THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

THE EFFECTS OF INTERNAL H3 TRUNCATIONS ON LSD1/COREST DEMETHYLASE ACTIVITY

JAMES LAKSHMAN JOHNSTON
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Biochemistry and Molecular Biology and Statistics
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Song Tan
Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Teh-hui Kao
Distinguished Professor of Biochemistry and Molecular Biology
Honors Adviser

Scott Selleck
Department Head for Biochemistry and Molecular Biology

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The LSD1 and CoREST proteins form a complex that specifically demethylates histone H3 at lysines 4 and 9 (D'Oto et al, 2017). Without CoREST, LSD1 cannot act on the nucleosome, though it can act on individual H3 proteins, and several studies have shown that CoREST makes crucial contact with the DNA surrounding the histone octamer (Kim et al, 2015). Our lab is attempting solve the structure of the LSD1/CoREST complex on the nucleosome, both through X-ray crystallography and various enzymatic assays. Previous experiments conducted in our lab demonstrated that 5 and 10 amino acid internal truncations were insufficient to reduce LSD1 activity (Kim et al, 2015). A recent, low-resolution structure of LSD1 on the nucleosome led to the suggestion that our previous structure was inaccurate. Our revised structure suggested an 18 amino acid internal truncation to histone H3 would be sufficient to remove the H3 tail from the LSD1/CoREST active site and eliminate any demethylase activity. To test this hypothesis, I designed histone H3 mutants with 12, 15, 18, and 20 amino acid internal truncations. I incorporated these truncated histone mutants into nucleosomes and was able to demonstrate through a demethylase assay that both 18 and 20 amino acid internal truncations to histone H3 were sufficient to eliminate the activity of LSD1. This finding lends credence to the structure proposed by our laboratory and will help the laboratory continue to fine tune the structure of LSD1 and CoREST on the nucleosome.

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

ACKNOWLEDGEMENTS

There are many people who I would like to thank for their help with this thesis. Firstly, I would like to thank Dr. Tan for everything he has done for me over my four years at Penn State. You taught me not only how to be a scientist but also made me a markedly better and more responsible person. Entering your lab during my freshman year is amongst the best decisions that I have made and I cannot properly thank you for all the help you have given me.

I would also like to thank Dr. Sang-Ah Kim for her help over the years. I have no idea how many questions I have asked you over the past four years but you answered every single one. You were beyond generous with your time and observing how you went about your work helped shape how I approach projects. I hope you one day get to open your café.

Additionally, I would like to express my sincere gratitude to all of the other current and former members of the Tan laboratory. Dr. Robert McGinty played a vital role in the completion of this thesis and I cannot thank him enough for his guidance. Without the aid of Kevin Thyne, Max Kruse, Mike Doyle, Bryan Tornabene, and Lauren McCall, I would never have been able to complete all of the expression and purification required for this thesis. I cannot thank them all enough for their help. Finally, I would like to thank my fellow undergraduates, specifically Turner Pecen and Victoria Spadafora. There were certainly some stressful times over the past four years and you both helped keep me grounded.

Special thanks go to my friends and family. Mom and Dad, I would not have gotten to this point without you. The values that you imparted on me throughout my childhood guide every decision I make in my life and I hope I can continue to make you proud. Mitchell and Olivia, you are the best siblings anyone could possibly ask for and I cannot imagine a better

support system. To my friends new and old, thank you for understanding when I was too busy to be around as much as I would like and for making me remember to relax and have fun every once in a while. I would not have completed this thesis without all of the amazing people in my life. I could not be more grateful to everyone who contributed to this thesis and to those who have shaped these wonderful past four years. Thank you all.

Chapter 1 : Introduction

1.1 DNA and the Genetic Code

Deoxyribonucleic acid (DNA) is the basis of all complex life on earth. From the simplest organisms to the most complex, the genome encoded by the seemingly simple combinations of the four base pairs of DNA spawns the incredibly diverse range of living organisms that we see today. DNA serves a number of functions in the cell, primarily as a template for the various types of RNAs and as a regulatory feature determining protein expression. However, perhaps most importantly, our chromosomal genome carries the DNA that encodes the genes that are translated to protein. The regulation of gene expression is the core mystery our laboratory strives to solve.

Gene expression determines the fates of cells. While a neuron and a liver cell contain the same DNA, they present with vastly different morphology and functions. Answering why and how specific genes are expressed in different cell types and different organisms is of the utmost importance, not only for the sake of understanding human development but also for understanding the roots of disease. For instance, cancers are caused by a dysregulation of the systems that regulate the cell cycle, systems defined by proteins encoded in the genome (Audia and Campbell, 2016). Take for example p53, a tumor suppressor. p53 acts as transcription factor for a variety of proteins that suppress constant growth and replication of cells. A mutation to p53 causes the cell to lose its ability to inhibit cell division as p53 target genes are no longer expressed, resulting in tumorigenesis (Fischer, 2017). Additionally, selective gene expression is

a large part of what gives humans such a wide range of inter-species diversity: humans share 99.5% of their DNA with unrelated humans.

To understand the complexity of gene expression, it is first important to understand the sheer size of the genome and how that size significantly impacts the expression of genes. Human beings have 6 billion base pairs of DNA within each cell. Stretched out in a thin strand, the DNA in a single cell would be about 2 meters in length. This would seem problematic: the average human cell is only about 25 micrometers (10^{-6}) in diameter and that is not even considering the fact that DNA is limited to the nucleus. How can DNA of such great length fit within the cell?

1.2 Chromatin

Chromatin is the packaging that allows DNA to be stored in the small enclosure that is the nucleus. There are three core purposes of chromatin. The first is the packaging of DNA so that it fits within the nucleus (McGinty and Tan, 2014). The second purpose is illustrated during mitosis: a change in chromatin packing allows for the condensation of chromosomes (Müller and Almouzni, 2017). Finally, and of most interest to our lab, chromatin modifications play a large role in gene expression (Pirrotta, 2016).

The basic unit of chromatin is called a nucleosome. Each nucleosome core particle consists of between 145 and 147 base pairs of DNA wrapped around an octamer of proteins (McGinty and Tan, 2014). The octamer consists of two copies of each of the following proteins: H2A, H2B, H3, and H4. H2A and H2B form a pair of dimers within the octamer while H3 and H4 form a single tetramer. The histone fold found in each of the four core histone proteins has

three alpha helices joined by two loops. As is the case with the vast majority of protein complexes, hydrophobic residues and domains occupy the core of the complex while the charged, in this case basic, and polar residues occupy the domains facing the environment. Dimerization of H2A and H2B occurs along the alpha-2 helix with antiparallel orientation. The dimer joins the H3/H4 tetramer through an interaction between H4 and H2B hydrophobic domains, which forms an internal hydrophobic cluster. Basic residues line the outside of the octamer, enabling the interaction with negatively charged DNA (McGinty and Tan, 2014). The basic amino acids populating the outside of the complex primarily consist of arginine and lysine (Audia and Campbell, 2016).

Each nucleosome is connected by a linker DNA. The length of this linker is variable, ranging from 160 to 240 base pairs, and the linking DNA is occasionally accompanied by an additional histone, H1 or H5 (McGhee and Felsenfeld, 1980). Chromatin is frequently compared to beads on a string, with each nucleosome core particle representing a bead and DNA representing the string (Figure 1-1).

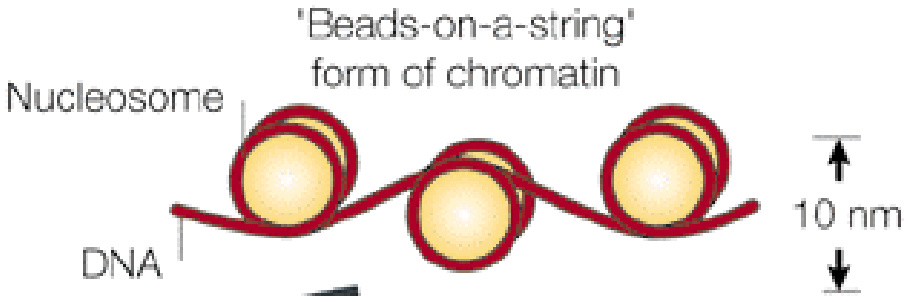

Figure 1-1: The wrapping of DNA around histones forms chromatin
The packaging is analogous to beads with string wrapped around them. Source: Nuclear compartmentalization and gene activity. Image Source: Francastel et al, 2000 (Figure 3)

The packing of these beads is at the core of gene expression. When a region of chromatin is in its highly condensed form, the genes in that region cannot be expressed as RNA polymerase cannot access the DNA.

1.3 Chromatin Modification

Chromatin fundamentally exists in two major states. Fluctuation between these two states plays a large role in regulating gene expression. One state is heterochromatin. In heterochromatin, the chromatin is extremely tightly packed. Because of the highly condensed nature, genes cannot be transcribed as the DNA is inaccessible. During mitosis, the condensed chromosomes are almost entirely made up of heterochromatin (Müller and Almouzni, 2017). Euchromatin is the active state of chromatin packing. The nucleosomes are more spread apart, allowing them to be accessed by RNA polymerase, transcription factors, and other proteins important for the expression of proteins. Figure 1-2 illustrates the difference between euchromatin and heterochromatin, specifically highlighting how proteins are able to access euchromatin.

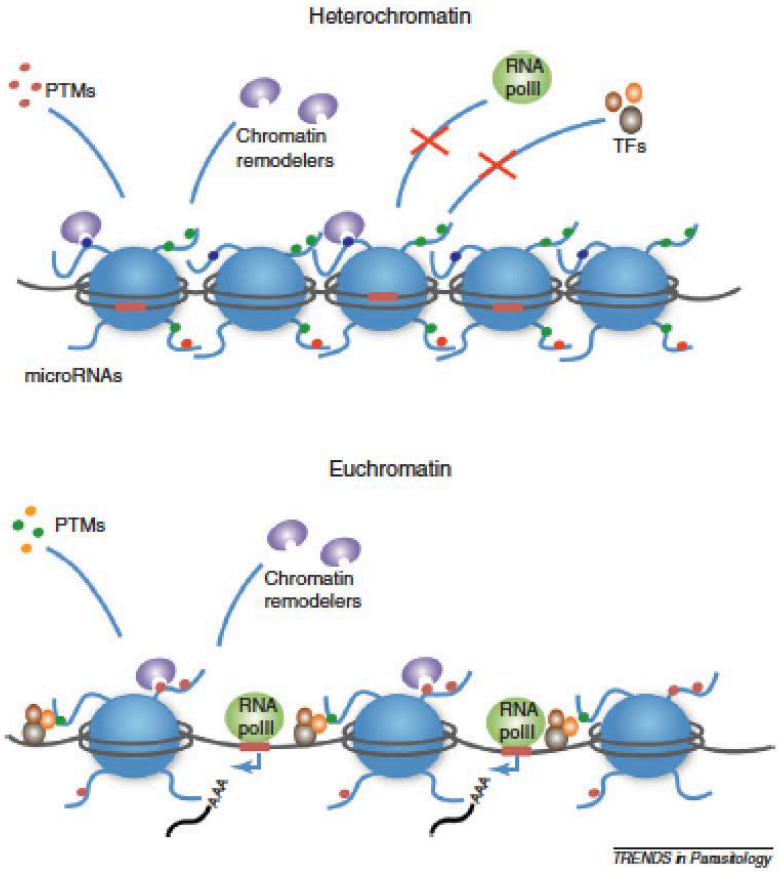

Figure 1-2: Euchromatin and Heterochromatin
Euchromatin is less densely packed. This allows for transcription factors and RNA polymerase to bind the DNA. Heterochromatin, on the other hand, is too densely packed for any transcription to occur. Image Source: Croken et al, 2012 (Figure 1)

A good example illustrating the importance of chromatin packing is cell differentiation in the human body. Each cell type carries the same DNA but has a unique epigenetic signature that regulates which genes are expressed. In the case of smooth vascular muscle cells, for instances, researchers were able to show that vascular smooth muscle cell marker genes were expressed due to a shift, regulated by epigenetic modifications to both DNA and histone tails, from heterochromatin to euchromatin in specific gene containing regions. Conversely, regions containing genes specific to other cell types were silenced via a shift from euchromatin to heterochromatin controlled by the same regulatory mechanisms (Gomez et al, 2015).

