. The demethylation catalytic cycle involves the reduction of the FAD co-factor to $FADH_2$ along with the formation of an imine intermediate on the lysine side chain of the substrate, which is then spontaneously hydrolyzed to produce formaldehyde and demethylated lysine (Figure 3.3). Molecular oxygen then oxidizes $FADH_2$ to FAD in a reaction that involves the release of stoichiometric amount of H_2O_2[26]. We used a horseradish peroxidase (HRP) dependent hydrogen peroxide coupled assay to follow the activity of LSD1 on the peptidic substrates[27,28]. The K_M and k_{cat} values observed for various substrates with full-length LSD1 (Table 1) were very similar to the previously reported kinetic parameters observed when the LSD1Δ1-184 truncant was used for the *in vitro* demethylation of H3 (1-21) peptide. This confirms that the unstructured N-terminal region of LSD1 is dispensable for demethylation of the peptidic substrates.

Since LSD1 displays weak binding (K_M = 3.21 ± 0.37 μM) even with a long peptidic substrate containing the first 21 amino acids of the H3 tail, we wondered if acetylation on individual Lys residues in this region may influence the binding and catalysis. Despite the availability of a crystal structure of the H3 (1-21) tail bound to LSD1[29], the individual contributions of the various side chains toward LSD1 binding are not clearly understood. To address this, we generated various mono- (H3K9ac, H3K14ac, H3K18ac) and hyper-acetylated (H3K9acK14ac, H3K9acK14acK18ac) versions of the H3K4me2 (1-21) peptide. As expected, hyperacetylation in the H3 tail completely abolished catalysis by LSD1, but interestingly acetylation at Lys14, which is farther away from the di-methylated Lys4 residue, resulted in poorer substrate specificity compared to Lys9 acetylation (Table 1). Surprisingly, the biggest difference was observed when acetylation was installed on Lys18 of the H3 tail. To understand this interesting result, we made use of a docking model generated by Yu et. al. which suggests that the H3 tail may insert itself between the SWIRM and the amine oxidase domain of LSD1 to achieve effective binding. Based on the docking model, Lys9 and Lys14 of the H3 tail experience a less charged and more hydrophobic environment than Lys18 of the same tail, which explains some of the differences in the observed K_M. Unfortunately, in the absence of a crystal structure of the H3 tail with full length LSD1, it is difficult to predict the precise contacts made by these lysine side chains. Further mutational analyses of LSD1 will have to be undertaken to rationalize the stark increase in K_M upon acetylation of Lys18 in H3.

Since the presence of CoREST was indispensable for demethylation of MNs isolated from mammalian nuclei, which are known to contain multiple PTMs simultaneously, we wondered whether CoREST could help LSD1 overcome the effect of neighboring PTMs. We performed demethylation assays in the presence of full-length CoREST, but as previously observed with the truncated form of CoREST[30], the addition of full-length CoREST failed to catalytically

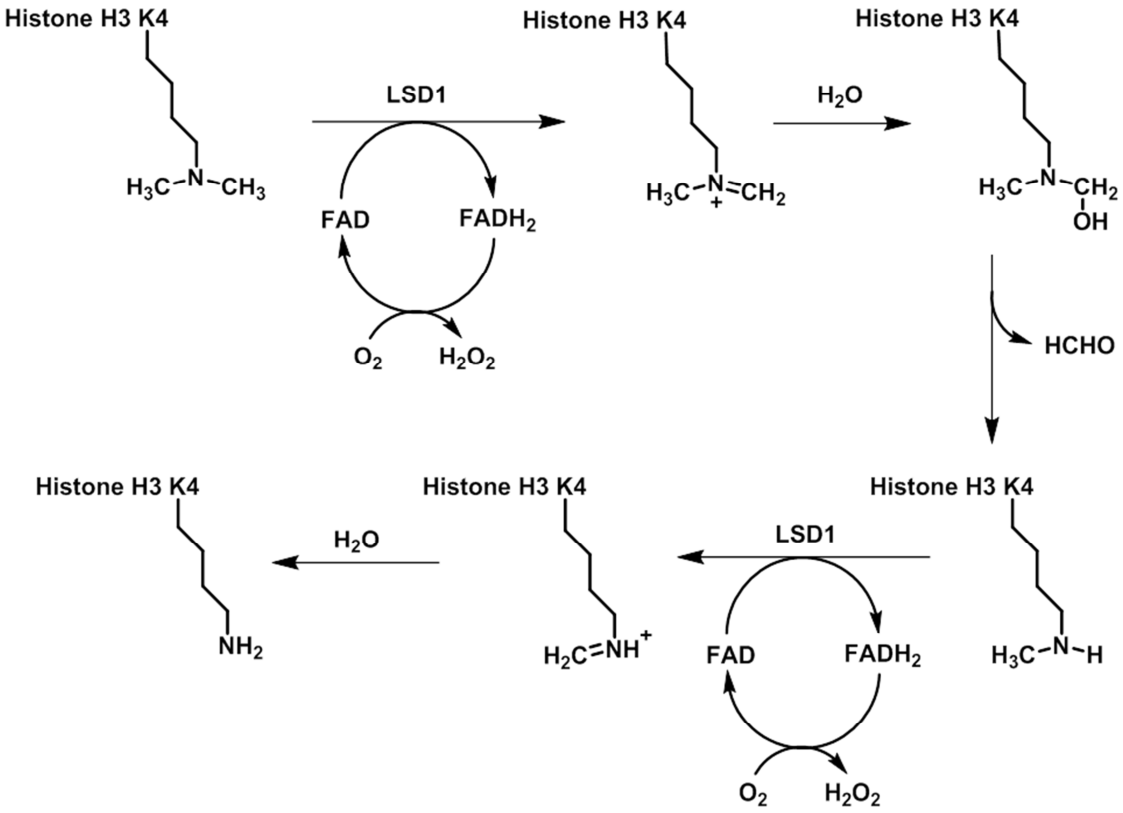


Figure 3.3. Diagram depicting the mechanism of the catalytic activity of LSD1 toward dimethylated H3 peptide resulting in a monomethylated species followed by another round of demethylation that results in complete reversal of methylation. The catalytic mechanism involves formation of an imine intermediate with concomitant reduction of the FAD co-factor. The imine is spontaneously hydrolyzed by molecular water and is released as formaldehyde. The co-factor is regenerated by molecular oxygen prior to another round of demethylation.

Substrate	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M (μM^{-1} min^{-1})
H3 (1-21) K4me2	3.21 ± 0.37	6.71 ± 1.85	2.10 ± 0.63
H3 (1-21) K4me2 K9ac	3.44 ± 0.72	1.65 ± 0.12	0.48 ± 0.11
H3 (1-21) K4me2 K14ac	0.68 ± 0.11	0.63 ± 0.02	0.92 ± 0.15
H3 (1-21) K4me2 K18ac	8.83 ± 2.65	2.08 ± 0.22	0.24 ± 0.07
H3 K_C4me2	3.12 ± 0.27	4.64 ± 0.50	1.51 ± 0.30
H3 K4me2	3.16 ± 0.77	10.94 ± 1.23	3.47 ± 0.60

Table 1. Kinetic parameters of demethylation of H3 peptides by full-length LSD1 in the presence of acetylation on neighboring lysine residues. Kinetic parameters for full-length native dimethylated H3 and its thialysine analog were also calculated.

Substrate	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M (μM^{-1} min^{-1})
H3 (1-21) K4me2	9.58 ± 1.50	8.56 ± 0.62	0.89 ± 0.15
H3 (1-21) K4me2 K9ac	2.22 ± 0.46	2.02 ± 0.12	0.91 ± 0.20
H3 (1-21) K4me2 K14ac	1.61 ± 0.22	0.90 ± 0.03	0.56 ± 0.08
H3 (1-21) K4me2 K18ac	32.83 ± 12.63	1.98 ± 0.39	0.06 ± 0.03

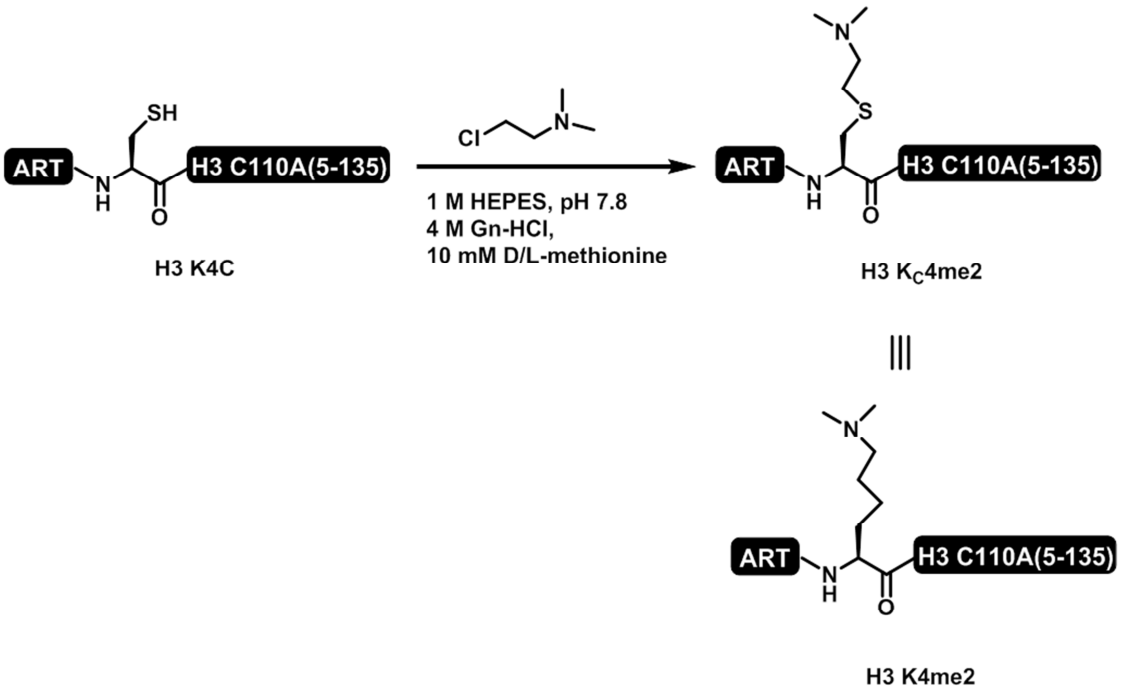
Table 2. Kinetic parameters of demethylation of H3 peptides by full-length LSD1 and CoREST complex in the presence of acetylation on neighboring lysine residues.

activate demethylation of the peptidic substrates (Table 2). All the spectrophotometric demethylation assays were initiated by me, but finally performed and completed by Patrick Shelton in the Chatterjee lab.

3.2.2 Synthesis and Catalysis of Native H3K4me2

Next, we sought to compare the catalytic activity of LSD1 toward the widely used methyl lysine analog versus the native full length H3 protein containing dimethylated Lys4. We synthesized the thia-lysine analog of methylation ($H3K_C4me2$) by using the methodology reported by Shokat and co-workers[31] (Figure 3.4a). To synthesize the native full-length dimethylated protein, we used solid-phase peptide synthesis (SSPS) and native chemical ligation to generate multi-milligram quantities of H3 K4me2. The synthesis was achieved by generating a 6 amino acid long peptide on 2-chlorotrityl chloride resin containing dimethylated Lys residue at the fourth position from the N-terminus. The peptide was cleaved from the resin as a hydrazide which was then derivatized into an activated thioester using 4-mercaptophenylacetic acid (Figure 3.4b). The C-terminal fragment of H3 containing A7C mutation was recombinantly obtained from *E. coli* and used to react with the peptide thioester in order to generate full length H3. Radical based desulfurization of the cysteine at the site of ligation to the native alanine residue resulted in native full length H3K4me2 protein. On comparing the demethylation kinetics of H3K4me2 protein with $H3K_C4me2$, we discovered that the native substrate is twice as efficiently demethylated by LSD1 (Table 1). We also compared the sensitivity of the commercially available antibodies toward H3K4me2 and $H3K_C4me2$ using dot-blot analysis, and discovered that the native substrate is detected significantly better than the analog (Figure 3.4c). Hence, we decided to incorporate the native substrate in MNs for all future demethylation studies.

a)



b)

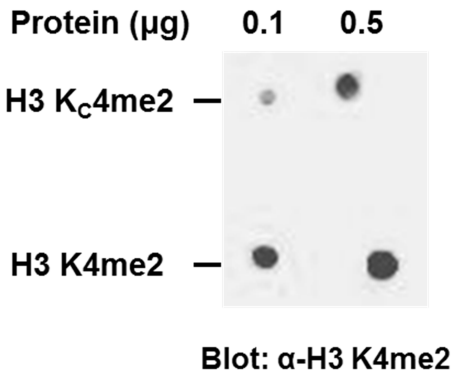


Figure 3.4. Synthesis and detection of full-length native H3 K4me2 and its thia-lysine analog. a) Synthesis of methyl-lysine analog of dimethylated H3 K4 through nucleophilic attack by the side-chain of a cysteine residue in H3. b) Dot-blot analysis of equal quantities of native H3 K4me2 and its thia-lysine analog using a commercially available antibody recognizing H3 K4me2.

