Identification of the Autoregulatory Properties and Molecular Interactions of the Lysine-Specific Histone Demethylase 1 Enzyme

Dulmi Senanayaka
Marquette University

IDENTIFICATION OF THE AUTOREGULATORY PROPERTIES AND MOLECULAR INTERACTIONS OF THE LYSINE-SPECIFIC HISTONE DEMETHYLASE 1 ENZYME

By

Dulmi Senanayaka, B.Sc. (Hons)

A Dissertation Submitted to the Faculty of the Graduate School,

Marquette University,

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

August 2023

ABSTRACT

IDENTIFICATION OF THE AUTOREGULATORY PROPERTIES AND MOLECULAR INTERACTIONS OF THE LYSINE-SPECIFIC HISTONE DEMETHYLASE 1 ENZYME

Dulmi Senanayaka, B.Sc. (Hons)

Marquette University, 2023

The conserved chromatin-remodeling enzyme, Lysine-Specific histone Demethylase 1 (LSD1) primarily demethylates Histone H3K4me$^{1/2}$, acting as a transcriptional repressor. It plays pivotal roles in various physiological processes including cancer and interacts with many regulatory proteins, non-coding RNAs, and small metabolites. However, little biochemical information is known about LSD1's N-terminal intrinsically disordered region (IDR) and how LSD1 interacts with various biomolecules in the context of nucleosome. I present evidence that the IDR of LSD1, containing multiple post translational modifications (PTMs) and a nuclear localization signal (NLS), can act as a reversible competitive autoinhibitor of LSD1's activity. This autoinhibition can be relieved by phosphorylation PTMs adjacent to NLS, suggesting synergistic and versatile roles of IDRs and PTMs. The combined results support a new role for phosphorylation mediated NLS regions that may function by fine tuning chromatin-remodeling enzyme activity in an auto regulatory manner.

Previous studies show that LSD1-G-quadruplex (GQ) RNA binding coincides with LSD1-nucleosome binding interface and acts as non-competitive inhibitor of LSD1's peptide substrate demethylation. Here, I present evidence that LSD1 specifically recognizes, and binds GQ-RNA. The GQ-RNA structure preferentially inhibits LSD1 activity on nucleosomal substrates suggesting an RNA structure-based effect on regulatory properties of LSD1. In addition, we propose a higher-order biophysical interaction between GQ TERRA RNA and LSD1, involved in regulating telomere maintenance. The cell localization data and in-vitro fluorescent-labeled studies reveal that TERRA and LSD1 can undergo phase separation. Phase separation appears RNA-structure dependent and suggests a model for how R loop formation and telomere maintenance can be regulated.

Lastly, LSD1's interaction network with anticancer drugs and Ewing-sarcoma oncogenic fusion protein EWS/FLI1 were examined to better understand LSD1's oncogenic mechanisms. For the first time, we show a direct interaction that occurs between LSD1 and EWS/FLI1 in vitro and in cancer cells. We find that this interaction leads to inhibition of LSD1 catalytic activity on nucleosomes. We examined LSD1's druggability and demonstrated that Seclidemstat used for treating Ewing-sarcoma is not efficient in inhibiting LSD1's nucleosomal demethylation in-vitro or in cancer cells highlighting the need for multiple validation experiments in epigenetic pharmacology. Altogether my studies advance an understanding of the wide range of molecular interactions and associated functional roles of LSD1.

ACKNOWLEDGMENTS

Dulmi Senanayaka, B.Sc. (Hons)

I would like to express my first and foremost gratitude to my academic advisor Dr. Nicholas J. Reiter for the tremendous support, encouragement, and great mentorship throughout my graduate school. This dissertation would not be possible without his continuous guidance and patience. I am very fortunate to have him as my dissertation advisor. I would like to thank my committee members, Dr. Chae Sung Yi, Dr. Martin St Maurice, Dr. Ofer Kedem, and Dr. Christopher Dockendorff for their insightful comments, thoughtful suggestions, and support throughout my thesis work.

I'd like to express my gratitude to collaborators of my projects: Professor Huaiying Zhang, Dr. Meng Xu, Carnegie Mellon University, Professor Aykut Üren, and Dr. Emre Deniz, at Georgetown University. I would like to thank all past and present members of the Reiter laboratory including Dr. William Martin and Dr. Danyun Zeng, for collaborative research, helpful discussions, and constructive criticism of my research. Many thanks to Marquette University's Department of Biological Sciences, Medical College of Wisconsin, Northwestern University, NMRFAM at the University of Wisconsin Madison, and LS-CAT at Argonne National Laboratory.

I thank all my schoolteachers from Grade 1 to Grade 13, lecturers at the University of Ruhuna, Sri Lanka. I extremely appreciate the financial assistantship from the United States National Institute of Health (NIH RO1 grant), Marquette University Graduate School, and the Department of Chemistry. I'm grateful to my family, especially my ever-loving brother Hasiru, my grandparents, and my friends for being there for me in each step of my life journey through ups and downs. Thank you for the sacrifices, endless love, friendship, encouragement, and joy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Structure and Function of Chromatin

The haploid human genome contains approximately 3 billion base pairs of DNA packaged into 23 chromosomes. When fully extended each cell consists of about two meters of DNA compacted into the microscopic space of the eukaryotic nucleus with an average diameter less than 10 μm. Specific type of proteins called histone proteins coordinate and compact chromosomal DNA through electrostatic interactions, resulting in a polymeric DNA – protein complex called chromatin. Interestingly chromatin not only serves to condense DNA within the nucleus of the cell but also permits the site-specific coordinated accessibility of transcription factors and regulatory apparatus to the genome and controls the gene expression by keeping segments of the genome transcriptionally active or repressed. (1-3)

1.1.1 Nucleosome Core Particle

The fundamental structural unit of chromatin is called the nucleosome. The nucleosome serves three primary functions. Firstly, it compacts ~ 200 bp DNA and serves as the first level of genomic compaction. DNA in eukaryotic chromatin is arranged as tightly packaged arrays of nucleosomes. Within the nucleus of the cell, DNA accessibility and packaging is primarily determined by the nucleosome core partice (NPC). (4) Secondly, the NPC provides a scaffold for binding chromatin modifying enzymes and allowing post translational modifications (PTMs) that regulates the

recruitment of chromatin enzymes. Finally, the NPC allows further compaction of the genome by assembling into higher order chromatin structures. (3)

Each ~ 200kDa disc shaped nucleosome consists of a nucleosome core of eight histone molecules (about 100kDa), linker DNA, linker histone protein and 145-147 base pairs of DNA (~100kDa) wrapped around the octamer core of histones to form between one and two left handed super helical turns.(2,4-6) The octamer consists of two copies of each four core histone proteins H2A, H2B, H3 and H4 (Figure 1-A, B). Nucleosome core particle is roughly cylindrical with a height of about 5.5 nm and diameter of 100Å.(6) Each of the highly conserved nucleosome occurs every 160-240 base pairs of the eukaryotic genome, nucleosome repeat lengths varying on the cell type, tissue type and species type. Repeating nucleosomes are further assembled into higher order oligonucleosome structures separated by 10-90 base pairs of linker DNA and are stabilized by linker histone protein H1 or H5. (4)

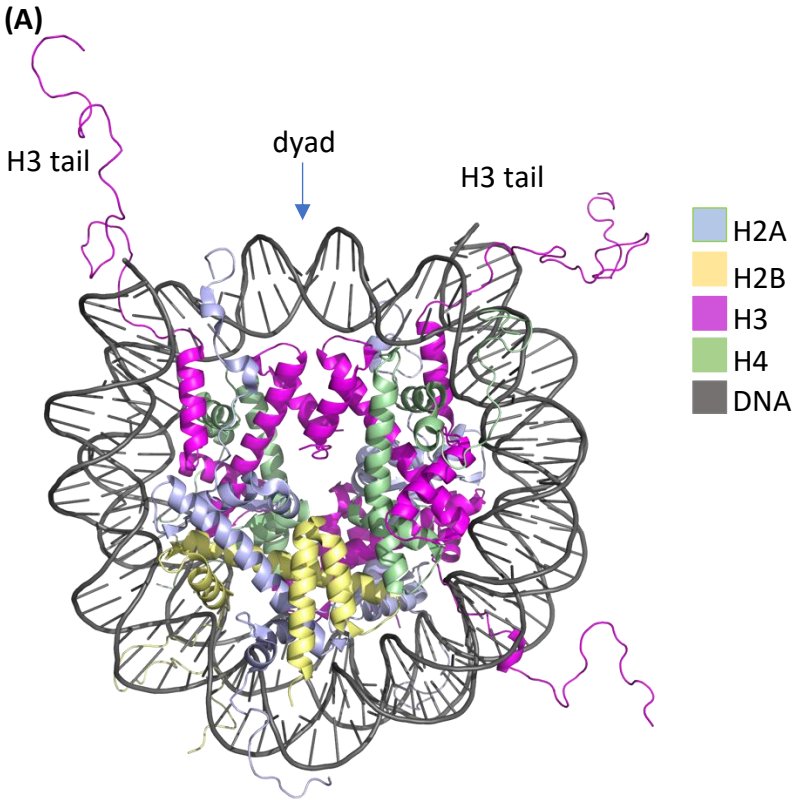
The first 2.8 Å resolution crystal structure of nucleosome and subsequent 1.9 Å resolution crystal structure of nucleosome shows the atomic details of histone protein octamer, 146 and 147 base pairs respectively, of human alpha satellite DNA arranged in 1.65 turns of flat left-handed super helical organization around the octamer.(4,7) The histone octamer contains two of each histone fold heterodimers H2A-H2B and H3-H4. Two H3-H4 dimers interact with each other in a head-to-head arrangement through H3 / H3 four helix bundle to form a $(H3/H4)_2$ tetramer (Figure 1-C, D). Each H2A-H2B dimer binds with half of the $(H3/H4)_2$ tetramer through a 4 -helix bundle formed by H4 and H2B histone fold motifs, to produce the octamer.(4) The two halves of nucleosome core

particle are related by a pseudo two-fold symmetry axis that passes through the central base pair in the 147 bp nucleosome and one of the two central base-pairs in 146 bp nucleosome. (4,7)

The amino acids sequence of histones is highly conserved within diverse forms of organisms throughout evolution. Histones are relatively small, ranging from 102-135 amino acids, highly basic proteins that are rich in arginine and lysine. Therefore, the highly basic surface of the histone octamer binds with the highly negative electrostatic surface of phosphate oxygen atoms within nucleosomal DNA. Additionally, within a groove on the H2A /H2B dimer surface, there is a negatively charged surface referred to as nucleosome acidic patch. Six acidic residues from H2A (E56, E61, E64, D90, E91 and E92) and two acidic residues from H2B (E102, E 110) contribute to the acidic patch surface. It is predicted that this acidic patch may be an electrostatic binding recognition element for chromatin factors.(3)

Each of the histone proteins consist of poorly ordered N-terminal tail region and well-ordered alpha helical region that forms the histone-fold motif. At defined locations, ten flexible N and C terminal tails protrude from the surface of the nucleosome either outside (H4 and H2A) or between (H3 and H2B) the DNA gyres. Eight of them are N-terminal tails from each of the eight core histone proteins. Two additional C terminal tails protrude out from where the DNA enters and leaves nucleosome and are contributed by the histone H2A protein. The flexible N terminal tails of core histones that stick out from the nucleosomes range from 15- 36 amino acids and comprise of roughly 25% of the mass of the histones. They contain many arginine and lysine residues

and can extend further away from nucleosome core particle to bind neighboring nucleosomes, impacting histone surfaces and intranucleosomal DNA. These post-translational modifications on histones are epigenetic markers because they directly regulate transcription, gene activity and DNA dependent processes. (8)



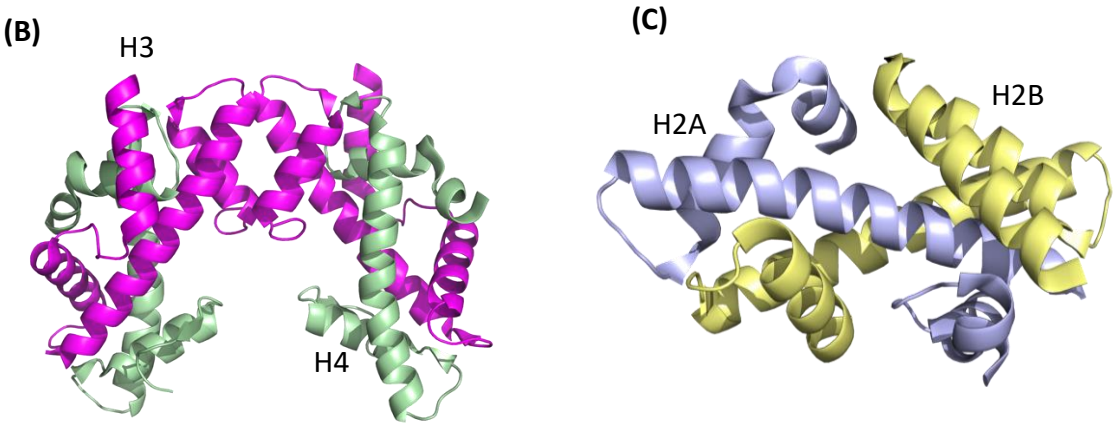


Figure 1 : Secondary structure of histone fold heterodimers and overview of the nucleosome architecture.
(A) Cartoon representation of histones and DNA in the nucleosome core particle with the dyad axis marked by an arrow (PDB ID 1KX5) **(B)** H3 H4 tetramer **(C)** H2A H2B dimer (3)

1.1.1.1 Recognition of the Nucleosome by Chromatin Factors and Enzymes

Recruitment of macromolecular chromatin factors to genomic loci is regulated by the degree of chromatin compaction. A variety of mechanisms including histone and DNA covalent modifications modulate chromatin condensation. Based on its level of compaction, interphase chromatin is classified as either euchromatin or heterochromatin. Euchromatin appears as an 11 nm fiber with less compact 'beads on a string' structure. In contrast, in the presence of linker histone or chromatin associated proteins, heterochromatin can be formed. (9) It is a more compact 30 nm fiber of condensed array of nucleosomes. By forming a repressive environment, these hierarchical chromatin structures are inaccessible to transcriptional machinery, and inhibit enzyme activities that require direct access to DNA template thus silencing genes by inhibiting transcription, DNA replication, repair, and recombination. In the lightly

packed euchromatin state, nucleosomes provide a platform for binding many macromolecular chromatin factors. Some transcription factors that bind DNA bind to specific DNA sequences in nucleosome free regions. (3) Chromatin factors use one or more of, histone octamer surface (and its H2A/H2B acidic patch), nucleosomal DNA or Histone N or C terminal tails to bind nucleosomes. (10)

The co-crystal structures of chromatin protein RCC1 (Regulator of Chromosome Condensation) with the nucleosome and first chromatin enzyme, the Polycomb repressive complex 1 (PRC1) E2-E3 ubiquitylation module bound to the nucleosome core particle, provide the first atomic views of how a chromatin factor and a chromatin enzyme interact with nucleosome. (11,12) In the structure of RCC1-nucleosome, RCC1 binds to H2A/H2B dimer component of the histone octamer. Meanwhile, in the crystal structure of nucleosome – ubiquitylation module of PRC1, a Polycomb group epigenetic enzyme that represses expression of developmentally regulated genes in higher eukaryotes, shows how a chromatin enzyme achieves substrate specificity by interacting with multiple nucleosome surfaces such as disc face of the octamer, nucleosomal DNA and histone tails that are spatially distinct to each other.(11,12)

Binding of chromatin factors and enzymes to histone tais is due to the molecular recognition of histone tails by catalytic domains of these enzymes that establish or remove histone post translational modifications (PTMs) as well as protein domains that bind histone tails with specific post translationally modified residues.

1.1.1.2 Post Translational Modifications (PTMs) of Histones

Gene expression patterns are positively or negatively regulated through post translational modifications of histone proteins or by chromatin remodeling complexes that fine tune structural properties of chromatin causing heterochromatin or euchromatin. PTMs modulate intrinsic nucleosome dynamics and influence binding of chromatin remodelers, histone chaperons and guide cellular factors to their specific destinations in the genome. (13) Histone flexible tails subject to post translational modifications that may constitute a 'histone code' that extend the genetic message into epigenetic information, beyond the DNA sequence. Eight different types of histone PTMs have been reported in more than 70 different sites of histones including random coil N terminal tails (Figure 2- A), C terminal tails and internal globular domains of histones. (14) Eukaryotic PTMs are added as simple covalent modifications such as methylation, acetylation and phosphorylation, ADP ribosylation, deimination and complex modifications like addition of carbohydrates or lipids, covalent linkage of small protein (e.g., Ubiquitin) and proteolysis or irreversible hydrolysis of a specific peptide bond between two amino acids, succinylation, malonyation, sumoylation, crotonylation as well as non-covalent proline isomerization. (9) . Specific enzymes catalyze dynamically added PTMs and most of them are enzymatically reversed by the opposite catalytic activities. Various PTMs display crosstalk between different modifications. Specific modifications synergistically regulate the integration of transcriptional output and alter the chromatin organization by affecting the binding of certain domains

thereby affecting the presence of a second modification elsewhere in a histone protein. (15) (16)

Amino acids including lysine (K), arginine (R), histidine (H), serine (S), threonine (T), tyrosine (Y), and glutamic acid (E) can be modified through PTMs. Among others, some of the PTMs include deamination and methylation of arginine, phosphorylation of serine, tyrosine, threonine as well as ADP ribosylation of glutamate residues. Besides methylation and acetylation, lysine amino acid residues can be sumoylated and ubiquitinated. (13) Due to their accessibility to modifying machinery, histone N- and C-terminal tails that protrude from the nucleosome are predominantly susceptible to most of these post translational modifications. (17)

(A)

(B)

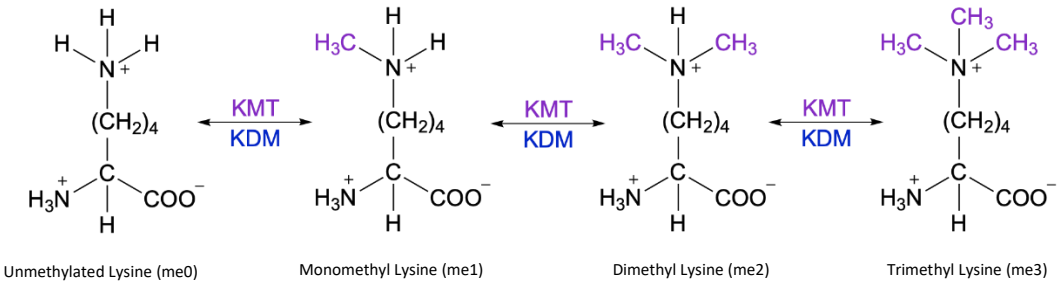


Figure 2: Examples of Histone post translational modifications.
(A) Post translational modifications at amino terminal tails of human core histones. Purple dotted residues represent methylation sites, red color represents phosphorylation sites and green color indicates known acetylation sites. Blue Ac-NHs represent N-terminal acetylation of H2A and H4 tails. (18) **(B)** Structures of methyl lysine, dimethyl lysine and trimethyl lysine amino acids.

Histone acetylation at conserved lysine residues is one of the most intensely studied, well characterized modification considered as a transcription activation mark that neutralize positive charge of histone tails and decrease the affinity between basic histones and negatively charged DNA.(16) Histone acetyltransferases (HAT) establish acetylation by addition of an acetyl-coA to the ε amino group of lysine side chains that ultimately creates an "open" chromatin structure, poised for transcriptional activation through the exposure of DNA-binding sites. The reverse reaction of HATs is carried out by histone deacetylases (HDACs) which removes the acetylation on lysine residues and mediate transcriptional repression and gene silencing. (15) So far, eighteen enzymes belonging to four classes of HDAC superfamily (class 1, class11, class 11a, class 11b, class 111) have been identified. (19) Similarly, phosphorylation of histone residues that add a negatively charged phosphoryl group which can neutralize the positive charge of histone tails, act as an activation mark and thereby decrease the interactions of histone tails with DNA. The addition of a phosphoryl group and removal of phosphorylation is regulated by protein kinases and phosphatases, respectively. In addition to regulation of transcriptional activity, several studies have suggested that histone phosphorylation is associated with DNA damage repair and control of chromatin compaction associated with mitosis and meiosis. (16) (19) Ubiquitination regulates most cellular functions in

many cell-signaling pathways in eukaryotes. During ubiquitination, histone ubiquitin ligases add 8.5 kDa protein to the amino acid residues and these can be removed by ubiquitin-specific peptidases known as deubiquitinating enzymes. (19)

The addition of one, two, or three methyl groups on amino acid side chain residues such as lysine, arginine, histidine, and glutamine can dramatically impact the functional properties of a protein. (20) Meanwhile, histone methylation occurs at various sites of histone proteins, predominantly on lysine and arginine residues making it one of the most studied examples of PTMs in histones. (21)

1.2 Histone Methylation

Post translational modification of histones by methylation has been shown to play a role in important biological processes such as DNA damage response, stem cell maintenance and differentiation, stress response, cell cycle regulation, transcription regulation, X inactivation, RNA processing as well as disease and aging. (22) (23) Chemically, histone lysine residues can be mono (me), di (me2) or tri (me3) methylated by addition of one two or three methyl groups to the ε nitrogen of a lysine side chain. Arginine residues can be monomethylated (me1), symmetrically dimethylated (me2s), or asymmetrically dimethylated (me2a) on their guanidino group.(14) (18) (24)

Methylation of histone and non-histone proteins generated by methyltransferases and demethylation is performed by demethylases. In human, protein methylation is mediated by a number of universal methyl group donor S-adenosylmethionine (AdoMet/SAM) dependent methyltransferases that facilitate the transfer of a methyl group from S-adenosyl-L-methionine to the ε-amino group of a

histone lysine thus converting SAM to S-adenosyl-L-homocysteine (SAH).(15)(25)(26) These methyltransferases belong to two distinct classes, the SET (Su(var)3–9, Enhancer of zeste and Trithorax) domain containing methyltransferases and the seven-beta-strand (7βS) domain containing methyltransferases that mediate the transfer of a methyl group to histones associated with chromatin, free histones and non-histone proteins.(24) Lysine specific methyltransferases mainly belong to the family of SET domain methyltransferases while 7βS family methyltransferases, characterized by a twisted beta sheet structure, act upon a wide range of substrates including lysine and arginine, nucleic acids, small metabolites, proteins and lipids. (27) The so called 7βS methyltransferases represent the largest methyltransferase class of enzymes while histone specific DOT1 is the first identified eukaryotic 7βS lysine methyltransferase.(28) Arginine methyltransferase (PRMT) proteins that have been shown to methylate arginines also belong to the 7βS family of methyltransferases. (22) Nine protein arginine methyltransferases (PRMTs) encoded in the mammalian genome mediate the three types of arginine methylations and serve as critical regulators of cellular responses including cancer development, aggressiveness, progression, hepatic gluconeogenesis, and T-lymphocyte activation. Among them, PRMT1 enzyme is responsible for asymmetric dimethylation and PRMT5 is the primary enzyme that symmetrically dimethylates arginine residues.(22) The most extensively studied histone arginine methylation sites include histone H3R2, H3R8, H3R17, H3R26 and H4R3.

Histone lysine methylation has been the most intensively studied example of the functional significance of protein methylation. Methylation of distinct lysine residues

along the Histone H3 N terminal tail is a widely studied example of Lysine methylation and well-known sites of lysine methylation include histone H3 lysine 4 (H3K4), H3K9, H3K27, H3K36, H3K79 and H4K20 mono di and tri methylation. Each of these methylation states represent a well-defined marker for biological function and chromatin localization. For example, methylation at H3K4 is associated with gene activation while H3K9, H3K27 and H3K20 methylation is associated with gene repression. In the eukaryotic genome, the promoter regions of actively transcribed genes contain H3K4 trimethylated nucleosomes and transcriptional enhancers contain monomethylated H3K4. (29) Cross talk between methylation and phosphorylation has been recognized.(15) Addition of methyl groups increase the hydrophobicity and size of protein lysine side chains. Therefore, proteins with aromatic cages interact with methylated arginines and methylated lysines. Unlike PTMs such as phosphorylation and acetylation, histone lysine methylation localized at specific locations of the genome does not alter the charge of histone proteins and is correlated with both gene silencing and activation depending on the location of the methyl lysine and the and degree of methylation (me1, me2 or me3) (Figure 2-C). (24)

1.2.1 Regulation of Histone Lysine Methylation and Demethylation

Histone lysine methylation occurs in both promoter and enhancer regions of genes and is often associated with active transcription. For example, primed enhancer elements contain H3K4me marks, active promoters and enhancers contain H3K4me2 marks and H3K4me3 marks can be found in actively transcribed promoters. (30) Some of the basic functions of lysine methylation marks on core nucleosomal histone H3 and

H4 include, transcriptional regulation, RNA processing, silencing, DNA repair, genome stability, DNA replication and DNA recombination. (27) Site and degree of histone lysine methylation is dynamically modulated by the balanced activities of lysine methyltransferases (KMTs) and lysine demethylases (KDMs). The dysregulation of KMTs and KDMs is often associated with neurological disorders, developmental abnormalities, and cancer. (31) In the human proteome, over 1000 proteins harbor lysine methylation through the enzymatic activity of lysine methyltransferases. Conserved from yeast to humans, there are reported to be over 50 lysine methyl transferases in the human genome that belong to SET domain family and 7βS domain family of KMTs. (28) The first ever histone lysine methyltransferase SUV39H1, discovered in 2000, deposits histone H3 lysine 9 di- and trimethylation (H3K9me2/3) marks forming heterochromatin. (27) In addition to the SUV39 class, SET1 family, SET2 family and RIZ family are well characterized human SET domain protein families with homologues SET domains. (15)(16) (26) (32) Two of the most studied members of the SET1 family of KMTs are EZH1 and EZH2. SET domain KMTs methylate N terminal tails of Histone H3 (at K4, K9, K27, K46). DOT1L is the only KMT enzyme that is responsible for mono, di and tri methylation of H3K79 which is in the core of H3 and is responsible for gene activation.(28)

Histone methylation was long thought to be an irreversible genomic imprint until the groundbreaking discovery of the first histone lysine demethylase enzyme LSD1 (Lysine specific histone demethylase 1) in 2004, that was previously identified as a member of several deacetylase complexes.(33) (24) (34) Since the discovery of the first

KDM, to date, two families of histone demethylase enzymes that demethylate methylated lysines have been reported. These are LSD family amine oxidases and Jumonji C (JmjC) family of Fe (II)- and 2-oxoglutarate (2-OG)-dependent KDMs. Both amine oxidase and Jumonji C histone demethylase (JHDM) families are highly evolutionarily conserved from yeast to human.(24)

The subfamilies of JmjC KDMs display different methylation site specificities (H3K4, H3K9, H3K27, H3K36) as well as methylation state specificities (me1, me2, me3) than LSD family demethylases. For example, H3K4 di and trimethylation marks are demethylated by Jumonji domain protein JARID1B while the same location mono and dimethylation is demethylated by LSD1. Although, ~30 members of proteins identified in humans belonging to JHDM family contain various other non-catalytic domains, they each feature a Jumonji C catalytic domain. The chemical reaction catalyzed by JmjC domain is the iron dependent oxidation of mono, di, or tri methyl groups by radical attack of a highly reactive oxoferryl species. This reaction forms an unstable carbinolamine intermediate which spontaneously dissociated into demethylated lysine and formaldehyde byproduct.

Interestingly it appears that there is coordination between histone lysine methyltransferases and demethylases as well as between different histone lysine demethylases. For example, both KDMs LSD1 and JHDM2 family members associate with androgen receptor and demethylate same methylated lysine at H3K9. (35) (36) In addition, long noncoding RNAs (lncRNAs) are suggested to play an important role in directing certain histone demethylases and methyltransferases to specific locations of

the genome. For example, Human lncRNA HOTAIR binds to PRC2 complex that is responsible for di and tri methylation of H3K27 and H3K4/9 demethylase LSD1 to coordinate the recruitment of them that leads an efficient repression of specific loci.(37)

1.2.2 Lysine Specific Demethylase (LSD) Family of Histone Demethylases

Lysine specific demethylase family of histone demethylases consist of FAD (flavin adenine dinucleotide) -dependent amine oxidase homologs LSD1 (KDM1a/AOF2) and LSD2 (KDM1b/AOF1), that use FAD dependent demethylation reaction to catalyze the demethylation of the mono and dimethylation marks. LSD family is characterized by presence of amino oxidase catalytic domain and a SWIRM (Swi3p, Rsc8p and Moira) domain. LSD1 contains catalytic amino oxidase domain, SWIRM domain and a coiled coil TOWER domain. However, LSD2, the only homolog of LSD1 in the human genome, contains amino oxidase domain and SWIRM domain but lacks a tower domain. In the region corresponding to LSD1s' unstructured N terminus, LSD2 consists of an amino terminal zinc finger domain. This distinct domain architecture facilitates the two LSD family proteins to associate different protein complexes and different genomic loci. Unlike LSD1, LSD2 is unable to form complex with REST co-repressor proteins (CoREST1-3).(38) LSD2 associates predominantly with gene bodies of actively transcribed genes, and it is known to specifically demethylate H3K4me1 and H3K4me2 within intragenic regions of its target genes. (39)

1.3 Lysine Specific Histone Demethylase 1 (LSD1)

1.3.1 Biological Functions of LSD1

LSD1, also known as KDM1a, is an 852 amino acid protein of 110kDa that belongs to the FAD dependent family of mono amine oxidases (MAO) and was noted to have some sequence homology with the family of polyamine oxidase enzymes. (30) LSD1's structure and function is conserved from yeast to human although it exhibits relatively modest levels of sequence conservation across eukaryotes. With remarkable specificity, LSD1 selectively removes mono or dimethyl groups of histone H3K4 (H3K4me$^{1/2}$). It is also reported that upon interaction with androgen receptor, LSD1 can demethylate H3K9me$^{1/2}$.(35) By demethylating H3K4me$^{1/2}$ and H3K9me$^{1/2}$ LSD1 can act as a transcriptional repressor or activator respectively. Mammalian LSD1 has four full length isoforms (LSD1, LSD1-8a, LSD1-2a, LSD1-2a/8a) from combinatorial retention of two exons (E2a, E8a) formed by alternative splicing by which LSD1 acquires its differential substrate specificities in neurons. (40) (41) Inclusion of exon E2a occurs in all tissues while retaining E8a is a neurospecific event. Therefore LSD1, LSD1-2a are ubiquitous and LSD1-8a, LSD1-2a/8a are restricted to neuronal histotype. Isoform LSD1+8a where the 4 amino acid exon E8a is internal to the amine oxidase domain resulting in a small protruding loop near the catalytic site, mediates H3K9me^{2} demethylation in collaboration with LSD1+8a interacting protein supervillin (SVIL) in neuronal cells.(42) (40) A separate study reported that LSD1+8a functions as an H4K20 demethylase in cortical neurons. In addition, LSD1 demethylates several non-histone protein targets such as Lys1096 on DNA methyltransferase 1 (DNMT1), K370me1 and K370me2 on p53

(tumor suppressor and transcriptional activator), Lys185 on E2F1, STAT3, TP53 and regulate their activity, protein interactions and stability. (38)(43) Interestingly there are no sequence similarities between LSD1 histone substrates and non-histone substrates suggesting regulatory roles for protein interaction partners in lysine demethylation.

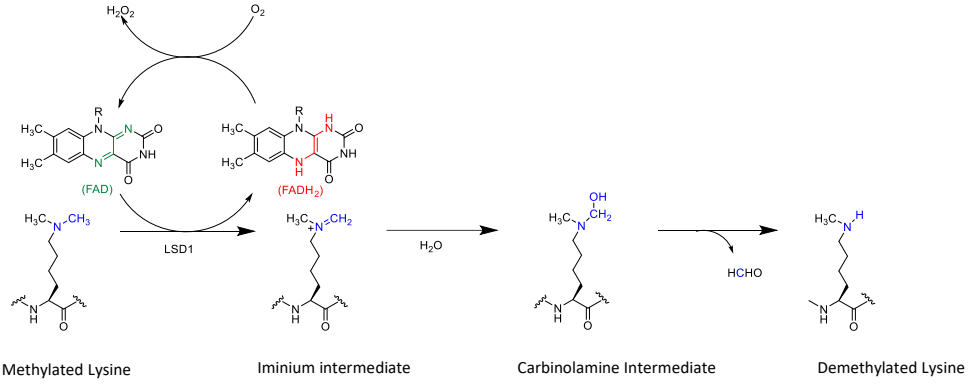
LSD1 silences neuronal specific genes in non-neuronal cells as well as plays an important role as a regulator of neural stem cell proliferation and neuronal development. While being highly expressed in undifferentiated human embryonic stem cells, and being progressively downregulated through differentiation, LSD1 is suggested to play a role in development and differentiation. Through the regulation of H3K4 methylation levels, LSD1 is reported to be a key player in maintenance of embryonic stem cell pluripotency. It has been shown to control hemangioblast formation as well as regulate the generation of hematopoietic stem cells in vertebrates. (30) Thus, LSD1 has also been reported to play a variety of physiologically relevant roles and is involved in the molecular mechanisms of DNA damage response, cancer (cell differentiation), and neurological disorders.

LSD1 is a subunit of transcriptional corepressor complexes like CoREST, nucleosome remodeling and deacetylase complex (NuRD), BRAF-35, CtBP, lysine methyltransferase mixed lineage leukemia (MLL) coactivator complex and it is an interaction partner of histone deacetylase HDAC2.(30)(44) CoREST complexes also enclose HDAC1 or HDAC2 together with LSD1 to positively influence LSD1 activity and increases the affinity of the entire complex for chromatin. Likewise, LSD1 enzymatic activity is often linked with optimal deacetylation in the cell, suggesting mutual,

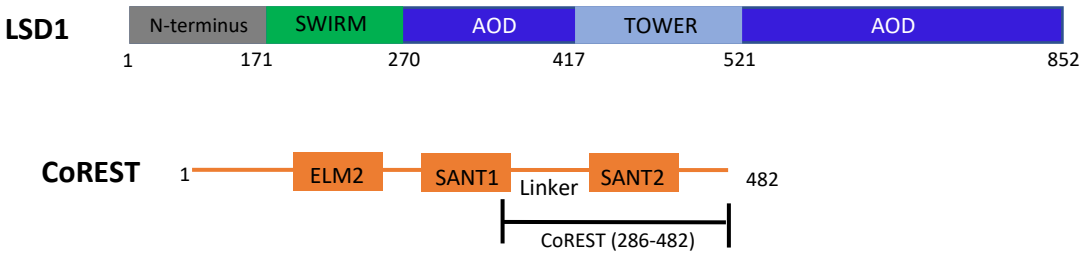
synergistic functions. The CoREST complex of LSD1 with HDAC1 or HDAC2 is known to remove histone acetylation and activate LSD1 demethylation on nucleosomes. (45)(46) Thus the association of LSD1 with corepressor proteins like CoREST and NuRD are essential for its optimal demethylation activity on nucleosomal substrates, and also is involved in regulating LSD1's activity on non-histone substrates. (47)

1.3.2 Structure of Human LSD1

(A)



(B)



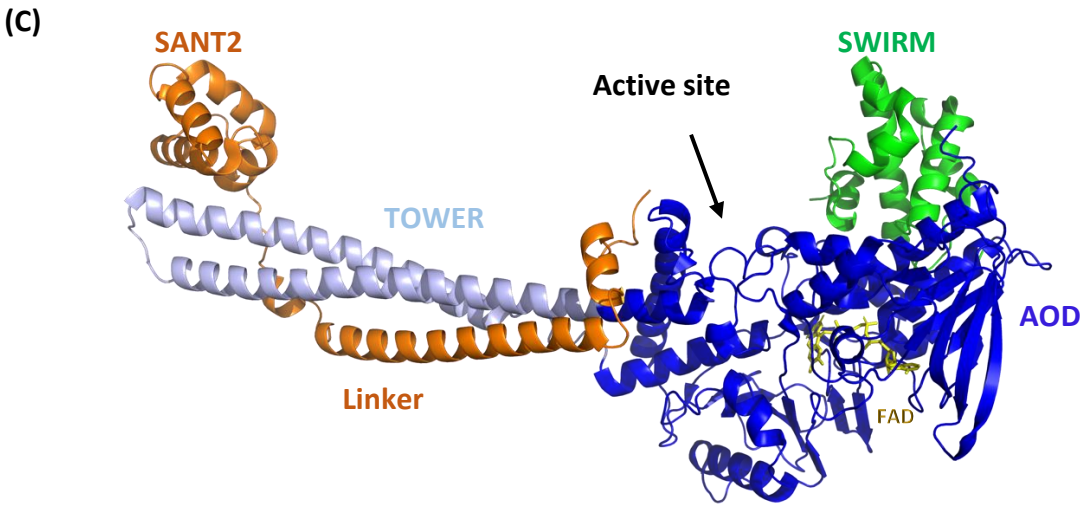


Figure 3: The chemical reaction, domain organization and crystal structure of LSD1-CoREST complex.
(A) The mechanism depicting LSD1 catalyzed demethylation of methylated lysine. **(B)** domain organization of human LSD1 and human CoREST with the boundaries of proteins used in crystallization. **(C)** The crystal structure of ΔN LSD1(171-852)-CoREST-C (286-482). PDB ID: 2IW5 (48) The AOD, TOWER, SWIRM domain and FAD in LSD1 structure are colored in blue, light blue, green, yellow respectively. The CoREST linker and SANT2 are colored in orange.

The first crystal structure of human LSD1 was published in 2006 by Stavropoulos *et al.* in 2006. (49) The human LSD1 comprises of 852 amino acids polypeptide chain divided into several functional regions. It consists of a N terminal 170 amino acids long disordered region, followed by a SWIRM domain, C terminal amino oxidase domain (AOD) and a TOWER domain that is inserted to AOD. (49) The 98 residue SWIRM domain (172-270) intimately bids atop AOD through hydrophobic interactions and helps to pack the FAD binding lobe of AOD. The SWIRM domain is responsible for protein-protein binding interactions such as LSD1 binding to androgen receptor but lacks the high affinity canonical DNA binding properties that other SWIRM domain proteins display.

(47)(50) Thus the N terminal SWIRM domain functions as a stabilizer of LSD1's structure and is considered a protein 'interaction hub' that resides adjacent to the catalytic domain. (47) The TOWER domain (417-521) is a helical hairpin structure about 100 Å in length that forms two long antiparallel α-helices (Iα1 and Iα2) bound tightly to each other by hydrophobic interactions forming a long stalk that projects away from AOD. This coiled-coil structural element is crucial for both the interaction of LSD1 with CoREST and facilitates binding of several homologous proteins. In this way, the TOWER domain enables optimal demethylase activity of LSD1 on its physiologically relevant nucleosome substrate. The amino oxidase domain of LSD1, which shares sequence homology to FAD-dependent oxidases, is divided into two separate functional lobes: one forming the substrate binding and recognition site and the other forming the FAD binding site that is separated by a 92 residue insert. Non covalently bound FAD sits deeper in the active site pocket and is optimally oriented through the interaction with positively charged Lys 661 residue that helps to stabilizes the anionic form of FAD.

1.3.2.1 Substrate Binding to the Active Site Cavity of LSD1

The amino oxidase domain of LSD1 (271-416, 522-852) is characterized by a remarkable 1245 $Å^3$ large, ∼15 Å deep, open active site cleft that is located at the interface of the two AOD lobes, spanning ∼25 Å at its widest opening with a highly negative electrostatic potential that exposes the isoalloxazine ring of FAD deeper in the pocket.(47) The surface of the cavity comprises about 50 evolutionarily conserved residues. The negatively charged residues in the rim of the active site cavity facilitate the

electrostatic interactions between LSD1 active site and the positively charged histone N terminal tail, guiding the substrate into the active site. (48)

Inside the cavity, several hydrophobic residues are required for the exact positioning of the FAD cofactor and methylated lysine substrate. The active site pocket creates distinct chemical environments capable of interacting with different residues and modifications in the substrate.(49) According to the crystal structure of LSD1-CoREST bound to H3K4M peptide, LSD1 recognizes and binds the first 16 N terminal amino acids of the H3 tail without any conformational change of the LSD1-CoREST complex upon substrate binding.(51) In this structure, the Arg2, Gln5, Ser10 amino acids of the histone peptide mimic engage in several critical intramolecular interactions with other peptide residues. Moreover, all peptide residues establish key intermolecular interactions with the active site cavity residues. The methyl group at Lys4 is about 3Å distance from the reactive N-5 atom of the flavin and the methylamino group of the Lys4 is fixed into the position through the interaction with LSD1s' Tyr 761 aromatic side chain facilitating flavin mediated oxidation of N-CH_3 group of methylated lysine4.(51) Many residues in the catalytic cavity are conserved in LSD1 orthologs and homologs. Among them, the Lys661 in the catalytic cavity plays a crucial role in LSD1 catalytic activity by H bonding to the N5 atom of the FAD via a conserved water molecule. Through the water bridge, Lys661 deprotonates H3K4me^2 by allowing the hydride to be transferred onto FAD via oxidative demethylation. These mechanistic studies suggest the importance of the conserved Lys661 amino acid, and the K661A point mutation has been shown to completely abolishthe LSD1 demethylation activity. (52)

1.3.2.2 Activity of LSD1

With the help of numerous sequence specific interactions, the H3 peptide in the active site adopts several secondary structural turns to fit snugly into deep inside the cavity. Through binding, the H3 peptide goes from being an IDR outside of the active site pocket to adopting a putative structural element when it is embedded in the active site cleft. Within the active site cavity of LSD1 the mono and dimethylation is removed through an FAD dependent amine oxidation reaction that involves the stepwise conversion of H3K4me^1/me^2 to a demethylated H3K4. LSD1 oxidizes the methyl ε -amine of lysine by the cleavage of the α carbon bond, with the simultaneous 2 electron reduction of FAD to FADH2 to form an imine intermediate. The intermediate is hydrolyzed via non enzymatic process to produce carbinolamine which spontaneously decomposes to form formaldehyde and the demethylated lysine. The reduced FADH2 produced in the initial two electron hydride transfer reaction step is rapidly re-oxidized to FAD by molecular oxygen producing a molar equivalent of H_2O_2. (36)(53)

For LSD1 to be catalytically active, there should be at least 16 amino acids in the substrate peptide, but the optimal LSD1 activity has been determined when using peptides with the first 21 amino acids, which represents the N terminal portion of histone H3 substrate. (51) Since the large active site pocket can accommodate several residues near the target lysine, it enables LSD1 to sense post-translational modifications in the vicinity of the target lysine and the surrounding local environment. Since the protein-substrate association is dominated by electrostatic interactions, the ability of LSD1 to physically interact and demethylate the substrate is influenced by such

mutations and nearby post translational modifications (PTM) in the substrate, providing another level of regulation that affect LSD1 activity.(30) For example, replacement of the N terminal amino group of H3 peptide with a methyl group or addition of acetyl group to the N terminus of H3 peptide greatly reduces its demethylation by LSD1.(53) The H3 N terminal tail PTM, Ser10 phosphorylation is an activating signal that disrupts recruitment of transcriptional repressors by Lys9 methylation mark. Moreover, Ser10 phosphorylation disrupts the productive binding geometry of the N terminal tail and totally abolishes the enzyme activity. Thus, Ser10 dephosphorylation is a prerequisite for LSD1 demethylation. (54) On the contrary, substrate Lys9 acetylation decreases the enzymatic activity by decreasing the substrate affinity and increasing the K_m value due to removal of the favorable interactions between positively charged Lysine and negatively charged residues in active site. Therefore, Lys9 deacetylation by HDAC1/2 is favorable for the LSD1 activity by facilitating the productive way of H3 binding. (45) However, Lys 9 methylation which is known to cause gene repression does not affect the H3K4 demethylation reaction of LSD1. (51)(54)

LSD1 can only demethylate H3K4me^1/me^2 that contain a lone pair of electrons present on the lysine ε-nitrogen atom in their deprotonated state, therefore cannot demethylate trimethylated H3K4. Even though LSD1 cannot demethylate H3K4me^3 peptide, unmodified (H3K4) and trimethylated (H3K4me^3) peptides are able to bind the active site cavity and act as competitive inhibitors suggesting that LSD1 cannot distinguish between H3K4/me^1/me^2/me^3 peptides. (54)(55) A previous study has shown that full length H3 without any post translational modification acts as a tight binding

competitive inhibitor of LSD1 approximately with a 100-fold higher inhibitor constant (Ki) than 21 mer product peptide suggesting a secondary binding site for Histone H3.(55) This is supported by the crystal structures of peptide substrate bound to LSD1 that reveal two main conformations where the C terminus of the peptide exiting the active site in two different orientations. Interestingly, it has been suggested that Tetrahydrofolate (THF) binds to LSD1 at the active site. A crystal structure shows THF in close proximity to FAD, where the THF metabolite has the ability to accept formaldehyde, forming 5,10-methylene-THF (a key metabolite in intracellular detoxification mechanisms). Here, the small metabolite was shown to sterically disrupt productive substrate binding conformation of the H3 tail. (56) (57) Although LSD1 alone can demethylate a 21 amino acid minimal H3K4 me^1/me^2 peptide substrate, the demethylation of H3K4me^1/me^2 on nucleosome substrates requires the CoREST transcription factor. The corepressor Co-REST heterodimerizes with LSD1, not only forming a stable complex that functions on nucleosomes, but also protects LSD1 from proteasomal degradation. (30)(48)(58)(45) Even measured with a small peptide substrate that does not interact with CoREST, there is a 2-fold increase in rate of catalysis when LSD1 is bound to corepressor CoREST. (51) As mentioned (sections 1.3.1 and 1.3.2), LSD1's interaction with CoREST is required for optimal H3K4 demethylation on nucleosomes during hematopoiesis, development and stem cell maintenance.

1.3.2.3 Structure of LSD1-CoREST

The LSD1 primary interacting partner CoREST consists of an ELM2 (Egl-27 and MTA1 homology 2) domain, two SANT (SWI-SNF, ADA, N-CoR, and TFIIIB) domains (termed SANT1 and SANT2), and a linker region (Figure 3-B). When complexed, LSD1 - CoREST forms an elongated structure of about 150 Å in length with the LSD1 tower domain directly interacting with CoREST through extended coiled coil interactions (Figure 3-C). In the crystal structure of ΔN LSD1(171-852)-CoREST-C (286-482) the inter-SANT linker sequence and SANT2 domain of CoREST interact with LSD1 via hydrophobic and ionic interactions. (48) It is found that C-terminally truncated CoREST (residues 293-482), containing SANT2 domain and the linker between SANT2 domains, binds the TOWER domain and is sufficient to orient LSD1 to demethylate nucleosomal substrates.(48)

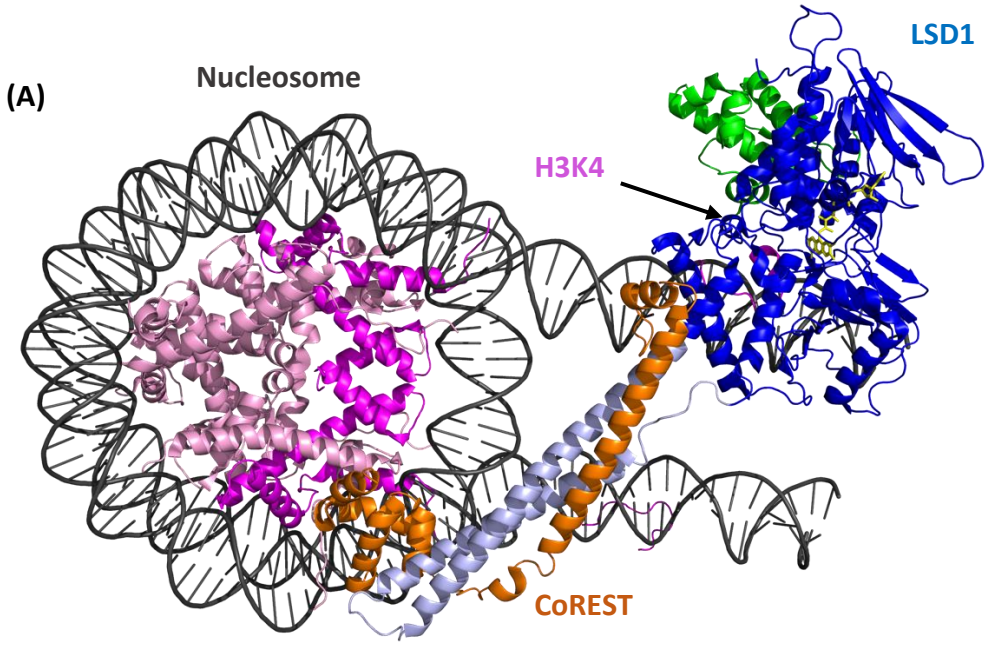
1.3.2.4 Structure of LSD1-CoREST-Nucleosome Complex

LSD1-CoREST binds nucleosome as a 1:1 complex.(29) During the productive engagement of LSD1-CoREST with nucleosome, the complex interacts with DNA around the nucleosomal dyad as well as extranucleosomal DNA one or two turns beyond the end of the nucleosomal core particle. LSD1-CoREST complex binds tighter to nucleosomes with extra nucleosomal DNA and it positively influences the catalytic activity of the complex on nucleosomal substrates.(29) The crystal structure of LSD1(171–852)-CoREST (286–440) complex bound to nucleosome containing 145-bp Widom 601 nucleosome core positioning sequence flanked by 23 bp of extranucleosomal DNA on either side (191 base pair DNA) was published in 2020 (Figure

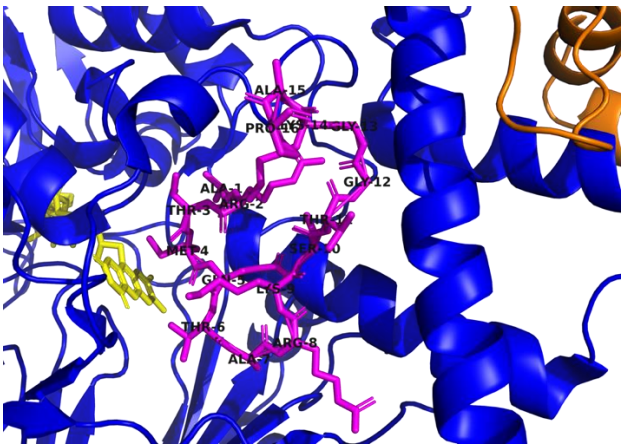
4-A, Binding mode 1).(59) In the structure, the amino oxidase domain of LSD1 binds the phosphate backbone of extranucleosomal DNA positioning 100Å away from the core of the nucleosome while its catalytic site binds H3 tail extending out from the nucleosomal core. In the 5Å crystal structure, about 1 turn away from the nucleosome core, the N terminal region of CoREST linker binds with extranucleosomal DNA while the SANT2 domain of CoREST binds extranucleosomal DNA about 1.5 turns away from nucleosomal dyad and Histone H4 on the octamer surface, providing details of the importance of CoREST in nucleosomal demethylation.(59) Nevertheless, LSD1 also adopts a different binding mode (Binding mode 2) where two molecules of LSD1 bind to one octamer face of the nucleosome and this binding mode is hypothesized to be used for H3K9 or H4K20 demethylation.(59) Interestingly, this mode also coincides with an RNA binding interface of LSD1 that our lab has previously identified. (60)

In addition to the above 2 modes of nucleosome binding, a third LSD1-nucleosome binding mode in the structure of CoREST complex (LSD1-CoREST-HDAC1) bound to nucleosome (185 bp H3K4-propargyl nucleosomes) was discovered in 2020 using Cryo-electron microscopy (Cryo-EM) (Figure 4-C). (44) Within this 26 Å resolution structure, LSD1 is bound to the nucleosome in a way that H3K4-proparagyl group is able to bind the LSD1 active site by covalently tethering to FAD through adduct formation. (53) Since HDAC1 in the ternary complex seems to be farther away from the nucleosome, it may not be able to deacetylate the target H3K9 acetylated tails of the same nucleosome. In this 3D structural model, the positively charged residues of LSD1 are positioned such that they can mediate the electrostatic interactions between LSD1

and negatively charged phosphate backbone of nucleosomal DNA dyad. Interestingly, LSD1's positively charged interaction surface that is ideally positioned to interact with nucleosomal DNA was previously shown to be an RNA binding interface. (60)



(B)



(C)

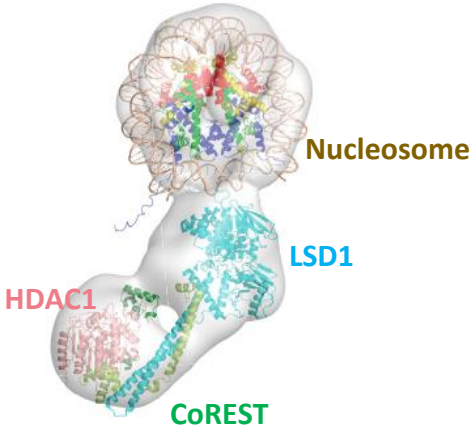


Figure 4: The crystal structure of LSD1-CoREST-Nucleosome complex and the substrate binding site of LSD1.
(A) Cartoon presentation of the crystal structure of LSD1-CoREST-Nucleosome complex at 5 Å resolution. In the nucleosome, histone H3 is in magenta H2A, H2B, H4 are in pink and nucleosomal DNA is in gray. The histone H3 tail bound to LSD1 catalytic pocket is colored in magenta. The LSD1 AOD, TOWER, SWIRM domain and FAD in LSD1 structure are colored in blue, light blue, green, and yellow respectively. The CoREST linker and SANT2 are colored in orange. PDB ID:6VYP **(B)** The 1-16 residue H3K4Met peptide substrate (Red) bound to active site of LSD1 (sky blue). The Methionine residue at position 4 is bound right in front of FAD with the side chain positioned toward the flavin ring of FAD. (PDB ID: 2V1D) **(C)** The crystal structures of nucleosome (PDB:1AOI, H3 in blue, nucleosomal DNA in beige), HDAC1, LSD1, CoREST fitted to the electron microscopic envelop (EMD-10626). LSD1, CoREST, and HDAC1 are in cyan, green, and pink respectively.

