RNA Recognition by the Histone Demethylase LSD1 and Proteinaceous RNase P: Characterization of the Binding of Highly Structured RNAs by Enzyme Complexes

By

William Martin

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

February 28, 2018

Nashville, Tennessee

Approved:

Neil Osheroff, Ph.D.

Nicholas Reiter, Ph.D.

Martin Egli, Ph.D.

David Cortez, Ph.D.

Gregor Neuert, Ph.D.

ACKNOWLEDGEMENTS

I must first acknowledge my mentor Nick Reiter; he has encouraged me at every turn, especially when science was disheartening, and provided constant support while allowing me the flexibility to direct my own research while encouraging the pursuit of a full life outside of research as well. I must also thank the other members of our lab especially Alex Hirschi, who began the LSD1 story and taught me many lessons in the lab. I would also like to acknowledge the members of my thesis committee for providing regular feedback and advice.

Vanderbilt has a reputation as an exceptionally collaborative environment for scientists. That reputation is well-deserved; I cannot mention all of the people who selflessly took the time to learn about my research and provide advice, training, materials, the use of equipment, or simply have discussions about my research. However, I would especially like to thank Manny Ascano and his lab for their critical assistance in setting up and performing PAR-CLIP; the Guengerich and York lab for generously allowing the use of their lab space and equipment; Hayes McDonald for his contributions to the mass spectrometry project; the Vickers lab for their assistance with cDNA preparation; Quanhu "Tiger" Sheng for his assistance with bioinformatics; and the Vanderbilt Center for Structural Biology for providing state-of-the art resources and training on those instruments.

Numerous organizations made the following work financially possible. They are the National Institute of General Medical Sciences of the NIH through award number R01GM120572 and the Vanderbilt Molecular Biophysics Training Program T32GM008320; the American Heart Association through grant number GRNT20380334; and the Vanderbilt Biochemistry Department.

Finally, I am grateful to my parents, Jay and Lezlie, and my wife Rebecca; without their constant support and patience this body of work would not have been possible.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

4SU	4-thiouridine
AOD	amine oxidase domain
ATM	ataxia telangiectasia mutated
ATR	ataxia telangiectasia and Rad3-related
CD	circular dichroism
ChIP	chromatin immunoprecipitation
CK2	casein kinase 2
COREST	corepressor of RE1 silencing transcription factor
CRB	CREB-binding protein
CtBP1	C-terminal binding protein 1
CytB	cytochrome B
DMS	dimethyl sulfate
DNMT1	DNA methyltransferase 1
dsDNA	double-stranded DNA
E2F1	E2F transcription factor 1E2F1
EMSA	electromobility shift assay
EPHB1	EPH receptor B1
FAD	flavin adenine dinucleotide
FAM57B	family with sequence similarity 57 member B
FMRP	fragile X mental retardation protein
fRIP	formaldehyde RNA immunoprecipitation
GQ	G-quadruplex
HDAC	histone deacetylase
HOTAIR	HOX transcript antisense intergenic RNA
HOX	homeobox
HP1	heterochromatin protein
HR	homologous recombination
lncRNA	long non-coding RNA
LSD1	lysine-specific demethylase-1
LSD2	lysine-specific demethylase-2
MALAT1	metastasis associated lung adenocarcinoma transcript 1
MRE11	MRE11 homolog, double strand break repair nuclease
MRN	Mre11/Rad50/Nbs1
MRPP1	mitochondrial RNase P protein 1; also, tRNA methyltransferase 10C (TRMT10C)
MRPP2	mitochondrial RNase P protein 2; also, hydroxysteroid 17-beta dehydrogenase 10 (HSD17B10)
MYO1B	myosin 1B
MYPT1	myosin phosphatase target subunit 1
NAI	2-methylnicotinic acid imidazolide
Nbs1	nibrin
ncRNA	non-coding RNA
ND5	NADH dehydrogenase subunit 5
NEAT1	nuclear paraspeckle assembly transcript 1

NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NHEJ	non-homologous end joining
PAR-CLIP	photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation
PDB	protein data bank
pre-tRNA	precursor tRNA
PPR	pentatricopeptide repeat
PRC2	polycomb remodeling complex 2
PRORP	proteinaceous RNase P protein
Rad50	RAD50 double strand break repair protein
RBD	RNA binding domain
RGG motif	arginine-glycine rich motif
RNase MRP	ribonuclease mitochondrial RNA processing
RNase P	ribonuclease P
RNF168	ring finger protein 168
RNP	ribonucleoprotein
RREB1	Ras Responsive Element Binding Protein 1
SANT	Swi3, Ada2, N-Cor, and TFIIIB domain
SNP	single-nucleotide polymorphism
ssRNA	single-stranded RNA
SV-AUC	sedimentation velocity analytical ultracentrifugation
SWIRM	Swi3, Rsc8, and Moira
TERRA	telomeric repeat-containing RNA
TPR	tetratricopeptide repeat
TSS	transcriptional start site
UTR	untranslated region
Xist	X inactive specific transcript
XL-MS	cross-linking and mass spectroscopy

ABSTRACT

Noncoding RNAs are increasingly recognized as central, structured components in the clockwork of life which interact with and regulate proteins. However, our mechanistic understanding of these interactions is currently limited. This study investigates the recognition of structured RNAs by two essential enzymes, proteinaceous RNase P (PRORP) and lysine-specific demethylase-1 (LSD1).

PRORP binds and cleaves human mitochondrial precursor tRNAs in a fundamental step for the generation of mature mitochondrial transcripts. The enzyme recognizes specific structured domains of the pre-tRNAs in a manner similar to but distinct from the unrelated nuclear RNase P complex. PRORP likely binds pre-tRNA substrate via the pentatricopeptide repeat (PPR) domain, an element that is found in a number of mitochondrial proteins. Structures of PPR domain proteins in complex with ssRNAs have previously been determined but PRORP appears to recognize RNAs through a discrete, shape-dependent manner. In order to move towards a crystal structure of the PRORP-RNA complex, minimal constructs were generated and biochemically validated with binding studies and *in vitro* activity assays and used in crystallization studies.

LSD1 modulates gene expression through enzymatic histone demethylation and also serves as a protein scaffold in various large protein complexes. Long noncoding RNAs (lncRNAs) regulate chromatin modifiers such as LSD1 through a number of mechanisms. The lncRNA telomeric repeat-containing RNA (TERRA) has previously been demonstrated to recruit LSD1 to deprotected telomeres where it promotes the recruitment of the nuclease MRE11. Here, it is shown that LSD1 specifically recognizes the G-quadruplex structure formed by TERRA and other RNAs and a G-quadruplex RNA binding region is identified in the regulatory SWIRM domain of LSD1. Together, these studies advance our understanding of the role of structured RNAs in RNA/protein interactions.

Note: Portions of this dissertation are reprinted with permission from Structural Roles of Noncoding RNAs in the Heart of Enzymatic Complexes (Martin WJ, Reiter NJ. *ACS Biochemistry*, 2017) and G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme (Hirschi A,* Martin WJ*, Luka Z, Loukachevitch LV, Reiter NJ. *RNA*, 2016;22(8):1250-60) *Co-1st authors.

CHAPTER 1

Introduction

1.1: Role of Structure in RNA Biology

Over billions of years of evolution, nature has embraced proteins as the major workhorse molecules of the cell. However, nearly every aspect of metabolism is dependent upon how structured RNAs interact with proteins, ligands, and other nucleic acids. Key processes, including telomere maintenance, RNA processing, and protein synthesis, require large RNAs that assemble into elaborate three-dimensional shapes. These RNAs can (i) act as flexible scaffolds for protein subunits, (ii) participate directly in substrate recognition, and (iii) serve as catalytic components.

Noncoding (nc) RNAs were instrumental in the evolution of the genetic code and likely served as key progenitor molecules for all life forms.[1–3] Examining the compositions of a few of these ancient RNAs, from the analyses of sequence conservation to secondary and tertiary structure determination, has yielded tremendous insight into our understanding of RNA biology. Detailed, biophysical studies of these RNAs in isolation or as part of a ribonucleoprotein complex have revealed a high degree of structural diversity, illuminating how large RNAs assemble into defined tertiary architectures and interact with their protein partners at the atomic level.[4] These structural studies serve as paradigms for improving our understanding of the versatile roles of RNAs and

the emerging functional roles of newly discovered ncRNAs in the modern world.

RNA molecules possess several evolutionary advantages that allow them to serve as a central regulatory molecule in the cell. Unlike proteins, RNAs can utilize both shape-based recognition and intermolecular base pair interactions with RNA and DNA to provide specificity.[5] In an almost symbiotic relationship, RNA can also act as a scaffold that bring proteins together and direct them to a nucleotide substrate while the proteins in turn stabilize RNA structures. (Figure 1) This added stabilization by conserved protein–RNA interactions plays a central role throughout the ribonucleoprotein (RNP) assembly process and fortifies key RNA conformations during the reaction cycle.

This work focuses on the role of RNA tertiary structure in two separate systems: the processing of pre-tRNAs by RNase P and the recruitment of the epigenetic regulator lysine-specific demethylase-1 (LSD1) to chromatin. Both processes depend upon the recognition of structured RNAs by proteins and represent novel and emerging biological mechanisms. The following sections of this chapter present the biological context of the RNase P and LSD1 enzymes and highlight the roles played by RNA structure.

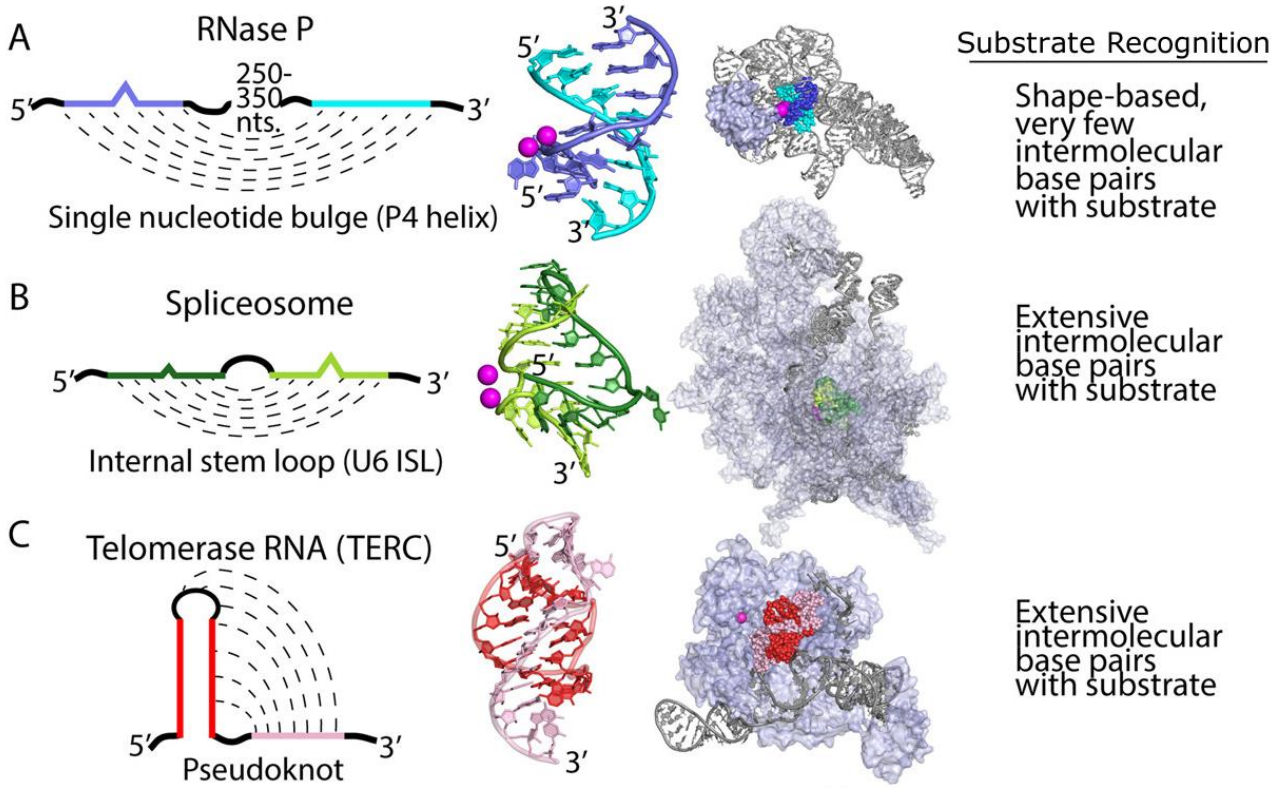


Figure 1: Structured RNA scaffolds are at the Core of RNP Complexes.