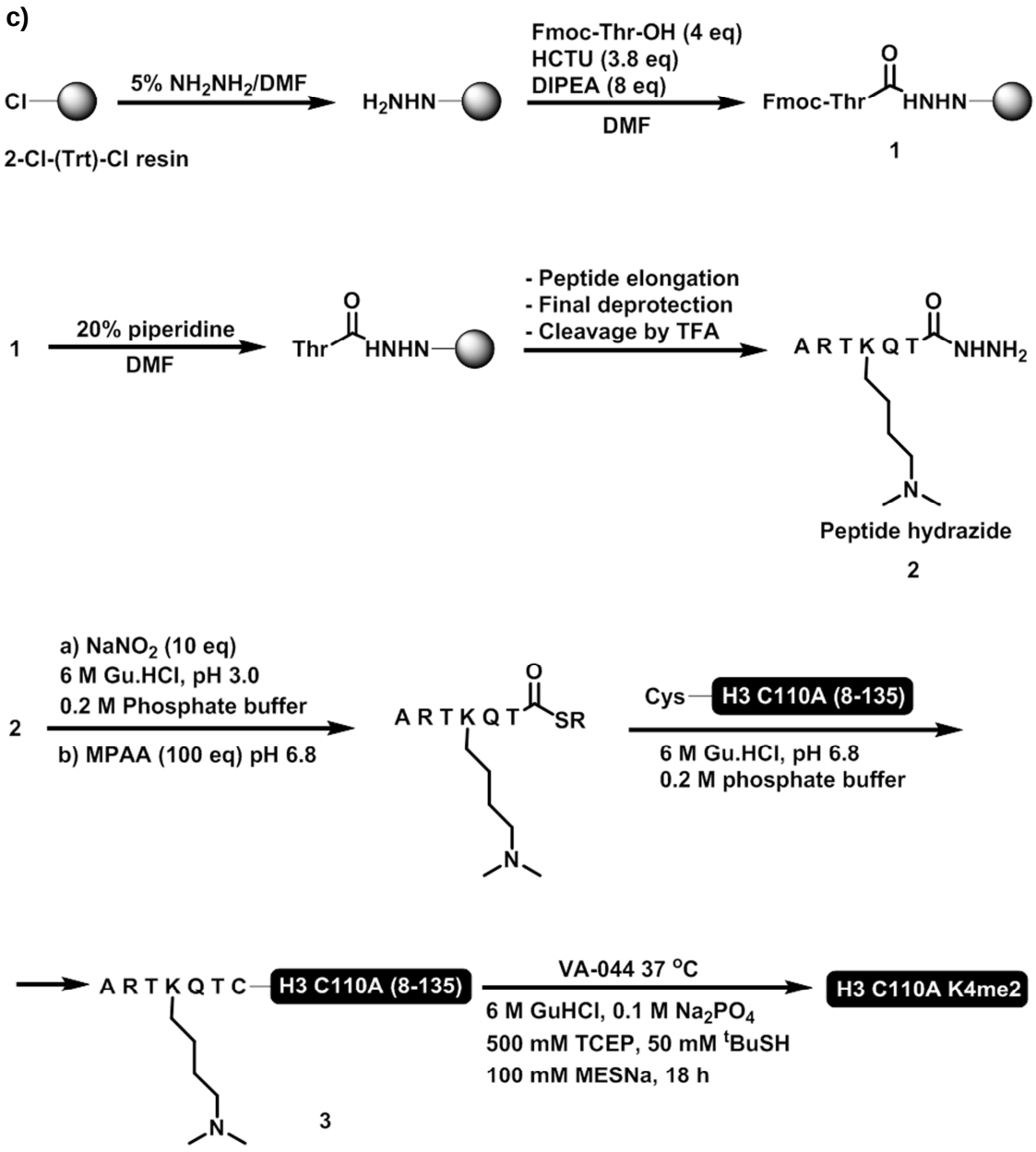
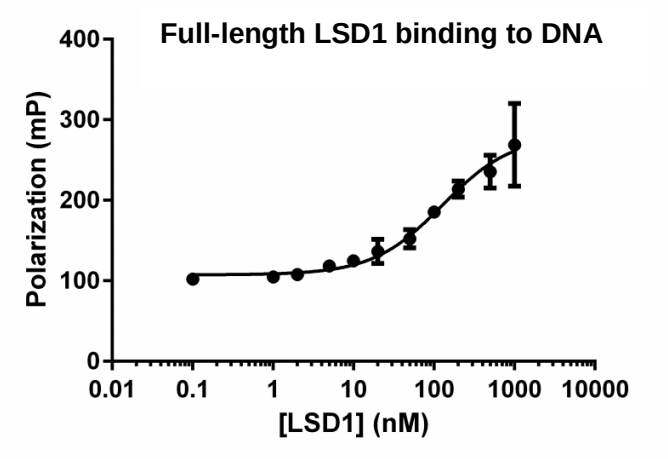


Figure 3.4. Synthesis and detection of full-length native H3 K4me2 and its thia-lysine analog. c) Synthetic scheme for the generation of full-length native H3 K4me2 using native chemical ligation. A 6-mer corresponding to the first six residues of the H3 tail, and containing dimethylation on Lys4, was synthesized on the solid-phase and its C-terminus was oxidized in the presence of external thiol to generate a thioester. NCL was performed with N-terminal cysteine containing H3 C110A (8-135) to generate full-length native H3 K4me2 after desulfurization.

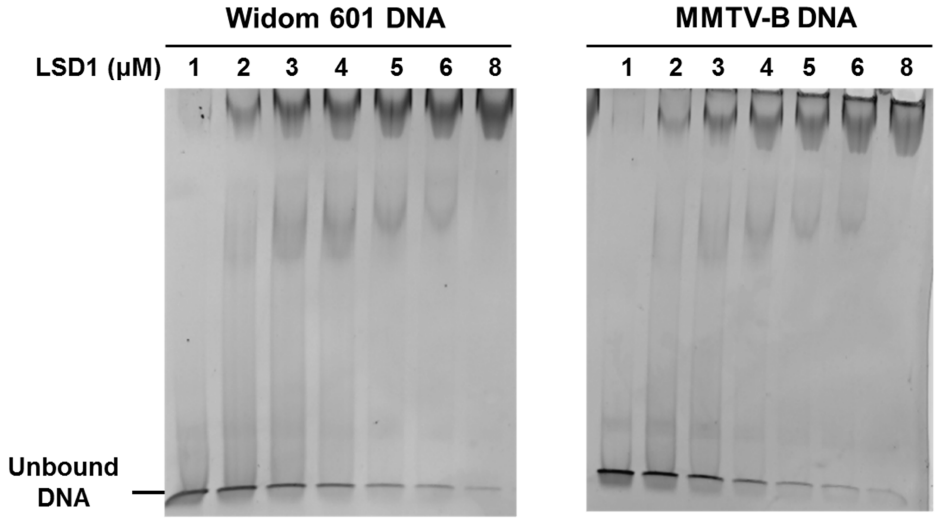
3.2.3 Binding of LSD1 to Linear DNA

By testing peptidic substrates, we established that acetylation in the N-terminal region of histone H3 drastically impedes the catalytic activity of LSD1. This led us to wonder whether the previously reported lack of demethylation activity on nucleosomal substrates could be a consequence of PTMs found in the H3 tail. All the previous demethylation studies employed nucleosomes extracted from mammalian cells which could potentially contain numerous PTMs in their histone tails[12,19]. Given the presence of a SWIRM domain in LSD1, we wondered whether LSD1 was capable of binding unmodified, reconstituted mononucleosomes. We first tested the binding of full length LSD1 with linear DNA by means of fluorescence polarization assays. We generated fluorescently labeled DNA by labeling one of the strands of a 147 bp long double stranded DNA, consisting of the Widom 601 sequence, with a single molecule of AlexaFluor 488 dye. To our surprise, we observed a robust increase in the fluorescence polarization signal upon the incubation of the fluorescent double stranded DNA with increasing amounts of full length LSD1 (K_d = 127 ± 20 nM) (Figure 3.5a). The widely employed N-terminal truncant form of LSD1 (GST-LSD1Δ1-170) failed to show appreciable binding with the same piece of DNA. We further confirmed these results by means of electrophoretic mobility shift assays (EMSA), which also showed clear biding of LSD1 to dsDNA. We discovered that the nature of this binding is independent of the DNA sequence as a 147 bp sequence of DNA based on the mouse mammary tumor virus (MMTV) bound LSD1 ($K_{1/2}$ = 3.4 ± 0.3 μM) as efficiently as the Widom 601 sequence of DNA ($K_{1/2}$ = 3.9 ± 0.4 μM) (Figure 3.5b). Also, the binding of LSD1 to DNA is electrostatic in nature because increasing the ionic concentration to 200 mM NaCl in the binding buffer significantly retarded the gel-shifts ($K_{1/2}$ = 7.24 ± 0.7 μM) (Figure 3.5c) and the fluorescence anisotropy signal.

a)


Full-length LSD1 binding to DNA
Polarization (mP)
400
300
200
100
0
0.01
0.1
1
10
100
1000
10000
[LSD1] (nM)

b)


Widom 601 DNA
MMTV-B DNA
LSD1 (μM)
1 2 3 4 5 6 8
1 2 3 4 5 6 8
Unbound DNA

c)

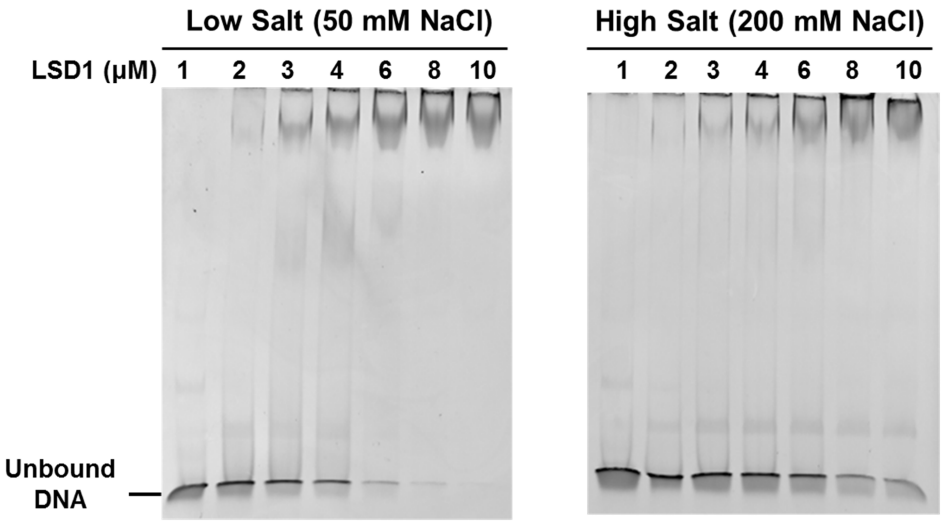


Figure 3.5. Analysis of the binding of full-length LSD1 to 147 bp dsDNA through electrophoretic mobility shift assays and fluorescence anisotropy measurement. a) Increase in fluorescence polarization of a 147 bp dsDNA labeled with Alexfluor488 dye was measured upon incubation with increasing amounts of full-length LSD1. b) Ethidium bromide stained 5% TBE gels showing the binding of full-length LSD1 to two different sequences of 147 bp dsDNA. c) Ethidium bromide stained 5% TBE gels showing the binding of full-length LSD1 to 147 bp dsDNA in buffers containing 50 mM NaCl (low salt condition) or 200 mM NaCl (high salt condition).

3.2.3 Binding of LSD1 to Nucleosomes in the Absence of CoREST

Recent reports have shown the LSD1 is also part of certain histone deacetylase containing complexes that lack any known DNA binding subunits, but are still critical in maintaining heterochromatin in Drosophila[32] and repressing transcription in association with the cohesion establishment factor, Eco1[33]. These reports led us to investigate whether full length LSD1 can demonstrate binding and demethylating activity on reconstituted nucleosomes *in vitro*, independent of any known DNA binding proteins. To this end, we generated a homogeneous pool of NCPs containing native H3K4me2 and other core histones which were recombinantly obtained from *E. coli*, and hence free from all other post-translational modifications. We first tested the binding of LSD1 to NCPs, since we had already observed that LSD1 could bind free linear dsDNA. Increasing amounts of full-length LSD1 were incubated in a solution containing 50 nM NCPs and the binding of LSD1 with NCPs was quantified using a native 5% TBE gel (Figure 3.6a). The electrophoretic mobility shift assays (EMSA) revealed a $K_{1/2}$ of 4.19 ± 0.86 μM for the full-length LSD1 whereas GST-LSD1Δ1-170 failed to show any binding with NCPs (Figure 3.6b). We also observed a salt dependence in the binding of LSD1 to NCPs. At higher salt concentrations of 200 mM NaCl, higher concentration of LSD1 were required to completely saturate NCPs, but there was no discernible difference in the binding pattern at different ionic concentrations. We also performed similar EMSA with NCPs containing wild-type histone H3, free from any modifications in its N-terminal tail. The lack of difference in the binding profile of NCPs containing wt-H3 or H3K4me2 suggested that dimethylation in the H3 tail does not result in increased binding opportunities for LSD1. This result is corroborated by the fact that wt-H3 (1-21) and H3K4me2 (1-21) peptides show very similar K_i and K_M values *in vitro*[34].