1.3.3 LSD1 as a Therapeutic Target

LSD1 plays a crucial role in epigenetic regulation of both normal and disease state transcriptional programs. (61) Although the relationship between histone demethylation and cancer initiation is still not fully understood, recent findings have implicated that reduction of methylation levels that lead to heterochromatin formation are the contributing factors for cancer development and tumorigenesis. (62) While LSD1 plays diverse biological roles in various physiological processes as cell proliferation, development, differentiation, cell cycle arrest, epithelial mesenchymal transition (EMT), chromosomal segregation and cell migration, its aberrant overexpression and dysregulation is closely associated with the initiation and differentiation of many human diseases including cancer and tumors such as acute myeloid leukemia (AML), prostate cancer, breast cancer, lung cancer and neurodegenerative diseases, cardiovascular diseases, inflammation, and viral infections. Nevertheless, it’s found to be involved in

differentiation, proliferation, invasion, metastasis, and poor prognosis of cancer cells. Therefore, since the inhibition and knockout of LSD1 can effectively suppress tumor development, LSD1 is consequently an active molecular target of therapeutic interventions. It has been found that small molecule mediated inactivation and RNAi-mediated downregulation of LSD1 can inhibit cancer development and tumor growth.(43) To date, many types of LSD1 inhibitors including irreversible and reversible inhibitors have been reported. Among them, several inhibitors have advanced to clinical trials as mono or combined therapy for diseases as small cell lung cancer (SCLC), acute myeloid leukemia (AML) and Ewing sarcoma. (43) (63) Targeting the root cause of Ewing sarcoma through LSD1 inhibition will be further discussed in Chapter 4 and 5.

Since LSD1 based complexes are key to transcription regulation in normal and cancer cells, the main focus of my studies were on LSD1's diverse interaction network. Specifically, LSD1's molecular interactions with intrinsically disordered regions (IDRs) and intrinsically disordered proteins (IDPs) (Chapter 2 and 5), different RNA structures (Chapter 3), oncogenic transcription factors (Chapter 5) and small molecules (Chapter 4) were examined to better understand chromatin remodeling enzyme structure, function, and nucleosome engagement.

CHAPTER 2 AUTOREGULATORY MECHANISM OF HISTONE DEMETHYLATION BY A NUCLEAR LOCALIZATION SIGNAL: INTERPLAY OF DISORDER, STRUCTURE, AND NUCLEOSOME BINDING WITHIN LSD1

2.1 Introduction

Gene sequences in the eukaryotic genome encode entire unstructured proteins or segments of proteins that lack well-structured regions. The LSD1 enzyme contains both highly structured regions and intrinsically disordered regions allowing it to interact with over 60 regulatory proteins and transcription factors (including the tumor suppressor p53 and key protein interaction partner CoREST), essential nutrients such as tetrahydrofolate (THF), and numerous nucleic acid structures including telomeric repeat containing RNA (TERRA), nucleosomes, cancer associated long non coding RNAs, pre messenger RNAs and R-loops (DNA:RNA hybrids). One such intrinsically disordered protein segment (IDR) in LSD1 structure, the N terminal portion of LSD1 (aa 1-170) is a very flexible region, thus its structure has not been solved yet even though the structure of ΔN LSD1-CoREST with H3 peptide substrate (1-21 amino acids) and with nucleosome substrate has been previously determined. The LSD1 N terminus is dispensable for catalysis with methylated H3K4me^2 peptide substrate in vitro and is predicted to be disordered with linear motifs that may play functional roles in association of LSD1 with a variety of other transcriptional protein complexes. (54) (64) Although the N terminus is a disordered region, it is predicted that the N terminal flexible modules may fold upon binding to their biological targets. Mutations within the N terminal flexible region have

been found in different cancers including several types of carcinomas, suggesting the importance of studying interactions and dynamics within this LSD1 N terminal IDR. (65)

The composition and sequence of disordered regions in chromatin modifying proteins can be highly conserved between species and often contain functional importance. (66) Similarly, the 1-170 amino acids (a.a) long N terminal disordered region of LSD1 contains an amino acid sequence from 103 – 151, conserved in nearly all vertebrates (Figure 5-B). Within the conserved region of LSD1 IDR, contains a putative nuclear localization signal (NLS) sequence (a.a. 112-117 RRARAK) and it is found to be important for the internalization of LSD1 into the nucleus (Figure 5-C). (49) (67) Bioinformatic studies reveal conserved electrostatic clusters within this conserved region of LSD1 that resides near the NLS sequence.

The flexible N terminus of LSD1 itself including the conserved region, that resides atop the catalytic active site, is subjected to several functionally relevant post translational modifications by various other enzymes suggesting similarities with the heavily modified, electropositive human H3 tail (Figure 5-B). The LSD1 N terminus is rich in Serine (Ser / S) and Threonine (Thr / T) residues suggesting a modulatory activity of LSD1 function by phosphorylation of those residues. Among the several other PTMs in the N terminus, NLS region contains three functionally important post translational modifications (T110, S111 phosphorylation and Lysine114me^2/K114 methylation) involved in androgen dependent transcription regulation.(68) Protein kinase A (PKA) phosphorylates threonine 110 and serine 111 while serine 111 can also be phosphorylated by protein kinase C (PKC). Threonine 110 and serine 111

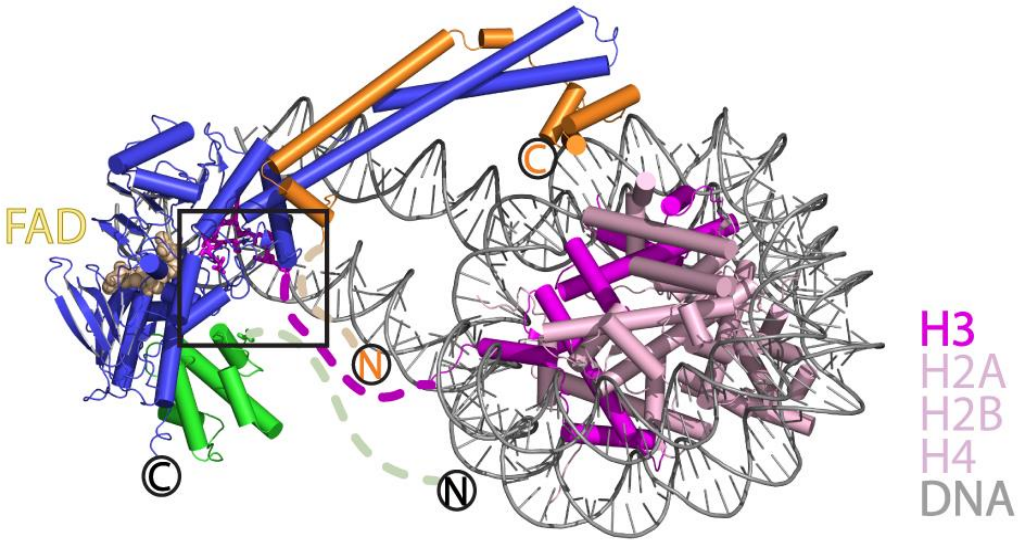
phosphorylation PTMs are required for the recruitment of HDAC and CoREST complexes to estrogen responsive genes as well as ser 111 phosphorylation is required for the activation of inflammatory response and for LSD1 induced transcriptional activity that helps in the induction of EMT and breast cancer cell metastasis. Euchromatic histone-lysine N-methyltransferase 2 (EHMT2) methylates lysine 114 of LSD1 that plays a role in favoring transcription. CHD1 reads the K114me^2 mark on LSD1 as a part of the androgen transcriptional program. (68) Furthermore, polo-like kinase 1 (PLK1) adds a phosphate group on ser 126 in the N terminal tail that releases LSD1 from chromatin during mitosis and preserves the balance between methylation and demethylation in chromatin during the cell cycle. LSD1 phosphorylation at ser 131 and ser 137 by casein kinase 2 (CK2α) is important for its direct recruitment to DNA damage sites that favor the activation of DNA damage repair. (65) Taken together, the presence of this variety of PTMs identify an extensive protein-protein interaction hub that is clustered within the conserved region of the N terminus. Aside from the post translational modifications, mass spectroscopic analysis of the LSD1-CoREST-HDAC ternary complex identified that LSD1 a.a. 117, 144 cross links with HDAC1 (a.a. 89, 123) and CoREST (a.a. 298, 300, 356). Taken together, the N terminal disordered region of LSD1, which lacks a stable 3D structure, is predicted to provide flexibility to LSD1 in targeting different chromatin proteins and acts as an interaction hub within residues 110-145.

In the recently published ΔN LSD1-CoREST- nucleosome structure, there are three intrinsically disordered regions (IDRs) including the N terminus of LSD1 (1-170), the linker region of CoREST (280-308) and the H3 tail (15-38) that extend out and cluster

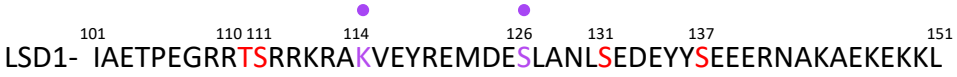
in close proximity to the active site cavity of LSD1 between 7-25 Å from each other and the catalytic pocket (Figure 5-A). (59) This suggests that these flexible regions may play regulatory roles in positioning substrate binding to the active site cavity, nucleosome docking, and fine-tuning the structural mechanism of activation.

The placement of the N terminus near the active site cavity ranging 10-25 Å between the active site cavity and the N terminus of SWIRM domain (a.a. 171), and the extent of protein-protein interactions within the LSD1 N terminus suggests the presence of tunable, dynamic elements within the nucleosome-demethylase activity interface of LSD1. It was hypothesized that this flexible LSD1 N terminal region influences the demethylation in an analogous way to that of the critical transcription factors such as SNAIL1, INSM1, gfi1 that possess patterns of positively charged residues that mimic the H3 substrate tail to inhibit the demethylation reaction of LSD1. Given the unique features of the N terminus and its close proximity to the active site cavity, in this chapter I sought to investigate the role of the N terminal tail within the context of the LSD1-CoREST structure and nucleosome binding. I describe the binding properties, structure and autoregulatory properties of the N terminal conserved region of LSD1 and address the roles of the putative non-catalytic region of LSD1 in terms of nucleosome binding, H3K4me^2 demethylation and molecular mechanisms of the dynamic LSD1 interaction network.

(A)



(B)



(C)

	103		151
Homo sapiens-	ETPEGRRTS**RRKRAK**	VEYREMDESLANLSEDEYYSEEERNAKAEKEKKL	
Mus musculus-	ETPEGRRTS**RRKRAK**	VEYREMDESLANLSEDEYYSEEERNAKAEKEKKL	
Bos Taurus-	ETPEGRRTS**RRKRAK**	VEYREMDESLANLSEDEYYSEEERNAKAEKEKKL	
Gallus gallus-	ETPEGRRTS**RRKRAK**	VEYREMDESLANLSEDEYYSEEERNAKAEKEKKL	
Xenopus laevis-	ETPEGRRTS**RRKRAK**	VEYREMDESLANLSEDEYYSEEERNAKAEKEKKV	
Danio rerio-	DSAEGRRTS**RRKRAK**	VEYREMDESLANLSEDEYYSEEERNAKAEKEKKQ	

Nuclear Localization Signal (NLS)

Figure 5: Structure of ΔN LSD1-CoREST-Nucleosome complex indicating the locations of N terminal tails of LSD1 (green), Histone H3 (magenta), and CoREST C-terminal tail (orange).

(A) Illustration of amino terminal tails of LSD1 and CoREST as well as C terminal tail of LSD1 protruding in near vicinity of the active site of ΔN LSD1-CoREST bound to nucleosome structure PDB ID: 6VYP. In the figure, LSD1 is in blue and green colors, and CoREST is in orange color. (59) **(B)** Post-translational modifications at amino terminal tail 101-151 amino acids of human LSD1. Purple dotted residues represent methylation

sites, red color represents phosphorylation sites. (67) **(C)** The conserved sequence of amino acids within the flexible N terminus of LSD1. Nuclear localization signal (NLS) sequence within the conserved region is bolded in the sequences.

Interestingly my results demonstrate that LSD1's own NLS region can function as a reversible, competitive inhibitor of LSD1's demethylation activity, acting in an autoregulatory manner. I show that, within the IDR conserved stretch (a.a. 103-151) of amino acids in the LSD1 N terminus, there is a transient α-helical structure (a.a. 135-151), and this conserved IDR region inhibits the demethylation on the H3K4me^2 nucleosome substrate. Further, my studies show that this autoinhibition can be relieved by phosphomimetic-specific substitutions, suggesting a synergistic and versatile role of IDRs and PTMs. This work expands the regulatory roles of disordered regions in the immediate vicinity of LSD1 active site cavity and provides insight into the development of PTM based peptidomimetic inhibitors that may target LSD1. Results from this chapter propose a new way for how an NLS region within a chromatin modifying protein can modulate the function of the enzyme itself.

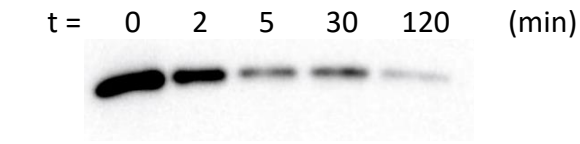
2.2 Results

2.2.1 ΔN LSD1/CoREST is More Active Than FL LSD1 -CoREST

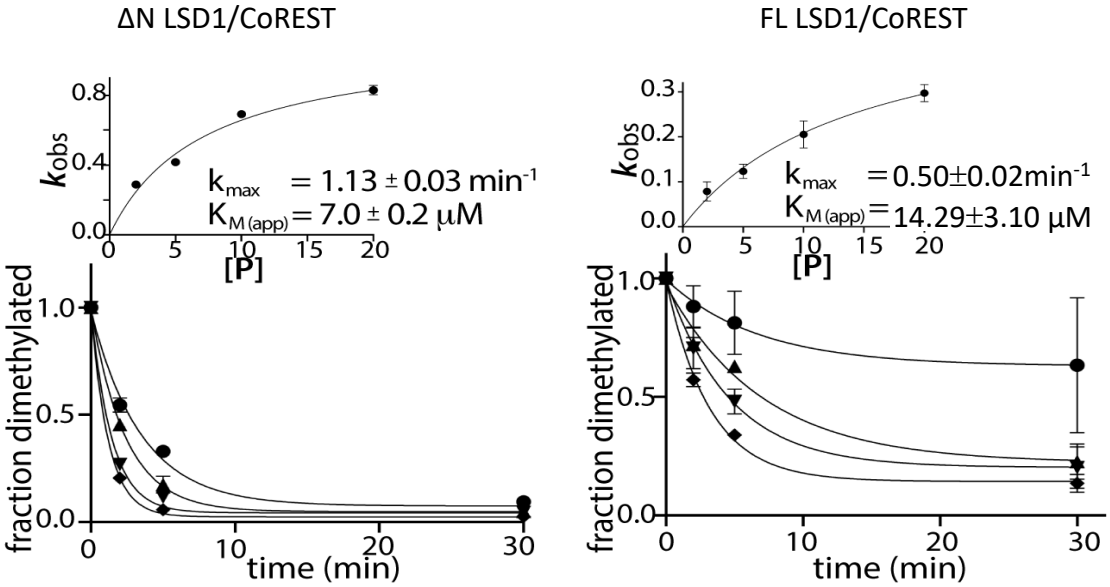
Even though activity assays to measure enzyme activities of FL LSD1 and ΔN LSD1 has been performed previously by different groups, a full enzyme kinetics study comparing activities of truncated ΔN LSD1(171-852) -CoREST (amino acids (a.a.) 286-482) and full length (FL) LSD1 (1-852)-CoREST (286-482) in the presence of nucleosomal substrate had not been performed before. Therefore, after simultaneously co-purifying truncated and full length LSD1 in complex with CoREST, I sought to demonstrate the enzyme activities of ΔN LSD1-CoREST and FL LSD1-CoREST in presence of H3K4me^2 mononucleosomal substrates in order to identify potential differences between FL and truncated LSD1 complexes. The kinetics of demethylation reactions were measured using single turnover conditions measured by quantitative western blots using anti-H3K4me^2 antibodies in the presence of 100nM H3K4me^2 mononucleosomes. The time course of LSD1 catalyzed demethylation was used to measure residual dimethylated H3K4me^2 at each timepoint, which was normalized using corresponding H3 antibody signal monitored by anti H3 antibody. (69)(70)(71) According to the demethylation reactions, ΔN LSD1-CoREST has a 1.13 min^{-1} maximum rate of catalysis (k_{max}) and a $K_{1/2}$ of 7.14 μM. In comparison, FL LSD1-CoREST has a k_{max} of 0.5072 min^{-1} and $K_{1/2}$ of 14.29 μM which is two times slower rate for both catalysis and substrate affinity (Figure 6). When compared the catalytic efficiency ($k_{max}/K_{1/2}$), presence of N terminal tail sequence decreases the catalytic efficiency 4.5 times than in N terminally truncated LSD1-CoREST.

These kinetic measurements are comparable with the other studies that used similar western blot approaches. (48)(59)(71)

(A)



(B)



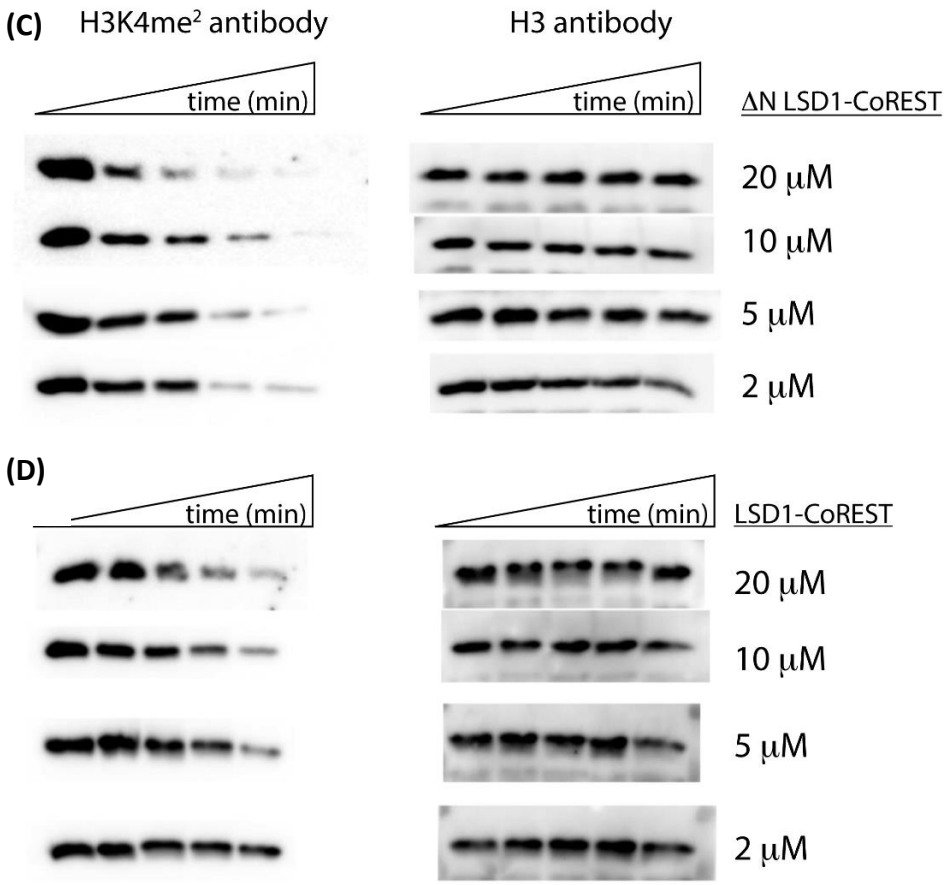


Figure 6: Nucleosome demethylation by ΔN LSD1/CoREST and FL LSD1/CoREST. **(A)** Representative image of time course of LSD1 catalyzed demethylation of H3K4me^2 nucleosomes followed with anti H3K4me^2 antibody. **(B)** Fraction dimethylated versus time graph plotted by quantitation of residual dimethylation from western blots ($n \geq 2$). The kinetic parameters of LSD1 catalytic activity determined by the LSD1 protein concentration ([P]) versus K_{obs} graph ($n \geq 2$). Circle, triangle, inverted triangle, and diamond shapes in the fraction dimethylated vs time graphs represent ΔN LSD1- CoREST or FL LSD1- CoREST enzyme concentrations of 2, 5, 10, 20 μM respectively. **(C) (D)** Representative western blots for activity assay of H3K4me^2 mononucleosome demethylation at varying concentrations of **(C)** ΔN LSD1-CoREST and **(D)** full-length LSD1-CoREST, and increasing time points (0, 2, 5, 30, 120 minutes).

2.2.2 N-Terminal IDR Peptide Bind ΔN LSD1-CoREST Complex

Given the functional complexity of the N terminus and since now it's known that LSD1 N terminus has an influence on its enzymatic activity, I focused on exploring whether the conserved N terminus contains any structure, and if it interacts with the nucleosome or perhaps its own LSD1 interface. First, surface plasmon resonance (SPR) studies were performed to investigate if there is any binding propensity between the ΔN LSD1-CoREST and the N terminus that accounts for the impeded LSD1 catalytic activity. According to the SPR binding curve (Figure 7-A), ΔN LSD1-CoREST binds the immobilized N terminal peptide consisting of 100-151 amino acids, with a dissociation constant (K_d) of 571 nM, corresponding to tight binding affinity between the conserved N terminal region and LSD1. (72) In addition, the CD spectroscopic data, backbone assignments from NMR, and x-ray diffraction data (PDB ID: 6WC6, see appendix) revealed formation of an α helical structural element (aa 135-151) that binds ΔN LSD1-CoREST amino oxidase domain. Despite this binding data and structural information, the functional relevance of this helical region of N terminal IDR binding to LSD1 structure has yet to be determined.

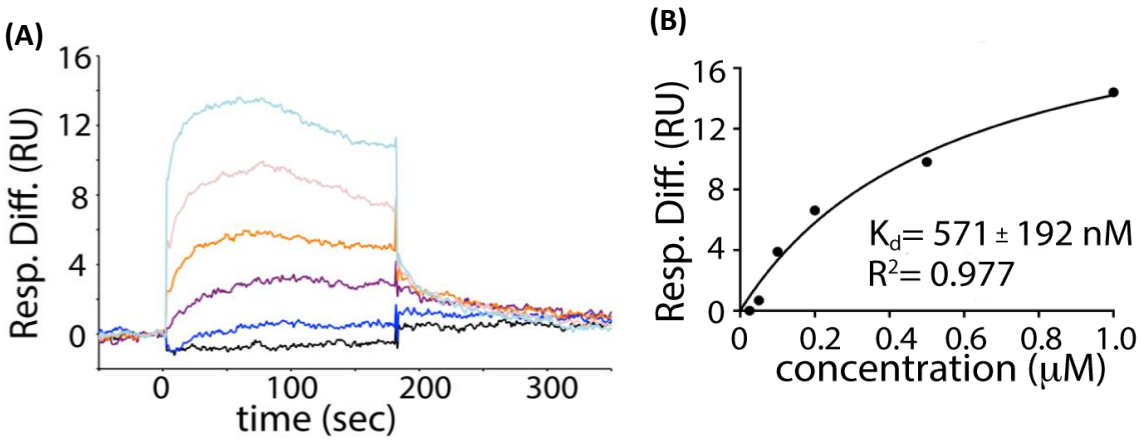


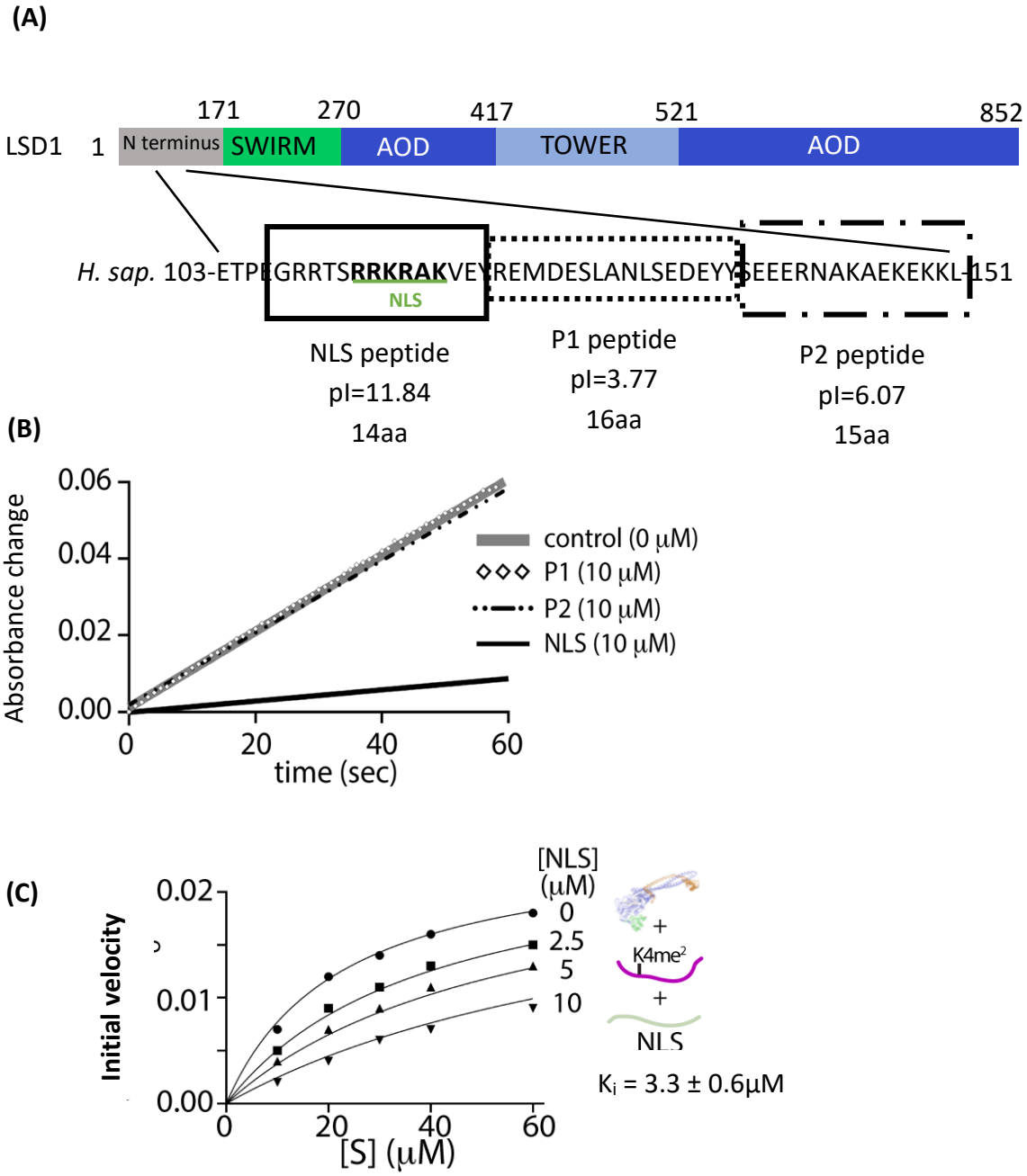
Figure 7: Intrinsically disordered N terminus of LSD1 binds ΔN LSD1-CoREST.
The surface plasmon data (A) and analysis (B) of N terminus (a.a. 100-151) binding to ΔN LSD1-CoREST. (A) Representative sensorgram data show the differential response units (RU) over time, measuring how the LSD1 complex interacts with N terminal conserved sequence peptides that are immobilized on a sensor chip. Immobilized peptides were subject to 5 nM (black), 25 nM (blue), 50 nM (purple), 100 nM (orange), 200 nM (light pink), and 300 nM (cyan), protein concentrations. (B) The plot of the difference in response units (RU) versus LSD1 protein concentration reveals the binding dissociation constants *(K_d)* for the conserved N-terminus of LSD1.

2.2.3 NLS Region of LSD1 IDR Autoinhibits LSD1 Activity in Presence of Peptide Substrate

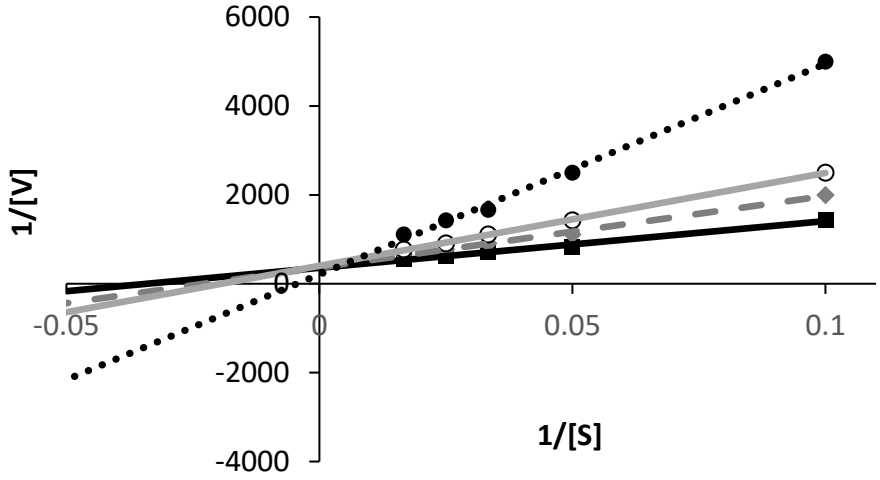
As the next step to determine the exact regions of N terminal sequence that bind and obstruct LSD1 catalyzed demethylation, three consecutive short peptides from the highly conserved region (a.a. 103-151) of LSD1 N terminus were prepared. The peptides are namely, NLS peptide (positively charged, aa 107-120, pI ~11.84) that contains the nuclear localization signal sequence, P1 peptide (negatively charged, a.a. 121-136, pI ~3.77) and P2 peptide (neutral, a.a. 137-151, pI ~6.07) (Figure 8-A).

The inhibition study was performed using HRP coupled demethylase assay in the presence of NLS, P1 and P2 peptides, using the 21 a.a H3K4me^2 peptide as substrate. (54) (45) Of the three peptides, only NLS peptide acted as an inhibitor of ΔN LSD1-

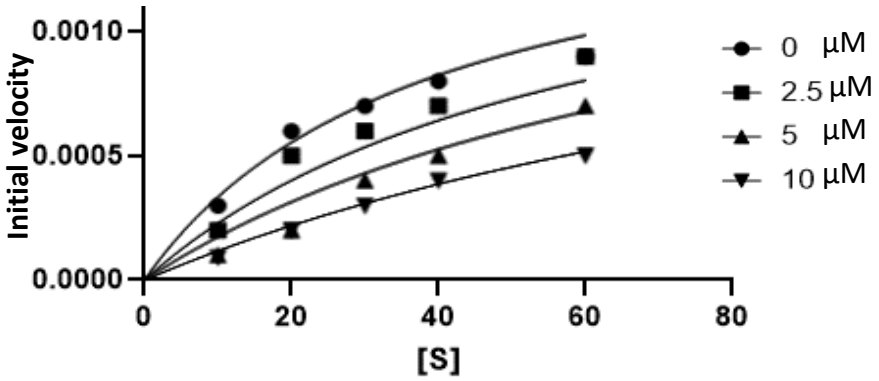
CoREST. The P1 and P2 peptides did not influence the demethylation reaction (Figure 8-B). When full inhibition study was performed using different concentrations of 21 a.a. H3K4me^2 peptide as substrate and different concentrations of NLS peptide inhibitor, an inhibition constant (K_i) of 3.3 ± 0.6μM was determined. This suggests that the NLS peptide is an inhibitor of LSD1 catalytic activity with comparable inhibition to other known peptide fragments whose proteins are known to directly interact with LSD1. (Figure 8-C and Table 1). The NLS peptides competitive mode of inhibition suggests direct binding to the ΔN LSD1-CoREST. When performing CD spectroscopy, this inhibitory NLS peptide (aa 107-120) showed a strong negative band near 200 nm of the spectrum indicating a lack of well identified secondary structural features (Figure 8 F). (73)


(A)
171
270
417
521
852
LSD1 1
N terminus
SWIRM
AOD
TOWER
AOD
H. sap. 103-ETPEGRRTSRRKRAKVEYREMDESLANLSEDEYYSEEERNAKAEKEKKL-151
NLS
NLS peptide
pI=11.84
14aa
P1 peptide
pI=3.77
16aa
P2 peptide
pI=6.07
15aa
(B)
Absorbance change
0.06
0.04
0.02
0.00
0
20
40
60
time (sec)
control (0 μM)
P1 (10 μM)
P2 (10 μM)
NLS (10 μM)
(C)
Initial velocity
0.02
0.01
0.00
0
20
40
60
[S] (μM)
[NLS] (μM)
0
2.5
5
10
K4me2
NLS
Ki = 3.3 ± 0.6μM

(D)


6000
4000
2000
0
-2000
-4000
1/[V]
-0.05
0
0.05
0.1
1/[S]

(E)


0.0010
0.0005
0.0000
Initial velocity
0
20
40
60
80
[S]
0 µM
2.5 µM
5 µM
10 µM

(F)

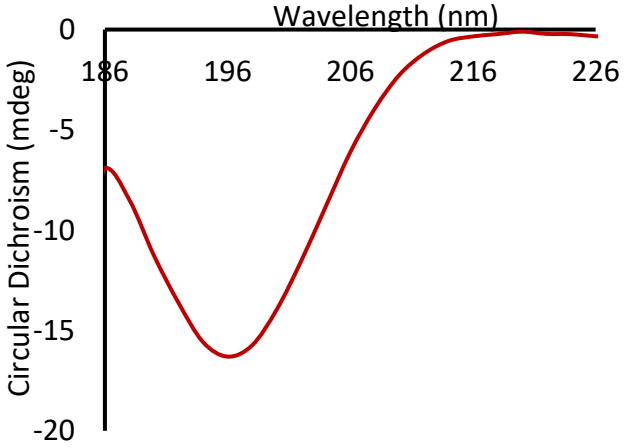

Wavelength (nm)
0
-5
-10
-15
-20
Circular Dichroism (mdeg)
186
196
206
216
226

Figure 8: The conserved region of LSD1 N terminus contains NLS sequence that inhibits LSD1 activity.
(A) Domain organization of LSD1 and zoomed view of conserved region of human LSD1 N terminus. Sequences of NLS, P1, and P2 peptides are illustrated in boxes with the corresponding isoelectric (PI) values. **(B)** Plot of absorbance change versus time graph using the HRP coupled assay. NLS peptide inhibits LSD1 activity on model peptide substrates while P1 and P2 peptides do not inhibit LSD1's activity. **(C)** The initial velocity versus H3K4me^2 21 a.a. peptide substrate concentration graph obtained by the HRP coupled assay in the presence of different NLS peptide concentrations shows that the NLS peptide acts as an inhibitor of LSD1 catalytic activity (Trial 1). **(D)** Lineweaver Burke plot and analysis of LSD1 catalyzed demethylation on a model peptide substrate. Increasing NLS peptide concentrations (0 (black-square), 2.5 μM (grey dash-diamonds), 5 μM (grey-open circle), and 10 μM (black-dotted, solid circle) indicate competitive inhibition. Lineweaver Burk plot of initial velocity measurements of LSD1-catalyzed demethylation on an H3K4me^2 peptide substrates at 0 μM (squares), 2.5 μM (grey diamonds), 5 μM (open circles), and 10 μM (solid circles) NLS peptide inhibitor. Linear regression analysis reveals that all traces intersect the y-axis, indicative of competitive inhibition kinetics. **(E)** The initial velocity versus H3K4me^2 21 aa peptide substrate concentration graph obtained by the HRP coupled assay in the presence of different NLS peptide concentrations shows that NLS peptide acts as an inhibitor of LSD1 catalytic activity with a K_i = 4.3 ± 0.9 μM (Trial 2). **(F)** CD spectrum of disordered NLS peptide (aa 107-120). CD spectrum of 5 μM NLS peptide in water recorded at room temprature exhibiting characteristics of disordered proteins with an intensive minimum in viscinity of 200 nm.

Table 1: Comparative (K_i ^{app}) data of peptide sequences that inhibit LSD1-catalyzed demethylation on H3K4me^2 model (peptide) substrate.

No	Peptide	Sequence	approximate K_i/μM
1	NLS (14 residues)	GRRTSRRKRAKVEY	3.3
2	NLSp*	GRRDDRRKRAKVEY	38.7
3	NLS-K114me	GRRDDRRKmeRAKVEY	-, ND[&]
4	[a] SNAIL (1-20)	PRSFLVRKPSDPNRKPNYSE	0.21
5	[a] INSM1 (1–20)	PRGFLVKRSKKSTPVSYRVR	0.24
6	[b] H3 (1-21)	ARTKQTARKSTGGKAPRKQLA	1.8
7	[b] H3 (1–12)	ARTKQTARKSTG	199
8	[d] H3 (1-21) K4M	ARTMQTARKSTGGKAPRKQLA	0.05
9	[c] H3(1-21)K4meR8me	ARTKmeQTARmeKSTGGKAPRKQLA	100
10	[c] H3(1-21)K4me3	ARTKme3QTARKSTGGKAPRKQLA	19.5
11	[c] H3(1-21) K4R	ARTRQTARKSTGGKAPRKQLA	0.41
12	[b] H3(5-21)	QTARKSTGGKAPRKQLA	87
13	[b] H3(1-21) K4me S10p	ARTKmeQTARKSpTGGKAPRKQLA	31
14	[e] H3(1-21) K4^{propargylamine}	ARTK^{prop}QTARKSTGGKAPRKQLA	16.6

All peptides, except (14), contain C-terminal amides. Unless denoted (&), all peptides exhibit competitive inhibition.

[a] Data from reference (77)
[b] Data from reference (54)
[c] Data from reference (81)
[d] Data from reference (51)
[e] Data from reference (82)

2.2.4 Autoinhibition of ΔN LSD1-CoREST by the NLS region in presence of H3K4me^2 nucleosome substrate

Once the binding and inhibition of NLS peptide with ΔN LSD1-CoREST was confirmed, I next tested the ability of NLS peptide to inhibit the activity of ΔN LSD1-CoREST in presence of its actual substrate- the H3K4me^2 nucleosome, a large poly anionic molecule about 100x larger than the model peptide substrate. For this, single turnover assays were conducted with 100 nM nucleosomal substrate and 2 μM ΔN

LSD1-CoREST in the reaction mixture. The observed rate constants (k_{obs}) of the demethylation reactions were determined by analyzing western blots with anti-H3K4me^2 and anti-H3 antibodies and quantitating the relative fraction dimethylated over time (Figure 9-A). The half maximal inhibitory concentration (IC_{50}) of NLS inhibition of demethylation reaction in the presence of nucleosomal substrate was determined (30.4 ± 3.2 μM) (Figure 9-B). (74)(75) It is speculated that the nucleosome, which contains both anionic (DNA) and electropositive (histone) features, can bind and neutralize the electropositive NLS peptide, reducing its ability to act as an inhibitor on demethylation reaction. This would help to explain why the IC_{50} value is an order of magnitude lower when compared to the K_i value in presence of short peptide substrate. Furthermore, the well-established relationship between K_i and IC_{50} for competitive inhibition mode (K_i=IC_{50}/2) suggests a calculated K_i value of 15.2 μM for NLS inhibition of demethylation reaction in the presence of nucleosomal substrate using western blot assay.

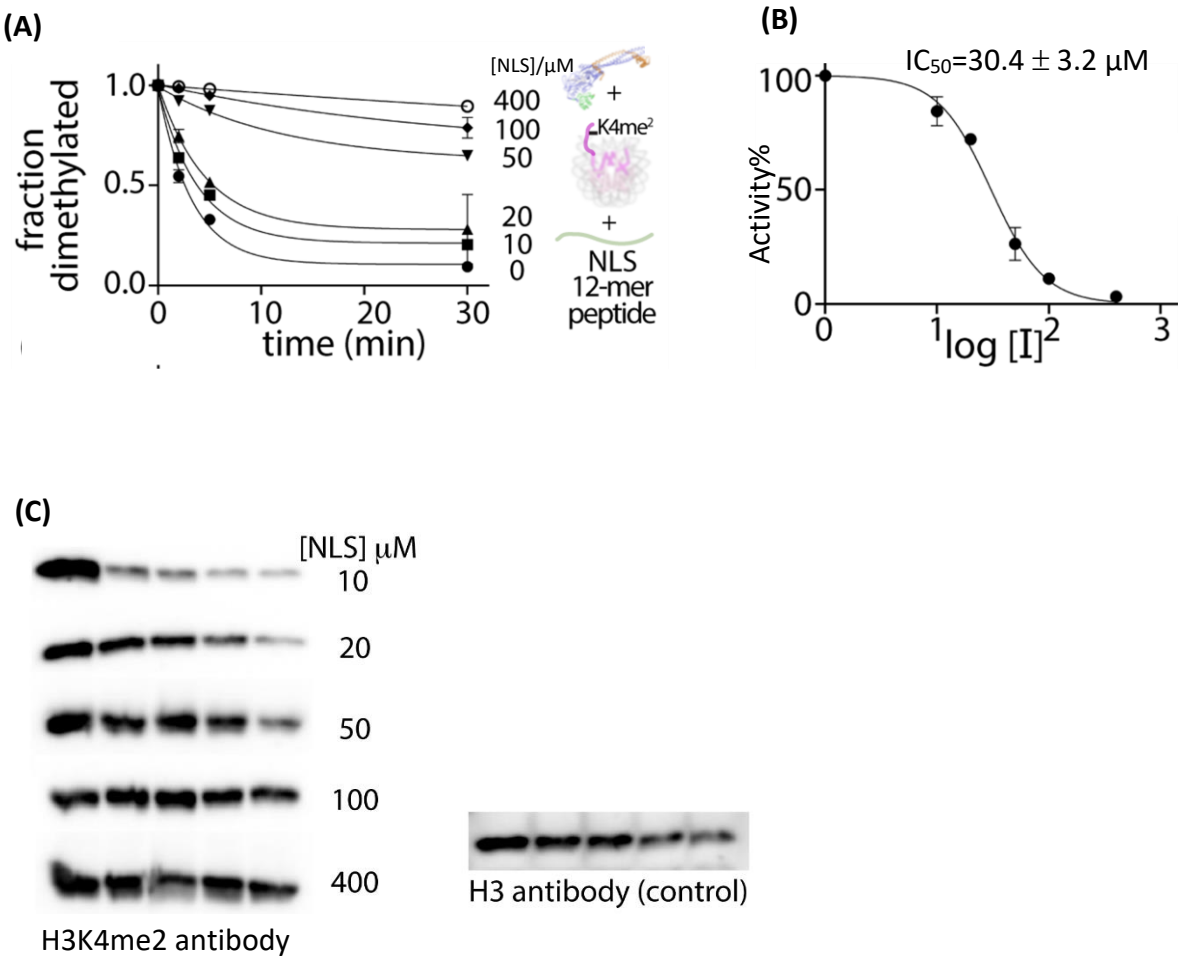


Figure 9: LSD1 catalytic activity on nucleosomes is inhibited by NLS peptide.
(A) Quantification of residual dimethylation on western blots at each timepoint. n≥2 **(B)** Activity % versus log inhibitor concentration ([I]) graph of NLS peptide inhibition of LSD1 activity in presence of H3K4me^2 nucleosomal substrate.

2.2.5 NLS Peptide Bind to ΔN LSD1-CoREST

In order to determine the binding between ΔN LSD1-CoREST and NLS peptide containing nuclear localization signal sequence, isothermal titration calorimetry (ITC) was performed. From ITC experiments, exothermic binding was demonstrated for a series of small aliquots NLS peptide injected to the ΔN LSD1-CoREST protein solution.

When the area of each peak was integrated and plotted against the molar ratio of NLS: ΔN LSD1-CoREST, a binding constant (K_d) of 15.4 ± 4.3 μM was determined from the binding isotherm data. This value is consistent with the tight nanomolar range affinity of N terminal conserved region (100-151) binding to ΔN LSD1-CoREST (Table 1).

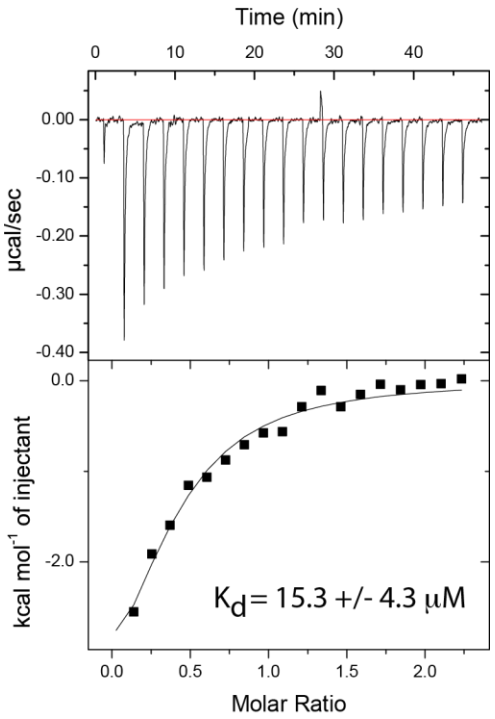


Figure 10: Isothermal titration calorimetry of NLS peptide (107-120) binding to ΔN LSD1-CoREST.

NLS peptide injection-induced deflection and decay patterns into ΔN LSD1 (171-852)-CoREST (286-482) (top). Determination of the dissociation constant (K_d = 15.4 ± 4.3 μM) based upon the exothermic binding isotherm (bottom).

2.2.6 Impact of Post Translational Modifications Near the NLS Sequence on LSD1 Catalyzed Demethylation

LSD1 can be heavily modified through various PTMs including phosphorylation, acetylation and ubiquitination where most of the PTMs are clustered at the N terminal IDR of LSD1. I used modified NLS peptides that contain PTMs at K114, T110 and S111 that are close to and in the nuclear localization region to determine how those PTMs affect the LSD1 catalyzed demethylation.

2.2.6.1 K114 Methylation on NLS Peptide Doesn't Impact the Repressive Role of NLS Peptide on LSD1 Catalyzed Demethylation

Among the three functionally important PTMs in or near the NLS sequence, K114 methylation by the protein EHMT suggests that there is the possibility of LSD1 being able to demethylate its own N terminus K114 dimethylation mark. This was previously examined, where protein constructs of a GST tagged LSD1 (GST-LSD1) or GST-LSD1 K114A were purified and evaluated enzymatically, revealing that the K114A modification had no competing impact on demethylation of H3K4me^2 peptide. (68) To determine if LSD1 could potentially demethylate its own K114 methylation mark, a K114 dimethylated NLS peptide (GRRTS**RRK^{me2}RAK**VY/ NLS K114me^2) was purchased and prepared and the demethylation assay using untagged ΔN LSD1-CoREST was performed. Consistent with the previous findings, when NLS K114me^2 was used in HRP coupled assay with the H3K4me^2 21 a.a. peptide substrate and purified ΔN LSD1-CoREST, the NLS K114me^2 acted with near identical levels of inhibition to that of the NLS peptide (Figure 11-A). Further, when the HRP coupled assay was performed using only NLS K114me^2 as

a potential substrate, I could not see any demethylation activity (Figure 11-A). These experiments indicate that the K114 dimethylation PTM on the N terminal IDR is not an active substrate of LSD1 catalyzed demethylation, and that it does not influence the NLS mediated auto-inhibition of LSD1 demethylase activity.

2.2.6.2 T110, S111 Phosphorylation on NLS Plays De-repressive Role in LSD1 Catalyzed Demethylation on Peptide Substrate

In order to see the impact of the two fuctionally important phosphorylation post translational modifications near the NLS at T110, S111 on the NLS mediated autoinhibition of LSD1 catalytic activity, the NLS peptide was modified to contain phosphomimetic mutations T110D and S111D (GRRDD**RRKRAK**VY/ NLSp*) and I performed demethylase peroxidase assay using model peptide substrate. The complete inhibition profile of NLSp* peptide suggests that introducing two electronegative phosphates at the specific location near the NLS sequence partially relieves inhibition by NLS playing a derepressive role in LSD1 catalyzed demethylation on nucleosomes (K_i ~ 38.7 ± 1.8μM, 12x less than the inhibition of NLS peptide) (Figure 11-B). It further confirms how presence of PTMs near NLS region can substantially influence LSD1 catalyzed demethylation reaction.

(A)

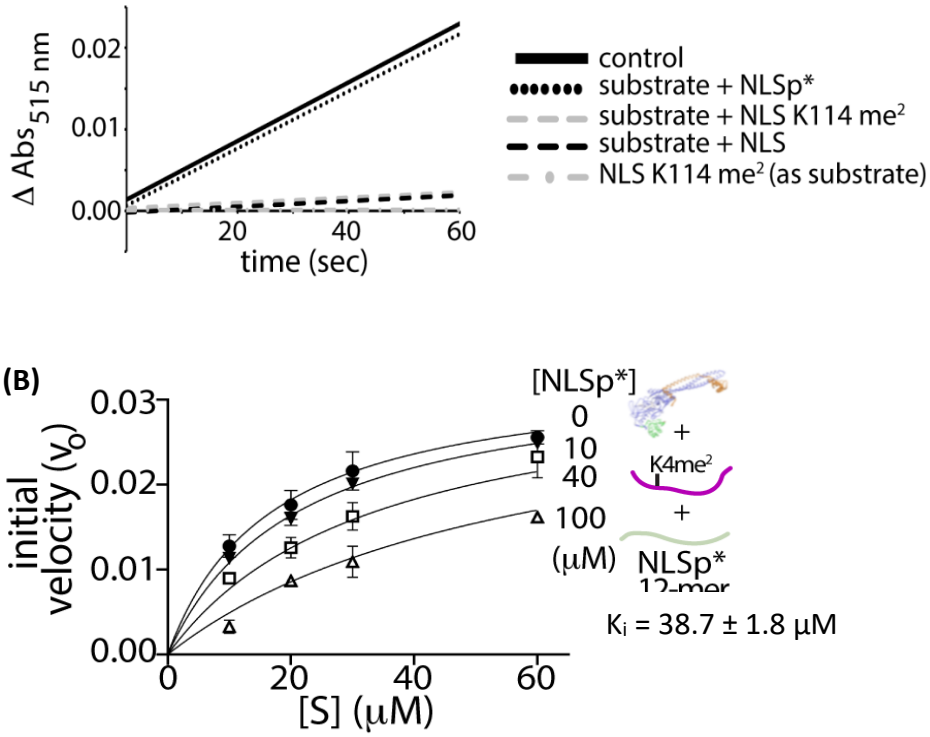


Figure 11: Phosphorylation PTM near the NLS sequence partially relieves the inhibition by NLS.

(A) The absorbance versus time graph of NLS peptides with or without different PTMs that affect and does not affect catalytic activity of LSD1. NLS peptide inhibit LSD1 activity on peptide substrate while NLSp* shows near identical absorbance with the control. NLSK114me^2 peptide inhibits LSD1 demethylation reaction similar to the NLS peptide. K114 di methylated NLS peptide does not act as a substrate of LSD1. **(B)** Inhibition assay of LSD1 performed using HRP coupled assay in presence of increasing amounts of NLSp* that partially relieves the inhibition by NLS.