Illustration of conserved RNA structures in essential RNP enzymes. Dashed lines represent base pairs. (A) Two conserved domains (blue and cyan) of the ribonuclease P (RNase P) enzyme are separated by 250–350 nucleotides but assemble to form the central scaffold, termed the P4 helix. The crystal structure of a bacterial RNase P holoenzyme with tRNA (PDB entry 3Q1Q) is shown, with the P4 helix (blue and cyan), active site metal ion location (magenta spheres), surrounding the P RNA structure (gray), and P protein (blue-white surface).[6] (B) An asymmetric internal stem–loop structure (green and lime) from the U6 small nuclear RNA (termed the U6 ISL) helps to form the activated spliceosome core. The cryo-EM-derived minimal structure is shown, with U6 ISL (green/lime), active site metal ions (magenta), surrounding snRNAs (gray), and protein components (blue-white surface) (PDB entry 5LJ3).[7] (C) The telomerase RNA component (TR) utilizes a conserved pseudoknot, where a stem–loop structure (red) intercalates with a single-stranded region (pink). This structural element is central to the assembly of an activated telomerase holoenzyme. A pseudoatomic experimental model of the catalytic core of the Tetrahymena telomerase illustrates the location of the pseudoknot (red/pink), the approximate active site metal (magenta), surrounding TR (gray), and core protein components (blue-white surface).[8] Figure adapted from Martin and Reiter, 2017.[9]

1.2: RNase P: An Ancient Ribozyme and Modern Protein Enzyme

Structured RNAs in RNA Processing

RNase P is an RNP complex composed of an essential RNA ribozyme subunit and one or more protein subunits required for its enzymatic function *in vivo*. Aside from the ribosome, RNase P constitutes the only known example of a multiple-turnover RNA enzyme that is required for cell viability in all domains of life.[10,11] Newly transcribed precursor tRNA (pre-tRNA) contains excess nucleotides at its 5′ and 3′ ends. The 5′ leader region is removed by RNase P, whereas the 3′ tail extension can be trimmed or cleaved by various protein enzymes.[11] Because of its vital role in cell metabolism, the bacterial RNase P RNP complex is a prime target for novel antibiotics.[12]

Numerous structural and biochemical studies have demonstrated that the large RNA component of RNase P can display versatility in terms of substrate recognition and that its functional complexity is comparable with that of multidomain protein enzymes.[11,13] The RNase P RNP complex not only catalyzes the maturation of the 5′ end of tRNA but also can exhibit broad specificity and act on many different RNAs, including viral and phage RNA, mRNAs, noncoding RNAs, rRNA, and riboswitches.[13] RNase P targets substrates that mimic portions of the distinctive tRNA secondary structure, termed t-elements.[11] Eukaryotic RNase P acts on t-element substrates which fold into structures containing analogs of the tRNA leader stem and T-domain stem-loop.[14] For example, mammalian long ncRNAs that contain tRNA-like elements, such as MALAT1 and NEAT1, undergo processing and become activated via an

RNase P-mediated mechanism.[15] Cleavage at specific RNase P-defined sites generates a triple helix at the 3' end, creating a polyadenylation-independent mechanism for stabilizing the 3' end of RNAs. This functionality can be artificially exploited to direct RNase P to cleave viral or disease-associated transcripts.[16] External guide sequence RNAs or antisense oligonucleotides can be designed to hybridize with transcripts of interest so as to fold into an intermolecular tRNA-like structure and serve as an RNase P substrate; this approach can be more specific than RNase H approaches, though it has been overshadowed by CRISPR advances in recent years.[17,18] The following study investigates the substrate specificity of an alternative, recently characterized human RNase P enzyme to determine how it recognizes structural motifs of pre-tRNAs and how it compares with the classical RNase P.

Proteinaceous RNase P

In 1988, a 5' tRNA-processing agent was observed in spinach chloroplast that was insensitive to nuclease treatment and exhibited density consistent with a protein enzyme, indicating that some organelles process pre-tRNAs through a ribozyme-free mechanism.[19] Distinct substrate specificity between the human nuclear and mitochondrial enzymes provided an additional indication of an alternative RNase P enzyme.[20] However, similar magnesium dependence in the mitochondrial and nuclear enzymes and the potential for small amounts of nuclear RNase P RNA contamination in mitochondrial extracts raised skepticism

of the existence of a proteinaceous RNase P until the human mitochondrial RNase P complex was conclusively identified and biochemically characterized in 2008.[21,22,23] The complex consists of three subunits, termed mitochondrial RNase P proteins 1, 2, and 3 (MRPP1, MRPP2, and MRPP3). MRPP1 and MRPP2 also have dehydrogenase and methyltransferase functions, respectively, and are also called 3-hydroxyacyl-CoA dehydrogenase type-2 and tRNA methyltransferase 10C.[21] For simplicity, this manuscript will refer to the proteins as MRPP1 and MRPP2.

The catalytic component of the human mitochondrial complex is proteinaceous RNase P (PRORP). PRORP requires the presence of MRPP1 and MRPP2 for full activity in humans although the *A. thaliana* homolog functions as a single subunit.[24] It is essential for life, and the loss of the gene in knockout mice is embryonically lethal.[25]

RNase P: An extraordinary case of convergent evolution

The broad mechanistic similarities between the two enzymes underscore the remarkable convergent evolution of the distinct RNase P enzymes. This convergence is best demonstrated by the remarkable ability of the *A. thaliana* PRORP to functionally substitute for the endogenous RNA-based RNase P in both yeast and bacteria.[26,27] It appears that the catalytic mechanisms of the ribozyme and proteinaceous forms of RNase P enzymes are similar, wherein both utilize two metal ions to perform a concerted hydrolytic cleavage

reaction.[6,28] (Figure 2). One metal (M1) serves to position and activate a hydroxide nucleophile, while a second metal (M2) stabilizes the transition state and coordinates the oxyanion leaving group. However, while the RNA-based RNase P mechanism consistently proceeds through inner-sphere coordination between the metals and the (pro-)Rp nonbridging oxygen of the target phosphate, PRORP coordinates with the (pro-)Sp oxygen.[22,29] (Figure 2)

The RNA and protein enzymes also appear to exhibit general similarities and specific differences in terms of substrate recognition. Both enzymes bind the TψC loop of the T-domain of pre-tRNA substrate, but biochemical data suggest that PRORP relies on interactions around the pre-tRNA cleavage site less than the RNA-based enzyme.[6,27,30] Human and *A. thaliana* PRORP enzymes likely utilize the pentatricopeptide repeat (PPR) domain to bind the RNA substrate but it is currently unknown how this domain interacts with structured RNA molecules.[31,32] Small angle X-ray scattering studies and analysis of point mutations suggest putative RNA-protein interactions, yet deeper insights into the mechanism of pre-tRNA recognition by PRORP await further structural studies.[31,33]

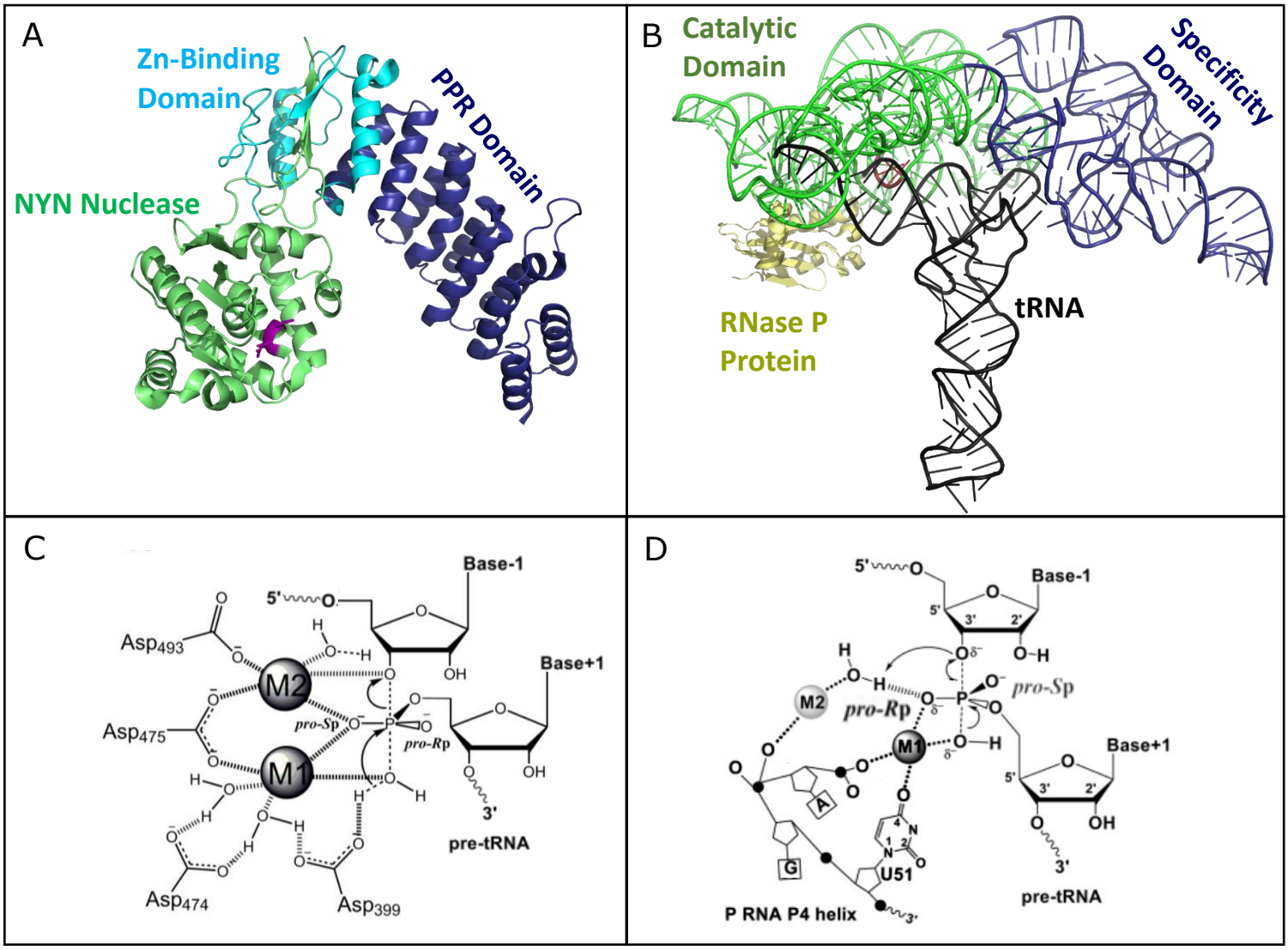


Figure 2: Structure and Mechanisms of RNase P Enzymes.
(A) *A. thaliana* PRORP1 crystal structure: the putative RNA-binding PPR domain (dark blue) is structurally connected with the catalytic NYN nuclease domain (green) by the central zinc-binding domain (cyan). The catalytic aspartates are colored purple. PDB: 4G26.[28] (B) Crystal structure of *T. maritima* bacterial RNase P enzyme in complex with tRNA substrate reveals tRNA binding through interactions between the specificity domain and T-loop and acceptor stem-loop. PDB: 3Q1Q[6]

Bottom: Current model of pre-tRNA cleavage by proteinaceous RNase P and RNA-based RNase P. The enzymes bind two catalytic metal cations, typically magnesium, through inner-sphere coordination with (C) conserved catalytic aspartates or (D) uridine residues and backbone phosphates. In addition to the metals, water molecules help stabilize the active site. Figure adapted from Howard et al., 2015, and Liu et al., 2017.[28,34]

In addition to better understanding the biology of pre-tRNA processing, studying the mechanics of proteinaceous RNase P holds larger implications for human health. There are a large number of pathological point mutants in mitochondrial tRNAs although the pathogenesis is largely unknown.[35] Many of the human mitochondrial proteins in addition to PRORP utilize PPR domains, and only a few RNA-bound PPR-domain structures have been determined.[36–38] The existing complexes are all with single-stranded RNAs, so it remains to be seen how the highly degenerate PPR protein sub-domain interacts with highly structured pre-tRNA substrates. An initial goal of my graduate career was to structurally define how the PPR domain of the protein-only RNase P interacts with tRNA.