In order to precisely quantitate LSD1 binding to nucleosomes, we employed a recently developed MicroScale Thermophoresis (MST) approach. Our MST assay is based on the

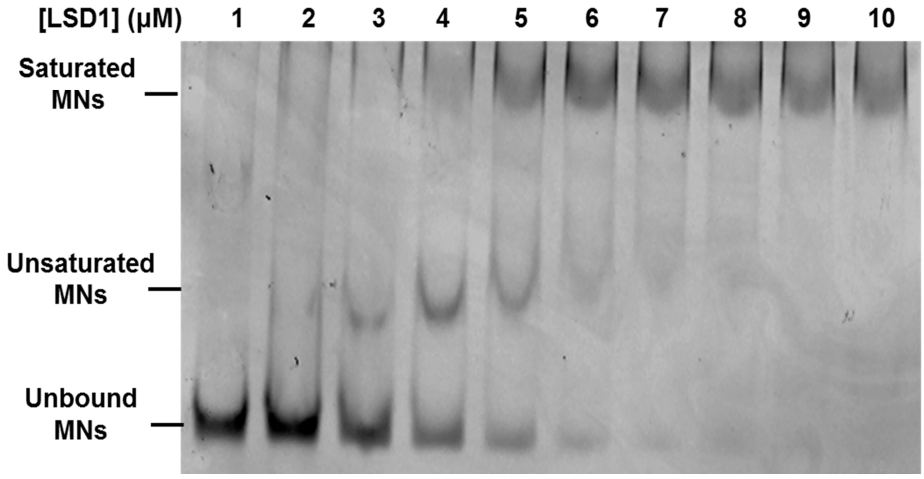
movement of mononucleosomes through temperature gradients, a physical effect called thermophoresis. The thermophoretic movement of nucleosomes varies with their size, charge and hydration shell. Upon LSD1 binding to a nucleosome, one or more of these properties changes leading to a change in movement behavior of the nucleosome, which can be used to calculate the equilibrium dissociation constant, K_d. Using methylated nucleosomes prepared with 601 DNA labeled at its 5'-end with the fluorescent dye Cy5 and by varying concentrations of LSD1 in the MST capillary, we measured a K_d of 6.9 ± 2.0 μM for LSD1 binding to MNs (Figure 3.6c). Furthermore, we calculated a Hill coefficient of 1.49 ± 0.6, which suggests that the binding of one LSd1 molecule may facilitate the binding of another LSD1 molecule to the nucleosome.

We next focused our attention on the binding of LSD1 toward nucleosomes in the presence of neighboring nucleosomes on a single piece of double stranded DNA (Figure 3.6d). To this end, we generated 12-mer nucleosome arrays using the Widom 601_177 bp DNA and histone octamers containing native H3K4me2. The incubation of 50 nM 601 binding sites with increasing amounts of full-length LSD1 resulted in a 4-fold reduction in the $K_{1/2}$ value ($K_{1/2}$ ~4 μM for MNs to ~1 μM for arrays), suggesting an avidity effect which could be more pronounced *in vivo* due to the large amount of chromatin.

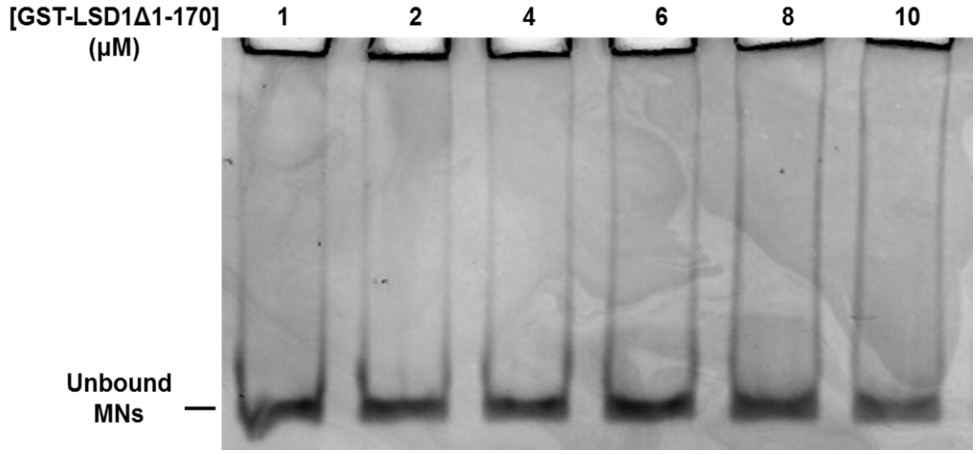
3.2.4 LSD1 Effectively Demethylates NCPs and Nucleosomal Arrays

We performed demethylation assays under single-turnover conditions to determine the rate at which full length LSD1 demethylates NCPs containing native H3 K4me2 independent of CoREST. The demethylation was followed by the means of quantitative western blot analysis using an antibody specific toward H3K4me2.

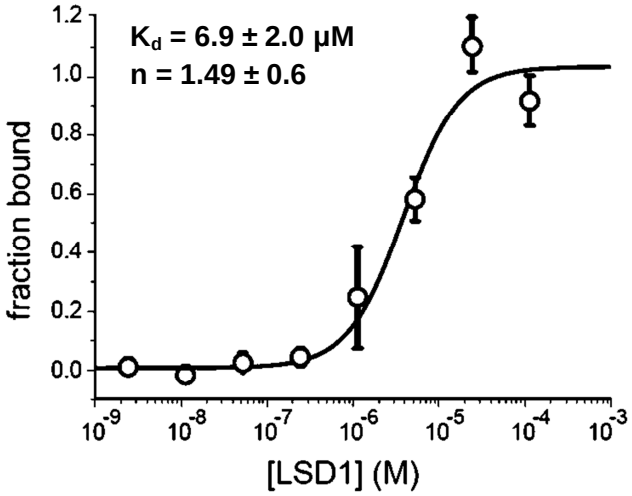
a)


[LSD1] (μM)
1
2
3
4
5
6
7
8
9
10
Saturated MNs
Unsaturated MNs
Unbound MNs

b)


[GST-LSD1Δ1-170] (μM)
1
2
4
6
8
10
Unbound MNs

c)



d)

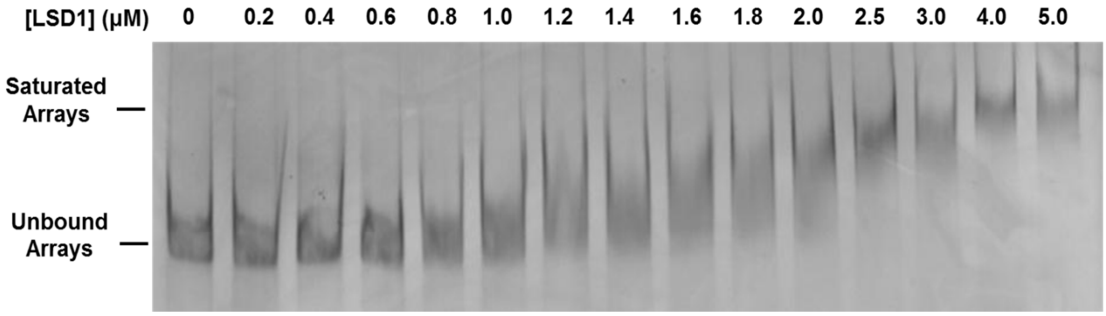


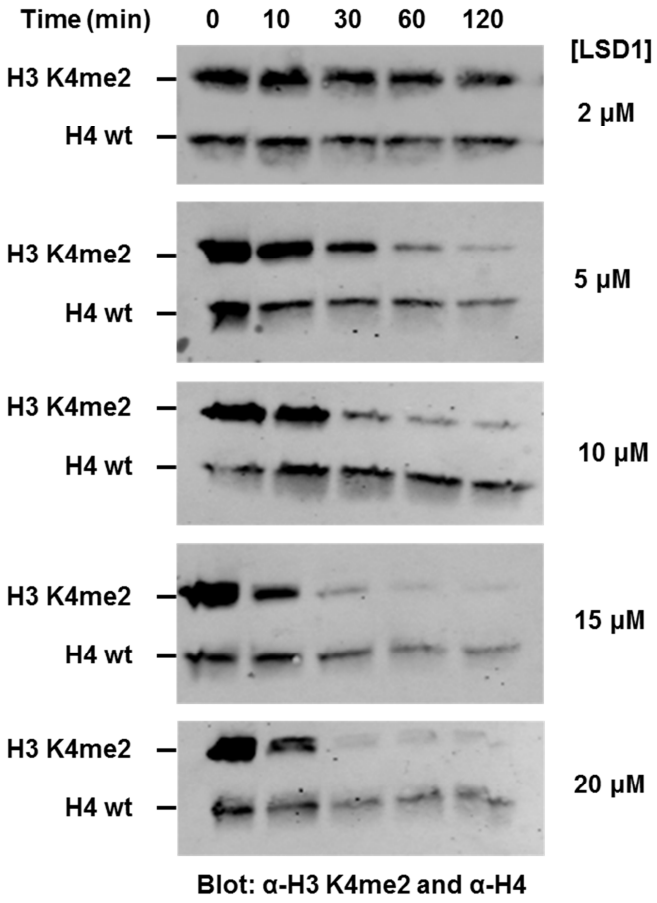
Figure 3.6. Analysis of the binding of full-length LSD1 to mononucleosomes and 12-mer nucleosome arrays through electrophoretic mobility shift assays and MicroScale thermophoresis (MST). a) Ethidium bromide stained 5% TBE gel showing the binding of full-length LSD1 to mononucleosomes (MNs). b) Ethidium bromide stained 5% TBE gel showing poor binding of GST-LSD1Δ1-170 to MNs. c) Binding curve of full-length LSD1 to fluorescently labeled mononucleosomes as measured by MST experiments. n=3, error bars show s.e.m. d) Ethidium bromide stained 1% APAGE gel showing the binding of full-length LSD1 to nucleosome arrays.

Nucleosomal substrates (50 nM) were incubated with varying concentrations of LSD1 (2 to 20 uM) and the reduction in H3K4me2 signal was measured as a function of time (Figure 3.7a). The observed rate of demethylation (k_{obs}) (Figure 3.7b) was plotted against LSD1 concentration and the graph was fit to $k_{obs} = (k_{max}X)/(X+K'_M)$ using non-linear regression to obtain an apparent K_M (K'_M) of 6.2 ± 4.9 μM and the maximum rate of catalysis as (k_{max}) 0.093 ± 0.02 min^{-1} (Figure 3.7c). We observed a strong dependence of LSD1 catalysis on the amount of salt in the demethylation assay. The rate of demethylation was greatly reduced when the amount of KCl in the demethylation assay was increased to 200 mM (Figure 3.7d). This is consistent with the decreased binding of LSD1 to MNs at high salt concentrations. We also performed demethylation of nucleosomal arrays but did not observe an increase in demethylation rates, possibly due to increase in off target binding to the 30 bp linker DNA between individual nucleosomes. We confirmed that DNA binding is essential for demethylation by performing some competition assays mentioned in the next section.

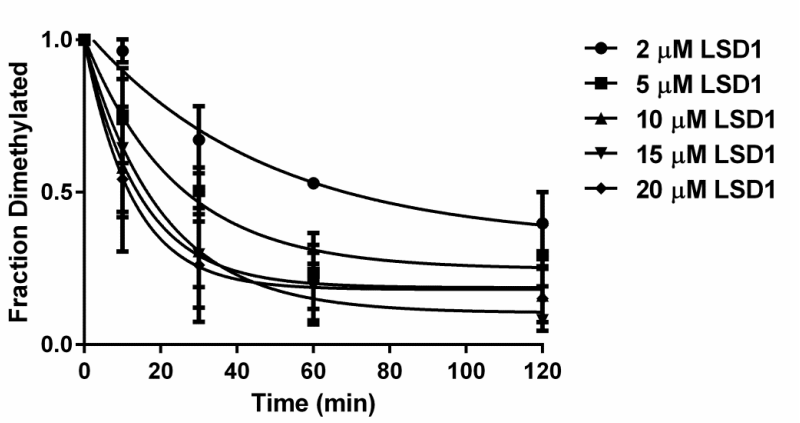
3.2.5 Mechanism of LSD1 Binding to MNs

We wanted to determine whether the demethylation activity of LSD1 was dependent on its nucleosome binding or if LSD1 could also access the protruding H3 tail freely in solution. We decided to compete the nucleosome binding of LSD1 with a member of the high motility group N (HMGN) family of proteins, which are known to bind nucleosomes through a highly conserved nucleosomal binding domain (NBD). In addition to the core sequence of the NBD, which interacts with the acidic patch formed by the H2A and H2B histone dimer, all HMGNs contain a negatively charged regulatory domain (RD) in their C-termini[35].

