THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

EXAMINING THE ROLE OF THE SANT DOMAIN IN BINDING OF LSD1-COREST TO THE H4 TAIL OF THE NUCLEOSOME

VICTORIA SPADAFORA
SPRING 2017

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

Reviewed and approved* by the following:

Song Tan
Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Ming Tien
Professor of Biochemistry
Honors Adviser

Scott Selleck
Department Head for Biochemistry and Molecular Biology

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

LSD1, a lysine specific demethylase enzyme, requires a corepressor, CoREST, to demethylate nucleosome substrates. LSD1-CoREST's demethylase activity provides its gene regulation abilities through transcriptional activation and suppression. Active genes are characterized by methylated lysine four in histone H3 and silenced genes by methylated lysine nine on histone H3. A preliminary low resolution structure of LSD1- CoREST in complex with the nucleosome has been solved in our lab, but further experiments are needed to determine all points of interaction between the two components. The SANT domain is a region of CoREST that is conserved among corepressors and has proven to have a role in chromatin remodeling through histone modifications. Our lab's preliminary structure suggests a potential interaction between the SANT domain of CoREST and the histone H4 tail of the nucleosome. Mutations to the SANT domain could potentially interfere with this interaction, hindering LSD1-CoREST's ability to bind to the nucleosome. I made mutations in the SANT domain, converting amino acids in the acidic patch to the neutral alanine, which could disrupt their ability to interact with the basic H4 tail. Nucleosome binding assays were performed to assess the change in binding affinity of the demethylase and corepressor complex to the nucleosome. Determining the presence and essentiality of aforementioned interactions is important for refining the model of and mechanism through which LSD1-CoREST and the nucleosome interact. It can also help determine target areas for disrupting or enhancing the binding of this complex. LSD1-CoREST's oncogenic properties make these targets a point of interest for therapeutic treatment for cancer.

TABLE OF CONTENTS

LIST OF FIGURES iv

LIST OF TABLES vi

ACKNOWLEDGEMENTS vii

Chapter 1 Introduction 1

1.1 Chromatin 1
1.2 Histone Methylation 3
1.3 Histone Demethylation 4
1.4 Lysine Specific Demethylase 1 (LSD1) 5
1.5 LSD1 and the Demethylation Reaction 7
1.6 CoREST 9
1.7 Application to Biology 13
1.8 Summary 13

Chapter 2 Materials and Methods 14

2.1 Bacteria Media and Strains 14
2.2 Subcloning 15
2.2.1 PCR Mutagenesis 15
2.2.2 PCR Screening 16
2.2.3 Plasmid Isolation 17
2.2.4 Restriction Mapping 19
2.3 Agarose Gel Electrophoresis 19
2.4 UV Quantification of DNA 21
2.5 DNA Sequencing 21
2.6 Creating Vectors and Inserts with Restriction Digest 21
2.7 Gel Purification 22
2.8 Ligation 23
2.9 Transformation 23
2.10 Protein Expression 24
2.10.1 SDS PAGE Electrophoresis 24
2.10.2 Expression Vector System 26
2.10.3 The T7 Expression System 27
2.11 Large Scale Overexpression 28
2.12 Auto Induction 29
2.13 Protein Purification 29

2.13.1 Cell Extract Preparation 29
2.13.2 Metal affinity column purification 30
2.13.3 TEV Protease cleavage 30
2.13.4 Talon Batch Purification 31
2.14 Enzyme Activity Assays 32
2.14.1 HI-FI Binding Assay 32

Chapter 3 Results/Discussion 35

3.1 PCR Mutagenesis and Subcloning 35
3.2 Protein Expression 37
3.3 Protein Purification 40
3.3.1 Batch Purification of LSD1 Complexes 41
3.3.2 Talon Column Purification 42
3.3.3. TEV Protease Cleavage 43
3.4 HI-FI Binding Assay 47

Chapter 4 Conclusion 62

4.1 Summary 62
4.2 Future Directions 63

REFERENCES 64

LIST OF FIGURES

Figure 1. Chromatin Structure 1

Figure 2. The nucleosome core particle. 2

Figure 3. Sites of Important Histone Methylation 4

Figure 4. Known Demethylase Enzymes and their Sites of Demethylase Activity 7

Figure 5. Histone methylation and demethylation reaction by various enzymes 8

Figure 6. LSD1/CoREST Complex Highlighting the Domains of Interest 10

Figure 7. Structure of the LSD1/CoREST complex and its proposed interaction with the nucleosome 11

Figure 8. LSD1/CoREST Structure Depicting Electrostatic Charge of the Protein 12

Figure 9. Polycistronic Expression Vector for Protein Expression in *E. coli* cells 27

Figure 10. HI-FI Scheme: The schematic of the setup of the HI-FI binding assay 34

Figure 11. Scanned HI-FI plate 34

Figure 12. hLSD1 complex v01 Talon Purification 38

Figure 13. Large Scale Expression of hLSD1 complex v35 (23ºC) 40

Figure 14. Batch Purification of hLSD1 complex v37 41

Figure 15. hLSD1 Complex v55 Talon Purification 43

Figure 16. TEV protease digest of hLSD1 complex v36 and v37 44

Figure 17. HPLC Generated Chromatogram 45

Figure 18. Source Q10 Anion Exchange Chromatography SDS-PAGE Gel 46

Figure 19. Talon Purification of hLSD1 complex v54 47

Figure 20. Raw fluorescence HI-FI data graphs. 49

Figure 21. HI-FI Data for hLSD1 complex v01 51

Figure 22. HI-FI data for hLSD1 v36 in triplicate. 53

Figure 23. HI-FI data for hLSD1 v36 on three nucleosomes 54

Figure 24. HI-FI data for hLSD1 v35 in singlet 56

Figure 25. HI-FI data for hLSD1 v37 in duplicate 58

Figure 26. HI-FI data for hLSD1 v37 on three nucleosomes.....59

Figure 27. HI-FI data for hLSD1 v55 in duplicate 60

LIST OF TABLES

Table 1. Construct Names and Corresponding Amino Acid Mutations:. 35

Table 2: hLSD1 Complex Version Numbers and their Corresponding DNA Constructs and Amino Acid Residue Mutations..... 39

Table 3. Summary of all the Kd Values for Each Mutant Complex and their Relative Affinity to hLSD1 Complex v01. 61

ACKNOWLEDGEMENTS

I would like to thank everyone who has shaped my research experience at Penn State University for the past three years. Science is a team effort and every member of the Tan lab past and present has been influential in helping me in some way the past three years. First, I would like to thank Sang Ah Kim for allowing me to work with her on the LSD1 project and for all of her guidance and advice during this time. In addition, I would like to thank the upperclassmen undergraduate students who trained me as I entered the lab including Ryan Henrici and Nick Rizer. The technicians in our lab throughout the past three years have been instrumental in both teaching me and aiding me in many of my experiments. They include Kevin Thyne, Max Kruse, Mike Doyle, Matt Baird, Lauren McCarl, and Bryan Tornabene. I cannot forget to thank my fellow undergraduate seniors who have provided me with many laughs throughout my time in the lab including Turner Pecen, James Johnston, and Dan Friedland. Finally, I would like to thank my thesis mentor, Dr. Song Tan. I have learned so much more than science from him over the past three years. While the techniques that I have developed in the lab will be helpful as I move on to my next step in my career, more importantly are the ways he taught me to think critically, problem solve, and to be confident in myself as a researcher. These are the lessons that will stick with me as I move on from Penn State, and the moments I will value most from my experience in the lab.

Chapter 1

Introduction

1.1 Chromatin

The human genome consists of 3 billion base pairs of DNA arranged in long double helices that stretch for meters in length. In order to fit into the nucleus of the cell, 10 microns in diameter, these DNA strands must be compressed in an orderly fashion (Shilatifard, 2006). DNA is packaged along with proteins into chromatin. Figure 1 shows an unwound chromatin structure and identifies each component involved in the complex.

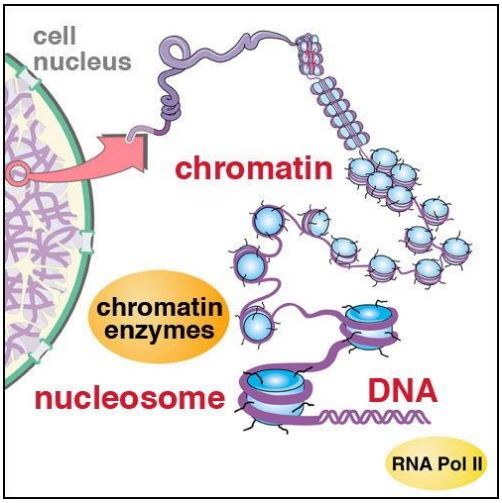


Figure 1. Chromatin Structure: A look into the chromatin structure that allows for the packaging of DNA molecules within the nucleus of the cell. The nucleosome core particles form a beaded string of chromatin with which chromatin enzymes can interact. Any exposed linker DNA allows for the binding of RNA polymerase for active gene transcription. ("Chromatin enzymes" *Tan Lab website*. Song Tan. Penn State University, April 30, 2015.*)*

As depicted in Figure 1, the nucleosome is the basic subunit of chromatin. Eight histone core proteins form an octamer consisting of two copies each of histones H2A, H2B, H3 and H4 (Knapp et al, 2016). Between 145 and 147 base pairs of DNA wrap around this octamer to form the nucleosome core particle (Knapp et al, 2016). These nucleosome core particles connect with each other via extranucleosomal, linker DNA to form a beaded chain that can fold in different conformations to either facilitate or inhibit transcription of genes within the DNA sequence (Igo-Kemenes et al, 1982). A fifth, linker histone, is also present within the structure at the point in the nucleosome core particle where the DNA duplex enters and exits the core and at the axis of symmetry within the nucleosome. This histone is thought to have a role in compaction of the chromatin as well as stabilization of all the components in the nucleosome structure (Kim et al, 2015). Figure 2 focuses on a singular nucleosome core particle and identifies each of the histones within the structure as well as the DNA wrapped around it.

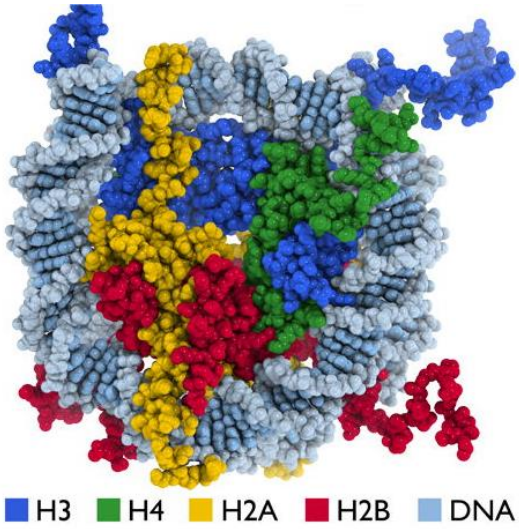


Figure 2. The nucleosome core particle: The nucleosome core particle consists of eight histones: two copies each of H2A, H2B, H3 and H4. Each pair is labeled in a different color and the respective colors are indicated in the key at the base of the figure (Figure 1.2 from McGinty and Tan from *Fundamentals of Chromatin,* 2014).

The nucleosome and chromatin structures are important for gene regulation. In order to transcribe a gene, RNA polymerase requires unhindered access to the gene of interest and the arrangement of the chromatin structure determines when certain genes are "turned on or off." Two configurations of chromatin, euchromatin and heterochromatin, can be formed. DNA within the euchromatin is easier to access because it is less compact, and therefore, this form of chromatin is associated with active transcription (Leontovyc et al, 2011)**.** Heterochromatin on the other hand, is denser and more compressed, making the DNA within it difficult to access and causing the associated genes to be silenced (Leontovyc et al, 2011). Modifications such as DNA methylation and epigenetic alterations of histone proteins in the nucleosome play a role in what chromatin structure exists in a cell and what genes are expressed or silenced (Leontovyc et al, 2011).

1.2 Histone Methylation

Since Vincent Allfrey's discovery in 1968 of histone acetylation, a wide variety of post translational modifications of histones have been studied (Bannister et al, 2011). Each of these modifications function differently in their mechanism of action and therefore have different effects on chromatin structure and overall gene regulation. The post-translational modification focused on in this project is methylation. Methylation differs from modifications such as acetylation and phosphorylation because it does not affect the histones' charge (Bannister et al, 2011). Also, rather than having a consistent phenotype, methylation can differ depending on the specific lysine residue being methylated or the degree of methylation on the lysine of interest, ranging from mono to trimethylation. The main methylation sites exist on the histone tails of H3

and H4. Trimethylation of H3K4, the fourth lysine on the histone 3 tail, is known to be important for transcriptional activation, specifically in the promoter region (Mosammaparast, 2010). Enhancer elements are more often associated with monomethylation of H3K4 than trimethylation whereas active genes in general have trimethylation on H3K4 in the promoter region. Trimethylation at H3K9 is associated with repression of transcription rather than activation. Figure 3 shows important sites of histone methylation on both the H3 and H4 tails. It also indicates the effect of the mono, di, or trimethylation of each amino acid residue on gene expression and whether the site is located at a promoter or an enhancer region.

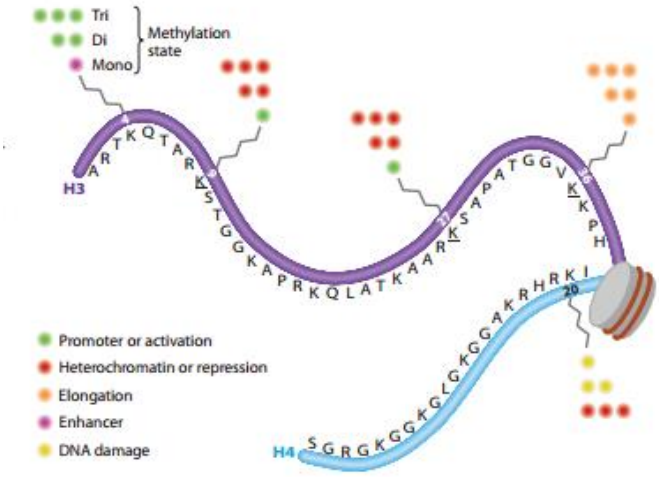


Figure 3. Sites of Important Histone Methylation: The important methylated lysine residues on the H3 and H4 tails are emphasized within this figure. The number of dots represent the extent of methylation, mono-, di-, or tri-, and the color of the dots indicate what type of methylation is being represented (*Figure 1 from Mosammaparast, 2010).*

1.3 Histone Demethylation

In early studies of histone modification, methylation was viewed as an irreversible modification. No histone demethylases were known, and the only apparent demethylation process occurred during histone turnover. During this process, methylated histones could be

replaced with unmodified histones to decrease the overall methylation in a cell. This seemed to be the most plausible scenario as histones and lysine methyl groups had the same half-life, indicating their turnover would occur at the same time (Shilatifard, 2006). Changing the percent of methylated histones within the cell can have a variety of effects. These can include changes in DNA replication, transcription, the overall structure of the nucleosome or the ability of protein complexes to recognize a specific genomic recruitment pattern and be targeted to the nucleosome (Shilatifard, 2006).

1.4 Lysine Specific Demethylase 1 (LSD1)

Because of histone methylation's significant role in gene expression, it is important to study the enzymes involved in regulating the methylation within a cell. The Tan laboratory as a whole is interested in chromatin enzymes and their interactions with the nucleosome to perform their modifying function and regulate gene expression. The focus of this project was the histone demethylase LSD1, and how it interacts with the nucleosome to both activate and repress expression of genes. An important factor in understanding how LSD1 interacts with the nucleosome is knowledge of its overall structure. LSD1 is composed of a SWIRM domain at its N terminus and an amine oxidase domain needed for the demethylation reaction. The SWIRM domain was initially named for being a conserved component of three proteins found in chromatin remodeling complexes Swi3, Rsc8, and Moira. More recently, this domain has also been identified in proteins involved in chromatin regulation such as Ada2 in a histone acetyltransferase and now in the LSD1 demethylase (Da, G. et al, 2005). The function of this domain can differ slightly in each of these proteins, but its consistent functions include aiding

chromatin associated enzymes in targeting the nucleosome and upholding the stability of these enzyme complexes (Da, G. et al, 2005). The amine oxidase domain has two binding purposes, one for binding of the substrate to the enzyme and one for flavin adenine dinucleotide (FAD). LSD1's active site sits between the two binding domains. The amine oxidase of LSD1 has specificity for larger peptides, requiring histone peptides with more than 16 amino acid residues to have efficient demethylase activity. It is large enough to fit the histone tails of the nucleosome into its binding pocket, which is rare for most amine oxidases (Culhane et al, 2007, Mosammaparast, 2010). It is believed that the first twelve amino acids of the H3 tail are important for interacting with acidic patches within the active site of LSD1 and that the SWIRM domain interacts with the N-terminal H3 tail that contains amino acid residues 1-20 (Tochio et al, 2006). The enzyme specifically removes mono and di methylation because the H3 tail changes conformation within the LSD1 active pocket, creating a tighter space within this region and inhibiting large substrates' ability to enter the site. (Mosammaparast, 2010) Despite this conformational change, the pocket still has the ability to fit the trimethylated conformation. As a result, the trimethylated substrates can function as a competitive inhibitor for demethylase activity as the enzyme's active site will be occupied but it will be unable to decrease the methylation count within the genome (Mosammaparast, 2010). To further understand how the pocket can accommodate the trimethylated substrate but cannot function to demethylate it, the full reaction mechanism must be studied.

1.5 LSD1 and the Demethylation Reaction

A compilation of some known histone demethylases and their sites of action within the H3 and H4 tails of the nucleosome are shown in Figure 4. LSD1 is shown as being specific for both H3K4 and H3K9 and therefore having an effect on both transcriptional activation and repression. It is also shown to have an effect on lysine residues methylated to different extents ranging from mono to tri methylation.