A variety of factors regulate the packing of chromatin. Perhaps chief among these factors is the “Histone Code” (Sun et al, 2001). The histone code refers to covalent modifications to the amino acid residues on the tails of the histones featured in the nucleosome core particles. These modifications change the conformation of the chromatin surround that particular histone. For instance, the acetylation of specific histone tail residues has been shown to correspond to the activation of genes located in that chromatin region (Audia and Campbell, 2016). The negatively charged acetyl group weakens the electrostatic interactions between the negatively charged DNA and the positively charged histones. Generally, lysine residues are most commonly targeted for acetylation. The lessened electrostatic attraction leads to more loosely packed chromatin and greater access to the DNA for both polymerases and transcription factors.

Histone methylation and demethylation is the focus of this thesis. Histones can be methylated on both arginine and lysine residues. Unlike acetylation, methylation of histones is tied to both the repression and activation of genes. The specific residue methylated as well as the number of methyl groups added (up to three methyl groups can be added to lysine) both dictate the effect of methylation of the genes expression. For this reason, methylation and demethylation activity is carried out by a wide range of protein with very specific activity. Of most interest to our laboratory currently is a demethylase known as LSD1.

1.4 LSD1/CoREST

Lysine Specific Demethylase 1, known for the duration of this paper as LSD1, is a nuclear protein that demethylates lysine residues on H3. My thesis work is part of a larger effort within our laboratory to determine the structure of LSD1 on the nucleosome. LSD1 is part of the

KDM1 family and can remove either methyl or dimethyl groups from either lysine 4 or lysine 9 on H3 (Marabelli et al, 2016). These residues are of specific interest to those who study epigenetics. Methylation of H3K4 has been closely associated with either active genes or promoter regions (Figure 1-3). In contrast, H3K9 methylation is closely associated to the repression of genes. Thus, LSD1 plays a major role in both activation and repression of genes.

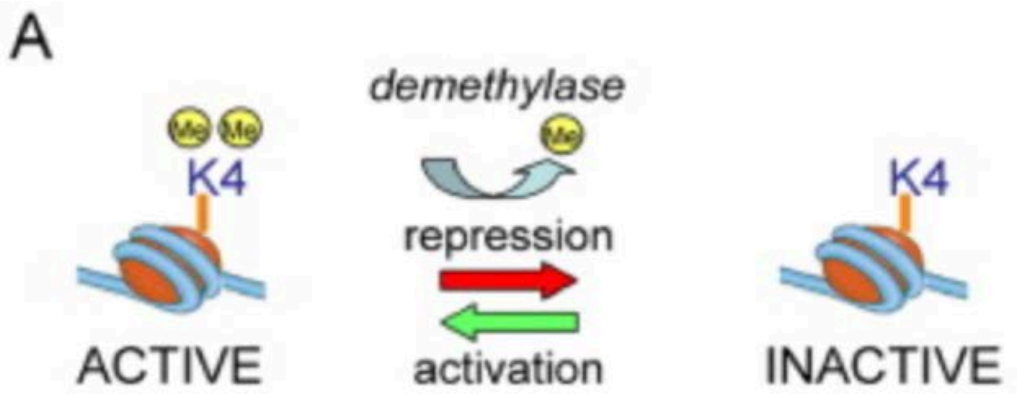

Figure 1-3: Effect of H3K4 Methylation State on Gene Expression
Removal of methyl groups from H3K4 is associated with the inactivation of genes. Image Source: Wysocka et al, 2005

LSD1 performs its demethylase activity in conjunction with another nuclear protein, REST co-repressor protein (CoREST). The catalytic domain of LSD1 removes methyl groups via an oxidation reaction, resulting in the formation of a free formaldehyde group. Specifically, LSD1 and CoREST interact to demethylate both dimethylated and monomethylated lysine 4. The demethylase activity is dependent on flavin adenine dinucleotide (FAD) and a pair of electrons donated by the ε nitrogen on lysine 4 (Marabelli et al, 2016). For this reason, LSD1 cannot act on trimethylated lysine. Additionally, it is worth noting that while LSD1 can demethylate free H3, it cannot act on the H3 tail when H3 is incorporated in the nucleosome without CoREST.

Despite significant similarity to traditional amine oxidases, both in terms of certain structural homology and chemical mechanism, LSD1/CoREST has several interesting features relating to its mechanism. LSD1 consists of an N terminal SWIRM domain as well as a C terminal amine oxidase domain (referred to as AOD for the remainder of this paper) (Yang et al, 2006). CoREST contains an ELM2 domain and two SANT domains. The second SANT domain, SANT2, binds DNA, and it has been shown that truncation mutants of CoREST containing only SANT2 and the linker region between the SANT domains is sufficient to lead to demethylation of H3K4 (Yang et al, 2006). The LSD1 AOD domain has 2 globular lobes: one of these lobes binds FAD while the other interacts with the substrate, H3K4. The AOD consists of 6 beta sheets and 5 alpha helices. The LSD1 active site is located where these two lobes interact (Yang et al, 2006). The "insert" of LSD1, two alpha helices that extend away from the AOD lobes, interacts with the SANT domain linker of CoREST (Figure 1-4). The outer rim of the LSD1 active site is lined with negative amino acids, similar to enzymes in the amine oxidase family (Yang et al, 2006).

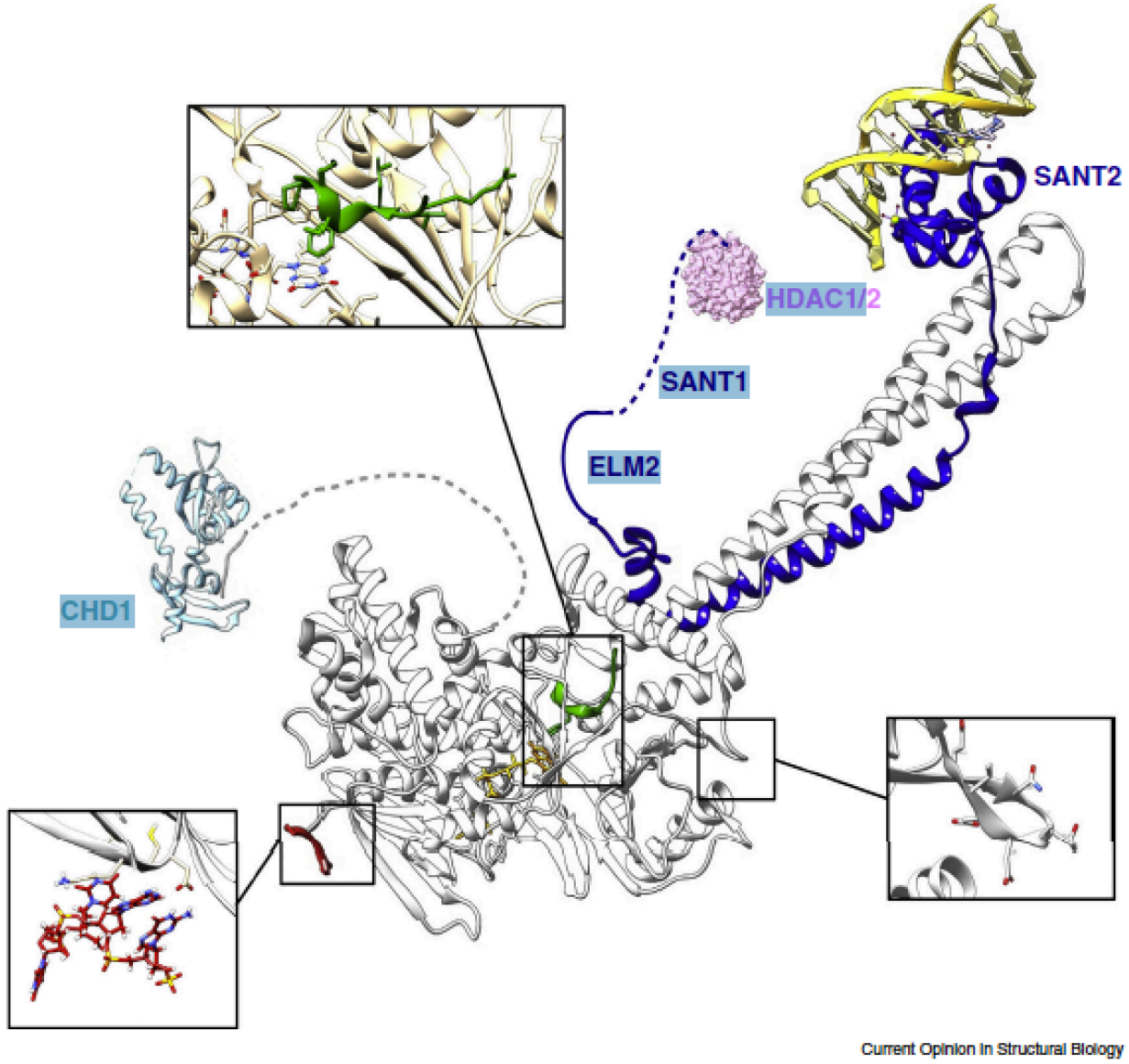

Figure 1-4: Crystal Structure of LSD1 and CoREST
Image Source: Marabelli et al, 2016 (Figure 1)

While the structure and mechanism of LSD1 and CoREST are well characterized, significantly less is known regarding how LSD1 acts on the nucleosome. Several groups have proposed structures of LSD1 on the nucleosome but these structures are not based on crystallographic evidence. One such structure is displayed in Figure 1-5. Our laboratory is

currently working to solve the structure of LSD1/CoREST on the nucleosome. The above figure shows where the H3 tail lies in the LSD1 active site. The location of the H3 tail within the LSD1 active constrains how LSD1 can bind to the nucleosome in an enzymatically active manner. Our current structure of LSD1/CoREST on the nucleosome shows that the LSD1 active site is 49 angstroms from the point where the H3 tail leaves the core of the nucleosome. Based on this structure, we hypothesized that 18 and 20 amino acid internal truncations to histone H3 will be sufficient to prevent the H3 tail from entering the LSD1 active site. To test this hypothesis, I created histone H3 mutants with 12, 15, 18, and 20 amino acid internal truncations. Previously, our laboratory has shown that 5 to 10 amino acid internal truncations to the H3 tail did not prevent or inhibit the activity of LSD1 on the nucleosome (Kim et al, 2015). The truncations start at amino acid 20 because we do not want to alter tail of H3: instead, these experiments seek to determine what level truncation is sufficient to remove the H3K4 from the LSD1 active site. The peptide chain length of an extended peptide is approximately 3.6 angstroms. Thus, the 12, 15, 18, and 20 amino acid truncations would shorten H3 by approximately 43, 54, 65, and 72 angstroms, respectively. The failure of LSD1/CoREST to demethylate nucleosomes contain the 18 or 20 amino acid truncated H3 provides structural constraints that can be used to challenge or validate our current proposed crystal structure of LSD1/CoREST on the nucleosome.