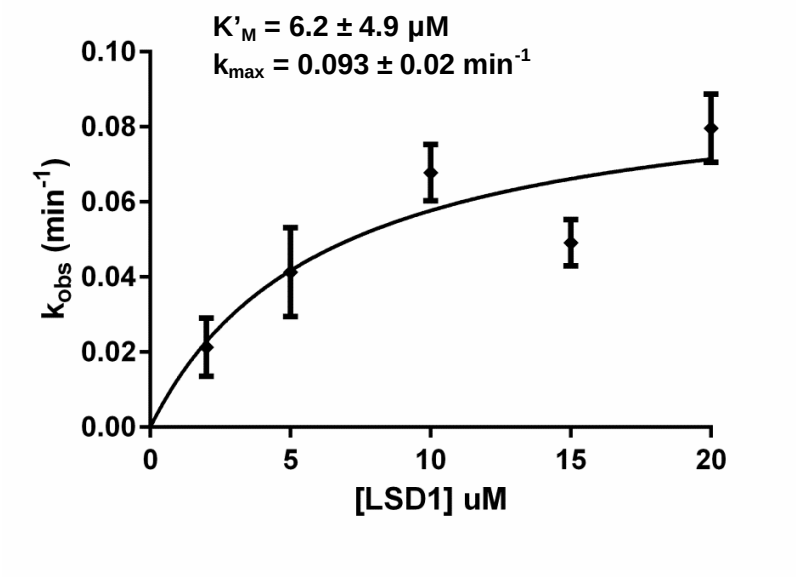
a)


Time (min)
0
10
30
60
120
[LSD1]
H3 K4me2
H4 wt
2 µM
5 µM
10 µM
15 µM
20 µM
Blot: α-H3 K4me2 and α-H4

b)


Fraction Dimethylated
1.0
0.5
0.0
0
20
40
60
80
100
120
Time (min)
2 µM LSD1
5 µM LSD1
10 µM LSD1
15 µM LSD1
20 µM LSD1

c)



d)

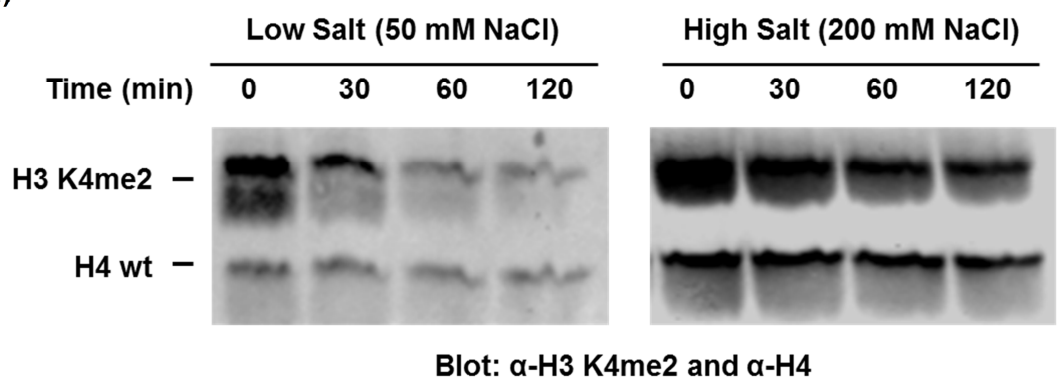


Figure 3.7. Demethylation of mononucleosomes by full-length LSD1 in the absence of CoREST. a) Western-blot showing the amount of dimethylated H3 K4 present in the assay mixture at various time points when full-length LSD1 is incubated with H3 K4me2 containing MNs. Wt-H4 is used as an internal standard to normalize the measured signal for H3 K4me2. b) Rates of demethylation (k_{obs}) at various concentrations of LSD1 were determined by plotting the fraction of dimethylated H3 left in the solution at various time points. c) Kinetic parameters for demethylation of mononucleosomes by LSD1 were determined by plotting the rates of demethylation against the concentrations of LSD1 used in the demethylation assays. d) Western-blot showing the salt dependence of LSD1 mediated demethylation of mononucleosomes. 5 µM LSD1 was incubated with 30 nM MNs at 25 ˚C in identical buffers containing two different NaCl concentrations.

One particular ~10 kDa member of this family, HMG14 (HMGN2), binds to each side of the nucleosome core with the C-terminal portion of the NBD making contacts with DNA near the nucleosomal entry/exit point, while the unstructured C-terminal region of HMG14 protrudes beyond the periphery of the nucleosomal DNA and can interact with the adjacent H3 tail[36]. We purified full-length HMG14 and confirmed its binding with MNs through gel-shift assays. Upon addition of 30 – 300 nM HMG14 to a demethylating assay containing 5 uM LSD1 and 30 nM MNs, we observed a dose-independent inhibition of the demethylating activity (Figure 3.8a). We hypothesized that the dose independent nature of the inhibition could be a result of HMG14's binding to the H3 tail which would occlude LSD1 from binding its target. To rectify this, we synthesized a 29 amino acid long peptide based on the sequence of the highly conserved NBD of HMG14 (Figure 3.4.1c). Unlike full-length HMG14, upon addition of low concentrations (15 nM – 300 nM) of the NBD containing 29-mer peptide, we observed no inhibition of demethylation owing to the lack of H3 tail binding capacity of the peptide (Figure 3.8b). However, at higher concentrations of 5 – 15 μM, the NBD peptide successfully competed away the binding of 5 μM LSD1 to 30 nM MNs, and a dose dependent decrease in demethylation was observed (Figure 3.8b). We confirmed that highest concentration of the NBD used was not detrimental to LSD1 activity as demethylation was observed for dimethylated H3 (1-21) peptides in the presence of 20 μM NBD peptide derived from HMG14.

3.2.6 Effect of CoREST on LSD1 Activity Toward MNs

Full-length CoREST shows poor stability at low ionic concentrations unlike the widely used CoRESTΔ281. All the demethylation assays employing CoREST were performed at 200 mM NaCl. At this ionic concentration, LSD1 shows poor demethylation activity with peptides and MNs as substrates.

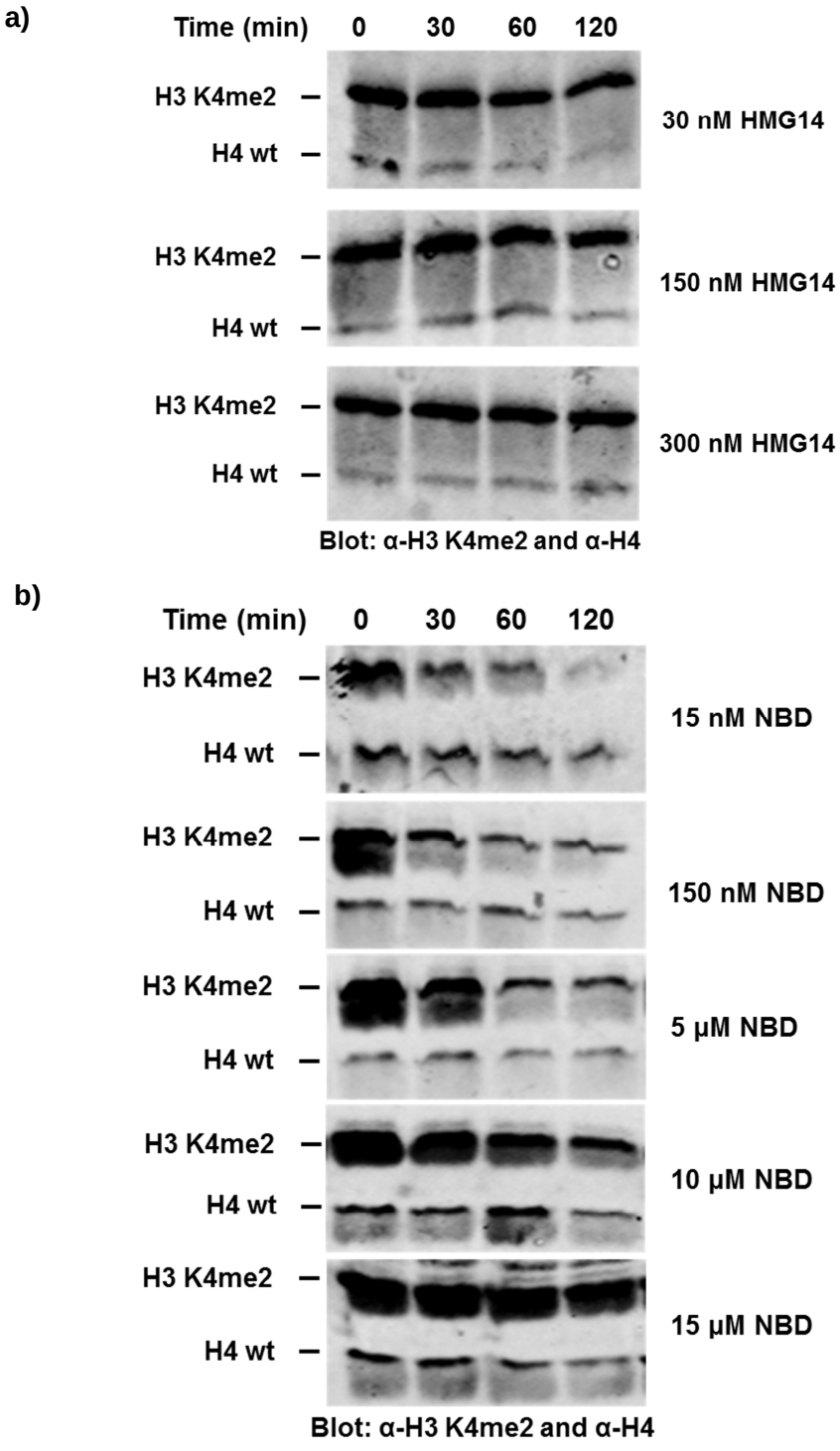


Figure 3.8. Demethylation of mononucleosomes by LSD1 in the presence of full-length HMG14 or its nucleosome binding domain (NBD). a) Western-blot showing the dose-independent inhibition of demethylation of MNs by the full-length HMG14 protein. b) Western-blot showing the dose dependent inhibition of demethylation of MNs by NBD of HMG14.

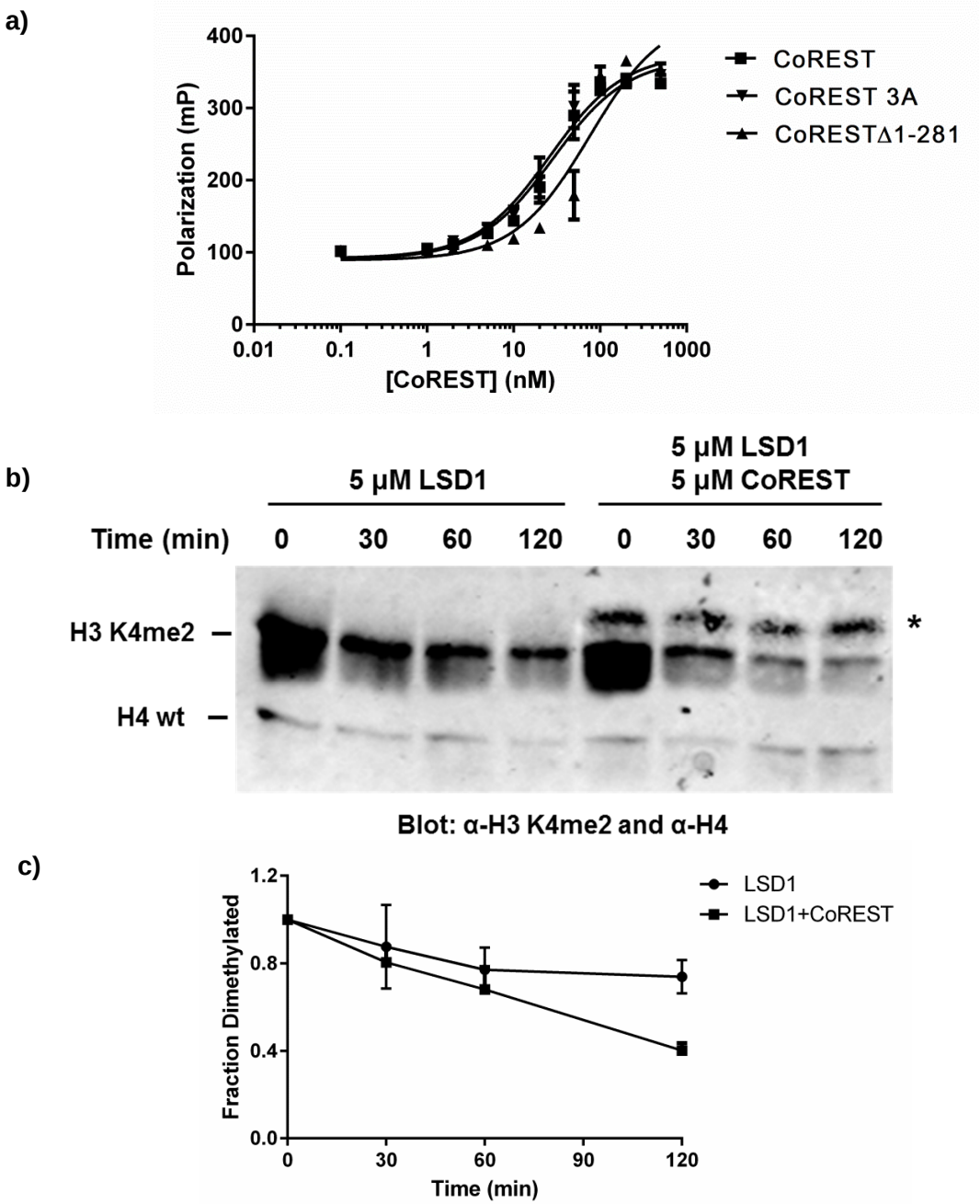


Figure 3.9. Effect of CoREST on the demethylation of MNs by LSD1. a) Increase in fluorescence polarization of a 147 bp dsDNA labeled with Alexfluor488 dye was measured upon incubation with increasing amounts of CoREST, CoRESTΔ1-281, or a mutant of CoREST incapable of binding SUMO-3 (CoREST3A). b) Western-blot showing the demethylation of MNs by LSD1 in the absence and presence of CoREST at 200 mM NaCl. Asterisk denotes a non-specific band from CoREST. c) Quantification of the amount of dimethylated H3 remaining in the demethylation assays of MNs with LSD1 in the presence and absence of CoREST. n=3, error bars show s.e.m.