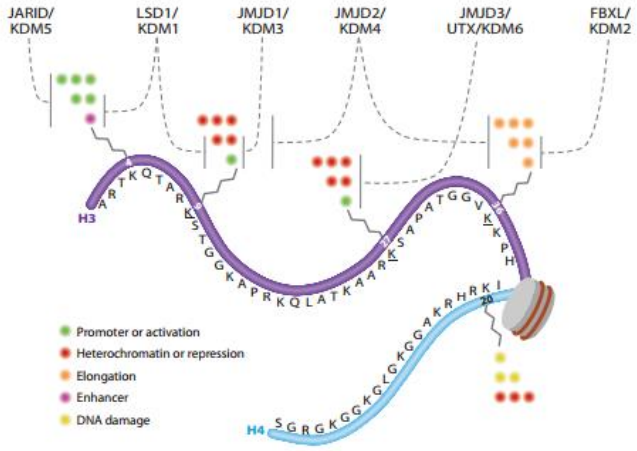


Figure 4. Known Demethylase Enzymes and their Sites of Demethylase Activity: The methylated lysine residues on the H3 tail are depicted along with the names of the demethylase enzymes that are specific for removing methyl groups from each of these sites. LSD1 is shown as being specific for both H3K4 and H3K9 and therefore having an effect on both activation and repression. It is also shown to have an effect on lysine residues methylated to different extents ranging from mono to tri methylation (*Figure 3 from Mosammaparast, 2006).*

It is important to understand the mechanism through which these enzymes demethylate their substrates to better understand the regulation of gene expression. Despite their common function, each demethylase enzyme participates in a different reaction to remove the desired methyl group from the nucleosome. Figure 5 depicts the reactions and byproducts of a few demethylase enzymes.

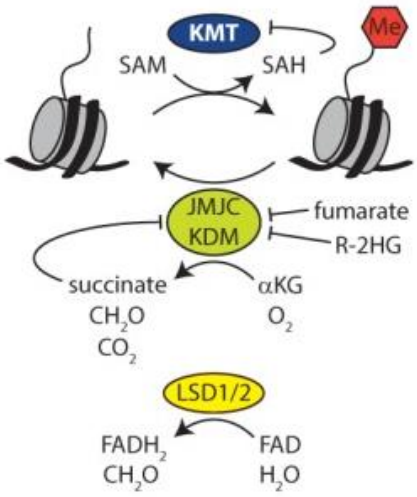


Figure 5. Histone methylation and demethylation reaction by various enzymes. Many enzymes are known for either their methyltransferase or demethylase function. This shows some of the enzymes and the functions that they perform including LSD1's demethylase activity and its production of formaldehyde (*Figure 4 from Knapp et al, 2016.)*

The LSD1 enzyme participates in a demethylase reaction using its amine oxidase domain. In this reaction, flavin adenine dinucleotide (FAD) is reduced and releases a formaldehyde molecule as a byproduct. This process is conserved among many different eukaryote species. LSD1 homologs have been found in *Drosophila (*flies), *C. elegans* (worms), *Arabidopsis* (plants), and *S.pombe* (a fission yeast). The *Saccharomycetales* budding yeast, on the other hand have no identifiable LSD1 homolog encoded in their genome (Mosammaparast, 2010). While this highly conserved reaction allows LSD1 to demethylate mono and dimethylated H3K4, it does not enable it to demethylate trimethylated H3K4. This reaction mechanism can be broken down further to understand why trimethylated lysines are not suitable substrates for LSD1. The FAD is initially oxidized, and the amine converted to an imine via a hydride transfer. A

hemiacetal is formed by hydrating the imine and this hemiacetal spontaneously breaks into the formaldehyde byproduct and the initial amine, resulting in an overall reduction of the FAD (Culhane et. Al, 2007). The amine oxidase of LSD1 is not able to react with a trimethylated lysine because with three methyl groups on the nitrogen, there is no lone pair to participate in the demethylase reaction. Because this is a chemical inhibition and not a kinetic inhibition, the demethylase reaction does not have a preferred substrate between mono and dimethylated lysine residues (Culhane et. Al, 2007).

1.6 CoREST

CoREST is the nonenzymatic protein that aids in LSD1's ability to demethylate the nucleosome and is connected to LSD1 via the tower domain made up of two alpha helices (Mosammaparast, 2010). CoREST proteins have two domains, the EGL-27 and MTA1 homology 2 (ELM2) and SWI3/ADA2/N-CoR/TFIII-B (SANT) that are conserved among regulatory factors of chromatin. Specifically, the second SANT domain focused on in this thesis, has been shown to play a role in histone modifying enzymes' interactions with the nucleosome (Meier et al, 2014). This is why the SANT domain was of interest when studying the binding points of LSD1/CoREST and the nucleosome (Meier et al, 2014). Figure 6 outlines the LSD1/CoREST structure and the domains of interest within this enzyme/corepressor complex.

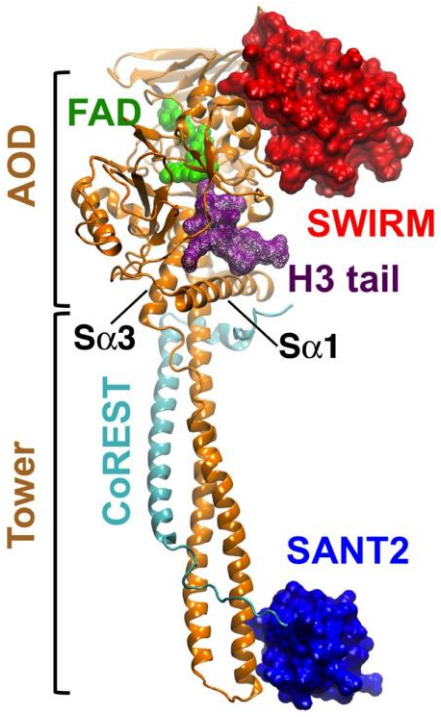


Figure 6. LSD1/CoREST Complex Highlighting the Domains of Interest. This figure highlights the amine oxidase (AOD) and SWIRM domains of LSD1 and the SANT2 domain of the CoREST. The LSD1 enzyme and its corepressor are connected via the indicated tower domain. Each highlighted domain is important for either complex formation, nucleosome binding, or demethylase activity (Baron, 2012).

The CoREST family has three proteins that are related: CoREST 1, 2, and 3. Enzymatic activity of all three CoREST proteins complexed with LSD1 have been studied. CoREST 3 has proven to have decreased demethylation activity in comparison to the other CoREST proteins. CoREST 1 is the protein associated with most LSD1 activity. It has demethylase and deacetylase activity, both creating a chromatin structure that represses transcription. CoREST is composed of LSD1, HDAC1/2 histone deacetylases, and a PHD finger protein (Meier et al, 2014).

By itself, the LSD1 enzyme can demethylate H3K4 on individual peptides and histone dimers or tetramers, but not on the full nucleosome core particle (Meier et al, 2014). CoREST is required for LSD1 to have the ability to interact with the nucleosome as a substrate and

demethylate its histone tails (Meier et al, 2014). Figure 7 presents a proposed crystal structure of the full LSD1/CoREST complex interacting with the nucleosome core particle as proposed by Yang 2006. The specific domains of this complex are once again highlighted on the LSD1/CoREST complex by itself so that these domains and their potential interactions can be identified within the proposed structure.

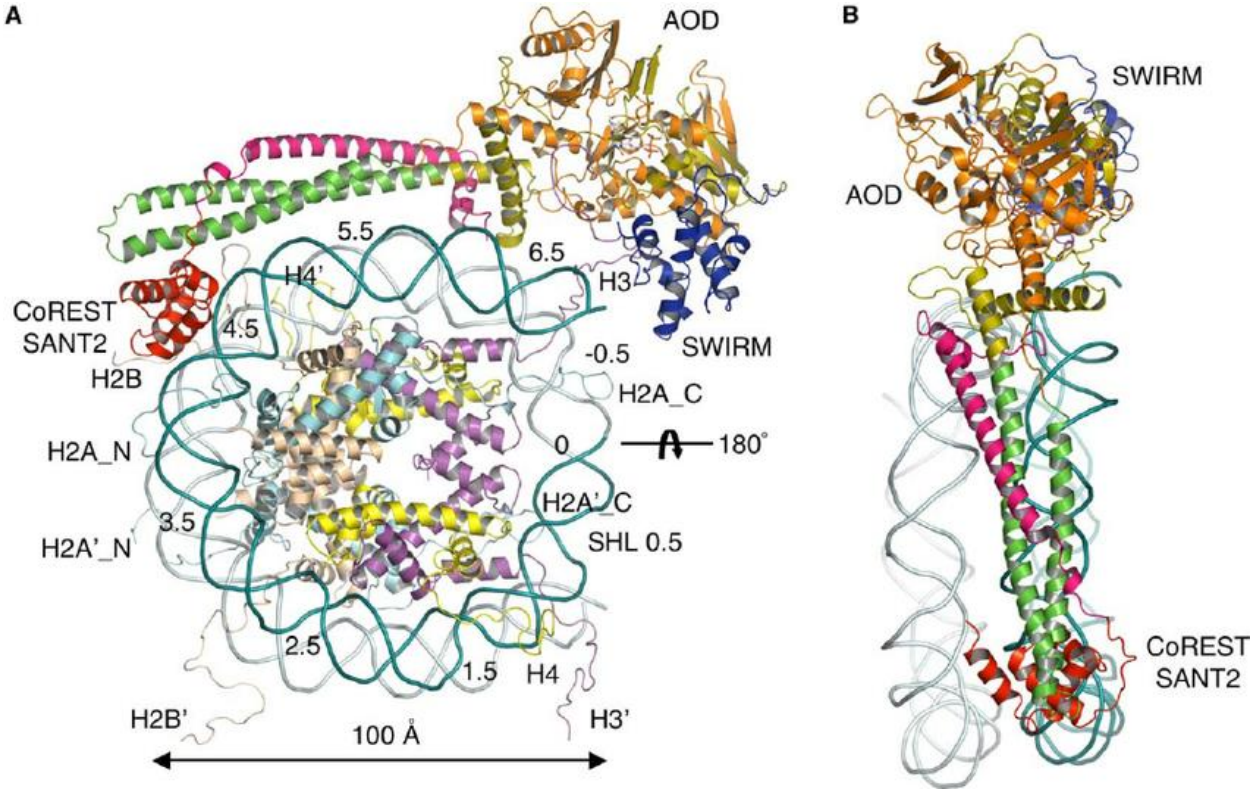


Figure 7. Structure of the LSD1/CoREST complex and its proposed interaction with the nucleosome: (A) A current proposed crystal structure of the LSD1/CoREST complex on the nucleosome by Yang et al, 2006. (B) LSD1/CoREST on its own highlighting the important domains including the amine oxidase domain of LSD1 and the SANT 2 domain of CoREST. These domains are color coded with respect to how they appear on the proposed structure so their interaction points with the nucleosome can be easily identified (*Figure from Yang et al, 2006).*

Previous studies showed that the second SANT domain of the CoREST was particularly critical for the demethylation activity of LSD1/CoREST on the nucleosome (Meier et al, 2014). Figure 8 is a PyMol molecular graphics image that depicts the charges of the amino acids throughout the LSD1/CoREST complex. Patches of basic amino acid residues are identified in blue, acidic residues in red, and neutral residues in white. This helps to visualize potential regions of interest within the complex and the important domains that could be influential in binding to the nucleosome. Within the SANT domain of interest, there is a large patch of acidic amino acid residues. Because the H3 and H4 tails consist of largely basic amino acid residues, this region of opposite charge could potentially interact with the tails of these histones. Therefore, I am interested in the acidic patch of the SANT2 domain of LSD1's corepressor CoREST and whether it is essential to the binding of the LSD1/CoREST complex to the nucleosome to perform its demethylase activity.

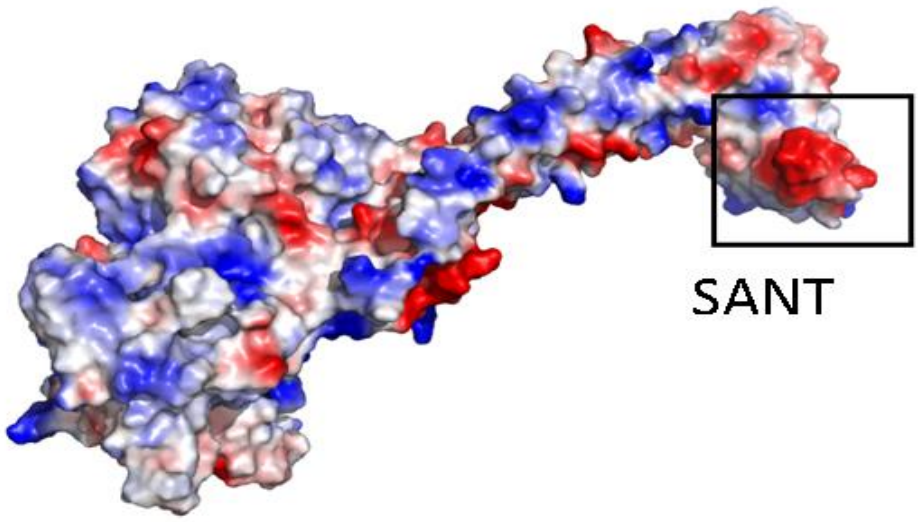


Figure 8. LSD1/CoREST Structure Depicting Electrostatic Charge of the Protein: The above structure of LSD1/CoREST depicts the charges of the amino acid residues within the protein complex. Any acidic residues are highlighted in red, basic in blue, and neutral in white. The SANT domain is further identified and the acidic patch of interest is highlighted in red. Figure prepared by Sang Ah Kim using PyMol molecular graphics software (Schrodinger, 2011).

1.7 Application to Biology

LSD1/CoREST can be significant in many types of cells. A large variety of eukaryotes including flies, worms, plants, and fission yeast, have an LSD1 homolog within their genome that plays an important role in the organism's developmental processes and progression through life. In *Arabidopsis,* the LSD1 homolog functions in the plant's transition to a flowering state, in *C. elegans* mutations in the homolog can cause sterility. In various types of stem cells demethylases have been shown to regulate differentiation of these cells (Mosammaparast, 2010). More recently, LSD1 has been identified as potentially playing a role in the methylation state of the tumor suppressor p53. When p53 is demethylated, it can no longer bind to DNA and therefore loses its ability to act as a tumor suppressor. This has brought LSD1 into question as a potential therapeutic target for cancer (Mosammaparast, 2010).

1.8 Summary

The LSD1/CoREST complex is important in the demethylation of lysine residues in histones H3 and H4 of the nucleosome. It has the ability to regulate gene expression by both activating and repressing transcription. Most recently, this complex has become of interest as a potential therapeutic target in cancer. In order to understand how this complex performs its demethylase function and how to inhibit this function if it plays a role in cancer promotion, its interaction with the nucleosome needs to be further studied. This project specifically focuses on the acidic patch of the SANT domain of CoREST and whether this region is essential to the binding of the LSD1/CoREST complex to the nucleosome.

Chapter 2

Materials and Methods

2.1 Bacteria Media and Strains

For all experiments that required liquid media such as plasmid preps and large scale protein expressions without autoinduction, 2x TY media was used. This media was composed of 1.6 % bactotryptone, 1% yeast extract, and 0.5% NaCl. The media was autoclaved and 50 mg/mL ampicillin was added for plasmid preps and 50 mg/mL ampicillin and 25 mg/ml chloramphenicol was added for large scale protein expression.

TYE media was used to make solid agar plates on which the cells were grown. This media consisted of 1% bactotryptone, 0.5% yeast extract, 0.8% NaCl, and 1.5% agar. The ingredients were added and autoclaved without mixing. Then the autoclaved flask was cooled in the 60ºC water bath while the appropriate antibiotics were added. When thoroughly cooled, the media was poured into petri dishes and allowed to dry and solidify overnight.

Different strains of *E. coli* cells were used for various steps in the project. TG1 cells were used for all of the DNA work including after PCR mutagenesis and DPN1 digestion and after ligation BL21(DE3)pLysS cells were used for protein expression.

2.2 Subcloning

In order to express the human LSD1/CoREST chromatin modifying complex in *E. coli* cells, an expression plasmid had to be created that coexpressed LSD1 and CoREST. Subcloning incorporates a variety of techniques with the end goal of creating a recombinant plasmid with DNA from two separate plasmids in the form of a vector and an insert. These specific subcloning experiments resulted in the incorporation of a mutated version of the CoREST corepressor into a vector containing the wildtype LSD1. Each of these steps is outlined below:

2.2.1 PCR Mutagenesis

Polymerase Chain Reaction (PCR) is a method to amplify segments of DNA using different temperature cycles to denature the double stranded DNA, anneal the forward and reverse primers, and elongate the primers to replicate the DNA. This method was used for PCR mutagenesis by incorporating the point mutations within the primers used to amplify the parent DNA. Forward and reverse primers were designed for five different mutagenesis experiments, changing amino acids within the acidic patch of the SANT domain to neutral alanines. The forward and reverse primers were complimentary and within 12-15 base pairs long. The mutations were introduced within the primers and the primers extended using the parent DNA as a template. The PCR reaction mix included 1 µl of 10 ng/µl parent plasmid, 0.5 µl of each of the forward and reverse primers, 5 µl of the enzyme reaction buffer, 0.5 µl of 2 units/µl polymerase, and 2 µl of 2.5 mM free nucleotides called dNTPs. The mutagenesis and amplification occurred in a thermocycler machine with programmed cycles for denaturing temperature, elongation, and amplification. 2 µl of the amplified PCR mix containing parent plasmid and the new desired

plasmid incorporating the mutation was saved in an Eppendorf tube and labeled ligation A. The remaining mix was digested with 0.5 µl of 10 units/µl DpnI to remove any parent plasmid left in the amplification mixture. DpnI digests methylated DNA only. Any DNA isolated from TG1 cells have methylated DNA at the adenine residue within the DpnI restriction site and all of the parent plasmids were isolated from TG1 cells. This digested mixture was labeled ligation B. Ligations A and B were transformed on two separate TYE plates with ampicillin and incubated at 37ºC for at least 10 hours.

2.2.2 PCR Screening

While ideally all the colonies on the B plate would express strictly the desired plasmid DNA, the colonies were still screened in case not all parent plasmid was digested and some colonies expressing parent DNA remained on the B plate. PCR screening primers were selected based on the ability to produce small PCR products that identify the presence or absence of the mutation. The area of the plasmid incorporating the mutation should be amplified with the screening primers. The reaction mix used for screening incorporates 10x enzyme buffer, 2.5 mM dNTPs, 10 µM of each the forward and reverse screening primers, and 2 units/µl of Pfu polymerase. Each colony being screened was mixed with 100 µl of water to create a cell suspension. 1 µl of this cell suspension was mixed with 19 µl of the reaction mix and amplified within the thermocycler. The PCR products of the screening amplification were digested by endonucleases that would cut at the restriction site introduced by the desired mutation. Any colonies that allow for the digestion to occur incorporate the mutation and can be chosen for further plasmid isolation.