2.2.8 Phosphorylated NLS Partially Relieves Inhibition on Nucleosomes

In confirming that NLSp* partially relieves the inhibition by NLS, I next tested the ability of FL T110D, S111D mutant LSD1-CoREST (FL DD mutant LSD1) to act upon H3K4me^2 nucleosomal substrate and catalyze the demethylation reaction to get an insight of how phosphorylation PTMs near the NLS site affect catalytic activity of LSD1. I

performed site directed mutagenesis on FL LSD1 plasmid to introduce phosphomimetic DD mutation at T110 and S111 positions near the NLS sequence. The kinetics of demethylation reaction of mutant FL LSD1-CoREST was measured using single turnover conditions measured by quantitative western blots using anti-H3K4me^2 antibodies in presence of 100nM H3K4me^2 mononucleosomes. The time course of LSD1 catalyzed demethylation was used to measure residual dimethylated H3K4me^2 at each timepoint, which was normalized using corresponding H3 antibody signal monitored by anti H3 antibody.

According to the demethylation reactions FL DD mutant LSD1/CoREST has a 0.97 min^{-1} maximum rate of catalysis (k_{max}) and a substrate binding constant ($K_{1/2}$) of 8.8 µM (Figure 12). When compared with the catalytic efficiency ($k_{max}/K_{1/2}$) values of ΔN LSD1-CoREST and FL LSD1- CoREST (Table 2) it suggests that introduction of phosphomimetic mutation near the NLS sequence of FL LSD1 partially relieves inhibition of LSD1 catalysed demethylation by N terminal NLS sequence in FL LSD1.

Table 2: Kinetic constants of different LSD1 complexes with H3K4me^2 nucleosome substrates, determined by quantitative western blots.

LSD1 complex	k_{max} (min^{-1})	$K_{1/2}$ (μM)	$k_{max}/K_{M(app)}$ ($\mu M^{-1}min^{-1}$)
ΔN LSD1-CoREST	1.13	7.14	0.1583
FL LSD1- CoREST	0.5072	14.29	0.0355
FL DD mutant LSD1-CoREST	0.97	8.8	0.1102

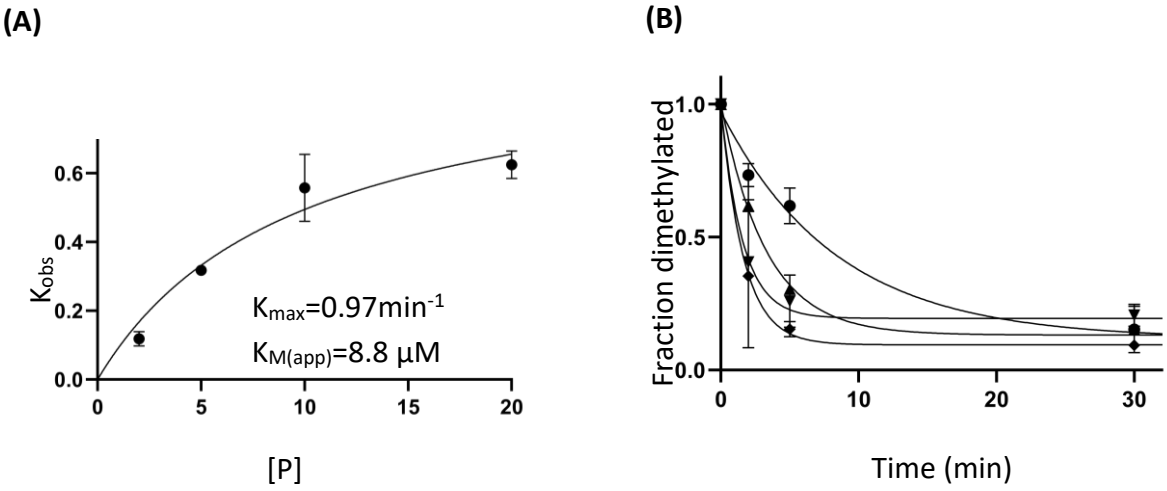


Figure 12: Nucleosome demethylation by FL DD mutant LSD1/CoREST suggests phosphorylation of T110/S111 relieves autoinhibition.

(A) The kinetic parameters of LSD1 catalytic activity determined by the protein concentration versus k_{obs} graph ($n \geq 2$). **(B)** Fraction dimethylated versus time graph plotted by quantitation of residual dimethylation from western blots ($n \geq 2$).

2.2.9 Model for Competitive Binding of the NLS Peptide to the LSD1 Active Site

Next, brief rounds of local Rosetta docking of 12 mer NLS peptide near the LSD1 active site were used to test the potential binding mode of the peptide's interaction with LSD1. Due to NLS peptide's competitive inhibition of LSD1 activity on model substrate, I propose that NLS peptide binds to the active site cavity of LSD1. The presence of conserved electrostatic clusters and functionally relevant PTMs near the NLS in the LSD1 N terminus conserved region suggest similarities with highly modified and electropositive human H3 tail (the isoelectric point/pI of 21 a.a H3 N terminal tail peptide is 12.31). Supporting this, initial randomization and perturbations of the NLS peptide prior to performing the local refinement developed a converged model and preliminary docked structure similar to the H3 histone tail orientation within the active site (Figure 13). Since both NLS peptide and H3 peptide are electropositive, and that H3 peptide substrate electrostatically recognizes and binds negatively charged active site, a side-by-side computational comparison of H3 peptide bound active site and NLS bound active site was performed. The docked models reveal that, the NLS region is accommodated within the active site and that the location of the NLS S111 residue closely overlaps with the location of the known phosphorylation site S10 in the Histone H3 tail. This modeling exercise has significant implications, since the Histone H3 S10 site phosphorylation acts a strong inhibitory role in LSD1 catalyzed demethylation. (76) Thus, I hypothesize that just as histone H3 S10 phosphorylation inhibits the H3 tail from entering the active site, the NLS S111 site phosphorylation restricts NLS inhibition of LSD1, thus serving as a way to relieve the auto inhibition of LSD1.

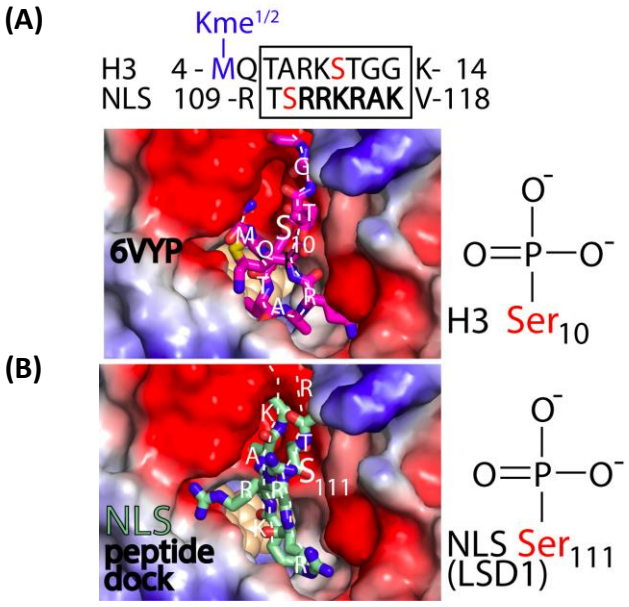


Figure 13: Phosphorylation, electrostatic, and sequence similarities between Human H3 N terminal peptide (A) and NLS peptide of LSD1 (B).
(A) H3 peptide bound to LSD1 active site (PDB 6VYP) is in magenta color. **(B)** Rosetta dock modeling of the NLS peptide (a.a res. 109-118) in green color, at the LSD1 active site. Docked model suggests the mode of auto-inhibitory binding to the active site. The location of the histone Ser10 (A) and Ser111 in LSD1 are shown as both serine residues known to be phosphorylated. Just as phosphorylated histone Ser10 impacts LSD1's activity on H3K4me2 nucleosomes, I propose that phosphorylated Ser111 within the N-terminus of LSD1 can regulate an autoinhibitory molecular mechanism.

2.2.10 The Co-crystal Structure of NLS Bound to ΔN LSD1-CoREST

To determine the exact binding location of NLS sequence in the crystal structure of LSD1, I performed co-crystallization studies with purified ΔN LSD1-CoREST with NLS peptide followed by further crystal soaking experiments with the obtained ΔN LSD1-CoREST-NLS co-crystals. Crystallization was performed using three different crystallization conditions screens that use sodium citrate, ammonium sulfate and sodium potassium tartrate as precipitation agents. (48)(51)(53) Among the three

different conditions screens used, 2 M sodium formate, 0.1 M sodium chloride, 0.1 M sodium citrate (pH 5.6) and 10 mM DTT gave best crystal morphology and diffraction (up to 2.6 Å diffraction crystals) (Figure 14-A). Upon X ray diffraction data processing by Phaser MR (full featured) in PHENIX the Fourier difference maps (|Fo|−|Fc|) to 3.5 Å show electron density corresponding to weak, partial occupancy of the NLS peptide at the active site of LSD1. A polder map was generated to obtain more pronounced and observable electron density that could be attribute to the NLS peptide (Figure 14-C), but the resolution of the structure and the weak density are of poor quality. This data suggests that the NLS peptide can localize at the active site and provides supportive evidence for direct binding to the active site of the enzyme.

To better predict NLS peptide binding and interaction at the LSD1 active site, peptide soaked crystals of LSD1 in conditions without high anionic salts (formate and sulfate) will be required. In addition, an all atomic molecular dynamics (MD) simulation of the active site with and without various H3 and NLS peptide will be needed. These experiments will help us to better assess the binding free energy of the complex through combined structural biology approaches and may be helpful for elucidating the proposed auto-regulatory mode of action of the NLS region of LSD1.

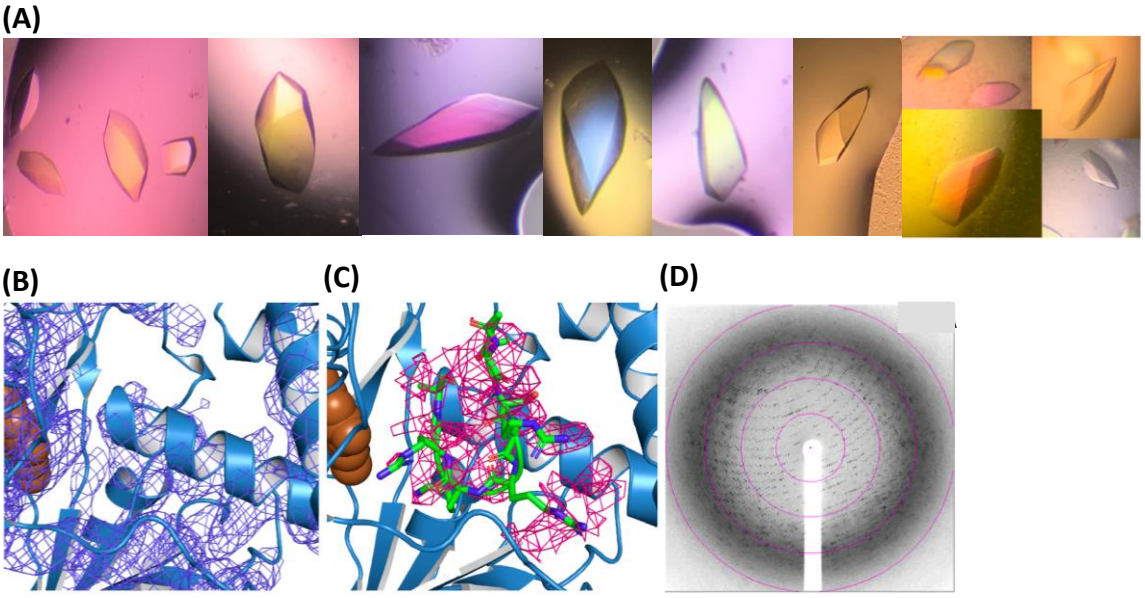


Figure 14: LSD1-CoREST-NLS crystallization.
(A) Optimized crystals of ΔN LSD1-CoREST-NLS.
(B) 2Fo-Fc density map (1.1 rmsd) at 3.5 Å. **(C)** Low resolution difference density (mFc-DFc), (pink mesh, 2.3 rmsd) with NLS peptide (green). **(D)** Diffraction pattern generated by ΔN LSD1-CoREST crystals.

2.2.11 FL LSD1-CoREST Binds Nucleosomes Tighter Than ΔN LSD1-CoREST

SPR studies were performed comparing quantitative nucleosome binding by ΔN LSD1-CoREST and FL LSD1-CoREST complexes. Titration experiments using high (~2400 response units) and low (~1000 surface units) surface densities were used to show that full length LSD1-CoREST binds unmodified nucleosomes immobilized via biotinylated 147 bp "601" DNA sequence on a sensor chip with 17 fold higher affinity when compared with the binding affinity of ΔN LSD1-CoREST (Figure 15). SPR sensorgram data of the response units (RU) versus time graphs show consistent 1:1 stoichiometry Langmuir model binding kinetics. According to the difference in RU versus protein

concentration at high and low surface immobilization densities, FL LSD1-CoREST has a higher affinity towards nucleosomes than ΔN LSD1-CoREST. These SPR data demonstrate that LSD1 N terminal region plays a direct role in nucleosome recognition and binding yet in a catalytically unproductive manner that does not facilitate the enzymatic activity of LSD1.

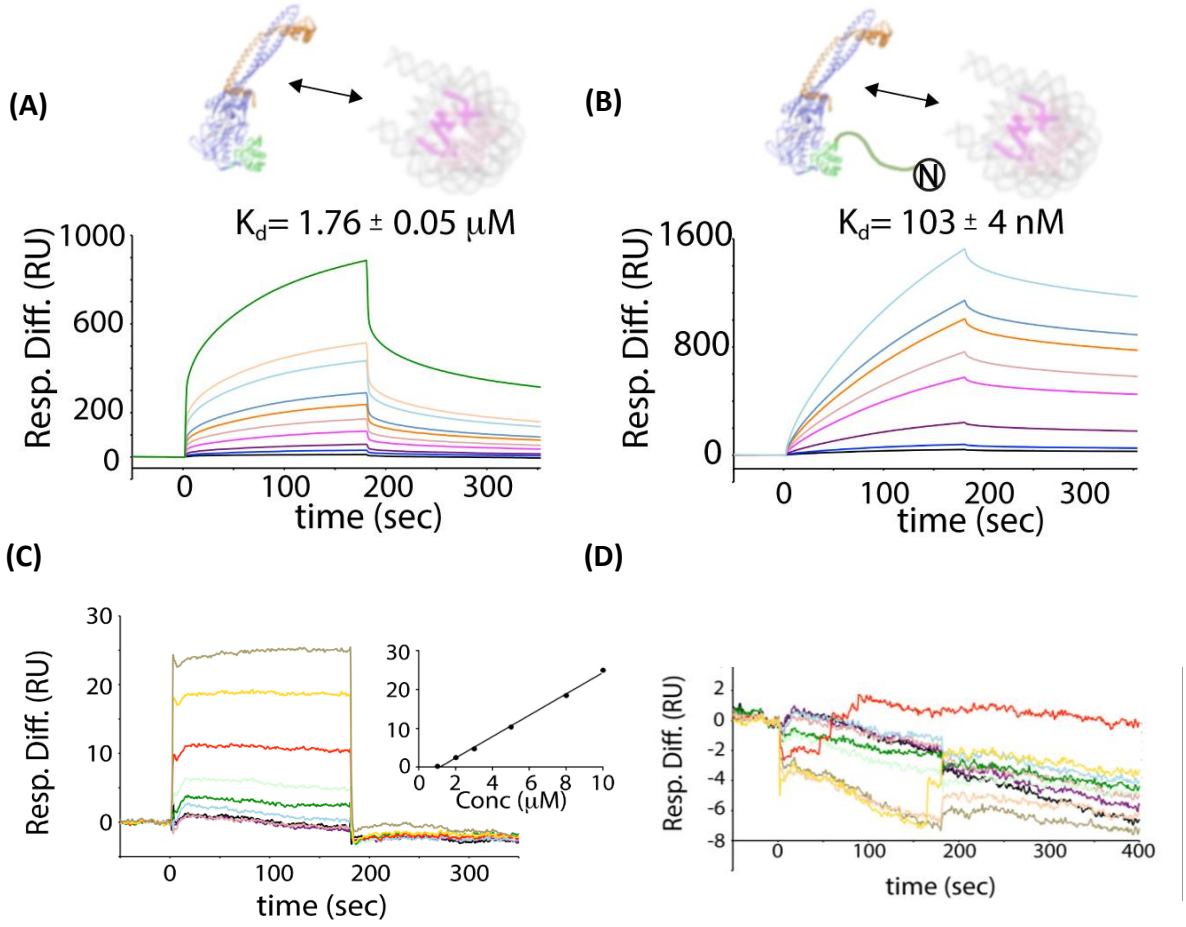


Figure 15: Surface plasmon resonance (SPR) experiments confirm the tight binding nature of ΔN LSD1-CoREST **(A)** compared to FL LSD1-CoREST **(B)** binding with nucleosomes.

SPR data of NT-LSD1 (100-151) binding weakly to nucleosomes, whereas the SWIRM domain of LSD1 does not bind nucleosomes. **(C)** N-terminal peptide (100-151) binds nucleosomes weakly but **(D)** the SWIRM does not bind mononucleosomes. The weak binding by N-terminus (residues 100-151) appears to contribute to increased binding of the FL-LSD1 to the nucleosome.

2.3 Discussion

My results led to an unexpected finding whereby LSD1's own NLS region can act as a reversible, competitive inhibitor of LSD1's catalytic activity. I found a differential enzyme activity and nucleosome binding of FL LSD1-CoREST and ΔN LSD1-CoREST whereas ΔN LSD1-CoREST is more catalytically active than the FL LSD1-CoREST with the N terminal IDR region. The FL LSD1-CoREST containing intrinsically disordered N terminus (aa 1-170) showed ~4.5 times lower catalytic activity than that of ΔN LSD1-CoREST, towards the H3K4me^2 nucleosomal substrate.

The reason for the difference in activity is most likely attributed to presence of the intrinsically disordered N terminus that resides immediately above the catalytic site and extends close to the substrate binding pocket. The N terminus of LSD1 is long enough to engage with nucleosome core particle and participate in regulatory interactions within LSD1 itself that effects its catalytic activity. From the SPR binding studies, it was evident that the N terminal conserved stretch of 100-151 amino acids binds tightly to the N terminally truncated LSD1 itself (K_d 571 ± 192nM). I sought to investigate the region of N terminus that affects the LSD1 demethylation and contributes in binding to the truncated enzyme. For this, three consecutive peptides were prepared namely: NLS peptide containing LSD1's nuclear localization signal, P1 and P2 from the conserved region of LSD1 N terminus. Determined by ITC experiments, of the three peptides, positively charged NLS peptide that contains the nuclear localization signal sequence binds to the ΔN LSD1-CoREST complex (K_d 15.4 ± 4.3μM) with affinities consistent with the N terminal LSD1 (103-151 aa) binding. Likely mimicking the heavily

modified electropositive histone H3 tail, the intrinsically disordered positively charged NLS peptide inhibits demethylation activity on H3K4me^2 peptide substrates (K_i ~ 3.3 ± 0.6μM) as well as large electronegative H3K4me^2 nucleosome substrates (IC_{50} ~ 30.4 ± 3.2μM). This inhibition by NLS region (aa 107-120) is the reason for the impaired catalytic activity of FL LSD1-CoREST.

In addition to the NLS, conserved region of the LSD1 N terminus contains multiple functionally important PTMs such as phosphorylation (at T110, S111, S131, S137), and methylation (at K114, S126) that influence DNA damage repair mechanisms, transcription networks and promote release of LSD1 from chromatin during mitosis. I observed that one of these functionally important PTMs, K114 dimethylation on NLS peptide does not affect NLS peptide's inhibition potential. In addition, the K114 dimethylation is not demethylated by LSD1 itself. Interestingly, determined by the HRP coupled assay, the introduction of phosphorylation PTMs at T110 and S111 of the identical NLS peptide (T110D, S111D phosphomimetic mutation on NLS peptide/ NLSp*) that adds electronegative modifications close to the electropositive NLS region, relieves the NLS peptides inhibition of LSD1 activity on model peptide substrate by tenfold (K_i 38.7 ± 1.8μM). This was further supported by the kinetics of demethylation of FL DD mutant LSD1-CoREST that display nearly restored catalytic efficiency, almost identical to the ΔN LSD1-CoREST. These findings demonstrate a regulatory mechanism of how phosphorylation PTMs that occurs proximal to the NLS sequence can play a de-repressive role in the electrostatic based interactions between the NLS and LSD1 catalytic cavity that leads to the autoinhibitory repressive role of the N terminus NLS

region. Supporting these results crystal structure of NLS peptide bound to ΔN LSD1-CoREST shows transient electron density of NLS peptide (3.5 Å) bound at the active site cavity of LSD1 further suggesting a competitive mode of inhibition. I performed multiple attempts of co crystallization and crystal soaking experiments with different peptide concentrations and series of different crystallization solutions in order to improve the resolution and occupancy of NLS peptide at the active site. I believe that since the crystallization solutions are abundant in negatively charged ions, the electrostatic driven binding of positively charged NLS peptide at the negatively charged active site is discouraged resulting a transient electron density of peptide bound to active site.

The LSD1 active site can differentially bind and interact with protein sequences with distinct biological functions (Figure 16 and Table 1). Similar to the NLS peptide, LSD1-CoREST bind to a positively charged 20 a.a peptide derived from SNAIL sequence that acts as a competitive inhibitor with an inhibitor constant (K_i) of 0.21 ± 0.07μM. (77) When compared, 21 a.a. H3 peptide (K_i= 1.8 ± 0.6μM) and SNAIL1 has conserved sequence similarity (Figure16, highlighted in magenta) and identity.(78) The same sequence similarity in inverse can be seen in the NLS sequence of the LSD1 N terminal "NLS peptide" suggesting that LSD1 recognize and bind these three peptides similarly. Moreover, the Ser10 phosphorylation site which accounts for disrupting the productive binding geometry of H3 N terminal tail substrate and lead to totally abolishing enzyme activity when phosphorylated is situated close to the S111 phosphorylation site of the NLS peptide when aligned as in the Figure 16. (54) This suggests that NLSp* relieves the

NLS inhibition of LSD1 activity by disrupting the productive binding of NLS peptide into the active site cavity, that is necessary for NLS inhibition.

(A)

SNAIL1 ^{1}-PRSFLVRKPSDPNRKPNYSE-^{20}	K_i =0.21 ± 0.07µM
Histone H3 ^{1}-ARTKQTARKSTGGK-^{14}	K_i= 1.8 ± 0.6µM
NLS ^{120}- YEVKARKRRSTRRG-^{107}	K_i= 3.3 ± 0.6µM
INSM1 ^{1}- PRGFLVKRSKKSTPVSYRVR-^{20}	K_i= 0.24 ± 0.1µM

Figure 16: Sequence alignment of LSD1 inhibitors SNAIL1 (1-20 a.a), Histone H3 (1-14 a.a), NLS (120-107 a.a) and INSM1 (1-20 a.a).
(A) Sequence alignment of the N terminal histone H3 peptide (1-14 a.a.), SNAIL1 N terminal peptide (1-20 a.a.), LSD1 N terminal NLS peptide (14 a.a.) and INSM1 peptide (1-20 aa) show sequence similarity with the presence of electropositive residues similar to H3 substrate. with their corresponding K_i values. ΔN LSD1(171-836)-CoREST (308-440) was used for inhibition assays of SNAIL1, Histone H3 and INSM1 (77) as well as ΔN LSD1(171-852)-CoREST (282-482) was used for NLS inhibition assays. The residues with sequence similarity are highlighted in magenta. H3 Lys4 methylation site is in blue. Sites of phosphorylation are in red.

Taken together, my results led to an unexpected observation whereby the nuclear localization signal region in LSD1's own IDR functions as a reversible, competitive inhibitor of its demethylation activity on peptide substrate and nucleosome substrate. It further introduces the possibility that, phosphorylation near the NLS sequence may regulate the LSD1's nuclear transport as well as functions to fine tune LSD1 chromatin remodeling enzyme activity in an autoregulatory manner by partially relieving the NLS inhibition.

Collectively, these results have high impact in the understanding of the structure-function relationships of LSD1 N terminus highlighting the versatile and

synergistic roles of PTMs, IDRs and structured regions near the active site cavity of LSD1. Furthermore, my findings are critical to understand LSD1's vast interaction network with implications for how oncogenic intrinsically disordered proteins such as p53 associate with LSD1. It would be interesting to test how LSD1's N terminus IDR may regulate non histone substrates such as P53 as this could help to expand LSD1's druggable space.

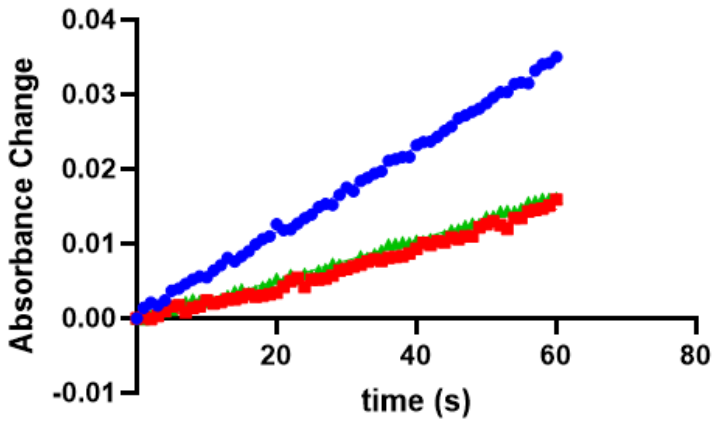


Figure 17: Peptide-based inhibition is independent of Horseradish peroxidase (HRP) concentration in the coupled demethylase reaction mixture.
This is a control experiment to assess the influence of the HRP concentration on the inhibition assays. For all three trials 20 µM H3K4me^2 peptide substrate was used. The LSD1-demethylase time course reaction at 0 µM NLS peptide inhibitor (blue), 5 µM NLS-peptide inhibitor and 1X HRP (1 µg/150 µL of reaction mixture) (red), and with 5 µM inhibitor and 2X HRP (2 µg/150 µL) (green). No change in activity due to HRP changes indicates the NLS peptide directly inhibits the LSD1 enzyme in the demethylation reaction.

2.4 Materials and Methods

2.4.1 Plasmid Construction, Cloning and Site Directed Mutagenesis

The N-terminally truncated ΔN LSD1 (LSD1 171–852) and CoREST-C (CoREST 286–482) recombinant plasmids were a generous gift of Dr. Cole (Johns Hopkins University). For creation of the recombinant plasmid, coding region for human ΔN LSD1 was cloned into pGEX-6P-1 expression vector with an ampicillin resistance marker. A pET28b expression vector with kanamycin resistance marker was used for cloning coding regions for human CoREST (286–482).

The recombinant plasmids for the full-length (FL) LSD1 (1–852) expression was a generous gift of Dr. Shi (Harvard University). In this plasmid, full length human cDNA for full length LSD1 was cloned into pET-15b bacterial expression vector with an ampicillin resistance marker.

For the N terminal phosphomimetic FL LSD1 (FL DD LSD1), T110D S111D point mutation were introduced to the FL LSD1 plasmid using PCR protocol that amplifies the entire plasmid. The site directed mutagenesis was performed on pET-15b bacterial expression vector carrying FL LSD1 (1–852). Amount of 140 ng of template FL LSD1 plasmid was amplified by 35 cycles in a reaction mixture of 50µL prepared by adding 25µL of CloneAmp™ HiFi PCR Premix and 4µL of each 3.7µM forward and reverse primers. Following amplification, the PCR product was exposed to 1µL of Dpnl restriction enzyme at room temperature for 12 hours, for removing the parent template from the PCR reaction. The digested product was further cleaned using Thermo scientific™

GeneJET DNA cleanup Micro kit. Purity of the product was assessed by 0.8% Agar gel with 0.002% Ethidium bromide. Three µL of cleaned DNA was transformed into DH5α *E. coli* competent cells by standard transformation protocol. The transformed cells were plated on LB agar plates with 100µg/mL ampicillin resistance and grown at 37˚C overnight. The positive colonies were grown in LB media with 100µg/mL ampicillin overnight and plasmid DNA was isolated using GenElute^{TM} HP plasmid Miniprep kit. The mutation in the plasmid was confirmed using DNA sequencing.

2.4.2 Co Expression of LSD1-CoREST complexes

2.4.2.1 Co Expression of ΔN LSD1 – CoREST

GST tagged human ΔN LSD1 (LSD1 171–852) and hexa-histidine tagged human CoREST-C (CoREST 286–482) were co-expressed in Rosetta (DE3) pLysS competent cells to produce recombinant ΔN LSD1- CoREST complex. The ampicillin resistant pGEX-6P-1 plasmid containing human ΔN LSD1 gene and kanamycin resistant pET28b plasmid containing human CoREST gene were both added into chloramphenicol resistant Rosetta (DE3) pLysS competent cells. The mixture was incubated on ice for 40 minutes before transforming the plasmids into competent cells by heat shock for 50 seconds at 42 ˚C. The transformed cells were plated on the LB agar plates containing 50 ng/µL ampicillin, kanamycin and 25ng/µL chloramphenicol and grown at 37˚C overnight. The next day 6 separate colonies were selected and grown in 10mL each of LB media with 200ng/µL ampicillin, kanamycin and 40ng/µL chloramphenicol concentrations, at 30 ˚C, overnight. The next day, six liters of ZYP-5052 autoinduction media (928 mL ZY media, 1 mL 1M $MgSO_4$, 20 mL 50 x 5052, 50 mL 20 x NPS and antibiotics for 1L of ZYP-5052

autoinduction media) with same antibiotic concentrations as starter cultures was inoculated using the starter cultures and they were grown at 25 ˚C with constant shaking at 225 r.p.m. After growing 30 hours, the cells were centrifuged at 7000 r.p.m. for 15 minutes in 4 ˚C. The cells were frozen and stored at -20˚C.

2.4.2.2 Co-Expression of FL LSD1-CoREST and FL DD Mutant LSD1- CoREST

Both hexa-histidine tagged human FL LSD1 (LSD1 1–852) and human CoREST-C (CoREST 286–482) were co-expressed in Rosetta (DE3) pLysS competent cells to produce recombinant FL LSD1- CoREST complex using ampicillin resistant pET-15b plasmid containing FL LSD1 gene and Kanamycin resistant pET28 plasmid containing CoREST gene. Similarly, the hexa-histidine tagged human FL DD mutant LSD1 (LSD1 1–852) and human CoREST-C (CoREST 286–482) were co-expressed in Rosetta (DE3) pLysS competent cells to produce recombinant FL DD mutant LSD1- CoREST complex. The starter cultures, and 6L flask inoculation growing and collecting cell palate were performed similar to the ΔN LSD1- CoREST co expression.

2.4.3 Purification of LSD1-CoREST Complexes

2.4.3.1 Purification of ΔN LSD1- CoREST

Cell pellets were re-suspended in ice cold lysis buffer (50 mM Tris base pH 8.0, 300 mM NaCl, 4 mM imidazole, 0.1mM PMSF, Lysozyme, FAD) and the extract was sonicated on ice at an output power of 6 for about 10 times with 1-minute intervals in between. The soluble extract was centrifuged at 18,000 r.p.m, for 45 minutes at 4 ˚C. Supernatant was separated and loaded on Ni-NTA resin that was equilibrated with wash buffer (50 mM Tris base pH 8, 300 mM NaCl with 4 mM imidazole). Lysate was rocked

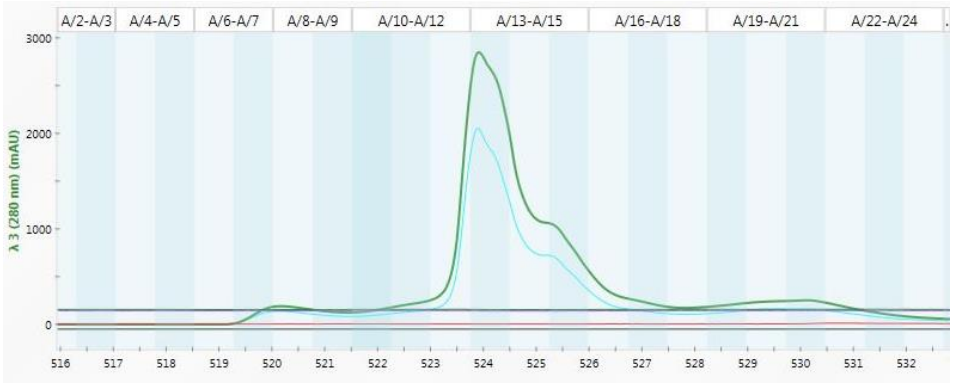
with resin for 1.5 hours at 4 ˚C. Resin was centrifuged at 4000 r.p.m for 10 minutes at 4 ˚C and supernatant was discarded. The resin was washed with wash buffer until the flow through is clear and the protein was eluted with elution buffer (50 mM Tris base pH 8.0, 300 mM NaCl with 300 mM imidazole). The yellow color in the elutions were used as a rough estimate of the ΔN LSD1- CoREST presence and concentration. The protein elution was loaded on glutathione Sepharose resin pre equilibrated with buffer A (50 mM Tris base pH 8.0, 300 mM NaCl). The glutathione Sepharose resin was washed with 4 column volumes of buffer A and protein was eluted using 10mM glutathione in buffer A. The Hexa-histidine and GST tags were removed by adding restriction grade thrombin and 3C protease into the elute together with 0.5mM TCEP and incubating overnight at 4 ˚C. The next day, protein solution was concentrated using Amicon 30kDa ultra centrifugal concentrator to remove glutathione and tags. Concentrated protein was loaded on a cleaned glutathione Sepharose resin pre equilibrated with buffer A. The flowthrough from the resin and washings were collected and concentrated further. The concentrated protein sample was injected and separated by a size exclusion column connected to NGC chromatography system and 750µL fractions were collected using 25 mM HEPES Na pH 7.5, 100 mM NaCl, 2 mM TCEP buffer. Fractions were analyzed by 10% acrylamide SDS-PAGE, and the desired fractions were concentrated. Protein concentrations were measured by Nano Drop™ as well as using the Bradford Assay. Glycerol was added to 40% (v/v) before the complex was flash frozen and stored at -20˚C.

2.4.3.2 Purification of FL LSD1- CoREST

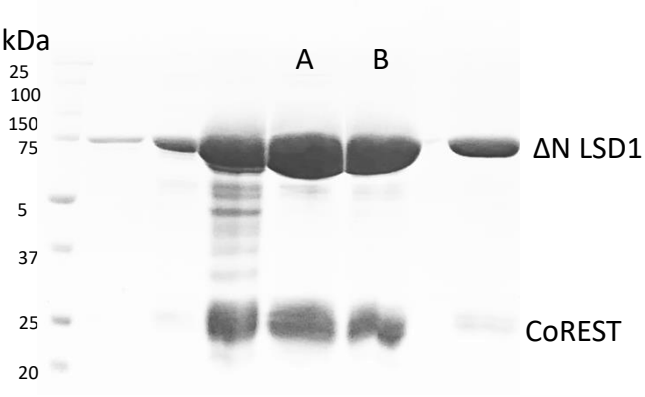
Cell pellets were re-suspended in ice cold lysis buffer (50 mM Tris base pH 8.0, 300 mM NaCl, 4 mM imidazole, 0.1mM PMSF, Lysozyme, FAD) and the extract was sonicated on ice at an output power of 6 for about 10 times with 1-minute intervals in between. The soluble extract was centrifuged at 18,000 r.p.m, for 45 minutes at 4 ˚C. Supernatant was separated and loaded on Ni-NTA resin that was equilibrated with wash buffer (50 mM Tris base pH 8, 300 mM NaCl with 4 mM imidazole). Lysate was rocked with resin for 1.5 hours at 4 ˚C. Resin was centrifuged at 4000 r.p.m for 10 minutes at 4 ˚C and supernatant was discarded. The resin was washed with wash buffer until the flow through is clear and the protein was eluted with elution buffer (50 mM Tris base pH 8.0, 300 mM NaCl with 300 mM imidazole). The yellow color in the elution was used as a rough estimate of the FL LSD1- CoREST presence and concentration. The protein solution was dialyzed in 25 mM HEPES Na pH 7.5, 100 mM NaCl, 0.1 mM TCEP buffer overnight at 4 ˚C. The next day, dialyzed protein was loaded into pre-packed Ni-NTA column connected to NGC chromatography system (Bio-Rad). After washing the column with 4 column volumes of buffer A (50 mM Tris base pH 8.0, 300 mM NaCl), protein gradient elution was performed using 0-100% imidazole buffer A (50 mM Tris base pH 8.0, 300 mM NaCl, 300 mM imidazole) at 0.5mL/min flow rate. Fractions were analyzed by SDS-PAGE, the desired fractions were pooled and 0.5mM TCEP was added. The hexa-histidine tags in protein complex were removed by adding restriction grade thrombin letting sit overnight at 4 ˚C. The next day, protein solution was concentrated and separated by a size exclusion column connected to NGC chromatography system, using

25 mM HEPES Na pH 7.5, 100 mM NaCl, 2 mM TCEP buffer. Fractions collected were analyzed further by SDS-PAGE, and 40% (v/v) glycerol was added to the desired fractions before storing in -20 ˚C. FL DD mutant LSD1-CoREST complex was also expressed and purified similarly to their wild-type FL LSD1-CoREST.

(A)



(B)



(C)

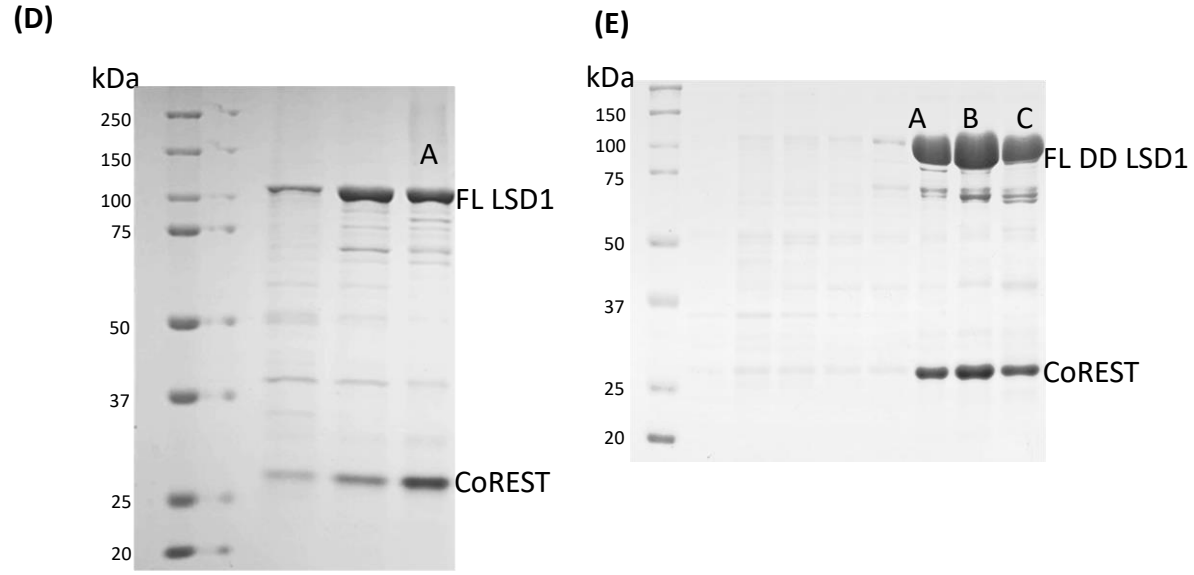


Figure 18: Purification of the LSD1-CoREST complexes.
(A) Chromatogram of ΔN LSD1-CoREST purified by Size-exclusion chromatography (SEC). Green shows absorbance for protein at 280nm and cyan line shows absorbance for nucleic acids at 255nm. **(B)** Corresponding 10% poly acrylamide SDS-PAGE gel of chromatography fractions. A and B fractions were pooled together for further assays. **(C)** Purified ΔN LSD1-CoREST protein. **(D)** SDS-PAGE gel of chromatography fractions of FL LSD1-CoREST purified by SEC. Fraction A was used for further assays. **(E)** 10% poly acrylamide SDS-PAGE gel of chromatography fractions from FL DD mutant LSD1-CoREST purification.

2.4.4 Removing Nucleic Acid Contamination of LSD1-CoREST Complexes

Nucleic acid contamination during purification was observed during size exclusion chromatographic step in purification (Figure 18-A), in purified LSD1-CoREST complexes during the measurement of concentration by NanoDrop™ and when the samples were analyzed for RNA presence using 8M Urea, 10 % polyacrylamide gel electrophoresis followed by toluidine blue nucleic acid specific staining of the gel. Therefore, the removal of nucleic acid contamination was performed during purification of LSD1-CoREST complexes. For removal of RNA contamination, RNaseT1 was used that specifically degrades single-stranded RNA at G residues by cleaving the phosphodiester

bond between 3'-guanylic residue and 5'-OH residue of adjacent nucleotides. During the purification of ΔN LSD1-CoREST, 1:10 diluted RNaseT1, 2µL was added to protein eluted from the first Ni NTA resin, and the sample was kept at 4 ˚C overnight before the GST column step. For the removal of RNA contamination in FL LSD1-CoREST, the purified protein was incubated at 4 ˚C for 2 hours with 1:10 diluted RNaseT1, 2 µL. After 2 hours, the sample was concentrated and buffer exchanged several times with 25 mM HEPES Na pH 7.5, 100 mM NaCl, 2 mM TCEP containing buffer.

2.4.5 HRP Coupled Assay for Measuring LSD1 CoREST Activity

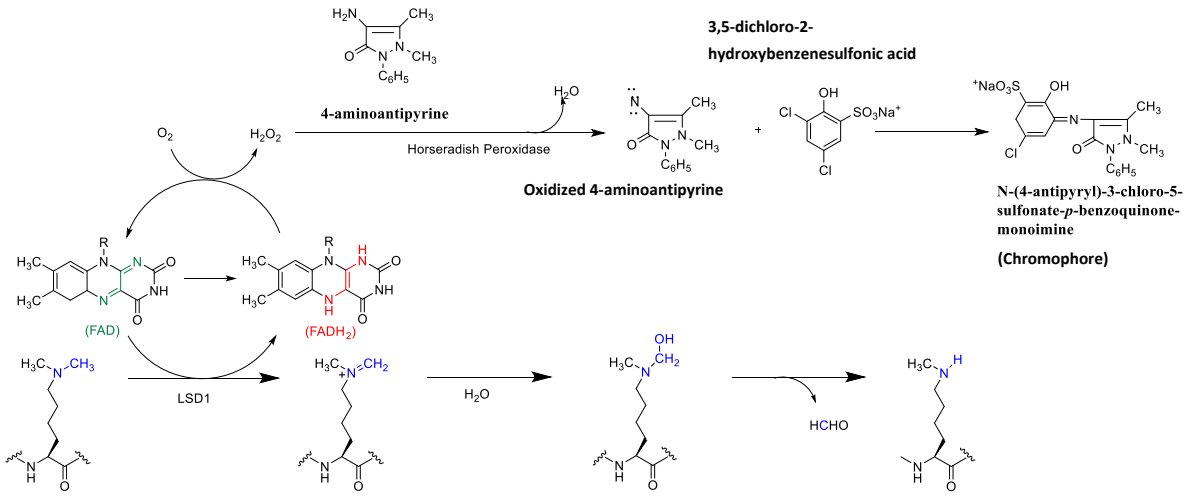


Figure 19: Horseradish peroxidase coupled assay for detection of hydrogen peroxide formation during demethylation reaction by LSD1.
Horseradish peroxidase (HRP) uses H_2O_2 produced during the reoxidation of $FADH_2$ to couple 4-aminoantipyrine and dichlorohydroxybenzenesulfonate to form a chromophore that is detected at 515 nm. (79)

LSD1-CoREST activity assays were performed using a peroxidase-coupled assay (Horseradish peroxidase -HRP coupled assay) that monitors H_2O_2 production under

aerobic conditions. Initial velocity measurements were measured using a Shimadzu UV-2600^{TM} UV-Visible spectrophotometer equipped with thermostated cell holder (T =25 ˚C). ΔN LSD1-CoREST was buffer exchanged into 25 mM HEPES pH 7.4, 100 mM NaCl, 5 mM TCEP buffer. A 150µL reaction mixture contained 50 mM HEPES(Na) pH 7.5, 0.3µM ΔN LSD1-CoREST, 1µg of horseradish peroxidase (HRP), 0.1mM 4-aminoantipyrine, 1.0 mM and substrate concentrations 10, 20, 30, 40, and 60µL. Enzyme reactions were initiated by the addition of substrate; H3K4me^2 peptide aa 1–21 or P1 peptide (107-120) dimethylated at K114, into the reaction mixture in quartz cuvette. Activity region was 0-30 and total time of each measurement was 180 seconds. Baseline correction was done at 505- 525 nm. Changes in absorbance were measured at 515 nm wavelength and initial velocity calculations were done using an extinction coefficient of 26,000 M^{-1} cm^{-1} by Graphpad Prism 8. Initial velocity values obtained from absorbance vs. time graph, and they were fitted to Michaelis-Menten equation non-linear regression and obtained V_{max} (maximum velocity), apparent k_{cat} (turnover number) and K_m (Michaelis constant).

2.4.6 LSD1 Inhibition Assays with N-Terminal Model Peptide Substrate Using HRP Coupled Assay

Peptide inhibition assays in presence of model substrate were conducted as follows using a peroxidase-coupled assay that monitors H_2O_2 production under aerobic conditions. A 150 µL reaction mixture contained 50 mM HEPES (Na) pH 7.5, 0.3µM ΔN LSD1-CoREST, 1µg of HRP, 0.1 mM 4-aminoantipyrine, 1.0 mM 3,5-dichloro-2-hydroxybenzenesulfonic acid, NLSK114me2 10µM (as inhibitor and substrate) / NLSp* 10µM / NLS 10µM / P1 10µM / P2 10µM and H3K4me2 21aa substrate 10µM. Enzyme

reactions were initiated by the addition of substrate; H3K4 me2 peptide aa (1–21) into the reaction mixture in quartz cuvette. Changes in absorbance were measured at 515 nm wavelength using UV-2600 Shimadzu UV-Visible spectrophotometer equipped with thermostated cell holder (T =25 ºC). Initial velocity calculations were done using an extinction coefficient of 26,000 M^{-1} cm^{-1} by Graphpad Prism 8.

2.4.7 LSD1-CoREST Inhibition Assays with NLS Peptide and Phosphomimetic N-Terminal Peptide NLSp* Using Model Substrate

Peptide inhibition assays using NLS and NLSp* were performed in presence of model substrate as follows using a peroxidase-coupled assay that monitors H_2O_2 production under aerobic conditions. A 150 µL reaction mixture contained 50 mM HEPES (Na) pH 7.5, 0.3 µM ΔN LSD1-CoREST, 1µg of HRP, 0.1 mM 4-aminoantipyrine, 1.0 mM 3,5-dichloro-2-hydroxybenzenesulfonic acid, NLS or NLSp* as inhibitor (NLS peptide -0 µM, 2.5 µM, 5 µM, 10 µM or NLSp* peptide- 0 µM, 10 µM, 40 µM, 100 µM), and substrate concentrations 10 µM, 20 µM, 30 µM and 60 µM. Enzyme reactions were initiated by the addition of substrate; H3K4 me2 peptide aa (1–21) into the reaction mixture in quartz cuvette. Changes in absorbance were measured at 515 nm wavelength using UV-2600 Shimadzu UV-Visible spectrophotometer equipped with thermostated cell holder (T =25 ºC). Initial velocity calculations were done using an extinction coefficient of 26,000 M^{-1} cm^{-1} by Graphpad Prism 8. Initial velocity values obtained from absorbance vs. time graph were fitted to competitive inhibition model using non-linear regression fit and obtained K_i (inhibitor constant) value.

2.4.8 Circular Dichroism (CD) Spectroscopy of LSD1 N-Terminal Peptide

CD spectra for 5µM LSD1 N terminal conserved region NLS peptide (aa 107-120) in water was recorded at room temperature on a Chirascan^{TM} V100 CD spectrophotometer with a 1 mm cell, 1nm band width, 4s time per point, and 0.1ms timed intervals. Spectra from 180–280 nm were averaged over three scans, and background from a matched water only sample was subtracted.

2.4.9 LSD1-CoREST Demethylation Assays with Nucleosome Substrates

Varying concentrations (2, 5, 10, 20 µM) of ΔN LSD1- CoREST or FL LSD1- CoREST or FL DD mutant LSD1- CoREST in 50 mM HEPES pH 8, 5% glycerol, 1 mM TCEP, 50mM KCl buffer were used for demethylation of H3K4me2 containing nucleosome substrate (100 nM) at 25 °C. After initiating demethylation reaction by adding nucleosomes (Recombinant human H3K4me2 nucleosomes with 147bp Widom 601 double stranded DNA sequence from EpiCypher^{TM}) to the reaction mixture, aliquots were withdrawn at t = 0, 2, 5, 30, 120 min followed by quenching with 4X Laemmli Dye and boiling for 2 min to stop demethylation reaction. Samples were run in 20% SDS PAGE get for 40 minutes at 200 V. Protein bands in SDS gel were transferred to immunoblot PVDF membranes and blocked with 4% fat free milk in PBS buffer. Corresponding blots were incubated with H3K4me2 and H3 specific primary antibodies overnight followed by goat anti rabbit secondary antibody for 1 h. The blots were visualized by chemiluminescence and analyzed using Amersham software. All demethylation and control experiments were performed in either duplicate or triplicate (N=2-3).

Dividing the H3K4^{me2} antibody signal by the H3 antibody signal allowed for accurate quantitation of each data point. The H3K4me2/H3 ratios were normalized at time zero and plotted as a function of time (minutes), and subsequently analyzed using nonlinear regression. Data were fit to the equation [H3K4me2] = $[H3K4me2]_{t=0}$ e^{-kobst} and the determined rate constant (k_{obs}) values were evaluated with LSD1-CoREST concentrations. Extraction of $K_{1/2}$ and k_{max} parameters were determined based on the equation: $k_{obs}= k_{max}$[Enzyme]/([Enzyme]+$K'_{1/2}$).

2.4.10 LSD1-CoREST Demethylation Assays in Presence of N-Terminal Peptide Using Nucleosome Substrate

Different concentrations of LSD1 N terminal NLS peptide (0 µM, 10 µM, 20 µM, 50 µM, 100 µM, 400 µM) were incubated with the 2µM ΔN LSD1-CoREST containing reaction mixture in absence of the nucleosome substrate at 25 °C for 60 minutes. Demethylation reactions were initiated by adding 100 nM nucleosome substrate into the reaction mixture. Aliquots of 10µL were withdrawn at t = 0, 2, 5, 30, 120 minutes time points and quenched using Laemmli dye followed by boiling for 2 minutes. The assay products were resolved by 20% SDS-PAGE and the degree of inhibition was quantified using H3K4me^2 specific antibody relative to the amount of H3 in each lane quantified using H3 specific antibody.

2.4.11 SPR Binding Assays for LSD1-CoREST Binding with Nucleosomes and NT-LSD1

SPR measurements were performed on a Biacore 3000 instrument (GE healthcare) at 298 K in running buffer (10 mM Tris, 100 mM NaCl, 1 mM EDTA, 2 mM TCEP, 0.005% Tween 20, pH 7.4). Biotinylated nucleic acid (Widom 601) was

immobilized on a streptavidin-coated sensor chip SA (GE healthcare). Streptavidin surfaces saturated only by biotin served as a reference for all collected data. For series of binding tests, biotinylated recombinant human nucleosomes (EpiCypher) were immobilized at surface densities between 4000-5000 response unit (RU). For binding experiments between ΔN and FL LSD1-CoREST with nucleosome ΔN and FL LSD1-CoREST complexes were titrated to nucleosomes whereas the concentrations of ΔN LSD1-CoREST complex were 25nM, 50nM, 100nM, 200nM, 500nM, 800nM, 1uM and 2μM; the full-length LSD1-CoREST were injected following the same concentration points as the ΔN LSD1-CoREST until 800nM, to avoid aggregation at higher concentrations.