We investigated whether CoREST can overcome this limitation. We first confirmed the binding of CoREST to DNA under high ionic concentrations of 200 mM NaCl. Employing the fluorescence polarization assays described in section 3.2.3, we observed significant decrease in the binding of LSD1 to DNA at 200 mM NaCl, whereas full length CoREST (K_d = 28 ± 7 nM) maintained robust binding similar to CoRESTΔ281 (K_d = 74 ± 38 nM) (Figure 3.9a). The decrease in the DNA binding of LSD1 explains the poor demethylation of MNs observed by us at high ionic concentrations when CoREST is absent from the demethylation assays. The addition of full-length CoREST rescued the rate of demethylation and faster demethylation of MNs was observed at 5 μM concentration of LSD1/CoREST complex as compared to 5 μM LSD1 at 200 mM NaCl (Figure 3.9b).

3.2.7 Generation of Isopeptide Linked SUMO-3-H4

We had previously employed disulfide linked SUMO-3-H4 ($suH4_{ss}$) for our biophysical studies of sumoylated chromatin. Unfortunately, the requirement of reducing agents in the demethylation assays precluded the use of $suH4_{ss}$ as the disulfide bond is susceptible to reduction. Our lab developed a novel synthetic strategy to generate isopeptide linked SUMO-3-H4 (suH4) which could be readily employed in biochemical studies. I performed the cloning, overexpression and purification of the recombinant proteins and the rest of the methodology was developed by Caroline Weller in the lab. We synthesized a H4 (1-14) peptidyl hydrazide, using Fmoc chemistry based SPPS, with Gly92 of SUMO-3 and the ligation auxiliary attached to Lys12 as part of the peptide fragment. The cleaved and deprotected peptide was ligated to the SUMO-3(2-91)C47S-α-thioester to yield the sumoylated peptide. Native chemical ligation was performed between the sumoylated H4 peptide α-thioester and a truncated H4(15-102) protein containing the A15C mutation at its N-terminus.

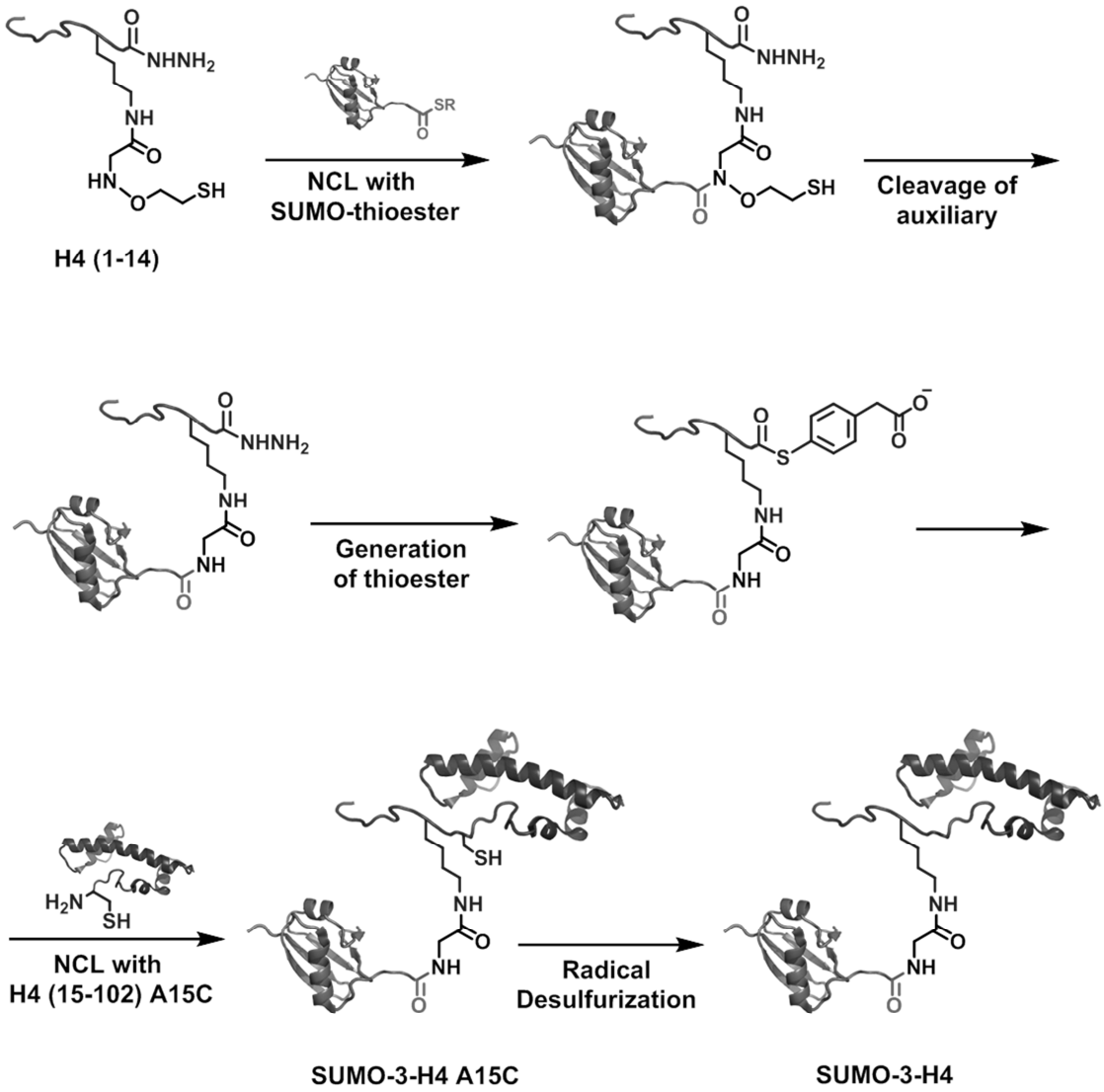


Figure 3.10. Scheme depicting the strategy for the semisynthesis of isopeptide linked SUMO-3-H4. N-terminal portion of H4 is synthesized using Fmoc-based SPPS with an auxiliary present at the site of SUMO attachment. NCL of the peptide with a SUMO C-terminal thioester results in a sumoylated peptide. A second NCL with the rest of the H4 protein is performed to generate a ligation product containing SUMO and H4 proteins. Cleavage of the auxiliary and desulfurization of the cysteine at the site of ligation results in a native SUMO-3-H4.

At the end of the 24 h NCL reaction the N-O bond between the auxiliary and the protein was cleaved at 25 ˚C for 24 h. As the final step of the synthesis, the full-length sumoylated histone H4 A15C mutant was desulfurized to yield the desired suH4 protein in multimilligram quantities (Figure 3.10).

3.2.8 Effect of Sumoylation on Demethylation

We simultaneously incorporated suH4 and H3 K4me2 into nucleosomes and performed demethylation assays with the LSD1/CoREST complex. Faster demethylation was observed for sumoylated MNs in comparison with non-sumoylated MNs. To confirm that the faster demethylation was due to the binding of CoREST1 with SUMO, we also performed the demethylation assays with a triple alanine mutant of CoREST, referred to as CoREST3A, which is incapable of binding to SUMO. Indeed, CoREST3A failed to recapitulate the faster demethylation observed in case of wild-type CoREST (Figure 3.11a). We calculated the p-values for the observed demethylation in case of sumoylated MNs in comparison to non-sumoylated MNs. At each time point the percentage of demethylation observed was significantly different ($p < 0.05$) (Figure 3.11b). The initial faster rate of demethylation of sumoylated MNs even with CoREST3A can be attributed to the intrinsic property of sumoylation to reduce aggregation of chromatin which would make it easier to access methylated MNs. The mutations did not affect the binding of CoREST3A to DNA, as determined by fluorescence polarization measurements ($K_d = 28 \pm 7$ nM for CoREST ; $K_d = 25.8 \pm 5.8$ nM for CoREST3A) (Figure 3.9a). These results provide evidence of a possible *trans* effect of sumoylation by which it can mediate gene repression.

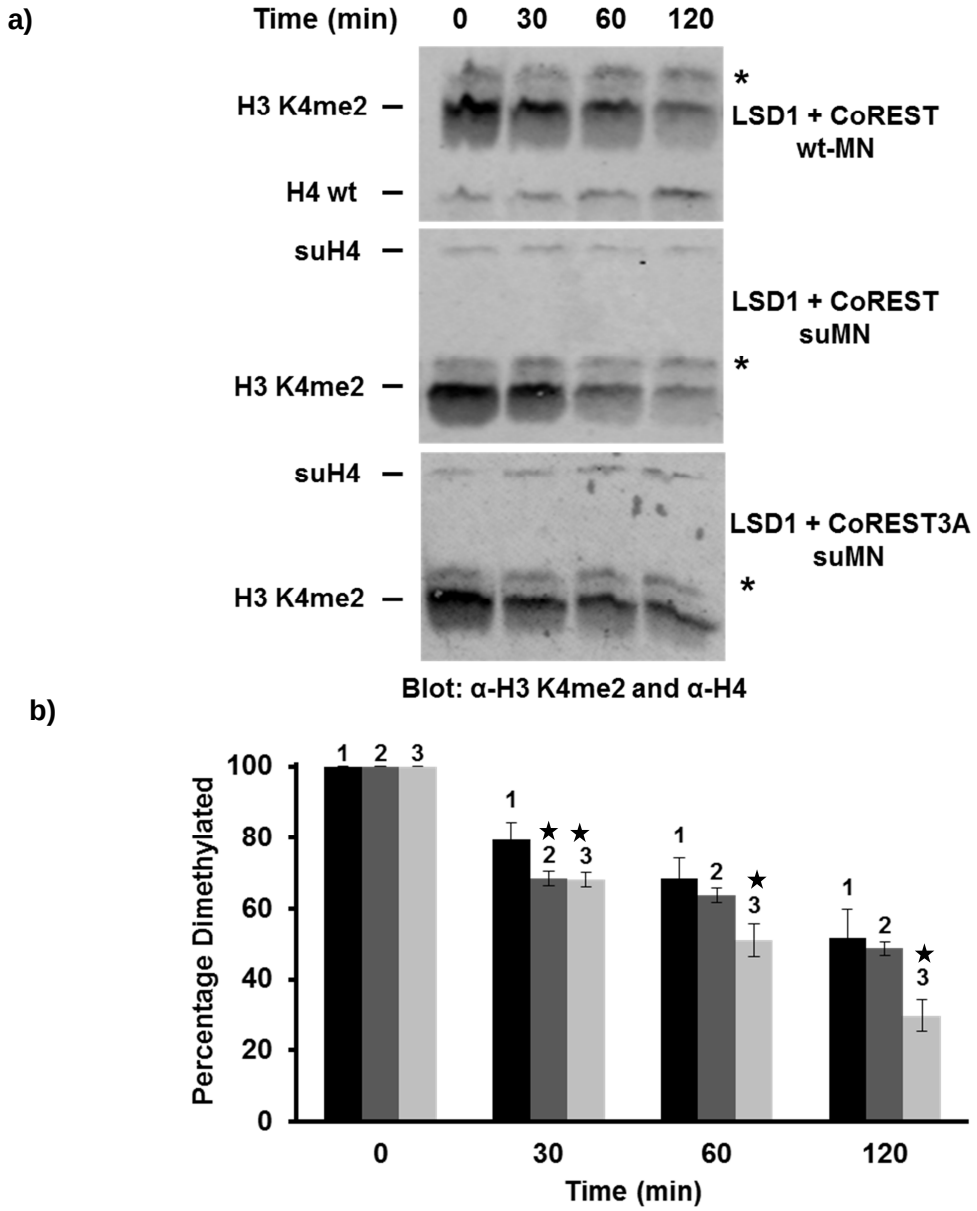


Figure 3.11. Effect of Sumoylation on the demethylation of MNs by the LSD1/CoREST complex. a) Western-blot showing the demethylation of sumoylated or non-sumoylated MNs by LSD1/CoREST complex. Last row shows the demethylation of sumoylated MNs by LSD1 in the presence of a CoREST mutant (CoREST3A) which is incapable of binding SUMO. Asterisk denotes a non-specific band from CoREST. b) Lane 1,2 and 3 denote demethylation of non-sumoylated MNs with LSD1/CoREST, demethylation of sumoylated MNs with LSD1/CoREST and demethylation of sumoylated MNs with LSD1/CoREST3A respectively. n=3 to 6, error bars show s.e.m. Asterisk denotes a p-value of less than 0.05 in comparison to Lane 1.