2.2.3 Plasmid Isolation

Based on PCR screening results, colonies were chosen from the transformation plate to inoculate a flask of 100 mL of 2xTY media with ampicillin. Each singular colony chosen was grown in a separate 500 mL Erlenmeyer flask. The cells were grown in a 37ºC shaking incubator overnight for at least 10 hours. The next day, the media and cells were poured into a 250 mL centrifuge bottle and spun down at room temperature at 4000 rpm for 5 minutes in a Heraeus #7570G rotor centrifuge. The supernatant was removed and the cell pellet was resuspended in 5 ml of lysis solution (50mM glucose, 25mM Tris-Cl pH 8.0, 10 mM EDTA, Na) to break the cells open and release their contents. In a 50 ml Falcon tube, the lysed cells were mixed with 10 ml of a solution of NaOH/SDS (0.2M NaOH, 1% SDS), shaken, and immediately incubated on ice for at least 3 minutes. This solution was used to denature the proteins with the SDS and the nucleic acids with the high pH from the sodium hydroxide, and to further break open any organelles that were still fully intact. After incubation, 10 ml of a prechilled mixture of 5M KAc /2.5 M HAc solution is added to the Falcon tube to neutralize the high pH and precipitate the DNA released by the lysed cells and organelles. Some larger RNA molecules and proteins also precipitated at this step. This solution was then mixed thoroughly and incubated on ice for at least three minutes. The circular DNA plasmids renature during this incubation process and are reintroduced to the supernatant of the mixture. After centrifugation for 5 minutes at 4000 rpm at room temperature in Heraeus #7570G rotor, both a pellet at the bottom of the tube and a disk floating above the supernatant are formed. Only the supernatant is transferred to a 50 mL polypropylene tube and mixed with 12.5 ml of isopropanol. The isopropanol precipitated any nucleic acids in the supernatant. After a five-minute incubation of this mixture, the polypropylene tubes were centrifuged in the Sorvall RC 5C Plus centrifuge in the SS-34 rotor at

13K rpm at 20ºC for five minutes to speed up the precipitation process. The supernatant was then removed as it no longer contained DNA, and 500 µl of 70% ethanol was added to the pellet to displace it from the tube wall. The pellet was transferred to an Eppendorf tube, resuspended in the 70% ethanol, and centrifuged for 1 minute at room temperature in the tabletop microcentrifuge. The supernatant was aspirated and the pellet centrifuged for another 30 seconds to extract any extra supernatant, which was once again aspirated off. This final pellet was resuspended in 120 µl TE (10.50) and 1.5 µl of 10 mg/mL RNase A added to the resuspension to digest any of the RNA that precipitated with the DNA. This solution was incubated at 37ºC for 15 minutes in a water bath. While the RNA was being digested, Sephacryl S400 HR spin columns were prepared for each 100 mL prep that was being performed. A 1.5 mL Eppendorf tube was cut in half and the top piece was placed inside of a 5 mL polypropylene tube. Siliconized glass wool was pushed into the bottom of a blue 1 mL pipette tip and this was placed inside the Eppendorf tube. Sephacryl S400-HR resin equilibrated in TE (10,0.1) was filled to the top of the pipette tip and this spun column was spun at 2000 rpm for 3 minutes at 20ºC in the Heraeus #7570G rotor centrifuge rotor. Any liquid that was collected in the bottom of the polypropylene tube was discarded after centrifugation. The digested plasmid solution was extracted with phenol/CIA and CIA alone. Two extractions with 150 µl phenol/CIA were performed first. The mixture was vortexed after each addition and spun at full speed in the microcentrifuge. The top aqueous layer was removed and placed in a new tube with the next 150 µL of phenol/CIA and the process was repeated. After these two extractions, the aqueous layer was extracted with 500 µL of CIA alone. After centrifugation, the aqueous layer of this extraction was placed onto the prepared Sephacryl S400 HR spun column and centrifuged at

2000 rpm for 3 minutes at 20ºC. This time the eluted liquid is the desired plasmid eluted in TE(10,0.1) and was transferred to a new 1.5mL Eppendorf tube for storage.

2.2.4 Restriction Mapping

Restriction mapping was performed after plasmid isolation to determine whether the isolated plasmids had the desired construct sequence. Each plasmid was mapped with two sets of two restriction enzymes. When the plasmid being isolated was the product of PCR mutagenesis, one set was to determine if the correct mutation was introduced into the construct and the second set was to confirm that there was no primer insertion during the amplification cycles of the mutagenesis. When the plasmid being isolated was the product of a subcloning experiment that combined a vector and insert, both sets were used to cut in different areas of the vector and insert to determine that the correct construct was created. The reaction mixture consisted of 2 µl of DNA, 1 µl of the corresponding enzyme buffer, 1 µl of 1 mg/ml BSA, 0.5 µl of 100 mM DTT, and 0.5 µl of each of the two restriction enzymes used for the digestion. This mixture was incubated for 1 hour in a 37ºC water bath. The digested mixture was incorporated with 2 µl of 6x gel loading buffer (GLB) (2.5 mg/ml bromophenol blue, 2.5 mg/ml xylene cyanol, 0.3 g/ml glycerol, 60 mM EDTA) and ran on an agarose gel to analyze the digestion products.

2.3 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to analyze DNA samples and to determine the size of the fragments within these samples. Depending on the fragment size being visualized, different concentrations of agarose gels could be prepared for the best analysis. DNA naturally

has a negatively charged backbone due to the phosphates within it. This natural negative charge allows for DNA molecules to travel towards the cathode, positively charged, within the agarose gel chamber when a voltage was applied. The distance migrated towards the cathode is relative to the molecular weight.

To create the gel matrix 30 mL of 0.5x TBE buffer (45 mM Tris base, 45 mM boric acid, 1.5 mM EDTA) was added to the appropriate concentration of agarose. 3 to 5 mL of deionized water was also added to make up for any water evaporated during the boiling of the mixture. The mixture was left to settle for 5 minutes and then microwaved for 90 seconds to boil the mixture and dissolve the agarose. 1.5 µL of ethidium bromide was added to the agarose solution in order to be able to visualize the bands of DNA after running the gel. The ethidium bromide intercalates within the phosphate backbone of DNA and is fluorescent upon exposure to UV light in a transilluminator. The agarose solution was once again allowed to sit and then poured into a casting mold and set for 40 minutes. A comb was added immediately after pouring to form the wells of the gel. After hardening into an even mixture, the comb was removed and the gel placed into the electrophoresis box filled with 0.5x TBE buffer. Each sample of DNA was loaded into a separate well along with one well designated for the DNA marker that would indicate bands of known DNA lengths for comparison to the unknown fragments. The gels were run at 125V for about 40 minutes and then analyzed after imaging on a UV transilluminator.

2.4 UV Quantification of DNA

The spectrophotometer was blanked with water and set to record at a 320, 280, and 260 nm. DNA absorbs at 260 nm while protein absorbs at 280 nm. The spectrophotometer generated a spectrum with a peak at 260 nm and calculated the concentration in nanomolar.

2.5 DNA Sequencing

Samples were prepared and submitted for sequencing by the Nucleic Acid Core Facility on campus. pST50Tr and pST44 plasmids were not diluted from their plasmid isolation concentrations. 5 µL of 1 µM T7 or T7 term primers were used to sequence the area of interest of the plasmid. 5 µL of the isolated plasmid were submitted along with the choice of primer. DNA sequencing results were analyzed by comparison to the sequence of desired DNA. Any discrepancies or undetected base pairs were analyzed further using an electropherogram. This diagram had a peak correlated to each of the bases and the highest colored peak represented the base pair that was present at that location within the DNA sequence.

2.6 Creating Vectors and Inserts with Restriction Digest

The vectors and inserts were created using asymmetric sticky end restriction sites. The restriction enzymes used left bases overhanging the restriction site and it did this in an asymmetric fashion so that the vector could not ligate on itself and the insert could not ligate into the vector in the reverse direction. The vector and insert are digested with the same restriction enzymes so that the sticky end overhangs line up together to be ligated back together. The vector

and insert were digested in a mixture including 0.2 µg/µl plasmid DNA, a New England Biolabs buffer suitable for the restriction enzymes of choice, 1 mg/ml BSA, 100 mM DTT, 10 to 20 units of each of the restriction enzymes being used. The mixture was placed in the 37ºC water bath for two hours and then the vector and insert were gel purified on an agarose gel.

2.7 Gel Purification

An agarose gel was prepared normally at the appropriate concentration in a preparative casting mold for purification. The largest well comb, the 8 well comb, was inserted in the gel matrix to allow for large amounts of sample to be loaded. A molecular weight marker was still loaded with the sample, but was often loaded with an empty well between the sample and the marker to ensure space for cutting when removing the desired product. As much of the sample as possible was loaded, generally about 30 µl, to ensure high concentrations after gel purification. The gel was again run at 125V for about 40 minutes. A filter apparatus was made containing small 0.5 ml Eppendorf tube within a 1.5 ml Eppendorf tube. The smaller tube was pierced with a heated needle and siliconized glass wool was added to the bottom of this 0.5 ml tube. Before purification, the gel was imaged to ensure the correct fragment was produced and to identify where within the gel the fragment lie. The gel was placed on a UV light box to identify the pink band illuminated by the light. This band was removed with a razor and placed into the purification assembly made previously. The assembly was spun down at 7000 rpm for five minutes. About 50 µl of liquid was eluted from this gel fragment and this eluent contained the desired DNA fragment for ligation.

2.8 Ligation

Sticky ended ligation was used to anneal the vector plasmid to the insert to complete a full plasmid expressing the mutated gene of interest and capable of expressing the protein within the BL21(DE3)pLysS cells. Two ligations were created. Both ligations incorporated 10xT4 DNA ligase buffer, 100 mM DTT, the gel purified vector, and 40 units/µl T4 DNA ligase. Ligation A contained no insert DNA as a control and ligation B had the insert DNA included. This mixture was incubated for about an hour and then transformed onto a 2xTY plate with ampicillin and incubated at 37ºC for at least 10 hours.

2.9 Transformation

Frozen competent cells were thawed out on ice. TG1 cells were thawed out for DNA experiments such as ligations and PCR mutagenesis. BL21(DE3) pLysS cells were thawed out for protein expression experiments. After about 15 minutes of thawing, 2 µl of DNA was added to 100 µl of TG1 cells and 1 µl of DNA was added to 100 µl of BL21(DE3)pLysS cells. The cells were left on ice with the DNA added for 15 minutes and then were heat shocked at 42ºC for 30 seconds. Immediately after heat shock, the cells were placed on ice for 20 seconds and 500 µl of 2xTYmedia was added to each cell suspension tube. Incubation in a shaking incubator at 37ºC was used to allow the cells to express the gene for antibiotic resistance to ampicillin that was coded within the plasmid. Finally, 300 µl of the cells were plated on a TYE plate with ampicillin and put into the 37ºC incubator for ten hours to grow into colonies.

2.10 Protein Expression

2.10.1 SDS PAGE Electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) was used to determine the size of the protein in the sample to help decide if the protein of interest was expressed. SDS is a detergent that both denatures the protein and gives it an overall negative charge, in which the amount of charge is relative to the size of the protein. This allows for the electrophoresis method to determine the size of the protein based on how far it migrates toward the positive electrode in comparison to proteins of known molecular weight in the SDS PAGE marker.

Polyacrylamide (18%) gels composed of a stacking gel and separating gel were used to separate and visualize the protein bands. The separating gel consisted of 36 mL of 30%/0.5% acrylamide, 120 µl of 1% bromophenol blue in ethanol, 15 ml of 3 M Tris-Cl pH 8.8 and 8 mL of water. This first mixture was degassed before 600 µl of 10% SDS, 60 µl of tetramethylethyldiamine (TEMED), and 240 µl of 25% ammonium persulfate (AMPS). After stacking plates in a gel pouring block, the separating gel was injected into the apparatus and the acrylamide polymerized at room temperature before overlaying this gel with water saturated 1-butanol to level the surface of the gel mix. This butanol mix was thoroughly rinsed off after sitting for at least 10 minutes. The stacking solution was mixed incorporating 10 mL of 10% acrylamide/0.5% bisacrylamide, 5 ml of water, and 4.8 ml of 0.5M Bis-Tris. This mixture was also deaerated before adding 200 µl of 10% SDS, 15 µl of TEMED, and 80 µl of 25% AMPS. The stacking solution was poured in the cracks of the gels of the gel block on top of the

solidified separating gel. Combs were quickly placed in between each plate, alternating between 10 well and 15 well combs, and the gel allowed to polymerize at room temperature.

Each piece of the Bio Rad gel apparatus was placed together and the center of the cooling core first filled with protein gel running buffer (PGRB). If no leaks were present, the rest of the buffer chamber was filled to just above the feet of the cooling core. After protein samples were collected, they were mixed with an equal volume of protein gel loading buffer (PGLB). The Eppendorf tubes were then boiled for 2 to 5 minutes. The comb of the gel was removed and seven ul of each sample was added to each well when using a 10 well SDS PAGE gel and 5 ul was added to each well when using a 15 well SDS PAGE gel. A single gel was run at 10 W for 30 minutes. The gel was removed from the apparatus and taken off of the glass plates.

To visualize the bands, the gels were first placed in FIX solution (45% ethanol, 9% glacial acetic acid) for about 5 minutes at room temperature and rocked on a platform during this time. The FIX solution was removed and saved and STAIN (0.5% Coomassie Blue R, 45% ethanol, 9% glacial acetic acid) was added at room temperature between 5 and 10 minutes while sitting on the bench. The STAIN was also poured off into a container for reuse and the gel was rinsed with deionized water to remove any extra unwanted STAIN solution. This was not saved but rather discarded. Finally, a DESTAIN solution (7% ethanol, 5% glacial acetic acid) was added along with a tissue and placed in the 60ºC shaking water bath for 30 to 60 minutes. The DESTAIN was then removed and the gel soaked in deionized water and visualized.

2.10.2 Expression Vector System

The Tan Lab has developed an expression vector system to express multicomponent protein complexes in a single plasmid instead of needing to express each component separately to then reconstitute the complex later. This more efficient system allows for four genes to be placed in four separate cassettes to transcribe a single RNA molecule with four separate coding regions. Each of these coding regions has its own Shine Dalgarno sequence for ribosome binding and a START and STOP codon for translation, but the plasmid as a whole only contains one promoter for T7 to carry out T7 RNA polymerase transcription. Each cassette is also flanked with a different restriction site at each end to allow for easy subcloning of genes into specific cassettes (Tan, 2000).

The second generation of this polycistronic expression system, the pST44 expression system, was used in this project. This new system allows for proteins to be expressed with affinity tags on either termini, such as the hexahistadine tag used on the LSD1/CoREST proteins. In order to insert genes into the expression vector, the vector needs a corresponding transfer vector with equivalent restriction enzyme pairs flanking the gene of interest. This transfer vector for pST44 is the pST50Tr which was used for all of the mutagenesis experiments before transferring the mutated genes into the expression vector (Tan 2005).

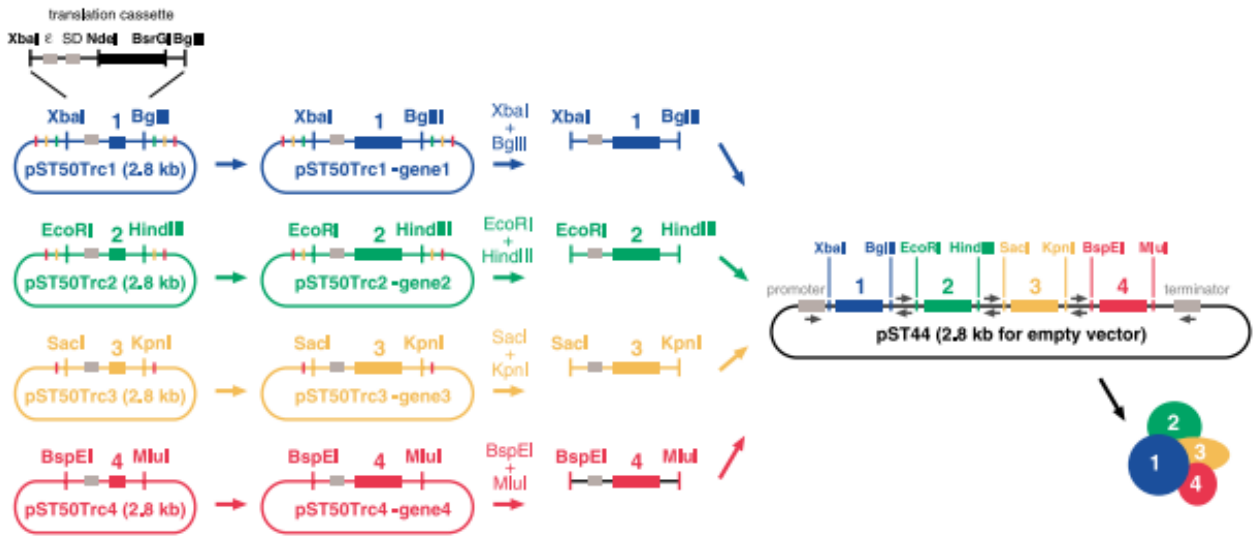


Figure 9. Polycistronic Expression Vector for Protein Expression in *E. coli* cells: Above is the scheme for creating the polycistronic pST44 expression vector. Restriction enzymes can be used to subclone up to four different genes into this plasmid to be coexpressed as a protein complex. *Figure from Tan et al 2005.*

2.10.3 The T7 Expression System

0.1 – 0.3 µg of the desired plasmid were transformed into BL21(DE3)pLysS cells and plated on TYE plates with chloramphenicol and ampicillin. The BL21 strain of *E. coli* has low protease activity and will therefore ensure minimal protein degradation during expression. These cells also contain a lambda prophage and a plasmid expressing the T7 lysozyme gene indicated by (DE3) and pLysS respectively. The T7 RNA polymerase expression system utilizes the lac operon to induce protein expression. The lambda prophage contains the T7 RNA polymerase gene which is regulated by the lac operator. When the lac repressor is bound to the operator, T7 polymerase cannot be expressed and gene expression is turned off. Isopropyl B-D-1-thiogalactopyranoside (IPTG) is a solution that can be added to the cells to induce expression. It has the ability to bind the lac repressor and causes it to release its bind to the lac operator,

allowing the expression of T7 RNA polymerase once again. The T7 polymerase then has the ability to a promoter and allow for the translation of the genes of interest and expression of the desired proteins. This allows for a more controlled expression of protein by IPTG. The T7 lysozyme expressed by the pLysS plasmid can inhibit any production of T7 RNA polymerase when the promoter is leaky, and expressing in the absence of lactose. This will inhibit any undesired RNA polymerase expression, though minimal, and ensure the polymerase is only expressed upon addition of IPTG.