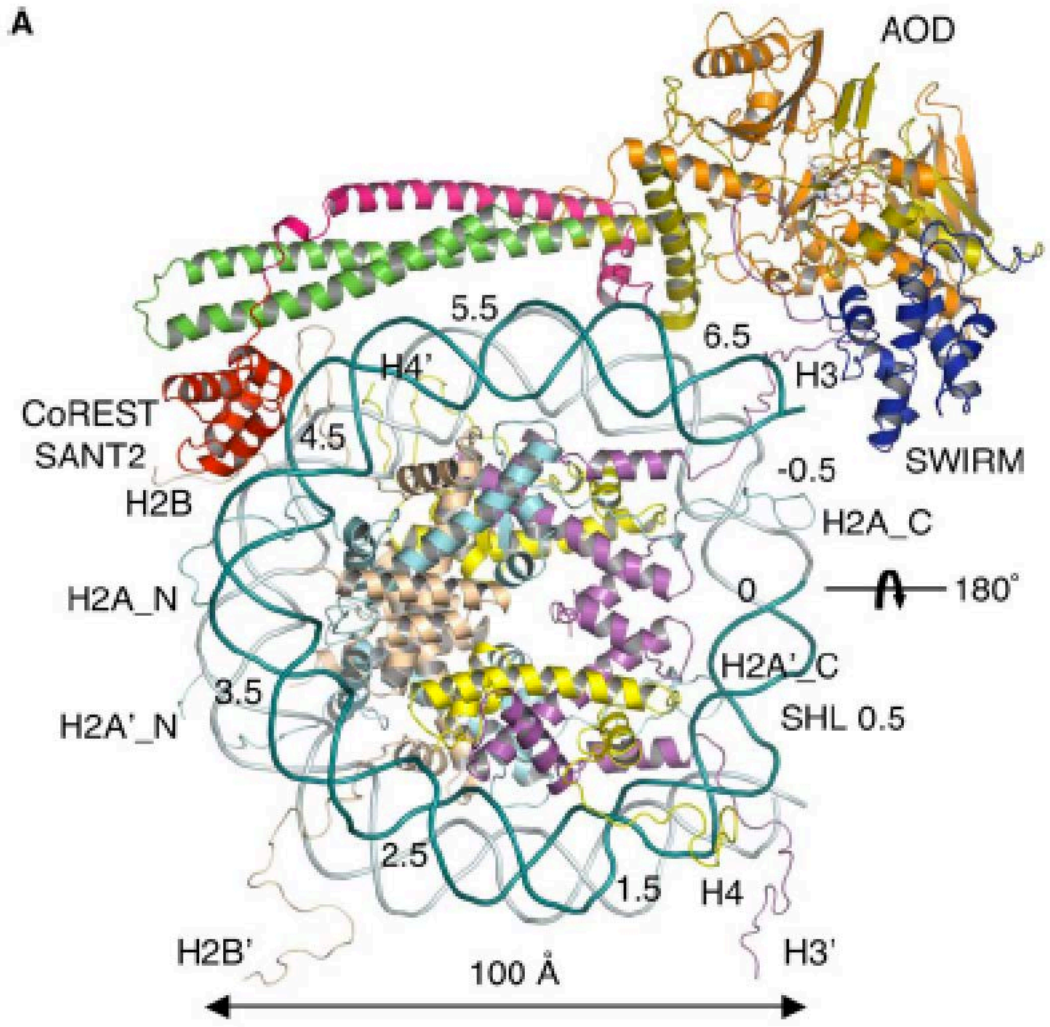

Figure 1-5: A proposed model of LSD1/CoREST on the nucleosome

Of particular note is the location of the H3 tail in relation to AOD. Also of note is the SANT2 interaction with the nucleosomal DNA. Image Source: Yang et al, 2006 (Figure 7A)

Several histone demethylases have been tied to the development of a variety of cancers. Cancer is caused by a loss of cell cycle control: overexpression of genes promoting cell division or repression of tumor suppressor genes leads to tumorigenesis (Audia and Campbell, 2016). LSD1 specifically has been shown to prevent the differentiation and maintain the malignant phenotype of neuroblastoma cells. Additionally, the protein has been shown to relevant activity in acute myeloid leukemia, breast cancer, and Ewing's Sarcoma (Audia and Campbell, 2016).

Further understanding the mechanism of LSD1's demethylase activity, specifically how it interacts with the nucleosome, could lead to the development more acutely targeted drugs.

1.5 Experiments Performed

In order to determine whether internal histone truncations could prevent LSD1/CoREST activity on the nucleosome, I created nucleosomes with a dimethyl lysine analog at position 4 on the truncated histone H3 variants. As stated in the prior section, four H3 mutants were created with amino acid truncations of 12, 15, 18, and 20 amino acids. Each of these mutants also had the lysine at position four mutated to a cysteine residue. Following the purification of these mutants, each mutant had an alkyl group ($C_2H_4NH_3$) added the sulfur of the cysteine residue using the method described by Simon et. al. This created a side chain resembling lysine on the cysteine residue, the only difference being a sulfur atom in place of the second carbon of a lysine side chain. Two methyl groups were then added to this residue. Using the lysine analog in place of a regular lysine allows for specific methylation of the amino acid at position four (Simon et al, 2007). Figure 1-6 demonstrates the alkylation method and shows the structure of the methyl lysine analog.

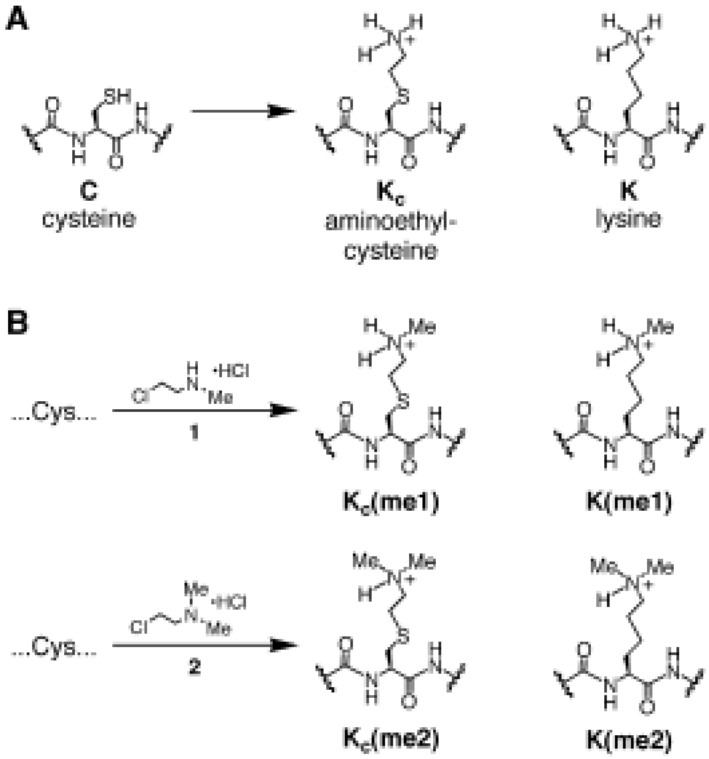

Figure 1-6: Formation of Methyl Lysine Analogs

This diagram shows the formation of methyl lysine analogs via alkylation of a cysteine residue. The system allows for a specific amount of methyl groups to be added to the aminoethylcysteine. For my experiments, 2 methyl groups were added to the lysine analog. Image Source: Simon et al, 2007 (Figure 1A-B)

Following alkylation of the H3 mutants, they were incorporated in nucleosomes with canonical H2A, H2B, H4, and 197 bp DNA. Prior studies conducted by our laboratory showed that longer extranucleosomal DNA increases the binding affinity of LSD1/CoREST to the nucleosome. The lone exception was the nucleosome containing xH3Δ8x3 (12 amino acid internal truncation). This nucleosome was included reconstituted with 157 bp DNA as the reconstitution occurred before extranucleosomal DNA was implicated in LSD1/CoREST binding to the nucleosome. Following this initial reconstitution, our new proposed structure suggested that a 12 amino acid internal H3 truncation would be insufficient to prevent LSD1 activity on the nucleosome. A demethylase assay was used to measure the activity of LSD1/CoREST on the

modified nucleosomes and demethylase activity was qualitatively measured via western blotting with a primary antibody specific for H3K4me2.

1.6 Summary

LSD1 is an important protein in gene regulation. It demethylates histone H3 primarily at lysine 4, acting on the nucleosome when in complex with CoREST, and leads to the repression of specific genes. Mutations to LSD1 have been implicated in the development of several different forms of cancer. While the structure of LSD1/CoREST is well characterized, much is still unknown regarding how LSD1/CoREST acts on the nucleosome. Our laboratory is currently working to solve the structure of LSD1/CoREST on the nucleosome. Crystallographic and chemical studies performed by Sang-Ah Kim, a post-doctoral student in the Tan laboratory, suggested that the active site of LSD1 is located 49 angstroms from the point where the H3 tail exits the nucleosome core. To confirm this, I worked with others in the laboratory to create histone H3 mutants with internal truncations between 12 and 20 amino acids and incorporated these mutants into nucleosomes. Via a demethylase assay, I was able to confirm that 18 and 20 amino acid truncations in the H3 tail, but not 12 or 15 amino acid truncations, eliminated LSD1/CoREST activity on the nucleosome.

Chapter 2 : Materials and Methods

2.1 Nomenclature Guide

- Δx following the name of a protein refers to the truncation number of that protein. For instance, xH3Δ7x3 refers to truncation number 7 of xH3x3
- The truncation number is followed by the protein version number. Using xH3Δ7x3 as an example again, the x3 refers to the fact that version 3 of xH3 was truncated
- The x prior to all histone proteins used in these experiments refers to Xenopus laevis, the species from which the histone DNA sequence was obtained. This sequence is highly homologous to human H3.
- pLysS refers to a plasmid carried by certain strains of *E. coli* that serves to lower background expression in the T7 expression system. It also weakens the cell wall to allow for easier lysing of the cells during purification.
- DE3 indicates that the *E. coli* strain has been modified to include a lac operon that when activated expresses the gene coding for RNA polymerase.
- HST indicates a 10x poly-histidine affinity tag.

2.2 Subcloning

Plasmids encoding histone mutants with internal amino acid truncations were made via Polymerase Chain Reaction (PCR) mutagenesis. These mutations were made on template DNA encoding xenopus H3 with a mutation at amino acid position four, converting the traditional lysine codon to one encoding cysteine.

2.2.1 PCR Mutagenesis

PCR is a process that enables a desired portion of DNA to be replicated in vitro. The reaction requires the following: *Taq* DNA polymerase, a buffer, free deoxyribonucleotides, and a set of primers consisting of a forward primer and a reverse primer. Fluctuating the temperature of the solution enables the exponential replication of DNA to be carried out over a number of cycles in three major steps: denaturation, annealing, and extension.

Denaturation is the first step in a round of PCR. At 95ºC, the complimentary strands of DNA separate. The solution is held at 95ºC for 30 seconds. Following denaturation, the temperature of the solution is dropped to about 60ºC for 1 minute. This stage is called the annealing stage because the primers, both forward and reverse, bind to the template DNA at the lower temperature. The binding of the primers to the template DNA allow for the polymerase to have a starting point for elongation of a strand complementary to the template. This stage, called elongation, occurs at 68ºC. The length of this stage is variable and depends on the length of the DNA being replicated. In general, 1000 bp of extension corresponds to 1 minute of elongation time. Both histone mutants produced were around 3300 basepairs, leading to an elongation time of 3 minutes and 20 seconds.

All rounds of PCR mutagenesis contained 5 ng of template DNA. Additionally, the solution contained 0.7 µl of 10 µM forward primer, 0.7 µl of 10 µM reverse primer, 2.5 µl of 2.5 mM deoxyribonucleotide solution, 2.5 µl of TheroPol buffer, and 0.4 µl of PfuTurbo DNA polymerase. Milli Q water was added to the reaction mixture to bring the final volume of the solution to 25 µl. Eighteen rounds of PCR were run.

2.2.2 DpnI Digestion

Following PCR mutagenesis, 2 µl of the solution was removed and placed in a separate Eppendorf tube as a control for transformation. What remained of original solution was then digested with 5 units of DpnI, a restriction enzyme, at 37ºC for 1 hour. DpnI recognizes the DNA sequence GATC and cleaves DNA at this location. However, DpnI can only act if the adenosine of this sequence is methylated. Methylation of DNA occurs in biological systems like *E. coli*, but because PCR takes place in vitro the DNA produced during mutagenesis will not be methylated. DpnI digestion allows for the cleavage of template plasmids that contain the GATC sequence because the template was isolated from a biological system at some point. Competent cells cannot uptake linearized DNA, thus allowing for the elimination of the parent strain and ideally guaranteeing only the uptake of the desired mutagenized plasmid.

2.2.3 Transformation

Mutagenized plasmid DNA digested with DpnI and the control DNA solution were transformed separately into competent TG1 *E. coli*. A 100 µl suspension of *E. coli* cells were thawed on ice. Competent cells are cells that are capable of the uptake of foreign DNA. After the cells had thawed, 2 µl of plasmid DNA solution was added to the cell suspension. The solution was incubated for 30 minutes on ice, after which it was heat shocked at 42ºC for 30 seconds. Heat shocking allows for the uptake of plasmid DNA by competent cells. Following 15 seconds on ice, 0.5 ml of 2x TY media was added to the solution and the solution was incubated at 37ºC in a shaking incubator for 30 minutes. After removal of the cell solution from the shaking incubator, cells were plated on 2xTY agar plates with ampicillin. Plates were then incubated for between 10 and 18 hours at 37ºC.

2.2.4 PCR Screening

Following transformation, colonies on the 2xTY plates were screened to determine whether they contained the desired plasmid DNA. Several colonies were selected from the plate containing cells transformed with the DpnI digested DNA. The individual colonies selected from the plate were swirled in 100 µl of MilliQ and streaked on a fraction of a 2xTY ampicillin plate. The restreak plate was incubated at 37ºC for 10 to 18 hours. A master mix was produced as the PCR screen solution. For 6 colonies screened, the master mix contained 90.3 µl of MilliQ water, 14 µl of Thermo Pol buffer, 14 µl of 2.5 mM dNTP, 7 µl of 10 µM forward primer, 7 µl of 10 µM forward primer, and 0.7 µl of Pfu polymerase. Six aliquots of 19 µl of the master mix were prepared and 1 µl of each cell suspension was added to a corresponding aliquot. PCR was

performed on the six prepared solutions and the products were run on an agarose gel to confirm the correct, mutagenized plasmid was taken up by the cells.