3.3 CONCLUSION AND OUTLOOK

Even decades after the discovery of SUMO as a post-translational modification[37], our understanding of the functional effects of this modification is limited. This is in part due to the transient nature of this modification and its low endogenous levels. The presence of SUMO in all eukaryotes and the fact that complex organisms have higher number of SUMO isoforms suggests that this modification is important in basic cellular functions[38]. Traditionally SUMO is best known as a modification to alter the localization and/or the activity of an enzyme, but recently we have started to uncover its role in DNA damage repair[39], maintaining homeostasis under oxidative and temperature induced stress[40], and regulating gene expression[3]. Unlike the ubiquitylation pathway, only a small number of SUMO E3 ligases have been identified and only a single E2 conjugating enzyme is known. This has limited our ability to generate sumoylated substrates *in vitro* for biochemical and biophysical analyses. Also, the simultaneous presence of multiple PTMs inside a cell makes is impossible to isolate homogeneous, site-specifically sumoylated substrates. Hence, to further investigate recent reports of a possible role of sumoylation in transcriptional repression, we made use of chemical biology tools to generate a homogeneous sample of isopeptide linked SUMO-3-H4 (suH4). We probed whether sumoylation could enhance demethylation of H3K4me2, a mark associated with actively transcribed promoters. The LSD1/CoREST repressive complex has been shown to be localized to regions of sumoylation but the molecular mechanisms of its functioning were still unknown[11]. We performed the first ever semi-synthesis of the full-length H3K4me2 and incorporated it into nucleosomes with and without suH4. Upon comparing the rate of demethylation in the presence and absence of sumoylation we discovered that SUMO facilitates the removal of this activating mark. We hypothesize that this effect is possibly due to the increased residence time of the LSD1/CoREST complex on to the nucleosome due to the presence of additional binding site for the sumo interaction motif (SIM) in CoREST. It is also possible that the interaction of SUMO on

the H4 tail with the SIM in CoREST orients the complex such that it can access the H3 tail easily, and hence result in faster demethylation.

This interesting cross-talk between sumoylation and demethylation has opened avenues for further studies into the pathways by which SUMO can mediate repression. Another component of the LSD1/CoREST complex, the histone deacetylase HDAC1/2 has well established roles in transcriptional repression. As the next step, we will investigate the effect of sumoylation on deacetylation. It is likely that the presence of SUMO at Lys12 of H4 will not allow the deacetylation of H4K16ac, the primary PTM associated with active transcription, but acetylation on the H3 tail at Lys9 and Lys14, which are also associated with activation can be removed. Another repressive complex, the Mi-2/NuRD complex also contains a SIM[11]. It would also be interesting to investigate the ATP-dependent remodeling activity of this complex in the presence of sumoylation. We have initiated some of these investigations in our lab to uncover the chromatin specific roles of SUMO.

3.4 EPERIMENTAL PROCEDURES

3.4.1 Overexpression and purification of full-length hLSD1 (KDM1A isoform b)

The pET15b-LSD1 plasmid was a kind gift from Dr. Yi Zhang. *E. coli* BL21(DE3) cells were transformed using the plasmid and cell stocks for pET15b-LSD1 expressing cells were made according to the protocol mentioned in section 2.4.1. For overexpression of His_6-LSD1, cells were grown at 37 ^oC in YT media containing 100 µg/mL ampicillin to an OD_{600} of ~1.0 and were cooled down to 25 ^oC before induction. Protein expression was induced by the addition of 0.3 mM IPTG to the growth media and the cells were further grown at 25 ^oC for 4 h. At the end of the induction period, the cells were harvested by centrifugation at 7,000xg, resuspended in PBS, 0.2 mM PMSF, 5% glycerol buffer and lysed by sonication till the viscosity of the solution

was reduced along with an observed change in color. The lysate was centrifuged at 20,000xg and the supernatant was applied to a nickel column at 4 ^{o}C for 1.5 h. Column was washed with 20 and 50 mM imidazole containing lysis buffer, 5 column volumes (CV) each, and full-length His_6-LSD1 eluted with 250 mM imidazole containing lysis buffer. Pure fractions were dialyzed against 20 mM Tris, pH7.5, 5% glycerol buffer for 3 h and further purified by anion exchange chromatography. The fractions were visualized on a 12% SDS-PAGE gel and the fractions containing the desired protein in 90% purity were combined and concentrated. The amount of active LSD1 was determined by using the FAD extinction coefficient at 458 nm as 10,790 $cm^{-1}M^{-1}$. The total protein concentration was determined using a Bradford assay. About 70% of the total protein was typically purified in the active form, based on co-factor occupancy.

3.4.2 Overexpression and Purification of full-length CoREST

The pET28b-CoREST plasmid was a kind gift from Dr. Grace Gill. *E. coli* BL21(DE3) cells were transformed using the plasmid and cell stocks for pET28b-CoREST expressing cells were made according to the protocol mentioned in 2.4.1. For overexpression of His_6-CoREST, cells were grown at 37 ^{o}C in YT media containing 30 µg/mL kanamycin to an OD_{600} of 0.5 and were cooled down to 16 ^{o}C. Protein expression was induced by the addition of 0.2 mM IPTG to the growth media and the cells were further grown at 16 ^{o}C for 6 h. At the end of the induction period, the cells were harvested by centrifugation at 7,000xg, resuspended in 50 mM Tris, pH 8, 500 mM NaCl, 0.2 mM PMSF, 10% glycerol buffer and lysed by sonication till the viscosity of the solution was reduced along with an observed change in color. The lysate was centrifuged at 20,000xg and the supernatant was applied to a nickel column at 4 ^{o}C for 1.5 h. Column was washed with 25 mM imidazole and 0.1% Triton-X containing lysis buffer and further washed with 100 mM imidazole containing lysis buffer (5 CV each). His_6-CoREST was eluted with 250 mM imidazole containing lysis buffer. The fractions were analyzed using a 12% SDS-PAGE gel and clean

fractions were dialyzed against 20 mM Tris, pH 8, 100 mM NaCl, 10% glycerol buffer for 2 h at 4 ^{o}C. At the end of the dialysis some precipitation of CoREST was observed and the soluble protein was further purified by anion exchange chromatography. The cleanest fractions were combined and concentrated. The amount of full-length CoREST was estimated using the extinction coefficient at 280 nm as 52,035 $cm^{-1}M^{-1}$ and also confirmed by Bradford assay.

3.4.3 Generation of human H3 K4C

Human histone H3 C110A was cloned between the NdeI and BamHI restriction sites in the pET15b vector with a Tobacco Etch Virus (TEV) protease cleavage sequence between the His_6-tag and the H3 N-terminus. Site-directed mutagenesis was then employed to generate the mutants H3 K4C and H3 (7-135)A7C from this plasmid. The primers used to perform this mutagenesis are mentioned in Table 1. *E. coli* BL21(DE3) cells were transformed with the mutant plasmids according to protocol mentioned in 2.4.1 and grown at 37 ˚C in 6 L YT medium until OD_{600} ~0.7. Protein expression was induced by the addition of 0.3 mM IPTG for 3 h at 37 ˚C. Cells were harvested by centrifugation at 7000xg, resuspended in 150 mM NaCl, 50 mM Tris, pH 7.5 and lysed by sonication. The lysate was centrifuged at 20,000xg for 20 min and insoluble histones were recovered from inclusion bodies with an extraction buffer consisting of 6 M guanadinium chloride (Gn-HCl), 100 mM NaCl, 50 mM Tris, pH 7.5. The re-solubilized histones were purified by Ni^{2+}-affinity chromatography. Briefly, proteins were bound to the Ni^{2+}-column overnight at 4 ˚C with gentle shaking and the column subsequently washed with 3 volumes of extraction buffer containing 25 mM Imidazole. Immobilized proteins were eluted using 500 mM imidazole in the extraction buffer. Histone-containing fractions were identified by 15% SDS-PAGE, combined and dialyzed into 1 mM DTT overnight at 4 ˚C. Purified His_6-TEV protease was then added in a 1:10 molar ratio (enzyme:substrate) and His_6-tag cleavage was performed overnight in a buffer containing 1 mM EDTA, 10 mM DTT, 10 mM Cys, 50 mM Tris,

pH 6.9. The cleavage products were dialyzed into extraction buffer and applied to a Ni^{2+}-column. This led to retention of the cleaved His6-tag, His_6-TEV and any uncleaved histones on the column. The eluted tag-less H3 K4C and H3 (7-135)A7C were further purified to homogeneity by RP-HPLC and characterized by ESI-MS (Calculated m/z $[M+H]^+$ = 15,201.7 Da, Observed = 15,200.54 ± 1.2 Da).

3.4.4 Generation of Thialysine Analog of H3 K4me2

The methyl lysine analog (MLA) was synthesized according to the protocol published by Simon et. al[31]. Briefly, Lyophilized H3 K4C (5 mg) was dissolved in 900 µL alkylation buffer, 1 M HEPES pH 7.8, 4 M Gn-HCl, 10 mM D/L-methionine; and DTT was added (20 µl of 1 M, dissolved immediately before use). The histones were reduced for 1 hr at 37 ˚C and (2-chloroethyl)-dimethylammonium chloride was added (50 µL of 1 M, dissolved immediately before use) to perform the alkylation. After the reaction proceeded for 2 hr at room temperature additional DTT was added (10 ml of 1 M) to reduce any oxidized histone H3 K4C. The reaction incubated at room temperature for 30 min and treated with additional alkylating agent (50 mL, 1M dissolved immediately before use) and allowed to proceed for an additional 2 hr at room temperature. The reaction was quenched with beta-mercaptoethanol (50 µL, 14.2 M) and then purified to homogeneity by RP-HPLC and characterized by ESI-MS (Calculated m/z $[M+H]^+$ = 15,271.7 Da, Observed = 15,271.54 ± 1.4 Da).

3.4.4 Generation of H3 K4me2

We made use of the expressed protein ligation strategy to synthesize H3 K4me2. Using SPPS we synthesized a 6-mer H3 N-terminal peptide (ARTKme2QT) containing dimethylated Lys4 and a C-terminal hydrazide. Briefly, amino acids were coupled to the 2-Cl-(Trt)-$NHNH_2$ using the

Liberty Blue microwave assisted peptide synthesizer (CEM corporation). The peptide was cleaved from the resin and low-temperature diazotization was performed to generate a C-terminal azide. Briefly, the peptide was dissolved in 0.2 ml of 0.2 M phosphate solution containing 6 M Gn-HCl, pH 3.0 and incubated at -20 ˚C. On the other hand, 2 mg of H3 (7-135)A7C prepared in 3.4.3 was mixed with 2 mg of 4-mercaptophenylacetic acid (MPAA) and dissolved in 0.2 ml of 0.2 M phosphate solution containing 6 M Gn·HCl, pH 3.0 by vortexing. The pH of this mixture was adjusted to 6.5 with 6 M NaOH and the solution was cooled to -20 ˚C. To oxidize the peptide hydrazide to the corresponding azide, 20 µL of 0.5 M $NaNO_2$ was added into the solution and gently agitated for 15 min at -20 ˚C. The 6-mer was then ligated to H3 (7-135)A7C polypeptide by mixing the two solution together and incubating them at room temperature for 30 min after adjusting the pH to 6.8 – 7.0. This yielded the full-length H3 K4me2 A7C mutant histone, which was desulfurized in 0.2 ml of 1.0 M TCEP, 40 µL of tBuSH and 20 µL of 0.1 M VA-044 solution at 37 ˚C for 2 h to yield wild-type H3 K4me2 in 17% yield. The product was purified to homogeneity by RP-HPLC and characterized by ESI-MS (Calculated m/z $[M+H]^+$ = 15,256.7 Da, Observed = 15,256.66 ± 0.67 Da) (Figure 3.4.1a).

3.4.5 Reconstitution of NCPs and 12-mer Nucleosome Arrays

NCPs and 12-mer nucleosome arrays were reconstituted as previously described in section 2.2.3. Briefly, individual human histones were overexpressed in *E. coli* BL21 (DE3) cells and purified by size exclusion chromatography on a Superdex 200 column followed by reverse phase high performance liquid chromatography (RP-HPLC). Histones were combined in equimolar amounts to generate octamers containing native H3K4me2.