1.3: TERRA: A Structured RNA at Telomeres

The ends of DNA pose unique problems for cells. Telomeres at the ends of chromosomes must be sequestered to prevent being recognized by and activating the DNA damage response pathways.[39] The 3' DNA strand is slightly longer, resulting in a 3' overhang. The shelterin complex folds the overhang back into the DNA by organizing the telomere into a large protected structure termed a T-loop. (Figure 3)

Telomeres shorten with each cycle of DNA replication in a phenomenon called the end replication problem. This causes genes to erode over a sufficient number of replication cycles, which can serve as a check on excessive proliferation. The telomerase enzyme elongates the telomeres through reverse

transcription by using an RNA template, resulting in a repetitive telomere sequence. This DNA sequence is conserved in vertebrates as TTAGGG, a sequence which readily forms G-quadruplex (GQ) structures.[40–42] (Figure 4)

The repetitive G-rich sequence, abundance of G-quadruplexes, and shelterin complex are all challenging elements for the DNA replication complex. During the cell cycle process, telomeres are remodeled to allow the DNA replication complex to pass through. One of the regulators of telomere protein composition is the recently discovered telomeric repeat-containing RNA (TERRA).[43] TERRA transcription begins during the G1 stage of the cell cycle and RNA levels decrease throughout the S-phase.[44] Transcription begins in the subtelomeric region and extends into the telomeric repeats, resulting in transcripts of varying length. In humans most TERRA originates from the 20q chromosome locus and is dispersed from there to the other chromosomes.[45] Deletion of the 20q locus in human cell lines results in a large decrease in TERRA levels, shortening and deprotection of telomeres in general, and a strong reduction in viability, emphasizing that TERRA is a functional lncRNA.[45]

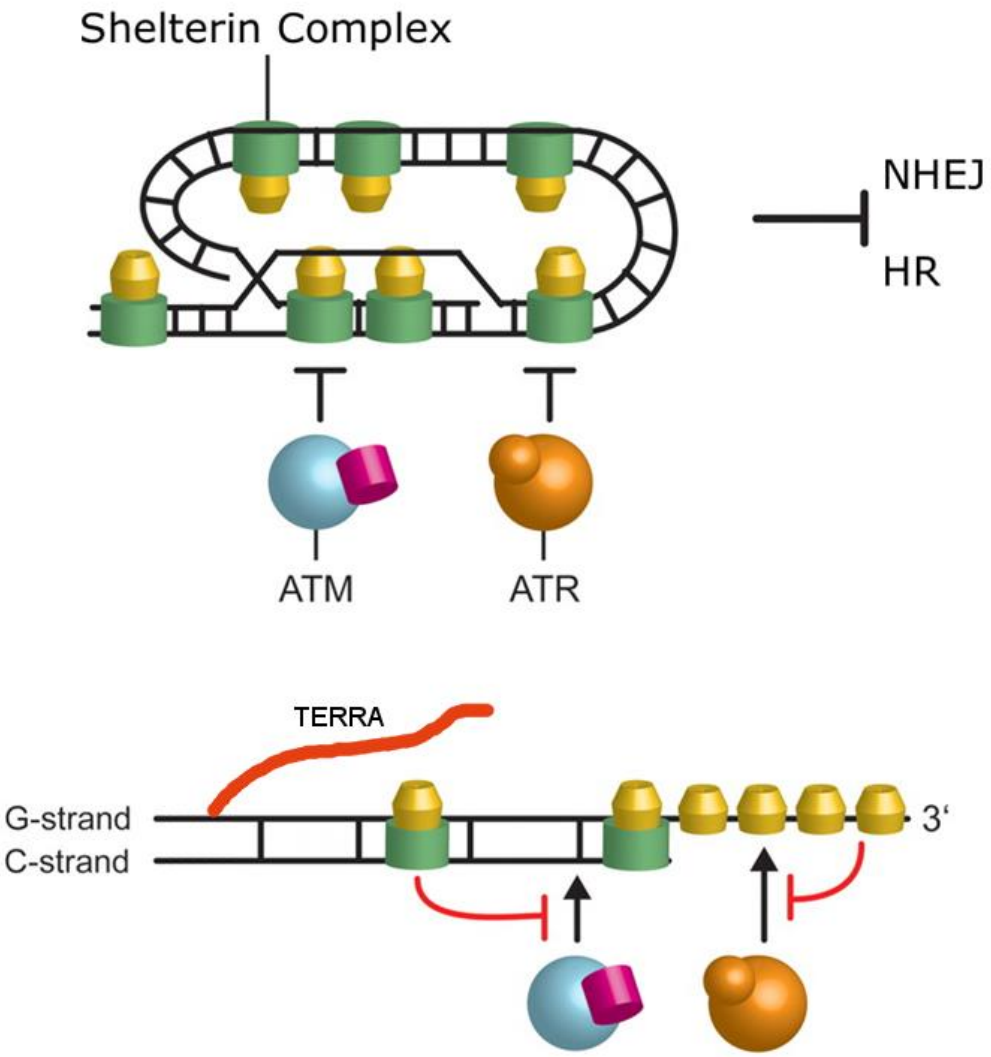


Figure 3: Telomere T-Loop Blocks the DNA Damage Response.

Top: The overhanging 3' telomere overhang strand invades the DNA duplex to form a T-loop structure. The shelterin complex organizes the T-loop structure and protects it from the DNA damage response initiator complexes ATM and ATR.

Bottom: Disruption of the T-loop structure and a decrease in the shelterin complex makes the telomere susceptible to the DNA damage response and promotes TERRA transcription.[46] Figure adapted from Longhese, M.P., 2008.[39]

RNA G-Quadruplex Formation

The telomeric repeats are known to fold into G-quadruplex (GQ) structures, intra- or inter-strand arrangements which consist of planar stacks of Hoogsteen-bound DNA or RNA guanosine tetrads.[47,48] (Figure 4) The tetrads are stabilized by monovalent ions, with potassium generally producing more stable structures than sodium. The linker regions connecting the guanosine stretches may be parallel or antiparallel in DNA G-quadruplexes. However, the extra hydroxyl group on the RNA ribose restricts the conformational space compared with DNA, allowing only the formation of parallel G-quadruplexes and resulting in a more uniform topology among different RNA quadruplexes.[49–51] The 2' hydroxyl also forms intra-strand hydrogen bonds and help to organize the surrounding water, providing increased stability.[51]

The loop sequences of G-quadruplex RNAs are variable in both sequence and length. G-quadruplex stability decreases with increasing loop length, with single nucleotide linkers being the most stable and cytidine nucleotides slightly destabilizing G-quadruplex formation.[49,50] Still, even sequences with loops approaching 15 nucleotides have been shown to possess the ability to form G-quadruplexes *in vitro* at physiological temperatures.[50]

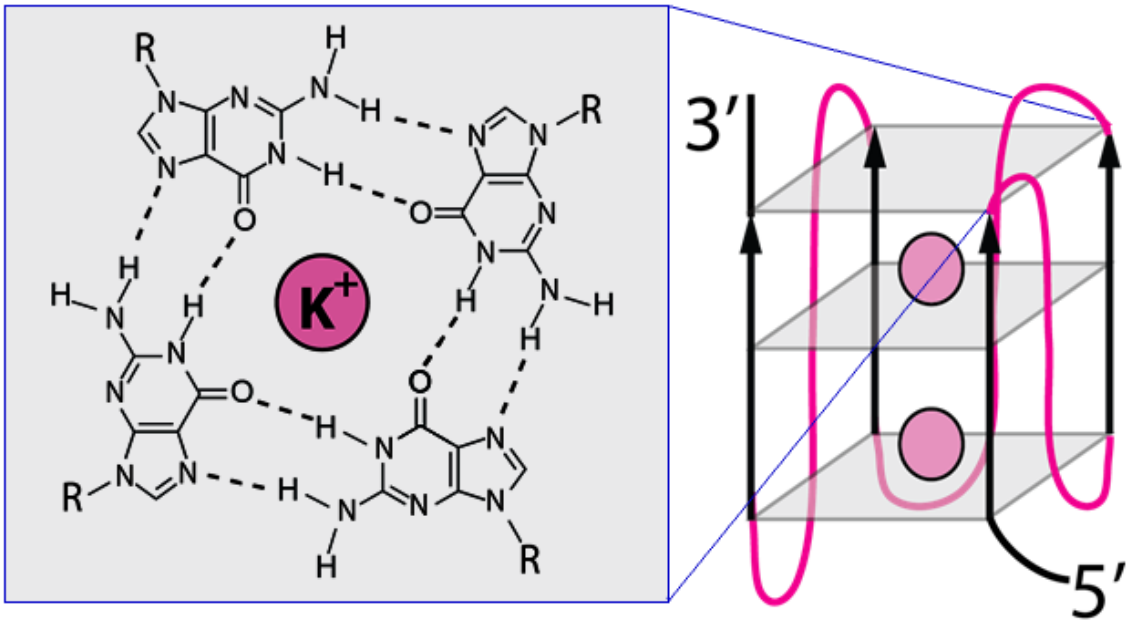


Figure 4: RNA G-Quadruplex Structure.

The GQ structure is formed by four stretches of guanosine triplets/tetrads which interact via Hoogsteen base pairing. The triplets are connected by linker regions (pink), though GQ RNAs may also be intermolecular. The structure must be stabilized by monovalent cations, usually potassium or sodium, which sit in the middle of the GQ where they help counteract the concentrated negative charge.

While it is undeniable that RNAs form G-quadruplexes *in vitro*, a recent paper from Dr. David Bartel's group cast doubt on the ability of RNAs to fold into G-quadruplexes *in vivo*.[52] The study used DMS and 2-methylnicotinic acid imidazolide (NAI) to probe G-quadruplex forming sequences *in vitro* and in cell culture, and found that while G-quadruplexes can fold in isolated RNA they are generally unfolded and available for modification in living cells. It is possible that G-quadruplexes exist only transiently, folding and re-folding, or only form in certain subcellular environments. Still, Bartel's conclusions are difficult to reconcile with the large body of evidence for the existence and function of G-quadruplexes in mammalian cells, as described below.

Functions of RNA G-Quadruplexes

Putative G-quadruplex forming elements are found throughout the genome and transcriptome, and are enriched at the regulatory 5' and 3' UTRs of mRNAs while being depleted in coding sequences compared with the expected occurence.[52,53] Immunofluorescence using G-quadruplex specific antibodies confirms the presence of G-quadruplexes in both the nucleus and cytoplasm.[54] The signal is RNase sensitive and amplified by application of RNA G-quadruplex binding small molecules, confirming the specificity of the antibody signal.

5' UTR G-quadruplexes are generally repressive, though a number of genes have conserved G-quadruplexes in the 5'-UTR which appear to promote translation as part of the cap-independent internal ribosome entry site.[55–58]

While the mechanisms are not clear, G-quadruplex RNAs are thought to exert their function through a number of G-quadruplex binding proteins.[59] Fragile X mental retardation protein (FMRP) is among the best characterized RNA G-quadruplex binding proteins. The FMRP protein regulates the localization and translation of target mRNAs and is critical for proper neural function.[60,61] Numerous cross-linking studies have found an enrichment of putative G-quadruplex forming sequences in target mRNAs.[60,61] Studies have found that FMRP binds to G-quadruplexes 5' of the translational start site of target transcripts, including the FMRP transcript, and represses translation.[62,63]

The arginine-glycine rich (RGG) motif of FMRP provides the only structure of a natural peptide in complex with an RNA G-quadruplex (PDB 5DE5).[64] The structure shows that the RGG peptide does not contact the G-quadruplex directly, but rather binds the transition between the duplex RNA and the G-quadruplex. These GQ-binding motifs contain clusters of the sequence RGG and are involved in both RNA and DNA binding. Other RNA binding proteins, such as the RGG-containing nucleolin protein, also bind both RNA and DNA G-quadruplexes.[65,66] At this time it is too early to speculate on whether other RNA G-quadruplex binding proteins employ similar binding mechanisms.

In 2014, Dr. Joachim Lingner reported that disruption of the shelterin complex induces the transcription of TERRA RNA which binds the protein lysine-specific demethylase-1 (LSD1).[46] The TERRA-LSD1 complex then recruits the DNA repair nuclease MRE11, which trims the overhang to resemble a blunt double-

stranded DNA break. The process is thought to promote non-homologous recombination of telomeres, leading to chromatin instability.

How does this interaction affect cell biology, and does the specific TERRA-LSD1 interaction serve a purpose in the cell? How common is this RNA-based recruitment strategy? This dissertation begins to answer how TERRA recruits LSD1 and whether additional RNAs might recruit the LSD1/CoREST complex to other regions of the genome.