2.2.5 Plasmid Isolation

Two flasks of 2xTY media were inoculated with two single colonies from the restreak plates. These two colonies came from clones that were screened to be positive during the PCR screen. Five hundred microliters of ampicillin were added to each flask. The cell cultures were grown overnight at 37ºC and 200 rpm in a shaking incubator for 15 hours. Following incubation, the cell cultures were spun at 4000 rpm in a tabletop centrifuge. The resultant supernatant was poured off and additional supernatant was removed via aspiration. The cell pellet was suspended in 5 ml of Lysis Buffer (50 mM glucose, 25 mM Tris-Cl pH 8, 10 mM EDTA, NA). After suspending the cells, 10 ml of NaOH/SDS (0.2 M NaOH, 1% SDS) were added. The resultant solution was shaken until the solution became clear. Immediately upon the mixture becoming clear, it was incubated in an ice bucket for 5 minutes. Ten milliliters of 5 M KAc/2.5 M HAc were added to solution. The solution was mixed by shaking and incubated on ice for another 5 minutes. Addition of the KAc caused the chromosomal DNA to precipitate, forming a pellet at both the top and bottom of the solution. The solution was centrifuged at 4000 rpm for 3 minutes. The supernatant, containing the desired plasmid, was transferred to a polypropylene tube and 12.5 ml of isopropanol was added to the solution. The solution was then centrifuged at 13,000 rpm for 5 minutes at 20ºC. The supernatant was poured off and any residual supernatant was aspirated. Half a milliliter of 70% ethanol was added to the tube to transfer the pellet to an Eppendorf tube. Following transfer of the pellet, the mixture was centrifuged at 13,000 rpm for 1

minutes in a micro centrifuge and the supernatant was aspirated. An additional spin was conducted to remove any residual supernatant. The pellet was suspended in 150 µl of TE(10,50). After suspension, 1.5 µl of RNAase A was added to the solution and the mixture was incubated at 37ºC for 15 minutes. The mixture was vortexed every 5 minutes to break up clumps. After 15 minutes, the solution was phenol/chloroform extracted.

A Sephacryl S400 column was prepared by adding a small portion of glass wool to a 1000 µl pipette tip and filling the tip with S400 resin equilibrated in TE(10, 0.1). The pipette tip was placed on top of a 5 ml polypropylene tube and was spun at 2000 rpm in a centrifuge at 20ºC to remove excess TE(10, 0.1). The collected liquid was disposed. The extracted solution was added to the Sephacryl S400 column and spun at 2000 rpm for 3 minutes at 20ºC. The effluent was collected. This solution contained the plasmid. Plasmid identity was confirmed via restriction mapping and plasmid sequencing.

2.2.6 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to analyze the size of DNA fragments. DNA has a negatively charged phosphate backbone. When a voltage is applied to an agarose gel with DNA loaded on the gel, the DNA will migrate towards the cathode, traveling through the pores of the 3-dimensional matrix of agarose molecules. Low molecular weight DNA fragments migrate more quickly through the matrix.

Agarose gels were prepared by dissolving 0.30 grams of agarose in 30 ml of 0.5x TBE buffer (45 mM Tris Base, 45 mM boric acid, 1.5 mM EDTA). The amount of agarose added to the TBE buffer depended on the size of the fragment being analyzed. Larger fragments were run

on agarose gels with lower concentrations of agarose: a lower agarose concentration leads to a less dense matrix and allows for faster migration of high molecular weight fragments. The mixture was heated in a microwave for 1 minute and 20 seconds to dissolve the agarose. Finally, 1.8 µl of ethidium bromide was added the solution and the solution was mixed via gentle swirling. Following 2 minutes of cooling time at room temperature, the solution was poured into a gel block and a 15 well comb was added. The gel was given 40 minutes to polymerize.

Once the gel had polymerized, the gel was placed in a gel electrophoresis box. The gel was then covered with 0.5x TBE buffer. For each PCR sample to be analyzed, 2 µl of 6x GLB was added to 10 µl of sample. Of this mixture, 10 µl was loaded into a well of the gel. The gel was run for 40 minutes at 125 V, though for smaller fragments the gel was run for a shorter period of time.

2.2.7 Ethanol Precipitation

Ethanol precipitation is a tool used both to further purify DNA and to concentrate and isolated plasmid sample. Precipitation of DNA was achieved by the addition of 0.1 volumes of 3M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol. The solution was vortexed for 5 seconds and the centrifuged at 13K for 10 minutes in a micro centrifuge. The supernatant was carefully aspirated off so as not disturb the DNA pellet. The sample was suspended in the volume of TE(10,0.1) to achieve the desired concentration.

2.2.8 Phenol Chloroform Extraction

Phenol/Chloroform extraction allows for the removal of organic materials such as proteins and RNase A from DNA solutions. An aliquot of 1:1 TE equilibrated phenol and chloroform of equal volume to the sample to be extracted is prepared. The sample is added to this aliquot, briefly vortexed, and the mixture is centrifuged at 13K for 1 minute in a micro centrifuge. The organic (clear), top layer, which contains the DNA, is removed and placed in a new Eppendorf tube. The prior step is then repeated with a fresh aliquot of phenol/chloroform. The top layer is once again extracted and added to a 500 ml aliquot of chloroform. The mixture was centrifuged at 13K for 1 minute and the final DNA sample, the top layer of the separated solution, was removed and stored in a new Eppendorf tube for further use.

2.2.9 UV Quantification of DNA

Following isolation of a plasmid, the solution must be quantified via UV quantification to determine the concentration of DNA in the sample as well as the purity of the sample. A spectrophotometer measured absorbance between 220 nm and 320 nm. The spectrophotometer was blanked with 1.5 µl of MilliQ water. Following blanking, 1.5 µl of the sample was loaded on the sensor. The UV reading returned the concentration of DNA in the solution.

2.2.10 Sequencing

Plasmids were sequenced at the Nucleic Acid Facility located on the fourth floor of Chandlee Laboratory at Pennsylvania State University. All plasmid solutions were diluted to a concentration of 200 ng/µl and 5 µl of each plasmid sample was submitted. The primer used for sequencing was T7 at a concentration of 1 µM. The Nucleic Acid Facility returned both a text file containing the sequence and a chromatogram (analyzed using 4peaks).

2.3 Protein Expression

2.3.1 Small Scale Protein Expression

All histone mutants were expressed from the pST50 transfer vector. These vectors use the T7 expression system outlined in the introduction. Vectors were transformed into Rosetta (DE3) pLysS cells and grown overnight of Tryptone-Yeast Extract (TYE) plates containing both ampicillin and chloramphenicol. A 100 ml flask of 2x TY with both ampicillin (100 µl, 50 µg/ml) and chloramphenicol (100 µl, 25 µg/ml) was inoculated with several colonies from the plate. The cells were grown in a shaking 37ºC shaking incubator at 200 rpm until the OD_{600} of the culture reached 0.5. Prior to induction, 250 µl of sample was removed as an un-induced sample. At this point, the culture was induced with 100 µl of freshly prepared Isopropyl β-D-1 thiogalactopyranoside (IPTG). At each hour, 100 µl of the sample was removed for time point

analysis. After 3 hours at 37ºC, the cells were harvested. The culture was split into two 50 ml aliquots. The aliquots were spun in an SS-34 rotor at 10,000 rpm for 10 minutes at room temperature. The supernatant was poured off and the pellet of cells was suspended in 10 ml of T100 (20 mM Tris-Cl pH 8.0, 100 mM NaCl, 0.5 mM (EDTA, Na_2), 1mM benzamidine, 10 mM 2-mercaptoethanol). The suspended cells were then flash frozen using liquid nitrogen and stored at -20ºC until ready for analysis.

Samples drawn before and during induction were analyzed via SDS-PAGE gel electrophoresis. Samples were mixed with an equal amount of PGLB and boiled prior to being loaded on the gel. The gel was run at 40V for 30 minutes.

2.3.2 Large Scale Expression of Proteins

Following small scale expression of histone mutants to optimize growth conditions, a 3 L large scale expression was performed. The T7-expression vector containing the histone mutant was transformed into Rosetta (DE3) pLysS competent cells as described in section 2.2.3. After overnight incubation, a 100 ml starter culture of 2x TY media was inoculated with approximately 5 colonies from the transformation plate. Antibiotics (100 µl of 50 µg/ml ampicillin and 100 µl of 25 µg/ml chloramphenicol) were added to the starter culture prior to inoculation. The culture was incubated a 37ºC, 200 rpm in a shaking incubator until it reached an OD_{600} of 0.1-0.2. At this point, 6 flasks contain 500 ml of 2x TY were inoculated with 5 ml of started culture. The flasks were then incubated in a large shaking incubator at 200 rpm at 37ºC until their OD_{600} reached 0.5. When the flasks reached the appropriate OD, 500 µl of sample was collected as a 0

hour time point and 500 µl of 0.2 M IPTG was added to each flask. At each hour, 250 µl of sample from two select flasks were collected, spun at 13K in a microcentrifuge, and the pellet was resuspended in 100 µl of PGLB. Following 3 hours of induction, the cells were transferred to centrifuge bottles and spun at 7,000 rpm in an Sorvall SLA-3000 rotor at room temperature for 5 minutes. The supernatant was poured off and the pellet was suspended in T100 (20 mM Tris-Cl pH 8.0, 100 mM NaCl, 0.5 mM (EDTA, Na_2), 1mM benzamidine, 10 mM 2-mercaptoethanol) at ratio of 25 ml of T100 per 1 L of culture.

2.4 Purification of Proteins

2.4.1 Solubility Test of Proteins

Frozen cells were thawed in a lukewarm water bath. The cells were then lysed via sonication. They were sonicated 2x at 40% power for 10 seconds. BL21 (DE3) pLysS contain the gene coding for T7 lysozyme. This gene is included to prevent leaky expression of T7 RNA polymerase before induction of the cells but also serves to weaken the cell wall of the bacteria, making them easier to lyse. Of the sonicated cell sample, 25 µl was removed and 25 µl of PGLB was added to the aliquot. This sample was the whole cell extract.

0.5 ml of sample was transferred to an Eppendorf tube and spun in a microcentrifuge at 13K for 5 minutes. The supernatant was transferred to a separate Eppendorf tube and the pellet was centrifuged at 13K in the microcentrifuge for an additional 30 seconds. Any additional

supernatant was discarded. The pellet was suspended in 0.5 ml of P300 – EDTA. 25 µl of both the suspended pellet solution and the supernatant solution were placed in separate Eppendorf tubes and 25 µl of each PGLB was added to each aliquot. The samples were analyzed on an SDS-PAGE gel. Histones are inclusion bodies and therefore are largely insoluble.

2.4.2 Small Scale TALON Affinity Purification

His tagged proteins were purified using TALON resin. Talon resin was suspended in its storage liquid via inversion of the storage container. 1.5 ml of the suspended resin was transferred to a 15 ml falcon tube and 10 ml of MilliQ water was added to the tube. The tube was inverted several times to mix the water and the resin and following inversion the tube was spun down at 1800 rpm in a table top centrifuge for 2 minutes. The supernatant was poured off and 10 ml of P300-EDTA + 8 M urea was added to the sedimented resin. The tube was again inverted several times to suspend the resin and the resultant mixture was spun at 1800 rpm in a tabletop centrifuge for 2 minutes. The supernatant was poured off and the wash repeated one additional time.

5 ml of sonicated cell extract, from the solubility test, was transferred to a 15 ml Falcon tube. The extract was spun in a table top centrifuge for 5 minutes at 4000 rpm and 20ºC. The supernatant was transferred to another 15 ml falcon tube and the pellet was spun at the prior conditions for an additional 1 minutes. The supernatant was discarded after the additional spin. The pellet was suspended in 1 ml of P300-EDTA + 8M urea. Following suspension, an additional 4 ml of P300-EDTA + 8M urea was added to the mixture. The mixture was allowed to

incubate at room temperature for 10 minutes. The sample was then spun in a tabletop centrifuge at 4000 rpm and 20ºC for 5 minutes. 25 µl of the supernatant was mixed with 25 µl of PGLB. This represented fraction 0. The remaining supernatant was transferred to the 15 ml Falcon tube with the washed Talon resin. The mixture was incubated for 20 minutes at room temperature with mixing by inversion every 5 minutes.