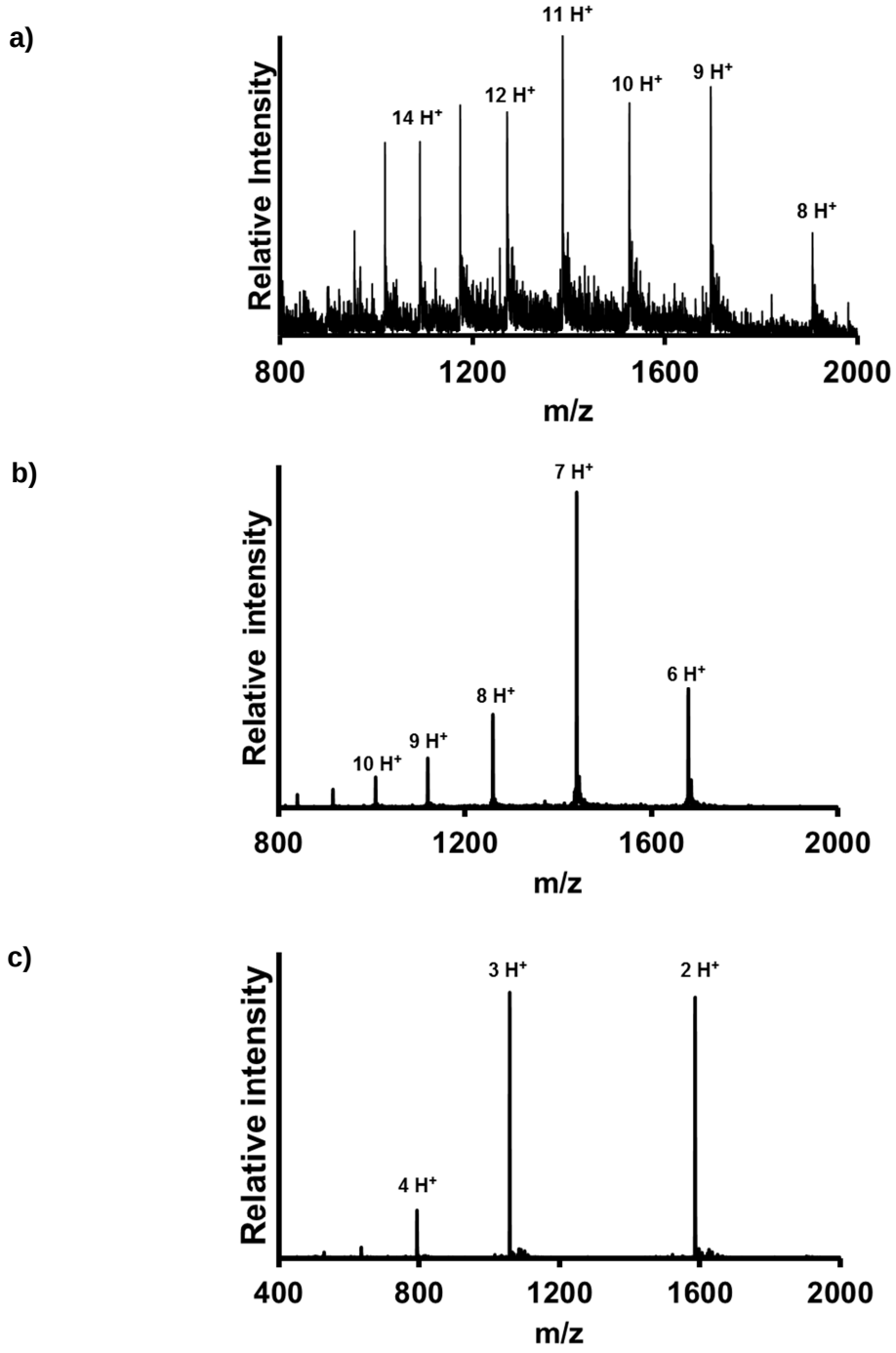


Figure 3.4.1. Characterization of H3 K4me2, H4 Δ14 A15C and NBD of HMG14. a) ESI-MS of H3 K4me2, calculated m/z $[M+H]^+$ 15,256.7 Da, observed 15,256.66 ± 0.67 Da. b) ESI-MS of H4 Δ14 A15C, calculated m/z $[M+H]^+$ 10,072.71 Da, observed 10,072.71 ± 1.2 Da. c) ESI-MS of 29 residues long nucleosome binding domain (NBD) of HMG14, calculated m/z $[M+H]^+$ 3,170.1 Da, observed 3,170.3 ± 0.6 Da.

Octamers and 147 bp DNA (1_147_601) were combined in high-salt refolding buffer followed by dialysis into low-salt buffer to generate NCPs which were verified by resolution on a Criterion 5% TBE gel run in 0.5x TBE buffer, followed by staining with ethidium bromide.

3.4.6 Microscale Thermophoresis and Electrophoretic Mobility Shift Assays

A buffered solution (50 mM HEPES, 1 mM DTT, 50 mM NaCl, 5% glycerol, pH 8.0) containing 50 nM final concentration of NCPs or 601 binding sites (in case of 12-mer arrays), was incubated with varying concentrations of LSD1 (0 to 12 μM) at room-temperature for 30 min. In case of NCPs, the mixtures were visualized on a 5% TBE gel run at room-temperature at 140 V for 50 min in 0.5x TBE buffer, fol-lowed by ethidium bromide staining. For 12-mer arrays, the mixtures were visualized on a 1% APAGE gel run at room-temperature at 90 V for 60 min in 0.25x TAE buffer, followed by ethidium bromide staining. The unbound fraction of NCPs or 601 binding sites was quantified at each concentration of LSD1. The $K_{1/2}$ values were determined by plotting the bound fraction against LSD1 concentration and fitting the graph to the following equation: Fraction bound = $[LSD1]^n/[LSD1]^n + (K_{1/2})^n$.

3.4.7 Demethylation Assays for Peptides and Proteins

Demethylation of peptides and full length proteins was performed under aerobic conditions and followed using a NanoDrop 2000c spectrophotometer. The 75 μL reactions were initiated by the addition of buffered substrate solution, varying in concentration form 2 μM to 25 μM, to reaction mixture consisting of 50 mM HEPES buffer (pH 7.5), 0.1 mM Amplex Red, 0.76 μM horseradish peroxidase and 500 nM full-length His_6-LSD1. Absorbance changes were monitored at 571 nm, and an extinction coefficient of 52,000 $M^{-1}cm^{-1}$ was used to calculate product formation. Double

reciprocal plot of the initial rate of product formation versus substrate concentration afforded the k_{cat} and K_m values for each substrate.

3.4.8 Demethylation Assays for NCPs and Nucleosomal Arrays

Demethylation of NCPs and nucleosomal arrays was performed by incubation of native H3K4me2 containing nucleosomes (30 nM) with varying concentrations of LSD1 (2 to 25 μM) in 50 mM HEPES, pH 8.0, 5% glycerol, 1 mM DTT, 50 mM KCl buffer at room temperature. Time points taken at t = 0, 30, 60 and 120 min were quenched using 6x Laemmli dye and boiled for 2 min to stop all enzymatic activity. Samples were run on 15% SDS PAGE gels at 200 V for 40 min and transferred onto Immunoblot PVDF membranes (Bio-Rad) for 1 hr at 100 V. Membranes were blocked at room temperature using 4% Fat free milk in PBS buffer for 60 min and then incubated overnight at 4 ˚C with H3K4me2 antibody (Up-state 07-442, Lot: 2017310) and H4 antibody (Active Motif 39269, Lot: 11908001) at 1:10,000 and 1:1,000 dilutions, respectively, in a PBST buffer containing 4% fat free milk. Membranes were subsequently washed and incubated with goat anti-rabbit secondary antibody (LI-COR 926-68021, Lot: C10628-01) at 1:20,000 dilutions in a PBST buffer containing 4% fat free milk for 30 min at room temperature. Following this, membranes were washed and visualized using Odyssey Application Software (LI-COR Biosciences).

3.4.9 Generation of human H4 Δ14 A15C

Site-directed mutagenesis was employed to generate the mutants H4 Δ14 A15C from human H4 cloned in the pET15b vector. *E. coli* BL21(DE3) cells were transformed with the mutant plasmids and grown at 37 ˚C in 6 L YT medium until OD_{600} ~0.7. Protein expression was induced by the addition of 0.3 mM IPTG for 2 h at 37 ˚C. Cells were harvested by centrifugation

at 7000xg, resuspended in 150 mM NaCl, 50 mM Tris, pH 7.5 and lysed by sonication. The lysate was centrifuged at 20,000xg for 20 min and insoluble histones were recovered from inclusion bodies with an extraction buffer consisting of 6 M Gn-HCl, 100 mM NaCl, 50 mM Tris, pH 7.5. The re-solubilized histones were purified by Ni2+-affinity chromatography. Briefly, proteins were bound to the Ni^{2+}-column overnight at 4 ˚C with gentle shaking and the column subsequently washed with 3 volumes of extraction buffer containing 25 mM Imidazole. Immobilized proteins were eluted using 500 mM imidazole in the extraction buffer. Histone-containing fractions were identified by 15% SDS-PAGE, combined and dialyzed into 1 mM DTT overnight at 4 ˚C. Purified His6-TEV protease was then added in a 1:10 molar ratio (enzyme:substrate) and His_6-tag cleavage was performed overnight in a buffer containing 1 mM EDTA, 10 mM DTT, 10 mM Cys, 50 mM Tris, pH 6.9. The cleavage products were dialyzed into extraction buffer and applied to a Ni^{2+}-column. This led to retention of the cleaved His_6-tag, His_6-TEV and any uncleaved histones on the column. The eluted tag-less H4 Δ14 A15C was further purified to homogeneity by RP-HPLC and characterized by ESI-MS (Calc 10,072.71 Da, Obs 10,072.3 ± 1.2 Da) (Figure 3.4.1b).

3.4.10 Generation of suH4

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

The peptide BocHN-SGRGKGGKGLG<u>K</u>GG-C(O)$NHNH_2$ corresponding to the first 14 N-terminal residues of the human histone H4 protein was synthesized by SPPS on a 0.25 mmol scale employing standard 9-fluorenylmethoxycarbonyl (Fmoc)-based N^{α}-deprotection chemistry. Briefly, 2-chlorotrity hydrazine resin was prepared by reacting 2-chlorotrityl chloride resin (1.33 mmol/g) in a 10% solution of hydrazine in DMF at 30 ˚C for 30 min. The reaction was repeated once with fresh hydrazine solution. The resin was then treated with a 10% methanol in DMF

solution for 10 min to cap any unreacted sites on the resin. The first amino acid, Gly, was coupled in 4-fold molar excess. The coupling reaction containing Fmoc-Glycine (1.0 mmol), *O*-(6-Chlorobenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium hexafluorophosphate (HCTU, 0.95 mmol) , and DIEA (2.0 mmol) proceeded for 60 min at 30 ˚C. From glycinyl 2-chlorotrityl hydrazine resin each remaining amino acid was coupled in 5 molar excess based on resin loading. Deprotection of the Fmoc group was achieved by treating resin with 20% piperidine in DMF for 3 min at 75 ˚C. Coupling reactions were undertaken for 5 min at 75 °C with a mixture of Fmoc-amino acid (1.31 mmol), HBTU (1.28 mmol) and DIEA (2.75 mmol) in DMF. For Arg, an additional coupling reaction was performed for 25 min at 75 °C. The Lys at position 12 was orthogonally protected with the ivDde protecting group. The peptide was protected at the α-NH_2 position with Boc group by reaction with di-*tert*-butyl dicarbonate (2.0 mmol) and DIEA (4.0 mmol) in DMF for 2 hours.

Attachment of the ligation auxiliary

Deprotection of the ivDde group was achieved by reacting resin bound peptide with a solution of 5% hydrazine in DMF for 5 min. This deprotection was repeated three times. The peptidyl resin was then coupled to bromoacetic acid (8-fold molar excess) with *N,N′*-Diisopropylcarbodiimide (DIC, 8-fold molar excess) in DMF for 45 min at room temperature. The coupling was repeated once. Subsequently, dry peptidyl resin was placed in a solution containing 9 equivalents of auxiliary (0.5 M in DMSO) and shaken for 24 hours at room temperature. Completion of the displacement was judged by test cleavage and subsequent ESI-MS analysis. Peptide was cleaved and deprotected by reaction of resin at 20 µL/mg with Reagent K (TFA: thioanisole: H_2O: phenol: 1,2-ethanedithiol 82.5:5:5:5:2.5 v/v) for 1.5 hours at room temperature, then precipitated and washed twice with cold diethyl ether. Dry peptide was dissolved in RP-HPLC

buffer A and purified by C18 preparative and semi-preparative RP-HPLC with a gradient of 0-25% B for KAK^{aux}I and 0-50% B for H4(1-14) ^{aux}-C(O)NHNH$_2$. This yielded 38% of the peptide-auxiliary conjugate KAK^{aux}I and 9% of the peptide-auxiliary conjugate H4(1-14)^{aux}-C(O)NHNH$_2$ - based on initial resin loading. ESI-MS of KAK^{aux}I. Calculated *m/z* $[M+H]^+$ 591.4 Da, observed 591.8 Da. ESI-MS of H4(1-14)^{aux}-CONHNH$_2$. Calculated *m/z* $[M+H]^+$ 1,363.6 Da, observed 1,363.8 Da.

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

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

Primer	DNA Sequence (5'- to -3')
hSUMO-3(1-91)C47S-FP	ATCGACGTGTTCCAGCAGCAGACGGGATGCATCACGGGAGATGCACTAGTTGCC
hSUMO-3(1-91)C47S-RP	GGCAACTAGTGCATCTCCCGTGATGCATCCCGTCTGCTGCTGGAACACGTCGAT

hUb(1-75)-AvaDnaE-AAFN-H6-FP	CTGCACCTGGTACTCCGTCTCAGAGGTTGCCTGAGCTATGATACCGAAGTGCTG
hUb(1-75)-AvaDnaE-AAFN-H6-RP	CAGCACTTCGGTATCATAGCTCAGGCAACCTCTGAGACGGAGTACCAGGTGCAG
hH4(15-102)A15C-FP	GGGAATTC*CATATG*GAAAACCTGTACTTCCAGTGCAAACGTCACCGTAAAGTTCTG (*NdeI*)
hH4-RP	GCCCGC*GGATCC*TCAACCACCGAAACCGTACAGGGTACGACCC (*BamHI*)

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

E. coli BL21(DE3) cells were transformed with the plasmid pTXB1-SUMO-3(2-91)C47S. Cells were grown in 6 L Luria-Bertani medium supplemented with 100 µg/mL of Ampicillin at 37 ºC with shaking at 250 rpm until OD_{600} ~0.6-0.8. Overexpression of the desired fusion proteins was induced by the addition of 0.3 mM IPTG and cells were grown for an additional 4 h at 25 ºC. The cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in lysis buffer: PBS, pH 7.2 containing 1 mM 2-mercaptoethanesulfonic acid sodium salt (MESNa). Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The lysate supernatant was passed through a 0.45 µm filter then applied to a 30 mL chitin column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 12 h at 4 ºC. The column was then washed with 20 column volumes (CV) of lysis buffer followed by 2 CV of PBS, pH 7.75. SUMO-3(2-91)C47S-MES was cleaved from its intein-CBD fusion by incubation with 1.5 CV of PBS, pH 7.75 containing 100 mM MESNa for 72 h at 4 ºC. The eluted α-thioester was purified by C18 preparative RP-HPLC employing a gradient of 30-60% B over 60 min. Fractions containing the desired thioester were identified by ESI-MS. We observed that the N-terminal Met of SUMO-3 is consistently processed *in vivo*, leading to the SUMO-3(2-91)-α-thioester product. Typical yield is 4-5 mg/L of cell culture.