1.4: Epigenetic Regulation of Gene Expression by LSD1

While LSD1 functions as a molecular scaffold at telomeres, it is best known as an essential epigenetic regulator. Dr. Yang Shi characterized LSD1 as the first known histone demethylase in 2004, demonstrating that histone methylation is a reversible and dynamic modification.[67] The LSD1 functions are vital, as demonstrated by the arrest of LSD1 knockout mouse embryos at only 7.5 days.[68] While it has many gene targets and functions, its general biological role is epigenetically reprogramming cells during differentiation, especially embryonic stem cells.[69] LSD1 is an oncogene which is commonly misregulated in a number of cancer types such as leukemia and non-small cell lung cancer, and is a significant therapeutic target.[70–72] A number of small-molecule LSD1 inhibitors are under development for cancer treatments.[73] A better understanding of LSD1 regulation has the potential to aide in the identification of the ideal candidates for LSD1 therapy and more specific targeting of LSD1

complexes.

Effects of Histone Demethylation

Histone methylation is a relatively subtle histone modification. Unlike phosphorylation and acetylation, it does not directly alter the residue's charge or significantly affect its interaction with DNA. Instead, histone methylation influences the chromatin state and transcription by regulating the recruitment of reader proteins and histone modifying proteins. It is associated with both heterochromatin and euchromatin, and its transcriptional effects are highly context-dependent.[74]

LSD1 acts upon mono- and dimethylated H3K4 and H4K9, two of the more commonly methylated histone residues. However, LSD1 knockdown also affects H3K4 trimethylation at target genes.[75] As many H3K4 methyltransferases cannot catalyze all the steps from H3K4 to H3K4me3 on their own, H3K4me1/2 are intermediates for the generation of H3K4me3.[76] H3K4 methylation is strongly associated with active enhancers and transcriptional start sites (TSS). H3K4 is predominantly trimethylated around the TSS but transitions to H3K4me2 further out, with mostly H3K4me1 modifications ~1 kb from the transcription start site.[77] Once deposited, methylated H3K4 recruits a variety of histone remodelers, transcription factors, and components of the RNA polymerase preinitiation complex to promote transcription.[78,79]

In contrast, H3K9 methylation is typically associated with heterochromatin and

gene repression. H3K9me3 is specifically recognized by heterochromatin protein 1 (HP1), which oligomerizes on chromatin to bridge nucleosomes and organize histones into heterochromatin.[80,81] However, while H3K9me2/3 are primarily found at silent genes, H3K9me1 is actually present at many active promoters.[77] Therefore, H3K9 demethylase activity is primarily viewed as disinhibiting of gene expression but in certain circumstances it may actually be repressive.

Relationship between LSD1 and Related Demethylases

LSD1 is a member of the flavin monoamine oxidase protein family, a group of proteins which oxidize various amines using the cofactor FAD.[82] (Figure 5) While histones are its the best known substrates, LSD1 also catalyzes the removal of methyl groups from a number of other proteins including DNA methyltransferase 1 (DNMT1) and the cell cycle regulatory proteins p53, E2F1, and myosin phosphatase target subunit 1 (MYPT1).[83–86] Demethylating specific lysine residues of these proteins alters their stability and function, allowing LSD1 to indirectly modulate cell fate, proliferation, and gene expression.

Additional histone demethylases have come to light since the discovery of LSD1. (Table 1) Jumonji domain-containing histone lysine demethylases utilize an alternative, iron-dependent mechanism that uses alpha-ketoglutarate as a cofactor in place of FAD.[82,87] (Figure 5) Unlike LSD1, the Jumonji domain proteins can use trimethylated lysine as substrate in addition to di- and

monomethylated residues and have a range of specificities.

LSD1 also has one human homolog, LSD2. While LSD2 also demethylates H3K4me2, it forms distinct complexes from LSD1 and localizes to the coding region of genes instead of promoters.[88,89] The mouse LSD2 has been demonstrated to also demethylate H3K9me2 in an NF-κB dependent manner.[90] LSD2 contains a zinc finger region not found in LSD1 that provides it with the potential to directly target and bind histones without the assistance of additional proteins. The zinc finger domain is also likely to function as an E3 ubiquitin ligase, making LSD2 an exceptionally versatile enzyme.[91]

LSD1 Forms Larger Chromatin-Modifying Complexes

Since LSD1 lacks the zinc finger of LSD2, it must form complexes with other proteins and RNAs in order to bind target genes. LSD1 contains a long, alpha-helical region termed the tower domain which is tightly bound by the corepressor of RE1 silencing transcription factor (CoREST).[92] CoREST is likely constituently bound to LSD1; it stabilizes LSD1 and its absence results in lower LSD1 levels.[93] CoREST also promotes substrate binding and the formation of larger protein complexes. This is accomplished with the aid of two SANT domains in CoREST which are involved in chromatin binding. While LSD1 is able to demethylate short histone peptides on its own, it requires CoREST in order to bind and demethylate nucleosomes.[93,94]

There are three high-homology CoREST paralogs termed CoREST 1,2, and 3, and while most studies- including this one- have focused on the more prevalent LSD1/CoREST1 complex, all three proteins are capable of binding LSD1.[95,96] Whereas the CoREST homologs have high conservation, the different LSD1/CoREST complexes have discrete functions and varying ability to complex with additional proteins.[96]

LSD1/CoREST forms a variety of complexes that regulate its localization and activity. The LSD1/CoREST complex associates with histone deacetylases 1 and 2 (HDAC1/2) through CoREST to form a repressive complex which inhibits the transcription of target genes.[97,98] HDAC1/2 are highly homologous and largely redundant.[99] By removing acetyl groups from histone lysines HDACs promote tighter binding of DNA, resulting in heterochromatin formation and generally repressing transcription. The LSD1/CoREST/HDAC complex is cooperative both *in vitro* and *in vivo*, and H3K4 demethylation and H3 deacetylation activity is inhibited by the absence or inhibition of either LSD1 or HDAC1.[98]

The LSD1/CoREST/HDAC complex can in turn bind the transcriptional corepressor C-terminal binding protein 1 (CtBP1) to form a larger complex.[100,101] The core of the CtBP1 complex is thought to include LSD1, CoREST, and histone deacetylases 1 and 2 (HDAC1/2).[97] CtBP1 functions as a molecular bridge between transcription factors such as zinc finger proteins, which provide specificity for regions of chromatin, and epigenetic regulators such as LSD1 which directly modify the 'histone code' by catalytically altering

the methylation and acetylation state of histones.[100] The targeting of the CtBP1/LSD1/CoREST/HDAC1/2 complex by zinc finger proteins has been shown to regulate tissue differentiation and oncogenesis *in vitro* and *in vivo*.[68,102,103] The complex suppresses target genes, at least in part by LSD1-induced demethylation of H3K4me1/2.

Although LSD1/COREST only demethylates H3K4 *in vitro*, it can also act upon H3K9 when bound by certain proteins.[42,67,94,104] The androgen receptor is a transcription factor which regulates a variety of genes involved in cell proliferation and communication by binding to androgen response elements, which are typically located near the transcriptional start sites of genes.[105] The androgen receptor normally regulates tissue development but is also a critical driver of prostate cancer. In the presence of a receptor agonist the androgen receptor binds LSD1 and serves as a transcriptional activator by removing repressive H3K9 mono/di-methyl groups at androgen receptor target sites.[104] Similarly, the zinc finger transcription factor RREB1 and estrogen-related receptor α have both been shown to form a complex with LSD1 and broaden its specificity to include H3K9me1/2 both in cell culture and *in vitro*.[106,107]

Table 1: Lysine Demethylase Enzymes.

Family	Enzymes	Mechanism	Histone Substrate
LSD1	LSD1 (KDM1A)	Flavin-dependent monoamine oxidase	H3K4 and H3K9
	LSD2 (KDM1B)		
UTX	UTX (KDM6A)	Fe(II)- and 2-oxoglutarate-dependent oxygenase	H3K27
JARID	JARID1B (KDM5B)		H3K4
JMJD	JMJD1A (KDM3A)		H3K9
	JMJD2A (KDM4A)		H3K9 and H3K36
	JMJD2B (KDM4B)		H3K9
	JMJD3 (KDM6B)		H3K27

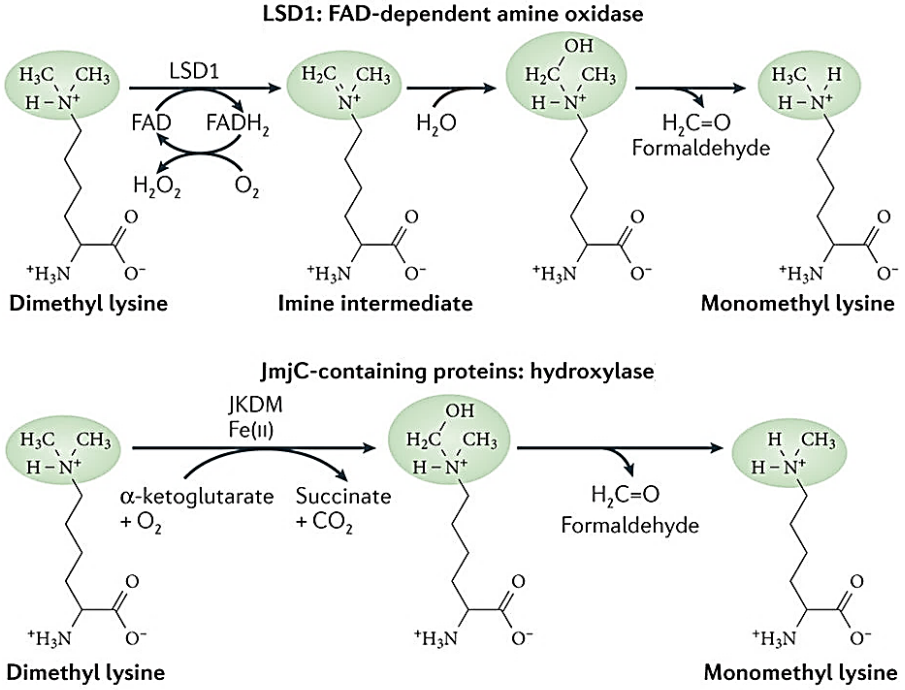


Figure 5: Mechanism of LSD1 and JmjC-class Demethylases.

Adapted from Hamamoto et al., 2015.[108]

LSD1-RNA Complexes

lncRNAs have long been known to recruit histone regulators to chromatin, the most notable instance being the lncRNA X inactive specific transcript (Xist). Xist coats the inactive X chromosome and uses a repetitive, structured repeat sequence termed RepA to recruit polycomb remodeling complex 2 (PRC2); PRC2 lays down repressive H3K27me3 marks to begin the silencing of most of the X-chromosome.[109]

In 2010 Dr. Howard Chang published a groundbreaking paper demonstrating that the lncRNA HOX transcript antisense intergenic RNA (HOTAIR) can function similarly in *trans*.[110] HOTAIR is transcribed from the HOX C gene locus on chromosome 12, tethers PRC2 and LSD1/CoREST to form an RNP complex, and translocates to chromosome 2 where PRC2 and LSD1 silence the HOX D cluster. Combined with the increased recognition of the scope and tissue-specificity of lncRNA expression following the publishing of the ENCODE project, this ignited an interest in lncRNAs as critical epigenetic regulators which use sequence and structure to recruit protein complexes.[111]

In 2009 it was demonstrated by RNA immunoprecipitation followed by microarray analysis (RIP-Chip) that a significant proportion of lncRNAs are physically associated with CoREST and PRC2, though it is not clear whether LSD1 was complexed with the CoREST.[112] A survey of RNA-binding proteins was then published in 2016 by Dr. John Rinn's group using a new technique termed formaldehyde RNA immunoprecipitation (fRIP).[113] This method uses

formaldehyde to covalently cross-link RNA/DNA with proteins and the complex is then immunoprecipitated. The method has significant potential for false positives due to the possibility of inadvertent co-immunoprecipitation of other RNA-binding proteins in complex with LSD1. The study is dependent on the specificity of the LSD1-targeting antibody, but it provides some information about the nucleic acid binding properties of LSD1-containing complexes. LSD1 was found to associate with significantly more DNA than RNA, as might be expected for a chromatin remodeling protein. The RNAs which immunoprecipitated with LSD1 were primarily mRNAs and LSD1 preferentially bound exons.

There is no clear functional role for this observed mRNA binding by LSD1, and further work is required to validate the finding. Biochemical studies have focused on lncRNA-LSD1 interactions, and found that LSD1 is recruited to specific genes in an RNA-dependent manner and then regulates gene expression by either H3K4 or H3K9 demethylation.[114–117] This dual specificity may reflect the recruitment of different LSD1-containing protein complexes by various RNAs or direct regulation of LSD1 by bound lncRNAs.