For determining the binding of NT-LSD1 to LSD1-COREST, biotinylated NT-LSD1 was immobilized on the sensor chip with a surface density of ~560 RU. Biotinylation of the NT-LSD1 peptide was generated by the EZ-Link™ Sulfo-NHS-LC-Biotinylation Kit (ThemoFisher Scientific) and purified via SpinOUT™ GT-100 1ml column (G-Biosciences). ΔN LSD1-CoREST complex was injected at a flow rate of 40μL/min for 2 minutes, with various concentrations of 25nM, 50nM, 100nM, 200nM, 500nM, and 1μM. The time course of dissociation was recorded for 4 minutes. At the end of each measurement, a regeneration of mononucleosome surfaces was conducted by applying a mixture of 0.1% Tween 20 and 0.1% NP40 diluted in running buffer (4 consecutive injections for 30 seconds), and subsequently equilibrated with running buffer (5 minutes).

Data processing and evaluation were firstly performed using BIA evaluation 4.1.1 (Biacore AB). The kinetic curves were fitted using the 1:1 Langmuir binding mode to determine k_a, k_d and K_d. For the titrations which were binding-saturated during the

association periods, further analysis of the binding specificities was also checked by plotting the RUs at the saturated versus protein concentrations in Prism 6.01 (GraphPad Software) and analyzed using the equation below where RU_{max} is the maximum binding response units, C is the ligand concentration, h is the Hill slope and K_d is the equilibrium dissociation constant when h=1.

$$RU = RU_{max} * C^h / (K_d^h + C^h) \quad \text{(Equation 1)}$$

2.4.12 N-Terminal Peptide Binding to ΔN LSD1-CoREST Complex

ITC measurements were performed at 25°C on a MicroCal ITC_{200} (Malvern) in 25 mM HEPES, pH 7.4, 100 mM salt, 2 mM TCEP, and 1 mM EDTA. The ΔN LSD1-CoREST complex and peptide were buffer exchanged against degassed buffer under identical conditions. Duplicate measurements were performed, where 55µM of ΔN LSD1-CoREST was filled in the sample cell and the peptide at 0.72mM was titrated by 18 automatic injections of 1.8µl that occurred beginning the second injection. An initial injection of 0.4µL was also injected but not incorporated into the data analysis. Spacing of 150 seconds was applied between each titration, with a continuous stir speed of 1000 rpm. A control, background measurement was obtained under the same condition except the protein sample been replaced by pure buffer in the sample cell. The resulting data were processed by MicroCal Origin software and a comparison of binding models were evaluated, with one set of sites mode chosen as the data fitting analysis step.

2.4.13 ΔN LSD1- CoREST Crystallization

Three different series of optimized crystallization conditions that resulted in larger crystals were used for co crystallization and crystal soaking experiments. For the first crystallization condition, ΔN LSD1/CoREST protein was buffer exchanged in to 25 mM HEPES-Na pH 7.5, 100 mM NaCl, 1 mM TCEP, 1 mM PMSF buffer. The final protein concentration was set to be between 70-100µM. NLS peptide in 50mM HEPES buffer was added to the ΔN LSD1 -CoREST complex to a final concentration of 50µM. After incubation on ice for 40 minutes, sodium borohydride was added to a final concentration of 1mg/mL. Proteins were set for crystallization by hanging drop method using a series of different conditions derived from 2 M sodium formate, 0.4 M sodium chloride, 0.1 M sodium citrate (pH 5.6) and 10 mM DTT. (53) The crystallization drops were set up by combining 1µL of protein mixture and 1µL of crystallization solution on the siliconized cover slip with 1mL of crystallization solution in the reservoir. Best crystals and better crystal diffraction of the LSD1 complex with peptide was generated using crystallization solution 1.9M sodium formate, 0.1 M sodium chloride, 0.1 M sodium citrate (pH 5.6) and 10mM DTT crystallization condition.

For the second crystallization condition, protein was buffer exchanged to 25 mM HEPES-Na pH 7.5, 100 mM NaCl, 1 mM TCEP, 1 mM PMSF buffer. The final protein concentration was set to be between 70-100µM. NLS peptide in 50mM HEPES buffer was added to the ΔN LSD1 -CoREST complex to a final concentration of 50µM and incubated on ice for 40 minutes. Proteins were set for crystallization using a series of

different conditions derived from 0.63M $(NH_4)_2SO_4$, 0.6M Li_2SO_4, 0.25M NaCl, 100mM Na Citrate pH 5.6, 5mM TCEP.(48)

The third crystallization condition was a series of crystallization solutions derived from 1.4 M sodium/potassium tartrate, 100 mM N-(2-acetamido)-2-iminodiacetic acid (ADA), pH 6.1. (51) For hanging drop crystallization, ΔN LSD1/CoREST protein was buffer exchanged in to 5% glycerol (w/v), 25mM potassium phosphate pH 7.2 buffer to a final concentration of 70-100 µM. NLS peptide in 50mM HEPES pH 7.5 buffer was added to the ΔN LSD1 -CoREST complex to a final concentration of 50µM and incubated on ice for 40 minutes before crystal trays were set up. The crystal trays were kept at ~20 ˚C, undisturbed for crystallization.

2.4.14 Post Crystallization Soaks and Harvesting

Crystals typically were grown for 7-14 days before further soaking with peptide and harvesting. Crystals grew in three dimensions with birefringence and sharp edges were harvested under the microscope using CryoLoops. First the crystals were transferred to a droplet containing 1.5µL of same crystallization solution and 1.5µL of NLS peptide solution and the crystal was soaked for 30 minutes/1 hour/3hours/24 hours. A solution of solution 1.9M sodium formate, 0.1 M sodium chloride, 0.1M sodium citrate (pH 5.6) and 10mM DTT, 23% (v/v) glycerol, up to 5mM NLS peptide was prepared as the cryo-protectant for crystal harvesting. The soaked crystals were transferred into a cryo-protectant drop before quickly harvesting and flash freezing in liquid nitrogen.

In order to get better diffraction of NLS peptide bound LSD1-CoREST crystals, I tried soaking out negatively charged ions (formate) by soaking crystals through a series of crystallization solutions with progressively decreasing concentrations of formate. But the attempts were failed since the crystals started melting after few conditions.

2.4.15 Data Collection and Structural Determination

Diffraction data collection was performed at sector 21 Life Sciences Collaborative Access Team (LS-CAT), G hutch and F hutch at Argonne National Laboratories. Data were processed using XDS, and the structure of ΔN LSD1-CoREST-N terminal peptide was solved by molecular replacement with Phenix, PhaserMR (full featured). (80) The molecular replacement was performed using the previously published crystal structure of ΔN LSD1-CoREST (PDB ID:2IW5). All molecular graphics were prepared with Coot and PyMOL softwares. All molecular modeling experiments of the NLS peptide bound to LSD1's active site was performed using an in-house ROSETTA-dock algorithm and implemented by our collaborator (Ben Brown, Meiler Lab, Vanderbilt University).

Project Contributions:

All LSD1-CoREST wild type and mutant constructs purification, all western blot nucleosome demethylation assays, all HRP coupled assays for peptide (NLS, P1, P2, NLSp*, NLS K114me^2) inhibition of LSD1, all western blot LSD1 inhibition assays using NLS peptide, LSD1-CoREST crystallization with and without NLS peptide, crystal soaking, harvesting and CD spectroscopy and relevant data analysis were performed by Dulmi Senanayaka. FL DD mutant LSD1 plasmid construct preparation, SPR binding between nucleosomes and LSD1-CoREST was performed by Dr. Danyun Zeng and Dulmi Senanayaka. SPR binding assay between N terminus (aa 100-151) and LSD1-CoREST, ITC of NLS binding to LSD1-CoREST, MD simulation, Rosetta model analysis was performed by Dr. Danyun Zeng. MD simulation and Rosetta model preparation was performed by Dr. Ben Brown (Vanderbilt University). Crystallization data collection and analysis were performed by Dulmi Senanayaka, Dr. Danyun Zeng, and Dr. Nicholas Reiter.

CHAPTER 3 ROLE OF STRUCTURED RNA ON LSD1 ACTIVITY AND LSD1 INVOLVED CELLULAR PROCESSES

3.1 Repeat G4 RNA Structures Preferentially Interact with the LSD1-CoREST Complex to Mask Nucleosome Recognition

3.1.1 Introduction

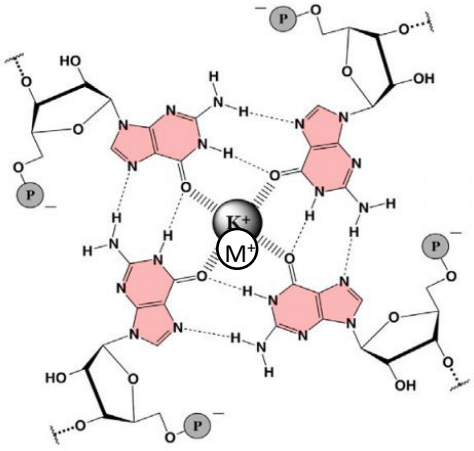
In addition to interacting with over 60 regulatory proteins, transcription factors (CoREST, p53, E2F1), key enzymes (MRE11, DNMT1, HDAC1/2) and essential nutrients (tetrahydrofolate), histone methyl regulatory enzyme LSD1, together with its interacting partner CoREST, functions as a nucleic acid binding protein. LSD1 interacts with numerous RNAs such as pre-messenger RNAs and higher order nucleic acid structures including nucleosomes, telomeric R loops, cancer associated long non-coding RNAs (lncRNAs) like HOTAIR as well as localize at telomeres to associate with G quadruplex forming Telomeric Repeat containing RNA (TERRA). In fact, long non-coding RNAs (lncRNAs) assist in a multitude of cellular roles including acting as decoys, guides, scaffolds, and signals. They directly interact with chromatin modifying proteins like LSD1 to influence the dynamic organization of the genome by controlling the chromatin structure and accessibility, modulating the gene expression. (83) These RNA-chromatin associated protein interactions are linked with a wide range of cellular processes including cancer, X inactivation and genomic imprinting. (84) Yet it is unclear how these RNA-protein structural interactions form and how these RNA mediated mechanisms alter gene expression and contribute to telomere maintenance.

In this chapter I discuss the guanine (G) rich RNA's higher order G quadruplex structure-based effect on the RNA-LSD1's interaction to mediate epigenetic changes in genome. While G quadruplexes' existence has been reported in human transcriptome and genome, the highest abundance of G quadruplex forming sequence motifs are located in telomeres as a repeat units. (85) (86) Telomeres, the DNA -protein complexes at the ends of linear eukaryotic chromosomes, which are essential for chromosome stability during mitosis and meiosis, determine the replicative lifespan of normal human somatic cells. Mammalian telomeres are comprised of 5'-TTAGGG-3' repeats ending with shorter G-rich single stranded overhangs. The double strand part of telomere is bound to a multiprotein complex called shelterin that prevents double-strand breaks in telomeres. Telomeres were originally considered to be transcriptionally silent until the discovery of telomeric repeat containing RNA (TERRA) in 2007. These G-quadruplex forming TERRA RNAs, primarily transcribed during the Alternative Lengthening of Telomeres (ALT) mechanism in human cells, are recognized to form an integral part of telomeric heterochromatin. (87)(88)

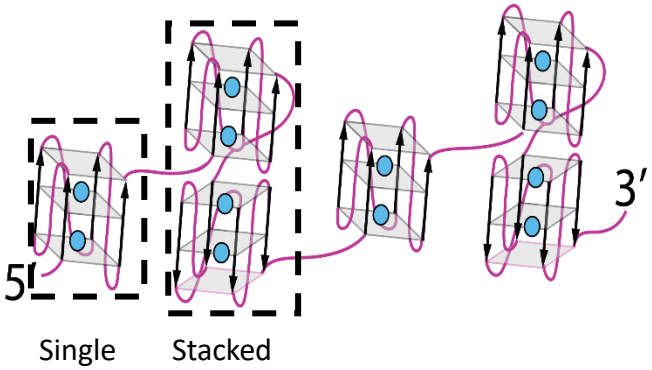
The human telomeric repeat containing RNA (TERRA) is transcribed by RNA polymerase II from the ends of human chromosome 18 and the pseudo autosomal regions Xq/Yq. It contains 5'-(UUAGGG)-3'$_{35\text{-}1500}$ repeats in heterogeneous length ranging from 100-9000 nucleotides. As these TERRA sequences are rich in guanines, they have been reported to form propeller type parallel stranded RNA G quadruplex structures in vitro and in vivo. (88) (89) These structures are organized in stacks of G quartets or G tetrads (GQ or G4) where four guanines are connected in a planar

structure with each other via hoogsteen base pairing (Figure 20-A). Three or more G quartets stack on top of each other to form a G quadruplex structure. Physiologically relevant monovalent metal cations such as K^+, Na^+, NH_4^+ can coordinate, stabilize the H-bonded G quartets to different extents, and enhance base stacking interactions by binding between the G tetrads or intercalating to the central anionic core of them. (90) Also, it is reported that TERRA G quadruplexes form higher order structures with dimerization of two stacked units.(91) (92) (93)

(A)

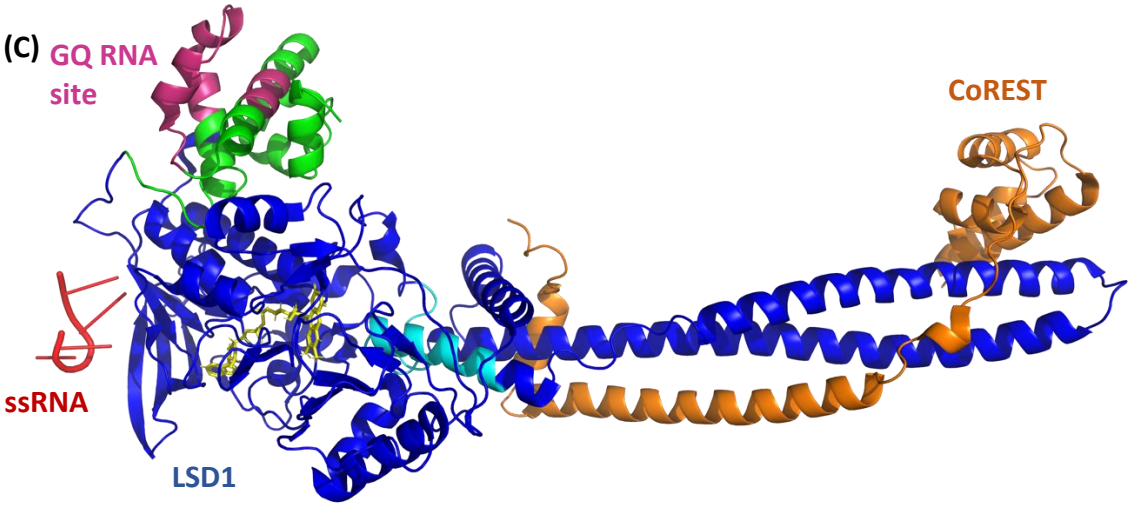


(B)



$M^+ = K^+ >> NH_4^+ \geq Na^+ >> Li^+$

(C)



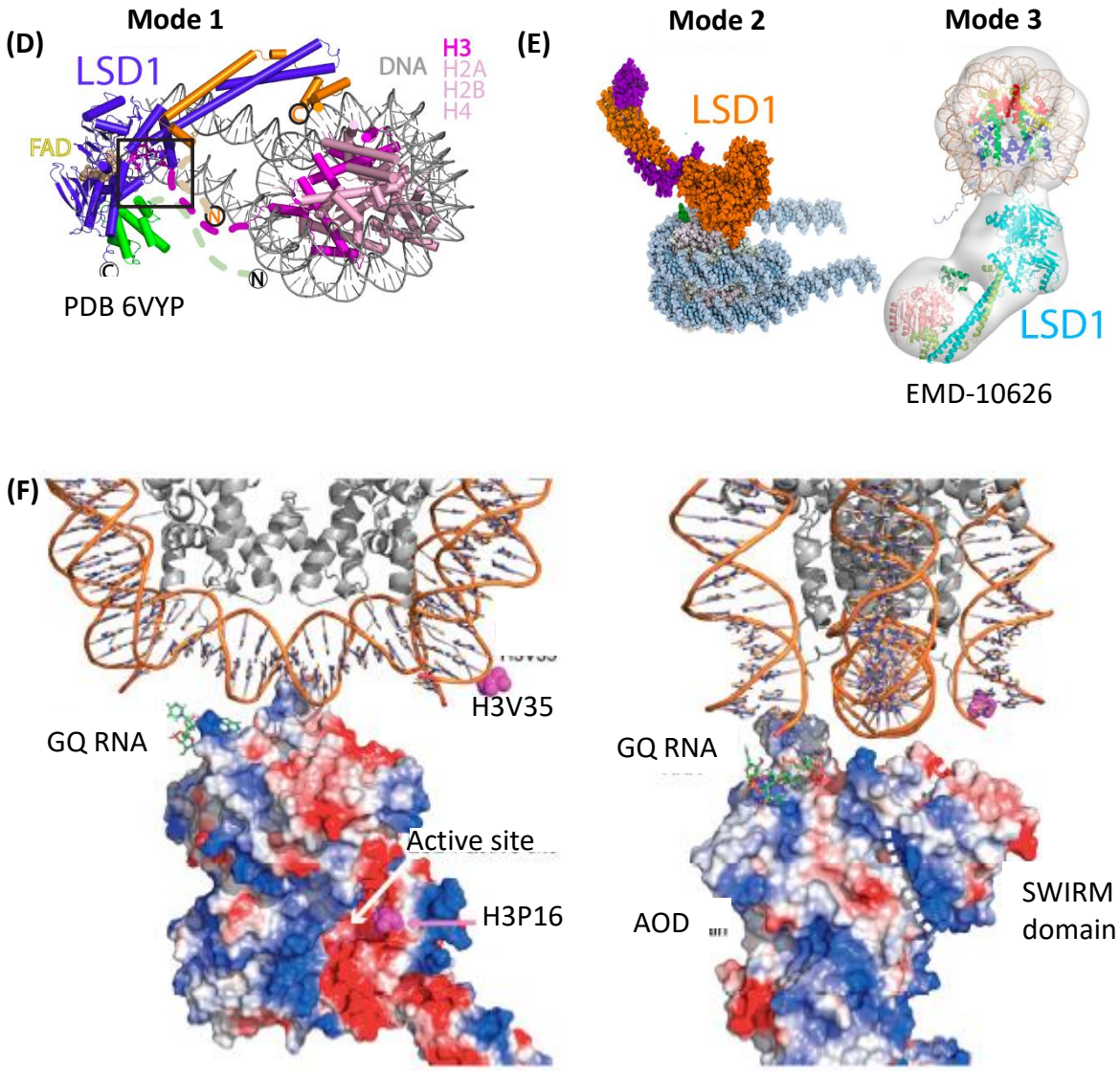


Figure 20: G-quadruplex structure, locations of RNA binding to LSD1 and LSD1-CoREST modes of nucleosome engagement.
(A) G quartet, the building block of G quadruplex, formed by the association of 4 guanines. **(B)** Single and stacked parallel stranded G quadruplex structures. **(C)** RNA binding regions mapped to LSD1-CoREST structure. The primary GQ RNA cross linking location is in SWIRM domain (aa 227–251, 210–216), colored in hot pink (rest of SWIRM domain is in green color). A minor LSD1-RNA adduct region exists at AOD (aa 527-550) (Cyan). A ss RNA bound adjacent to GQ RNA binding site, at the AOD of LSD1 identified via X ray crystallography is colored in red (RNA sequence- UUAGG). In the structure rest of LSD1 is in blue color and CoREST bound to LSD1 is colored in orange. (PDB ID: 4XBF) (60) **(D)** LSD1-CoREST-Nucleosome structure with IDRs clustered near the active site. The LSD1 in this structure is colored in blue and green, CoREST is colored in orange and

H3 N terminal tail, N terminus of LSD1 and Liner region of CoREST is in magenta, green and orange respectively.(44) **(E)** Two alternative orientations where LSD1 contacts the nucleosome as determined by X ray crystallography (mode 2, LSD1 in orange, CoREST is in magenta and nucleosome is in grey) (59) and Cryo-EM (mode 3, LSD1 is in Cyan, nucleosomal DNA is in beige) (44). **(F)** Charged LSD1 surface positioned as seen from the Mode 3 of nucleosome binding (Cryo-EM structure). Basic GQ RNA binding site (blue) interacts with the acidic phosphate backbone of nucleosomal DNA (orange) (44).

TERRA forms DNA/RNA hybrids *in vitro* and *in vivo* and acts as a negative regulator of telomere length in human cells. Similar to other lncRNA, TERRA participates in fine regulation of cell biology. They localize to telomeres, remain associated with their parental chromatin and is important for normal telomere function, regulating telomere length, heterochromatinization, and telomerase activity while acting as a negative regulator of telomerase. (94) (95) Genome instability and cellular senescence are promoted by the altered telomeric length associated with the changes in TERRA expression level. Moreover, TERRA G-quadruplexes play an important role in protecting telomere ends by acting together with heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) that bind G quadruplex loops and facilitate telomere capping. (91)(96)

In addition to interacting with helicases, DNA replication proteins, telomerase components, and several other critical epigenetic regulators such as methyltransferase PRC2 (Polycomb Repressive Complex 2), TERRA interacts with demethylase enzyme LSD1 and recruits LSD1 to telomeres.(60)(97)(98) Previously it has been shown that LSD1 binds UUAGGG repeat array of TERRA transcripts through its SWIRM and Amino oxidase domains. (99) I hypothesize that TERRA's higher order G quadruplex structure and its dynamic physical properties such as length and number of G repeats may confer

selectivity to enhance multivalent TERRA-protein interactions. TERRA interactions with LSD1 help to activate the telomeric DNA damage response (DDR) pathway, facilitating non homologous end joining (NHEJ) of uncapped telomeres, although its mechanism of activation has not yet been determined. Direct interaction of TERRA with LSD1 occurs when TERRA levels are increased due to loss of TRF2 (Telomeric Repeat Factor 2), a component of shelterin protein that protects telomeres from DNA repair mechanism. Stimulated by the TERRA-LSD1 interaction, the nuclease activity of double strand break repair protein MRE11 trim the 3' G overhangs at uncapped telomeres. (100) Altogether these suggest the importance of G quadruplex forming TERRA RNA binding to LSD1 at dysfunctional telomeres and its role in recruiting chromatin modifying proteins to modulate heterochromatinization at chromosome ends.

Previous studies have shown that LSD1 can specifically recognize and has strong binding preference (K_d = 96 nM) to stacked parallel stranded intramolecular G quadruplex structure (GG[UUAGGG]$_8$UUA) that represents minimal TERRA RNA, than single GQ RNA ([UUAGGG]$_4$U), GQ DNA ([TTAGGG]$_4$T), or an unstructured single-stranded RNA demonstrating that LSD1 can distinguish between different nucleic acid types (DNA Vs RNA) as well as structured versus unstructured RNA.(60)(100) Independently identified by cell-based and in vitro cross-linking mass spectrometry studies, the primary binding site of GQ RNA (GG[UUAGGG]$_8$UUA) binding to LSD1 exists within the SWIRM domain (residues 227−251, 210−216) and a minor adduct region exists in the amino oxidase domain (aa 527-550) (Figure 20-C). (60)(101) Identified by X-ray crystallographic studies, the ssRNA 5'-UUAGG-3' binding region is adjacent to the GQ

binding site in SWIRM/AOD. Furthermore, measured by the catalytic activity of LSD1 in presence and absence of GQ-forming RNA ($(UUAGGG)_4U$), RNA acts as a potential noncompetitive inhibitor of LSD1-catalyzed demethylation of peptide substrate.(60) However, it is unclear whether the TERRA RNA can inhibit LSD1 catalyzed demethylation on nucleosomes.

Recently discovered LSD1-CoREST structures in complex with nucleosome (5 Å - ~26 Å) shows three relative modes of nucleosome recognition (Figure 20-D, E). Interestingly, modes 2 and 3 suggest that previously mentioned RNA binding site coincides with an LSD1-nucleosome binding interface (Figure 20-E, F, Figure 25). (44) (59) The noncompetitive inhibition of LSD1 activity on $H3K4me^2$ peptide substrate by RNA can be explained by the fact that the observed GQ RNA binding site overlaps with the discovered nucleosome interaction site. (60)(101)

Here in the first half of this chapter, PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation) study was used to identify RNAs that preferentially interact with LSD1 across the human transcriptome. Using quantitative western blot studies, I show how PARCLIP identified G rich GQ forming RNAs mask nucleosome recognition and preferentially inhibit LSD1 demethylation on nucleosomal substrate over other RNA structures and un-structured RNAs, suggesting an RNA GQ structure-based effect on the LSD1 nucleosome engagement and regulatory properties of LSD1 function.

3.1.2 Results

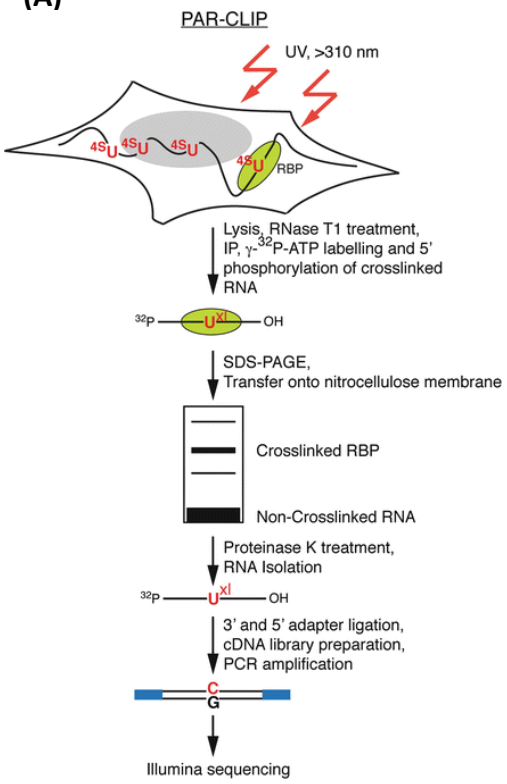
3.1.2.1 LSD1 Bind PAR-CLIP Identified GQ RNA TERRA, FAM57B and MYO1B

In order to broadly understand the relationship between RNA, chromatin and gene architecture as well as to provide insight into how RNA influences the function of chromatin remodeling enzyme complexes like LSD1-CoREST, photoactivatable ribonucleoside crosslinking immunoprecipitation (PARCLIP) was performed to characterize the RNA-LSD1 networks that occur in HEK293 cells (Figure 21-A). PARCLIP study was used to identify RNAs that preferentially interact with LSD1 across the human transcriptome. Analysis of PARCLIP data showed that 3'-UTR transcripts are enriched over two-fold in both CoREST and LSD1 datasets. In addition, microRNAs and lncRNA transcripts are also enriched among the LSD1 dataset. Within the PARCLIP identified transcripts, G rich regions were located in pre-mRNAs and 3' UTRs. In the intronic regions of FAM57B, MYO1B, and EPHB1 gene transcripts, three GQ forming motifs were identified. I performed CD spectroscopy to monitor the formation of GQ RNA structures in the PARCLIP identified G rich transcript TERRA as well as G rich FAM57B and MYO1B found in the intronic regions of gene transcripts (Figure 21-B). Characterized by the presence of a peak at 263 nm and trough at 240 nm, all three G rich RNAs (TERRA, FAM57B and MYO1B) form parallel stranded propeller type GQ structures in the presence of potassium (K^+) and sodium (Na^+) ions. (88)(102)(103) (Figure 21-B)

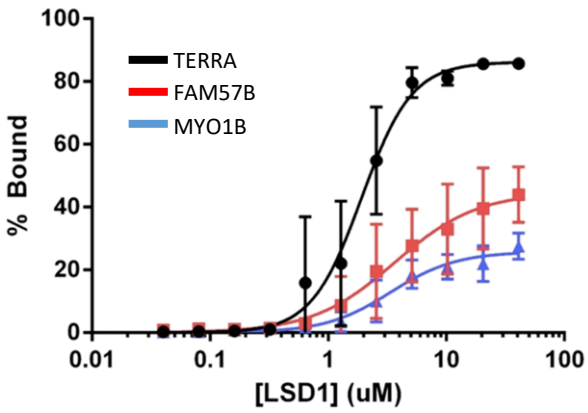
From previous studies, it is known that LSD1 RNA binding is dependent upon the RNA's ability to form a GQ RNA conformation and that GQ RNA preferentially bind LSD1–CoREST.(60) Therefore, in order to determine the affinities between LSD1 and

PAR-CLIP identified GQ RNAs TERRA, FAM57B and MYO1B, Electrophoretic Mobility Shift Assays (EMSA) were performed. EMSA experiment determined that GQ RNAs TERRA, FAM57B, and MYO1B bind with LSD1 with comparably similar affinities (Figure 21-C).

(A)



(C)



RNA	K_D(μM)	B_{max}
TERRA	1.9±0.18	86.2
FAM57B	3.7±1.2	44.8
MYO1B	3.4±0.6	25.8

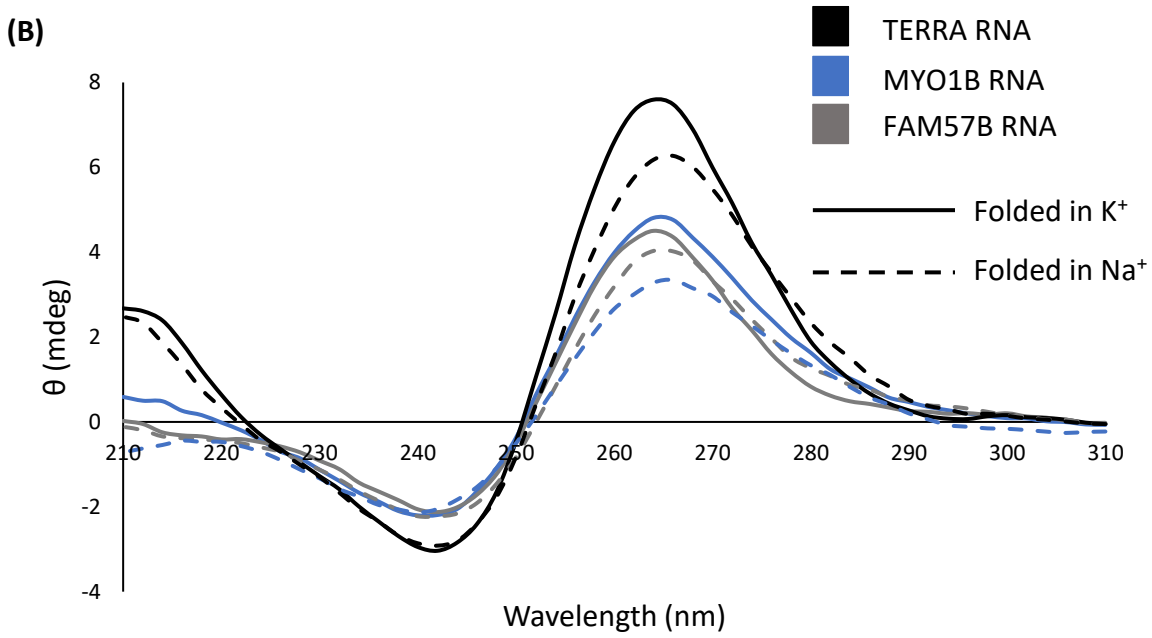
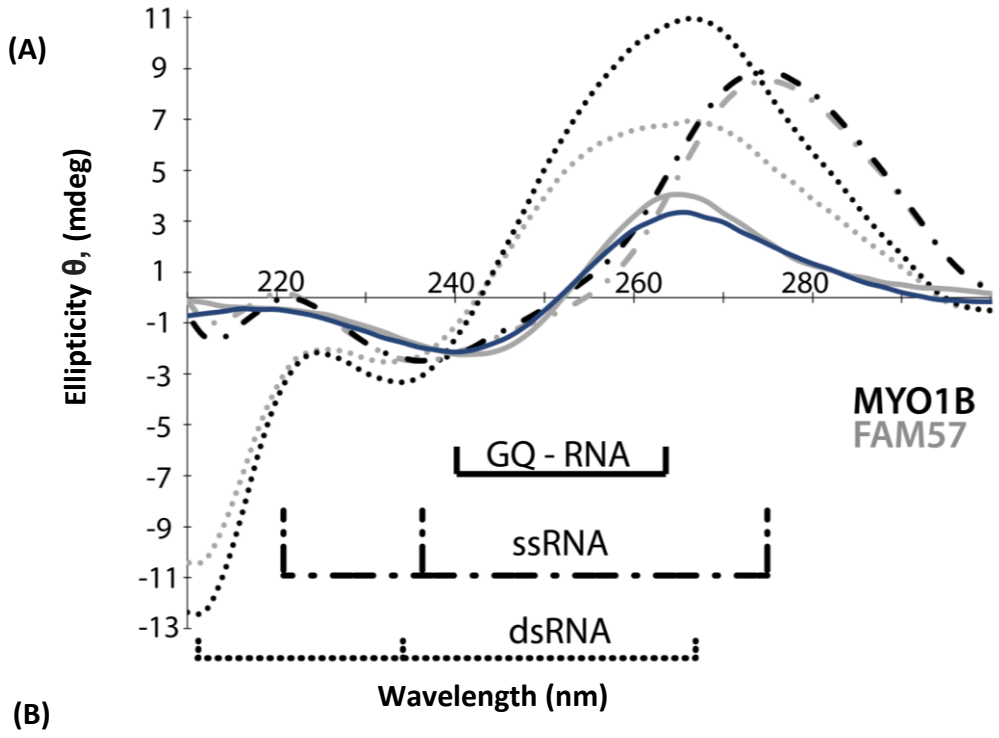


Figure 21: LSD1-CoREST bind PAR-CLIP identified GQ motifs FAM57B, MYO1B, and TERRA.
(A) PARCLIP workflow **(B)** Circular dichroism spectroscopy demonstrates that parallel stranded GQ RNA structures are formed by PARCLIP identified TERRA, MYO1B, FAM57B RNAs folded in presence of K^+ and Na^+ containing buffers. Consistent with previous studies, parallel propeller type GQ formation is characterized by a peak at 263nm and trough at 240 nm. (88)(102)(103)
(C) Determined by Electrophoretic Mobility Shift Assays (EMSA) LSD1 bind to PAR-CLIP identified GQ RNAs TERRA, FAM57B and MYO1B with comparably similar affinities but variable maximum binding.

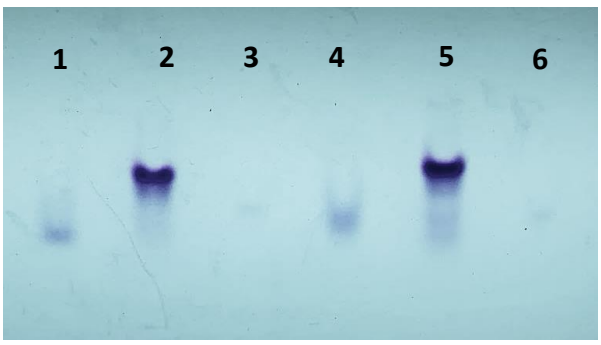
3.1.2.2 RNA Inhibits LSD1 Catalyzed Demethylation in a Structure Dependent Manner

To examine how different RNA structures affect catalytic activity of LSD1 in presence of its actual H3K4me^2 nucleosomal substrate, I used GQ RNA, similar length double stranded (ds) RNA and single stranded (ss) RNA. PARCLIP identified GQ forming RNAs TERRA (51mer), FAM57B (28mer) and MYO1B (33mer) were folded into GQ structures in presence of K^+ as well as double stranded (ds) MYO1B, ds FAM57B were

annealed in 10 mM Tris–HCl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$ by combining equimolar amounts of sense and antisense strands. As a control, single stranded (ss) PolyU RNA was treated similarly as GQ RNA using the same folding protocol. The folded and annealed RNAs were resolved using 8% poly acrylamide Native PAGE gel (Figure 22 B) to confirm the intact GQ, ds and ss structure formation. CD spectroscopy was used to further confirm the RNA topologies using characteristic peaks and troughs for GQ, ds and ss RNA (Figure 22 A).(104)



(B)

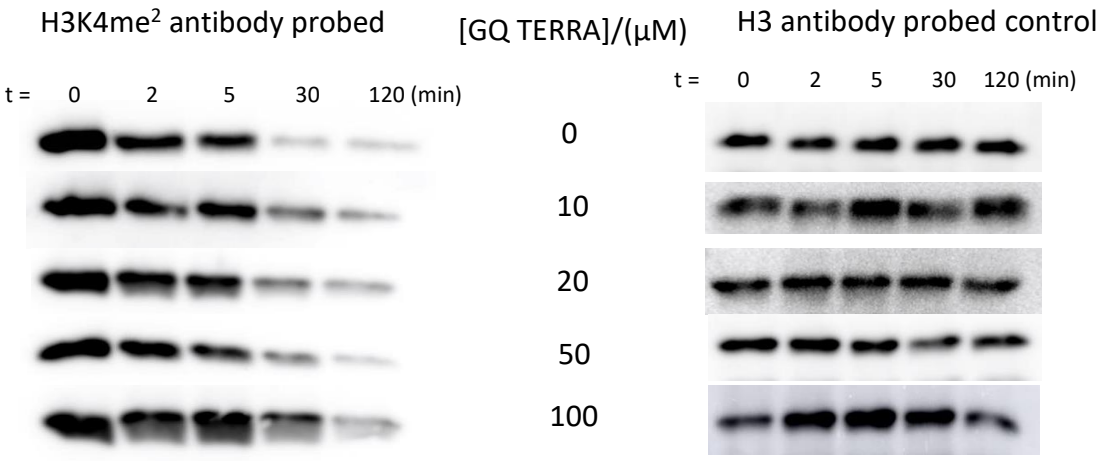


1. GQ FAM57B RNA
2. ds FAM57B RNA (double stranded)
3. ss RNA (antisense FAM57B)
4. GQ MYO1B RNA
5. ds MYO1B RNA
6. ss MYO1B RNA (antisense MYO1B)

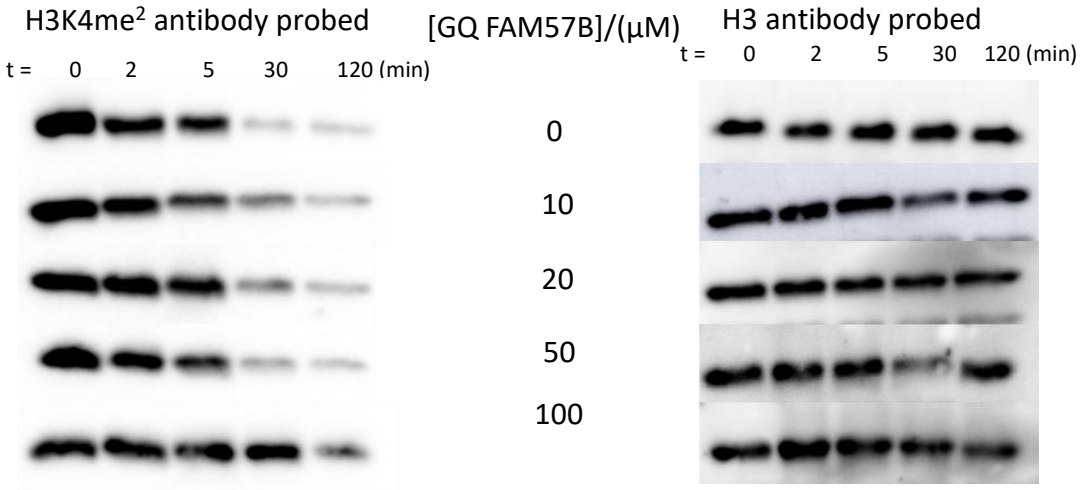
Figure 22: Different topologies of GQ RNA, single stranded (ss) RNA, and double-stranded (ds) RNA structures monitored using Circular Dichroism (CD) spectroscopy. **(A)** Parallel propeller type GQ formation in MYO1B and FAM57B is characterized by a peak at 263nm and trough at 240 nm. RNA:RNA duplex in ds MYO1B and ds FAM57B is characterized by a positive band at 265 nm and negative peak at 210nm. Ss RNA is characterized by the presence of positive peak at 275 nm and a trough at 235 nm. **(B)** FAM57B and MYO1B RNA duplex formation analyzed by Native PAGE gel electrophoresis. RNA samples were resolved using 8% polyacrylamide, 1xTHE, 50 mM NaCl Gel.

Quantitative western blot assays under single turnover conditions were performed in presence of GQ TERRA, GQ FAM57B, GQ MYO1B, ds FAM57B, ds MYO1B and ss PolyU RNA to test the ability of different RNA structures to inhibit demethylation in presence of nucleosomal substrate (Figure 23). Relative kinetics ($k_{observed}$) of the LSD1 catalyzed demethylation reaction was obtained from a series of time course assays measuring the fraction of dimethylated nucleosomes quantitated using H3K4me^2 specific antibody relative to the amount of H3 in each lane quantified using H3 specific antibody (Figure 23, 24, 25). Western blot study shows that PARCLIP identified G4 repeat RNAs preferentially inhibit LSD1-catalyzed demethylation on H3K4me^2 nucleosome substrates (IC_{50} of ~15 µM), whereas a similar sized double stranded RNA inhibits about twice lower than GQ RNA (IC_{50} of ~30 µM) while similar sized single stranded RNA weakly inhibits demethylation on nucleosomes nearly 10-fold lower than GQ RNA (IC_{50} ~150 µM) suggesting an RNA structure-based inhibition of LSD1 activity (Figure 26, Table 3).

(A)


H3K4me2 antibody probed
[GQ TERRA]/(μM)
H3 antibody probed control
t = 0 2 5 30 120 (min)
t = 0 2 5 30 120 (min)
0
10
20
50
100

(B)


H3K4me2 antibody probed
[GQ FAM57B]/(μM)
H3 antibody probed
t = 0 2 5 30 120 (min)
t = 0 2 5 30 120 (min)
0
10
20
50
100

(C)

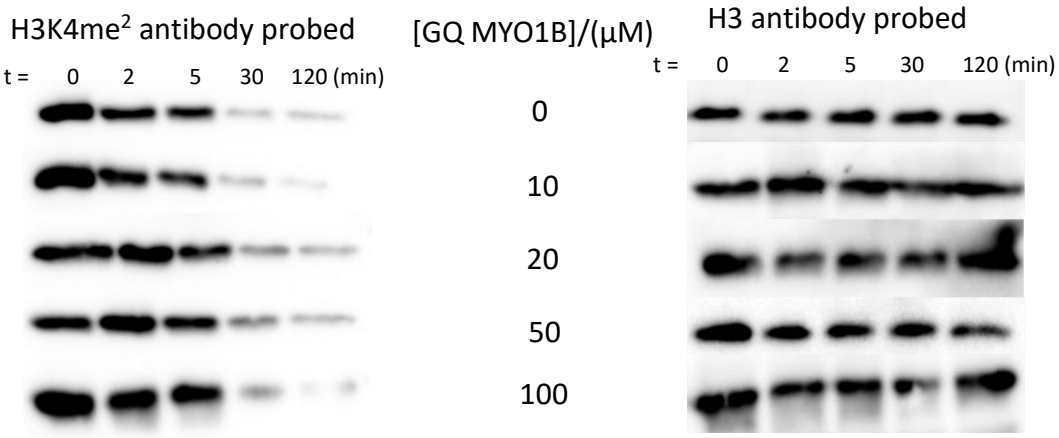

H3K4me2 antibody probed
[GQ MYO1B]/(μM)
H3 antibody probed
t = 0 2 5 30 120 (min)
t = 0 2 5 30 120 (min)
0
10
20
50
100

Figure 23: Representative western blot images for LSD1 demethylase assay on H3K4me^2 nucleosomes with increasing concentrations of GQ RNA:
(A) GQ TERRA RNA 0, 10, 20, 50, 100 µM **(B)** GQ FAM57B RNA 0, 10, 20, 50, 100 µM **(C)** GQ MYO1B RNA 0, 10, 20, 50, 100 µM.

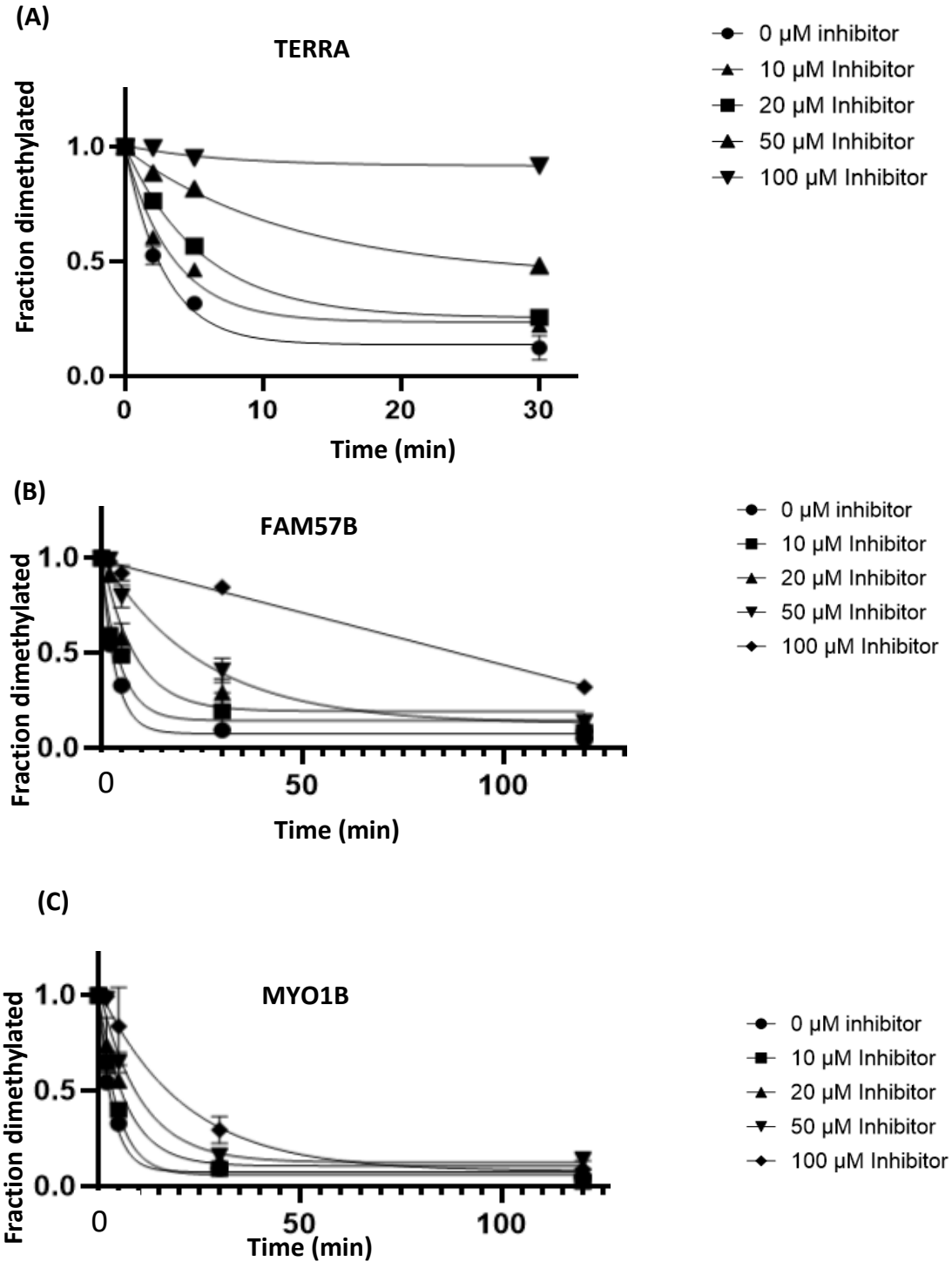
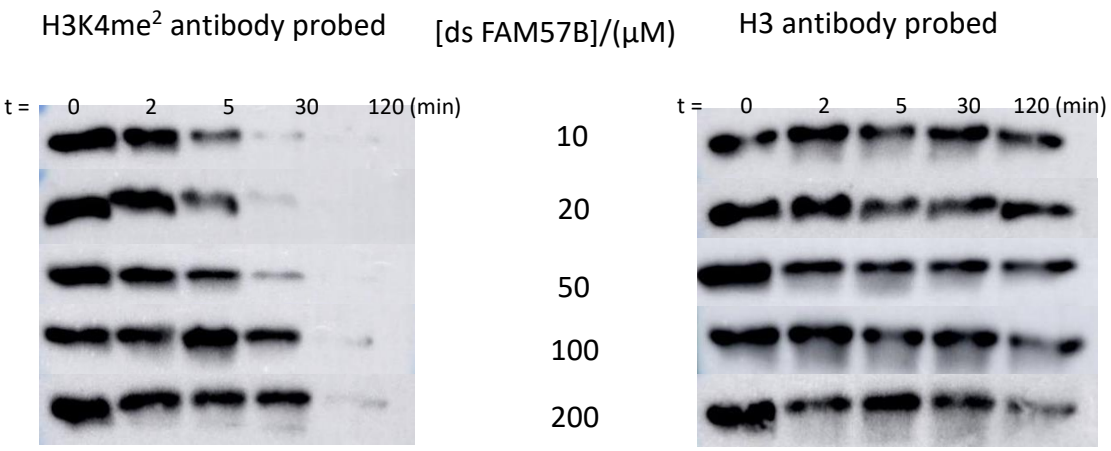
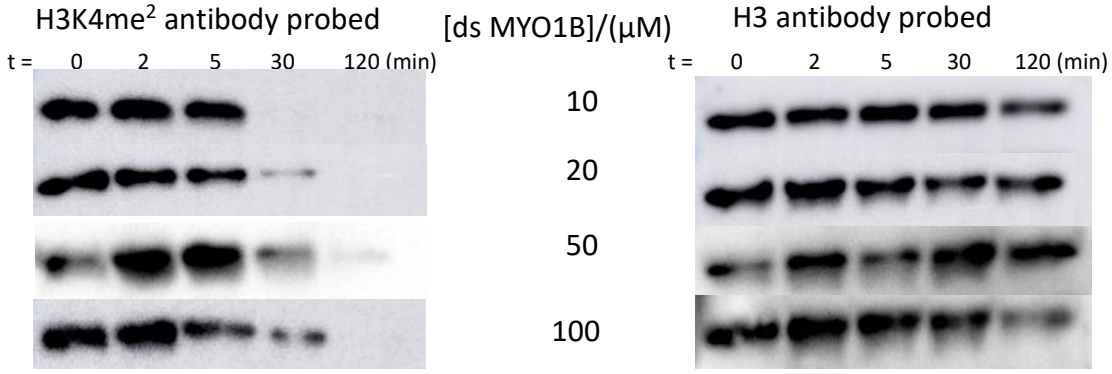


Figure 24: LSD1 catalytic activity inhibition by GQ RNA. Quantification of residual dimethylation on western blots at each time point. N ≥ 2 **(A)** GQ TERRA RNA **(B)** GQ FAM57B **(C)** GQ MYO1B RNA.