Following incubation, the mixture was spun at 1800 rpm in a table top centrifuge for 5 minutes in order to sediment the resin. The supernatant was transferred to a separate 15 ml falcon tube using a pipette so as to not disturb the sediment. The supernatant containing falcon tube was labeled “Talon Flow Through”.

10 ml of P300-EDTA + 8M urea was added to the sediment and the contents of the falcon tube were mixed by inversion. The mixture was spun at 1800 rpm in a table top centrifuge for 5 minutes at room temperature. The supernatant was transferred to a new falcon tube labeled “Wash A”. The washing procedure was performed a second time, with the supernatant being transferred to a falcon tube labeled “Wash B”.

The resin was suspended in 3 ml P300-EDTA + 8M urea. The solution was transferred to a BioRad BioSpin column. Any flow through was collected in a 15 ml falcon tube labeled “Wash C”. The bound protein was then eluted with P300-EDTA + 8M urea + 100 mM imidazole, added in 3 1 ml increments. Six 0.5 ml fractions were collected.

2.3.3 Alkylation of H3

The amino acid at position four of H3, which was mutated to cysteine, was dimethylated in order to create methyl lysine analogs that closely mimic physiological dimethylated lysine 4. Ten mg of lyophilized H3 variant was dissolved in 930 µl of alkylation buffer (1M HEPES pH 7.8, 4M guanidine-HCl, 10 mM D/L methionine). 20 µl of fresh 1M dithiothreitol (DTT) was added and the mixture was incubated at 37ºC for 1 hour. All steps following the incubation were performed in the dark. 50 µl of fresh 1M 2-chloroethyl-dimethylammonium chloride was added to the histone mixture. The mixture was incubated for 2 hours at room temperature. Following incubation, 10 µl of 1M DTT was added to the solution and the resultant solution was incubated for an additional 30 minutes at room temperature. The second methyl group was added via the addition of 50 µl of 1M 2-chloroethyl-dimethylammonium chloride. The solution was incubated for 2 hours at room temperature. The resultant solution was quenched with 50 µl of 14.4M 2-mercaptoethanol.

The modified histones were dialyzed in H100. Following dialysis, the histones were purified over a SourceS cation-exchange column. The fractions were pooled and dialyzed in 5 mM 2-mercaptoethanol. The dialyzed fractions were then lyophilized and stored at -80ºC. Success of the alkylation was confirmed by LC-MS.

2.3.4 Nucleosome reconstitution/purification

Histone tetramer and dimer were reconstituted from individual histones and purified by laboratory technicians. Nucleosomes were reconstituted using 197 bp DNA with core Widon 601 positioning sequence. H2A/H2B dimer, H3/H4 tetramer, and DNA were mixed in a 2:1:1 ratio in RB high buffer (10 mM Tris-Cl pH 7.5, 1 mM EDTA, 1 mM DTT, 2M KCl). The mixture was then dialyed in RB low buffer (10 mM Tris-Cl pH 7.5, 1 mM EDTA, 1 mM DTT, 250 mM KCl) using a peristaltic pump.

Reconstituted nucleosomes were purified via SourceQ anion exchange HPLC. Fractions containing nucleosome were pooled and dialyzed in NCP storage buffer (10 mM potassium cacodylate pH 6.5, 0.1 mM EDTA). The dialyzed fractions were concentrated using a Vivaspin centrifugal concentrator (10,000 MWCO) and the concentrated, quantitated nucleosomes were stored at 4ºC.

2.3.5 SDS-PAGE

Proteins were visualized via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). SDS is a detergent that carries a negative charge, allowing it to both denature proteins and assign a uniform negative charge to those proteins. The assigned uniform negative charge causes proteins to migrate towards the cathode in a manner relative to their size when in the polyacrylamide matrix. Gels were loaded into a Mini-Protean II electrophoresis apparatus. The apparatus was filled with protein gel loading buffer (10 mM Tris, 76 mM glycine, 0.02% SDS). All protein samples were mixed with an equal volume PGLB (125 mM Bis-Tris pH 6.8,

20% glycerol, 4% SDS, 15% 2-mercaptoethanol, 0.04% bromophenol blue) and boiled for 5 minutes prior to being loaded onto the gel. Following the loading of samples, the gel was electrophoresed for 30 minutes at 10 watts.

Following electrophoresis, the gel was soaked in FIX (45% ethanol, 9% acetic acid) for roughly five minutes. The gel was rinsed with DI water and soaked in stain (0.5% Coomassie Blue, 45% ethanol, 9% acetic acid) for 5 minutes. The stain was poured off and the stained gel was soaked in DESTAIN (7% ethanol, 5% acetic acid) in a 60ºC water bath for an appropriate amount of time.

Gels were prepared by members of our laboratory. The gels used for these experiments were 18% polyacrylamide gels. The separating gel solution was prepared using 8 ml of MilliQ water, 36 ml of 30% acrylamide/0.5% bisacrylamide, 120 µl of bromophenol blue in ethanol, and 15 ml of 3M Tris-Cl pH 8.8. The solution was then deaerated and 600 µl of 10% SDS, 60 µl of tetramethylethyldiamine (TEMED), and 240 µl of 25% Ammonium persulfate (AMPS) were added to the solution. A gel block was set up and the solution was slowly injected into the block using a 15 ml syringe. Water saturated butanol was poured over the gel block. The acrylamide solution was allowed to polymerize at room temperature. Following polymerization, the staking gel solution was prepared containing 5 ml of MilliQ water, 10 ml of 10% acrylamide/5% bisacrylamide, and 4.8 ml of 0.5M Bis-Tris solution. Following the deareating of the solution, 200 µl of 10% SDS, 15 µl of TEMED, and 80 µl of 25% AMPS were added to the solution. The water saturated butanol was poured off the gel block, the gel block was rinsed with MilliQ water, and the stacking gel mix was poured over the gel block. The gels polymerized at room temperature. They were separated following polymerization and were stored at 4ºC surrounded by damp paper towels.

2.4 Enzymatic Activity Measurement

2.4.1 Demethylase Assay

Nucleosomes containing dimethylated H3K4 analogs were diluted to a concentration of 200 nM using 20 mM Hepes pH 7.5, 50 mM NaCl. LSD1/CoREST complex was diluted to 200 nM as well using 20 mM Hepes pH 7.5, 50 mM NaCl. The enzyme was then added to the nucleosome substrate and gently mixed. Aliquots were taken from the mixture at times 0, 5, 10, and 20 minutes. The aliquots were quenched with an equivalent volume of PGLB.

2.4.2 Western Blotting

Samples were run on an SDS protein gel under standard conditions. Following completion of electrophoresis, the gel was soaked in 30 ml of transfer buffer (25 mM Tris-Cl pH 8.3, 192 mM glycine, 20% methanol, 1% SDS) for roughly 5 minutes. Additionally, the nitrocellulose filter was soaked in a separate store of transfer buffer. The filter was placed on top of the gel and the pair were surrounded on both sides by a piece of blotting paper and a fiber pad, also soaked in transfer buffer. The proteins were transferred at 100V for 60 minutes in the cold room in an apparatus filled with constantly stirred transfer buffer and an ice block. Following

transfer, the membrane was equilibrated in 30 ml of TBE buffer (25 mM Tris-Cl pH 8, 150 mM NaCl) for 5 minutes at room temperature. The membrane was then transferred to pre-incubation buffer (25 mM Tris-Cl pH 8, 150 mM NaCl, 2% nonfat dry milk) and incubated on a rocker at room temperature for 30 minutes. The membrane was then washed 2 times in TTBS (25 mM Tris-Cl pH 8, 150 mM NaCl, 0.05% Tween 20). Following the second wash, the membrane was incubated in 10 ml of TTBS with 2 µl of rabbit antibody specific for H3Kme2 (dimethylated H3) for 1 hour on a rocker for 1 hour. The membrane was then washed 3 times with 50 ml TTBS. The washed membrane was incubated at room temperature on a rocker for 1 hour in 30 ml of TTBS with 3 µl of secondary antibody (donkey anti-rabbit antibody conjugated to horseradish peroxidase). The membrane was washed three times with 50 ml TTBS. ECL detection solution were mixed evenly and pipetted on the membrane. The membrane was allowed to sit at room temperature for 2 minutes and was then wrapped in Saranwrap. The wrapped membrane was then exposed to X-ray film for 30 seconds and developed.

Chapter 3 : Results and Discussion

3.1 Expression of Histone Mutants

The two histone mutants that I subcloned, expressed, and purified were xH3Δ7x3 (2-20, 36-136, C110A, K4C) and xH3Δ8x3 (2-20, 33-136, C110A, K4C), which encoded Xenopus laevis histone H3 with 15 and 12 amino acid internal truncations, respectively. The amino acids deleted were outside of the histone tail region. Additionally, both mutants had the cysteine at position 110 mutated to an alanine and the lysine at position 4 mutated to cysteine. The cysteine at position 4 was later alkylated to create a methyl lysine analog. The cysteine at position 110 was mutated to an alanine to avoid alkylation of this residue. Each histone mutant was expressed in Rosetta (DE3) pLysS cells in a 3 liter expression at 37ºC. Our laboratory typically expresses histones in BL21 (DE3) pLysS cells. There is no specific reason why the histone mutants that I expressed were expressed in Rosetta (DE3) pLysS *E. coli*. Because our laboratory frequently expresses histones, the ideal conditions for over expression of the protein were already established. Following a 3 hour induction, xH3Δ7x3 expressing cells and xH3Δ8x3 expressing cells were harvested (Figure 3-1). The solubility of both histones was also tested (Figure 3-2). Both recombinant histones were expressed on the initial attempt without complications. xH3Δ10x3 (2-20, 39-136, C110A, K4C) and xH3Δ11x3 (2-20, 41-136, C110A, K4C) were subcloned by another undergraduate in our laboratory, Turner Pecen. These H3 mutants contained 18 and 20 amino acid internal truncations,

respectively, outside of the H3 tail region. The additional mutants were expressed by Bryan Tornabene.

Figure 3-1

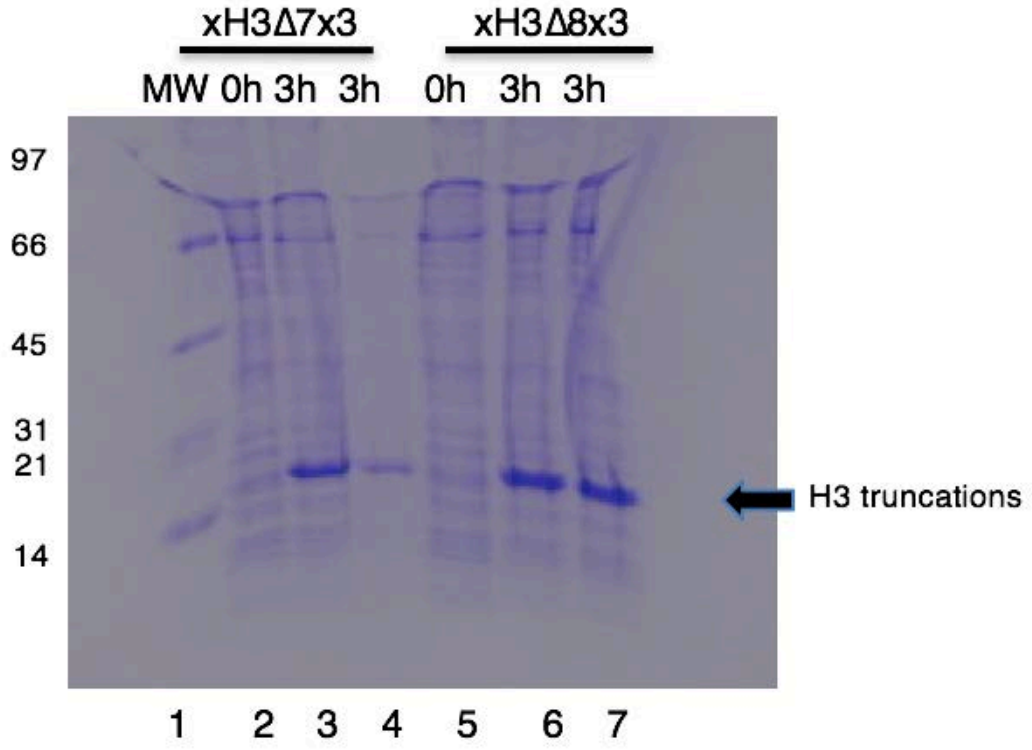

Figure 3-1: Expression of HST tagged xH3Δ7x3 and xH3Δ8x3 in Rosetta (DE3) pLysS cells at 37ºC

The gel was an 18% polyacrylamide SDS gel stained with Coomassie blue. Lanes 2-4 represent xH3Δ7x3, which migrates around 14 kD. Lane 2 contains the uninduced sample while lanes 3 and 4 are samples taken from two separate flasks after 3 hours of induction. Lanes 5-7 contain xH3Δ8x3, which also migrates around 14 kD. Lane 5 contains an uninduced sample while lanes 6 and 7 contain samples from two flasks 3 hours after induction.