ESI-MS for SUMO-3(2-91)C47S-MES. Calculated *m/z* $[M+H]^+$ 10,444.7 Da, observed 10,445.8 ± 3.6 Da.

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

Purified 13 (15.6 mg, 11.6 µmol) and SUMO-3(2-91)C47S-MES (20.1 mg, 1.9 µmol) were dissolved in 7.2 mL of a buffer consisting of 6 M Gn-HCl, 100 mM Na_2HPO_4, and 10 mM TCEP, pH 7.3. Ligation proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C18 preparative RP-HPLC employing a gradient of 25-50% B over 60 min to give 12.6 mg (56%). ESI-MS of H4(1-14)$^{Su(C47S)(aux)}$-CONHNH$_2$. Calculated *m/z* $[M+H]^+$ 11,666.1 Da, observed 11,666.9 ± 2.8 Da.

Overexpression and purification of TEV protease

E. coli BL21(DE3) cells containing the plasmid pRK793-His$_6$-TEV(1) were grown in 1 L LB supplemented with 100 µg/mL of Ampicillin at 37 ºC with shaking at 250 rpm until OD_{600} ~0.6. Overexpression was induced by the addition of 0.3 mM IPTG and cells were grown for an additional 6 h at 25 ºC. The cells were harvested by centrifugation at 7,000x*g* for 15 min. The cell pellet was resuspended in 15 mL lysis buffer: 20 mM tris, 50 mM NaCl, pH 7.2. Cells were lysed by sonication then centrifuged at 20,000x*g* for 15 min. The lysate supernatant was passed through a 0.45 µm filter then applied to ~5 mL Ni-NTA column pre-equilibrated with lysis buffer. Proteins were bound to the column over a period of 3 h at 4 ºC. The column was then washed thoroughly with lysis buffer containing 200 mM imidazole. His$_6$-TEV was eluted with lysis buffer containing 500 mM imidazole, and dialyzed into 4 L of 20 mM tris, 50 mM NaCl, 1 mM DTT, pH 7.5 for 18 h at 4 ºC.

Overexpression and purification of H4(15-102)A15C

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

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

Ligation was accomplished by first converting the C-terminal hydrazide of H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$ to an acyl azide with $NaNO_2$ via the diazotization reaction. Subsequent addition of MPAA served to both quench the remaining $NaNO_2$ and generate a highly reactive C-terminal thioester for the ligation reaction. Purified H4(1-14)$^{Su(C47S)(aux)}$-C(O)NHNH$_2$ (1.7 mg, 0.15 μmol) was dissolved at 1 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl, pH 3, and kept at -20 °C for a minimum of 20 min. To this solution was added 4.5 μL of a 500 mM solution of $NaNO_2$ in water. The reaction was briefly mixed, then kept at -20 °C for 15 min. Then, a solution of H4(15-102)A15C (3 mg, 0.3 μmol) dissolved at 1.25 mM in 200 mM Na_2HPO_4, 6 M Gn-HCl, 200 mM MPAA, pH 6.5, was added to the reaction. The mixture was allowed to come to room temperature, and the pH adjusted with 3 M NaOH to 6.8-7.0 in order to form the MPAA α-thioester. The ligation reaction proceeded with gentle shaking at 25 °C for 24 h. Ligation product was purified by C4 semi-preparative RP-HPLC employing a gradient of 30-70% B over 45 min to give 2.1 mg (66%). ESI-MS of H4(A15C)$^{Su(C47S)(aux)}$. Calculated *m/z* $[M+H]^+$ 21,704.9 Da, observed 21,711.0 ± 5.9 Da.

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

Purified H4(A15C)$^{Su(C47S)(aux)}$ (1.4 mg, 64 nmol) was dissolved in 7 mL of 200 mM MPAA, 100 mM Na_2HPO_4, pH 7.3, and incubated in a 15 mL conical tube at 25 °C for 24 h. After incubation, the reaction volume was reduced with a 10,000 MWCO spin concentrator (GE Healthcare Life Sciences, Pittsburgh, PA) and buffered TCEP, pH 7.3 was added to a concentration of 50 mM and the reaction incubated at 4 °C for 30 min. Ligation product was purified by C4 analytical RP-HPLC employing a gradient of 30-70% B over 30 min. ESI-MS of H4(A15C)$^{Su(C47S)}$. Calculated *m/z* $[M+H]^+$ 21,628.7 Da, observed 21,634.0 ± 9.1 Da.

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

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

3.5 REFERENCES

(1) Koutelou, E., Hirsch, C. L., and Dent, S. Y. R. (2010) Multiple faces of the SAGA complex. *Curr. Opin. Cell Biol. 22*, 374–382.

(2) Hathaway, N. a, Bell, O., Hodges, C., Miller, E. L., Neel, D. S., and Crabtree, G. R. (2012) Dynamics and memory of heterochromatin in living cells. *Cell 149*, 1447–60.

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

(4) McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G., and Muir, T. W. (2008) Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature 453*, 812–6.

(5) Shilatifard, A. (2006) Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu. Rev. Biochem. 75*, 243–69.

(6) Nguyen, A. T., and Zhang, Y. (2011) The diverse functions of Dot1 and H3K79 methylation. *Genes & Dev. 25*, 1345–1358.

(7) Bossis, G., Malnou, E., Farras, R., Andermarcher, E., Hipskind, R., Rodriguez, M., Schmidt, D., Muller, S., Jariel-encontre, I., and Piechaczyk, M. (2005) Down-Regulation of c-Fos / c-Jun AP-1 Dimer Activity by Sumoylation. *Mol. Cell. Biol. 25*, 6964–6979.

(8) Ross, S., Best, J. L., Zon, L. I., and Gill, G. (2002) SUMO-1 Modification Represses Sp3 Transcriptional Activation and Modulates Its Subnuclear Localization. *Mol. Cell 10*, 831–842.

(9) Yang, S., Jaffray, E., Hay, R. T., and Sharrocks, A. D. (2003) Dynamic Interplay of the SUMO and ERK Pathways in Regulating Elk-1 Transcriptional Activity *Mol. Cell 12*, 63–74.

(10) Yang, S.-H., and Sharrocks, A. D. (2004) SUMO promotes HDAC-mediated transcriptional

repression. *Mol. Cell 13*, 611–7.

(11) Ouyang, J., Shi, Y., Valin, A., Xuan, Y., and Gill, G. (2009) Direct binding of CoREST1 to SUMO-2/3 contributes to gene-specific repression by the LSD1/CoREST1/HDAC complex. *Mol. Cell 34*, 145–54.

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

(13) Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004) Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell 119*, 941–953.

(14) Mosammaparast, N., Kim, H., Laurent, B., Zhao, Y., Lim, H. J., Majid, M. C., Dango, S., Luo, Y., Hempel, K., Sowa, M. E., Gygi, S. P., Steen, H., Harper, J. W., Yankner, B., and Shi, Y. (2013) The histone demethylase LSD1/KDM1A promotes the DNA damage response *J. Cell Biol. 203*, 457–470.

(15) Yang, M., Gocke, C. B., Luo, X., Borek, D., Tomchick, D. R., Machius, M., Otwinowski, Z., and Yu, H. (2006) Structural Basis for CoREST-Dependent Demethylation of Nucleosomes by the Human LSD1 Histone Demethylase. *Mol. Cell 23*, 377–387.

(16) Forneris, F., Binda, C., Dall'Aglio, A., Fraaije, M. W., Battaglioli, E., and Mattevi, A. (2006) A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. *J. Biol. Chem. 281*, 35289–35295.

(17) Forneris, F., Battaglioli, E., Mattevi, A., and Binda, C. (2009) New roles of flavoproteins in molecular cell biology: Histone demethylase LSD1 and chromatin. *FEBS J. 276*, 4304–4312.

(18) Marmorstein, R., and Trievel, R. C. (2009) Histone modifying enzymes: Structures,

mechanisms, and specificities. *Biochim. Biophys. Acta - Gene Regul. Mech. 1789*, 58–68.

(19) Lee, M. G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005) An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature 437*, 432–5.

(20) Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L., Li, R., Li, Y., Zhang, Y., Li, Q., Yi, X., and Shang, Y. (2009) LSD1 Is a Subunit of the NuRD Complex and Targets the Metastasis Programs in Breast Cancer. *Cell 138*, 660–672.

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

(22) You, a, Tong, J. K., Grozinger, C. M., and Schreiber, S. L. (2001) CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proc. Natl. Acad. Sci. U. S. A. 98*, 1454–8.

(23) Da, G., Lenkart, J., Zhao, K., Shiekhattar, R., Cairns, B. R., and Marmorstein, R. (2006) Structure and function of the SWIRM domain, a conserved protein module found in chromatin regulatory complexes. *Proc. Natl. Acad. Sci. U. S. A. 103*, 2057–62.

(24) Stavropoulos, P., Blobel, G., and Hoelz, A. (2006) Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat. Struct. Mol. Biol. 13*, 626–32.

(25) Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E., and Mattevi, A. (2005) Human histone demethylase LSD1 reads the histone code. *J. Biol. Chem. 280*, 41360–5.

(26) Culhane, J. C., and Cole, P. a. (2007) LSD1 and the chemistry of histone demethylation. *Curr. Opin. Chem. Biol. 11*, 561–8.

(27) Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A., and Battaglioli, E. (2005) Histone

demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett. 579*, 2203–7.

(28) Culhane, J. C., Szewczuk, L. M., Liu, X., Da, G., Marmorstein, R., and Cole, P. A. (2006) A mechanism-based inactivator for histone demethylase LSD1. *J. Am. Chem. Soc. 128*, 4536–7.

(29) Yang, M., Culhane, J. C., Szewczuk, L. M., Gocke, C. B., Brautigam, C. A, Tomchick, D. R., Machius, M., Cole, P. A, and Yu, H. (2007) Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. *Nat. Struct. Mol. Biol. 14*, 535–539.

(30) Barrios, A. P., Gómez, A. V, Sáez, J. E., Ciossani, G., Toffolo, E., Battaglioli, E., Mattevi, A., and Andrés, M. E. (2014) Differential properties of transcriptional complexes formed by the CoREST family. *Mol. Cell. Biol. 34*, 2760–2770.

(31) Simon, M. D., Chu, F., Racki, L. R., de la Cruz, C. C., Burlingame, A. L., Panning, B., Narlikar, G. J., and Shokat, K. M. (2007) The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell 128*, 1003–12.

(32) Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schafer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T., and Reuter, G. (2007) Heterochromatin Formation in Drosophila Is Initiated through Active Removal of H3K4 Methylation by the LSD1 Homolog SU(VAR)3-3. *Mol. Cell 26*, 103–115.

(33) Choi, H. K., Kim, B. J., Seo, J. H., Kang, J. S., Cho, H., and Kim, S. T. (2010) Cohesion establishment factor, Eco1 represses transcription via association with histone demethylase, LSD1. *Biochem. Biophys. Res. Commun. 394*, 1063–1068.

(34) Tortorici, M., Borrello, M. T., Tardugno, M., Chiarelli, L. R., Pilotto, S., Ciossani, G., Vellore, N. a, Bailey, S. G., Cowan, J., O'Connell, M., Crabb, S. J., Packham, G., Mai, A., Baron, R., Ganesan, A, and Mattevi, A. (2013) Protein recognition by short peptide reversible inhibitors of

the chromatin-modifying LSD1/CoREST lysine demethylase. *ACS Chem. Biol. 8*, 1677–82.

(35) Kugler, J. E., Deng, T., and Bustin, M. (2012) The HMGN family of chromatin-binding proteins: Dynamic modulators of epigenetic processes. *Biochim. Biophys. Acta - Gene Regul. Mech. 1819*, 652–656.

(36) Kato, H., van Ingen, H., Zhou, B.-R., Feng, H., Bustin, M., Kay, L. E., and Bai, Y. (2011) Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc. Natl. Acad. Sci. U. S. A. 108*, 12283–8.

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

(38) Cubeñas-Potts, C., and Matunis, M. J. (2013) SUMO: A Multifaceted Modifier of Chromatin Structure and Function. *Dev. Cell 24*, 1–12.

(39) Li, Y.-J., Stark, J. M., Chen, D. J., Ann, D. K., and Chen, Y. (2010) Role of SUMO:SIM-mediated protein-protein interaction in non-homologous end joining. *Oncogene 29*, 3509–18.

(40) Bettermann, K., Benesch, M., Weis, S., and Haybaeck, J. (2012) SUMOylation in carcinogenesis. *Cancer Lett. 316*, 113-125.