(A)



(B)



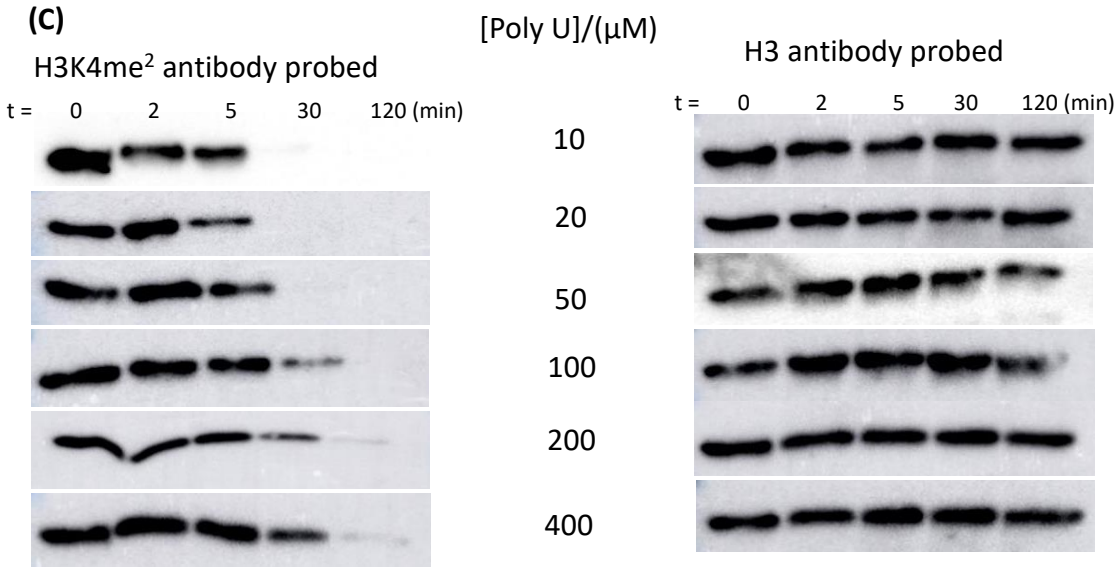
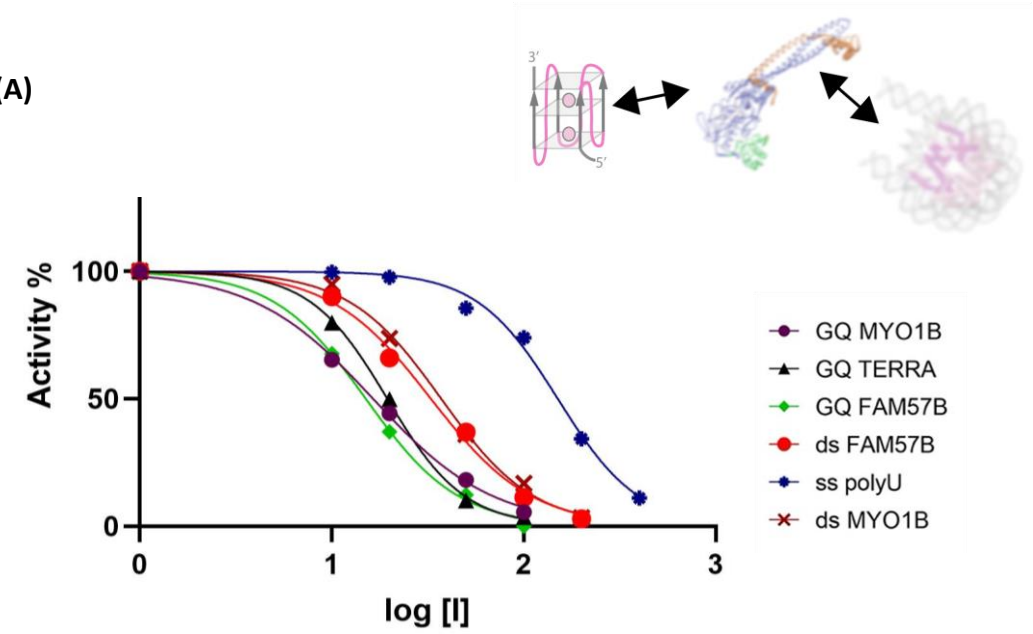


Figure 25: Representative western blot images for LSD1 demethylase assay on H3K4me2 nucleosomes with increasing concentrations of ds RNA and ss PolyU RNA.
(A) ds FAM57B RNA 10, 20, 50, 100, 200 µM **(B)** ds MYO1B RNA 10, 20, 50, 100 µM **(C)** ss PolyU RNA 10, 20, 50, 100, 200, 400 µM.

(A)



(B)

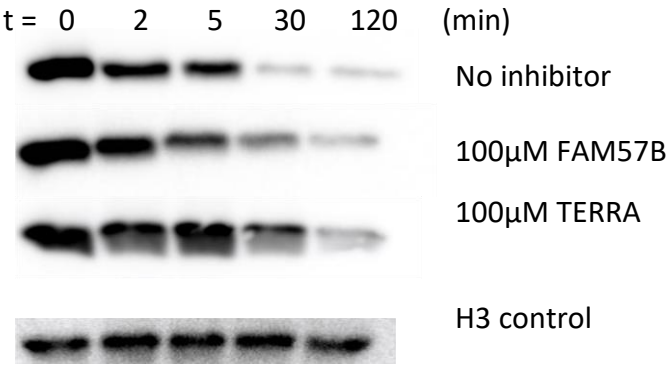


Figure 26: RNA structure-based inhibition of LSD1 catalytic activity on H3K4me^2 nucleosome demethylation.
(A) Activity% versus log [I] graph of various structured RNAs (GQ TERRA, GQ FAM57B, GQ MYO1B vs double-stranded (DS) MYO1B, DS FAM57B) and unstructured single stranded (ss) PolyU RNA inhibition of LSD1 catalytic activity on H3K4me^2 nucleosomes.
(B) Representative western blots for inhibition assay of mono-nucleosome demethylation at varying concentrations of RNA.

Table 3: Half maximal inhibitor concentration (IC_{50}) values of GQ RNA, ds RNA, ss RNA inhibition of LSD1 catalytic activity on demethylation of H3K4me^2 nucleosomal substrate.

	RNA	IC_{50}(μM)
GQ forming	TERRA	~19
RNA	FAM57B	~15
	MYO1B	~16
ds RNA	ds FAM57B	~33
	ss MYO1B	~37
ss RNA	Poly U	~150

3.1.3 Discussion

Understanding the interplay of LSD1 with lncRNAs and pre-mRNA transcripts can help facilitating discovering the molecular mechanisms between additional chromatin-based proteins and RNA that impact genome integrity. In this study, I examine the interactions of LSD1 with different RNA structures, with a focus on the GQ forming telomeric repeat containing RNA (TERRA containing UUAGGG repeats). Through biochemical approaches, I address what are the determinants of RNA-LSD1 binding specificity and how does RNA influence the active site of LSD1 in the context of nucleosome demethylation.

PARCLIP experimental results show that LSD1 binds more than 800 RNAs with a bias towards the G rich transcripts. PARCLIP identified G rich transcripts MYO1B, FAM57B and TERRA can fold into stable G quadruplexes in presence of potassium (K^+) and that these GQ forming RNAs can bind LSD1 with comparable affinities. Moreover, providing insight into the functional consequences of GQ RNA binding, PARCLIP identified GQ forming RNAs inhibit LSD1 activity on nucleosomal substrate (GQ FAM57B

RNA IC_{50}= 15 µM, GQ MYO1B RNA IC_{50}= 16 µM, GQ TERRA RNA IC_{50}= 19 µM) stronger than similar length double stranded RNA that does not form GQ structures (ds FAM57B RNA IC_{50}= 33 µM, ds MYO1B RNA IC_{50}= 37 µM) as well as single stranded unstructured RNA that weakly inhibits LSD1 catalytic activity (ss poly U RNA IC_{50}= 150 µM) suggesting a RNA structure based effect in binding and inhibiting LSD1's catalytic activity on nucleosomes (Figure 26).

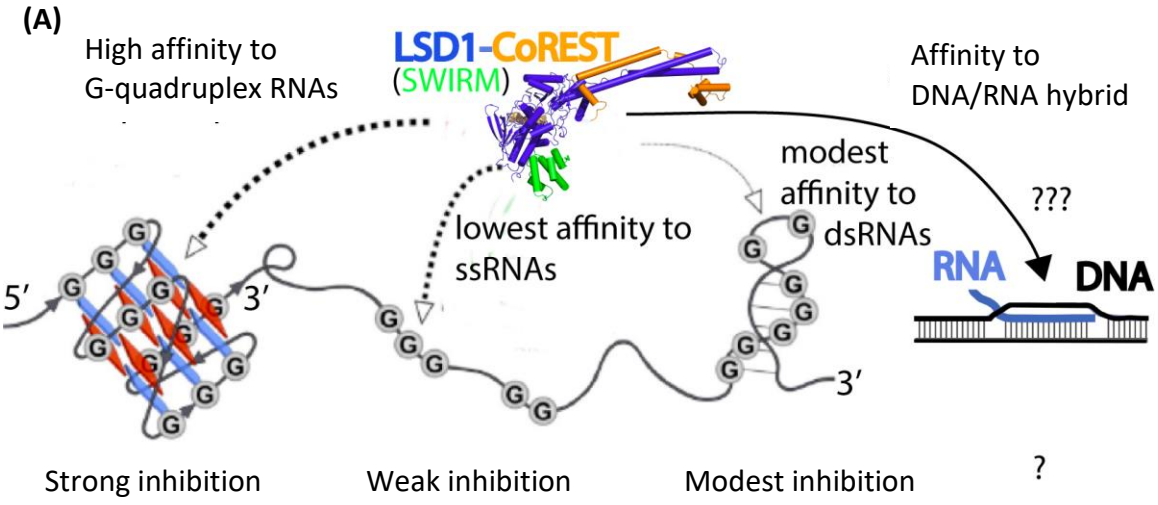


Figure 27: Summary of RNA structure based RNA:LSD1 interactions and TERRA's proposed role in regulating chromosomal ends.

(A) EMSA binding and inhibition of LSD1 nucleosome demethylation assays demonstrate that GQ RNA specifically interact with LSD1 with high affinity whereas ds RNA shows modest affinity and ssRNAs show lowest affinity to LSD1. It is yet to discover the LSD1's affinity to R-loops structures. **(B)** Suggested role for TERRA in the regulation of chromosomal ends. G-rich RNAs like TERRA recruit epigenetic regulators and form DNA-RNA hybrids (R loops) at telomeres to maintain and regulate telomeric chromatin structure.

Collectively PARCLIP, EMSA, and western blot approaches demonstrate that LSD1 specifically recognizes G quadruplex forming RNA. LSD1-CoREST-nucleosome structures (5 Å - ~20 Å) shows three relative modes of nucleosome recognition, two of which (model 2 and 3 in Figure 28) coincide with LSD1-RNA binding interface. This supports previous findings that shows GQ forming TERRA RNA is a noncompetitive inhibitor of LSD1 activity on peptide substrate. (59)(60)(101) Here I examined the role of RNA structure in the context of LSD1's activity on nucleosomal substrate. My work demonstrated that GQ RNA preferentially inhibits LSD1 catalyzed H3K4me^2 nucleosomal substrate demethylation suggesting an RNA structure-based effect on the LSD1-CoREST function defining the structural mechanisms associated with LSD1 activity. These findings validate previous work (60) and provide new insights into how LSD1 recognizes and directly interacts with GQ forming TERRA at telomeres. Finally, this work suggests a molecular mechanism for how the function and localization of chromatin remodeling enzyme LSD1 can be modulated through RNA structure-dependent manner by GQ TERRA RNA. Logical next step experiments would be to biochemically examine the interplay of LSD1 and TERRA in telomere maintenance, as it relates to the nuclease activity of the double strand break repair protein MRE11A (MRE11), or to how the LSD1

and TERRA interaction impacts the alternative lengthening of telomere (ALT) pathway in telomerase-free cancer cells. (99) (105) From this biochemical study, I propose a molecular mechanism for how the function and localization of a chromatin remodeling enzyme can be modulated through an RNA structure-dependent manner. Furthermore, I propose that TERRA's repetitive G-rich and dynamic structure serves to recruit regulators such as LSD1 to maintain and regulate telomeric chromatin structure (Figure 27-B).

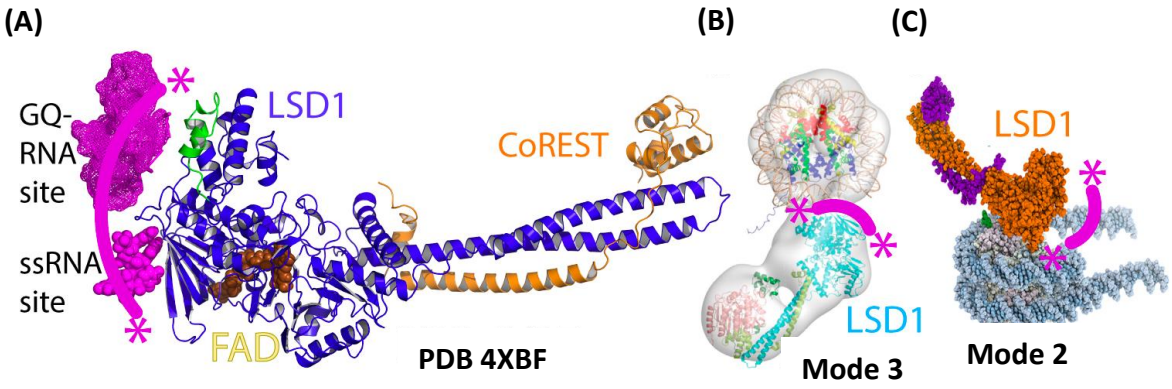


Figure 28: Structural model of LSD1-RNA interaction.
(A) Identified RNA binding model based on X-ray crystallography and XL-MS data. (PDB ID 4XBF) **(B)(C)** RNA binding site overlaps with observed nucleosome binding in alternative nucleosome recognition mode 2 and 3. (44)(59)(60)(101)

3.1.4 Materials and Methods

3.1.4.1 G-Quadruplex (GQ) RNA Folding and RNA Duplex Annealing

TERRA 8x RNA (51 mer, 8 repeats UUAGGG), FAM57B RNA (28 mer), MYO1B RNA (33 mer), Poly U RNA (32 mer), antisense FAM57B and antisense MYO1B RNA pallets were purchased from IDT™ and desalted using Bio-Rad desalting spin columns. TERRA 8x, FAM57B, MYO1B RNA were diluted in 10 mM Tris–HCl pH 7.5, 100 mM KCl buffer and were folded using a standard GQ RNA folding protocol (2 min at 95 ˚C, 5 min at 85 ˚C, 5 min at 75 ˚C, 5 min at 55 ˚C, 15 min at 37 ˚C, and then placed on ice). (60)

For RNA duplex formation, FAM57B with antisense FAM57B and MYO1B RNA with antisense MYO1B were annealed separately by mixing equimolar amounts of RNA oligonucleotide solutions in annealing buffer (10 mM Tris–HCl pH 7.5, 50 mM NaCl 5 mM $MgCl_2$). The solution was heated at 94 °C, 4 minutes and gradually cooled down to room temperature for duplex formation.

PolyU RNA (ss) was treated the same way as GQ RNA. It was diluted in 10 mM Tris–HCl pH 7.5, 100 mM KCl buffer and was folded using a standard GQ RNA folding protocol (2 min at 95 ˚C, 5 min at 85 ˚C, 5 min at 75 ˚C, 5 min at 55 ˚C, 15 min at 37 ˚C, and then placed on ice).(60)

3.1.4.2 Native Page Gel

Folded and annealed RNA (GQ FAM57B RNA, ds FAM57B RNA, ss RNA (antisense FAM57B), GQ MYO1B RNA, ds MYO1B RNA, ss MYO1B RNA (antisense MYO1B)) were resolved in a 50 mM NaCl, 8% Acrylamide, 1xTHE native PAGE gel. The gel was run in

1xTHE, 50 mM NaCl running buffer at 50V in 4 ˚C. Native PAGE gel RNA bands were stained in toluidine blue for visualization.

3.1.4.3 Circular Dichroism (CD) Spectroscopy

CD spectra for folded GQ RNA, single stranded (ss) antisense RNA, and annealed double stranded (ds) RNA were recorded at room temperature on a Chirascan^{TM} V100 CD spectrophotometer with a 1 mm cell, 1nm band width, 4s per point, and 0.1ms timed intervals. Spectra from 310–210 nm were averaged over three scans, and background from a matched buffer-only sample was subtracted.

3.1.4.4 LSD1-CoREST Inhibition Assays Using Different RNAs as Inhibitors in Presence of H3K4me^2 Nucleosomal Substrate.

Different concentrations of GQ TERRA RNA (0, 10, 20, 50, 100 μM), GQ FAM57B RNA (0, 10, 20, 50, 100 μM), GQ MYO1B RNA (0, 10, 20, 50, 100 μM), ds FAM57B RNA (0, 10, 20, 50, 100, 200 μM), ds MYO1B RNA (0, 10, 20, 50, 100 μM), ss PolyU RNA (0, 10, 20, 50, 100, 200, 400 μM) were incubated with the 2μM ΔN LSD1- CoREST containing reaction mixture in absence of the nucleosome substrate, at 4 °C for 60 minutes. Demethylation reactions were initiated by adding 100nM nucleosomal substrate into the reaction mixture. Aliquots of 10 μL were withdrawn at 0, 2, 5, 30, 120 minutes time points and the reactions were quenched using Laemmli dye followed by boiling for 2 minutes to stop the demethylation reaction. The assay products were resolved by 20% SDS-PAGE gel for 90 minutes at 200V.

Protein bands in SDS gel were transferred to immunoblot PVDF membranes and blocked with 4% fat free milk in PBS buffer. Corresponding blots were incubated with

anti-H3K4me^2 (EMD Millipore, #07-030), and anti-H3 (Abcam, #ab1791) specific primary antibodies overnight followed by goat anti-rabbit (HRP conjugate) secondary antibody for 1 h. The blots were visualized by chemiluminescence and analyzed using Amersham software. All demethylation and control experiments were performed in duplicate (N≥2). The degree of LSD1 catalytic activity inhibition was quantified using H3K4me^2 specific antibody signal relative to the amount of H3 in each lane quantified using H3 specific antibody. Dividing the H3K4me^2 antibody signal by the H3 antibody signal allowed for accurate quantitation of each data point. The H3K4me^2/H3 ratios were normalized at time zero and plotted as a function of time (minutes), and subsequently analyzed using nonlinear regression. Data were fitted to the equation [H3K4me^2] = [H3K4me^2]$_{t=0}$ e^{-kobst} and the determined rate constant (k_{obs}) values were evaluated with LSD1-CoREST concentrations. The $K_{1/2}$ and k_{max} parameters were determined based on the equation: k_{obs}= k_{max} [Enzyme]/([Enzyme]+$K_{1/2}$).

Project Contributions:

All LSD1-CoREST constructs purification, all western blot nucleosome demethylation assays, different RNA folding/annealing and their CD spectroscopy, all western blot LSD1 inhibition assays using different RNAs, Native PAGE gel analysis of RNAs, and relevant data analysis were performed by Dulmi Senanayaka. The PARCIP study, EMSA assay and relevant data analysis were performed by Dr. William Martin.

3.2 TERRA-LSD1 Condensates Promote R-loop Formation for Telomere Maintenance in ALT Cancer Cells

3.2.1 Introduction

In addition to binding with RNAs across the genome, LSD1 function is also regulated by R-loop/ DNA:RNA hybrid formation, linking LSD1 to G quadruplex RNAs and DNA:RNA hybrid structures. (106) Although G rich RNAs and R loops exist at telomeres in cancer cells, very little is known about the nucleation steps of these higher order interaction networks. In this part of the study, the interplay of LSD1 with the GQ forming telomeric repeat containing RNA (TERRA) at telomeres was further examined based on the hypothesis that highly structured G rich, GQ forming TERRA RNA-chromatin associated LSD1 interaction serves as an important nucleation step in the epigenetic regulation of the genome. Findings from this study reveal that at telomeres, LSD1 and TERRA mutually enrich each other in an enzyme activity independent manner to facilitate R loop formation for ALT telomere maintenance. This collaborative study suggests a new mechanism for how chromatin associated RNAs and histone modifiers act together to impact the chromatin environment.

Non-coding transcripts transcribed by genome physically associate with other RNAs, chromatin, chromatin modifying enzymes, and nucleosome remodeling factors, through RNA–RNA, RNA–DNA, and RNA–protein interactions thus playing important roles in regulating chromatin architecture, transcription, protein translation, and localization, as well as affect accessibility to genetic information. (107) These RNA hybridize with DNA to form DNA-RNA hybrid structures called R loops that can transiently form at promoters and enhancer regions and can play functional roles such

as pausing RNA polymerase II to repress transcription. (108) But R loop formation is deleterious and can cause genomic instability due to replication forks collisions or transcription factors associated with replication machinery. (109)

Transcribed from chromosome ends and most prominently generated during the ALT biological mechanism (see below), TERRA lncRNA is important for normal telomere function and is highly expressed in ~15% of cancers. TERRA regulates telomere maintenance and telomeric chromatin structure by protecting shortened telomeres and facilitating telomere DNA replication by increasing euchromatin formation. Chromatin associated RNAs such as TERRA not only act as scaffolds to recruit transcription factors and chromatin modifiers such as PRC2 and LSD1 to chromatin (110) (111) (112) but preferentially localize to short telomeres through the formation of R loops. (113) (114) In addition to interacting with histone modifying enzymes such as LSD1, TERRA interacts with a number of histone associated heterochromatin proteins that include: telomerase components, DNA repair factors, and higher order chromatin remodeling complexes (PRC2 and LSD1). (98)

Due to rapid proliferation and subsequent critical telomere shortening in each cell cycle, cancer cells elicit DNA damage response (DDR) that trigger cellular senescence to prevent tumorigenesis. (115) Therefore, to prevent premature senescence, cancer cells need to acquire replicative cellular immortality by upregulating telomerase or hijacking recombination mediated alternative lengthening of telomere (ALT) pathway. In absence of telomerase, cancer cells rely on maintaining telomere length via ALT pathway that uses homology directed DNA synthesis which require many DNA

recombination and repair proteins. (105) Upregulated and abundant in ALT cancer cells, TERRA contributes to telomere maintenance by forming R loops or DR-loops (a structure containing DNA-DNA and DNA-RNA hybrids) to promote DNA damage response (DDR) and activate homology directed DNA synthesis for telomere maintenance in ALT pathway. (114) (116) In fact, how and whether TERRA scaffold proteins in ALT pathway and the molecular mechanism underlying in ALT pathway is still poorly understood. However, it is believed that manipulating TERRA and its protein partners may serve as a promising treatment in ALT-positive cancers.

In this collaborative study, the proteins that localize at telomeres were screened in a TERRA dependent manner and we find that LSD1 contributes to ALT by interacting with TERRA. We proposed a novel function of LSD1 in mediating telomere maintenance independent of its H3K4me$^{1/2}$ demethylase enzymatic activity. The cell-based data show that TERRA is required for LSD1 localizing at ALT telomeres. Furthermore, upon LSD1 knockdown, telomere elongation signatures such as telomere clustering and telomeric DNA synthesis become impaired. It was found that G quadruplex forming TERRA RNA preferentially drives LSD1 phase separation in a TERRA structure and length dependent manner, and in turn, LSD1 enriches TERRA R loop stimulating proteins at telomeres, facilitating ALT mechanism. Taken together, this work presents a TERRA mediated LSD1 phase separation mechanism for R loop formation in ALT positive cancer cells.

3.2.2 Results

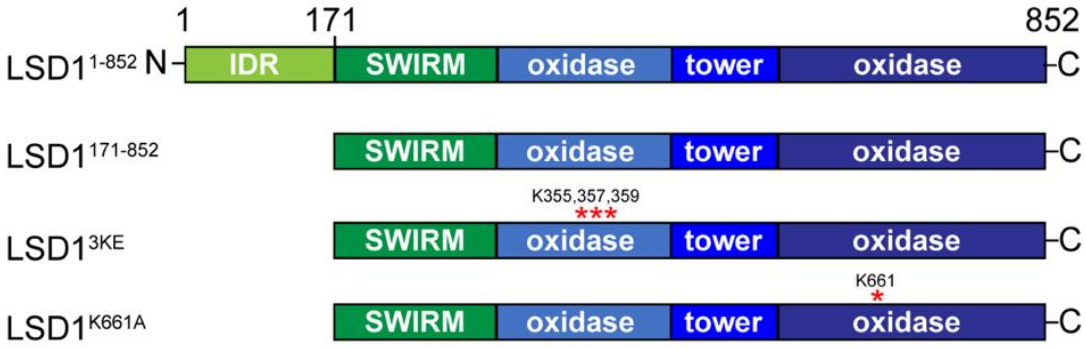


Figure 29: Schematic of full-length (LSD1 1-852), Δ N (LSD1 171-852), and mutant ($LSD1^{3KE}$, $LSD1^{K661A}$) LSD1

In this study, in order to test how LSD1 and critical mutations impact TERRA's role at telomers, the catalytically active ΔN LSD1 (171-852) lacking N terminal IDR region, FL LSD1 (1-852) with N terminal IDR, catalytic mutant LSD1 ($LSD1^{K661A}$) and LSD1 AOD triple mutant (LSD1 K355E, K357E, K359E or $LSD1^{3KE}$) that mutates a basic strip along the AOD were overexpressed and their effects in phase separation were examined (Figure 29). $LSD1^{3KE}$ mutant contains LSD1 DNA binding region mutation and it is known for largely reducing binding to extra nucleosomal DNA. (59) In addition, this DNA binding region may also coincide with LSD1 GQ RNA binding site on the surface of AOD. (60)

3.2.2.1 TERRA Promotes LSD1 Phase Separation In-Vitro

For quantitatively assessing TERRA driven LSD1 phase separation, LSD1 and TERRA phase behavior was examined in vitro by using purified LSD1-CoREST (complexed with CoREST for stability) and various synthesized repeats of UUAGGG that mimic TERRA RNA. Under experimental conditions, FL LSD1 alone did not phase separate, but the

addition of TERRA initiated LSD1 to undergo phase separation forming spherical condensates (Figure 30-A). According to the experimental-derived phase diagram generated by our collaborators, the longer the TERRA, the lower the LSD1 phase separation boundary (Figure 30-B, C, D). Longer 20X TERRA can phase separate FL LSD1 even at 5μM protein concentration (Figure 30-D). Since TERRA length is estimated to be 100-9000 bases long and contains 5'-(UUAGGG)-3' repeats, a longer TERRA is likely to be effective in driving phase separation even at very low protein concentrations on telomeres.

Since in vivo data showed that ΔN LSD1 (171-852/ IDR deletion mutant) is sufficient for LSD1 phase separation, we sought to test the phase separation properties in vitro. Interestingly, ΔN LSD1 (171-852/ IDR deletion mutant) appeared to be more efficient in forming condensates than FL LSD1 (Figure 30-B vs G), and even formed non-spherical condensates with 8X TERRA RNA (Figure 30-E), suggesting that IDR may play a regulatory role in shifting phase separation boundary to maintain condensate fluidity. To determine that phase separation is due to the liquid properties of the interaction, we monitored formation of ΔN LSD1-TERRA condensates fusion over time (Figure 30-H). To assess whether the LSD1 and TERRA localize in the droplets and examine the hydrodynamic properties of the LSD1-TERRA droplet condensates, LSD1 was fluorescently labeled (FITC-labeled LSD1) and Cyanin 3 (Cy3)- UTP labeled TERRA was transcribed. A FITC-labeled LSD1 and a Cy3 Uridine-labeled 12X TERRA were added at previously defined concentrations that gave rise to condensates and the fluorescence images were quantitated. We observed fluorescence images that were merged and

overlapped, suggesting TERRA and LSD1's uniformly mixed localization in droplets (Figure 30-I). Next, Fluorescence recovery after photobleaching (FRAP) assay, was used to measure the kinetics of diffusion. The FRAP results of partial fluorescent intensity recovered over time indicate that LSD1-TERRA condensates are not simple liquids but have complex material properties (Figure 30-J).

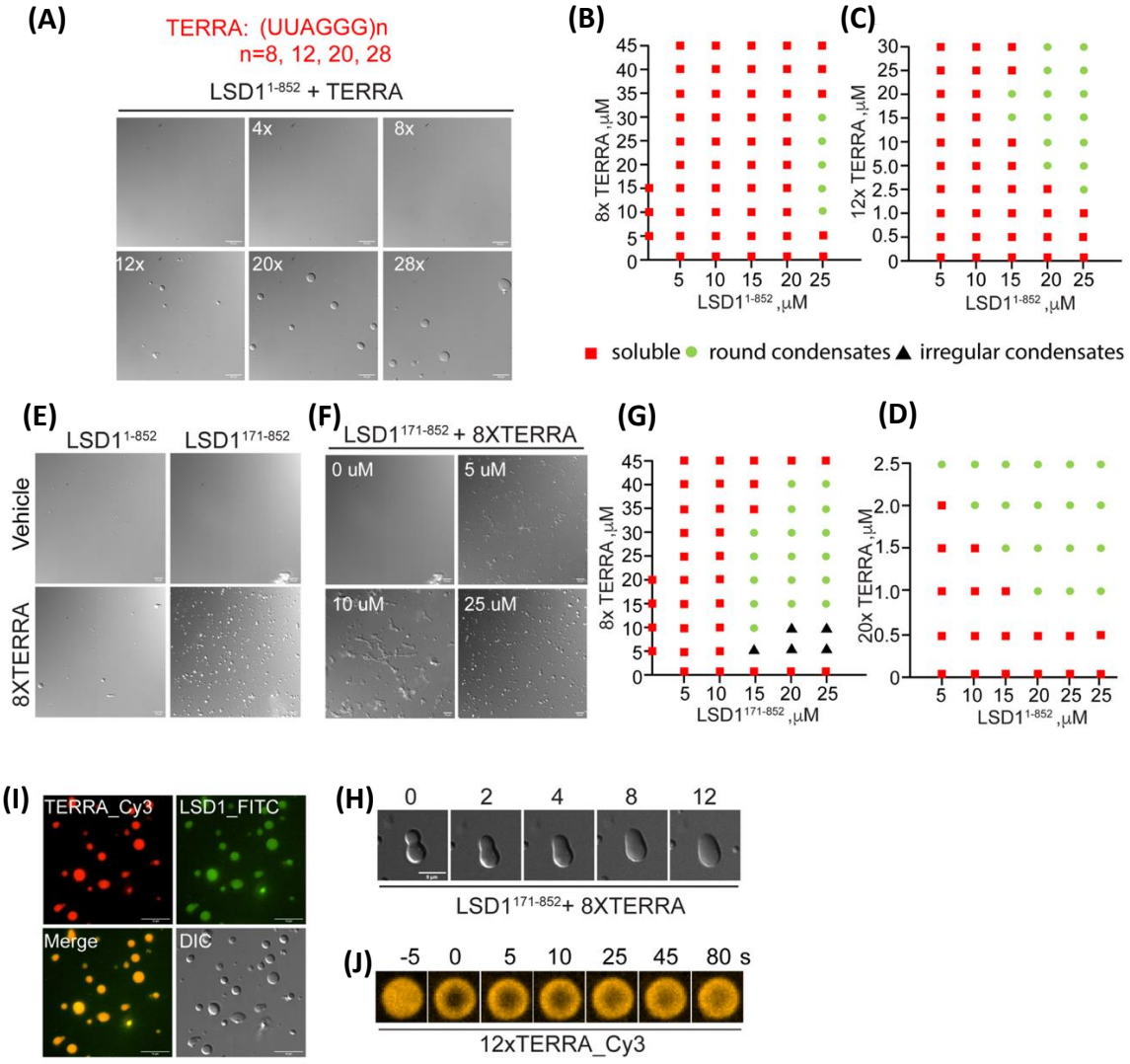


Figure 30: TERRA drives LSD1 phase separation in vitro.
(A) Representative DIC images of condensates formed by a mixture of purified LSD1 at 25uM and different UUAGG repeats as mimics of TERRA at 2.5uM. **(B, C, D)** Phase diagram of full-length LSD1 with the addition of 8x (B), 12x (C), 20x (D) TERRA. **(E-F)**

Representative DIC images and **(G)** phase diagram of LSD1171-852 with 8x TERRA. **(H)** Dynamic fusion of LSD1 condensates with each other and coalesce into a larger one. **(I)** Fluorescent images of Cy3-labeled TERRA and FITC-labeled LSD1. **(J)** FRAP assay of 12xTERRA_cy3 positive LSD1 condensates shows partial fluorescent intensity recovered over time after photobleaching.

3.2.2.2 TERRA-LSD1 Phase Separation Require LSD1 Nucleic Acid Binding Domain

Next, the effect of nucleic acid binding and demethylation on LSD1 phase separation behavior was assessed in vitro. The LSD1 AOD K661A mutant, which is often considered as a catalytically inactive mutant that directly affects the demethylation ability of LSD1, and nucleic acid binding mutant $LSD1^{3KE}$ were used for this study. Surprisingly LSD1 AOD K661A mutant maintains its phase separation properties with TERRA while nucleic acid binding mutant $LSD1^{3KE}$ totally abolishes LSD1's phase separation with TERRA. This suggests that specific nucleic acid binding to LSD1 AOD is necessary for the TERRA stimulated LSD1 phase separation but that LSD1's catalytic activity is not critical for phase separation (Figure 31).

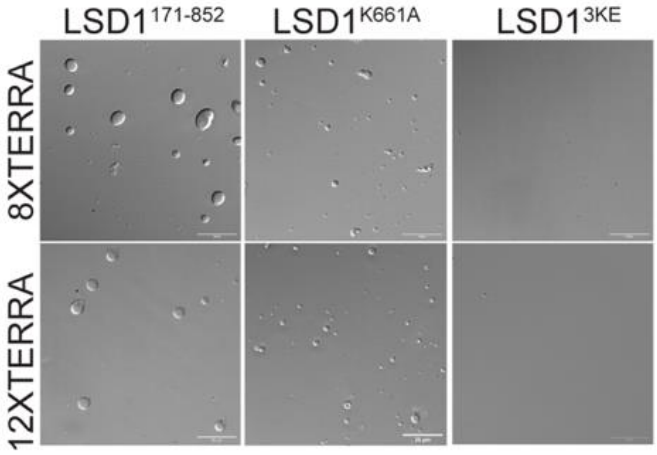


Figure 31:RNA binding is required for LSD1 phase separation and LSD1 functional rescue.

(A) Representative DIC images of LSD1^{3KE}, and LSD1 K661A with the addition of TERRA. LSD1^{3KE} totally abolished droplet formation with TERRA.

3.2.2.3 TERRA G-Quadruplex Structure is Important for LSD1 Droplet Condensate Formation.

With its G rich UUAGGG repeats, TERRA is prone to form GQ structures in vitro and in vivo. (89) (93) (102) My previous studies show that LSD1 preferentially binds GQ forming TERRA. This raises the question if TERRA driven LSD1 condensate formation is structure dependent. To test this, ss RNAs (anti FAM57B, anti MYO1B and PolyU) with similar lengths to 8X TERRA but with no known structure formation was used as control RNA. In addition, large non-coding RNaseP RNA which has a non-GQ forming but elaborate tertiary structure was used. (117) Microscopic images and turbidity measurements show that LSD1 phase separation is driven GQ RNA conformation dependently by TERRA RNA but none of the other non-GQ RNA (Figure 32 A, B).

For further assessing the TERRA GQ structure-based phase separation a small molecule (N-methylmesoporphyrin IX / NMM) that preferentially binds TERRA G quadruplex and inhibits the interaction between TERRA and its binding proteins was used. (118) By treatment of a low dosage of NMM before combining TERRA with LSD1 for phase separation assay, LSD1 transition into a more solid phase was observed, indicated by transitioning from spherical to irregular shape morphology. While a high dosage of NMM fully aggregated LSD1, indicating the importance of TERRA GQ structure in binding LSD1 promoting phase separation (Figure 32 C).

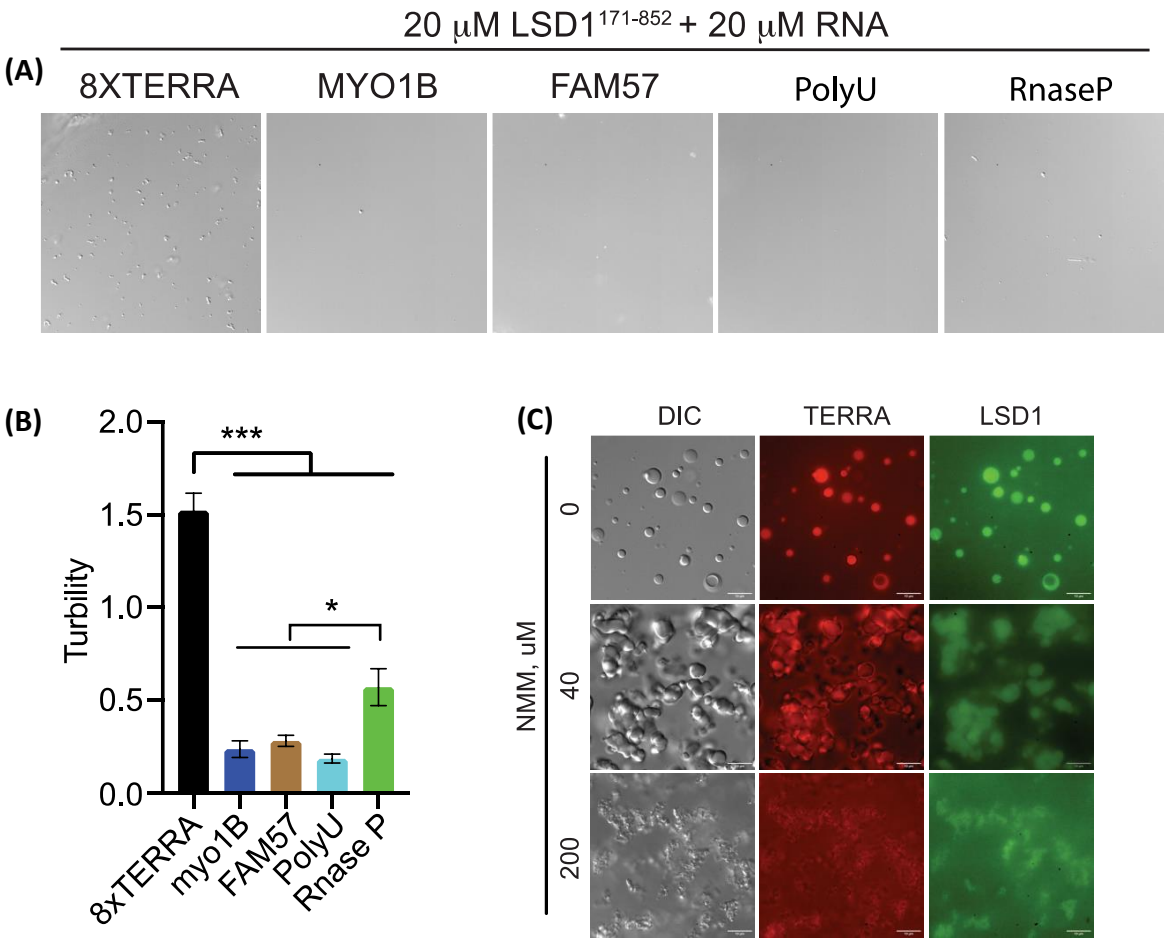


Figure 32:TERRA G-quadruplex structure is required for LSD1 phase separation and ALT functions.

(A-B) Representative DIC images **(A)** and turbidity assay **(B)** of LSD1 with the addition of different RNAs. Turbidity is shown as mean+- SEM (n=2). **(C)** Representative images of fluorescent LSD1 with TERRA pretreated with G4 inhibitor, N-methyl mesoporphyrin IX, (NMM).

3.2.2.4 LSD1-TERRA Condensates Enrich R-loop Stimulating Protein Rad51AP1

Finally, experiments were conducted to test weather LSD1 condensates promote R loop formation only by enriching TERRA on telomeres or if condensate formation also enriches the factors that may help stimulate R loop formation such as recombination proteins Rad51 and Rad51AP1 that have been shown to help TERRA R loop formation by invading dsDNA. Upon in vitro combining purified Rad51AP1 with LSD1 and TERRA separately it was discovered that Rad51AP1 does not promote LSD1 phase separation but makes irregular shaped condensates with TERRA. Interestingly, in presence of TERRA and Rad51AP1, LSD1 can partition into condensates. These condensates maintained a round shape and high in quantity than LSD1-TERRA condensates or Rad51AP1-TERRA condensates (Figure 33). Therefore, this suggests that likely by interacting with different regions of TERRA, Rad51AP1, and LSD1 synergistically phase separate with TERRA.

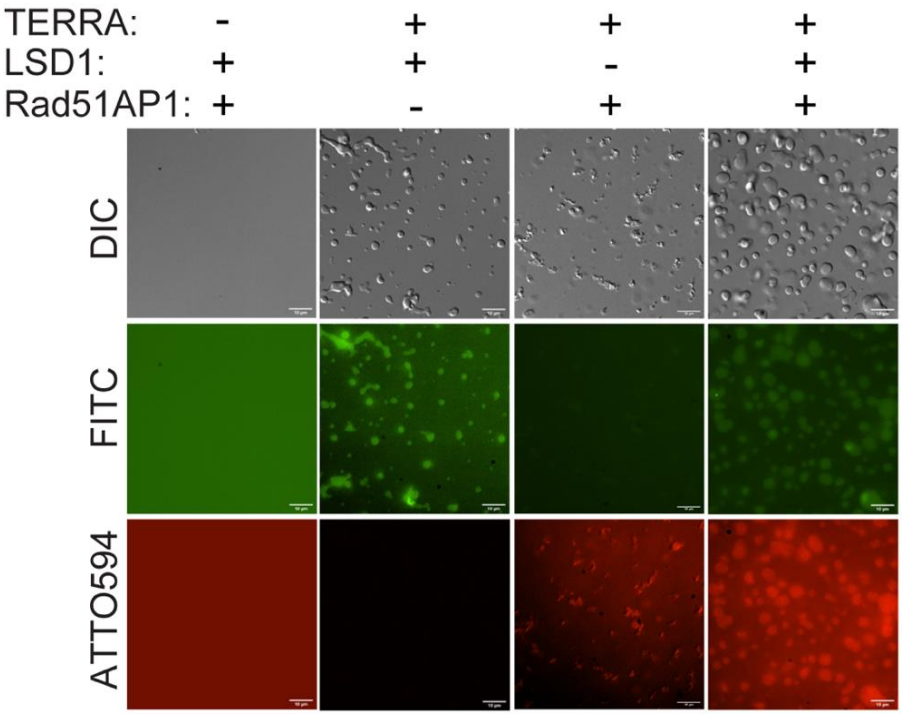


Figure 33: LSD1 promotes R-loop formation at ALT telomeres.
Fluorescent images of condensates after combination of ATTO594-labeled Rad51AP1 (30 µM), FITC-labeled LSD1 (50 µM) and 8x TERRA (50 µM).

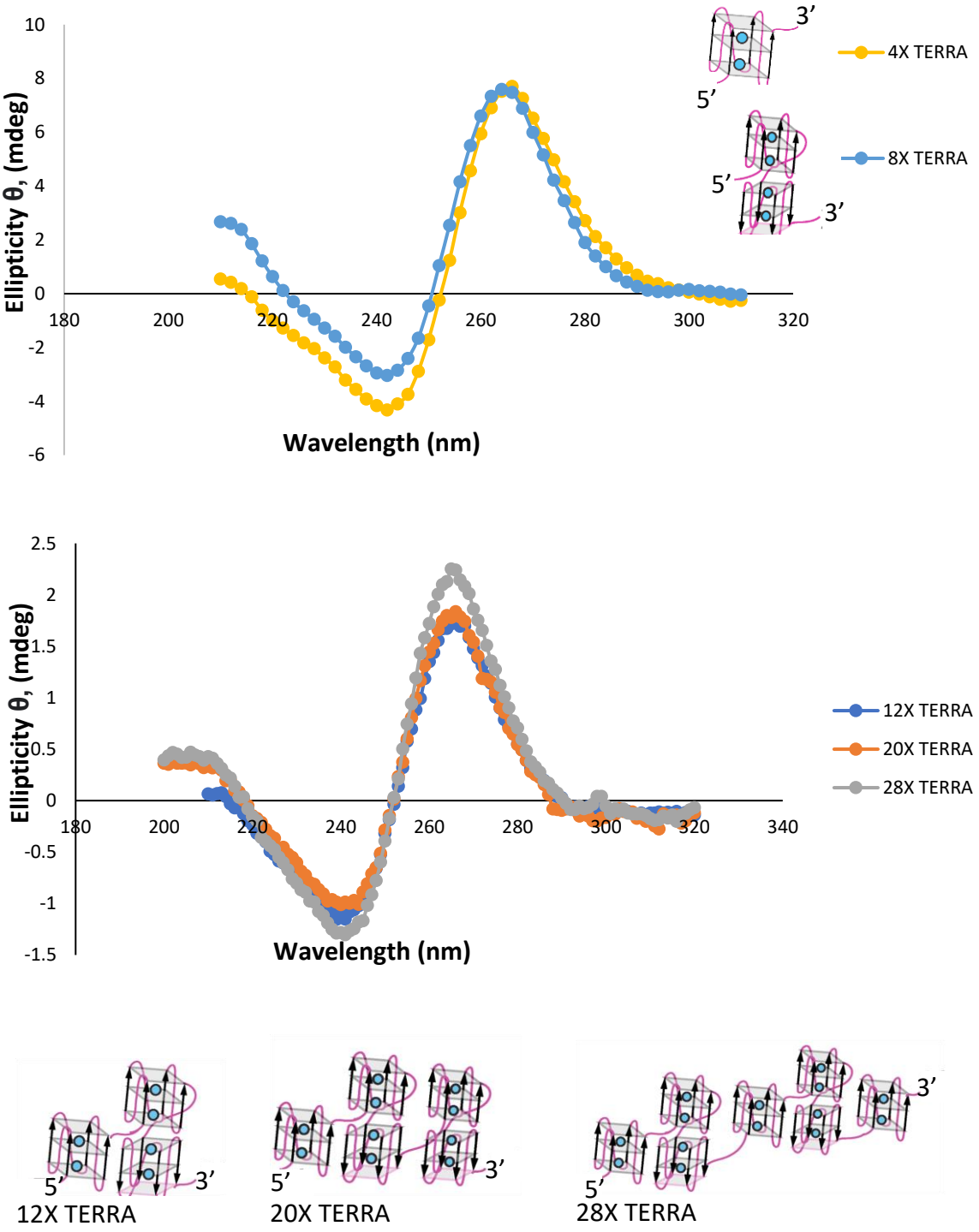


Figure 34: CD spectra of 4X, 8X, 12X, 20X, 28X TERRA RNA.
All RNA were folded in 10 mM Tris–HCl pH 7.4, 1 mM TCEP and 100 mM KCl buffer.

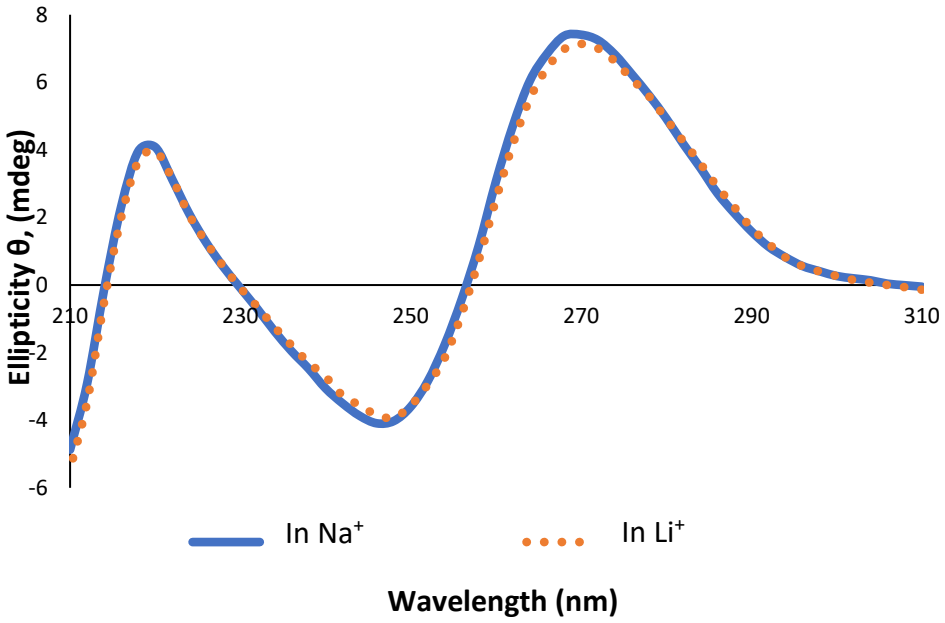


Figure 35: CD spectra of 4X TERRA DNA:RNA hybrid (duplex) annealed in K+ and Li+. Unlike GQ RNA, when annealed in Na^+ (10 mM Tris–HCl pH 7.4, 50 mM NaCl 5 mM $MgCl_2$) and Li^+ (10 mM Tris–HCl pH 7.4, 50 mM LiCl 5 mM $MgCl_2$) containing buffers, DNA:RNA hybrids contain similar characteristic CD spectrums with positive peaks at 270 nm, 220 nm and negative peak at 250 nm.

3.2.3 Discussion

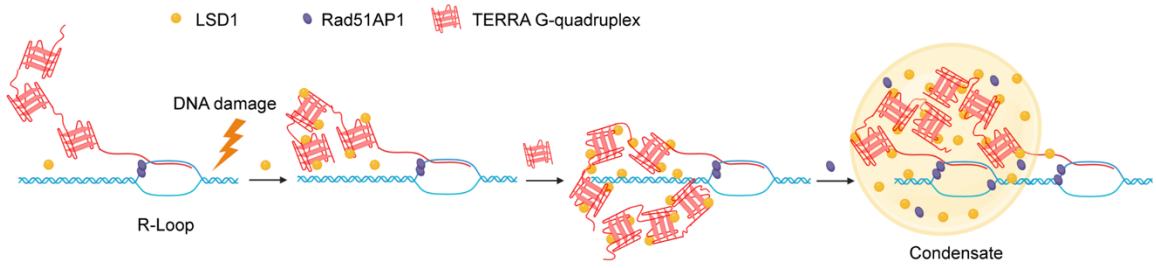


Figure 36: Working model of how TERRA drives phase separation for R-loop formation on telomere to maintain ALT.

Telomere maintenance, a hallmark of cancer cells, provides them immortality and is observed in a significant fraction of tumors. Cancer cells contain critically short

telomeres that are unstable, prone to replicative stress, and contain upregulated TERRA levels that form R loops mediated DNA damage. (114)(113) Replication machinery encountering R loops mediate replication stress and as a result, double strand breaks (DSB) are induced that trigger homology directed repair (HDR). Mediated by homology-directed repair mechanisms, alternative lengthening of telomeres (ALT) is a telomere elongation mechanism unique to cancer cells that is used to gain unlimited proliferation potential.

This study shows that TERRA contributes to ALT through interacting with LSD1. This work further suggests that TERRA R loops anchor TERRA on telomere to recruit LSD1 in response to DNA damage. This triggers TERRA mediated LSD1 local phase separation which may help enrich TERRA and other TERRA interacting proteins such as R loop stimulating protein Rad51AP1 to promote R loop formation for efficient telomere maintenance in ALT cells. (Figure 36)

Previously it has been determined that TERRA recruitment to shortened/stressed telomeres depends on R loop formation promoted by Rad51 and Brac2 recombinase proteins. (113) It is interesting to see if TERRA-LSD1 condensate formation also contributes to TERRA recruitment in stressed, shortened telomeres in non ALT telomeres. Although TERRA-LSD1 interaction is important for stimulating double strand break repair protein MRE11 nuclease activity to trim uncapped telomeres, TERRAs protein interaction network and role at shorted telomeres remains undefined. Furthermore, highlighting the diverse functional roles of chromatin modifiers, this study shows strong evidence how LSD1's enzymatic activity independent function that uses its

ability to interact with GQ forming TERRA RNA is used to enrich TERRA on telomeres and promote R loop formation in ALT cells.

IDRs and PTMs that can strengthen or impair multivalent interactions underlying phase separation can alter condensate formation through influencing RNA-protein interactions. In addition, displacement of a protein IDR enabling RNA binding and subsequent phase separation is well documented. Deleting IDR region of LSD1 that contains highly charged amino acid clusters, functionally relevant post translational modifications shows to enhance the TERRA mediated LSD1 phase separation. Therefore, it may be that IDR region regulates LSD1 phase separation with TERRA on telomeres. Although the molecular mechanisms and roles of LSD1 N terminus remain largely unknown, biochemical characterization reveals that LSD1-RNA interactions largely reduce in presence of the LSD1 IDR (see Chapter 2). In Chapter 2, I have shown that the phosphorylation sites on LSD1 N terminal IDR regulating LSD1 catalytic activity by influencing LSD1 substrate binding. In a similar manner, it is possible that electrostatic mediated altercations such as phosphorylation PTMs across the LSD1 IDR may influence protein-RNA interactions mediated phase separation at ALT telomeres. Collectively, these studies provide an example of how G quadruplex structured RNA (TERRA) phase separate with chromatin modifier (LSD1) in a GQ RNA length and structure dependent manner facilitating compartmentalization on chromatin to regulate the genome stability.

3.2.4 Materials and Methods

3.2.4.1 TERRA RNA Synthesis

Various TERRA RNA containing 28 mer (4xTERRA), 51 mer (8x TERRA), 74 mer (12x TERRA), 122 mer (20X TERRA), 170mer (28x TERRA) UUAGGG RNA repeats, as well as ssRNAs (anti-FAM57B RNA (28 mer), anti-MYO1B RNA (33 mer), Poly U RNA (32 mer), purchased as DNAs from IDT™ and transcribed as follows using the standard protocol. The DNA annealing was performed by adding 10µM template DNA, 12 µM T7 promoter, 1x transcription buffer together and annealing at 95 ˚C, 2 minutes, at 50˚C, 2 minutes and at 25 ˚C, 5 minutes respectively by using the thermocycler. RNA transcription reaction was set up by adding 5 mM DTT, 5 mM ATP/CTP/UTP, 7 mM GTP, 1x transcription buffer, annealed DNA, T7 polymerase together and incubating the reaction mixture at 37 ˚C for 4 hours. The RNA was precipitated by adding 0.1 volumes of 3 M sodium acetate pH 7.5, five volumes of ice cold 100% ethanol and keeping the reaction mixture in -20˚C overnight. The next day, RNA was spined down at 4300 r.p.m for 30 minutes. The pallet was collected and dried.