2.11 Large Scale Overexpression

0.1-0.3 µg of the desired plasmid was transformed into BL21(DE3)pLysS cells and plated on TYE plates with chloramphenicol and ampicillin and incubated overnight at 37ºC. Two 100 mL flasks were inoculated with six colonies from the transformation plates at two times during the afternoon, approximately four hours apart, and incubated in a shaking incubator at 21ºC and 220 rpm. The OD 600 of each of these flasks were monitored the next morning. If not between 0.1 and 1.0, the flasks were moved to a 37ºC incubator until the OD 600 increased to between 0.1 and 1.0. At this point 6 to 12 flasks of 500 mL 2xTY were containing 0.5 mL 50 mg/mL ampicillin and 0.5 mL 25 mg/mL chloramphenicol were inoculated with 3 mL of the preculture. These flasks were then grown at 37ºC until their OD increased to between 0.05 and 0.15 and the temperature was decreased to 23ºC. They OD600 was continuously monitored until it reached between 0.6 and 1.0 where expression was induced using 0.5 mL of 0.2M IPTG. For LSD1 expression, the flasks were left in the shaking incubator for 5 hours after induction and then harvested.

Cells were centrifuged at 7000 rpm for 5 minutes at 20ºC for harvest. They were resuspended in P300 EDTA and flash frozen in liquid nitrogen.

2.12 Auto Induction

Towards the end of the project, a method for protein expression was introduced to the lab. This was an autoinduction system that did not require the addition of IPTG to induce protein expression within *E. coli*. The media preparation included a slightly different base media and required an additional preparation of a carbon source to add to the media after autoclaving and before addition of antibiotics and cells. This base media consisted of 2% bactotryptone, 0.5% yeast extract, 0.5% NaCl, 0.6% Na2HPO4, and 0.3% KH2PO4. The 50x carbon source was made with 30% glycerol, 2.5% glucose, and 10% lactose. Both the base media and the carbon source were autoclaved. 10 mL of the 50x carbon source was added to each flask of 500 mL of autoinduction media prior to the addition of any antibiotics or cells. Then 0.5 mL of 50 mg/ml ampicillin and 0.5 ml of 25 mg/ml chloramphenicol was added to each flask along with about 6 ml of the cells from the preculture.

2.13 Protein Purification

2.13.1 Cell Extract Preparation

Cells were thawed in the 30ºC water bath for an hour. The cell extract was split into 5 beakers containing about 40 mL of extract each. They were sonicated at 70 percent power with 14, half second pulses with half a second in between using the Branson Digital Sonifier. Each

beaker was sonicated two times and placed on ice between sonication rounds. The sonicated extracts were poured into polypropylene tubes and centrifuged for twenty minutes at 18,000 rpm and 4ºC in the SS34 rotor. 15 uL of this extract was saved before centrifugation to use as the whole cell extract sample.

2.13.2 Metal affinity column purification

The Talon cobalt affinity column was equilibrated in the cold room. Three buffers were made for the purification. These consisted of a P300-EDTA base buffer (50mM sodium phosphate pH 7.0, 300 mM NaCL. 1mM benzamidine, 5mM B-mercaptoethanol), an elution buffer with high salt concentration (100 mM imidazole), and a wash buffer with low salt concentration (15 mM imidazole). The column was equilibrated with the P300-EDTA base buffer. The soluble extract was then loaded onto the Talon column and the column was washed with the low salt buffer and the protein eluted with the high salt buffer. Every third fraction was sampled and ran on an SDS PAGE gel to see which fractions would be pooled.

2.13.3 TEV Protease cleavage

The pooled fractions were placed in a dialysis tube and dialyzed against 2 liters of T100 (20mM Tris-Cl pH 8, 100 mM NaCl, 10mM B-mercaptoethanol) buffer overnight. Prior to this dialysis, a sample of the undigested tagged protein was taken for later analysis. During dialysis, TEV protease enzyme was added to cleave the hezahistidine affinity tag. The next day the dialyzed pooled sample was centrifuged for 15 minutes at 18,000 rpm at 10ºC using the SS34 rotor. The sample was then loaded onto the SourceQ column and the complex was eluted using a

simple gradient. SDS PAGE was again used to analyze the fractions and the desired fractions were separated and pooled. These fractions were again dialyzed, this time into 2L of H100 (10mM HEPES, 100mM NaCl, 5 mM B-mercaptoethanol). The Vivaspin centrifugal concentrator (30,000 MWCO) was used to concentrate the sample and dynamic light scattering was utilized to ensure there was no aggregation of the complex within the sample. The complex was flash frozen with liquid nitrogen and stored at 20ºC after glycerol was added to 20% weight per volume.

2.13.4 Talon Batch Purification

For some of the mutant LSD1 purifications, an alternative Talon purification method was performed. In this method, the cell extract was mixed with the resin in batch rather than passing the cell extract through the column. Both of these procedures produced comparable purification results for the LSD1 complexes. In this method, the resin was prepared using about ten ml of resin for each version that was purified. The resin was washed once with water and once with guanidine. The resin was then washed again with water and then once with P300-EDTA. Every time the resin was washed with a new solution, it was spun down at 1800 rpm for 5 minutes in the tabletop centrifuge. To bind the sample to the resin, the resin was split into 50 mL falcon tubes. The sample was added on top of the resin and incubated on a turntable in the cold room for 2 hours. After this incubation, the samples were spun down again, the flow through was saved and the resin was transferred to a singular 50 mL falcon tube. The resin was washed two times with P300 EDTA and then 3 times with P300 EDTA containing 20mM imidazole. The

total wash volume was approximately 40 ml. The resin was again transferred, this time to a 15 ml Falcon tube, and washed six times with 6 ml of P300 EDTA with 20 mM imidazole.

2.14 Enzyme Activity Assays

2.14.1 HI-FI Binding Assay

The high-throughput interactions by fluorescence intensity (HI-FI) assay was used to determine the strength of the interaction between the LSD1/CoREST complex and the nucleosome. Each HI-FI plate used, consisting of 384 wells each, was passivated over the course of two days. On the first day, the wells were filled with 1% Hellmanex and incubated at room temperature for 20 minutes. The Hellmanex was then removed by dumping and shaking, and the wells were washed four times with MilliQ water. 1M KOH was then added to each of the wells and the plate again sat for 20 minutes at room temperature. The KOH was then removed from the plate in the same fashion as the Hellmanex, and the plate rinsed again with water four times. The plate was allowed to dry overnight in the hood. The next day, a solution was prepared containing 49 ml (98%) heptane and 1 ml (2%) 1,7- dichlorooctamethyltetrasiloxane and 100 µl of this solution was added to each well, then incubated for 1 minute. A multichannel pipette was used for this step and the plate was passivated in sections in order to minimize the variance in the incubation times. After all sections were passivated, the solution was removed via shaking and the plate washed with MilliQ water at least four times. The plate was again allowed to dry overnight in the ventilated hood.

Fluorescent nucleosomes labeled with Oregon Green 488 at specific sites within the nucleosome were diluted in T75 HI-FI buffer (20 mM Tris-Cl ph 7.6, 5 mM DTT, 5% glycerol,

0.01% NP40, 0.01% CHAPS, 0.1 mg/mL BSA, and 75 mM NaCl) to 20 nM for the mutant experiments and 4 nM for the wildtype experiment. The LSD1/CoREST complexes were compared in a titration that included 14 concentrations of the complex starting at 40 µM down to 0.5 um. The initial dilutions of 80 um LSD1/CoREST complex was prepared in T0 HI-FI (20 mM Tris-Cl ph 7.6, 5 mM DTT, 5% glycerol, 0.01% NP40, 0.01% CHAPS, and 0.1 mg/mL BSA) and adjusted to T75 with the addition of 1M NaCl. The rest of the titrations were prepared with the T75 HI-FI buffer. The LSD1/CoREST complex was added to each well of the plate, 20 ul per well, with the 0 µm control at both the beginning and end of the titration. Twenty µl of the fluorescent nucleosome was then added on top of this titration in each well. The complex and nucleosome were mixed as the plate was spun down at 500 rpm for one minute. The plate was then vortexed at 1200 rpm for two minutes and incubated for 20 minutes at room temperature. The plate was taped and then scanned to detect the differences in fluorescence.

The Typhoon scanner was used to measure the fluorescence. The settings measured excitation at a wavelength of 488 nm and emission at a wavelength of 526 nm. The experiments were performed in triplicate and the data analyzed using the single binding isotherm equation.

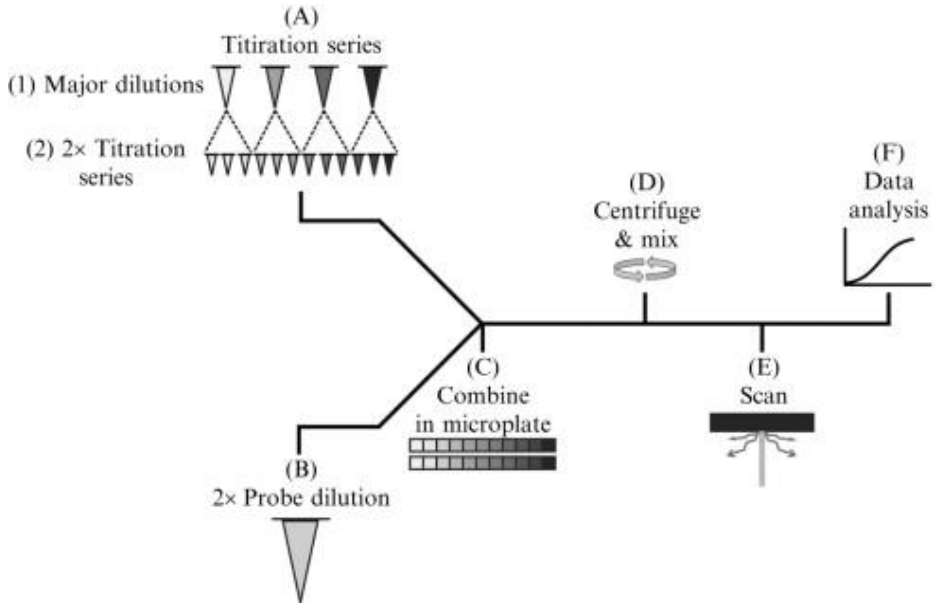


Figure 10. HI-FI Scheme: The schematic of the setup of the HI-FI binding assay. The titration series represents the different concentrations of the LSD1/CoREST complex added to the assay plate and the probe dilution is that of the fluorescent nucleosomes of varying lengths used in the experiments.

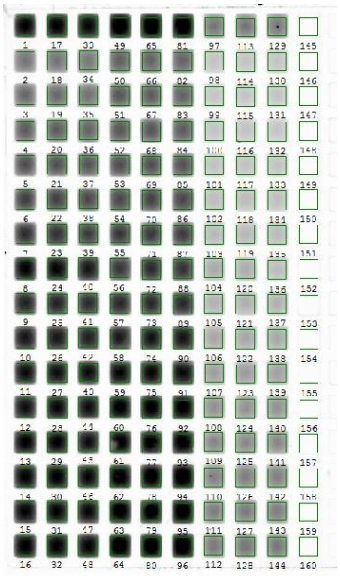


Figure 11. Scanned HI-FI plate: This is an image of the scanned HI-FI plate from the Typhoon scanner. The varying shades of black and grey are representative of the fluorescence change with increasing concentration of the LSD1/CoREST complex binding to the labeled nucleosome.

Chapter 3

Results/Discussion

3.1 PCR Mutagenesis and Subcloning

In order to test whether LSD1/CoREST's ability to bind to the nucleosome is affected by an interaction between the basic H4 tail of the nucleosome and the acidic patch of the SANT domain, five new LSD1/CoREST constructs were created. Each construct incorporated a different set of mutations in the acidic patch, and each individual mutation changed an acidic amino acid in the patch to a neutral alanine. The constructs ranged from incorporating two mutations up to five mutations as seen in Table 1.

Construct Name	Amino Acid Mutations
hLSD1Δ1/CoRESTΔ1x19	D431A/E432A
hLSD1Δ1/CoRESTΔ1x20	E436A/E438A/E440A
hLSD1Δ1/CoRESTΔ1x21	E444A/E445A
hLSD1Δ1/CoRESTΔ1x22	E466A/E467A/E468A
hLSD1Δ1/CoRESTΔ1x23	E466A/E467A/E468A/D469A/E470A

Table 1. Construct Names and Corresponding Amino Acid Mutations: This table displays the name of each of the five mutant constructs created to test binding of the SANT domain of CoREST to the H4 tail of the nucleosome. The corresponding amino acid mutations for each construct are also displayed in the table.

The mutations were made within the pST50Trc2 transfer vector and then subcloned into the pST44 expression vector using the restriction enzymes EcoRI and HindIII. This enzyme selection enabled the mutated CoREST segment to be incorporated into the second cassette of

the expression vector. The detailed schematic and vector can be seen in Section 2.10.2 of this thesis.

If the SANT domain of CoREST interacts with the H4 tail of the nucleosome and is important to the binding of the LSD1/CoREST complex to the nucleosome, these mutations would have an effect on the binding affinity. This is because the H4 tail of the nucleosome is made up of basic amino acid residues which interact with acidic amino acid residues, like the ones in the acidic patch of the SANT domain. If these acidic amino acid residues are neutralized in these new constructs, the interaction between the SANT domain and the H4 tail of the nucleosome may be weakened causing decreased binding of the LSD1/CoREST complex to the nucleosome. A decrease in binding affinity could also potentially lead to a reduction in the complex's demethylase function.

The names for each DNA construct dictate exactly what protein will be expressed when transformed into BL21(DE3)pLysS *E. coli* cells. The "Δ1" following both LSD1and CoREST in the constructs' names indicate that these sequences have been truncated and the full length protein will not be expressed in the complex. These truncated regions were amplified by Song Tan from HeLa cDNA and cloned into the transfer vectors in which mutagenesis could be performed. The region amplified to create LSD1Δ1 was LSD1 amino acid residues 171-852 and the region amplified to produce CoRESTΔ1 was CoREST amino acid residues 286-482. The "x19", "x20", "x21", "x22", and "x23" following only CoREST in the construct names indicates a mutation has been made to the truncated CoREST and the number corresponds to the construct version. Each construct version has specific amino acid residues that were mutated via PCR mutagenesis. As the mutations were only made in the SANT domain of the CoREST, this

indicates that the wild type truncated version of LSD1 will be expressed along with a mutant version of the CoREST.

3.2 Protein Expression

Each protein complex was expressed in BL21(DE3)pLysS cells using the pST44 expression vector and the T7 expression system as previously discussed in Section 2. The LSD1/CoREST complex was found to express best at 23ºCelsius with a five-hour induction period by a Post Doc in the lab, Sang Ah Kim. Uninduced samples and samples after 5 hours of induction by IPTG were sampled during the expression of each LSD1 mutant construct. The CoREST component was known to have a molecular weight of about 22 kD and the LSD1 component was known to have a molecular weight of about 75 kD. The bands corresponding to these proteins can be seen in Figure 12, an SDS PAGE gel for the expression of the wildtype LSD1/CoREST complex. The 22 kD CoREST band migrated between the 21 kD and 31 kD molecular weight marker bands and the 75 kD LSD1 band migrated between the 66 kD and 97 kD marker bands. After each protein complex was expressed in *E. coli* cells, it was given a version number, and with each new variation of the complex, a new version was added. hLSD1 complex v01 (version 1) is the complex that includes the wildtype CoREST and LSD1 truncations. The SDS PAGE gel in Figure 12, ran by Song Tan, shows the expressed and purified hLSD1 complex v01 which was later used to compare the migration pattern of the LSD1 and CoREST bands for future mutant complexes to help identify these bands. The version numbers of each of the five complexes created in this project and the corresponding DNA construct names and amino acid changes are included in Table 2.

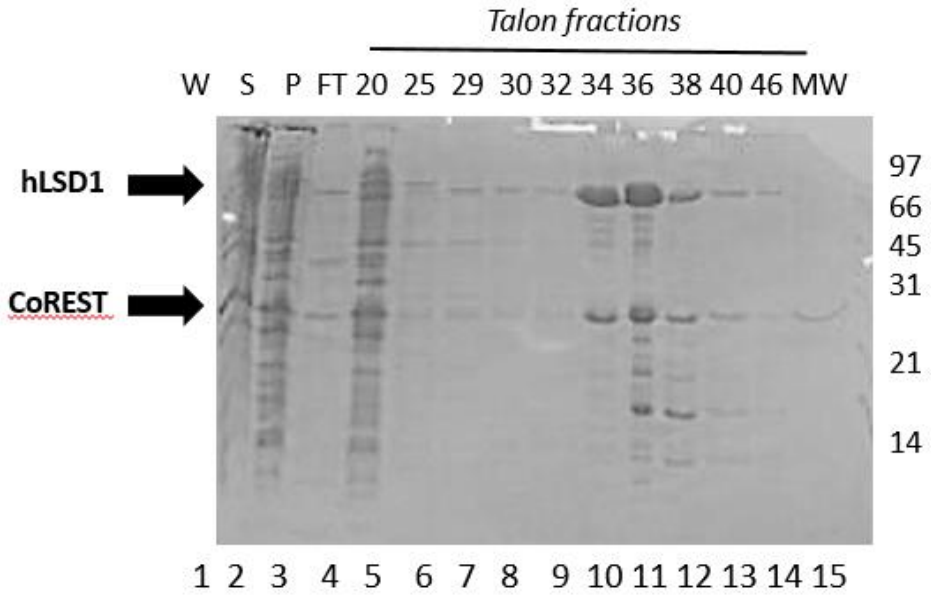


Figure 12. hLSD1 complex v01 Talon Purification: This SDS PAGE gel shows the LSD1/CoREST complex that was first expressed and purified in the lab via metal affinity chromatography in 2006. This purification was completed by Song Tan and was used to compare the location of the complex on future SDS PAGE gels. (Wells of the gel are labeled: MW- marker, W- whole cell extract, P-pellet, S- supernatant, FT- flow through)

hLSD1 Complex Version	DNA Construct Name	Altered Amino Acid Residues
hLSD1 Complex v35	pST44-STRaHISNhLSD1Δ1-hCoRESTΔ1x20	E436A/E438A/E440A
hLSD1 Complex v36	pST44-STRaHISNhLSD1Δ1-hCoRESTΔ1x19	D431A/E432A
hLSD1 Complex v37	pST44-STRaHISNhLSD1Δ1-hCoRESTΔ1x22	E466A/E467A/E468A
hLSD1 Complex v54	pST44-STRaHISNhLSD1Δ1-hCoRESTΔ1x23	E466A/E467A/E468A/D469A/E470A
hLSD1 Complex v55	pST44-STRaHISNhLSD1Δ1-hCoRESTΔ1x21	E444A/E445A

Table 2: hLSD1 Complex Version Numbers and their Corresponding DNA Constructs and Amino Acid Residue Mutations. After expression of each construct within the cell, the name of the protein complex was given in the order the preparation was completed. This chart indicates the corresponding DNA constructs for each protein complex expressed and the amino acids that were changed in these constructs to express neutral amino acids in the final protein complex.