Figure 3-2

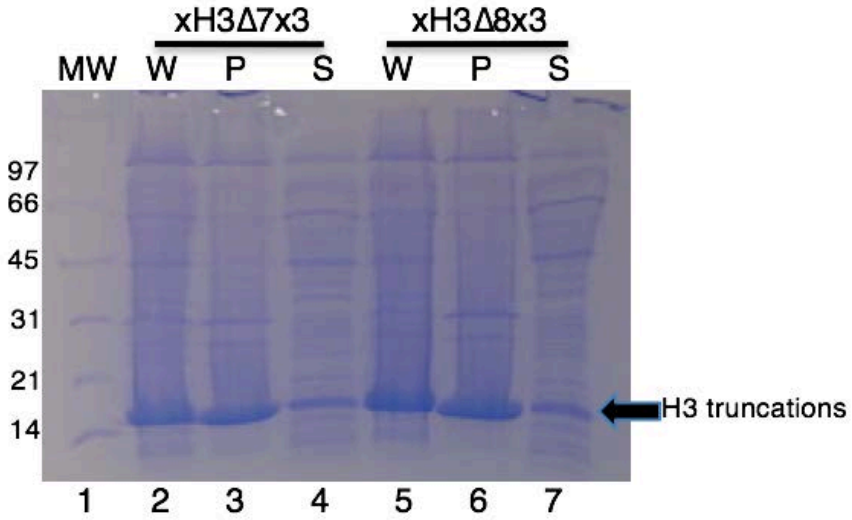

Figure 3-2: Solubility test of both xH3Δ7x3 and xH3Δ8x3

Lanes 2-4 contain the whole cell extract, pellet, and supernatant, respectively, from the solubility test for xH3Δ7x3. As can be seen from the gel, most material is in the pellet which indicates the histone mutant is insoluble. Lanes 5-7 contain the whole cell extract, pellet, and supernatant, respectively, from the solubility test for xH3Δ8x3. As was the case with the other recombinant histone, most of the material is insoluble. This result was expected as histones are inclusion bodies.

3.2 Purification of Histone Mutants

I partially purified the histone H3 truncations using an inclusion body preparation to isolate the insoluble fraction of the cells (Figure 3-3). Further purification using size exclusion and cation-exchange chromatography both performed under denaturing conditions were performed by Kevin Thyne and Michael Doyle, research technicians in our laboratory. The yields of all recombinant histone H3 versions made for this experiment are listed in the table below. Histone mutant yields are listed in Table 3-1.

Table 3-1: Yield of protein purification for histone H3 truncation mutants

Histone H3 Version	Yield (milligrams)
xH3Δ7x3	69.8
xH3Δ8x3	91.9
xH3Δ10x3	Yield unavailable
xH3Δ11x3	Yield unavailable

Figure 3-3

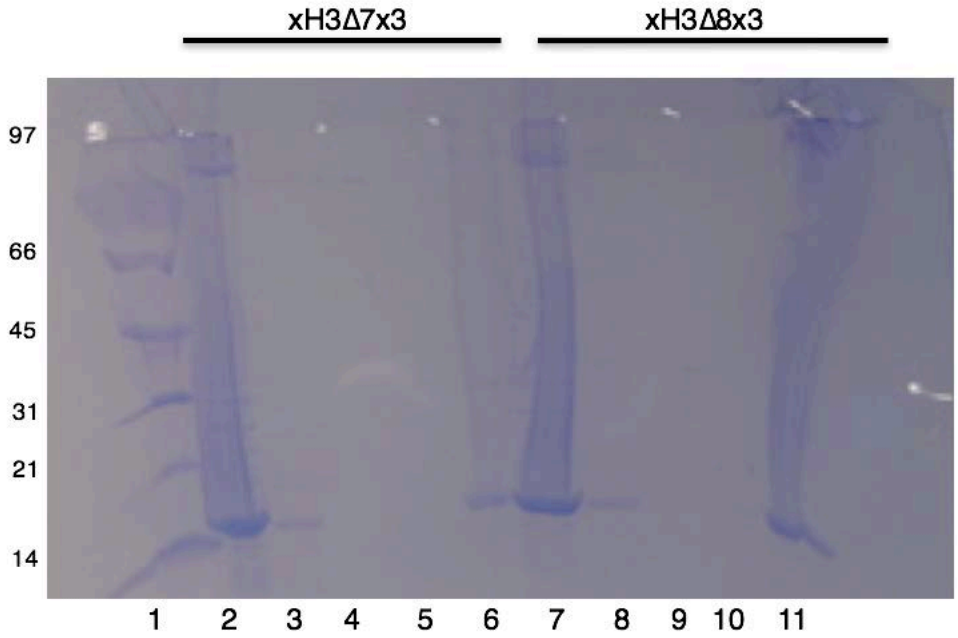

Figure 3-3: Inclusion body prep of xH3Δ7x3 and xH3Δ8x3

Lane 1 is the molecular weight marker. Lane 2 is the whole cell extract which shows the xH3Δ7x3, with a molecular weight ~14kD, is still in the pellet. Lanes 3-5 are the triton washes for xH3Δ7x3 while lane 6 is the whole cell extract wash for xH3Δ7x3. Lane 7 is the Triton whole cell extract which shows the xH3Δ8x3, with a molecular weight ~14kD, is still in the pellet. Lanes 8-10 are the triton washes for xH3Δ8x3 while lane 11 is the whole cell extract wash for xH3Δ8x3.

The purified histone H3 truncations were alkylated to form methyl lysine analogs representing H3K4me2. I alkylated xH3Δ7x3 and xH3Δ8x3 while Bryan Tornabene alkylated full length xH3x3, xH3Δ10x3, and xH3Δ11x3. In order to create methyl lysine analogs, the cysteine at position 4 was alkylated. 2-chloroethyl-dimethylammonium, an electrophile, attacked the sulfur of the cysteine side chain (Simon et al, 2007). The reaction led to the formation of HCl and aminoethylcysteine, an analog to lysine. Additional methyl groups were added by quenching the reaction with reducing agent DTT followed by repeated addition of 2-chloroethyl-dimethylammonium. Alkylation was confirmed to be successful by mass spectrometry following the purification of the alkylated histones (figure 3-6). Dimethylated histones were purified using HPLC. The yields of the alkylated recombinant histones listed in Table 3-2.

Table 3-2: Yield of alkylated histone H3 truncation mutants

Histone H3 Version	Yield (milligrams)
xH3Δ7x3me2	33.6
xH3Δ8x3me2	37.2
xH3Δ10x3me2	35.4
xH3Δ11x3me2	37.2

Figure 3-4

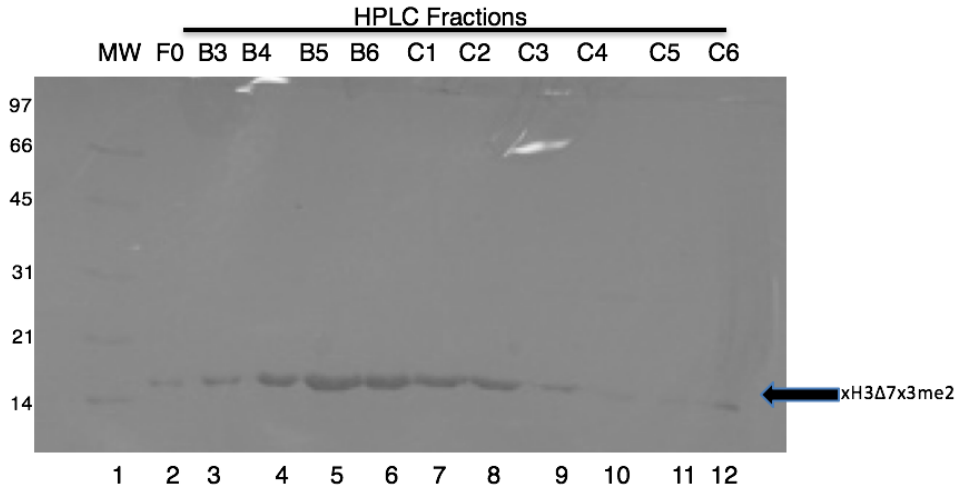

Figure 3-4: SDS protein gel showing elution of alkylated xH3Δ7x3 in HPLC fractions

The Coomassie-stained HPLC gel showed that the alkylated histone eluted in fractions B3-6 and C1-6. The protein was purified on a Source S10 cation-exchange urea column. These fractions were pooled and dialyzed.

Figure 3-5

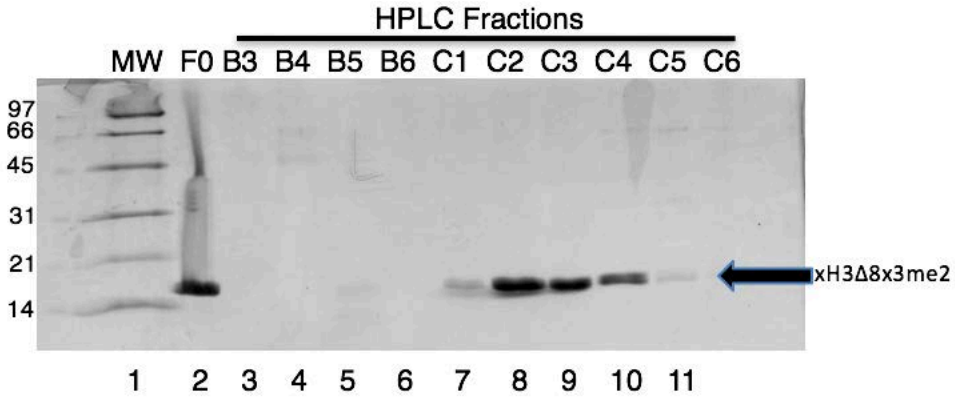

Figure 3-5: SDS protein gel showing elution of alkylated xH3Δ8x3 in HPLC fractions

The Coomassie-stained HPLC gel showed that the alkylated histone eluted in fractions C1-5. The protein was purified using a Source S10 cation-exchange urea column. These fractions were pooled and dialyzed.

Figure 3-6

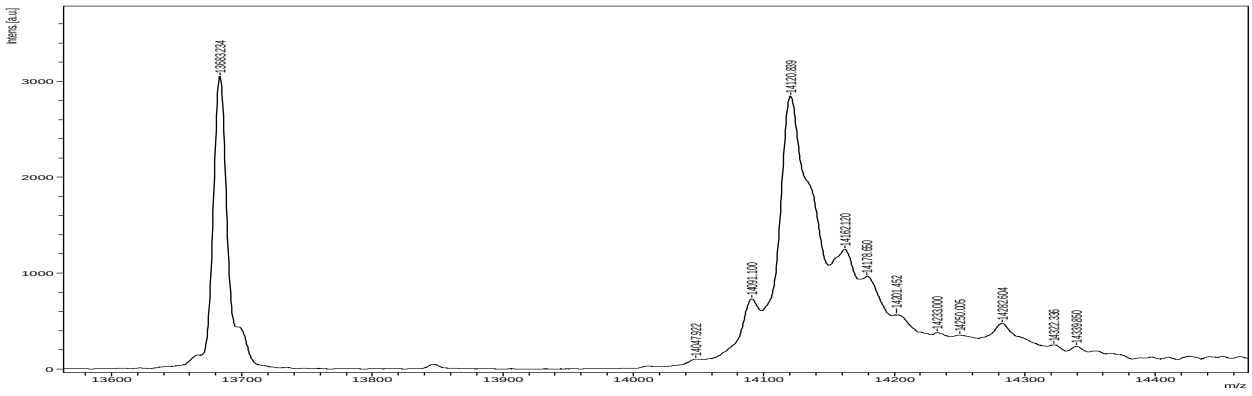

Figure 3-6: Mass spec readout of xH3Δ8x3me2

The figure shows that xH3Δ8x3 had the correct group added to the amino acid at position four. The first peak represents the native xH3Δ8x3. The second major peak is xH3Δ8x3me2. Because the alkylated histone was stored in a buffer containing urea, it is likely that the peaks following the second major peak represent varying levels of carbamylation. Carbamylation can be reduced by immediate dialysis of the alkylated histones.