3.2.4.2 Cy3-UTP Labeled TERRA RNA Transcription

The Cy3-labeled 12X TERRA RNA was purified and prepared in-house as described above, with an altered *in vitro* transcription reaction that contained UTP (3 mM) and the Cyanine-3-labeled UTP (0.8 mM) (Enzo life sciences).

3.2.4.3 TERRA RNA Purification

Dried RNA pallet was loaded to 8M urea, 1x TBE, 8% - 20% acrylamide (depending on the length of RNA) gel. The bands corresponding to desired RNA was cut out and RNA was extracted using 0.3M sodium acetate pH 7.5, and 100% ethanol before desalting using Bio Rad G50 desalting spin columns.

3.2.4.4 G-Quadruplex RNA Folding, Duplex Formation

GQ RNAs were diluted in 10 mM Tris–HCl pH 7.4, 1 mM TCEP and 100 mM KCl and was folded using a standard protocol (2 min at 95 ºC, 5 min at 85 ºC, 5 min at 75 ºC, 5 min at 55 ºC, 15 min at 37 ºC, and then placed on ice).

For RNA duplex formation, sense and antisense RNAs were annealed by mixing equimolar amounts of RNA oligonucleotide solutions in annealing buffer (10 mM Tris–HCl pH 7.4, 50 mM NaCl 5 mM $MgCl_2$). The solution was heated at 94 °C, for 4 minutes and gradually cooled down to room temperature for duplex formation.

Previous folding protocols were used to anneal the bacterial RNase P RNA from *T. maritima* which serves as a control RNA with a high degree of tertiary structure. (117)

3.2.4.5 CD Spectroscopy for RNA Global Confirmation Validation

Global conformations of each RNA were evaluated using circular dichroism (CD) spectroscopy, where CD spectrums were recorded at room temperature on a Chirascan^{TM} V100 CD spectrophotometer with a 1 mm cell, 1 nm band width, 4 s per point, and 0.1ms timed intervals.

3.2.4.6 Protein Purification

Human full-length (FL) LSD1 (1-852) and Δ N LSD1 (171–852) with CoREST (286–482) proteins were co-expressed in Rosetta (DE3) pLysS competent cells using either pET-15b (FL-LSD1) or pGEX-6P-1- Δ N LSD1 plasmids with pET28-CoREST plasmid, as previously described (60) with the following modifications. Rosetta (DE3) pLysS competent cells were grown at 30 ºC in auto induction media containing 200 µg/mL ampicillin, kanamycin, and 40 µg/mL chloramphenicol antibiotic concentrations. After sonication, cell lysate was purified using Ni-affinity chromatography followed by GST affinity chromatography. Restriction-grade thrombin and 3C precision protease digestion were used to remove histidine and GST tags followed by an additional GST affinity chromatography column and subsequent Superdex 200 size exclusion chromatography was used to further purify the LSD1-CoREST complex.

Two different point mutant constructs of LSD1 containing K661A or K355E/K357E/K359E with CoREST (aa 286-482) and pST44STRaHISNhLSDΔ1x32-hCORESTΔ2 hLSD1Δ1(K661A) and pST44STRaHISNhLSD Δ 1x17-hCORESTΔ2 hLSD1Δ1 (K355E, K357E, K359E) were gifts from Dr. Song Tan and were purified as described. (59) In all cases, protein was stored in 25 mM HEPES Na pH 7.5, 100 mM NaCl, and 5 mM TCEP buffer. The concentration of protein samples was determined by the BCA method (BCA Protein Assay kit, Pierce) with bovine serum albumin as a standard and validated with UV-VIS-spectroscopy, using the extinction coefficient for the FAD cofactor of LSD1 at 450 nm (ε = 11300 $M^{-1}cm^{-1}$). Protein spectra were recorded on Shimadzu UV-2600

Spectrophotometer and protein purity was assessed by SDS electrophoresis with Coomassie staining.

3.2.4.7 In-vitro Phase Separation Assay

Recombinant protein LSD1 in buffer 25 mM HEPES pH 7.5, 100 mM NaCl, 2 mM TCEP was added to solutions at varying concentrations. The TERRA RNA was added into the protein solution at indicated concentrations and mixed well. Then the protein-RNA solution was immediately loaded onto a homemade chamber comprising silicone isolator (VWR, 100490-928) and a glass slide (Fisher Scientific 22-266-858p) with a coverslip. The Slides were imaged with a Nikon confocal microscope with a 100x objective. Unless indicated, images presented are of condensates settled on the glass coverslip.

Project Contributions:

All LSD1-CoREST constructs purification, all RNAs synthesis (fluorescently labeled and unlabeled), RNA purification/folding/annealing and their CD spectroscopy, Native PAGE gel analysis of RNAs, and relevant data analysis were performed by Dulmi Senanayaka. All other cell localization, cell-based and in vitro based fluorescent-labeled studies and relevant data analysis were performed by Dr. Meng Xu (Carnegie Mellon University, in the laboratory of Dr. Huaiying Zhang).

CHAPTER 4 ANTICANCER DRUGS OF LSD1 DISPLAY VARIABLE INHIBITION ON H3K4^{me2} NUCLEOSOME SUBSTRATES

4.1 Introduction

Abnormal epigenetic regulation by enzymes can lead to altered gene expression and malignant transformation. Therefore, enzymes like LSD1, which mediate epigenetic modifications are therapeutic targets for cancer treatment. LSD1 has clear roles in cell differentiation in normal and disease pathways, is subjected to allosteric regulation, and serves as a predictor of a variety of cancers including acute myeloid leukemia, neuroblastoma, as well as prostate, bladder, breast, liver, and colorectal tumors. LSD1's elevated expression is reported in mesenchymal tumors including the rare and aggressive bone associated malignancy called Ewing Sarcoma. Here, high expression of LSD1 is correlated to the aggressiveness of the tumor and poor prognosis of the disease. (119) (120) Thus, LSD1 may be thought of as an oncogene in Ewing Sarcoma because its overexpression induces the proliferation of non-transformed human cells such as mesenchymal cells in the bone marrow. (121)

Ewing sarcoma (EWS) is a pediatric bone tumor that relies on the transcriptional activity of EWS/ETS (Erythroblast Transformation Specific) family of translocation derived fusion oncoproteins including the most common fusion EWS/FLI1 (Friend leukemia integration 1), which is an oncogenic fusion and a transcription factor that functions as a transcriptional activator and repressor to mediate oncogenic transformation. (122) In the EWS/FLI1 fusion protein structure, the N terminal low

complexity intrinsically disordered portion (EWS portion) that recruits transcriptional co-regulators is derived from the EWSR1 (Ewing Sarcoma breakpoint region 1/EWS RNA binding protein 1) gene. The FLI portion of the C terminus of the fusion protein is derived from ETS domain family gene member FLI1 gene and it contains a conserved DNA binding domain that binds in high affinity to canonical ETS binding motif (5'-ACCGGAAGTG-3'). (123) The oncofusion EWS/FLI1 protein is required to maintain the growth of Ewing sarcoma cell line. In Ewing sarcoma, both EWS/FLI mediated repression of target genes as well as upregulation of EWS/FLI1 target genes is required for its oncogenic activity. (124) (125) Determined by co-immunoprecipitation experiments, full-length EWS/FLI bind two components (MTA2 and CHD4) of the Nucleosome Remodeling Deacetylase (NuRD/CoREST) complex, suggesting that EWS/FLI1 directly interacts with the NuRD complex to mediate transcriptional repression. (124) EWS/FLI1 may recruit LSD1 as a member of NuRD complex which leads LSD1 to repress tumor suppressors and enact transcriptional activation through an unknown molecular mechanism. (123) (124) (126)(122)

Since it is not yet clinically feasible to disrupt transcription factor EWS/FLI1 mediated gene regulation through targeting the fusion protein, groups have targeted a critical transcriptional function of EWS/FLI1 via blockage of LSD1. LSD1 is a target for EWS because it is a known, validated drug target and represents a catalytic subunit of the NuRD repressor complex. Thus, an alternative therapeutic strategy for EWS treatment to reverse the aberrant roles of the EWS/FLi1 oncoprotein is to develop targeted LSD1-based inhibitors.(126)

LSD1 inhibition can lead to increased levels of H3K4 methylation and has been shown to reactivate tumor suppressor genes in several cancer types. As such, selective modulation of LSD1 has high therapeutic value due to diverse interaction pathways with various biomolecules including transcription factors, nucleosomes, oncoproteins, DNA, non-coding RNAs, splicing factors, and tumor suppressors. However, the number of diverse interaction pathways and LSD1's molecular complexity makes LSD1 based therapeutic approaches difficult. Also, LSD1 molecular inhibitors can exhibit poor selectivity, potency, and efficacy as well as poor off target toxicity that may limit the interrogation of LSD1 as a viable therapeutic strategy.

Nonetheless, to date, multiple small molecule inhibitors have been tested as clinical candidates for LSD1 targeted therapy for the treatment of Acute Myeloid Leukemia (AML), Small Cell Lung Cancer (SCLC), prostate cancer and myelodysplastic syndrome. (127) Among the many mono amine oxidase (MAO A and B) inhibitors, clinically validated antidepressants and non-selective inhibitors tranylcypromine (TCP/ trans-2-phenylcyclopropyl- amine) and pargyline as well as their derivatives are known irreversible LSD1 inhibitors. (128) TCP binds FAD and irreversibly weakly inhibits LSD1 activity (K_i of 243μM) by forming a tetracyclic adduct in the AOD catalytic cavity and acting as a mechanism-based suicide inhibitor. (129)(130) (131) (132) Six TCP based LSD1 inhibitors including TCP, GSK-2879552, ORY-1001 (iadademstat), ORY-2001, INCB059872, IMG-7289 (bomedemstat) have currently entered clinical trials. (130) Even though TCP has low efficacy, it is found that LSD1 inhibition by TCP impairs the growth of Ewing sarcoma cell lines in vitro. (133)(122) On the contrary Iadademstat, a

tranylcypromine derivative, acts as a covalent highly potent and selective inhibitor of LSD1 and is currently in clinical trials for the treatment of AML. (134) (135)

Since LSD1 plays critical roles in maintaining cellular function, homeostasis, and LSD1 ablation causes embryonic lethality, irreversibly inhibiting it may lead to increased on target toxicity. (136) Therefore, LSD1 reversible inhibitors can provide therapeutic outcomes while limiting on target toxicity to normal cells. (137) Until now reversible inhibitors seclidemstat (SP-2577), and pulrodemstat (CC-90011) have entered clinical evaluation for the mono or combined therapy of several diseases including Ewing sarcoma, acute myeloid leukemia (AML), and small cell lung cancer (SCLC). (63) (138)

The discovery of Pulrodemstat, a reversible inhibitor for LSD1, which showed highly potent inhibition of LSD1 provides a novel pathway for the treatment of LSD1 associated cancers and tumors such as AML, neuroendocrine tumors and potentially Ewing sarcoma. It is currently in phase 1 clinical trials for treating advanced solid tumors and phase 2 clinical trials for SCLC and has shown safety and tolerability in treatments. (137) Moreover, through structure-based high throughput virtual screening followed by biochemical screening Sorna *et al.* discovered novel, reversible N'-(1-phenylethylidene)-benzohydrazide based noncompetitive inhibitor HCL-2509 for the inhibition of LSD1 (LSD1 IC_{50} 13 nM, K_i 31 nM). (139)(126) HCL-2509 shows disrupted global oncogenic activity when used for the treatment of Ewing sarcoma cell line as well as suppressed growth of Ewing sarcoma xenografts in mice, suggesting that targeting LSD1 inhibition by HCL-2509 may be a therapeutic target in treating Ewing sarcoma. (122) The HCI-2509 analogue, SP-2577 (Seclidemstat), is also thought to be a potent noncompetitive,

selective, and reversible inhibitor of LSD1 (IC_{50} ~25-50 nM) (139)(140) Previously, mouse models revealed that SP-2577 treatment led to 80% reduction in tumor morphology suggesting a highly potent drug for inhibiting LSD1 demethylation on nucleosomes. (141)(139) In 2020, SP-2577 was FDA approved for the treatment of Ewing Sarcoma as well as it is in clinical trials for the treatment of advanced solid tumors. (141)

However, later studies through tumor regression analysis and H3K4 dimethylation analysis show that SP-2577 has little pre-clinical activity and inconsistent pharmacodynamic effects against Ewing sarcoma, questioning the efficacy of the drug. (121) Compared to other reversible LSD1 inhibitors in clinical trials, SP-2577 showed low stability at 37˚C and showed solubility issues under the conditions tested. (138) In addition, the exact molecular mechanism of the seclidemstat binding and mode of inhibition of H3K4me^2 nucleosome demethylation is unknown. A crystal structure of the drug bound to LSD1 has not been determined and it is unclear how SP-2577 impacts LSD1 function.

In this chapter I examine the LSD1's druggable space, demonstrating that some anticancer drugs in clinical trials display variable LSD1 inhibition in Ewing sarcoma treatment. I utilize a small molecule comparative kinetics approach to evaluate the potency of irreversible inhibitors tranylcypromine/iadademstat, reversible inhibitors Seclidemstat (SP-2577) and Pulrodemstat (CC-90011) in inhibiting LSD1 catalytic activity in the presence of H3K4me^2 nucleosomal substrate *in vitro* and in cancer cells. Furthermore, this study compares the drug's efficacy in inhibiting LSD1 activity on small

peptide and nucleosomal substrates in vitro to understand the “druggability” of LSD1 for the development of an effective cancer treatment for Ewing sarcoma.

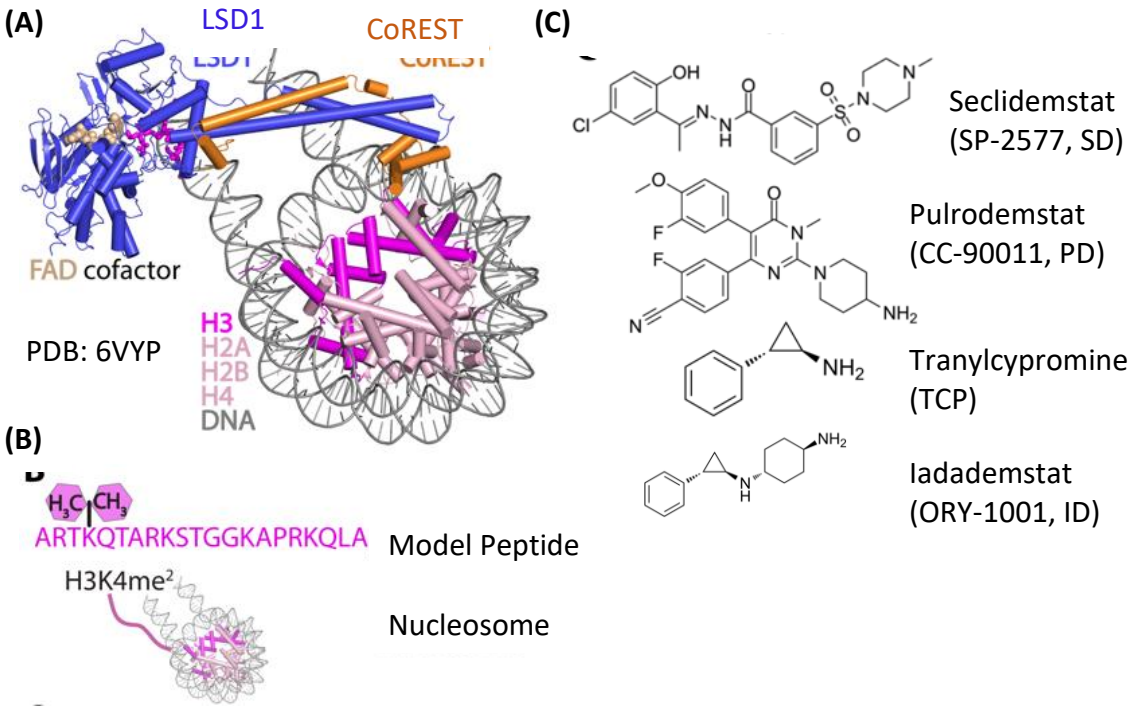


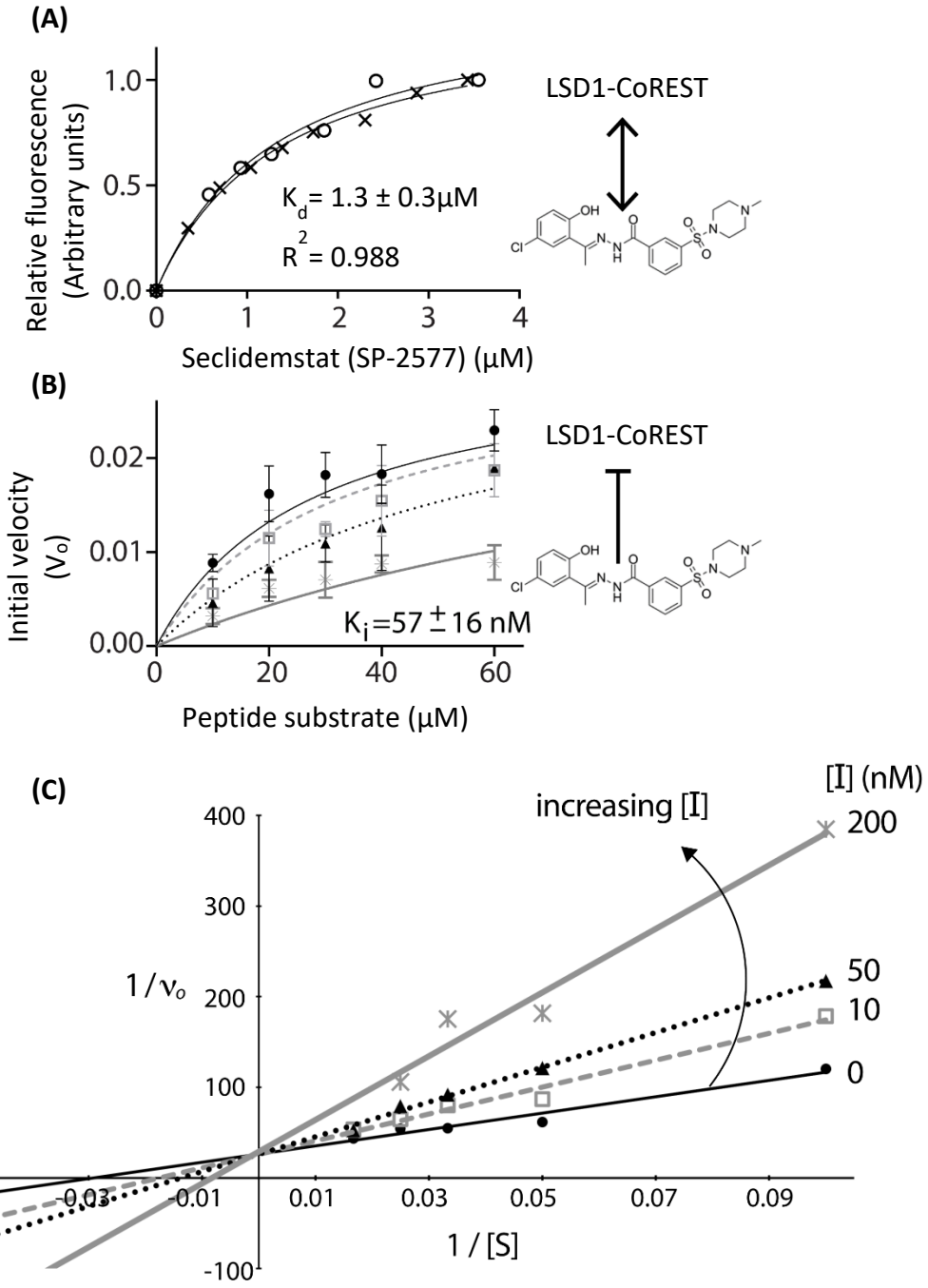
Figure 37: Structure of the LSD1-CoREST complex with mononucleosomes, and the LSD1 substrates and anticancer drugs used in this study.

(A) Structure of the complex with nucleosomes (PDB 6VYP) reveals LSD1’s mode of engagement. LSD1 (blue), CoREST (orange), and cofactor FAD (beige) interact and demethylate H3K4me$^{1/2}$ nucleosomes, comprised of DNA (grey) and the histone octamer (H3 (magenta), H2A, H2B, and H4 (light pink). **(B)** Schematic of substrates used in this study. Residues of the H3K4me^2 model peptide substrate (top) and H3K4me^2 containing nucleosome (bottom) are shown. **(C)** Structure of anticancer drugs that interact with LSD1. Seclidemstat (SP-2577) and pulrodemstat (CC-90011) drugs represent reversible ligands that interact with LSD1, whereas tranylcypromine (TCP) and iadademstat represent irreversible inhibitors that reacts with FAD at the active site.

4.2 Results

4.2.1 Seclidemstat Binds ΔN LSD1-CoREST with Low Micromolar Affinity

Using a label free approach, ligand Seclidemstat binding to ΔN LSD1-CoREST complex was measured by the intrinsic fluorescence of FAD cofactor deeply embedded non-covalently within the amine oxidase domain of LSD1 (Figure 37-A). Purification and fluorescence of the enzyme reveals a 1:1 stoichiometric ratio between LSD1 and FAD. Using fluorescence spectroscopy, the optimal FAD emission was monitored at 520nm, and Seclidemstat solution was titrated into an enzymatically active FAD -ΔN LSD1-CoREST solution. As previously established and based on control experiments, the increased fluorescence intensity upon titration was attributed to Seclidemstat binding near the FAD moiety. (142) The actual concentrations for ΔN LSD1-CoREST and Seclidemstat were corrected for the dilution factor at each titration measurement and all the titration measurements were performed in duplicate. From this approach, I was able to determine the ΔN LSD1-CoREST - Seclidemstat dissociation constant K_d = 1.3 ± 0.3 µM, suggesting that seclidemstat binds near the active site cleft of LSD1 (Figure 38-A).


(A)
Relative fluorescence (Arbitrary units)
1.0
0.5
0.0
0
1
2
3
4
Seclidemstat (SP-2577) (μM)
K_d= 1.3 ± 0.3μM
R^2 = 0.988
LSD1-CoREST
(B)
Initial velocity (V_o)
0.02
0.01
0.00
0
20
40
60
Peptide substrate (μM)
K_i=57 ± 16 nM
LSD1-CoREST
(C)
[I] (nM)
increasing [I]
200
50
10
0
1 / v_o
400
300
200
100
-100
-0.05
-0.03
-0.01
0.01
0.03
0.05
0.07
0.09
1 / [S]

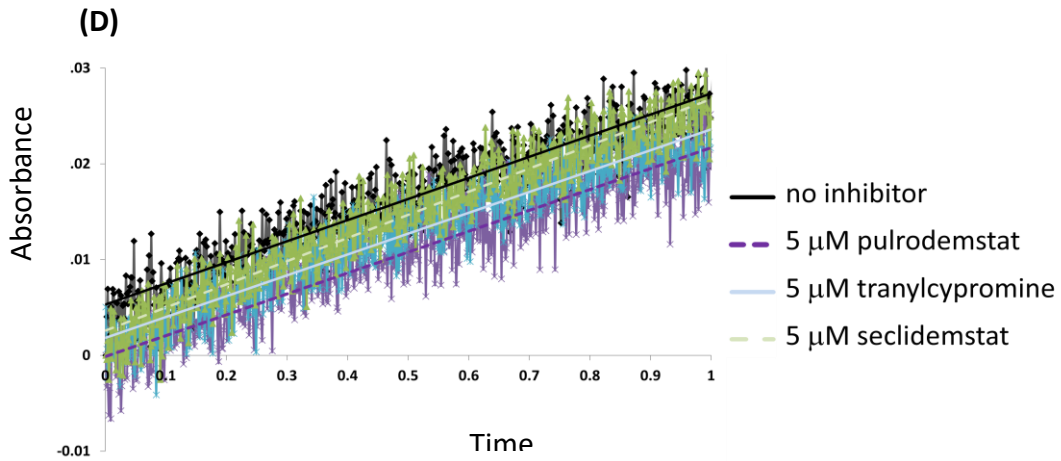


Figure 38: Seclidemstat binds LSD1 and inhibits demethylation using a H3K4^{me2} peptide substrate.
(A) Relative fluorescence changes with increasing seclidemstat concentrations enables determination of the apparent dissociation constant (K_d ~1.3 ± 0.3 μM) for the LSD1-seclidemstat binding interaction. Two independent experiments (noted as X, O) were individually fit to a 1:1 binding isotherm. **(B)** Data analysis of LSD1-catalyzed demethylation assays in presence of an H3K4me^2 peptide substrate and increasing concentration of seclidemstat. Reactions with 0.3 μM LSD1-CoREST were incubated with increasing concentrations of seclidemstat (0 (black circle), 10 (open square), 50 (triangle), 200 (asterisk) μM) and initiated with peptide substrate (10, 20, 30, 40, and 60 μM). Data of the Initial velocity vs. substrate concentration allowed for the determination of an apparent inhibition constant (K_i). **(C)** Lineweaver Burk plot showing inhibition of H3K4me^2 model peptide demethylase activity with increasing concentrations of the anticancer drug SP-2577. The inhibition data (K_i ~ 57 nM) correlate with past studies reporting on seclidemstat's ability to inhibit the ΔN LSD1-CoREST complex (K_i ~ 31 nM, (139)). **(D)** Seclidemstat and other anticancer drugs are not false positives in an H_2O_2 induced HRP coupled assay. A time course reaction of the HRP assays was monitored in the presence of 15 μM H_2O_2 and in the absence of LSD1 (black). Incubation of 5 μM of each anticancer drug, seclidemstat (light blue), pulrodemstat (purple), tranylcypromine (light dashed blue) with the HRP reagents occurred prior to the addition of H_2O_2 and an increase in absorbance was measured. All three anticancer drugs do not appreciably affect the rate of absorbance. This control indicates each compound does not impact the colorimetric reagents and do not react with H_2O_2. Rather, the drugs primarily impact LSD1 catalyzed demethylation on H3K4me^2 model peptide substrates, as previously observed.

4.2.2 Seclidemstat Inhibits ΔN LSD1-CoREST in Presence of H3K4me^2 Model Substrate

HRP coupled assay was used to measure the inhibition kinetics of the demethylation reaction between ΔN LSD1-CoREST and a 21 amino acid H3K4me^2 model substrate in the presence of increasing amounts of seclidemstat. (54) An apparent inhibition constant (K_i) of 57± 16 nM was determined using triplicated data of the initial velocity and substrate concentration (Figure 38-B). This study further validates the previous biochemical studies of LSD1-seclidemstat interaction.(139) When the data were analyzed using Lineweaver Burk plot (Supplemental Figure 38 D), it fitted consistent with fluorescent data revealing that seclidemstat binds LSD1. As a control experiment to eliminate the false positives in the HRP coupled assay I incubated seclidemstat with all the components of the HRP assay except ΔN LSD1-CoREST complex and collected the data after initiating the reaction by adding H_2O_2 to the reaction mixture (Figure 38 C). The data showed negligible changes in the absorbance, indicating that seclidemstat primarily impacts LSD1 catalyzed demethylation of model peptide substrate as previously observed but does not impact the H_2O_2 levels in the assay mixture.

4.2.3 Putative Anticancer Drugs of LSD1 Display Variable Inhibition on H3K4me^2 Nucleosome Substrates

Quantitative western blot assays under single turnover conditions were developed to test the ability of seclidemstat to inhibit the H3K4me^2 demethylation in the presence of nucleosomal substrates.(59)(71) For this assay, in addition to

seclidemstat (SP-2577) another reversible inhibitor pulrodemstat (CC-90011) and an irreversible inhibitor tranylcypromine were used as controls. Relative kinetics ($k_{observed}$) of the LSD1-catalyzed demethylation reaction was obtained from a series of time course assays measuring the fraction of dimethylated nucleosomes quantitated using H3K4me^2 specific antibody relative to the amount of H3 in each lane quantitated using H3 specific antibody (Figure 39-A). A smaller fraction of dimethylation corresponds to a stronger demethylase activity of LSD1 (Figure 39-B). Surprisingly, dimethylation profiles at increasing SP-2577 concentrations (0, 0.01, 0.02, 0.05, 0.1, 100, and 200 μM) reveal that it does not impact LSD1 demethylation of nucleosomal substrate even at very high concentrations of SP-2577. In contrast, addition of reversible inhibitor pulrodemstat (0, 2, 4, 8, 10 nM) impacts LSD1 demethylation of nucleosomes with a potent half maximal inhibitory concentration (IC_{50}) of 3.9 ± 0.3 nM. Further, addition of the known lead scaffold compound and weak inhibitor tranylcypromine (TCP), displayed appreciable levels of inhibition on nucleosomes (IC_{50} = 14.1 ± 1.3 μM). (Figure 39-C)

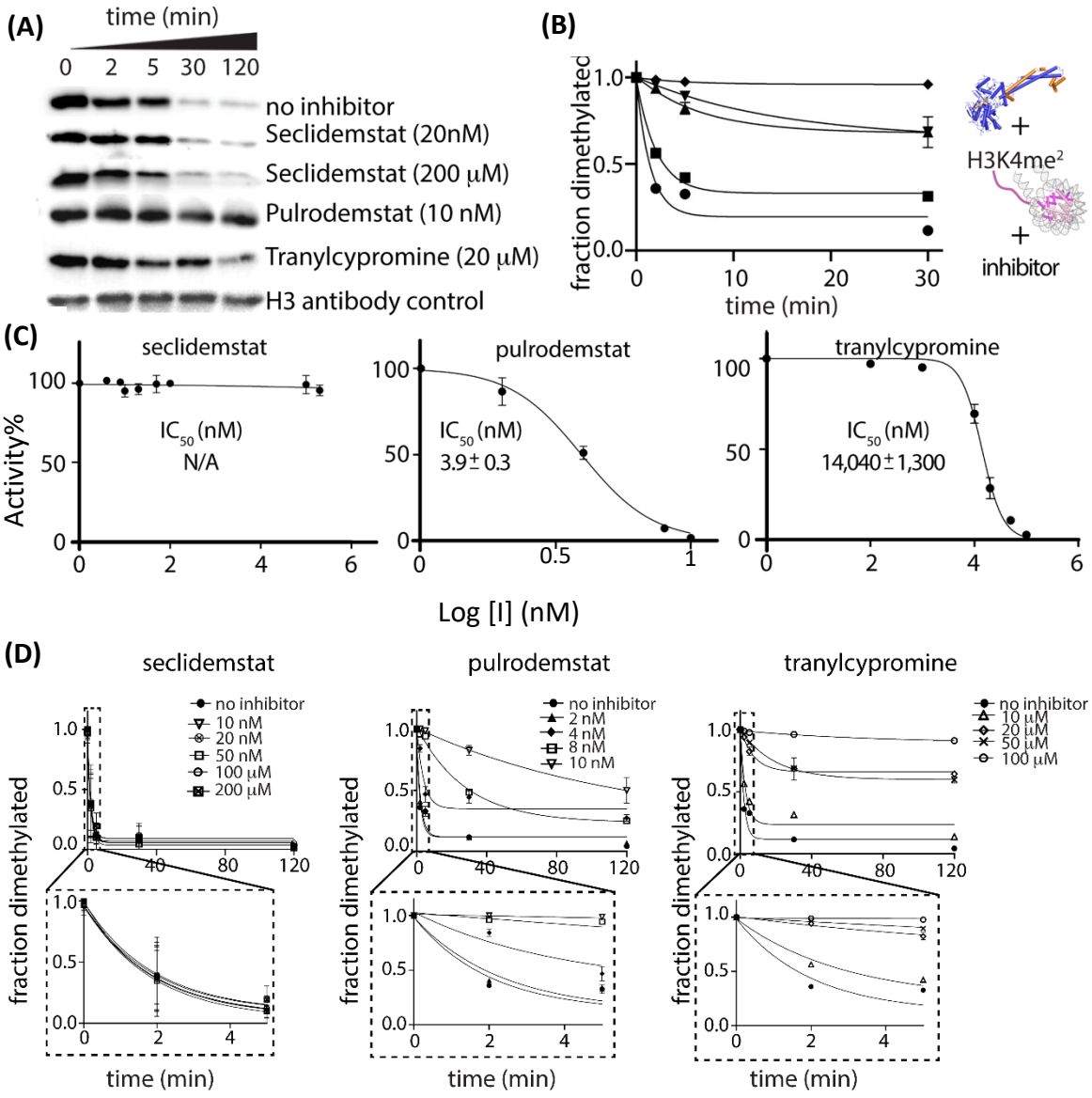
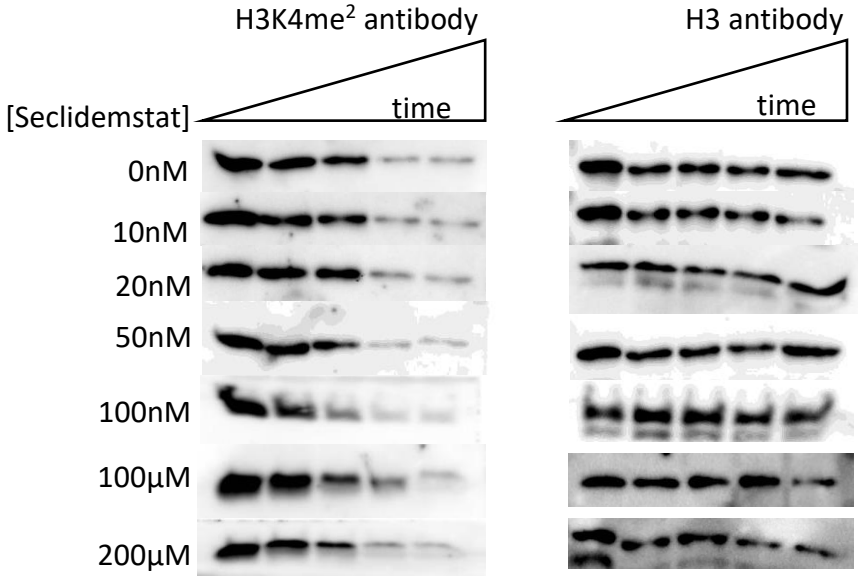


Figure 39: Three putative anticancer drugs of LSD1 display variable inhibition on H3K4me^2 nucleosome substrates.
(A) Histone demethylase activity on nucleosomes analyzed using a time-course western blot assay. Over 2 hours, the fraction of dimethylated nucleosomes with and without different anticancer drug concentrations was monitored using H3K4me^2 antibodies. Weaker signals correspond to strong demethylase activity. A time course experiment with identical reagents was performed and evaluated using anti-H3 antibodies as a control. **(B)** Representative plot of the fraction dimethylated over time. Dimethylation profiles of 'no inhibitor' (black circles) and increasing inhibitor concentrations (black squares to open circles) reveals an apparent $k_{observed}$ value, providing a measure for the potency of a molecule to inhibit LSD1's activity on H3K4me^2 nucleosomes. **(C)** Half

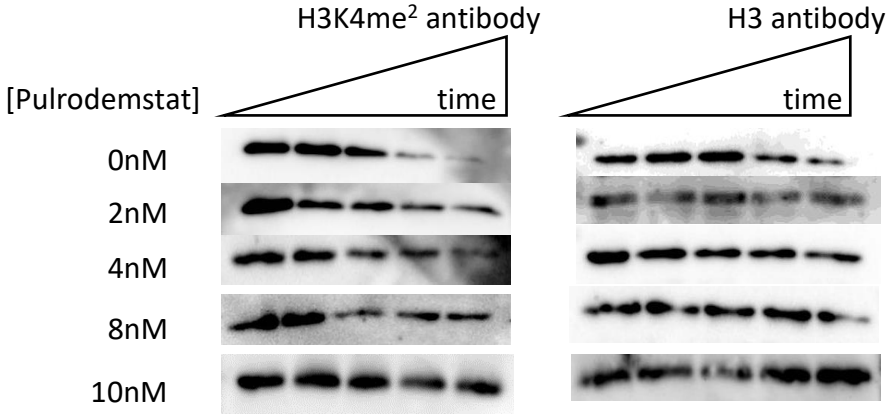
maximal inhibitory concentration (IC_{50}) values were determined based upon the analysis of the percent activity (y-axis) and drug concentration (log_{10}, x-axis). A wide range of seclidemstat, pulrodemstat, and tranylcypromine drug concentrations (shown in nM) were tested, revealing a wide range in potency. All reactions contain 2 µM ΔN LSD1-CoREST and 100 nM H3K4me^2 nucleosomes. (D) Quantitation of dimethylated H3K4me2 nucleosomes from western blot analysis. Representative data showing the fraction of dimethylated H3K4me2 nucleosomes as monitored with increasing concentrations of seclidemstat (0 – 200 µM), pulrodemstat (0 – 10 nM), or tranylcypromine (0 – 100 µM). Analyzed data from Figure 39 (A-C). A zoom of the fitted data for the first five minutes of the reaction are highlighted (dashed boxes), emphasizing the observed variability in H3K4me^2 nucleosome inhibition. Error bars (N=2-3) show the standard deviation.

Figure 40: Histone demethylase activity on nucleosomes analyzed by western blotting using H3K4me^2 and anti-H3 antibodies with increasing concentrations of an anticancer drug. Weaker signal corresponds to strong demethylase activity. All reactions contain 2 µM ΔN LSD1-CoREST and 100 nM H3K4me^2 nucleosomes.

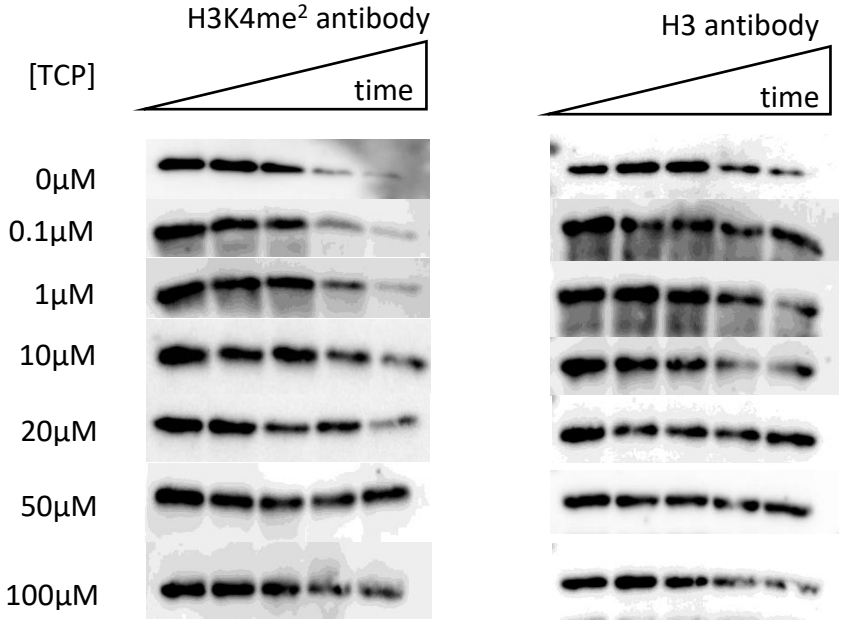
(A) Seclidemstat inhibition of ΔN LSD1-CoREST catalyzed demethylation



(B) Pulrodemstat inhibition of ΔN LSD1-CoREST catalyzed demethylation



(C) Tranylcypromine inhibition of ΔN LSD1-CoREST catalyzed demethylation



4.2.4 Seclidemstat is Not Potent at Demethylating H3K4me^2 Nucleosomes in Cancer Cells

Next, each of the LSD1 inhibitors' potency and selectivity were tested in different cell lines. For this assay, Ewing sarcoma (TC32) and osteosarcoma (U2OS) cancer cell lines and HEK293 (Human Embryonic Kidney) cell line were utilized as models. For testing the cytotoxic effects of LSD1 inhibitors, seclidemstat, pulrodemstat, TCP, and iadademstat IC_{50} values were calculated using cell titer blue assay 48 hours post-inhibitor treatment of all cell lines. Upon treatment, TCP did not induce cell death of the three cell lines even at the maximum concentration used (500μM) (Figure 41-B). Seclidemstat was selectively cytotoxic only for TC32 cells (IC_{50}= 4 μM) but not U2OS and HEK293 cells (Figure 41-A). Pulrodemstat (TC32 IC_{50}= 52 μM, U2OS IC_{50}= 57 μM, HEK293 IC_{50}= 12 μM) and iadademstat (TC32 IC_{50}= 216 μM, U2OS IC_{50}= 349 μM, HEK293 IC_{50}= 120 μM) showed cytotoxicity on cell lines at various concentrations (Figure 42-A, Supplementary Figure 41-A).

In order to determine the effects of drug induced LSD1 inhibition in Ewing sarcoma cancer cells, the remaining H3K4me^2 levels on nucleosomes were monitored. For this, Ewing sarcoma cancer cell (TC32) lysates were treated with seclidemstat, pulrodemstat, or iadademstat at two different concentrations (0.75 μM, 1.5 μM), or left untreated (control), for 48 hours (Figure 42-B). The abundance of H3K4me^2 in the cell lysates was assessed by western blot analysis and the residual H3K4me^2 in each lane was quantitated relative to both H3 and actin loading controls in each corresponding lane. Inhibition of LSD1 is represented by the increase of H3K4me^2 levels in cell lysates. Interestingly, consistent with my in vitro inhibition kinetics data where seclidemstat

shows no effect in inhibiting LSD1 activity on nucleosomes, in Ewing sarcoma cancer cells seclidemstat shows no effect in inhibiting LSD1 activity in cancer cells. But pulrodemstat and TCP derivative iadademstat (ID) show comparatively similar LSD1 inhibition potency (Figure 42-B). In addition to this, same assay was performed in presence of 3 different concentrations (Seclidemstat- 0.75, 1.5, 3 μM and Pulrodemstat- 10, 20, 40 μM) and analyzed the abundance of $H3K4me^2$ in the cell lysates (Figure 42-C). Both these data further suggest that Seclidemstat is not an effective drug in inhibiting LSD1 activity on nucleosomal substrates while Pulrodemstat shows appreciable levels of LSD1 inhibition in Ewing sarcoma cancer cells.

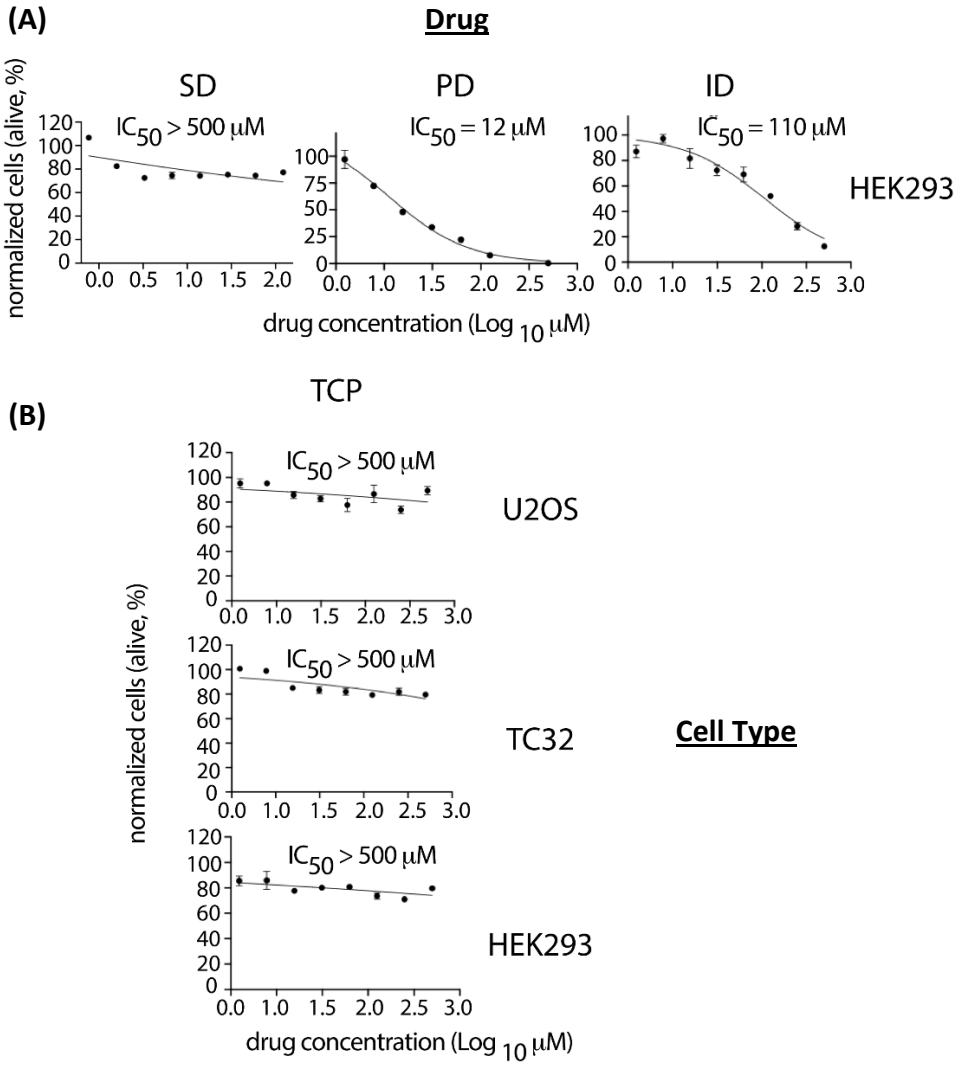


Figure 41: Cell viability of anticancer drugs.
(A) Representative cell viability IC_{50} determination of the anticancer drugs Seclidemstat (SD), pulrodemstat (PD and iadademstat (ID) in HEK293 cells performed in duplicate (N=2). SD does not show specificity to HEK293 but TC32. **(B)** TCP cell viability IC50 determination in U2OS, TC32 and HEK293 cells. TCP shows no appreciable effect on viability in all cell lines up to 500 μM concentrations.

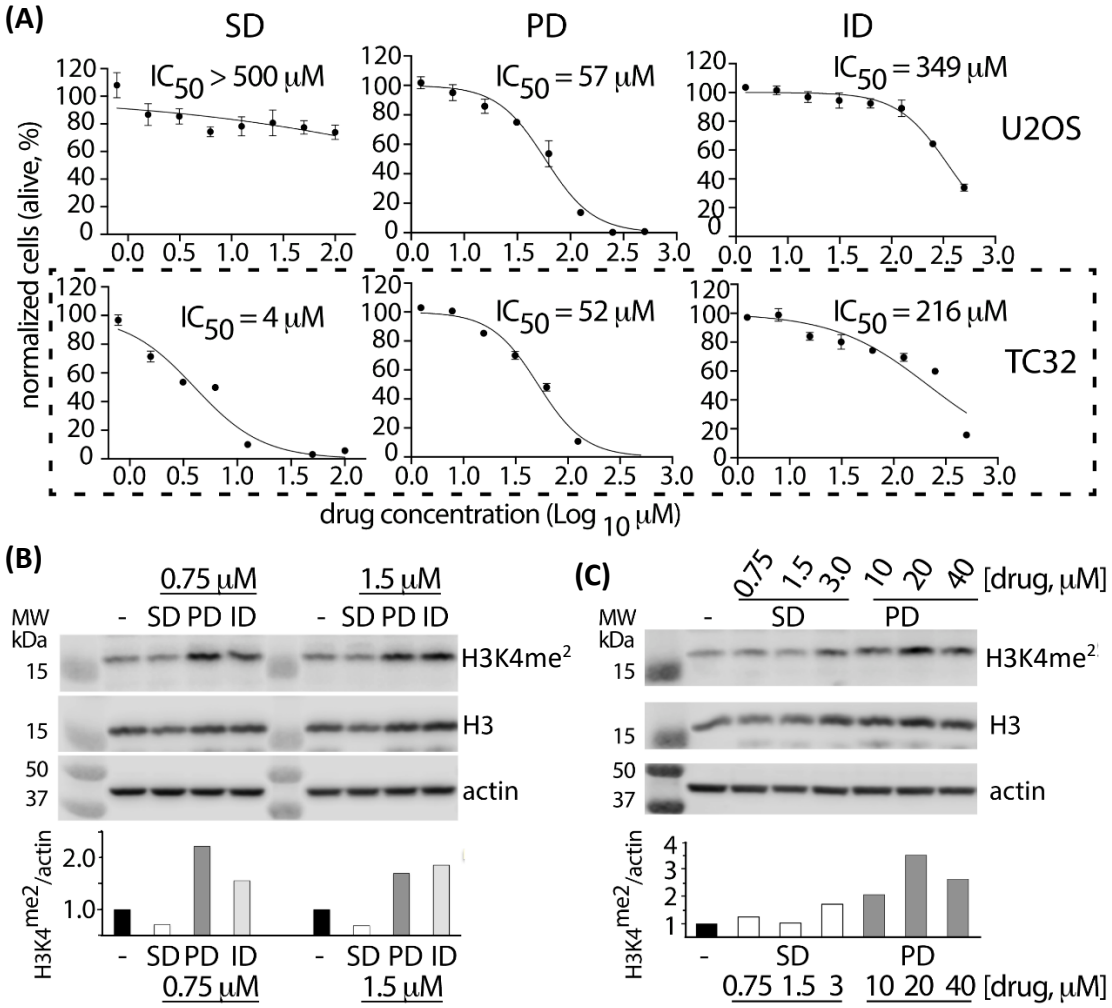


Figure 42: The cancer cell lines treated with different doses of LSD1 inhibitors Iadademstat (ID), Pulrodemstat (PD), and Seclidemstat (SD).
(A) The cytotoxicity of ID, PD, and SD in Osteosarcoma (U2OS) and Ewing sarcoma cancer cells (TC32) assessed using cell titer blue assay at 48 hours post-treatment and the calculated IC_{50} values. **(B)** Ewing sarcoma cancer cells were treated with ID, PD, and SD for 48 hours at 0.75 μM and 1.5 μM inhibitor concentrations. Endogenous levels $H3K4me^2$ at each inhibitor concentration was detected using $H3K3me^2$ specific antibody. The $H3K3me^2$ antibody signals were quantitated relative to corresponding anti H3 antibody signal and anti actin antibody signal at each concentration. Anti H3 antibody blots and actin antibody blots were used as controls. The bar graph represents the quantitation of each anti-$H3K4me^2$ antibody signal relative to the corresponding anti actin antibody signal at each inhibitor concentration. SD shows no effect on inhibiting LSD1. **(C)** Ewing sarcoma cancer cells were treated with SD and PD for 48 hours, each at three different inhibitor concentrations. The bar graph represents the quantitation of each anti-$H3K4me^2$ antibody signal relative to the corresponding anti actin antibody signal at each inhibitor concentration. SD shows no effect on inhibiting LSD1.

4.3 Discussion

From my experiments I validate that the putative Ewing's sarcoma and ovarian cancer drug seclidemstat (SP-2577) can bind LSD1 weakly (K_d = 1.3 ± 0.3 µM) and inhibit its demethylation reaction on H3K4me^2 (21aa) model peptide substrate (K_i = 57 ±16 nM). But the quantitative inhibitor based western blot assay demonstrates that seclidemstat does not influence LSD1 catalyzed demethylation on H3K4me^2 nucleosomal substrates. Whereas in vitro inhibition assays using pulrodemstat and TCP reveal that both of these drugs act as inhibitors of LSD1 demethylation reaction on H3K4me^2 nucleosomes (IC_{50} = 4 nM and 14.1 µM, respectively) even though no appreciable inhibition was observed with seclidemstat (SP-2577). Moreover, seclidemstat, and pulrodemstat as well as tranylcypromine derivative Iadademstat were used in cancer cell assays to determine their potency in inhibiting global LSD1 levels. Further supporting my in vitro assays seclidemstat did not inhibit LSD1 demethylation in cancer cells. But pulrodemstat and Iadademstat showed comparatively similar LSD1 inhibitions in Ewing sarcoma cancer cells.