Before each protein complex was expressed in *E. coli* cells, a sample of uninduced cell culture was taken to compare to the samples after five hours of induction. An SDS PAGE gel was used to determine what proteins were present in each sample and in what relative amounts. In Figure 13, the expression of hLSD1 complex v35, containing two neutral amino acids in the acidic patch of the SANT domain, can be seen after induction in comparison to the uninduced sample. All of the protein bands are very faint in the uninduced sample. On the contrary, within the five-hour induction period, each of the protein bands increase in darkness. Two specific

bands' intensity increased significantly relative to that of the other bands within each lane. These two bands represent the LSD1 and CoREST components that were induced by the addition of IPTG. They can be seen in Figure 13 at their approximate molecular weights of 22 kD for CoREST and 75 kD for LSD1.

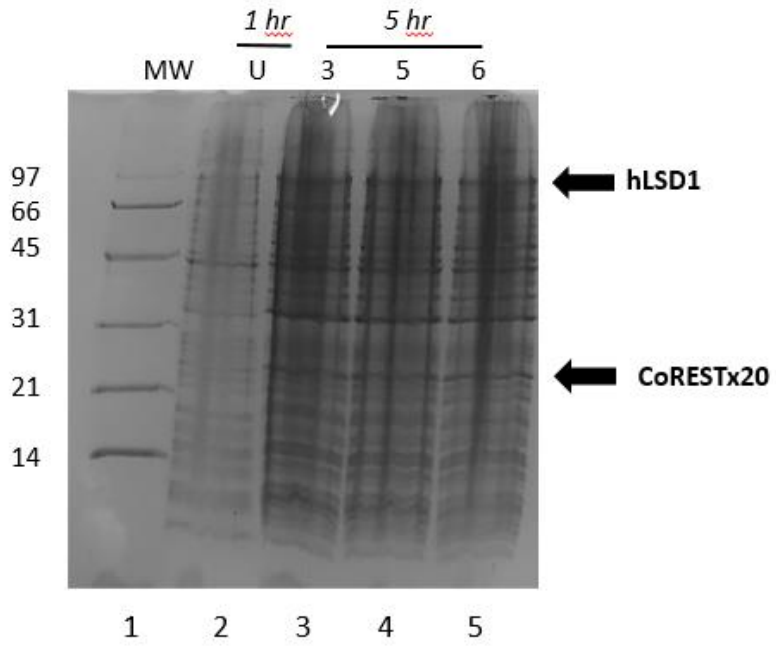


Figure 13. Large Scale Expression of hLSD1 complex v35 (23ºC): Expression of hLSD1 complex v35 at 23ºCelsius. An uninduced sample is included along with samples from three flasks after the 5hr induction of protein expression. In the induced samples there appears to be a band approximately where LSD1 would migrate and where CoREST would migrate. (The well labeled U stands for the uninduced sample.)

3.3 Protein Purification

LSD1 complexes hLSD1 v35, hLSD1 v36, and the first prep of hLSD1 v37 were purified in batch and complexes hLSD1 v54, hLSD1 v55, and the second prep of hLSD1 v37 were purified using metal affinity chromatography over the Talon column.

3.3.1 Batch Purification of LSD1 Complexes

Batch purification uses the same metal affinity resin as the Talon column but the purification is performed in a Falcon tube and only six fractions of 6 mL of eluted fractions are collected in total. Because of this, the elution pattern for the protein complexes are very different from that of the protein complexes after purification by the Talon column. The complex must fully elute within six washes of elution buffer and was eluted mainly within the first four fractions collected from the batch purification. Figure 14 shows an example of the SDS PAGE gel run after the batch purification of hLSD1 complex v37. This version of the complex contained three neutral amino acids in the SANT domain acidic patch.

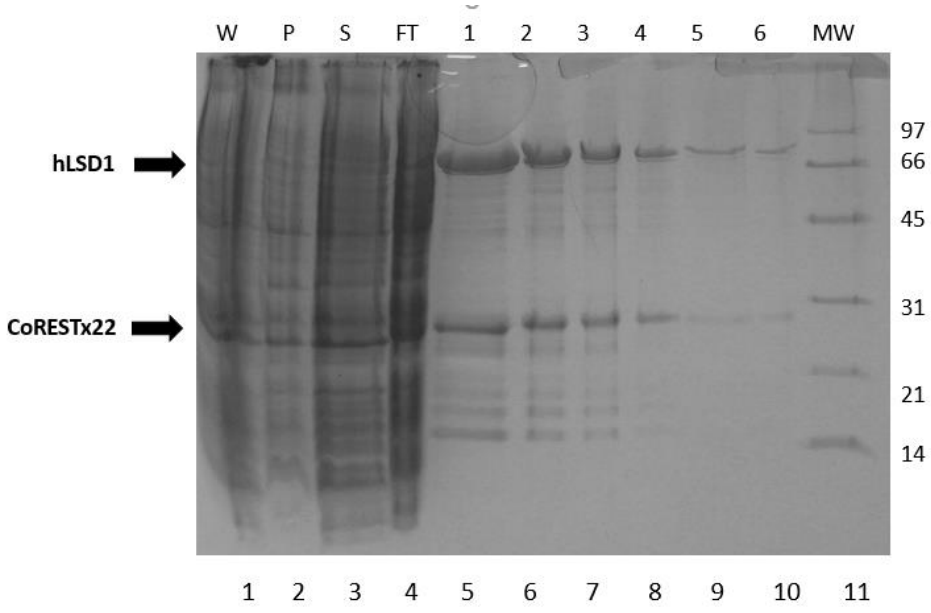


Figure 14. Batch Purification of hLSD1 complex v37: SDS PAGE gel of the metal affinity batch purification of hSLD1 complex v37 with three neutral amino acids in the acidic patch of the SANT domain. A large percentage of the LSD1/CoREST complex eluted in the first four fractions of the batch preparation. (Wells of the gel are labeled: MW- marker, W- whole cell extract, P-pellet, S- supernatant, FT- flow through, *Purification and gel performed by Mike Doyle).*

3.3.2 Talon Column Purification

This purification method uses the same cobalt metal affinity resin as the batch purification but was performed using a Talon column in the cold room. After equilibration and wash methods were performed, the high salt elution buffer was used to elute 44 fractions of 6 mL each. Unlike the batch purification, the complex was not eluted immediately but has rather been shown to elute between the 20^{th} to 40^{th} fractions by many members of the Tan lab, the exact range depending on the specific complex version that was being purified. An example of an SDS PAGE Gel run after a Talon column purification of hLSD1 complex v55 is seen in Figure 15. For this version, the complex was eluted in fractions 19 to 40. Also incorporated on the gel is the whole cell extract, the pellet, supernatant, and flow through to help identify where some of the total protein may have been lost during the purification process. This could indicate where improvements could be made with recovery in future purification of the same LSD1/CoREST complex as well as if the mutations in the SANT domain caused the complex to become insoluble.

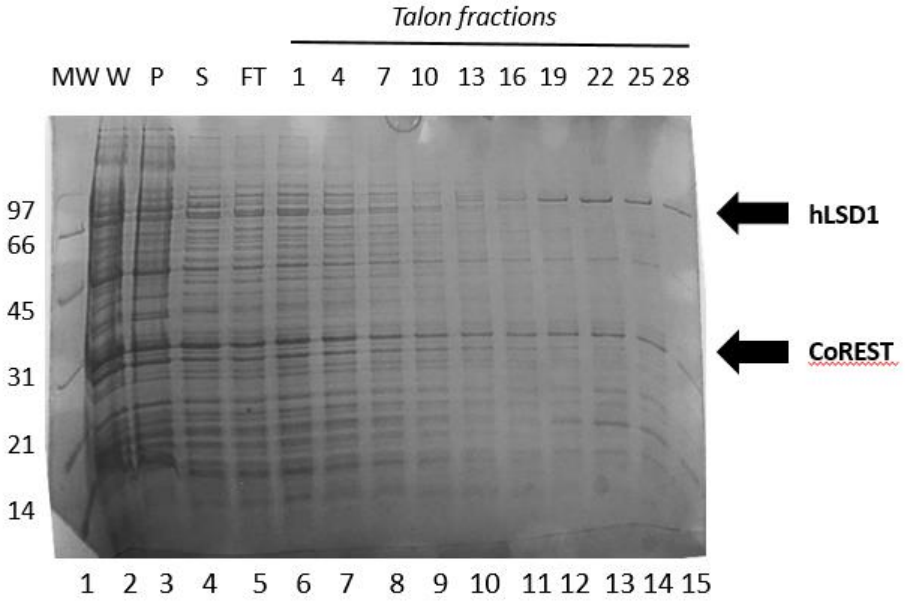


Figure 15. hLSD1 Complex v55 Talon Purification: The above SDS PAGE gel shows a Talon metal affinity chromatography purification of hSLD1 complex v55. This purification was performed over a column and not in batch. Fractions beginning at 19 up to 43 (on a second SDS PAGE gel) were pooled for dialysis and further purification by high performance liquid chromatography (HPLC). (Wells of the gel are labeled: MW- marker, W- whole cell extract, P- pellet, S- supernatant, FT- flow through)

3.3.3. TEV Protease Cleavage

In both purification methods, a hexahistidine tag was used to purify the LSD1/CoREST complex from the rest of the cell extract. This tag, however, could have the potential to interfere with binding and functional assays performed after purification. In order to prevent this disruption, the hexahistidine tag was removed with a TEV protease enzyme. To ensure that the tag was fully removed, an SDS PAGE gel was run for each complex on samples before and after the protease was added to the purified protein. A molecular weight change of about l kD was expected to be seen on the gel indicating the removal of the tag. An example of this can be seen in Figure 16. The SDS PAGE gel includes the undigested and digested samples for both the

hLSD1 v36 and hLSD1 v37 complexes. A very minimal downward shift of the protein bands can be seen with the addition of TEV protease.

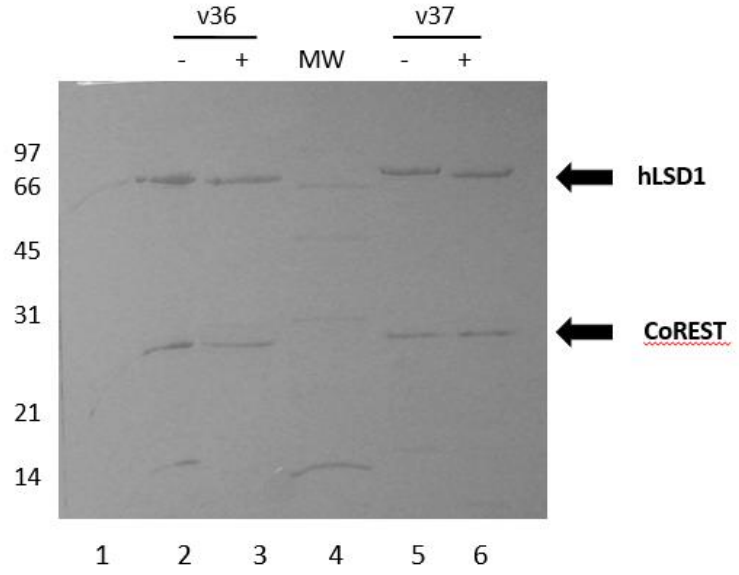


Figure 16. TEV protease digest of hLSD1 complex v36 and v37: The column labeled (-) for each of the mutant constructs represents the complex without the addition of TEV protease, and therefore the undigested protein. The column labeled (+) for each construct shows the complex after digestion by TEV protease and the cleavage of the hexahistidine affinity tag. The well with MW is the marker *(Gel run by Mike Doyle).*

Each of the five complexes were further purified using high performance liquid chromatography (HPLC) with a Source Q10 Anion Exchange column. A chromatogram was generated by the instrument and used to choose which fractions to pool for dialysis and the final concentration of the collected protein. This was decided based on the peaks in the chromatogram and their corresponding fractions, indicating where the protein was eluted from the column. An example of one of these chromatograms can be seen in Figure 17. The blue line in the chromatogram represents the absorbance of UV light at 280 nm, the same wavelength at which proteins absorb UV light. Any peak at this wavelength represents a protein being eluted from the column.

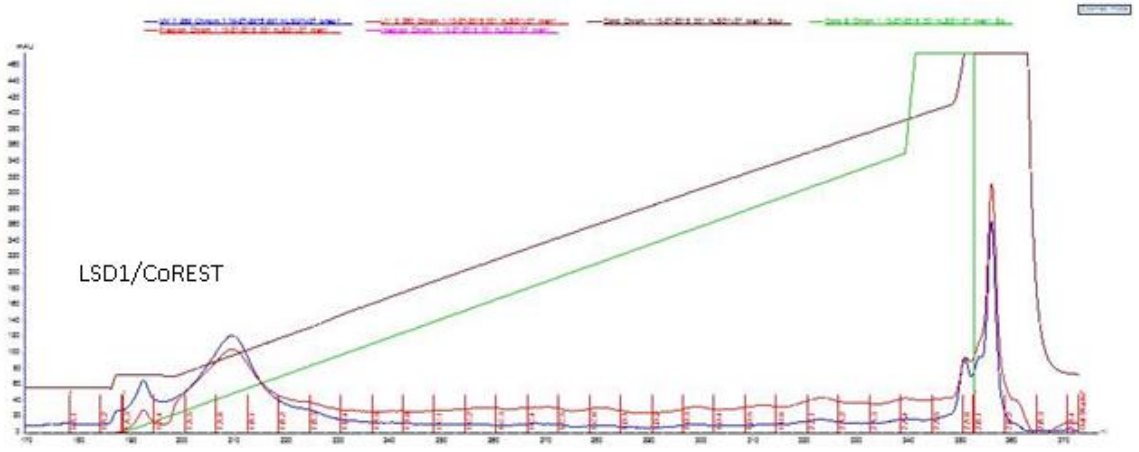


Figure 17. HPLC Generated Chromatogram. This is an example of a chromatogram generated by the HPLC. The absorbance was measured at 280 nm and the fractions were sampled from both peaks within the gradient. The peak labeled LSD1/CoREST was determined to be the sole peak that contained the complex.

In the first run of each HPLC purification, samples from each of the peaks within the chromatogram were ran on a gel to identify in which peak the desired protein complex appeared. LSD1/CoREST complexes as a whole were found to elute fairly early in the gradient, most often showing up in fractions ranging from 1A3 to 1C1. An SDS PAGE gel including the sampled fractions from the purification of hLSD1 complex v37 is shown in Figure 18.

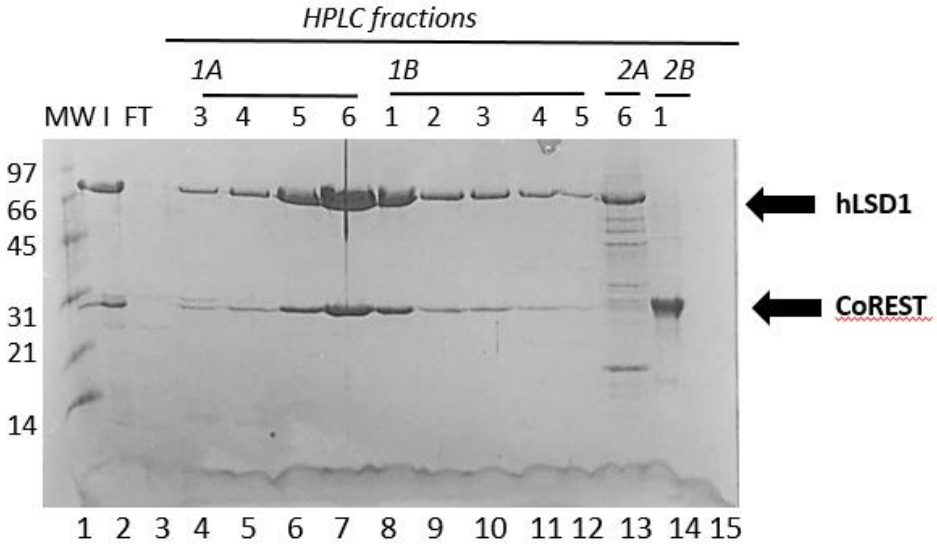


Figure 18. Source Q10 Anion Exchange Chromatography SDS-PAGE Gel: The hLSD1 v37 LSD1/ CoREST complex was further purified over the SourceQ anion exchange chromatography column. Samples from the fractions contained within the two peaks on the chromatogram were analyzed by SDS PAGE and fractions 1A5 to 1B1 were pooled for dialysis and concentration. (Wells of the gel are labeled: MW- marker, I-input, and FT- flow through)

The hLSD1 complex v54 contained five point mutations within the acidic patch of the SANT domain. The constructs that were successfully purified via the Talon column or via batch purification incorporated three mutations in this domain at most. After Talon column purification of hLSD1 complex v54, the bands representative of hLSD1 and CoREST in the whole cell extract, pellet, and supernatant on the SDS PAGE gel in Figure 19 were compared to each other to analyze their relative intensities. The pellet's LSD1 and CoREST bands were representative of over fifty percent of the intensity of the whole cell extract's respective bands. The supernatant on the other hand, had LSD1 and CoREST bands with very weak intensities, representing about 20 percent of the whole extract's relative intensity. This indicated that a majority of the protein was remaining in the pellet during the whole cell extract preparation instead of remaining in the supernatant for further purification. It is possible that the increased amount of acidic amino acid

residue mutations to neutral residues within the acidic patch of the SANT domain caused the complex to become partially insoluble. This will have to be further investigated in later experiments.