Although RNA has been shown to recruit LSD1 to gene targets, many unknowns remain. The RNA binding domain of which interfaces with the lncRNAs in unknown, and it is not clear whether these RNA interactions are occurring through LSD1 directly, CoREST, or another factor in one of the LSD1 complexes. The determinants that dictate which RNAs are bound by LSD1 also remain poorly defined. Finally, it has not been clearly demonstrated what

effect, if any, RNA has on LSD1 catalytic activity and substrate specificity. The following study investigates these questions regarding the mechanisms of TERRA-RNA interactions using a variety of structural biology and biochemical approaches.

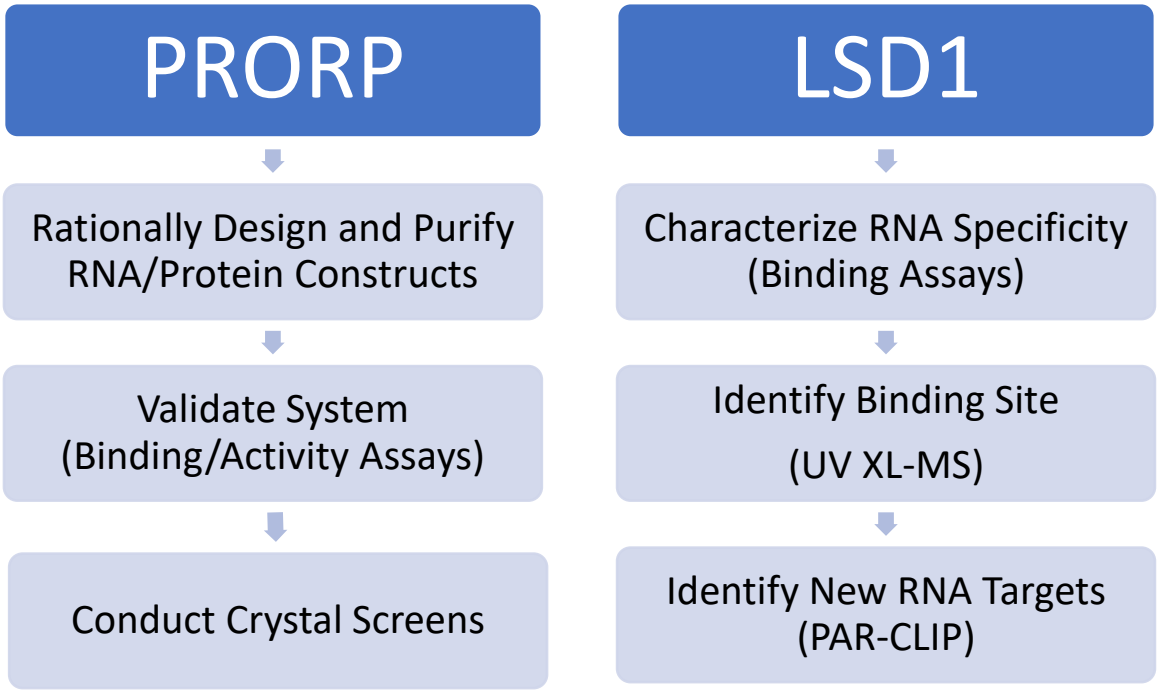


Figure 6: Overview of Approach.

Flowchart of the general strategies described in this manuscript that were used to better characterize the binding of proteinaceous RNase P/LSD1 with structured RNAs.

CHAPTER 2

Substrate Recognition by Proteinaceous RNase P

2.1 Introduction

tRNAs and the Mitochondrial Transcriptome

Transfer RNAs (tRNA) are transcribed with excess nucleotides at the 5' and 3' ends. The surplus oligonucleotides, termed the 5' leader and 3' trailer sequences, must be cleaved at the appropriate location to allow further tRNA processing and aminoacylation. The 5' leader is typically cleaved first by the RNase P enzyme, followed by removal of the 3' leader by RNase Z.[118] This ensures a uniform acceptor stem for subsequent aminoacylation. A ribozyme RNase P enzyme was likely present in the last common ancestor and is found in all domains of life, though some species have either replaced it with an unrelated protein or, in rare instances of severely constricted genomes, dropped the need for RNase P altogether by tightly regulating the start of tRNA transcription at the desired nucleotide.[24,119,120]

Human mitochondria present one such example of the replacement of the RNA-based RNase P by an alternative protein, PRORP. PRORP is a nuclear-encoded protein present is higher eukaryotes and is transported to human mitochondria.[121] The reason for the loss of the ancient RNA-based enzyme is unclear, but may be related to evolutionary pressures to reduce the organelle's genome. The emergence of a proteinaceous RNase P was not inevitable, as evidenced by the continued use of ribozyme RNase P in the mitochondria of

other eukaryotes such as *S. cerevisiae* which encodes the RNA ribozyme in the mitochondrial genome and imports a nuclear-encoded protein cofactor.[122]

When the human mitochondrial genome was sequenced, it was found that mitochondrial tRNAs (mt-tRNA) do not adhere to the same rules as nuclear tRNAs.[123] Human mt-tRNAs lack many of the conserved sequences and tertiary interactions found in nuclear tRNAs, such as the normally 7-nucleotide TψC loop ranging from three to nine nucleotides in mt-tRNA. As these noncanonical features are found in human mt-tRNAs but not the more conventional mt-tRNAs of *S. cerevisiae*, it is possible that the alternative RNase P enzyme loosened the structural restraints on mitochondrial tRNAs.[124]

Mt-tRNAs diverge from their nuclear counterparts in other ways as well. The mitochondrial genome is transcribed into polycistronic transcripts punctuated by tRNAs which are cleaved at the 5' and 3' ends by RNase P and RNase Z, respectively.[125–127] Proper mitochondrial RNase P activity is not only necessary for the maturation of pre-tRNA but the generation of the other mitochondrial gene transcripts. Separating the transcripts allows them to be individually regulated and polyadenylated as required. The ND5 and CytB genes pose an exception; the two adjacent genes lack a tRNA spacer and are likely separated by the ribonuclease pentatricopeptide repeat domain protein 2 (PPRD2).[128] Between 1) this added mt-tRNA functionality, 2) a higher mitochondrial mutation rate, and 3) a lack of redundancy (the human mitochondrial genome encodes 22 tRNA genes while the nuclear genome

contains ~500 tRNA genes), mt-tRNAs are much more disease-associated than nuclear tRNAs. In fact, all characterized disorders arising from tRNA mutations in the cell involve mt-tRNA.[108]

Animals inherit multiple maternal mitochondria. As the mitochondria proliferate they are dispersed unevenly throughout the body, resulting in heteroplasmy that varies from tissue to tissue. This complicates the study of mitochondrial disease pathogenesis. Nevertheless, it is clear that mt-tRNA are hot spots for mitochondrial diseases. The majority of pathogenic mitochondrial mutations occur are mt-tRNA SNPs; these mutations are often associated with neuromuscular and cardiomyopathy diseases.[129–131] Mutations in the Leu(UUR) mt-tRNA are particularly associated with a range of complex disorders, including mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS) syndrome.[132] MELAS mutations and other mt-tRNA SNPs are associated with the accumulation of unprocessed transcripts and are cleaved less efficiently by RNase P, suggesting that disruption of tRNA folding inhibits the ability of PRORP to bind and cleave pre-tRNA.

PPR Domains and RNA Processing

The RNA-based RNase P has been well studied and biochemical and structural work provides a basis for understanding pre-tRNA binding.[6,13,133] Much less is understood about substrate binding by proteinaceous RNase P. This study focuses on both the human and plant PRORP enzymes. There are three PRORP

paralogs in *A. thaliana* (PRORP1, 2, and 3). The enzymes are highly similar in sequence and structure but differ in localization, with PRORP1 functioning in chloroplast and mitochondria while PRORP2 and PRORP3 are localized in the nucleus of plant cells.[24,28,134,135] Comparison of the various human and plant PRORP proteins provides insight into how the enzymes are regulated. The enzymes consist of three domains: 1) a pentacotripeptide repeat (PPR) domain made up of a series of 35-amino-acid motifs, each of which forms a helix-turn-helix structure; 2) a magnesium-dependent metallonuclease domain; and 3) a central zinc-binding domain which orients the other two domains relative to each other.[136]

Numerous studies have revealed clues to the question of why the human PRORP must form a complex with MRPP1/2 to function. There are only three α-helices in the PPR domain of human PRORP whereas eleven α-helices exist in the plant PRORP1. Shortening of the PRORP1 PPR domain to three helices causes a dramatic reduction of both RNA binding and cleavage, which suggests that a shorter PPR domain is less functional.[28] However, a chimeric form of the human enzyme containing the PRORP1 catalytic domain is active on its own *in vitro,* indicating that the shorter human PPR domain is functional and is not in itself the root cause of human PRORP's dependence on the MRPP1/2 proteins.[137]

Comparing the crystal structures of the catalytic domains of plant and human PRORPs reveals that the human catalytic domain alone folds into an inhibited

conformation, though it is possible that this is a crystal artifact and does not answer the question of how it might be restructured by MRPP1/2.[28,137] One hypothesis stems from structural studies of the protein phosphatase 5 (Ppp5). The Ppp5 tetratricopeptide repeat (TPR) domain, which is closely related to the PPR domain, interferes with the active site and autoinhibits catalysis. Ppp5 is only activated when the heat shock protein Hsp90 disinhibits the enzyme by binding the TPR domain.[138] The PPR domain may play a similar role in human PRORP.

The primary function of the PPR domain is RNA binding. The PPR domain likely evolved in eukaryotes from the more ubiquitous TPR motif. PPR-containing proteins function in chloroplast or mitochondrial organelles, especially in plants.[136] All seven humans PPR protein are localized to mitochondria where they function in RNA transcription, processing, stability, and translation, though the molecular mechanisms are largely unclear.[139] Better knowledge of PRORP-pre-tRNA interactions holds implications for understanding the multi-functional roles of PPR protein domains.

Repeats of the 35-residue helix-turn-helix motif combine to form a super-helical structure which can bind RNA with some sequence specificity. PPR motifs can exhibit an amino acid code similar to the modular transcription activator-like effector nuclease (TALEN) and Pumilio family (PUF) motifs, raising interest in its use as a method to target specific RNA sequences. The 1st and 6th amino acids of each helix-turn-helix motif interact with single-

stranded RNA bases, and varying these residues can tailor specificity to certain sequences.[140] Crystal structures of plant PPR proteins in complex with single-stranded RNAs demonstrate how the protein conforms to bind the RNA and establish backbone and base-specific interactions.[38,140]

Despite this detailed knowledge and structures of PPR proteins in complex with single-stranded RNA substrates, attempts to identify a PPR code in the context of pre-tRNA recognition have failed, suggesting that the system is more complex.[30] The identity of the PRORP amino acids in the base-recognition positions as defined by the PPR code are: 1) not universally conserved; 2) not canonical base-recognition residues; and 3) not the most important PPR residues for pre-tRNA binding.[69] Therefore, questions remain about how PPRs recognize highly structured RNAs such as pre-tRNA.

It has been proposed that the PPR domain could bind the single-stranded 5' leader.[28] However, as PRORP has been shown to cleave leaders as short as 5 bases, it would be difficult for the short substrate leader to engage in important interactions with the PPR domain and also span the distance to the active site.[31] Modeling studies suggest that is more likely that the PPR domain orients the substrate by binding the stacked TψC and D arms of pre-tRNA, as these motifs are a prerequisite for proper cleavage and contain conserved residues which may interact with the PPR domains.[31] The 'elbow' formed by the TψC/D loops are also critical for proper orientation of pre-tRNA in the ancient ribozyme RNase P.[6] Further studies of the PRORP-pre-tRNA complex

will improve our understanding of the processing of mitochondrial transcripts by PPR proteins in general and how mitochondrial mutations lead to defects in the RNA processing pathways.

2.2: PRORP Activity and pre-tRNA Binding

While the initial goal of this study was to combine biochemical assays with crystallography in order to better understand the mechanisms of proteinaceous RNase P, a number of subsequent biochemical studies shed light on the residues and potential modes of RNA recognition as described previously.[29,31–33,69,141] Our focus therefore adjusted to concentrate on optimizing conditions to promote crystallization of a PRORP-RNA complex.

As the N-terminal region of PRORP3 is predicted to be disordered, a truncated PRORP construct (ΔN PRORP3) was cloned and purified. The ΔN PRORP3 lacks the first 86 amino acids preceding the first alpha-helix, and the truncation does not appear to have a significant effect on catalytic activity. (Figure 7) As disordered protein regions are commonly thought to impair crystal formation or reduce diffraction by crystals, the removal of the N-terminal unstructured region was expected to improve crystallization prospects.