Since seclidemstat exhibit selective mono amine oxidase inhibitory properties, it was originally considered as an optimal candidate drug for LSD1 related cancer therapy to promote anti-tumor immunity in ovarian cancer and suppress the proliferation in colorectal and breast cancer cell lines.(119) For Ewing sarcoma, which is a pediatric aggressive bone tumor characterized by reciprocal chromosomal translocation that results in EWS/ETS fusion oncoprotein, seclidemstat was characterized to reversibly and selectively inhibit LSD1, presumably reducing aberrant gene expression and EWS

progression. (119) However, recently published preclinical testing shows that no in vivo changes in H3K4me^2 levels and no tumor regression was observed in pediatric sarcoma models. (121) In this study, little Seclidemstat induced differentiation and anti-tumor activity was observed across nearly all tested sarcoma xenograft models suggesting that the drug had inconsistent and limited activity in nearly all studied pediatric sarcomas. (121) These in vivo findings are consistent with my quantitative western blot assays that show seclidemstat having no inhibition of LSD1 activity on nucleosomes as well as in Ewing sarcoma (EWS) cell line LSD1 inhibition studies that shows Seclidemstat does not inhibit endogenous LSD1 in EWS cancer cells. Collectively these results demonstrate that Seclidemstat is not a potential candidate for LSD1 inhibition in Ewing sarcoma cancer treatment. Also, most importantly, both in vitro and in cancer cell results show that pulrodemstat which is a reversible LSD1 inhibitor is a better candidate for reversible LSD1 targeted therapy than Seclidemstat in Ewing sarcoma. Furthermore, these findings highlight the need to perform both cell and biochemical substrate selectivity studies on physiologically relevant substrates. This will help us to better understand the complex epigenetic mechanisms associated with cancer pathology and will help us to find better drugs that target these mechanisms.

In conclusion, seclidemstat weakly associates with LSD1 and in contrast to other known reversible and irreversible inhibitors, it is not an effective drug for inhibiting LSD1 activity on H3K4me^2 nucleosomes. Although it may act as a potential Ewing's Sarcoma cancer therapy independent of LSD1 inhibition, seclidemstat should not be considered as an LSD1 inhibitor. I also emphasize the potential drawbacks of using only HRP

coupled assays to determine drug inhibition potency. This work emphasizes the importance of including nucleosome substrate selectivity-based assays, and the need for additional mode of action studies of anticancer drugs targeting chromatin-modifying enzymes.

4.4 Materials and Methods

4.4.1 LSD1-CoREST Inhibition Assays with Seclidemstat Using Model Substrate

Peroxidase-coupled assay that monitors H_2O_2 production under aerobic conditions was used to measure LSD1 catalytic activity in presence of Seclidemstat and 21 a.a. model peptide substrates (H3K4me^2 21 a.a.). A 150 µL reaction mixture containing 50 mM HEPES (Na) pH 7.5, 0.3 µM LSD1-CoREST, 1µg of HRP, 0.1 mM 4-aminoantipyrine, 1.0 mM 3,5-dichloro-2-hydroxybenzenesulfonic acid, varying concentrations of Seclidemstat (0, 10, 50, 200 µM) and peptide substrate concentrations 10, 20, 30, 40 and 60 µM was used for the assay. Enzyme reactions were initiated by the addition of substrate; H3K4me^2 peptide (a.a. 1–21) into the reaction mixture in a quartz cuvette. Cary UV-Visible spectrophotometer was used to measure the changes in absorbance, measured at 515 nm wavelength, at room temperature. Initial velocity calculations were performed using an extinction coefficient of 26,000 M^{-1} cm^{-1} by Graphpad Prism 8. Initial velocity values obtained from absorbance vs. time graph were fitted to competitive inhibition model using non-linear regression fit and obtained K_i (inhibitor constant) value.

4.4.2 LSD1-CoREST Inhibition Assays Using Seclidemstat, Pulrodemstat and Tranylcypromine as Inhibitors in Presence of H3K4me^2 Nucleosomal Substrates

Different concentrations of Seclidemstat (0, 10, 20, 50, 100 nM, 100, 200 µM) or pulrodemstat (pulrodemstat benzenesulfonate, MCE MedChemExpress) (0, 1, 2, 4, 8, 10 nM) or tranylcypromine (0, 0.1, 1, 10, 20, 50, 100 µM) were incubated with the 2 µM ΔN

LSD1- CoREST containing reaction mixture in absence of the nucleosome substrate, at 25 °C for 60 minutes.

Demethylation reactions were initiated by adding 100 nM nucleosomal substrate into the reaction mixture. Aliquots of 10 µL were withdrawn at 0, 2, 5, 30, 120 minutes time points and the reactions were quenched using Laemmli dye followed by boiling for 2 minutes to stop the demethylation reaction. The assay products were resolved by 20% SDS-PAGE gel for 90 minutes at 200 V.

Protein bands in SDS gel were transferred to immunoblot PVDF membranes and blocked with 4% fat free milk in PBS buffer. Corresponding blots were incubated with H3K4me^2 and H3 specific primary antibodies overnight followed by goat anti rabbit secondary antibody for 1 h. The blots were visualized by chemiluminescence and analyzed using Amersham software. All demethylation and control experiments were performed in duplicate (N≥2). The degree of LSD1 catalytic activity inhibition was quantified using H3K4me^2 specific antibody relative to the amount of H3 in each lane quantified using H3 specific antibody. Dividing the H3K4me^2 antibody signal by the H3 antibody signal allowed for accurate quantitation of each data point. The H3K4me^2/H3 ratios were normalized at time zero and plotted as a function of time (minutes), and subsequently analyzed using nonlinear regression. Data were fitted to the equation [H3K4me^2] = [H3K4me^2]$_{t=0}$ e^{-kobst} and the determined rate constant (k_{obs}) values were evaluated with LSD1-CoREST concentrations. The $K_{1/2}$ and k_{max} parameters were determined based on the equation: $k_{obs}= k_{max}$ [Enzyme]/([Enzyme]+$K'_{1/2}$).

4.4.3 Fluorescence Spectroscopy to Measure Binding Between Seclidemstat and ΔN LSD1- CoREST

Intrinsic fluorescence spectroscopy (Photon Technology International (PTI) spectrofluorometer) enabled quantitative equilibrium binding measurements of the seclidemstat - LSD1 interaction. A scanning emission spectrum of 0.1μM ΔN LSD1-CoREST was collected by exciting FAD at 380nm. Fluorescence emission was monitored at 450 – 650 nm wavelength, with optimal FAD emission monitoring at 520 nm. The wavelength scan speed was 60 nm/min while the bandwidth was 4 nm for both excitation and emission light. Samples containing identical LSD1-FAD concentrations were monitored in the absence and presence of increasing amounts of SP-2577 (0, 0.12, 0.23, 0.58, 0.63, 1.27, 1.85, 2.42, 3.55 μM), enabling direct comparison of the flavin emission before and after addition of the ligand. Each concentration of seclidemstat was incubated for 15 min with the 0.1 μM ΔN LSD1- CoREST solution in 25 mM HEPES pH 7.5 containing 100 mM NaCl, 2 mM TCEP in a total volume of 1.7 mL at 25 ^{o}C. As control, a buffer and Seclidemstat alone control was also performed with identical incubation times and excitation measurements at each titration point. Negligible fluorescence was detected in the presence of buffer and increasing amounts of seclidemstat.

As an additional control, a solution containing LSD1-FAD was performed with equivalent amount of DMSO to mimic each titration point. Thus, the relative fluorescence changes can be attributed solely to the interaction between seclidemstat and the FAD moiety within LSD1.

The actual LSD1 and seclidemstat concentrations were corrected for the dilution factor at each titration measurement. The corrected, relative fluorescence changes allowed for determination of the ΔN LSD1- CoREST – seclidemstat dissociation constant. The average intensity change upon each ligand addition was plotted against the concentration of seclidemstat and two independent experiments were fit to a 1:1 binding isotherm (GraphPad Prism).

4.4.4 Cell Viability and IC_{50} Study for Inhibitor Drugs

Cell viability determinations and IC_{50} calculations for determining the cytotoxic potency and selectivity of anticancer drugs were evaluated using Ewing sarcoma (TC32) and osteosarcoma (U2OS) cancer cells and human embryonic kidney (HEK293) cells. (143) Cells were treated in 96-well cell culture plates at the time of their 60–80% confluency with serially diluted 8 different concentrations of the inhibitor drugs pulrodemstat, TCP, seclidemstat and iadademstat (ranging between 0.781 µM and 500 µM) for 48 hours. At the end of the treatment, cell viability was determined using CellTiter-Blue® reagent (Promega, #G8080) by following the manufacturer's protocol. Data was acquired using a Synergy H4 Biotek. The IC_{50} calculations were performed by nonlinear regression analysis (log[inhibitor] vs. normalized response) using GraphPad Prism software.

4.4.5 Western Blot Assay for Cell-based LSD1 Inhibition

In order to determine the effects of drug-induced LSD1 inhibition on total H3K4me^2 levels, the cellular lysates of TC32 cancer cells were treated with seclidemstat, pulrodemstat, or iadademstat at two different concentrations (0.75 μM; 1.5 μM) or left untreated for 48 hours in 6cm cell culture dishes. Total cellular lysate preparations and immunoblotting steps were performed as previously described. (144) The anti H3K4me^2 (EMD Millipore, #07-030), anti H3 (Abcam, #ab1791) primary antibodies, Goat anti rabbit-HRP secondary antibody, anti-ACTIN-HRP (Abcam, #ab49900, AC-15), anti-Rabbit IgG-HRP (Cytiva, # NA934), and anti-Mouse IgG-HRP (Cytiva, # NA9311) antibodies were used for visualizing western blots using chemiluminescence. The blotting images were captured using the LI-COR Odyssey XF imaging system (LI-COR Biosciences) and densitometry data were acquired using Image Studio software (LI-COR Biosciences). The abundancy of H3K4me^2 at each drug concentration was assessed by western blot analysis by quantitating the H3K4me^2 antibody signal relative to the corresponding anti-H3 or anti-Actin antibody signal.

Project Contributions:

LSD1-CoREST constructs purification, all LSD1 inhibition western blot assays using Seclidemstat, TCP and Pulrodemstat, Seclidemstat - LSD1 inhibition by HRP coupled assay, Seclidemstat -LSD1 co-crystallization trials were performed by Dulmi Senanayaka. Fluorescence spectroscopy binding study between LSD1 and Seclidemstat performed by Dulmi Senanayaka and Dr. Danyun Zeng.
Ewing sarcoma cancer cell based LSD1 inhibition assays using Iadademstat, pulrodemstat and Seclidemstat were performed by Dr. Emre Deniz (Georgetown University).

CHAPTER 5 FUTURE DIRECTIONS

5.1 Discovery of LSD1-EWS/FLI1 Fusion Interactions

Understanding the oncogenic mechanisms of EWS/FLI1 and inhibiting LSD1 that ally in oncogenesis is important in uncovering novel pathways to pharmacologically block EWS/FLI1 function. As a first step to pharmacologically develop LSD1 targeted therapy in treatment, I sought to understand the molecular level interactions between histone demethylase LSD1 and oncogenic EWS/FLI1. This project was performed collaboratively with Professor Aykut Uren at Georgetown University, School of Medicine for collaborative cell-based studies with the goal to develop a structure-based understanding of LSD1's interfaces such that the drug discovery pipeline can be improved. The key collaborative questions of the study were as follows.

- Does LSD1 interact with the oncogenic fusion EWS/FLI1?
- Does LSD1-EWS/FLI1 interaction affect LSD1 catalytic activity?
- Is LSD1 inactivation the best approach for Ewing sarcoma therapy?
- Is it possible to develop ways to disrupt the oncogenic properties of EWS-FLI1 and LSD1?

This study discovered for the first time that a direct binding interaction occurs between LSD1 and intrinsically disordered Ewing sarcoma oncoprotein EWS/FLI1 that is found in most of Ewing sarcoma patients and that this interaction inhibits the catalytic activity of LSD1.

5.1.1 ΔN LSD1-CoREST Preferentially Binds EWS/FL1

To determine the in vitro binding between LSD1 and EWS/FLI1, surface plasmon resonance (SPR) experiments were performed on a Biacore 4000 Instrument (GE Healthcare). The purified full-length EWS/FLI1 (68 kDa) was immobilized on an NTA chip (GE Healthcare). The SPR responses were reported as resonance units (RU) where 1 RU corresponds to a 0.0001 degree of change in the reflected light angle. In order to doubly correct for nonspecific binding, each SPR response was subtracted to the buffer-only control injection and also to the control surface with no protein immobilization. (145)

From the experimental data analysis, it was determined that the intrinsically disordered FWS-FLI1 protein (68 kDa) binds tightly with ΔN LSD1-CoREST complex (K_d ~90 nM) suggesting that LSD1-CoREST complex directly interact with FWS-FLI1 (Figure 43-A). As a control, intrinsically disordered transmembrane protein CD99 (25 kDa) that is abundantly expressed in Ewing sarcoma cells and is a diagnostic marker for the disease was used. It is involved in regulating cell differentiation and migration of tumor cells. (146) When measured the binding between CD99 and ΔN LSD1-CoREST, in contrast to EWS/FLI1, CD99 did not bind with ΔN LSD1-CoREST (Figure 43-B).

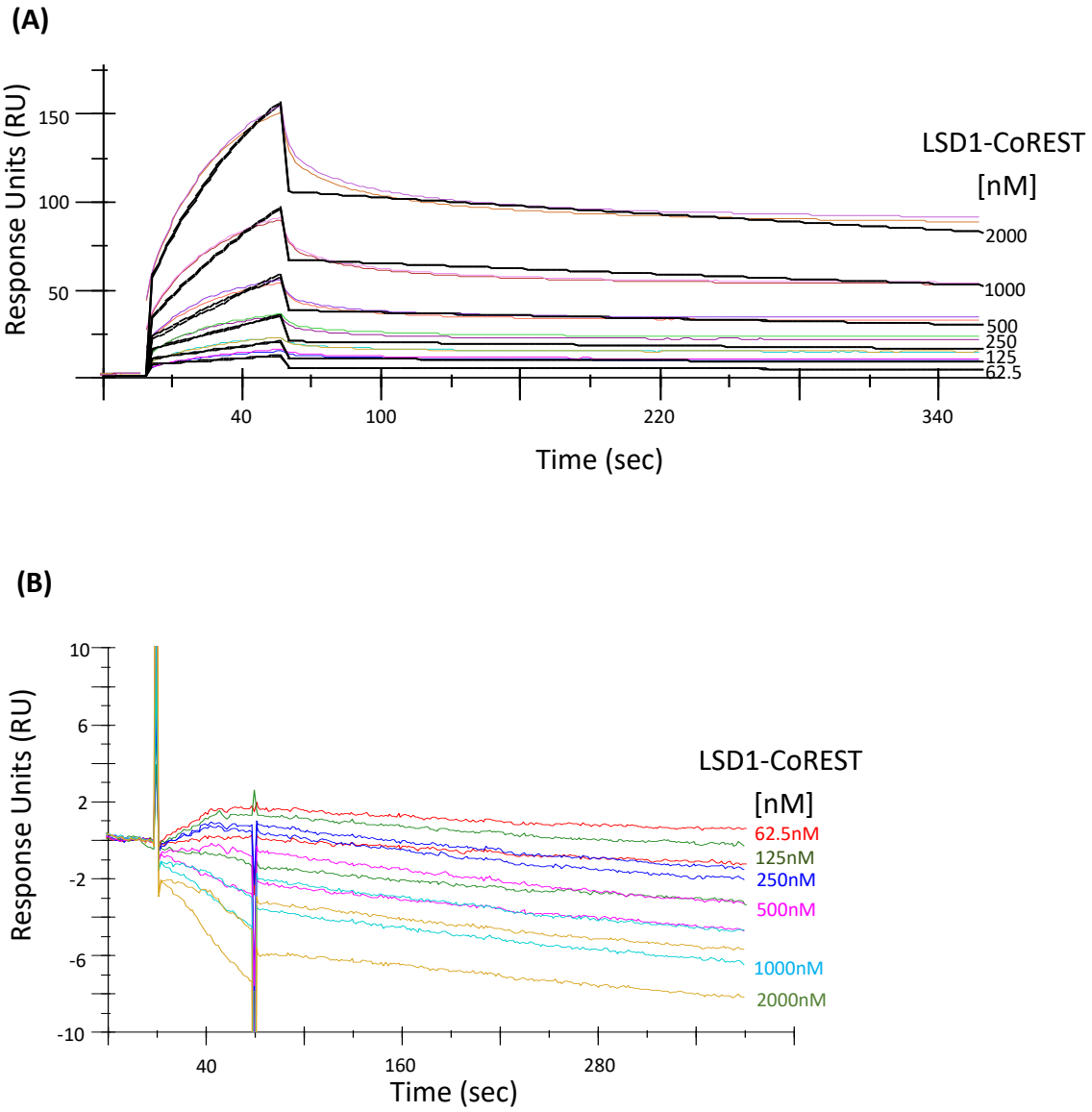


Figure 43: Δ N LSD1-CoREST specifically interacts with EWS-FL1 with high affinity. **(A)** SPR analysis of titration series of Δ N LSD1-CoREST interacting with biotinylated EWS-FL1 immobilized on a sensor chip. (N=2) Plots of the difference in response units (RU) versus time at different protein concentration reveal the apparent binding dissociation constant (K_d = 90 nM). **(B)** SPR analysis of titration series of Δ N LSD1-CoREST with immobilized CD99 on a sensor chip shows that Δ N LSD1-CoREST does not bind intrinsically disordered protein CD99. N =2 (K_d = N/A) (72)

5.1.2 Collaborative Cell-based Studies

Collaborative cell-based studies were performed using Ewing Sarcoma cell line TC32 to determine in cell interactions between EWS/FLI1 and LSD1.

5.1.2.1 LSD1-CoREST Bind with EWS/FLI1 in Ewing Sarcoma Cancer Cells Revealing an "in-cell" Interaction

In vivo binding interaction between EWS/FLI1 and LSD1-CoREST was determined by co-immunoprecipitation of LSD1 and CoREST bound to EWS-FLI1 in TC32 Ewing sarcoma cancer cell line. The protein lysates from TC32 cells were immunoprecipitated with anti-FLI1 antibodies. Immune complexes were resolved using PAGE and immunoblotted for LSD1 and CoREST using anti-LSD1 and anti-CoREST antibodies followed by anti-rabbit HRP secondary antibody. (147)(148)(149) Chemiluminescence analyzed by LiCOR imaging was used to detect bands corresponding to LSD1 and CoREST co-immunoprecipitated with EWS/FLI1 (Figure 44-B, C). For further confirming the presence of LSD1, the bands were cut and (Figure 44-B) protein in each sample were in gel digested with trypsin/Lys-C. After extracting and cleaning the digests, they were analyzed by nano UPLC-MS/MS in HCD mode. Upon database search against the human proteome database, the presence of LSD1 in the corresponding band was confirmed with high confidence. This study further confirmed the binding interaction between LSD1 and EWS/FLI1 in the Ewing sarcoma cells.

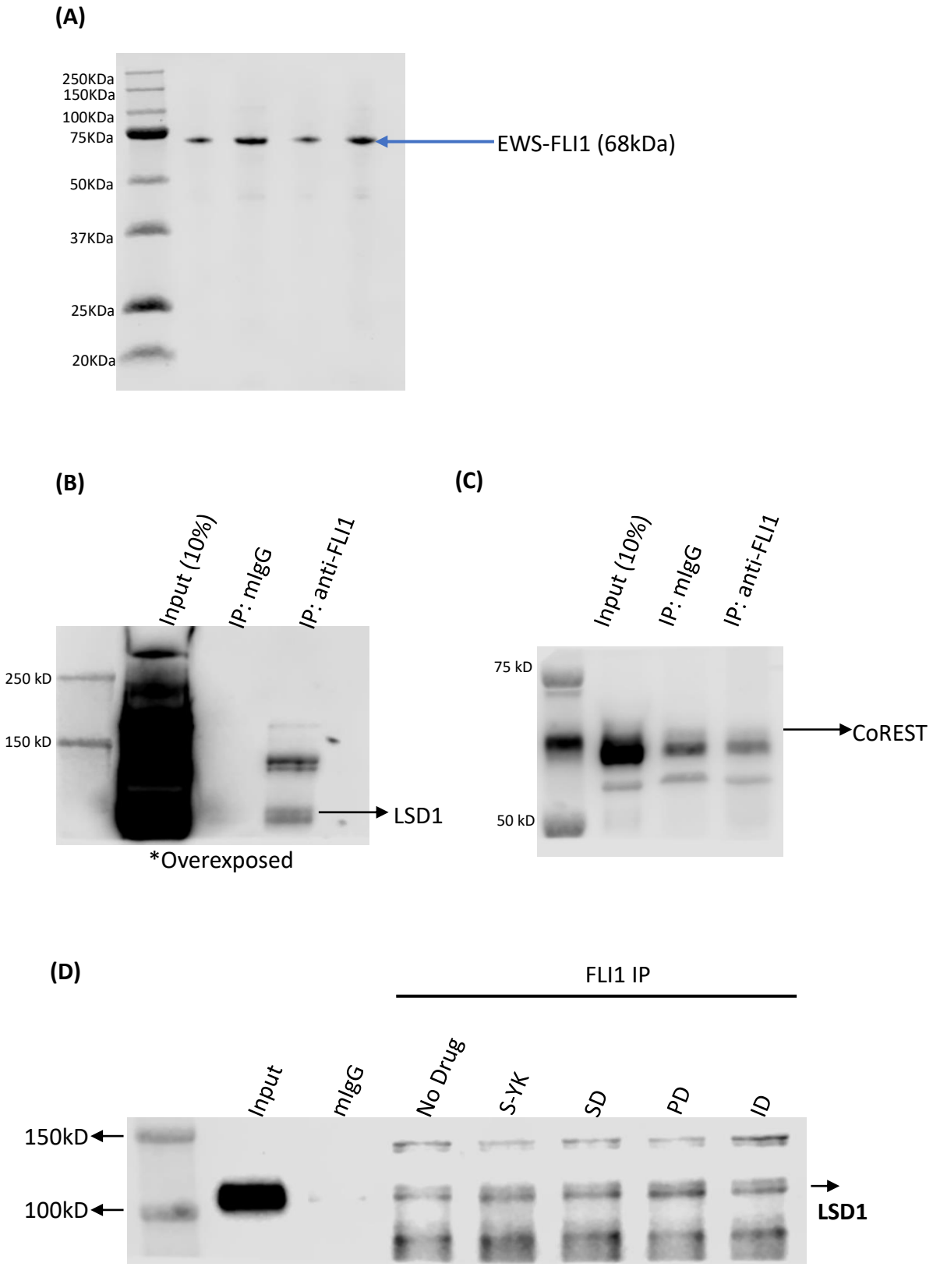

(A)
250KDa
150KDa
100KDa
75KDa
50KDa
37KDa
25KDa
20KDa
EWS-FLI1 (68kDa)
(B)
Input (10%)
IP: mIgG
IP: anti-FLI1
250 kD
150 kD
LSD1
*Overexposed
(C)
Input (10%)
IP: mIgG
IP: anti-FLI1
75 kD
50 kD
CoREST
(D)
FLI1 IP
Input
mIgG
No Drug
S-YK
SD
PD
ID
150kD
100kD
LSD1

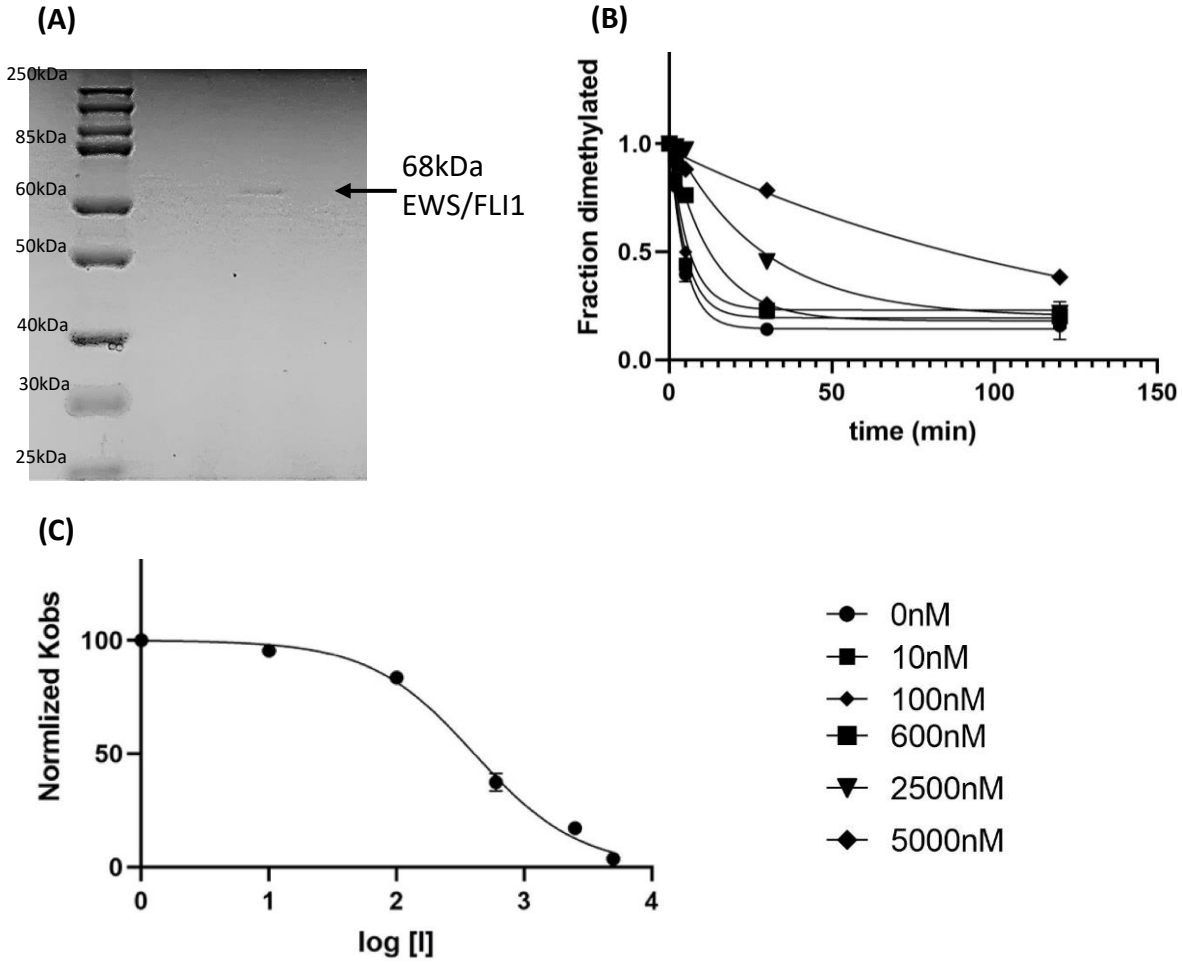
Figure 44: Co-immunoprecipitation of LSD1 and CoREST with EWS/FLI1. **(A)** 10% acrylamide purity gel of EWS/FLI1 purified from inclusion bodies using conventional chaotropic denaturant 8M Urea purification method. (144) **(B)** Representative western blotting for the EWS/FLI1 - LSD1/CoREST complex isolated from Ewing sarcoma cancer cell nuclear extract. LSD1 coimmunoprecipitated with EWS/FLI1 was detected using anti LSD1 antibody. **(C)** CoREST co-immunoprecipitated with EWS/FLI1 was detected using anti CoREST antibody. **(D)** EWS/FLI1 Co-IP in presence of LSD1 inhibitors. The lysate was treated with seclidemstat (SD), pulrodemstat (PD), iadademstat (ID) and S-YK-4-279, an orally bioavailable inhibitor that delays Ewing sarcoma tumor growth.

Next, further co-immunoprecipitation experiments were performed by including seclidemstat (6 μM), pulrodemstat (84 μM) and iadademstat (16 μM) drugs in the cell lysate during co-immunoprecipitation. According to immunoblot images, in presence of inhibitor drugs, LSD1 was still present in the pull-down sample indicating none of these LSD1 targeted inhibitor molecules influence the binding between EWS/FLI1 and LSD1 (Figure 44 D).

5.1.3 EWS/FLI1 Inhibits LSD1-CoREST Catalytic Activity on Nucleosomal Substrate

Since EWS/FLI1 protein was in 8 M urea, 20 mM Na_2HPO_4, 500 mM NaCl buffer after its purification I slowly refolded the protein into 20mM Na_2HPO_4 pH 7.5, 50 mM KCl buffer using 34 nucleotides double-stranded DNA containing canonical ETS binding motif as a chaperon. This DNA contains repetitive GGAA motif that can be targeted by EWS/FLI1. It is found that EWS/FLI1 binds this DNA as a homodimer. (150) Therefore I used this ETS binding DNA as a chaperone to fold EWS/FLI1 stably during the dialysis process. The presence of intact EWS/FLI1 in the sample was confirmed by using 20% SDS PAGE gel (Figure 45-A). Refolded protein was used in quantitative western blot assays under single turnover conditions to test the ability of EWS/FLI1 in inhibiting $H3K3me^2$

demethylation reaction in presence of nucleosomal substrates. Relative kinetics ($k_{observed}$) of LSD1 catalyzed demethylation reaction was obtained from a series of time course assays measuring the fraction of dimethylated nucleosomes quantitated using H3K4me^2 specific antibody relative to the amount of H3 in each lane quantified using H3 specific antibody (Figure 45-D). Dimethylation profiles at increasing EWS/FLI1 concentrations (0, 10, 100, 600, 2500, 5000 nM) revealed that the EWS/FLI1 fusion interacting with LSD1 impacts LSD1 demethylation of nucleosomes with a potent half maximal inhibitory concentration (IC_{50}) of 408 ± 60 nM (Figure 45-C)



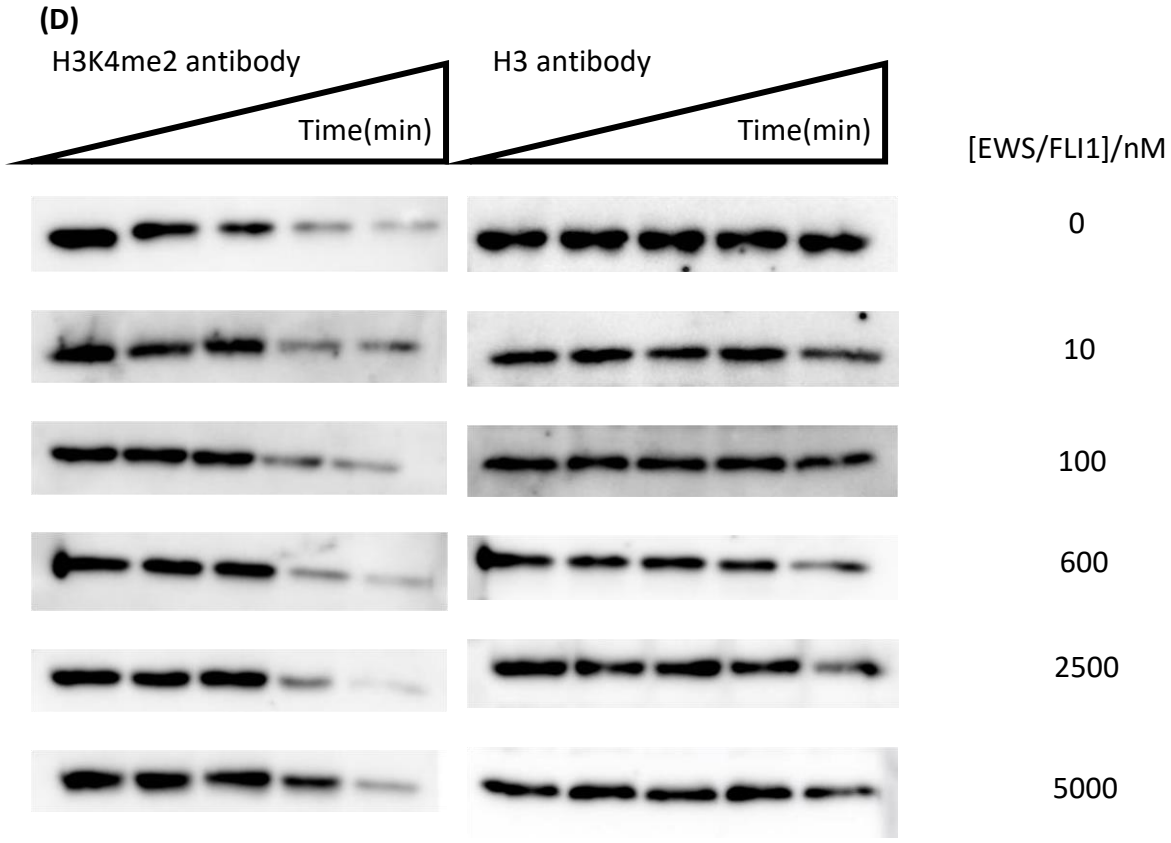


Figure 45: Inhibition of LSD1 catalytic activity by EWS/FLI1 on H3K4me2 nucleosome substrates.
(A) 20% acrylamide SDS PAGE gel of EWS/FLI1 (68 kDa) after its slow refolding into 20 mM KH_2PO_4 pH 7.5, 50 mM KCl buffer using EWS/FLI1 DNA as a chaperon. **(B)** Fraction dimethylated versus time graph plotted by quantification of residual dimethylation from western blots (n=2). **(C)** Activity % versus log [I] graph of EWS/FLI1 inhibition of LSD1 activity in presence of $H3K4me^2$ nucleosomal substrate. **(D)** Representative western blots for EWS/FLI1 inhibition of LSD1/CoREST catalyzed demethylation.

5.1.4 Discussion and Future Directions for Studying LSD1-EWS/FLI1 Interactions

This collaborative work was able to determine the first aim of the study: the interactions between catalytically active LSD1-CoREST complex and oncogenic mediated intrinsically disordered protein (IDP) EWS/FLI1. The SPR binding data detected tight

binding (K_d ~ 90 nM) between LSD1-CoREST and EWS/FLI1 in vitro. *In vivo* co-immunoprecipitation assays further confirmed that both LSD1 and CoREST directly bind with EWS/FLI1 in TC32 Ewing sarcoma cell line suggesting in cell interaction between LSD1, CoREST and EWS/FLI1. Next aim of this study was to determine if LSD1-fusion protein interaction affect the demethylase activity of LSD1. From the quantitative western blot studies demonstrated that the binding between EWS/FLI1 and catalytically active LSD1-CoREST exert an inhibitory effect on LSD1's nucleosomal substrate demethylation ability (IC_{50}= 408.7 ± 60 nM) *in vitro* suggesting that this in cell interaction can regulate LSD1 catalytic activity and its associated regulation of gene expression in Ewing sarcoma. In fact, this inhibition can be of key importance in understanding EWS/FLI1 function. These results open new pathway for conducting cancer based biochemical studies of the LSD1 complex and explore LSD1's druggability in treatment of Ewing sarcoma.

As a next step of this project, LSD1-EWS/FLI1 binding interactions can be further validated using cross linking mass spectrometry that can be used to map local protein-protein connectivity information within the LSD1- EWS/FLI1 complex. (151)(152) One of the next aims of this project is to perform small molecule screen for disruptor or destabilizer of the EWS-FLI1 – LSD1 complex. Once the protein-protein interaction (PPI) surfaces between EWS-FLI1 and LSD1 is determined by cross linking mass spectrometry, these interactions can be modulated by more ways other than direct inhibition of LSD1, through noncompetitive modes. Although my work in chapter 4 suggests Pulrodemstat as a potential lead molecule in EWS therapy through inhibition of LSD1 catalytic activity,

targeting root cause of EWS can be performed through disruptors as well. Disrupting protein-protein complex formation through direct competition at the PPI interface by orthosteric disruptors or with allosteric destabilizers that modulate the complex protein-protein interactions by binding remotely can be used as new avenues for targeting EWS-FLI1 – LSD1 complex in Ewing sarcoma therapy. Once the small molecule screen is performed, the positive hit compounds can be used for cell viability assay testing for the cytotoxic effect of the compounds. (153)

As a preliminary experiment we performed SPR studies of LSD1 binding to EWS/FLI1 in presence of previously identified LSD1 nucleosomal substrate demethylation inhibitor pulrodemstat. According to the quantitative western blot assays pulrodemstat exerted better inhibition on nucleosome demethylase activity compared to seclidemstat. Preliminary SPR data using pulrodemstat shows that compared to LSD1 binding with EWS/FLI1, pulrodemstat does not bind with immobilized EWS/FLI1 (Figure 46 A). Whereas LSD1, in presence of pulrodemstat (a non-disruptor), bind EWS/FLI1 similar to the LSD1 binding in absence of pulrodemstat. As a control experiment, SPR binding study was performed between EWS/FLI1 and ETS inhibitor TK216 that directly binds EWS/FLI1 inhibiting protein interactions (Figure 46 B). Preliminary data demonstrated tight binding between EWS/FLI1 and TK216 similar to the binding between EWS/FLI1 and LSD1. In presence of both LSD1 and TK216, binding affinity between LSD1 and EWS/FLI1 is lower suggesting that TK216 act as a disruptor of protein-protein binding interactions between EWS/FLI1 and LSD1. This method can be used as a way to further confirm the binding disruptors and destabilizers.

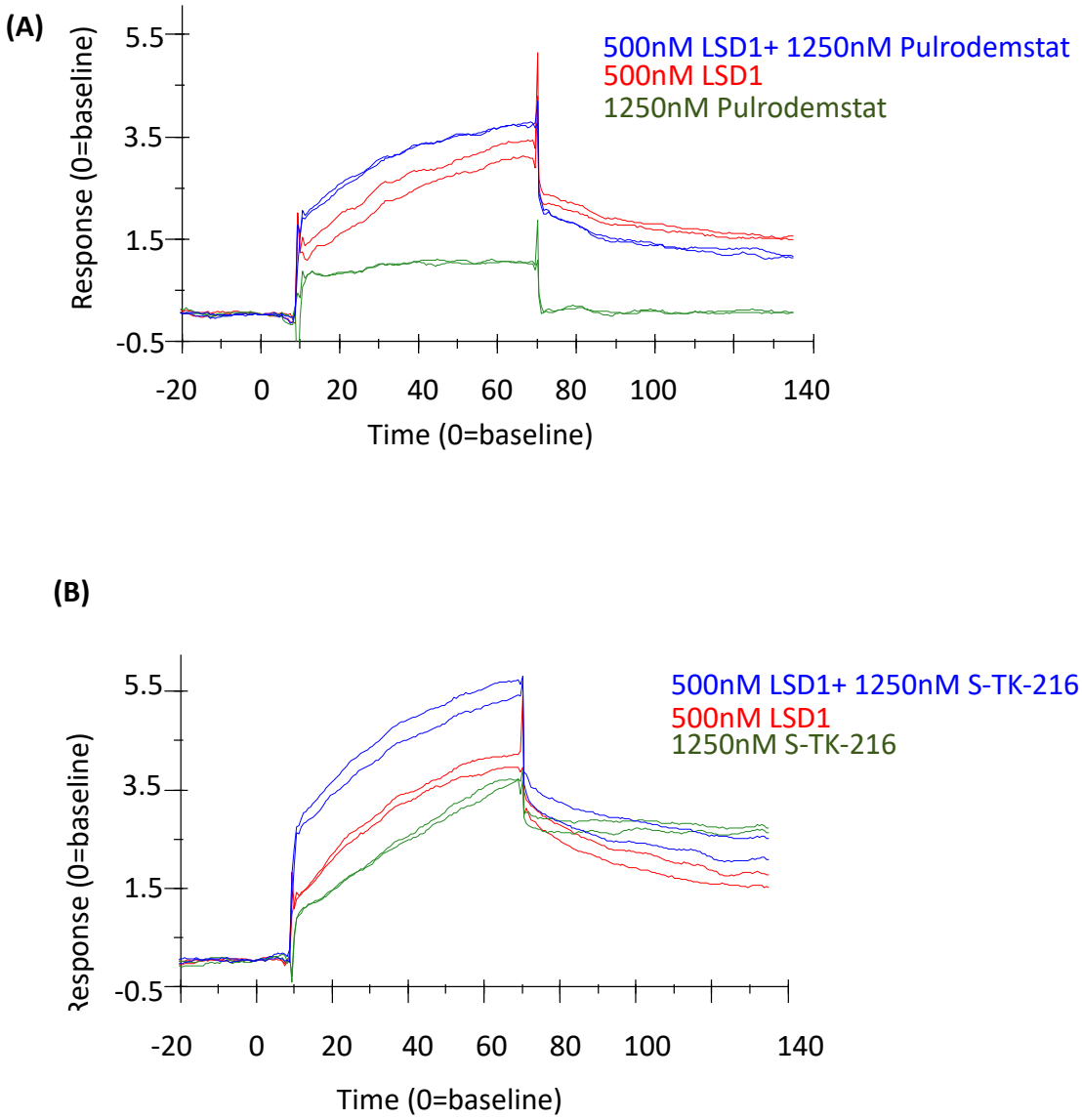


Figure 46: Preliminary SPR binding study to determine EWS/FLI1- LSD1 binding disruptors.
(A) Preliminary SPR data showing that LSD1 binding immobilized EWS/FLI1 in presence of pulrodemstat, similar to the LSD1 itself binding to EWS/FLI1. Pulrodemstat alone does not bind EWS/FLI1. **(B)** Control experiment showing ETS inhibitor TK216 binding to EWS/FLI1 similar to LSD1 binding EWS/FLI1. Decreased binding between LSD1 and EWS/FLI1 in presence of TK216.

In addition to these findings of LSD1 interacting with EWS/FLI1, previous studies have shown that EWS/FLI1 interact with HDACs suggesting a combination approach in targeting LSD1 and HDAC inhibition. (154) HDAC2 and HDAC3 that serve as catalytic subunits of NuRD transcriptional corepressor complex, play an important role in indirect transcriptional repression mediated by NKX2.2 transcription factor in Ewing sarcoma. Also, in Ewing sarcoma cells, EWS/FLI1 have co-immunoprecipitated with histone deacetylases HDAC2 and HDAC3 suggesting direct binding between HDAC2 and 3 with EWS/FLI1. (124) Furthermore, it is shown that HDAC inhibitor vorinostat treatment causes de-repression of EWS/FLI bound and downregulated genes suggesting that HDACs are involved in direct transcriptional repression of genes including TGFBR2 and LOX by EWS/FLI in Ewing sarcoma. Furthermore, in vivo, HDAC inhibitors have shown efficacy in some pre-clinical xenograft models.(155) Therefore, taken together it is possible that single agent chemotherapy cannot serve as the sole treatment for the Ewing sarcoma patients. Synergistic approaches including combining NuRD inhibitors such as LSD1 inhibitor pulrodemstat and HDAC inhibitors that potentiate Ewing sarcoma chemotherapy seems to be a better strategy to achieve maximal effect in treating Ewing sarcoma. (119)

Another aim for this project is to find the mode of interaction of EWS/FLI1-FL LSD1 complex including CoREST and HDAC using structural biology and to test disruptor/destabilizer small molecules targeting these protein-protein interactions. As a first step, negative stain electron microscopy followed by cryo-electron microscopy (Cryo-EM) can be used to determine the structure of EWS/FLI1-LSD1-CoREST-HDAC

complex and the physical relationship between members of the complex. Real time NMR assay can be used to characterize the crosstalk between EWS/FLI1, LSD1 and HDAC in the CoREST complex. (44) Upon identification of a potential small molecule disruptor/destabilizer or inhibitor targeting these interactions, moving towards a mouse xenograft model system that can detect the efficacy of the molecule in disrupting/destabilizing EWS/FLI1-LSD1-CoREST-HDAC complex or inhibiting the LSD1/ HDAC activity in order to know whether Ewing sarcoma tumor will respond to the specific therapeutic regime is the final goal of this project.

5.1.5 Materials and Methods

5.1.5.1 EWS-FLI1 Refolding Using EWS/FLI1 DNA as a Chaperon

Purified EWS/FLI1 fusion protein (68 kDa) in final 8 M urea, 20 mM Na_2HPO_4 pH 7.5, 500 mM NaCl buffer was refolded using slow buffer exchange into no urea buffer in presence of EWS/FLI1 DNA for facilitating the folding. Button dialysis of 40 µL of EWS/FLI1 fusion combined with 10 uL of 20 uM double stranded EWS/FLI1 DNA (20 kDa) was performed using a 3 kDa dialysis membrane. The buttons were dialyzed first into 6M urea, 20mM Na_2HPO_4 pH 7.5, 200mM KCl, 1mM BME solution, secondly into 4 M urea, 20 mM KH_2PO_4 pH 7.5, 50 mM KCl, 1 mM βME solution followed by 2 M urea, 20 mM KH_2PO_4 pH 7.5, 50 mM KCl 1 mM βME solution and finally into 0 M urea, 20 mM KH_2PO_4 pH 7.5, 50 mM KCl solution respectively to remove urea and refold the protein. Control sample of 40 µL 8 M urea, 20 mM Na_2HPO_4 pH 7.5, 500 mM NaCl buffer was dialyzed in same conditions as the protein samples after combining with 10 µL of EWS/FLI1 DNA. The stable protein presence was confirmed by 20% SDS PAGE gel.

5.1.5.2 LSD1-CoREST Demethylation Assays in Presence of EWS/FLI1 Using H3K4me^2 Nucleosome Substrate

Different concentrations of EWS/FLI1 protein (0, 10, 100, 600, 2500, 5000 nM) were incubated with 2µM ΔN LSD1-CoREST containing reaction mixture in absence of nucleosomal substrate at 25 ˚C for 60 minutes. Demethylation reactions were initiated by adding 100 nM nucleosome substrate into the reaction mixture. Aliquots of 10 µL were withdrawn at t=0, 2, 5, 30, 120 minutes time points and quenched using Laemmli dye followed by boiling for 2 minutes. The assay products were resolved by 20% SDS PAGE and the degree of inhibition was quantified using H3K4me2 specific antibody relative to the amount of H3 in each lane quantified using H3 specific antibody. Control experiment was performed at 0 nM EWS/FLI1 protein concentration but using 40 µL of buffer dialyzed with 10 µL of 20 µM EWS/FLI1 DNA.

Project Contributions:

LSD1-CoREST constructs purification, EWS-FLI1 refolding and all LSD1 inhibition western blot assays using EWS-FLI1, controls and corresponding data analysis were performed by Dulmi Senanayaka.
Ewing sarcoma cancer cell based LSD1 and CoREST coimmunoprecipitation, all SPR binding experiments and corresponding data analysis were performed by Dr. Emre Deniz (Georgetown University).

5.2 TERRA RNA Structure Elucidation by NMR Spectroscopy

5.2.1 Introduction

Previous studies demonstrated that LSD1 can recognize and preferentially bind stacked, parallel-stranded G-quadruplex (GQ) ($K_D \approx 96$ nM, Chapter 3), and that this GQ RNA binding site overlaps with a nucleosome binding site within the SWIRM/AOD of LSD1. (60) Later phase separation studies suggested that GQ TERRA RNA binds LSD1, and preferentially drive LSD1 phase separation in a length and structure dependent manner for ALT telomere maintenance through R loop formation (Chapter 3). But it's not much known about how the dynamics of higher order G rich RNAs affect gene regulation. I think that further characterization of a higher order TERRA structure in condensates is important for determining how LSD1 impact TERRA in solution and for determining the RNA structural transitions and sequence requirements that drive the protein-RNA interactions leading condensate formation at telomeres. Therefore, the aim of this project is to understand how higher order GQ RNA structures rely on phase separation to engage with chromatin remodeling enzyme complexes as LSD1 to regulate gene expression through an NMR spectroscopic approach.

Using NMR spectroscopy, the local and global conformations of TERRA RNA structure in solution and in condensed phase can be examined to understand how the chromatin remodeling enzyme complexes as LSD1 and GQ forming RNA interact to work together and regulate chromatin functions mechanistically. The previously determined TERRA structure at 2.2 Å (PDB 3IBK) shows sequence complexity and canonical GQ stereochemistry but does not provide the flexibility and hydrogen bond tertiary contacts

that occur within and between the RNA motifs. However, solution NMR studies that probe ^{1}H proton environment only, have revealed that there is some variability in the ribose sugar pucker in 8X TERRA with bimolecular stacked GQ containing three G quartet tiers. But it's hard to obtain key details such as H bonding across the G quadruplex or the dynamics of the loop region (UUA) without a ^{13}C, ^{15}N labeled RNA. To date no studies of the local H bonds that form across the TERRA G quartet or about the existence of variable backbone inversions in higher order repeats are available.

Therefore, NMR studies of TERRA RNA that define the new H bond couplings as well as the conformational flexibility at the interface between higher order RNA motifs will help us to better understand the TERRA's role in phase separation.

5.2.2 Results and Discussion

5.2.2.1 Defining the Structural Transitions of TERRA in Solution

Two ^{13}C, ^{15}N guanosine, ^{13}C, ^{15}N adenosine labeled TERRA samples were prepared (51 mer/8X TERRA and 10 mer/mini-TERRA) and NMR spectra were collected for comparison. The spectra revealed that the short 10mer r(GGGUUAGGGU) is similar, but not identical, to the stacked 8X TERRA [5'-r(UUAGGG)$_8$ UUA-3'] (Figure 47). Through this kind of experiments, the length dependent effect of TERRA in protein-RNA interactions that leads to condensate formation at telomeres can be determined.

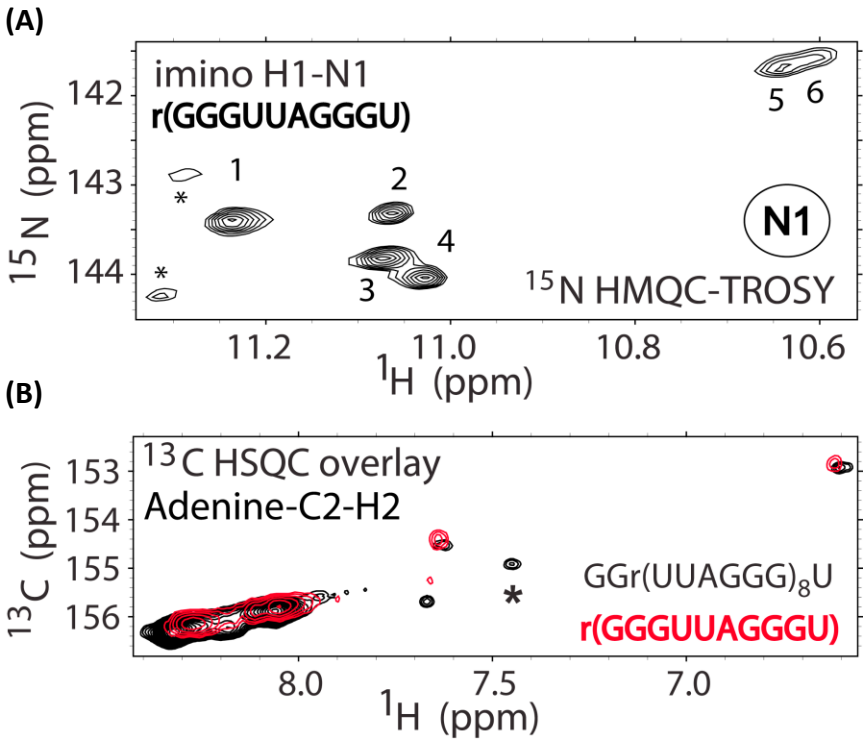


Figure 47: ^{13}C, ^{15}N TERRA ^{15}N HMQC-TROSY and ^{13}C HSQC spectra overlay. **(A)** ^{15}N HMQC-TROSY reveals 6 imino (N1) from 6 G nucleotides in mini-TERRA **(B)** ^{13}C HSQC spectral overlay of 11mer (red) and 51 mer (black). Differences are noted (asterisks). Comparative NMR helps us to understand the role of RNA structure in TERRA-LSD1 LLPS.

In order to monitor H bond couplings across the G quartet, selective HCN pulses were developed (Figure 48). In this spectrum weak couplings less than 1Hz can be observed across the G quadruplex nucleobases that allows to monitor H bonding properties of TERRA. Therefore, this spectrum is also helpful in understanding interactions between LSD1 and GQ RNA within condensates.