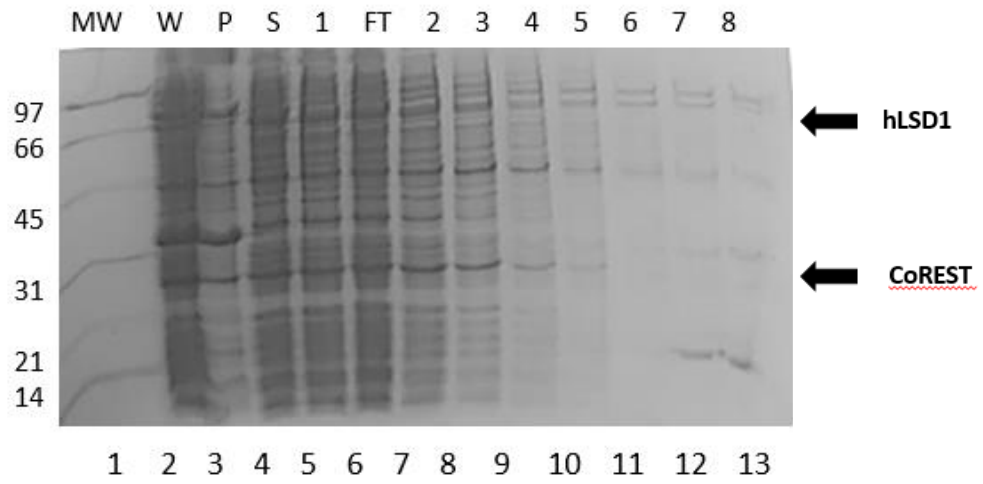


Figure 19. Talon Purification of hLSD1 complex v54: The extent of the mutations in the acidic patch may have caused this protein to become insoluble. This can be seen as a large portion of material from the whole cell extract shows up in the pellet fraction on the gel. (Wells of the gel are labeled: MW- marker, W- whole cell extract, P-pellet, S- supernatant, FT- flow through)

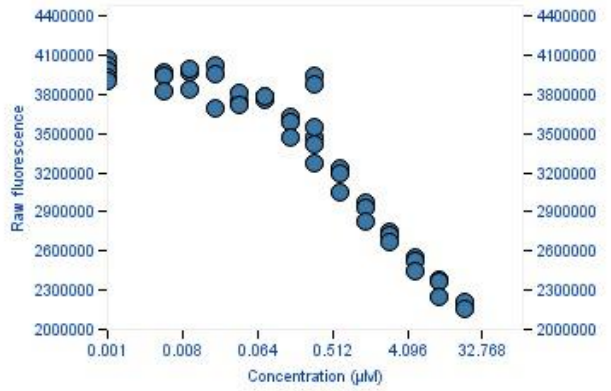
After each LSD1/CoREST complex was purified by HPLC, the protein was concentrated to a range between 20 and 40 mg/mL. The hLSD1 complex v35 had a total yield of 1.8 mg of protein and the hLSD1 complex v36 had a yield of 12.2 mg of protein. The hLSD1 complex v37 was purified twice. The first prep produced 6.8 mg of protein and the second prep using autoinduction media produced 79.2 mg of protein. Finally, hLSD1 complex v55 yielded 6.1 total mg of the protein complex.

3.4 HI-FI Binding Assay

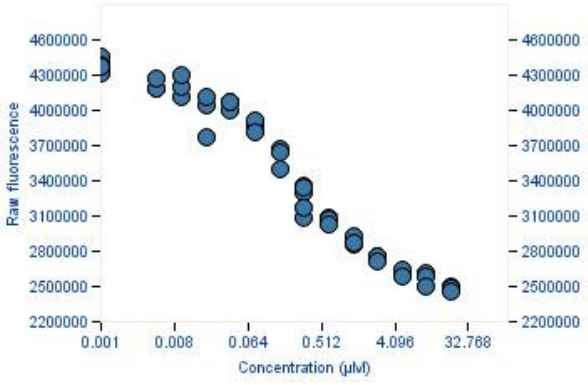
High throughput interactions by fluorescence intensity (HI-FI) assays were performed to analyze the binding capacity of the LSD1/CoREST complex to the nucleosome. hLSD1 complex v01 was tested first to have a baseline binding measure to compare to the mutant constructs' values. While the version one complex expresses truncations of the protein complex, these do not affect the binding of the complex to the nucleosome. After original analysis of the complex structure, truncations were made to shorten the complex for better expression results, but all essential binding and functional domains were kept intact for binding and demethylase assays. To determine if the acidic patch of the SANT domain is essential for the binding of the LSD1/CoREST complex to the nucleosome, the mutant construct's binding ability was compared to that of the v01 complex using the HI-FI binding assay. The neutral amino acid alterations in the mutant constructs have the ability to weaken the negative charge within the acidic patch of the SANT domain and inhibit its interactions with any positively charged domain on the nucleosome such as the tail of histone H4 of the nucleosome. This could affect the binding of the complex to the nucleosome and therefore would yield a weaker fluorescent signal in the raw data generated by the HI-FI assay. Fluorescently labeled nucleosomes were titrated with the version one complex and each of the mutant LSD1 complexes. The titrations that resulted in the largest amount of LSD1/CoREST bound to the nucleosome would have the greatest ability to quench the fluorescent signal on the nucleosome, resulting in a decreased raw fluorescence value. Using these data points, the single binding isotherm equation (Winkler et al) was used to generate values for the binding affinities (K_d). Each of the mutant complexes' K_d values were compared to that of version one. The lower the K_d value, the greater the affinity of the complex for the nucleosome.

In Figure 20, the raw fluorescence data relative to the concentration of LSD1 complex before using the single binding isotherm equation (Winkler et al) equation is displayed. This particular data is for hLSD1 complex v36 with three different nucleosomes. Each nucleosome consists of a different length of extranucleosomal DNA on either side of the nucleosome ranging from 147 base pairs, indicating one additional nucleotide on either side of the nucleosome, to 181 bp indicating 18 nucleotides on either side of the core nucleosome.

A)



B)



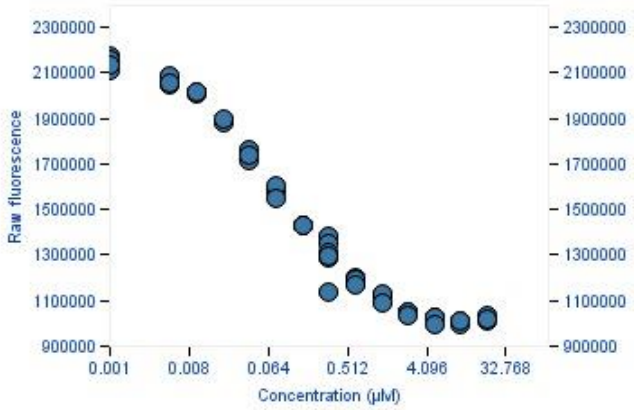


Figure 20. Raw fluorescence HI-FI data graphs: Raw fluorescence plots were created from the data generated by the Typhoon scanner and plotted against the concentrations of LSD1/CoREST complex titrated into the wells. This data was then used to calculate the K_d and the Hill coefficient for binding of the complex to the nucleosomes. The above graphs show the data for hLSD1 complex v36 on three different nucleosomes: 147 bp (A), 157 bp (B) and 181 bp (C).

After gathering the raw fluorescence data relative to the LSD1 complex concentrations, the Hill Coefficients and binding affinities were calculated using the single binding isotherm equation. Binding curves were generated for the interactions between the complex and each of the nucleosomes with varying extranucleosomal DNA lengths individually. A collective binding curve overlay was also generated for each mutant complex to compare the binding affinities between each of the nucleosomes and the LSD1/CoREST complex. This comparison was first analyzed for hLSD1 complex v01 data using the 157 bp and 181 bp nucleosomes. Figure 21 shows the compilation of data for the wild type complex including each of the individual binding curves and the overlay of the two. From this data it can be seen that the 157 bp nucleosome had a higher K_d value and therefore was more likely to dissociate from the LSD1/CoREST wildtype complex. The 181 bp nucleosome on the other hand had a lower K_d and therefore a higher binding affinity for the wildtype complex. Further analysis of the wildtype complex with more nucleosomes of various extranucleosomal DNA lengths have been completed by Sang Ah Kim in our lab, and she has been able to show a correlation between extranucleosomal DNA length and binding affinity. Sang Ah Kim found the wildtype LSD1/CoREST complex to interact with the 157 base pair nucleosome with a Kd of 0.862 μM and with the 181 base pair nucleosome with a Kd of 0.392 μM. This is a lower binding affinity, but still comparable to the 0.64 μM and 0.3 μM values calculated in this thesis for those respective nucleosomes.

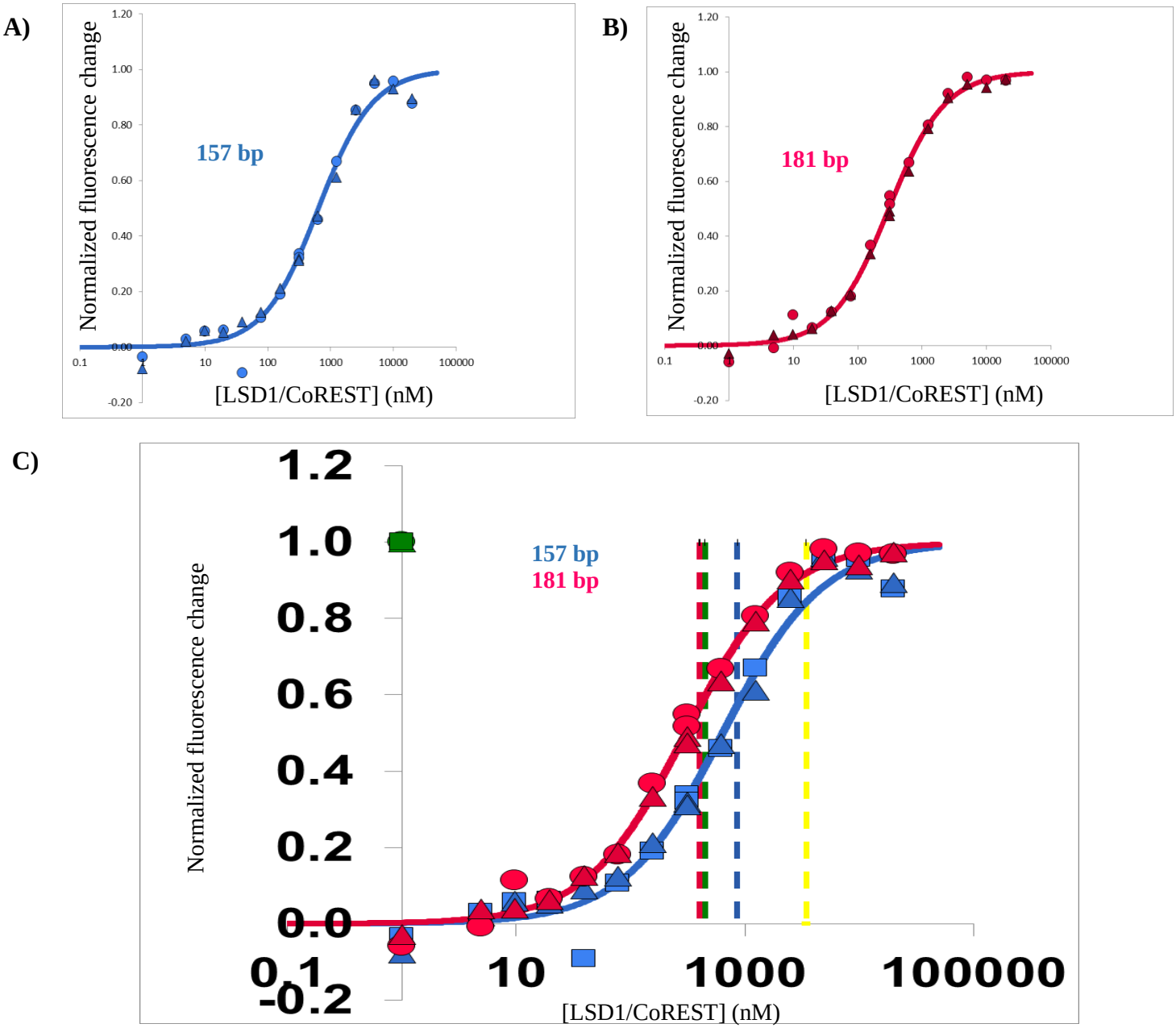
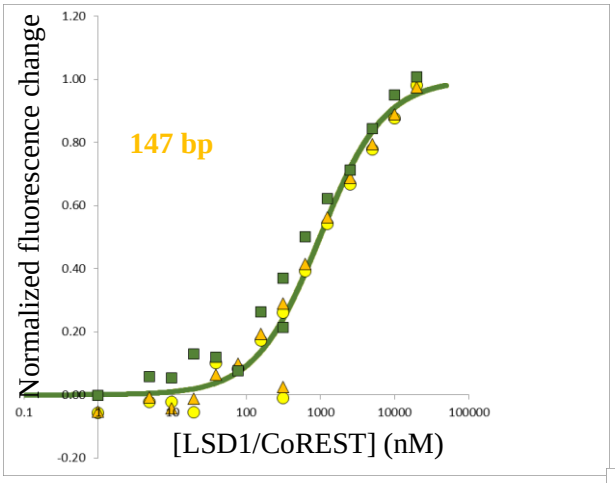


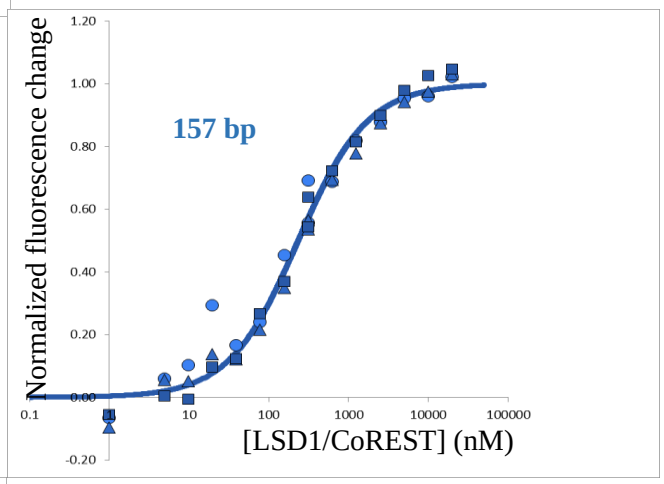
Figure 21. HI-FI Data for hLSD1 complex v01: (A) Binding curve for wildtype LSD1/CoREST complex interaction with the 157 bp nucleosome. (B) Binding curve for wildtype LSD1/CoREST complex interaction with 181 bp nucleosome. (C) Overlay of the two binding curves of wildtype LSD1 for comparison of different length extranucleosomal DNA on LSD1/CoREST binding to the nucleosome. The x axis represents the nM concentrations of LSD1/CoREST complex and the y axis measures change in fluorescence.

After HI-FI data was generated and analyzed for the wildtype complex, the mutant complexes were tested to compare their interactions with each of the nucleosomes. These results were then compared to those of wildtype complex and of the other mutant complexes to determine if any change in binding affinity occurred due to the mutations. In Figure 22, hLSD1 v36 was tested with three different nucleosomes: 147bp, 157bp, and 181 bp. This version of the complex had three of the acidic amino acid residues in the SANT domain mutated to neutral alanines. Once again, each nucleosome interaction was analyzed separately and then compared as a whole. The overlayed binding curve for hLSD1 complex v36 can be seen in Figure 23. The 147 base pair nucleosome had the largest K_d and the 181 base pair nucleosome the lowest K_d of the three. This indicates that the 181 base pair nucleosome also had the highest binding affinity to the hLSD1 complex v36 as it did with v01. In comparing whether the mutations had a significant effect on the binding affinity of hLSD1 v36 in comparison to v01, the specific K_d values must be compared between the two for each nucleosome tested. The hLSD1 v01 complex bound to the 157 base pair nucleosome with a K_d of 0.64 µM and to the 181 bp nucleosome with a K_d of 0.30 µM. The hLSD1 complex v36 bound to the same nucleosomes with K_ds of 0.23 and 0.08 µM respectively. When comparing the binding affinities, hLSD1 complex v36 has about a 2.8 times greater affinity for 157 bp nucleosome than v01 and an approximately 3.75 times greater affinity for the 181 bp nucleosome. This is because smaller K_d values indicate a larger affinity for the substrate.

A)



B)



C)

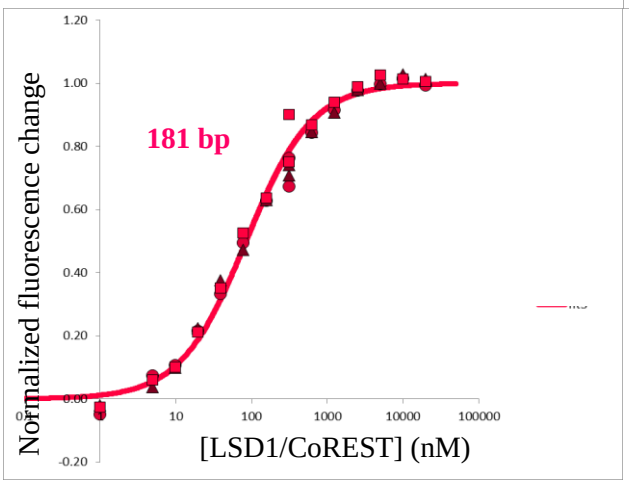


Figure 22. HI-FI data for hLSD1 v36 in triplicate: (A) Binding curve for v36 interaction with 147 bp nucleosome. (B) Binding curve for v36 interaction with 157 bp nucleosome. (C) Binding curve for v36 interaction with 181 bp nucleosome.