3.3 Nucleosome Reconstitution

Nucleosomes were reconstituted by Mike Doyle, Kevin Thyne, and Bryan Tornabene with the exception of xNCP v308. xNCP v308 contained canonical H2B, H2A, and H4 as well as xH3Δ11x3me2 and 197 bp DNA. This nucleosome contained the dimethylated H3 mutant with a 20 amino acid truncation. The prep yielded 1.56 milligrams of nucleosome xNCP v308 which represents a 35.3% yield. The yields of the other nucleosomes are in the table below. All other nucleosomes were reconstituted with the same H2B, H2A, H4, and DNA with the except of xNCP v278. xNCP v278 has 157 bp DNA. The nucleosome was reconstituted prior to the development of our current structure of LSD1 on the nucleosome which suggested both that a 12 amino acid internal H3 truncation would not be sufficient to curtail LSD1 activity on the nucleosome and that extranucleosomal DNA is crucial in the binding of LSD1/CoREST to the nucleosome (Kim et al, 2015). Because the 15 amino acid internal H3 truncation failed to eliminate LSD1 activity, it does not appear that the varied length of DNA changes the conclusions drawn from this experiment.

Table 3-3: Yield of reconstituted nucleosomes containing H3 truncation mutants

Histone H3 Version	Yield (milligrams)
xNCP v278 (12 aa H3 truncation)	2.3
xNCP v306 (15 aa H3 truncation)	2.3
xNCP v307 (18 aa H3 truncation)	1.9
xNCP v308 (20 aa H3 truncation)	1.6

3.4 Demethylase Assay

LSD1/CoREST activity on the nucleosome was measure via the demethylase assay outlined in the methods section. The demethylation activity was measured via western blotting with a primary antibody specific to dimethylated H3K4. LSD1 activity on five nucleosomes was measured: xNCP v295 (wild type H3), xNCP v278 (12 amino acid H3 truncation), xNCP v306 (15 amino acid H3 truncation), xNCP v307 (18 amino acid H3 truncation), and xNCP v308 (20 amino acid H3 truncation). LSD1 was added to an aliquot of each nucleosome and samples were taken at 0 minutes, 5 minutes, 10 minutes, and 20 minutes. The samples were quenched with an equal volume of PGLB and amount of dimethylated H3K4 remaining at each time point was analyzed through western blotting. LSD1 was able to demethylate xNCP v295, xNCP v278, and xNCP v306. However, as hypothesized, LSD1 was not able to act on xNCP v307 or xNCP v308. Assay results can be viewed in figures 3-7, 3-8, and 3-9.

Figure 3-7

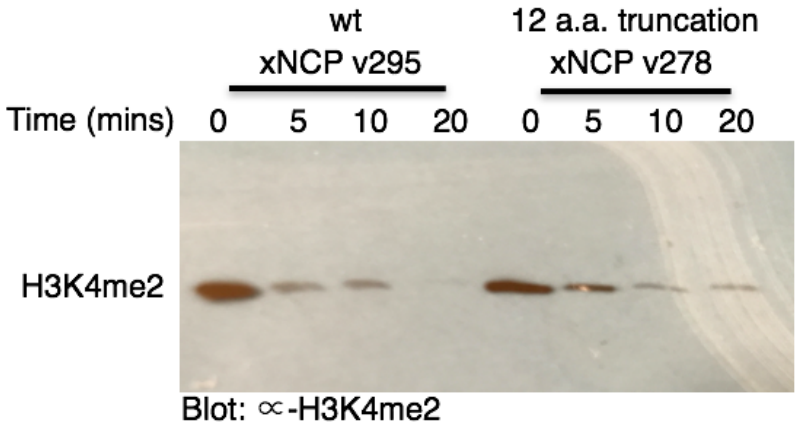

Figure 3-7: Demethylase assay of xNCP v295 (wild type H3) and xNCP v278 (H3 with 12 amino acid truncation)

For both xNCP v295 and xNCP v278, the level of H3K4me2 decreases over time. Interestingly, there is slight more H3K4me2 after 20 minutes of LSD1 exposure in the nucleosome contain H3 with a 12 amino acid internal truncation. Overall, it is clear that the 12 amino acid H3 truncation fails to eliminate LSD1 activity on the nucleosome.

Figure 3-8

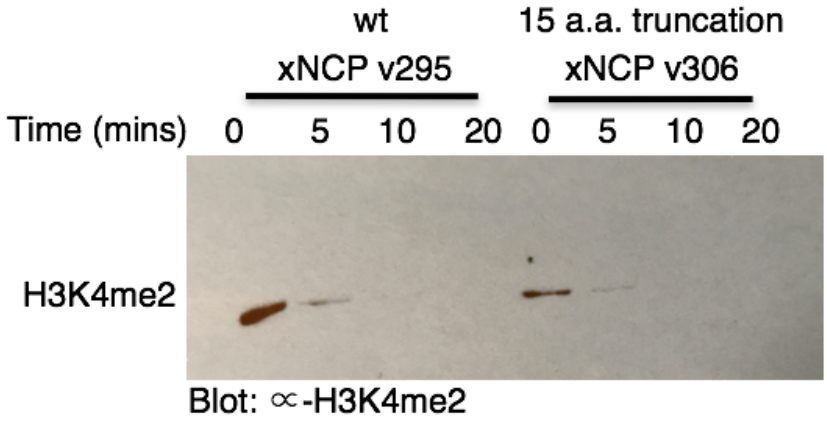

Figure 3-8: Demethylase assay of xNCP v295 (wild type H3) and xNCP v306 (H3 with 15 amino acid truncation)

For both xNCP v295 and xNCP v306, the level of H3K4me2 decreases over time. In both cases, there is no observable amount of H3K4me2 remaining on either nucleosome following 10 minutes of LSD1 exposure. Thus, it is clear that a 15 amino acid internal H3 truncation fails to remove the H3 tail from the LSD1 active site.

Figure 3-9

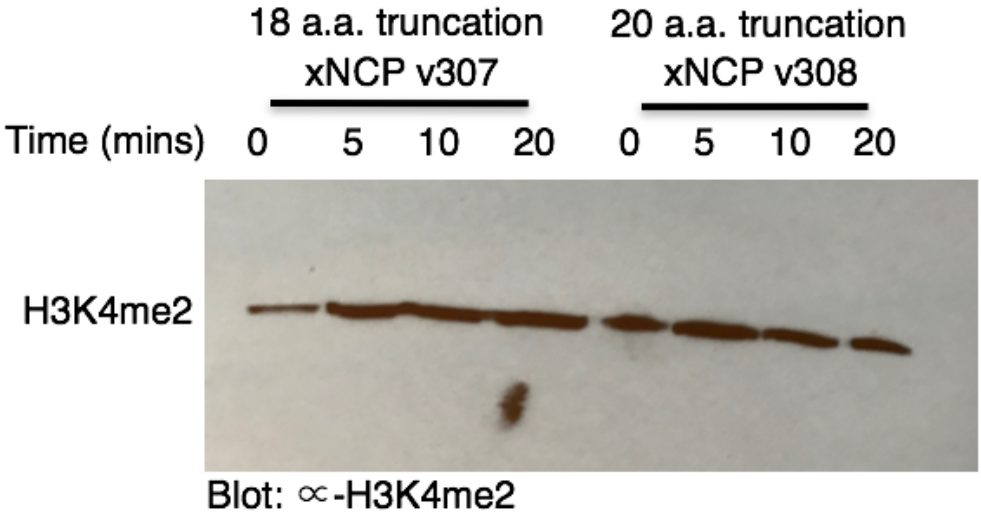

Figure 3-9: Demethylase assay of xNCP v307 (H3 with 18 amino acid truncation) and xNCP v308 (H3 with 20 amino acid truncation)

For both xNCP v307 and xNCP v308, the level of H3K4me2 stays steady over time. Slight deviations in blot intensity over time points can be attributed to slightly different volumes of samples loaded in different gel wells. Additionally, the appears to be slightly incomplete transfer of the band at time point 0 for xNCP v307. Nevertheless, this figure demonstrates that 18 and 20 amino acid internal H3 truncations prevent the activity of LSD1 on the nucleosome. The demethylase assay shown in this figure was run concurrently with the assay in Figure 3-8 with the same LSD1. The demethylation of both nucleosomes in Figure 3-8 shows that the LSD1 was active.

Chapter 4 : Conclusion

4.1 Future Directions and Summary

I was able to express and purify the histone H3 truncations with help from research technicians in the laboratory. The demethylase assay showed that 18 amino acid truncations are sufficient to prevent LSD1/CoREST activity on the nucleosome. Thus, it is not necessary to perform any further experiments involving internal truncations to H3.

Our current structure of LSD1/CoREST on the nucleosome, based on low resolution crystallographic data, shows that the active site of LSD1 resides 49 angstroms from the point where the H3 tail exits the core of the nucleosome when LSD1/CoREST is on the nucleosome. The activity of LSD1/CoREST on nucleosomes containing H3 with 15 amino acid internal truncations paired with the inactivity of LSD1/CoREST on nucleosomes containing H3 with 18 amino acid truncations suggests that this proposed active site distance is of the correct magnitude. The 15 amino acid truncation should reduce the H3 length by roughly 54 angstroms. While this value is greater than 49 angstroms, 54 angstroms is just an estimate and some flexibility should be assumed regarding the LSD1 linker regions. In summary, the elimination of LSD1/CoREST activity on nucleosomes containing 18 or greater amino acid internal H3 truncations lends credence to our current proposed structure of LSD1 and CoREST on the nucleosome.

One possible future experiment that may yield valuable information would involve incorporating the truncated H3 mutants into nucleosomes with various lengths of DNA. Prior studies conducted in our laboratory have shown that extranucleosomal DNA had a measurable

positive impact the binding affinity of LSD1/CoREST to the nucleosome (Kim et al, 2015). The 15, 18, and 20 amino acid internal H3 truncation mutants were all reconstituted in nucleosomes with 197 bp DNA. It is possible that the increased binding affinity observed with additional extranucleosomal DNA is due to some change in the positioning of LSD1/CoREST on the nucleosome. It would be interesting to see whether the 15 amino acid internal truncation would become sufficient to arrest LSD1/CoREST activity on the nucleosome with shorter extranucleosomal DNA extensions. It also is possible that LSD1/CoREST would be able to demethylate a nucleosome containing the H3 mutant with the 18 amino acid internal truncation should the extranucleosomal DNA length be shorter.

My hypothesis is that the extranucleosomal DNA length does not significantly change the distance between the active site of LSD1 and the point where the H3 tail exits the nucleosome. The nucleosome featuring the H3 mutant with the 12 amino acid internal truncation and 157 bp DNA still saw demethylation via LSD1. Additionally, LSD1 acts on H3 when K4 is dimethylated. Physiologically, H3K4 is dimethylated when local genes are being expressed and chromatin is in its active form. This means that the chromatin is in euchromatin form and there should be stretches of extranucleosomal DNA available, in the form of linker DNA, for which LSD1 or CoREST to interact.

The ultimate goal of my project was to help solve the structure of LSD1 on the nucleosome. The additional future work regarding solving the structure of LSD1/CoREST on the nucleosome lies in determining where the specific contacts are between the protein complex and the nucleosome as well as the crystallization of LSD1/CoREST on the nucleosome. Our hypothesis that 18 amino acid internal truncations to H3 would eliminate LSD1 activity on the nucleosome was based on X-ray crystallography data gathered by Dr. Sang-Ah Kim, a

postdoctoral researcher in our laboratory. She has continued to work towards crystallizing LSD1 on the nucleosome at a higher resolution. Overall, I was successful in designing H3 mutants containing internal amino acid truncations and I demonstrated that both 18 and 20 amino acid internal truncations to H3 were sufficient to remove H3K4 from the LSD1 active site.