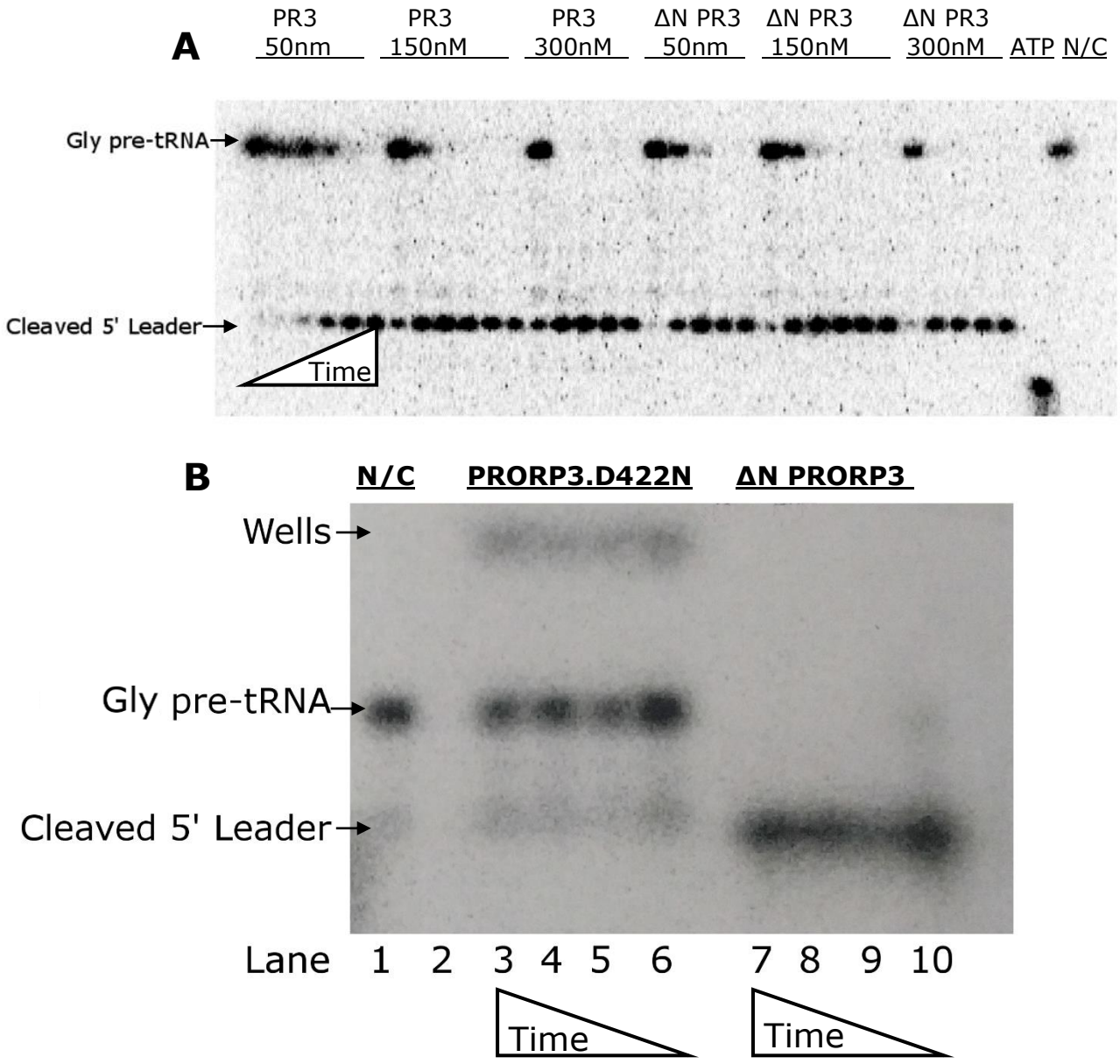


Figure 7: PRORP3 pre-tRNA Cleavage Assays.

(A) Full-length and truncated PRORP3 have comparable enzymatic activity. Enzyme was incubated with 5' ^{32}P-labeled Gly pre-tRNA at room temperature for 0.25, 4, 16, 40, and 120 minutes at room temperature under single-turnover conditions and run on a denaturing 8M urea, 1xTBE polyacrylamide gel. ^{32}P-ATP and protein-free pre-tRNA are included as markers and a negative control.
(B) Gly pre-tRNA was incubated with 50 nM PRORP3.D422N (lanes 3-6) or ΔN PRORP3 (lanes 7-10). Samples were incubated at room temperature for 10 seconds, 90 seconds, 3 minutes, or 10 minutes. NC: Protein-free Lane 1 contains no protein. While ΔN PRORP3 displays activity comparable to WT PRORP3, PRORP.D422N displays no significant enzymatic cleavage and appears to form a stable complex with the substrate which remains at the top of the gel.

Two approaches were taken to deactivate *A. thaliana* PRORP3 and stabilize the enzyme-substrate complex for crystallization. Site-directed mutagenesis of the active site aspartate to asparagine abolished catalytic activity in PRORP3.D422N without affecting global protein stability. Activity assays of the active-site variant confirmed that the enzyme was inactive. (Figure 7) Furthermore, it was observed that some of the PRORP3.D422N-incubated pre-tRNA failed to properly enter the gel, indicating that it formed a stable RNP complex.

PRORP is magnesium dependent, and its catalytic activity was inhibited by the presence of calcium. (Figure 8) This allows us to directly monitor and quantify binding in order to assess the co-crystallization of RNAs with wild-type and mutant PRORP. Electrophoretic mobility shift assays (EMSAs) revealed that PRORP3 binds tyrosine pre-tRNA substrate with micromolar affinity.

Previous studies demonstrated that removal of the anticodon stem-loop has no significant effect on PRORP binding or cleavage, and reducing the D-domain to a simple bulge had only a modest effect on the PRORP interaction.[30] With this in mind, the *A. thaliana* glycine pre-tRNA was modified into a minimal RNA with no anticodon domain and a kissing loop-containing bulge in place of the D domain. (Figure 9) Additional binding assays confirmed that PRORP3 binds a minimal glycine pre-tRNA with the same affinity as full-length pre-tRNAs. (Figure 8)

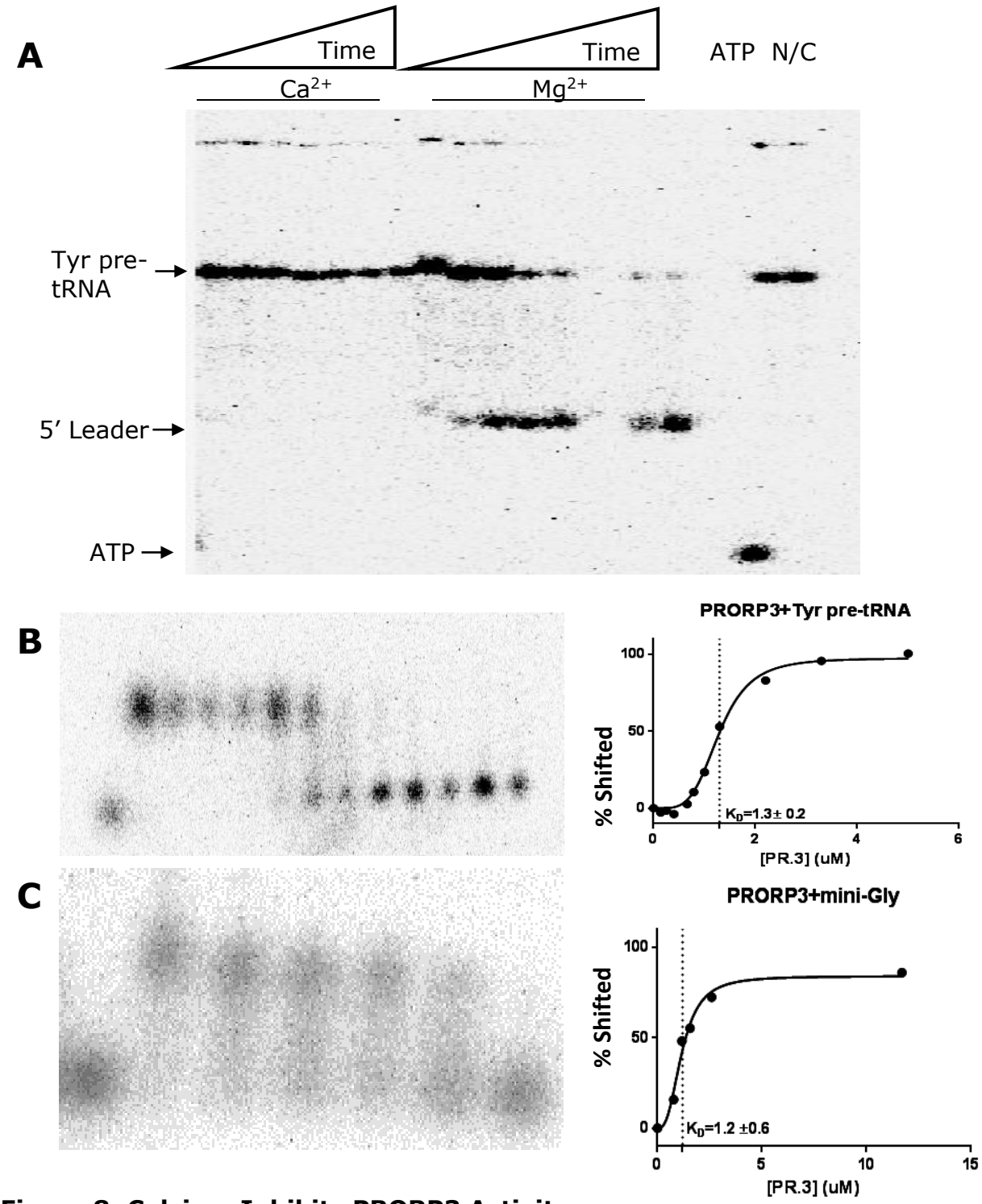


Figure 8: Calcium Inhibits PRORP3 Activity.

(A) PRORP3 activity assay was conducted with 1 μM PRORP3 and tyrosine pre-tRNA using either 10 mM $CaCl_2$ or 10 mM $MgCl_2$. Timepoints were taken at 0.25, 1, 4, 16, 25, 40, and 135 minutes. Calcium effectively inhibits the cleavage reaction. EMSA gels analyzed binding between PRORP3 and tyrosine (B) and mini-glycine (C) pre-tRNAs in the presence of calcium yielded similar K_D values of 1.3 μM and 1.2 μM, respectively.

2.3: Crystallization Screens

Both the human PRORP and *A. thaliana* PRORP3 were used in crystallography screens. The *A. thaliana* enzyme is not only able to function without additional proteins but is generally more stable in solution, making it the primary focus of crystallization efforts. In order to obtain protein crystals, it is desirable to obtain pure, concentrated samples. Glycerol promotes protein solubilization and allows greater concentration, but is detrimental at high levels in crystal screens as glycerol generally inhibits the nuclearization process and can lead to changes in the drop volume over time.[142] It was observed that in buffers with low glycerol content (~2.5% v/v) human PRORP was not soluble beyond ~0.4 mg/mL, more dilute than preferred for crystallization. This was largely remedied by the addition of tRNA that greatly promoted the stabilization and solubility of human PRORP.

A variety of RNA substrates were included in crystallization screens, including both native substrates and highly engineered RNAs. Two approaches were applied to the design of engineered RNAs: 1) Elimination of unnecessary elements to reduce RNA bulk; and 2) Promoting intermolecular contacts by incorporating elements such as kissing loops and tetraloops. Kissing loops and tetraloops dimerize into tertiary structures and are effective tools for promoting crystallization.[4,143] Both kissing loops and tetraloops were incorporated in the design of variants of the minimal glycine pre-tRNA which was validated in binding assays for crystallization trials. Full length pre-tRNAs were also altered by varying the anticodon stem-loop, replacing the loop with

the tetraloop-forming sequence GUGA and switching the native anticodon stem sequence out for a panel of more stable stems composed of GC base pairs. Finally, varied U1A binding sites were inserted into the anticodon domain of full-length glycine pre-tRNAs with variable length anticodon stems. U1A is a spliceosomal protein commonly used to stabilize engineered RNAs in structural studies, thus potentially facilitating both crystallization and phase determination.[144] The U1A RBD protein was purified and crystal trays were prepared with samples containing engineered glycine pre-tRNA, U1A protein, and PRORP3 constructs.

Crystal trays were set-up with a variety of conditions using both custom and commercial screens in an effort to obtain a crystal structure of PRORP either alone or co-crystalized with RNA. A number of mutations were screened, including inactive variants with point mutations at the active site and truncated proteins lacking the flexible N-terminal region. Unfortunately, the only successful crystal was of *A. thaliana* PRORP3. (Figure 10) This crystal diffracted to ~2.8 Å but consisted of two connected crystal domains with separate orientations, resulting in a twinned dataset which we were unable to successfully solve. Attempts to reproduce the crystal proved unsuccessful.