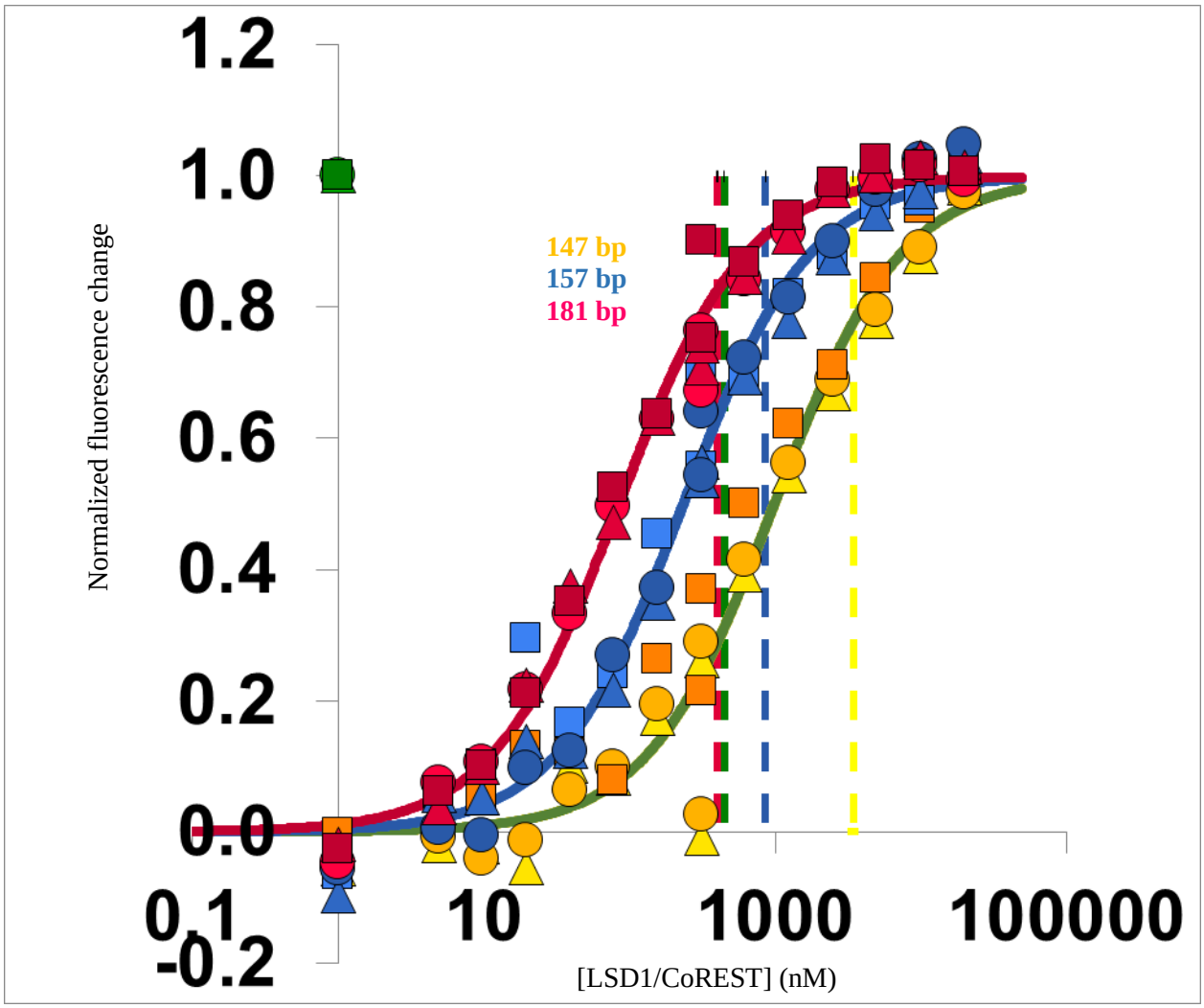


Figure 23. HI-FI data for hLSD1 v36 on three nucleosomes: An overlay of HI-FI data for hLSD1 v36. This is to compare the three nucleosomes with different length extranucleosomal DNA and the mutant constructs' ability to bind to the nucleosomes. This overlay can be compared to that of the wildtype to see the effect of the mutations on the interaction with the nucleosomes.

At this point, the binding of hLSD complex v35 to nucleosomes has only been examined in one experiment, rather than the desired triplicate experiments. hLSD1 v35 was also only tested for nucleosomes with extranucleosomal DNA lenghts of 147 base pairs and 157 base pairs. In Figure 24, the individual and overlayed binding curves can be compared. The same trend has been upheld with this mutant construct in regards to extranuclesomal DNA length. Once again, the 147 bp nucleosome has the lowest binding affinity with the largest K_d value and the 157 base pair has a higher affinity. With the experiments completed thus far on both the mutant and wildtype LSD1/CoREST complexes, it has been observed that the nucleosomes with longer extranucleosomal DNA have higher binding affinities to the LSD1/CoREST complexes. The hLSD1 complex v35 only has two neutral alanines within the acidic patch in comparison with the three altered in hLSD1 complex v36. Even so, hLSD1 complex v35 has an approximately 3.2 times greater affinity for the 157 bp nucleosome substrate than hLSD1 complex v01. This is a similar relative affinity to that of hLSD1 complex v36 which had a 3.75 times greater affinity for the 157 bp nucleosome than hlSD1 complex v01. Both of these results indicate that the mutant complexes with increased neutral residues in the acidic patch of the SANT domain increase the binding affinity of LSD1/CoREST to the nucleosome. In comparison to Sang Ah Kim's wildtype LSD1 data, hLSD1 complex v35 also has a greater binding affinity to the 147 bp nucleosome. Her results indicated an average of about 2.5 μM Kd for the wildtype complex with the 147 bp nucleosome whereas the v35 complex had a 4.7 times greater binding affinity with a Kd of 0.53 μM.

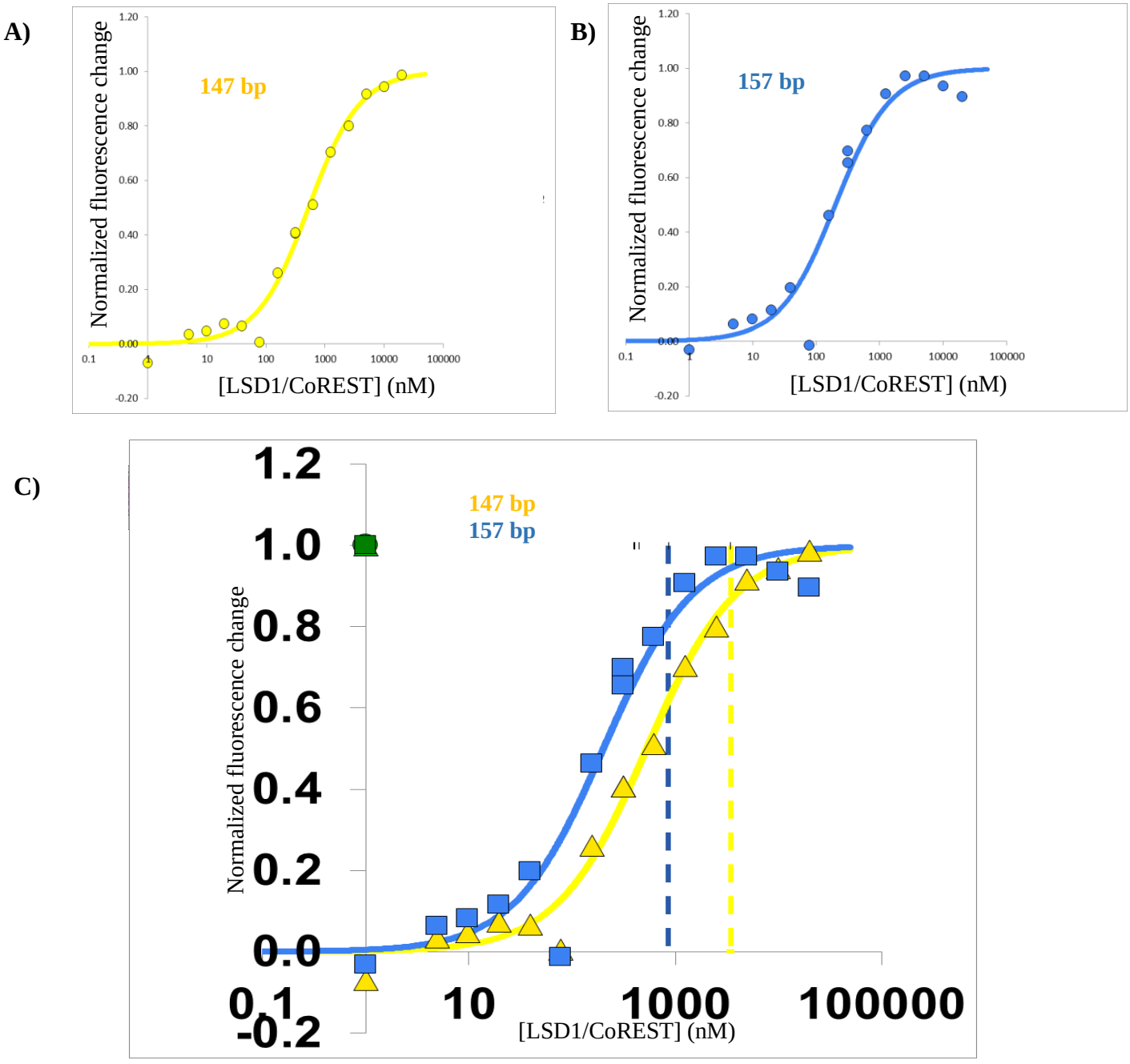
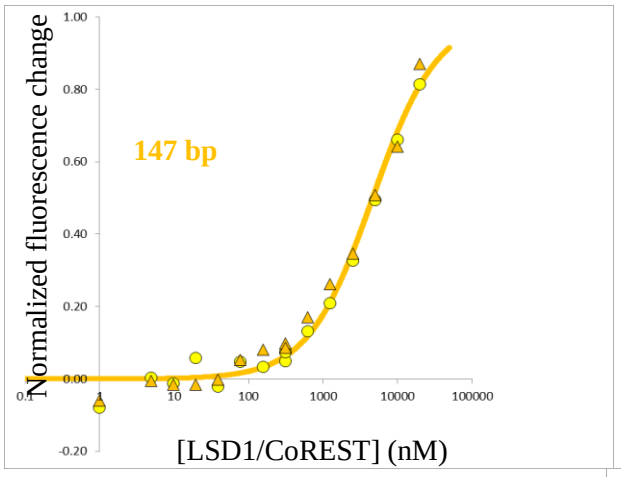


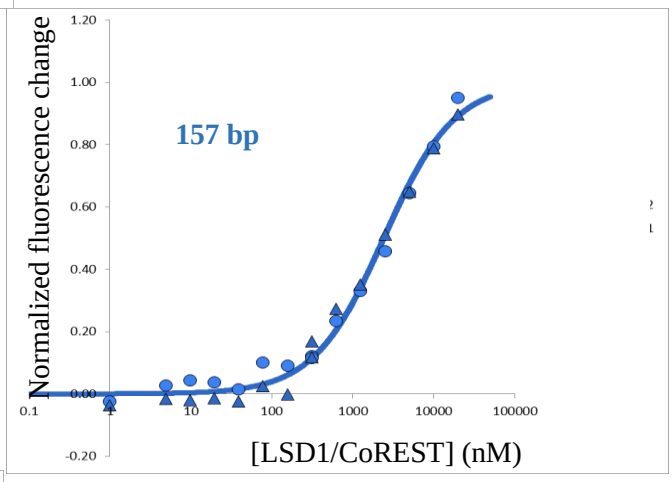
Figure 24. HI-FI data for hLSD1 v35 in singlet: (A) Binding curve for hLSD1 v35 to 147 bp nucleosome. (B) Binding curve for hLSD1 v35 to 157 bp nucleosome. (C) Overlay of hLSD1 v35 interaction with 147 bp and 157 bp nucleosome to compare the change in DNA length with the ability of hLSD1 v35 complex to bind.

hLSD1 complex v37 had three mutations within the acidic patch of the SANT domain. These three mutations were also incorporated in the five mutation combination within hLSD1 complex v54. Because hLSD1 complex v54 appeared to be insoluble with the extent of mutations that were made in the SANT domain, it was interesting to see the effects of a subset of these mutations with the binding assays involving hLSD1 complex v37. Unlike the previous two mutant complexes discussed, hLSD1 v37 followed the hypothesized trend that an increase in mutations within the SANT domain would decrease binding to the nucleosomes of varying DNA lengths. The wildtype LSD1/CoREST complex bound both the 147 bp and 181 bp nucleosome with a 2 times greater affinity than that of the hLSD1 complex v37. It also was calculated that the wildtype complex bound the 157 bp complex with a 3.7 times greater affinity than the hLSD1 complex v37. In comparison to the mutant complexes that contained only two mutations within the SANT domain and appeared to increase binding, this data could indicate that more mutations are required in bulk to decrease the binding of the LSD1/CoREST complex to the nucleosome. At the same time, with the hLSD1 complex v35 also having three mutations in the SANT domain but increasing the binding affinity of the complex, the specific location of these mutations within the domain could be of upmost importance as well.

A)



B)



C)

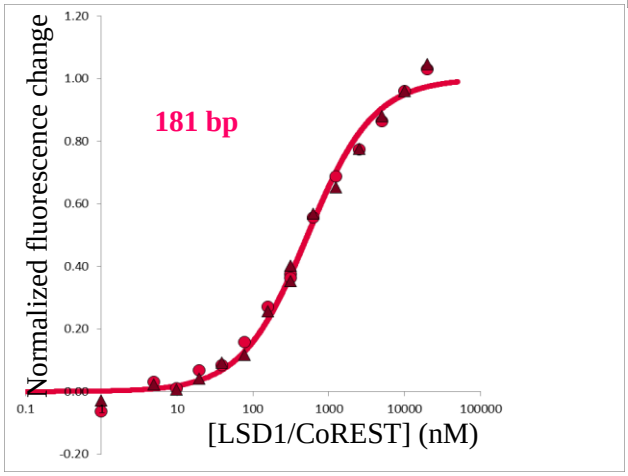


Figure 25. HI-FI data for hLSD1 v37 in duplicate: (A) Binding curve for v37 interaction with 147 bp nucleosome. (B) Binding curve for v37 interaction with 157 bp nucleosome. (C) Binding curve for v37 interaction with 181 bp nucleosome.

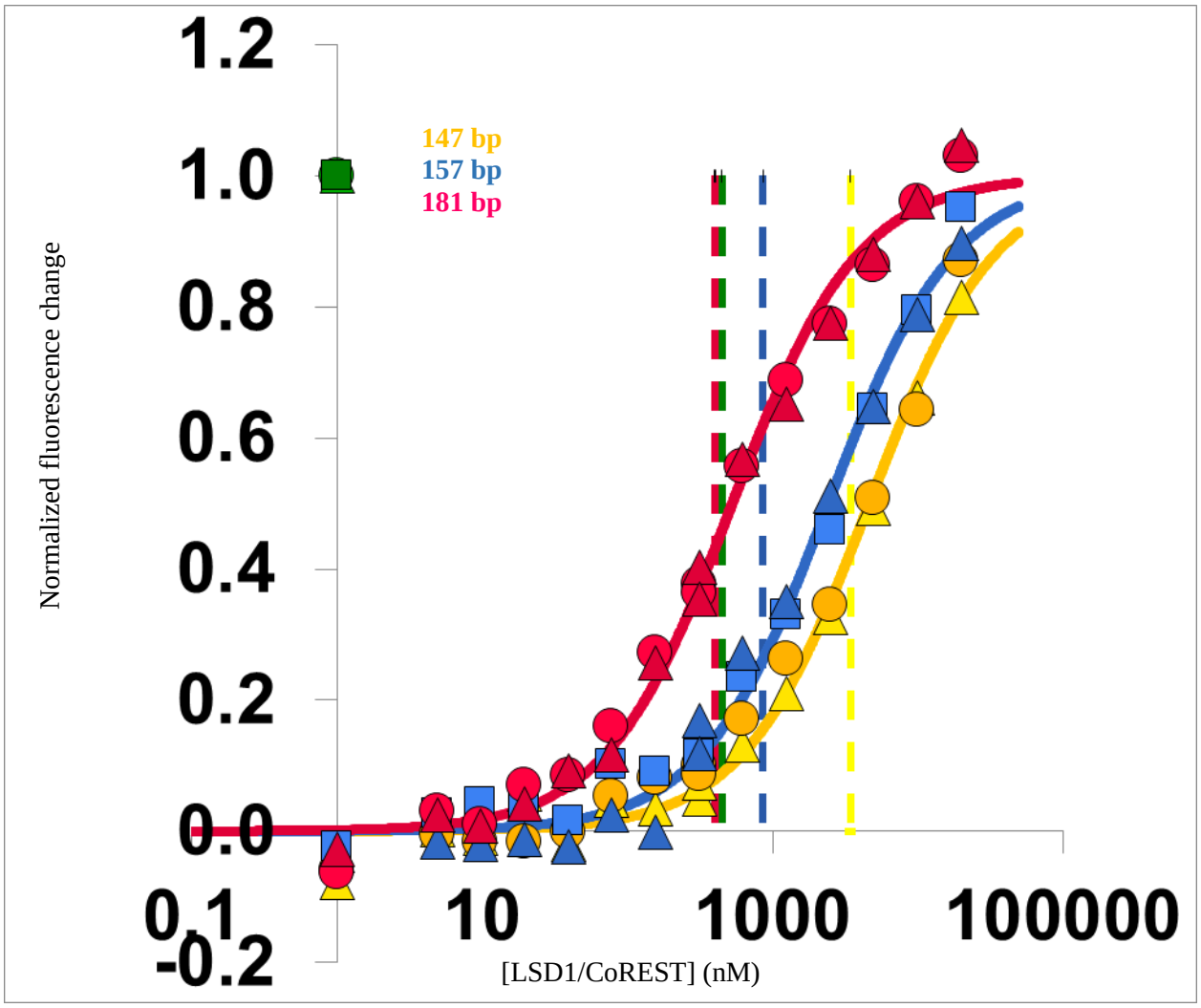


Figure 26. HI-FI data for hLSD1 v37 on three nucleosomes: An overlay of HI-FI data for hLSD1 v37. This is to compare the three nucleosomes with different length extranucleosomal DNA and the mutant constructs' ability to bind to the nucleosomes. This overlay can be compared to that of the wildtype to see the effect of the mutations on the interaction with the nucleosomes.

Finally, the hLSD1 v55 complex was the last explored with the HI-FI binding assays. This complex was only tested on the 181 bp nucleosome in duplicate. Due to an older HI-FI binding plate being used for this experiment, the fluorescent signal on the scan was more washed out than in previous experiments, and this could have potentially affected the results. This complex had two mutations in the SANT domain that converted acidic amino acids to neutral amino acids. Based on the HI-FI binding assay it was calculated that these mutations decreased the binding affinity of the LSD1/CoREST complex with the nucleosome in comparison to the wildtype complex. When comparing the Kd values directly, the wildtype complex was shown to bind the 181 bp nucleosome with a 3.3 times greater affinity than that of the hLSD1 complex v55.