4.2 Relevance

Epigenetic regulation plays a large role in a variety of maladies, most famously cancer. Further understanding of the mechanism through which LSD1/CoREST binds the nucleosome and demethylates H3K4 could help develop therapeutic treatments for a variety of cancers. Any treatments based on our structure are a way off: LSD1 plays a vital role in healthy cells and finding how it interacts with the nucleosome is just a step towards being able to target the demethylase as a therapeutic target. Still, LSD1 has been implicated in a wide variety of cancers and could emerge as a target for drugs further down the road. Our work will help lead to a greater understanding of epigenetic regulation and help prevent proteins like LSD1 from causing any number of disorders.

Appendix A

Sequence of Histone H3 Truncation Mutants

xH3Δ7x3

```
   1  M  A  R  T  C  Q  T  A  R  K  S  T  G  G  K  A  P  R  K  Q  20
   1 ATGGCCCGTACATGTCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAG 60
       .    |    .    |    .    |    .    |    .    |    .    |

  21  V  K  K  P  H  R  Y  R  P  G  T  V  A  L  R  E  I  R  R  Y  40
  61 GTCAAGAAACCTCACCGTTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTAC 120
       .    |    .    |    .    |    .    |    .    |    .    |

  41  Q  K  S  T  E  L  L  I  R  K  L  P  F  Q  R  L  V  R  E  I  60
 121 CAGAAATCCACCGAGCTGCTCATCCGCAAACTGCCTTTCCAGCGCCTGGTCCGGGAGATC 180
       .    |    .    |    .    |    .    |    .    |    .    |

  61  A  Q  D  F  K  T  D  L  R  F  Q  S  S  A  V  M  A  L  Q  E  80
 181 GCTCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAG 240
       .    |    .    |    .    |    .    |    .    |    .    |

  81  A  S  E  A  Y  L  V  A  L  F  E  D  T  N  L  A  A  I  H  A 100
 241 GCCAGCGAGGCTTATCTGGTCGCTCTCTTTGAGGACACCAACCTGGCTGCCATCCACGCC 300
       .    |    .    |    .    |    .    |    .    |    .    |

 101  K  R  V  T  I  M  P  K  D  I  Q  L  A  R  R  I  R  G  E  R 120
 301 AAGAGGGTCACCATCATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGG 360
       .    |    .    |    .    |    .    |    .    |    .    |

 121  A                                                         121
 361 GCT 363
```

xH3Δ8x3

```
   1  M  A  R  T  C  Q  T  A  R  K  S  T  G  G  K  A  P  R  K  Q  20
   1 ATGGCCCGTACATGTCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAG 60
       .    |    .    |    .    |    .    |    .    |    .    |

  21  T  G  G  V  K  K  P  H  R  Y  R  P  G  T  V  A  L  R  E  I  40
  61 ACCGGCGGAGTCAAGAAACCTCACCGTTACCGGCCCGGCACAGTCGCTCTCCGCGAGATC 120
       .    |    .    |    .    |    .    |    .    |    .    |

  41  R  R  Y  Q  K  S  T  E  L  L  I  R  K  L  P  F  Q  R  L  V  60
 121 CGCCGCTACCAGAAATCCACCGAGCTGCTCATCCGCAAACTGCCTTTCCAGCGCCTGGTC 180
       .    |    .    |    .    |    .    |    .    |    .    |

  61  R  E  I  A  Q  D  F  K  T  D  L  R  F  Q  S  S  A  V  M  A  80
 181 CGGGAGATCGCTCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCT 240
       .    |    .    |    .    |    .    |    .    |    .    |
```

```
  81  L  Q  E  A  S  E  A  Y  L  V  A  L  F  E  D  T  N  L  A  A  100
 241 CTGCAGGAGGCCAGCGAGGCTTATCTGGTCGCTCTCTTTGAGGACACCAACCTGGCTGCC 300
       .    |    .    |    .    |    .    |    .    |    .    |

 101  I  H  A  K  R  V  T  I  M  P  K  D  I  Q  L  A  R  R  I  R  120
 301 ATCCACGCCAAGAGGGTCACCATCATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGA 360
       .    |    .    |    .    |    .    |    .    |    .    |

 121  G  E  R  A                                                  124
 361 GGCGAGAGGGCT 372
       .    |
```

xH3Δ10x3

```
   1  M  A  R  T  C  Q  T  A  R  K  S  T  G  G  K  A  P  R  K  Q  20
   1 ATGGCCCGTACATGTCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAG 60
       .    |    .    |    .    |    .    |    .    |    .    |

  21  P  H  R  Y  R  P  G  T  V  A  L  R  E  I  R  R  Y  Q  K  S  40
  61 CCTCACCGTTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCC 120
       .    |    .    |    .    |    .    |    .    |    .    |

  41  T  E  L  L  I  R  K  L  P  F  Q  R  L  V  R  E  I  A  Q  D  60
 121 ACCGAGCTGCTCATCCGCAAACTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGAC 180
       .    |    .    |    .    |    .    |    .    |    .    |

  61  F  K  T  D  L  R  F  Q  S  S  A  V  M  A  L  Q  E  A  S  E  80
 181 TTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAG 240
       .    |    .    |    .    |    .    |    .    |    .    |

  81  A  Y  L  V  A  L  F  E  D  T  N  L  A  A  I  H  A  K  R  V  100
 241 GCTTATCTGGTCGCTCTCTTTGAGGACACCAACCTGGCTGCCATCCACGCCAAGAGGGTC 300
       .    |    .    |    .    |    .    |    .    |    .    |

 101  T  I  M  P  K  D  I  Q  L  A  R  R  I  R  G  E  R  A        118
 301 ACCATCATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCT 354

       .    |    .    |    .    |    .    |    .    |
```

xH3Δ11x3

```
   1  M  A  R  T  C  Q  T  A  R  K  S  T  G  G  K  A  P  R  K  Q  20
   1 ATGGCCCGTACATGTCAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAG 60
       .    |    .    |    .    |    .    |    .    |    .    |

  21  R  Y  R  P  G  T  V  A  L  R  E  I  R  R  Y  Q  K  S  T  E  40
  61 CGTTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAG 120
       .    |    .    |    .    |    .    |    .    |    .    |

  41  L  L  I  R  K  L  P  F  Q  R  L  V  R  E  I  A  Q  D  F  K  60
 121 CTGCTCATCCGCAAACTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAG 180
       .    |    .    |    .    |    .    |    .    |    .    |
```

```
 61  T  D  L  R  F  Q  S  S  A  V  M  A  L  Q  E  A  S  E  A  Y  80
181 ACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAGGCTTAT 240
       .    |    .    |    .    |    .    |    .    |    .    |

 81  L  V  A  L  F  E  D  T  N  L  A  A  I  H  A  K  R  V  T  I 100
241 CTGGTCGCTCTCTTTGAGGACACCAACCTGGCTGCCATCCACGCCAAGAGGGTCACCATC 300
       .    |    .    |    .    |    .    |    .    |    .    |

101  M  P  K  D  I  Q  L  A  R  R  I  R  G  E  R  A            116
301 ATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCT 348

       .    |    .    |    .    |    .    |    .
```

BIBLIOGRAPHY

1. D'oto A, Tian QW, Davidoff AM, Yang J. Histone demethylases and their roles in cancer epigenetics. J Med Oncol Ther. 2016;1(2):34-40.
2. Kim SA, Chatterjee N, Jennings MJ, Bartholomew B, Tan S. Extranucleosomal DNA enhances the activity of the LSD1/CoREST histone demethylase complex. Nucleic Acids Res. 2015;43(10):4868-80.
3. Audia JE, Campbell RM. Histone Modifications and Cancer. Cold Spring Harb Perspect Biol. 2016;8(4):1-31.
4. Fischer M. Census and evaluation of p53 target genes. Oncogene. 2017; Epub:1-14.
5. McGinty, Robert K. and Song Tan. "Histone, Nucleosome, and Chromatin Structure." *Fundamentals of Chromatin*, edited by L. Jerry Workman and M. Susan Abmayr, 1-28. New York, NY: Springer New York, 2014.
6. Müller S, Almouzni G. Chromatin dynamics during the cell cycle at centromeres. Nat Rev Genet. 2017;18(3):192-208.
7. Pirrotta V. The Necessity of Chromatin: A View in Perspective. Cold Spring Harb Perspect Biol. 2016;8(1):1-16.
8. Sun FL, Cuaycong MH, Elgin SC. Long-range nucleosome ordering is associated with gene silencing in Drosophila melanogaster pericentric heterochromatin. Mol Cell Biol. 2001;21(8):2867-79.
9. Gomez D, Swiatlowska P, Owens GK. Epigenetic Control of Smooth Muscle Cell Identity and Lineage Memory. Arterioscler Thromb Vasc Biol. 2015;35(12):2508-16.

10. Marabelli C, Marrocco B, Mattevi A. The growing structural and functional complexity of the LSD1/KDM1A histone demethylase. Curr Opin Struct Biol. 2016;41:135-144.
11. Mcghee JD, Felsenfeld G. Nucleosome structure. Annu Rev Biochem. 1980;49:1115-56.
12. Yang M, Gocke CB, Luo X, et al. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell. 2006;23(3):377-87.
13. Simon MD, Chu F, Racki LR, et al. The site-specific installation of methyl-lysine analogs into recombinant histones. Cell. 2007;128(5):1003-12.
14. Francastel C, Schübeler D, Martin DI, Groudine M. Nuclear compartmentalization and gene activity. Nat Rev Mol Cell Biol. 2000;1(2):137-43.
15. Croken MM, Nardelli SC, Kim K. Chromatin modifications, epigenetics, and how protozoan parasites regulate their lives. Trends Parasitol. 2012;28(5):202-13.
16. Wysocka J, Swigut T, Milne TA, et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell. 2005;121(6):859-72.

ACADEMIC VITA

Academic Vita of James Lakshman Johnston
jameslakjohn@gmail.com

EDUCATION

Pennsylvania State University (Schreyer Honors College), University Park, PA

B.S, Biochemistry and Molecular Biology (Cell Biology Option), Statistics (Applied Option)
Honors: Biochemistry and Molecular Biology
Thesis Title: The Effects of Internal H3 Truncations of LSD1/CoREST Demethylase Activity
Thesis Supervisor: Dr. Song Tan
Awards: Dean's List Fall 2013-Fall 2016
Academic Excellence Award (Schreyer Honors College)
Provost Scholarship (Pennsylvania State University)
Braddock Scholarship (Eberly College of Science, Pennsylvania State University)
Elizabeth S. and William Henry Gaeckle Alumni Memorial Scholarship (Schreyer Honors College)
Phi Beta Kappa

WORK EXPERIENCE

Penn State Department of Biochemistry and Molecular Biology Spring 2014 – Spring 2017

Undergraduate Researcher, Laboratory of Dr. Song Tan

- Performed bench research in structural biology, particularly its role in epigenetics, using techniques such as PCR, Western blotting, protein expression, and protein purification techniques.
- Co-authored a paper on the construction of DNA marker plasmids to be submitted to peer-reviewed journals.
- Mentored new laboratory students by helping them learn a variety of biochemical techniques.

Children's Hospital of Philadelphia, Division of Plastic Surgery Jun- Aug 2014,2015

Research Fellow

- Conducted clinical research involving cleft palate repair and velopharyngeal insufficiency.
- Co-authored two manuscripts, one of which was published in the Cleft Palate Craniofacial Journal and the other was recently submitted to the same journal.
- Shadowed various CHOP surgeons and clinicians both summers.

ACTIVITIES

Penn State Bioethics Club | *President, Founder*

- Founded the Bioethics Club at Penn State.
- Developed discussion topics and presentations for club meetings.

Science Lionpride | *Member*

- Served as a representative of the College of Science at various recruiting events.
- Gave tours to potential students highlighting the facilities and opportunities available to science majors.

Nexus | *Member*

- Serve as a mentor for incoming science students.
- Worked as part of a committee to create a mentoring program for young science students.

Mount Nittany Hospital | *Clinical Volunteer*

- O**f**fered patients emotional and physical assistance during the discharge process.
- Provided support to medical personnel.

Science 297 | *Learning Assistant*

- Instructed a group of 10 freshmen/sophomore students on getting involved in research at Penn State.
- Determined class activities as part of team of instructors and fellow learning assistants.

PUBLICATIONS

Swanson JW, Johnston JL, Mitchell BT, Alcorn K, Taylor JA. Perioperative Complications in Posterior Pharyngeal Flap Surgery: Review of the National Surgical Quality Improvement Program Pediatric (NSQIP-PEDS) Database. Cleft Palate Craniofac J. 2015 Sep 24. PubMed PMID: 26402723.