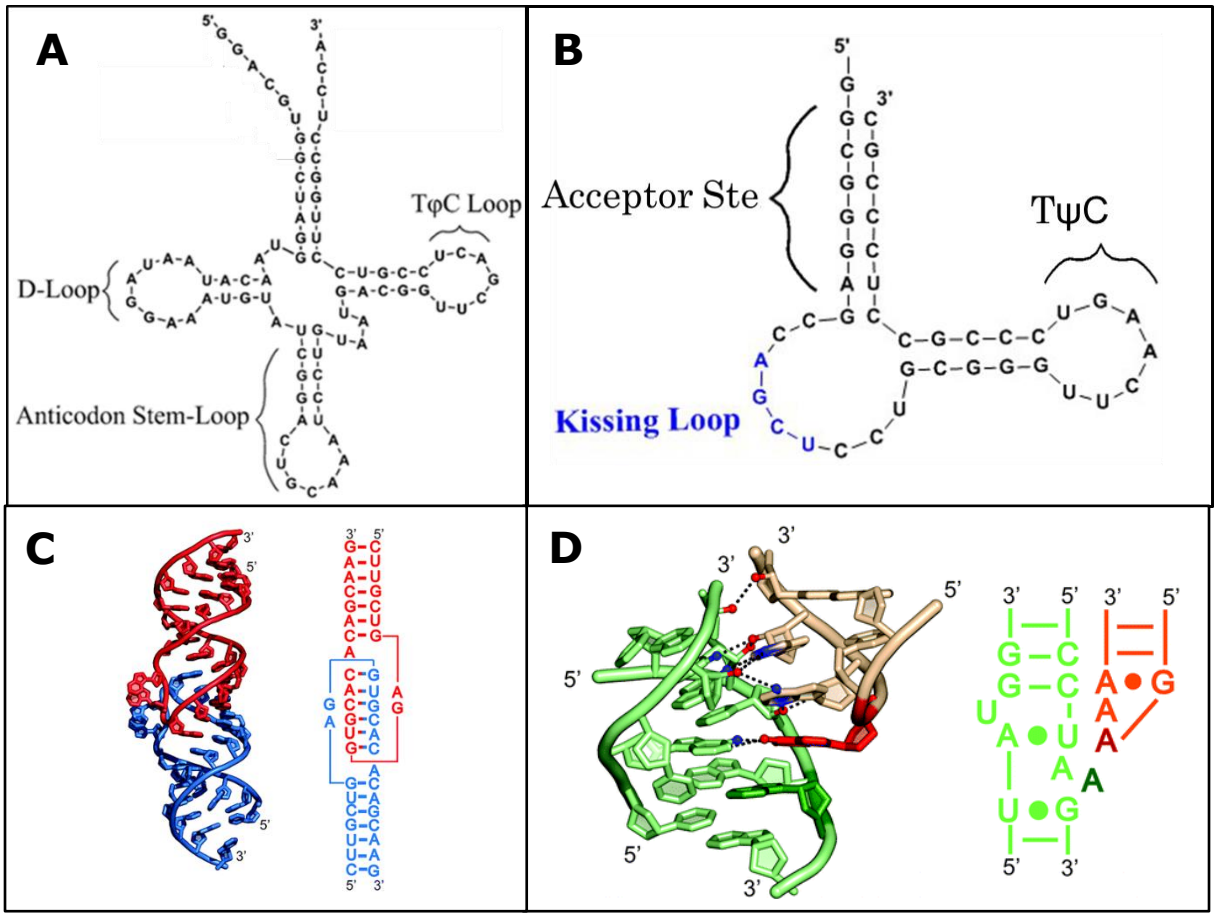


Figure 9: tRNA Design for Crystallographic Studies.

Examples of tRNAs used in biochemical and structural experiments include the full-length *T. thermophilus* glycine pre-tRNA (A) and a mini-glycine RNA (B) designed based on work from the Rossmanith group.[30] The mini-glycine construct eliminates the anticodon stem-loop, which is not involved in PRORP binding, and replaces the D-loop with a bulge containing a kissing loop sequence. The kissing loop motif (C) forms intermolecular tertiary interactions with other kissing loops, encouraging potential lattice contacts. Tetraloops also form structured tertiary intermolecular contacts (D), and the anticodon stem-loop was replaced by variants containing the tetraloop-forming sequence GUGA. Figures C and D were adapted from Butcher and Pyle, 2011.[4]

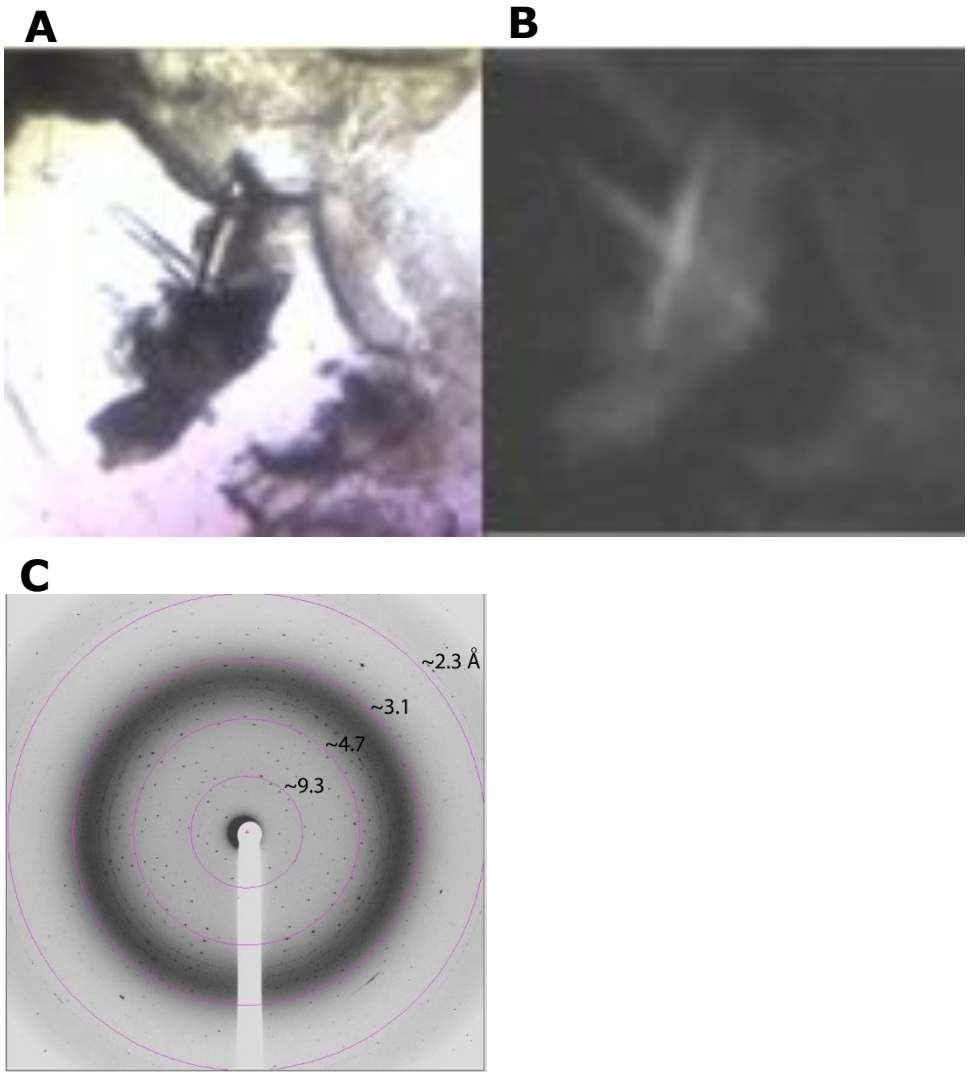


Figure 10: PRORP3 Crystal.

Top: PRORP3 crystal viewed under (A) visible and (B) UV light. (C) The crystal isotopically diffracted past 2.8 Å.

The crystal was grown in 37.5% PEG 4000, 50mM Tris (pH 8), and 75mM LiCl, 5mM $MgCl_2$.

2.4: Discussion

Although the crystallization of PRORP in complex with substrate was ultimately unsuccessful, numerous insights were gained which can aide future attempts. The reservoir solution which yielded the diffracting PRORP1 crystal (37.5% PEG 4000, 50mM Tris pH 8, 75mM LiCl, and 5mM $MgCl_2$) was significantly different from conditions previously published for PRORP1 crystals (18% PEG 3350, 100 mM sodium citrate pH 5.5), indicating that it might prove a distinct crystal form.[28] Truncation of the unstructured N-terminal region had no significant impact on PRORP3 activity. Inactivation of PRORP3 via calcium or inactivating mutations was consistent with previously reported results with PRORP1.[141] Here, we demonstrate that these strategies provide viable mechanisms to form stable complexes.

A rigorous suite of activity assays was not conducted, as numerous labs have previously conducted studies kinetic profiles of PRORP enzymes and our experiments confirmed enzymatic activity and were broadly consistent with the reported values.[29,30,141,145] These kinetic studies reveal that the proteinaceous RNase P surprisingly displays similar or worse efficiency than the ancient RNA enzyme *in vitro*, providing a biochemical rational for why the ribozyme has endured in most life. The enzymes recognize similar elements, best typified by the necessity of the T-domain of precursor tRNA for recognition by both enzymes, presenting a textbook case of convergent evolution by two unrelated enzymes. These insights allowed the generation of minimal and artificial pre-tRNA substrates in this study for crystal screens.

Numerous engineered RNAs were cloned that contain stabilizing motifs capable of forming tertiary intermolecular contacts. The addition of these elements to a minimal, structured RNA substrate did not inhibit PRORP3 binding. The resulting constructs hold potential for future structural studies of proteinaceous RNase P and other tRNA systems.

While cocrystallization of large RNA-protein complexes is technically challenging, a PRORP-tRNA crystal structure is likely in the coming years. Even with such a structure, a combination of computational and experimental techniques such as electron paramagnetic resonance will be required to fully understand the protein dynamics and conformational changes involved in binding and cleavage. Cryo-electron microscopy may be required to obtain a structure of the full human mitochondrial RNase P complex and mechanistically understand complex formation and substrate binding.

While much has been learned about proteinaceous RNase P in the decade since its discovery in humans, big picture questions remain. The variable reliance on protein or RNA-based RNase P in different systems and the consequences of the resulting tRNA/enzyme coevolution are still unanswered and must be addressed with biochemical and bioinformatic approaches.

2.5: Materials and Methods

Protein Purification

The *A. thaliana* PRORP3 gene in a pET15b plasmid was expressed in DE3 Gold *E. coli* cells grown overnight at 20 °C in TB media with 1 mM IPTG. Cells were harvested, resuspended in lysis buffer (150 mM NaCl, 50 mM HEPES pH 7.4, 1 mM PMSF), and lysate was loaded onto a sepharose FF 6 nickel column for IMAC purification. (Figure 11) Importantly, the protein was then washed on the resin with 1.5 M NaCl, 50 mM HEPES pH 7.4 buffer prior to elution to remove PRORP-associated RNAs which otherwise co-purify with the protein.

PRORP3 tended to precipitate out on DEAE or Superdex resin. Instead, after elution with 350 mM imidazole buffer, PRORP3 was purified on a phenyl sepharose column using a gradient from 1.5 M $(NH_4)_2SO_4$, 50 mM HEPES pH 7.4 binding buffer to 50 mM NaCl, 50 mM HEPES pH 7.4 elution buffer. (Figure 11) PRORP3 eluted beginning at 50% elution buffer.