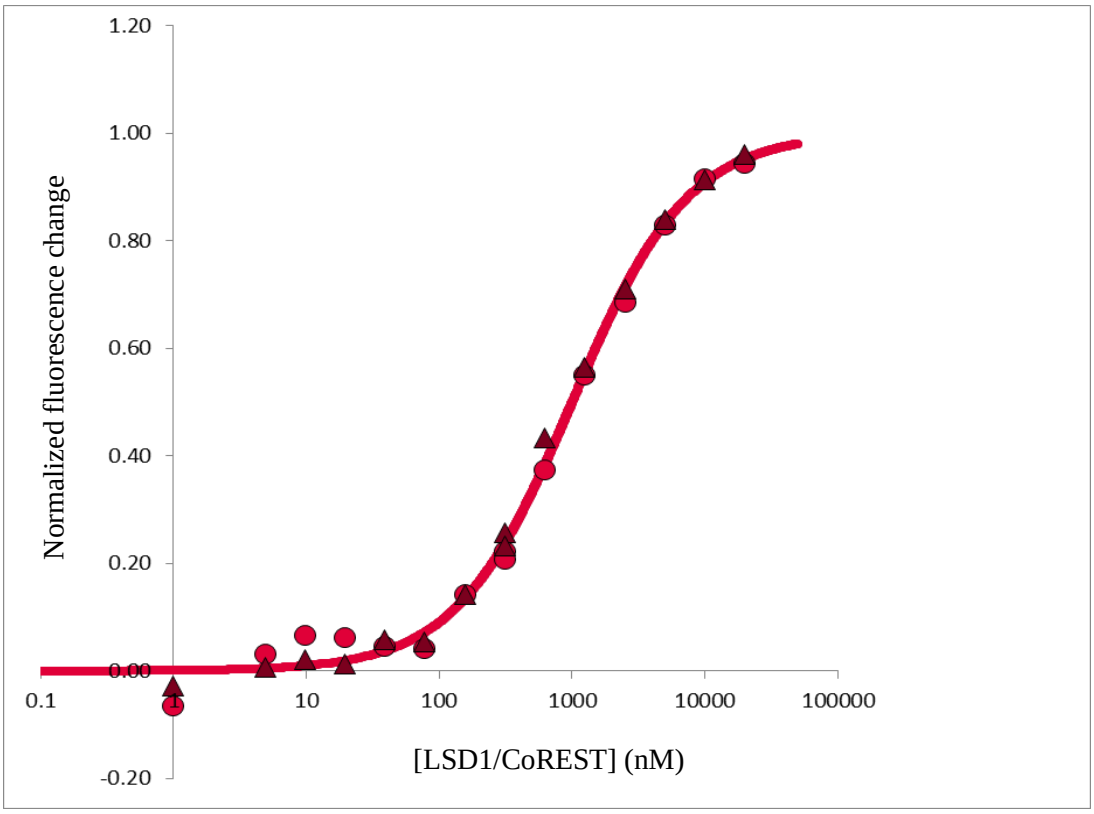


Figure 27. HI-FI data for hLSD1 v55 in duplicate: Binding curve for hLSD1 v55 to 181 bp nucleosome.

LSD1/CoREST Mutant Complex	CoREST Mutations	DNA length (bp)	DNA layout	Kd (μM)	Relative to WT LSD1/CoREST
hLSD1 complex v01	WT	147	1 + 145 + 1	~2.5*	1.0
hLSD1 complex v01	WT	157	6 + 145 + 6	0.64	1.0
hLSD1 complex v01	WT	181	18 + 145 + 18	0.30	1.0
hLSD1 complex v35	E436A/E438A/E440A	147	1 + 145 + 1	0.53	4.7 x
hLSD1 complex v35	E436A/E438A/E440A	157	6 + 145 + 6	0.20	3.2 x
hLSD1 complex v36	D431A/E432A	147	1 + 145 + 1	0.99	2.5 x
hLSD1 complex v36	D431A/E432A	157	6 + 145 + 6	0.23	2.8 x
hLSD1 complex v36	D431A/E432A	181	18 + 145 + 18	0.08	3.75 x
hLSD1 complex v37	E466A/E467A/E468A	147	1 + 145 + 1	4.63	0.54 x
hLSD1 complex v37	E466A/E467A/E468A	157	6 + 145 + 6	2.41	0.27 x
hLSD1 complex v37	E466A/E467A/E468A	181	18 + 145 + 18	0.53	0.57 x
hLSD1 complex v55	E444A/E445A	181	18 + 145 + 18	1	0.3 x

Table 3. Summary of all the K_d Values for Each Mutant Complex and their Relative Affinity to hLSD1 Complex v01: The above table describes each mutant complex, the mutations that were made within it, the nucleosomes that it was titrated on in the HI-FI assay and the K_d values for each binding assay. The relative K_d values were also calculated to compare each of the mutant complexes to the wildtype hLSD1 complex v01.
*This value was from the experiments completed by Sang Ah Kim

Chapter 4 Conclusion

4.1 Summary

The purpose of this thesis was to study the role of the SANT domain in the binding of LSD1/CoREST to the nucleosome. The prominent acidic patch of the SANT domain was mutated in various DNA constructs in order to analyze whether this region had any potential interactions with the basic tail of histone H4 in the nucleosome. Five different constructs were made incorporating neutral amino acid residues in place of the native acidic amino acids, and binding assays were performed to assess if any change occurred as a result. From the experiments completed thus far, it appears as though the SANT domain does have some effect on the binding of the LSD1/CoREST complex to the nucleosome. In both the hLSD1 complex v35 and v36, the replacement of acidic amino acids with neutral residues increased the binding affinity of the LSD1/CoREST complex to the nucleosome, contrary to what was hypothesized. This increase in affinity was approximately 3 times that of the wildtype complex. This was seen on nucleosomes of varying extranucleosomal DNA length. In both hLSD1 complex v37 and v55, the mutations within the SANT domain decreased their binding affinities to the nucleosome in comparison to the wildtype complex. The extent of this decrease ranged between 2 and 3 times that of hLSD1 complex v01. Different effects on binding affinity were observed with both the number of mutations made in the SANT domain as well as with the location of the altered amino acids within this domain. Further regions and quantities of mutations within the SANT domain will need to be explored to determine the factor that has the greatest effect on the binding of the LSD1/CoREST complex to the nucleosome.

4.2 Future Directions

Each of the HI-FI experiments completed were performed in either singlet, duplicate, or triplicate, depending on the amount of LSD1/CoREST protein available for that particular mutant. The experiments were performed only one time per mutant complex. In order for any trend to be shown, these results need to be replicated in triplicate for each mutant complex. The replicates should be completed on separate days to prove that the results are consistent and can be repeated at different times. The HI-FI experiments only evaluate the ability of the LSD1/CoREST complex to bind to the nucleosome. It does not determine if this change in binding has any effect on the demethylase activity of the LSD1/CoREST complex once bound to the nucleosome. A change in the interaction of the complex with the nucleosome potentially has the ability to enhance or inhibit the demethylase activity of the enzyme and there are specific demethylase assays that can test for this. These are future experiments that would be performed with each of the mutant complexes to further characterize the role of the SANT domain in the function of the LSD1/CoREST complex. If with these additional experiment the SANT domain proves to be a region of interest within the LSD1/CoREST complex, more studies can be done to analyze a wider range of binding sites within the domain itself.

REFERENCES

Bannister, Andrew J., and Tony Kouzarides. "Regulation of Chromatin by Histone Modifications." *Cell Research* 21.3 (2011): 381-95. Web.

Baron, R., and N. A. Vellore. "LSD1/CoREST Is an Allosteric Nanoscale Clamp Regulated by H3-histone-tail Molecular Recognition." *Proceedings of the National Academy of Sciences* 109.31 (2012): 12509-2514. Web.

Conaway, Joan W. "Introduction to Theme â□œChromatin, Epigenetics, and Transcriptionâ€• ." *Annual Review of Biochemistry* 81.1 (2012): 61-64. Web.

Culhane, Jeffrey C., and Philip A. Cole. "LSD1 and the Chemistry of Histone Demethylation." *Current Opinion in Chemical Biology* 11.5 (2007): 561-68. Web.

Da, G., J. Lenkart, K. Zhao, R. Shiekhattar, B.r. Cairns, and R. Marmorstein. "Structure and Function of the SWIRM Domain, a Conserved Protein Module Found in Chromatin Regulatory Complexes." *Journal of Biological Chemistry* (2006): n. pag. Web.

Igo-Kemenes, T., W. Horz, and H. G. Zachau. "Chromatin." *Annual Review of Biochemistry* 51.1 (1982): 89-121. Web.

Kang, Mk, S. Mehrazarin, N-H Park, and C-Y Wang. "Epigenetic Gene Regulation by Histone Demethylases: Emerging Role in Oncogenesis and Inflammation." *Oral Diseases* (2016): n. pag. Web.

Kim, Jin-Man, Kyunghwan Kim, Vasu Punj, Gangning Liang, Tobias S. Ulmer, Wange Lu, and Woojin An. "Linker Histone H1.2 Establishes Chromatin Compaction and Gene Silencing through Recognition of H3K27me3." *Nature News*. Nature Publishing Group, 19 Nov. 2015. Web. 30 Mar. 2017.

Kim, S.-A., N. Chatterjee, M. J. Jennings, B. Bartholomew, and S. Tan. "Extranucleosomal DNA Enhances the Activity of the LSD1/CoREST Histone Demethylase Complex." *Nucleic Acids Research* 43.10 (2015): 4868-880. Web.

Knaap, Jan A. Van Der, and C. Peter Verrijzer. "Undercover: Gene Control by Metabolites and Metabolic Enzymes." *Genes & Development* 30.21 (2016): 2345-369. Web.

Leontovyc, I., T. Koblas, L. Pektorova, K. Zacharovova, Z. Berkova, and F. Saudek. "The Effect of Epigenetic Factors on Differentiation of Pancreatic Progenitor Cells Into Insulin-Producing Cells." *Transplantation Proceedings* 43.9 (2011): 3212-216. Web.

McGinty, Robert K. and Song Tan. "Histone, Nucleosome, and Chromatin Structure." In Fundamentals of Chromatin, edited by L. Jerry Workman and M. Susan Abmayr, 1-28. New York, NY: Springer New York, 2014.

Meier, Karin, and Alexander Brehm. "Chromatin Regulation: How Complex Does It Get?" *Epigenetics* 9.11 (2014): 1485-495. Web.

Mosammaparast, Nima, and Yang Shi. "Reversal of Histone Methylation: Biochemical and Molecular Mechanisms of Histone Demethylases." *Annual Review of Biochemistry* 79.1 (2010): 155-79. Web.

Pilotto, Simona, Valentina Speranzini, Marcello Tortorici, Dominique Durand, Alexander Fish, Sergio Valente, Federico Forneris, Antonello Mai, Titia K. Sixma, Patrice Vachette, and Andrea Mattevi. "Interplay among Nucleosomal DNA, Histone Tails, and Corepressor CoREST Underlies LSD1-mediated H3 Demethylation." *Proceedings of the National Academy of Sciences* 112.9 (2015): 2752-757. Web.

Schrodinger, LLC. "The Pymol Molecular Graphics System." 1.4 (2011).

Shilatifard, Ali. "Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression." *Annual Review of Biochemistry* 75.1 (2006): 243-69. Web.

"Tan Lab at Penn State." *Tan Lab at Penn State*. N.p., n.d. Web. 30 Mar. 2017.

Tan, Song. "A Modular Polycistronic Expression System for Overexpressing Protein Complexes in Escherichia Coli." *Protein Expression and Purification* 21.1 (2001): 224-34. Web.

Tan, Song, Ronald C. Kern, and William Selleck. "The PST44 Polycistronic Expression System for Producing Protein Complexes in Escherichia Coli." *Protein Expression and Purification* 40.2 (2005): 385-95. Web.

Tochio, Naoya, Takashi Umehara, Seizo Koshiba, Makoto Inoue, Takashi Yabuki, Masaaki Aoki, Eiko Seki, Satoru Watanabe, Yasuko Tomo, Masaru Hanada, Masaomi Ikari, Miyuki Sato, Takaho Terada, Takahiro Nagase, Osamu Ohara, Mikako Shirouzu, Akiko Tanaka, Takanori Kigawa, and Shigeyuki Yokoyama. "Solution Structure of the SWIRM Domain of Human Histone Demethylase LSD1." *Structure* 14.3 (2006): 457-68. Web.

Winkler, D. D., K. Luger and A. R. Hieb. "Quantifying Chromatin-Associated Interactions: The HI-FI System." Methods Enzymol 512, (2012): 243-74.

Yang, M., C. B. Gocke, X. Luo, D. Borek, D. R. Tomchick, M. Machius, Z. Otwinowski and H. Yu. "Structural Basis for CoREST-Dependent Demethylation of Nucleosomes by the Human LSD1 Histone Demethylase." Mol Cell 23, no. 3 (2006): 377-87.

ACADEMIC VITA

Academic Vita of Victoria Spadafora
vws5110@psu.edu

Goal: To further my education and build experiences that will lead to a successful MD/PhD career.

Education: Pennsylvania State University; *Summer 2013 – Current*

1) Schreyer Honors College Scholar
    a. Member of a nationally ranked Honors College; maintain a cumulative GPA of 3.4
2) Millennium Scholar
    a. Member of a scholarship program directed towards students planning to obtain PhDs or MD/PhDs; maintain a cumulative GPA of 3.5
    b. Weekly meetings and additional networking and science related events each semester
    c. Required 6-week summer bridge including classes and seminars; *Summer 2013*
3) Academic
    a. Majors – Biochemistry, Eberly College of Science and Toxicology, College of Agricultural Sciences
    b. Dean's List Summer 2013, Fall 2013, Spring 2014, Spring 2015, Fall 2015

Research Experience

1) Pennsylvania State University- *Fall 2014 - Current*
    a. Dr. Song Tan; Center for Eukaryotic Gene Regulation
    b. Project Focus: Role of the SANT domain in binding of the LSD1-CoREST demethylase complex to the nucleosome
    c. Minimum 10 hours per week
    d. Presented my research at the Spring 2015 Undergraduate Research Exhibition at Penn State
2) University of California San Diego- *Summer 2016*
    a. Dr. Paul Insel; Department of Pharmacology
    b. Project Focus: The Role of GPRC5A in Pancreatic Cancer Cells
3) Scripps Research Institute- *Summer 2015*
    a. Dr. Dan Salomon; Molecular and Experimental Medicine Department; Transplant Immunology
    b. Project Focus: Lentivirus; T Cell Regulation; AIM2
    c. Participated in Poster Exhibition at the conclusion of the program
4) Hershey Medical Center, Hershey , PA- *Summer 2014*
    a. Dr. Chris Yengo's Lab; Cellular and Molecular Physiology
    b. Project Focus: Characterization of Cardiac Myosin Motor Function in Heart Failure
    c. Part of the Schreyer Honors College MD/PhD Summer Exposure program
    d. 10 week program, over 8 hours per day/5 days per week
    e. Participated in Poster Exhibition at the conclusion of the program
    f. Presented my research at the Fall 2015 Eberly College of Science Undergraduate Exhibition
5) Forensic Mentors Institute, Willow Grove, PA- *Summer 2012*
    a. Forensic Mentors Institute, Willow Grove, PA
    b. Project Focus: Method Development and Validation of Dimethylamylamine in Nutritional Supplements Using Gas Chromatography Mass Spectrometry
    c. Partnered with a graduate student of forensic science from Arcadia University to work on her Master's Thesis

d. 8 week program/40 hours per week

Clinical Experience

1) Hershey Medical Center, Hershey, PA- *Summer 2014*
    a. Dr. Diane Thiboutot; Dermatology, Nyes Road Clinic
    b. Shadowed a dermatologist for approximately 3 hours per week for 8 weeks
    c. Part of the Schreyer Honors College MD/PhD Summer Exposure Program
2) Penn State Hershey University Park Regional Campus, University Park, PA- *Spring 2016*
    a. Shadowed multiple Family and Community Medicine physicians
    b. Approximately 4 hours per week for 5 weeks

The Pennsylvania State University

Leadership

1) Scholar Advancement Team (Schreyer Honors College ambassador)- *Spring 2013-Current*
    a. Volunteer at alumni events for the Honors College and give tours
    b. Plan and work student events for the Honors College
    c. Work with Honors College faculty and staff
2) Scholar Advancement Team Group Leader- *Fall 2015-Current*
    a. Run orientation for new team members
    b. Meet once a month with Honors College faculty to be link between faculty and SAT team
    c. Represent the Honors College and the SAT team at alumni events
3) Millennium Society THON Chair-*Fall 2015- Current*
    a. Helped create the first Millennium Scholars Club and jumpstart the THON component
    b. Organize fundraising trips and other efforts to raise money for pediatric cancer research
4) Schreyer Honors College Shaping the Future Summit Event Co-Chair- *Fall 2014-Spring 2014*
    a. Help organize, set up, and direct events
    b. Lead volunteers, organize arrival times and duties
    c. Weekly meetings to help brainstorm, plan, and organize future events
5) Schreyer Honors College Student Council- *Fall 2013- Spring 2014*
    a. Participation in events such as Homecoming, Relay for Life, Student Recruitment, etc.
6) Member of Apollo Alternative Fundraising Team- *Fall 2014-Spring 2014*
    a. Weekly meetings to brainstorm new alternative fundraising ideas besides canning
    b. Assigned two to three projects/events at a time to see through
    c. Contact possible fundraising partners and organize events that you are assigned
7) Schreyer Honors College Student Council Service Committee Member- *Fall 2013- Spring 2013*
    a. Organized service events
    b. Planned and implicated fundraising methods for Relay for Life

Service

1) Apollo (THON org)- *Fall 2013- Current*
    a. Special interest org to raise money for pediatric cancer research and the Four Diamonds Fund
    b. Weekly meetings and a few events each month
    c. Fundraising trips up to three weekends per year
2) Rules and Regulations THON committee member- *Fall 2013- Spring 2013*
    a. Weekly meetings to prepare for THON (dance marathon for pediatric cancer)
    b. Attend THON events throughout the year
    c. Security Team during THON weekend- multi hour shifts throughout the weekend

Other Extracurriculars

1) Bioethics Club- *Fall 2015- Current*
    a. First bioethics club at Penn State involving weekly discussions of current ethical issues
2) IM Sports (once a week) : Fall IM Football (*2014)*, Fall/Spring IM Volleyball (*2013-2015*)
3) Club Sports (daily practice): Cross Country (*2013*)

4) CHAARG Fitness Club for Women

Awards and Honors

S.C. Johnson 2013 Sons and Daughters Scholarship-*Fall 2013-Current*
Provost Award- *Fall 2013-Spring 2014*
Dunkin Donuts Philadelphia Regional Scholarship- *Fall 2013-Spring 2014*
Junior League of Greater Princeton Scholarship (Award for Community Service)-*Fall 2013*
3rd in my class of 541 students at C.B East High School, Graduated Summa Cum Laude GPA 4.41- *Spring 2013*

Professional Society Membership

American Society for Pharmacology and Experimental Therapeutics (ASPET) Member- Summer 2016-cu