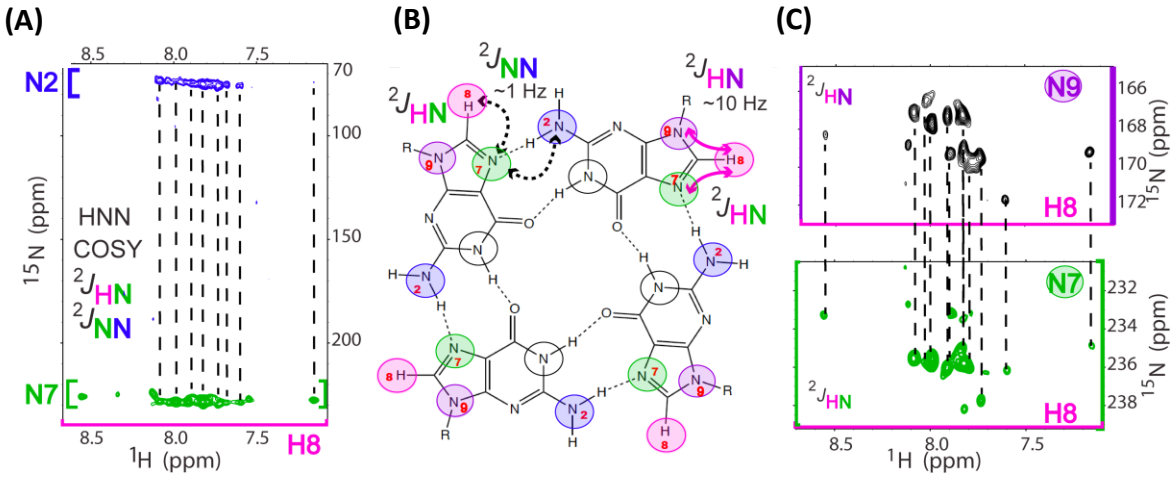


Figure 48: Novel scalar couplings across a GQ tetrad.
(A) Long range HNN COSY reveals two step couplings ($^2J_{HN}$, $^2J_{NN}$), directly detecting GQ Hoogsteen pairing.
(B) One G-quartet with color coded atoms corresponding to connectivities (A, C).
(C) HCN couplings ($^2J_{HN}$) from selective ^{15}N HMBC experiments validate HNN COSY data.

Since SWIRM domain is shown to primarily interact with RNA, chemical shifts of RNA can be mapped upon addition of SWIRM domain of LSD1. This will further the understanding of the specific regions of GQ RNA interacting with SWIRM domain of LSD1 during phase separation.

5.2.3 Materials and Methods

5.2.3.1 TERRA RNA Transcription

10mer TERRA (mini-TERRA) and 51 mer (8x TERRA) transcription reactions were performed as follows. The DNA annealing was performed by adding 10µM template DNA, 12µM T7 promoter, 1x transcription buffer together and annealing at 95˚C, 2 minutes, at 50˚C ,2 minutes and at 25 ˚C, 5 minutes respectively by using the

thermocycler. RNA transcription reaction was set up by adding 5mM DTT, 5mM ATP/CTP/UTP, 7mM GTP, 1x transcription buffer, annealed DNA, T7 polymerase together and incubating the reaction mixture at 37 ˚C for 4 hours. The RNA was precipitated by adding 0.1 volumes of 3M sodium acetate pH 7.5, five volumes of ice cold 100% ethanol and keeping the reaction mixture in -20˚C overnight. The next day, RNA was spined down at 4300 r.p.m for 30 minutes. The pallet was collected and dried.

5.2.3.2 Adenine, Guanine, C^{13} N^{15} Labeled TERRA RNA Transcription

RNA transcription reaction was set up by adding 5mM DTT, 5mM $C^{13}N^{15}$ATP/CTP/UTP, 7mM $C^{13}N^{15}$GTP, 1x transcription buffer, annealed DNA, T7 polymerase together and incubating the reaction mixture at 37 ˚C for 4 hours. The RNA was precipitated by adding 0.1 volumes of 3M sodium acetate pH 7.5, five volumes of ice cold 100% ethanol and keeping the reaction mixture in -20˚C overnight. The next day, RNA was spined down at 4300 r.p.m for 30 minutes. The pallet was collected and dried. Cy3-UTP labeled 8x TERRA and mini-TERRA (10mer) RNA were transcribed by adding 5mM DTT, 5mM ATP/CTP/Cy3-UTP,3mM UTP, 7mM GTP, 1x transcription buffer, annealed DNA, T7 polymerase, together and performing the transcription similarly as above described.

5.2.3.3 TERRA RNA Purification

Dried RNA pallet was loaded to 8M urea, 1x TBE, 8% - 20% acrylamide (depending on the length of RNA) gel. The bands corresponding to desired RNA was cut out and RNA was extracted using 0.3M sodium acetate pH 7.5, and 100% ethanol before desalting using BIO-RAD Micro Bio-spin columns.

5.2.3.4 Nuclear Magnetic Resonance (NMR) Spectroscopy for TERRA RNA

Mini TERRA unlabeled and $C^{13}N^{15}$ A, G labeled RNA 600µM samples were prepared in 70mM KCl, 20mM KH_2PO4 pH 7 buffers containing 100% D_2O and 100% H_2O respectively. TERRA 8x unlabeled and $C^{13}N^{15}$ A, G labeled RNA 700µM and 900µM samples were prepared in 70mM KCl, 20mM KH_2PO4 pH 7 buffers containing 99.9% D_2O and 100% H_2O respectively.

Project Contributions:

All RNA constructs synthesis, purification and NMR sample preparation were performed by Dulmi Senanayaka.
NMR experiments designing and data analysis was done by Dr N. J. Reiter.

CHAPTER 6 APPENDIX

6.1 The Conserved IDR of the N-Terminus Binds ΔN LSD1 and Contains Alpha Helical Structure

Given the functional complexity of the LSD1, I sought to determine whether its own conserved N-terminus (NT residues 100-151) contains any structural features and whether it interacts with nucleosome core particle (NCP) or its own LSD1 interface. Spectroscopic studies (CD, NMR) and X-ray crystallography (PDB 6WC6) of N terminal region (aa 100-151) reveal formation of an α-helical structure (NT-helix) although the relevance of this single helix has yet to be established.

The SPR studies show that the LSD1's conserved IDR (residues 100-151) binds weakly to the nucleosome, but strongly to ΔN LSD1-CoREST. This suggests the conserved NT region (aa 100-151) of LSD1 interacts with itself. The CD spectroscopic analysis of N terminal peptide (aa 100-151) revealed maximum negative ellipticity at 210nm and 220nm indicative of global α-helical structural features residing within the N-terminus of LSD1 (Figure 49). In addition, a series of short-range and medium range inter-residue proton-amide proton NOE resonances assigned in ^{15}N-NOESY-HSQC spectra acquired for N-T LSD1 identified the formation of a stable α-helix between 135-146 amino acids (P2 peptide region) in the N terminus of LSD1.

The X-ray crystallographic structural analysis of LSD1-CoREST bound to 15 amino acids P2 peptide (aa 137-151) reveals an intramolecular stable P2-α-helix wedged to a binding site within a relatively large, preformed cleft in the amine oxidase domain of

LSD1 suggesting a novel protein-LSD1 regulatory interface (PDB 6WC6) (Figure 50). This binding location is a bi lobed cleft made from flexible loops and β-turns and it is structurally conserved among FAD dependent mono amine oxidases. In the recent publication of the crystal structure of ΔN LSD1-CoREST bound to nucleosome, there is a second binding mode where LSD1 interacts with histone H2B at the octamer surface of nucleosome through the same catalytic AOD - P2 α helix binding site. (59) This suggests that the alpha helical based electron density may represent a functionally relevant binding location of P2 peptide. Furthermore, the identification of structure and intramolecular contacts of N terminal regions suggest how the flexible, dynamic N-terminus contributes to the LSD1 interaction network.

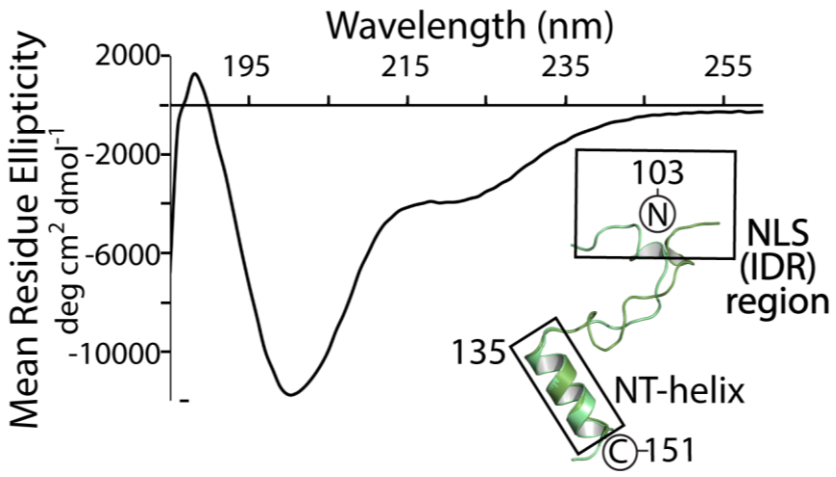


Figure 49: Spectroscopy of LSD1 (residues 100-151) reveals a transient α-helix structure. Collectively NMR and CD spectroscopy indicate a short 15 residues helix (residues: 135-151).

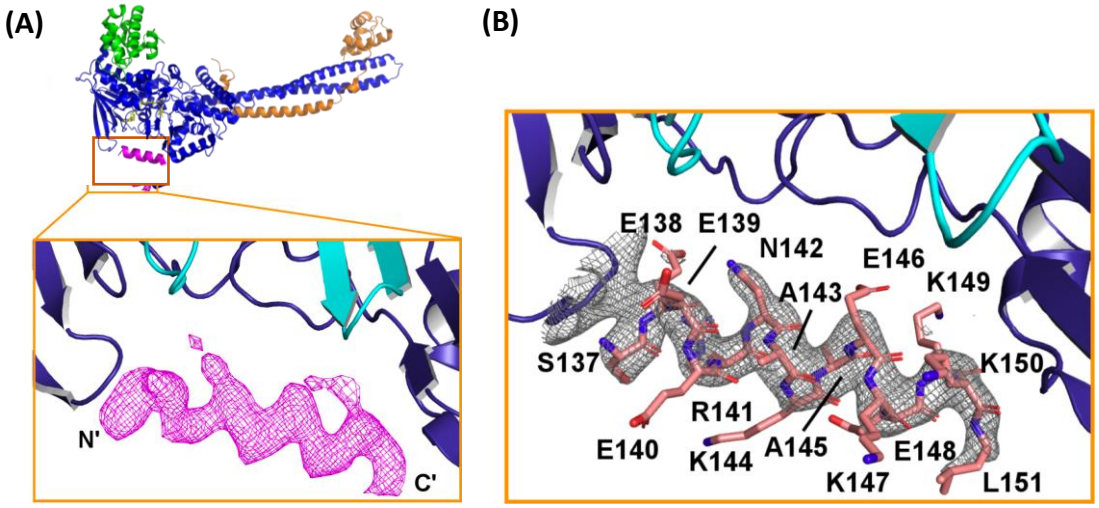


Figure 50: The co crystal structure of alpha helical P2 peptide bound to LSD1 structure at a surface groove of AOD that may represent a functionally relevant binding location. (PDB ID: 6WC6)

(A) The structure of the LSD1-CoREST complex in the presence of N-terminal residues 137-151. The view is centered on an α-helix that was visible in a 3.10 Å difference Fourier map (mFo-DFc), magenta mesh at 2.85 r.m.s.d contour levels. **(B)** The refined electron density map (2Fo-Fc) and the directionality of the α-helix (PDB 6WC6). The map is represented as a dark grey mesh, contoured at 1.38 r.m.s.d. The peptide model is shown as sticks, with the colors corresponding to atoms: carbon-pink, nitrogen-blue, and oxygen-red.

Table 4: Data Collection and Refinement Statistics of the structure of human LSD1-CoREST in complex with NT-LSD1 peptide (residues 137-151), PDB: 6WC6.

X-Ray Data	
Wavelength, (Å)	0.97872
No. Of Unique reflections	46959
Resolution range, (Å)	142.46-3.10 (3.21-3.10)
Space group	*I222*
Temperature, K	100
Cell Dimensions	
a (Å)	121.78
b (Å)	179.16
c (Å)	234.91
Number of molecules per asymmetric unit	1
No. of unique reflections	46,959 (4568)
Completeness (%)	99.9 (88.1)
I/sigma (I)	13.7 (1.6)
Redundancies	15.0 (15.3)
R_{pim}	0.060 (0.656)
CC1/2 (last shell)	0.998 (0.764)
Refinement statistics	
Resolution range	100.72-3.10
No. of reflections used in refinement	46920
R_{work}, R_{free}	0.216, 0.236
Number of Refined atoms	
Protein	6417
Ligand	53
Average B-factors (Å^2)	
Protein	102.2
Ligand	65.8
R.m.s. deviations	
Bonds(Å)	0.003
Angles(^{o})	0.587

6.1.1 Materials and Methods

6.1.1.1 ΔN LSD1-CoREST Crystal Soaking with P2 Peptide

The LSD1-CoREST complex was crystallized as above. Briefly, the LSD1-CoREST complex was buffer exchanged in Amicon Ultra 50K Centrifugal Filters into a solution containing 25 mM HEPES-Na, pH 7.5, 100 mM NaCl, 5 mM TCEP, 1 mM PMSF and crystallized by the hanging drop method at 20^{o}C. The LSD1-CoREST complex [10-12 mg/ml concentration in 25 mM HEPES-Na, pH 7.5, 100 mM NaCl, 5 mM TCEP, 1 mM PMSF] was mixed with the reservoir solution [0.60 M Li_2SO_4, 0.63 M $(NH4)_2SO_4$, 0.25 M NaCl, 100 mM Na-citrate, pH 5.6, 10 mM DTT] at 1:1 ratio. The crystals containing the N-terminal peptide belong to the orthorhombic I222 space group (a = 123.86 Å, b = 179.37 Å, c = 235.05 Å) and diffract to 3.1 Å resolution. The N-terminal peptide region of LSD1 (P2 peptide/aa residues 137-151) purchased from Genscript Inc. (purity > 95%), was dissolved, filtered, buffer-exchanged, and subsequently introduced into the crystals by the soaking method. (60) (156) The LSD1-CoREST complex crystals were incubated for 0.5-1 hour in the peptide ligand in a solution containing 5 mM NT-peptide, 0.76 M Li_2SO_4, 0.63 M $(NH4)_2SO_4$, 0.25 M NaCl, 5mM TCEP, and 100 mM Na-citrate, pH 5.6. Soaked crystals were cryoprotected in an identical solution that contained 23% (v/v) glycerol. After soaking, crystals were harvested and flash-cooled in liquid nitrogen.

6.1.2 ΔN LSD1-CoREST-P2 Crystal X-Ray Data Collection

For the P2 peptide soaked ΔN LSD1-CoREST crystals, diffraction data were collected at 100 K at LS-CAT beamline 21 (F hutch), advanced proton source, Argonne National Laboratory. Data were processed and scaled using the DIALS (157). Molecular replacement was applied using Phaser (80) in CCP4 program suite (158) and the structure of LSD1-CoREST complex (PDB 4XBF) was recruited as the initial template. The difference Fourier map ($|F_o|-|F_c|$) identified the location of an N-terminal peptide site within the LSD1-CoREST crystals. A view of the monoamine oxidase domain adjacent to the FAD binding site reveal regions of density characteristic of a right-handed helical structure with amino acids spaced approximately 3.6 Å per turn. Coot was used for model building throughout the refinement and the P2 peptide (residues 137-151 of LSD1) was built manually. (159) (160) (161) Data collection and data processing statistics are summarized in Table 4.

Project Contributions:

LSD1-CoREST constructs purification and crystallization were performed by Dulmi Senanayaka. Crystal soaking experiments and data collection performed by Dulmi Senanayaka and Dr. Danyun Zeng. CD spectroscopy performed by Dr Danyun Zeng. X-ray diffraction data analysis and PDB deposition by Dr. N. J. Reiter and Dr. Danyun Zeng.

6.2 Molecules Tested in the Studies

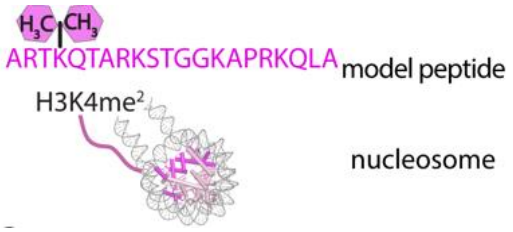
Table 5: Peptide fragments tested in the Chapter 2.
The nuclear localization signal (NLS) regions (light blue) and amino acid modifications (red) are indicated, with numbering corresponding to human LSD1. With exception of NT-LSD1, all peptides contain acetylation and amidation modifications at the N- and C-termini, respectively.

LSD1

Name	Peptide region
NLS-LSD1 (NLS)	^{107}GRRTSRRKRAKVEY^{120}
Nt LSD1-P1 (P1)	^{121}REMDESLANLSEDEYY^{136}
Nt LSD1- P2 (P2)	^{137}SEEERNAKAEKEKKL^{151}
NLS-LSD1 phosphomimic (NLSp*)	^{107}GRRDDRRKRAKVEY^{120}
NLS-LSD1 K114me (NLS-K114me2)	^{107}GRRDDRRKme2RAKVEY^{120}
N-terminus peptide (NT-LSD1)	^{100}GIAETPEGRRTSRRKRAKVEYREMDESLANLSEDEY YSEEERNAKAEKEKKL^{151}
N terminal α helical region	^{135}YYSEEERNAKAEKEKKL^{151}

H3

Name	Peptide region
H3	^{1}ARTKme2QTARKSTGGKAPRKQLA^{21}



model peptide (^{1}ARTKme2QTARKSTGGKAPRKQLA^{21})

nucleosome (H3K4me^{2} mono nucleosome)

Figure 51: LSD1 substrates used in Chapter 3.
21 amino acid H3 N terminal tail peptide containing H3K4me^{2} and recombinant Human H3K4me^{2} mono nucleosome.

Table 6: DNA, RNA transcripts used in Chapter 4 and Chapter 5.

RNA Name	Number of nucleotides	Sequence
Mini TERRA	10	5'-r(GGGUUAGGGU)-3'
4X TERRA	28	5'-r(UUAGGG)$_4$ UUAU-3'
8X TERRA	51	5'-r(UUAGGG)$_8$ UUA-3'
12X TERRA/ Cy3-UTP labeled 12X TERRA	74	5'-GG (UUAGGG)$_{12}$ -3'
20X TERRA	122	5'-GG (UUAGGG)$_{20}$ -3'
28X TERRA	170	5'-GG (UUAGGG)$_{28}$ -3'
FAM57B	28	5'-r(GGUGGAGGGUGGGAGGGUGUACGGGGAA)-3'
MYO1B	33	5'-r(UGGGUGUGGGUAGGGAAGAGGGGAACACUGGGU)-3'
PolyU	32	5'-r(UUUUUU)$_5$ UU -3'

EWS/FLI1 DNA (34 bases)-contain canonical ETS binding motif 5'-ACCGGAAG-3' and repetitive GGAA motifs targeted by EWS/FLI1 in binding genes.

5'-ACCGGAAGGAAGGAAGGAAGGAAGGAAGGAAGT-3'

EWS/FLI1 DNA reverse complement (34 bases)-

5'-CACTTCCTTCCTTCCTTCCTTCCTTCCTTCCGGT-3'

Table 7: LSD1 and CoREST constructs used in my studies.

Human Protein	Mutation/deletion	Name
FL LSD1	LSD1(1-852)	Wild type
ΔN LSD1	LSD1(171-852) (Δ 1-170)	N terminal IDR deletion mutant
FL DD LSD1	LSD1 T110D, S111D (1-852)	Phosphomimic mutant
LSD1^{3KE}	LSD1 K355E, K357E, K359E (171-852) CoREST (286-440)	Nucleic acid binding site mutant
LSD1^{K661A}	LSD1 K661A (171-852) CoREST (286-440)	Catalytically inactive mutant
CoREST	CoREST-C (286-482)	CoREST C terminus

REFERENCES

(1) Appling, Dean, Anthony-Cahill Spencer, Mathews Christoper. Biochemistry Concepts and connections. : Pearson; 2016.

(2) Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. Nature (London) 2003 May 08,;423(6936):145-150.

(3) McGinty RK, Tan S. Nucleosome Structure and Function. Chemical reviews 2015 Mar 25,;115(6):2255-2273.

(4) Richmond TJ, Luger K, Mäder AW, Richmond RK, Sargent DF. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature (London) 1997 Sep 18,;389(6648):251-260.

(5) Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. Epigenetics & Chromatin 2014 -11-20;7(1).

(6) Mcghee JD, Felsenfe G. Nucleosome Structure. Annual Review of Biochemistry 1980(49):1115-56.

(7) Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. Journal of Molecular Biology 2002 -06;319(5):1097.

(8) Vogler C, Huber C, Waldmann T, Ettig R, Braun L, Izzo A, et al. Histone H2A C-Terminus Regulates Chromatin Dynamics, Remodeling, and Histone H1 Binding. PLoS Genet 2010 -12-09;6(12).

(9) Marmorstein R, Trievel RC. Histone modifying enzymes: Structures, mechanisms, and specificities. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 2009 -01;1789(1):58.

(10) Mcginty RK, Tan S. Recognition of the nucleosome by chromatin factors and enzymes. Current Opinion in Structural Biology 2016 -04;37:54.

(11) Makde RD, England JR, Yennawar HP, Tan S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. Nature 2010 -09-30;467(7315):562.

(12) Mcginty RK, Henrici RC, Tan S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature 2014 -10-30;514(7524):591.

(13) Bowman GD, Poirier MG. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. Chem Rev 2014 -11-26;115(6):2274.

(14) Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol 2007 -11;14(11):1025.

(15) Kouzarides T. Histone methylation in transcriptional control. 2002;12:198–209.

(16) Miller JL, Grant PA. The Role of DNA Methylation and Histone Modifications in Transcriptional Regulation in Humans. Subcellular Biochemistry 2012 -06-28.

(17) Voigt P, Reinberg D. Histone Tails: Ideal Motifs for Probing Epigenetics through Chemical Biology Approaches. ChemBioChem 2010 -11-24;12(2):236.

(18) Kouzarides T. Chromatin Modifications and Their Function. Cell 2007 -02;128(4):693.

(19) Alaskhar Alhamwe B, Khalaila R, Wolf J, Von Bülow V, Harb H, Alhamdan F, et al. Histone modifications and their role in epigenetics of atopy and allergic diseases. Allergy Asthma Clin Immunol 2018;14(1).

(20) Davydova E, Shimazu T, Schuhmacher MK, Jakobsson ME, Willemen HLDM, Liu T, et al. The methyltransferase METTL9 mediates pervasive 1-methylhistidine modification in mammalian proteomes. Nat Commun 2021 -02-09;12(1).

(21) Wilkins JF. Genomic imprinting. 1. Aufl. ed. New York, NY: Springer Science + Business Media [u.a.]; 2008.

(22) Yang Y, Bedford MT. Protein arginine methyltransferases and cancer. Nature reviews. Cancer 2013 Jan;13(1):37-50.

(23) Shi Y, Whetstine JR. Dynamic Regulation of Histone Lysine Methylation by Demethylases. Molecular Cell 2007;25(1):1-14.

(24) Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 2012 -04-03;13(5):343.

(25) Schluckebier G, O'Gara M, Saenger W, Cheng X. Universal Catalytic Domain Structure of AdoMet-dependent Methyltransferases. Journal of molecular biology 1995 Mar 17,;247(1):16-20.

(26) Stirpe A, Guidotti N, Northall SJ, Kilic S, Hainard A, Vadas O, et al. SUV39 SET domains mediate crosstalk of heterochromatic histone marks. 2021;0.

(27) Husmann D, Gozani O. Histone lysine methyltransferases in biology and disease. Nat Struct Mol Biol 2019 -10;26(10):880.

(28) Morera L, Lübbert M, Jung M. Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. Clinical Epigenetics 2016;8(57):57.

(29) Kim S, Chatterjee N, Jennings MJ, Bartholomew B, Tan S. Extranucleosomal DNA enhances the activity of the LSD1/CoREST histone demethylase complex. Nucleic acids research 2015 May 26,;43(10):4868-4880.

(30) Maiques-Diaz A, Somervaille TC. LSD1: biologic roles and therapeutic targeting. Epigenomics 2016 Aug 01,;8(8):1103-1116.

(31) Luo M. Chemical and Biochemical Perspectives of Protein Lysine Methylation. Chemical reviews 2018 Jul 25,;118(14):6656-6705.

(32) Rea S, Eisenhaber F, Nal O'carroll DÂ, Strahl² BD, Sun² Z, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. NATURE 2000;406.

(33) Shi Y, Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, et al. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. Cell 2004;119(7):941-953.

(34) Zhang Y, Martin C. The diverse functions of histone lysine methylation. Nature reviews. Molecular cell biology 2005 November;6(11):838-849.

(35) Peters AHFM, Metzger E, Yin N, Wissmann M, Schüle R, Müller JM, et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature 2005 Sep 15,;437(7057):436-439.

(36) Anand R, Marmorstein R. Structure and Mechanism of Lysine-specific Demethylase Enzymes. The Journal of biological chemistry 2007 Dec 07,;282(49):35425-35429.

(37) Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010 -04;464(7291):1071.

(38) Kooistra SM, Helin K. Molecular mechanisms and potential functions of histone demethylases. Nature reviews. Molecular cell biology 2012 Apr 4,;13(5):297-311.

(39) Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, et al. Human LSD2/KDM1b/AOF1 Regulates Gene Transcription by Modulating Intragenic H3K4me2 Methylation. Molecular Cell 2010 Jul 30,;39(2):222-233.

(40) Laurent B, Ruitu L, Murn J, Hempel K, Ferrao R, Xiang Y, et al. A Specific LSD1/KDM1A Isoform Regulates Neuronal Differentiation through H3K9 Demethylation. Molecular Cell 2015 Mar 19,;57(6):957-970.

(41) Laurent B, Shi Y. Expression, Purification, and Biochemical Analysis of the LSD1/KDM1A Histone Demethylase. Methods in enzymology 2016;573:241-259.

(42) Zibetti C, Adamo A, Binda C, Forneris F, Toffolo E, Verpelli C, et al. Alternative Splicing of the Histone Demethylase LSD1/KDM1 Contributes to the Modulation of Neurite Morphogenesis in the Mammalian Nervous System. The Journal of neuroscience 2010 Feb 17,;30(7):2521-2532.

(43) Fu X, Zhang P, Yu B. Advances toward LSD1 inhibitors for cancer therapy. Future Medicinal Chemistry 2017 Jul 1,;9(11):1227-1242.

(44) Song Y, Dagil L, Fairall L, Robertson N, Wu M, Ragan TJ, et al. Mechanism of Crosstalk between the LSD1 Demethylase and HDAC1 Deacetylase in the CoREST Complex. Cell reports (Cambridge) 2020 Feb 25,;30(8):2699-2711.e8.

(45) Shi Y, Matson C, Lan F, Iwase S, Baba T, Shi Y. Regulation of LSD1 Histone Demethylase Activity by Its Associated Factors. Molecular Cell 2005;19(6):857-864.

(46) Lee MG, Wynder C, Bochar DA, Hakimi M, Cooch N, Shiekhattar R. Functional Interplay between Histone Demethylase and Deacetylase Enzymes. Molecular and Cellular Biology 2006 Sep 01,;26(17):6395-6402.

(47) Yong Chen, Yuting Yang, Feng Wang, Ke Wan, Kenichi Yamane, Yi Zhang, et al. Crystal Structure of Human Histone Lysine-Specific Demethylase 1 (LSD1). Proceedings of the National Academy of Sciences of the United States of America 2006 Sep 19,;103(38):13956-13961.

(48) Yang M, Gocke CB, Luo X, Borek D, Tomchick DR, Machius M, et al. Structural Basis for CoREST-Dependent Demethylation of Nucleosomes by the Human LSD1 Histone Demethylase. Molecular Cell 2006;23(3):377-387.

(49) Hoelz A, Blobel G, Stavropoulos P. Crystal structure and mechanism of human lysine-specific demethylase-1. Nature Structural & Molecular Biology 2006 Jul;13(7):626-632.

(50) Yoneyama M, Tochio N, Umehara T, Koshiba S, Inoue M, Yabuki T, et al. Structural and Functional Differences of SWIRM Domain Subtypes. Journal of molecular biology 2007 May 25,;369(1):222-238.

(51) Federico Forneris, Claudia Binda, Antonio Adamo, Elena Battaglioli, Andrea Mattevi. Structural Basis of LSD1-CoREST Selectivity in Histone H3 Recognition. Journal of Biological Chemistry 2007 Jul 13,;282(28):20070-20074.

(52) Chen Y, Yang Y, Wang F, Wan K, Yamane K, Zhang Y, et al. Crystal Structure of Human Histone Lysine-Specific Demethylase 1 (LSD1). Proceedings of the National Academy of Sciences - PNAS 2006 Sep 19,;103(38):13956-13961.

(53) Szewczuk LM, Tomchick DR, Yang M, Machius M, Cole PA, Gocke CB, et al. Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. Nature Structural & Molecular Biology 2007 Jun;14(6):535-539.

(54) Federico Forneris, Claudia Binda, Maria Antonietta Vanoni, Elena Battaglioli, Andrea Mattevi. Human Histone Demethylase LSD1 Reads the Histone Code. Journal of Biological Chemistry 2005 Dec 16,;280(50):41360-41365.

(55) Burg JM, Gonzalez JJ, Maksimchuk KR, McCafferty DG. Lysine-Specific Demethylase 1A (KDM1A/LSD1): Product Recognition and Kinetic Analysis of Full-Length Histones. Biochemistry 2016 Mar 22,;55(11):1652-1662.

(56) Luka Z, Pakhomova S, Loukachevitch LV, Calcutt MW, Newcomer ME, Wagner C. Crystal structure of the histone lysine specific demethylase LSD1 complexed with tetrahydrofolate. Protein Science 2014 Jul;23(7):993-998.

(57) Luka Z, Moss F, Loukachevitch LV, Bornhop DJ, Wagner C. Histone Demethylase LSD1 Is a Folate-Binding Protein. Biochemistry 2011 May 31,;50(21):4750-4756.

(58) Shiekhattar R, Lee MG, Wynder C, Cooch N. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature 2005 Sep 15,;437(7057):432-435.

(59) Kim S, Zhu J, Yennawar N, Eek P, Tan S. Crystal Structure of the LSD1/CoREST Histone Demethylase Bound to Its Nucleosome Substrate. Molecular cell 2020 Jun 04,;78(5):903-914.e4.

(60) Hirschi A, Martin WJ, Luka Z, Loukachevitch LV, Reiter NJ. G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme. RNA (New York, N.Y.) 2016 Aug;22(8):1250-1260.

(61) Tamaoki J, Takeuchi M, Abe R, Kaneko H, Wada T, Hino S, et al. Splicing- and demethylase-independent functions of LSD1 in zebrafish primitive hematopoiesis. Scientific reports 2020 May 22,;10(1):8521.

(62) D'oto A, Tian Q, Davidoff AM, Yang J. Histone demethylases and their roles in cancer epigenetics.

(63) Song Y, Zhang H, Yang X, Shi Y, Yu B. Annual review of lysine-specific demethylase 1 (LSD1/KDM1A) inhibitors in 2021. European journal of medicinal chemistry 2022 Jan 15,;228:114042.

(64) Forneris F, Battaglioli E, Mattevi A, Binda C. New roles of flavoproteins in molecular cell biology: Histone demethylase LSD1 and chromatin. The FEBS journal 2009 Aug;276(16):4304-4312.

(65) Peng B, Wang J, Hu Y, Zhao H, Hou W, Zhao H, et al. Modulation of LSD1 phosphorylation by CK2/WIP1 regulates RNF168-dependent 53BP1 recruitment in response to DNA damage. Nucleic Acids Research 2015 Jul 13,;43(12):5936-5947.

(66) Wright PE, Dyson HJ. Intrinsically unstructured proteins and their functions. Nature reviews. Molecular cell biology 2005 Mar;6(3):197-208.

(67) Perillo B, Tramontano A, Pezone A, Migliaccio A. LSD1: more than demethylation of histone lysine residues. Experimental & molecular medicine 2020 Dec 14,;52(12):1936-1947.

(68) Metzger E, Willmann D, Mcmillan J, Forne I, Metzger P, Gerhardt S, et al. Assembly of methylated KDM1A and CHD1 drives androgen receptor–dependent transcription and translocation. Nat Struct Mol Biol 2016 -01-11;23(2):132.

(69) Pillai-Kastoori L, Schutz-Geschwender AR, Harford JA. A systematic approach to quantitative Western blot analysis. Analytical biochemistry 2020 Mar 15,;593:113608.

(70) Shiau C, Trnka M, Bozicevic A, Torres I, Al-Sady B, Burlingame A, et al. Reconstitution of Nucleosome Demethylation and Catalytic Properties of a Jumonji Histone Demethylase. Chemistry & biology 2013 Apr 18,;20(4):494-499.

(71) Dhall A, Shelton PMM, Delachat NF, Leonen CJA, Fierz B, Chatterjee C. Nucleosome Binding by the Lysine Specific Demethylase 1 (LSD1) Enzyme Enables Histone H3 Demethylation. Biochemistry 2020 -06-26;59(27):2479.

(72) Munari F, Soeroes S, Zenn HM, Schomburg A, Kost N, Schröder S, et al. Methylation of Lysine 9 in Histone H3 Directs Alternative Modes of Highly Dynamic Interaction of

Heterochromatin Protein hHP1β with the Nucleosome. The Journal of biological chemistry 2012 Sep 28,;287(40):33756-33765.

(73) Na J, Lee W, Yu YG. How Do We Study the Dynamic Structure of Unstructured Proteins: A Case Study on Nopp140 as an Example of a Large, Intrinsically Disordered Protein. International Journal of Molecular Sciences 2018 Jan 27,;19(2):381.

(74) Cha S. Tight-binding inhibitors—I: Kinetic behavior. Biochemical pharmacology 1975;24(23):2177-2185.

(75) Cheng Y, Prusoff WH. RELATIONSHIP BETWEEN THE INHIBITION CONSTANT (&) AND THE CONCENTRATION OF INHIBITOR WHICH CAUSES 50 PER CENT INHIBITION (Iso) OF AN ENZYMATIC REACTION*. Biochemical Pharmacology 1973;22:3099.

(76) Nam H, Boo K, Kim D, Han D, Choe H, Kim C, et al. Phosphorylation of LSD1 by PKCα Is Crucial for Circadian Rhythmicity and Phase Resetting. MOLECULAR CELL 2014 Mar 06,;53(5):791-805.

(77) Tortorici M, Borrello MT, Tardugno M, Chiarelli LR, Pilotto S, Ciossani G, et al. Protein Recognition by Short Peptide Reversible Inhibitors of the Chromatin-Modifying LSD1/CoREST Lysine Demethylase. ACS chemical biology 2013 Aug 16,;8(8):1677-1682.

(78) Baron R, Binda C, Tortorici M, McCammon JA, Mattevi A. Molecular Mimicry and Ligand Recognition in Binding and Catalysis by the Histone Demethylase LSD1-CoREST Complex. Structure (London) 2011;19(2):212-220.

(79) Hayward D, Cole PA. LSD1 histone demethylase assays and inhibition. Methods in enzymology 2016;573:261-278.

(80) McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of Applied Crystallography 2007 Aug;40(4):658-674.

(81) Forneris F, Binda C, Dall'Aglio A, Fraaije MW, Battaglioli E, Mattevi A, et al. A highly specific mechanism of histone H3-K4 recognition by histone demethylase LSD1. The Journal of Biological Chemistry 2006;281(46):35289-35295.

(82) Culhane JC, Szewczuk LM, Liu X, Da G, Marmorstein R, Cole PA. A Mechanism-Based Inactivator for Histone Demethylase LSD1. Journal of the American Chemical Society 2006 Apr 12,;128(14):4536-4537.

(83) Wang K, Chang H. Molecular Mechanisms of Long Noncoding RNAs. Molecular Cell 2011 Sep 16,;43(6):904-914.

(84) Khalil AM, Rinn JL. RNA–protein interactions in human health and disease. Seminars in cell & developmental biology 2011 Jun 01,;22(4):359-365.

(85) Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Research 2015 Oct 15,;43(18):8627-8637.

(86) Biffi G, Di Antonio M, Tannahill D, Balasubramanian S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nature chemistry 2014 Jan;6(1):75-80.

(87) Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. Telomeric Repeat-Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. Science 2007;318(5851):798-801.

(88) Martadinata H, Phan AT. Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric RNA Sequences in K+ Solution. Journal of the American Chemical Society 2009 Feb 25,;131(7):2570-2578.

(89) Xu Y, Suzuki Y, Ito K, Komiyama M. Telomeric repeat-containing RNA structure in living cells. Proceedings of the National Academy of Sciences - PNAS 2010 Aug 17,;107(33):14579-14584.

(90) Fay MM, Lyons SM, Ivanov P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. Journal of molecular biology 2017 Jul 07,;429(14):2127-2147.

(91) Liu X, Xu Y. HnRNPA1 Specifically Recognizes the Base of Nucleotide at the Loop of RNA G-Quadruplex. Molecules (Basel, Switzerland) 2018 Jan 22,;23(1):237.

(92) Xu Y, Kaminaga K, Komiyama M. G-Quadruplex Formation by Human Telomeric Repeats-Containing RNA in Na+ Solution. Journal of the American Chemical Society 2008 Aug 20,;130(33):11179-11184.

(93) Collie GW, Parkinson GN, Neidle S, Rosu F, De Pauw E, Gabelica V. Electrospray Mass Spectrometry of Telomeric RNA (TERRA) Reveals the Formation of Stable Multimeric G-Quadruplex Structures. Journal of the American Chemical Society 2010 Jul 14,;132(27):9328-9334.

(94) Wang C, Zhao L, Lu S. Role of TERRA in the Regulation of Telomere Length. International Journal of Biological Sciences 2015;11(3):316-323.

(95) Redon S, Reichenbach P, Lingner J. non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. Nucleic acids research 2010;38(17):5797-5806.

(96) Liu X, Ishizuka T, Bao H, Wada K, Takeda Y, Iida K, et al. Structure-Dependent Binding of hnRNPA1 to Telomere RNA. J Am Chem Soc 2017 -06-07;139(22):7533.

(97) Wang X, Goodrich KJ, Gooding AR, Naeem H, Archer S, Paucek RD, et al. Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. Molecular cell 2017 Mar 16,;65(6):1056-1067.e5.

(98) Montero JJ, López-Silanes I, Megías D, F. Fraga M, Castells-García Á, Blasco MA. TERRA recruitment of polycomb to telomeres is essential for histone trymethylation marks at telomeric heterochromatin. Nat Commun 2018 -04-18;9(1).

(99) Porro A, Feuerhahn S, Lingner J. TERRA-Reinforced Association of LSD1 with MRE11 Promotes Processing of Uncapped Telomeres. Cell reports (Cambridge) 2014 Feb 27,;6(4):765-776.

(100) Porro A, Feuerhahn S, Delafontaine J, Riethman H, Rougemont J, Lingner J. Functional characterization of the TERRA transcriptome at damaged telomeres. Nature communications 2014 Oct 31,;5(1):5379.

(101) He C, Sidoli S, Warneford-Thomson R, Tatomer D, Wilusz J, Garcia B, et al. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. Molecular cell 2016 Oct 20,;64(2):416-430.

(102) Martadinata H, Phan AT. Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K+ Solution. Biochemistry (Easton) 2013 Apr 02,;52(13):2176-2183.

(103) Balaratnam S, Basu S. Divalent cation-aided identification of physico-chemical properties of metal ions that stabilize RNA g-quadruplexes. Biopolymers 2015 Jul;103(7):376-386.

(104) Chauca-Diaz AM, Jung Choi Y, Resendiz MJE. Biophysical properties and thermal stability of oligonucleotides of RNA containing 7,8-dihydro-8-hydroxyadenosine. Biopolymers 2015 Mar;103(3):167-174.

(105) Zhang J, Yadav T, Ouyang J, Lan L, Zou L. Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. Cell reports 2019 Jan 22,;26(4):955-968.e3.

(106) Pinter S, Knodel F, Choudalakis M, Schnee P, Kroll C, Fuchs M, et al. A functional LSD1 coregulator screen reveals a novel transcriptional regulatory cascade connecting R-loop homeostasis with epigenetic regulation. Nucleic acids research 2021 May 07,;49(8):4350-4370.

(107) Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen L, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol 2023 -01-03.

(108) Hegazy YA, Fernando CM, Tran EJ. The balancing act of R-loop biology: The good, the bad, and the ugly. The Journal of biological chemistry 2020 Jan 24,;295(4):905-913.

(109) Belotserkovskii BP, Tornaletti S, D'Souza AD, Hanawalt PC. R-loop generation during transcription: Formation, processing and cellular outcomes. DNA repair 2018 Nov;71:69-81.

(110) Thakur J, Henikoff S. Architectural RNA in chromatin organization. Biochemical Society Transactions 2020 -09-08;48(5):1967.

(111) Davidovich C, Cech TR. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA (Cambridge) 2015 Dec;21(12):2007-2022.

(112) Tsai M, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. Science (American Association for the Advancement of Science) 2010 Aug 06,;329(5992):689-693.

(113) Feretzaki M, Pospisilova M, Valador Fernandes R, Lunardi T, Krejci L, Lingner J. RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops. Nature 2020 -11-12;587(7833):303.

(114) Graf M, Bonetti D, Lockhart A, Serhal K, Kellner V, Maicher A, et al. Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. Cell 2017 Jun 29,;170(1):72-85.e14.

(115) Tian H, Gao Z, Li H, Zhang B, Wang G, Zhang Q, et al. DNA damage response – A double-edged sword in cancer prevention and cancer therapy. Cancer letters 2015 Mar 01,;358(1):8-16.

(116) Ouyang J, Yadav T, Zhang J, Yang H, Rheinbay E, Guo H, et al. RNA transcripts stimulate homologous recombination by forming DR-loops. Nature 2021 -06;594(7862):283.

(117) Reiter NJ, Osterman AK, Mondragón A. The bacterial ribonuclease P holoenzyme requires specific, conserved residues for efficient catalysis and substrate positioning. Nucleic Acids Research 2012 Nov 01,;40(20):10384-10393.

(118) Mei Y, Deng Z, Vladimirova O, Gulve N, Johnson FB, Drosopoulos WC, et al. TERRA G-quadruplex RNA interaction with TRF2 GAR domain is required for telomere integrity. Scientific reports 2021 Feb 10,;11(1):3509.

(119) Theisen ER, Pishas KI, Saund RS, Lessnick SL. Therapeutic opportunities in Ewing sarcoma: EWS-FLI inhibition via LSD1 targeting. Oncotarget 2016 -02-02;7(14):17616.

(120) Serce N, Gnatzy A, Steiner S, Lorenzen H, Kirfel J, Buettner R. Elevated expression of LSD1 (Lysine-specific demethylase 1) during tumour progression from pre-invasive to invasive ductal carcinoma of the breast. BMC Clinical Pathology 2012 Aug 24,;12(1):13.

(121) Kurmasheva RT, Erickson SW, Han R, Teicher BA, Smith MA, Roth M, et al. In vivo evaluation of the lysine-specific demethylase (KDM1A/LSD1) inhibitor SP-2577 (Seclidemstat) against pediatric sarcoma preclinical models: A report from the Pediatric Preclinical Testing Consortium (PPTC). Pediatric blood & cancer 2021 Nov;68(11):e29304-n/a.

(122) SANKAR S, THEISEN ER, BEARSS J, MULVIHILL T, HOFFMAN LM, SORNA V, et al. Reversible LSD1 Inhibition Interferes with Global EWS/ETS Transcriptional Activity and Impedes Ewing Sarcoma Tumor Growth. Clinical cancer research 2014;20(17):4584-4597.

(123) Romo-morales A, Aladowicz E, Blagg J, Gatz SA, Shipley JM. Catalytic inhibition of KDM1A in Ewing sarcoma is insufficient as a therapeutic strategy. Pediatr Blood Cancer 2019 -06-17;66(9).

(124) Sankar S, Bell R, Stephens B, Zhuo R, Sharma S, Bearss DJ, et al. Mechanism and relevance of EWS/FLI-mediated transcriptional repression in Ewing sarcoma. Oncogene 2013 Oct 17,;32(42):5089-5100.

(125) Owen LA, Kowalewski AA, Lessnick SL. EWS/FLI Mediates Transcriptional Repression via NKX2.2 during Oncogenic Transformation in Ewing's Sarcoma. PLoS ONE 2008 Apr 16,;3(4):e1965.

(126) Theisen ER, Selich-Anderson J, Miller KR, Tanner JM, Taslim C, Pishas KI, et al. Chromatin profiling reveals relocalization of lysine-specific demethylase 1 by an oncogenic fusion protein. Epigenetics 2020 -08-25;16(4):405.

(127) Yang G, Liu Y, Ding L, Tao F, Zhu M, Shi Z, et al. A state-of-the-art review on LSD1 and its inhibitors in breast cancer: Molecular mechanisms and therapeutic significance. Frontiers in pharmacology 2022 Sep 16,;13:989575.

(128) Lee MG, Wynder C, Schmidt DM, McCafferty DG, Shiekhattar R. Histone H3 Lysine 4 Demethylation Is a Target of Nonselective Antidepressive Medications. Chemistry & biology 2006;13(6):563-567.

(129) Schmidt DMZ, McCafferty DG. trans-2-Phenylcyclopropylamine Is a Mechanism-Based Inactivator of the Histone Demethylase LSD1. Biochemistry (Easton) 2007 Apr 10,;46(14):4408-4416.

(130) Dai X, Liu Y, Xiong X, Xue L, Zheng Y, Liu H. Tranylcypromine Based Lysine-Specific Demethylase 1 Inhibitor: Summary and Perspective. Journal of medicinal chemistry 2020 Dec 10,;63(23):14197-14215.

(131) Binda C, Valente S, Romanenghi M, Pilotto S, Cirilli R, Karytinos A, et al. Biochemical, Structural, and Biological Evaluation of Tranylcypromine Derivatives as Inhibitors of Histone Demethylases LSD1 and LSD2. Journal of the American Chemical Society 2010 May 19,;132(19):6827-6833.

(132) Mimasu S, Sengoku T, Fukuzawa S, Umehara T, Yokoyama S. Crystal structure of histone demethylase LSD1 and tranylcypromine at 2.25 A. Biochemical and biophysical research communications 2008 Feb 01,;366(1):15-22.

(133) Bennani-Baiti IM, PhD, Machado I, MD, Llombart-Bosch A, MD, Kovar H, PhD. Lysine-specific demethylase 1 (LSD1/KDM1A/AOF2/BHC110) is expressed and is an epigenetic drug target in chondrosarcoma, Ewing's sarcoma, osteosarcoma, and rhabdomyosarcoma. Human pathology 2012;43(8):1300-1307.

(134) Salamero O, Montesinos ;P, Willekens C, Antonio Pérez-Simón J, Pigneux A, Récher C, et al. First-in-Human Phase I Study of Iadademstat (ORY-1001): A First-in-Class Lysine-Specific Histone Demethylase 1A Inhibitor, in Relapsed or Refractory Acute Myeloid Leukemia. J Clin Oncol ;38.

(135) Maes T, Mascaró C, Tirapu I, Estiarte A, Ciceri F, Lunardi S, et al. ORY-1001, a Potent and Selective Covalent KDM1A Inhibitor, for the Treatment of Acute Leukemia. Cancer cell 2018 Mar 12,;33(3):495-511.e12.

(136) Foster CT, Dovey OM, Lezina L, Luo JL, Gant TW, Barlev N, et al. Lysine-Specific Demethylase 1 Regulates the Embryonic Transcriptome and CoREST Stability. Molecular and Cellular Biology 2010 Oct 15,;30(20):4851-4863.

(137) Kanouni T, Severin C, Cho RW, Yuen NY-, Xu J, Shi L, et al. Discovery of CC-90011: A Potent and Selective Reversible Inhibitor of Lysine Specific Demethylase 1 (LSD1). Journal of medicinal chemistry 2020 Dec 10,;63(23):14522-14529.

(138) Sacilotto N, Dessanti P, Lufino MMP, Ortega A, Rodríguez-Gimeno A, Salas J, et al. Comprehensive in Vitro Characterization of the LSD1 Small Molecule Inhibitor Class in Oncology. ACS pharmacology & translational science 2021 Dec 10,;4(6):1818-1834.

(139) Sorna V, Theisen ER, Stephens B, Warner SL, Bearss DJ, Vankayalapati H, et al. High-Throughput Virtual Screening Identifies Novel N'-(1-Phenylethylidene)-benzohydrazides as Potent, Specific, and Reversible LSD1 Inhibitors. Journal of medicinal chemistry 2013 Dec 12,;56(23):9496-9508.

(140) Soldi R, Ghosh Halder T, Weston A, Thode T, Drenner K, Lewis R, et al. The novel reversible LSD1 inhibitor SP-2577 promotes anti-tumor immunity in SWItch/Sucrose-NonFermentable (SWI/SNF) complex mutated ovarian cancer. PloS one 2020;15(7):e0235705.

(141) SP-2577 (Seclidemstat) for the Treatment of Relapsed or Refractory Ewing Sarcoma Pediatric Subcommittee of the Oncologic Drugs Advisory Committee. 2020 -06-17.

(142) Schmitt, M. L., Ladwein, K. I., Carlino, L., Schulz-Fincke, J., Willmann, D., Metzger, E., Schilcher, P., Imhof, A., Schüle, R., Sippl, W., and Jung, M. Heterogeneous antibody-based activity assay for lysine specific demethylase 1 (LSD1) on a histone peptide substrate. SLAS Discovery 2014; 973–978.

(143) Balaraman K, Deniz E, Nelson E, Pilicer SL, Atasoy S, Molotkova A, et al. Design, synthesis and biological evaluation of Nucleosidic CD99 inhibitors that selectively reduce Ewing sarcoma viability. Eur J Med Chem 2023; 251:115-244.

(144) Üren A, Tcherkasskaya O, Toretsky JA. Recombinant EWS-FLI1 Oncoprotein Activates Transcription. Biochemistry (Easton) 2004 Oct 26,;43(42):13579-13589.

(145) Wang Z, Tiwari PB, Üren A, Brelidze TI. Identification of undecylenic acid as EAG channel inhibitor using surface plasmon resonance-based screen of KCNH channels. BMC Pharmacology & Toxicology 2019 Jul 17,;20(1):42.

(146) Manara MC, Pasello M, Scotlandi K. CD99: A Cell Surface Protein with an Oncojanus Role in Tumors. Genes 2018 Mar 13,;9(3):159.

(147) Toretsky JA, Erkizan V, Levenson A, Abaan OD, Parvin JD, Cripe TP, et al. Oncoprotein EWS-FLI1 Activity Is Enhanced by RNA Helicase A. Cancer Research 2006 Jun 01,;66(11):5574-5581.

(148) Toretsky JA, Erkizan HV, Kong Y, Merchant M, Schlottmann S, Barber-Rotenberg JS, et al. A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. Nature medicine 2009 Jul;15(7):750-756.

(149) Lin J, Lai E. Protein–Protein Interactions: Co-Immunoprecipitation. Methods in Molecular Biology 2017;1615:211.

(150) Uchiumi F, Miyazaki S, Tanuma S. The possible functions of duplicated ets (GGAA) motifs located near transcription start sites of various human genes. Cell Mol Life Sci 2011 Jun 01,;68(12):2039-2051.

(151) Piersimoni L, Kastritis PL, Arlt C, Sinz A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions–A Method for All Seasons. Chem Rev 2021 -11-19;122(8):7500.

(152) Yu C, Huang L. Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. Analytical chemistry (Washington) 2018 Jan 02,;90(1):144-165.

(153) Fischer G, Rossmann M, Hyvönen M. Alternative modulation of protein–protein interactions by small molecules. Current Opinion in Biotechnology 2015;35:78-85.

(154) Welch D, Kahen E, Fridley B, Brohl AS, Cubitt CL, Reed DR. Small molecule inhibition of lysine-specific demethylase 1 (LSD1) and histone deacetylase (HDAC) alone and in combination in Ewing sarcoma cell lines. PLoS ONE 2019 Sep 24,;14(9):e0222228.

(155) Jaboin J, Wild J, Hamidi H, Khanna C, Kim CJ, Robey R, et al. Cell and Molecular Biology Section, Pediatric Oncology Branch. CANCER RESEARCH 2002 -11-01;62:6108.

(156) Hassell AM, An G, Bledsoe RK, Bynum JM, Carter III HL, Deng SJ, et al. Crystallization of protein-ligand complexes. Acta Crystallographica Section D: Biological Crystallography 2007 Jan;63(1):72-79.

(157) Otwinowski Z, Minor W. [20] Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology 1997:307.

(158) Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP 4 suite and current developments. Acta crystallographica. Section D, Biological crystallography. 2011 Apr 01,;67(4):235-242.

(159) Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM, Cheng ;X, et al. [32] Collaborative Computational Project, Number 4: Providing Programs for Protein Crystallography. Nature (London) 1993;362.

(160) Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta crystallographica. Section D, Biological crystallography. 2004 Dec;60(12-1):2126-2132.

(161) Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D: Biological Crystallography 2011 Apr 01,;67(4):355-367.