A truncated PRORP3 construct lacking the first 85 amino acids, termed ΔN PRORP3, were cloned into a modified pSV281 vector with a Tev-cleavable N-terminal 6xHis tag and purified with the same protocol as full-length PRORP3. However, the Tev cleavage site proved inaccessible and could not be effectively removed by Tev protease. The U1A A1-98 Y31H/Q36R RBD plasmid was a generously gift from Dr. Adrian Ferre-D'Amare (NIH). U1A was expressed in DE3 Gold *E. coli* and purified as previously described.[146]

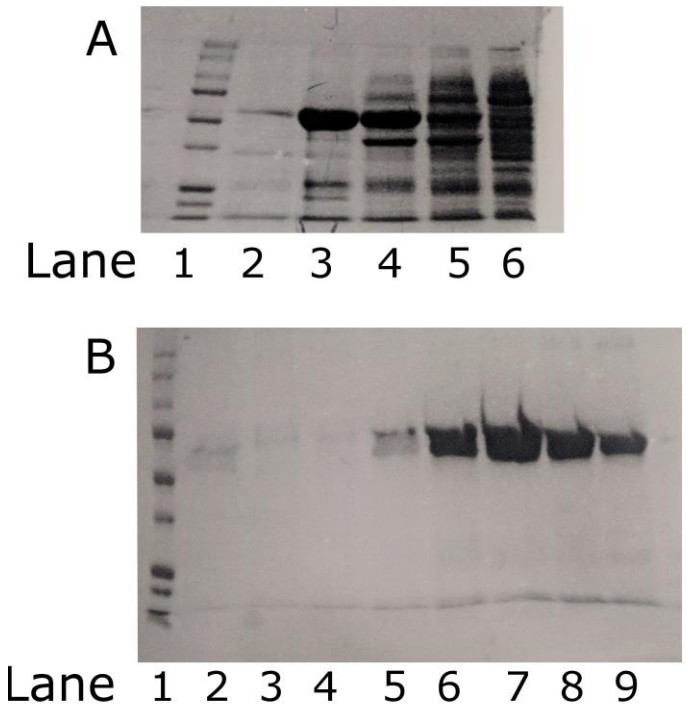


Figure 11: Representative PRORP3 Purification.

The PRORP3.D422N variant was purified as described in the materials and methods section.
(A) SDS-PAGE gel of fractions from nickel resin elution.
Lane 1: Precision Plus Ladder (Bio-Rad # #1610373)
Lane 2: 1 M imidazole elution Lane 3: 250 mM imidazole elution
Lane 4: 75 mM imidazole wash Lane 5: 40 mM imidazole wash
Lane 6: 15 mM imidazole wash.
The 350 mM imidazole fraction was collected and further purified as described below.

(B) Phenyl sepharose purification of 350 mM imidazole fraction from part A were diluted in 10 volumes of phenyl sepharose binding buffer (BB) containing 1.5 M $(NH_4)_2SO_4$, 50 mM HEPES pH 7.4, and eluted in a step-wise gradient with elution buffer (EB) of 50 mM NaCl, 50 mM HEPES pH 7.4. Fractions 7 and 8 were combined for biochemical assays and crystallography screens.
Lane 1: Precision Plus Protein Ladder Lane 2: Flow-through
Lane 3: 60% BB 40% EB Lane 4: 50% BB 50% EB
Lane 5: 40% BB 60% EB Lane 6: 30% BB 70% EB
Lane 7: 20% BB 40% EB Lane 8: 10% BB 90% EB
Lane 9: 100% EB

Enzymatic Activity and EMSA Assays

RNAs were 5’ radiolabeled using T4 Polynucleotide Kinase (NEB) and [γ-^{32}P] ATP (6000 Ci/mmol, 10 mCi/mL, PerkinElmer). BioGel P6 desalting spin columns were used to remove ATP and unincorporated radioactivity. Enzymatic cleavage assays were conducted at room temperature in 1xTHE, 10 mM $MgCl_2$, 200 mM KCl, 2 U/μL Promega RNase inhibitor. RNA substrate was heated to 50 °C for 10 minutes and then cooled to room temperature. Aliquots were taken during the reaction process at regular timepoints and quenched in 8M urea, 1xTHE buffer containing bromophenol blue and xylene cyanol dye. The samples were then loaded onto 8M urea, 1xTHE polyacrylamide gels for electrophoresis and imaged using an Imaging Screen-K (Kodak) phosphor imager. The screen was visualized with a Pharos FX Plus Molecular Imager (Bio-Rad) and analyzed with the Quantity One 4.6.9 (Bio-Rad) software.

EMSAs were prepared similar to the activity assays but with 10 mM calcium in place of magnesium. Samples were incubated for 10 minutes at room temperature and then put on ice before being run on 1xTHE, 0.9% agarose gel and analyzed in the same manner as the activity assays.

CHAPTER 3

Structure-Specific Recognition of G-Quadruplex RNA by LSD1

3.1 Introduction

Long noncoding RNAs (lncRNAs) are proposed to assist in a myriad of roles in the cell, acting as guides, scaffolds, decoys, or signaling molecules.[147] Although it is well established that distinct ncRNAs can act as gene regulators, recognize defined targets, and even function as catalysts (ribozymes), key mechanistic questions remain regarding how lncRNAs interact with and recruit chromatin-associated protein complexes to specific regions of the genome.[5]

LSD1 is an essential chromatin-remodeling enzyme conserved from yeast to humans and is also known to interact with lncRNAs.[46,67,112,148,149] A primary function of LSD1 is to influence gene expression and chromatin structure by catalyzing the removal of mono- and dimethyl functional groups from histone 3 proteins at lysine positions 4 and 9 (H3K4/K9).[67,150,151] LSD1 interacts with over 60 gene regulatory proteins, including transcription factors (CoREST, REST, p53, E2F1) and key enzymes (DNMT1, MRE11, HDAC1/2), as well as essential nutrients (tetrahydrofolate [THF]).[152–154] Of these, CoREST is the primary interacting partner required for post-translational LSD1 stabilization and is required for H3K4 demethylation during development, hematopoiesis, and stem cell maintenance.[67,155,156]

While distinct LSD1-containing protein complexes repress or activate gene transcription, it is unknown how lncRNAs bind and modulate these complexes.[67,151] The telomeric repeat containing RNA (TERRA) is an integral component of telomeric heterochromatin, acts as a negative regulator of telomere length in human cells, and interacts with critical epigenetic regulators that include LSD1 demethylase and SUV39H1 methyltransferase enzymes.[46,157–160] In addition, TERRA has a strong propensity to form intramolecular, parallel-stranded G-quadruplex (GQ) RNA structures due to its repeating UUAGGG sequence.[160]

Like many lncRNAs, TERRA remains associated with its parental chromatin, allowing cotranscriptional modulation of gene expression by epigenetic regulators. Telomeric gene silencing appears to correlate with methylation and demethylation patterns of H3K4/K9 histone modifications and it has been established that the TERRA–LSD1 interaction enhances the telomeric DNA damage response pathway.[46,158] Upon depletion of a shelterin component, the telomeric repeat factor 2 (TRF2), global TERRA levels increase in the cell and TERRA interacts directly with LSD1. Through an unknown mechanism, this RNA–LSD1 interaction subsequently stimulates the nuclease activity of the double strand break repair protein MRE11 to trim the 3′ G overhangs at uncapped telomeres (Figure 12).[46] MRE11 is a 3′-to-5′ exonuclease and a component of the dsDNA damage response-associated MRE11/RAD50/NBS1 (MRN) complex. The precise role of the complex at telomeres is unclear but it has been associated with both telomerase maintenance through promotion of

telomerase activity and the removal of the 3' telomere overhang upon telomere disruption.[161]

Previous studies have found LSD1 localized at sites of dsDNA breaks and necessary for proper DNA damage response and survival.[162,163] According to the current model, the ATM response leads to LSD1 phosphorylation by the protein kinase CK2 and the phosphorylated LSD1 is then recruited to sites of dsDNA breaks by the ubiquitin-protein ligase RNF168 during the ATM-mediated DNA damage response.[162,164] It then not only demethylates H3K4me2 but promotes the recruitment and ubiquitination of the critical response factor p53 binding protein-1 (53BP1).[162,163] Therefore, it is not completely unexpected that recruitment of LSD1 to telomeres may result in inappropriate activation of the DNA damage response at the ends of chromosomes.

These interactions suggest that: (i) RNA binding to LSD1 is important at dysfunctional telomeres, and that (ii) TERRA may scaffold chromatin modifying enzyme complexes in a manner that is similar to other chromatin-associated lncRNAs such as HOTAIR and Xist.[46,110,165,166] Taken together, these data support a role for TERRA in recruiting proteins to modulate heterochromatin formation at chromosome ends. The following chapter demonstrates that LSD1/CoREST has a strong preference to bind a stacked GQ-forming RNA and reveals that the primary binding site of the GQ RNA exists within the SWIRM/amine-oxidase domain of LSD1. Binding and cross-linking mass spectrometry (XL-MS) data collectively indicate that structured RNAs can

function as important regulators in LSD1-mediated pathways and support an emerging theme that structure-specific RNA binding can influence the function of chromatin-associated proteins.

3.2: TERRA Forms a Stacked G-Quadruplex in K^+

Quadruplex-forming nucleic acid sequences require specific monovalent ions for structural stability. In particular, potassium ions stabilize GQs, while lithium ions destabilize GQs.[167] Circular dichroism (CD) spectroscopy was used to monitor the effect of monovalent ions on GQ formation. CD spectra with a peak at 263 nm and a trough at 240 nm are characteristic of a parallel, propeller-type GQ conformation.[167–169] In both four and eight UUAGGG repeats (in the presence of K^+ and Na^+), we observe the formation of stable parallel-stranded GQ structures (Figure 13). In contrast, Li^+ destabilizes GQ structures and likely promotes a heterogeneous RNA architecture.

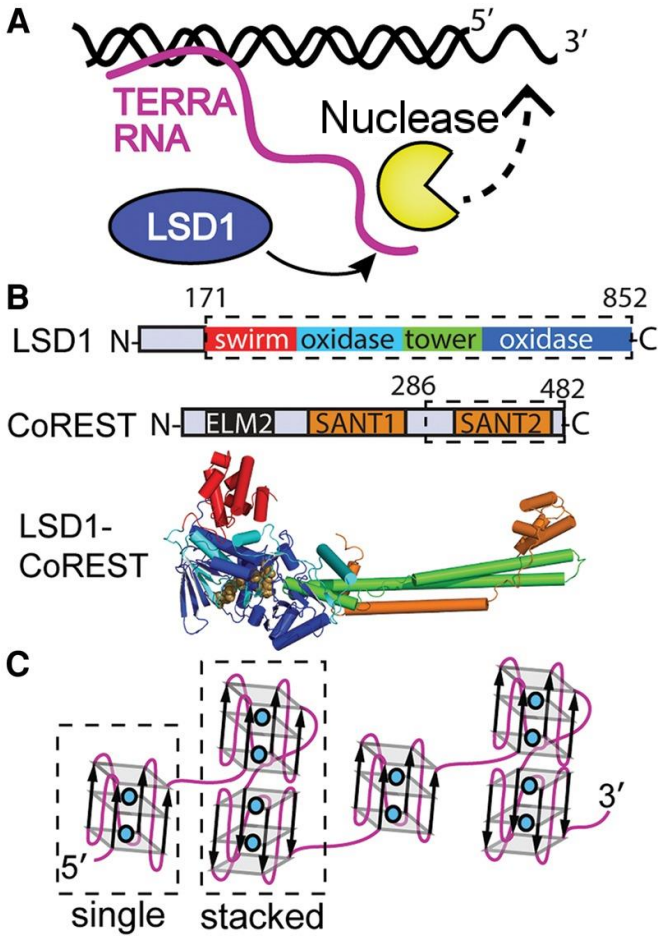


Figure 12: TERRA RNA recruits LSD1 to deprotected telomeres.
(A) A functional role for the TERRA RNA in the processing of uncapped telomeres, as previously reported.[46] TERRA can serve as a scaffold for LSD1–nuclease associations. (B) Protein constructs used in this study include LSD1 (aa 171–852) and CoREST (aa 286–482 plus 6xHis-tag sequence). LSD1 consists of a SWIRM domain (red), an intertwined monoamine oxidase domain (cyan/blue), and a tower domain (green), based on PDB 2IW5.[92] LSD1 biological function requires the presence of CoREST (shown in orange). (C) A GQ RNA is stabilized by specific monovalent ions (including K^+ and Na^+ denoted as blue spheres). A single GQ RNA unit and a stacked GQ RNA were prepared to investigate binding affinity and specificity of the GQ RNA–LSD1 interaction. The UUAGGG repeat elements of TERRA can form a stable parallel-stranded GQ RNA *in vivo* and a model of the higher order TERRA RNA architecture has been biochemically demonstrated.[169,170]

Sedimentation velocity analytical ultracentrifugation (SV-AUC) was also performed to assess RNA tertiary folding in solution, GQ oligomerization, and global compaction in the presence of different monovalent ions. Analysis of SV-AUC data reveals that a four-repeat UUAGGG RNA has completely different tertiary shapes in the presence of K^+, Na^+, or Li^+. (Figure 13) Whereas a four-repeat UUAGGG RNA forms a monomeric GQ structure in Na^+, the identical RNA forms a heterogeneous profile in the presence of K^+. Further, SV-AUC analysis of the four-repeat RNA with K^+ reveals a sedimentation profile that is positioned between the four-repeat RNA (Na^+) and an eight-repeat UUAGGG RNA (K^+).

Bayesian analysis suggests that the four repeat UUAGGG RNA with K^+ exists in equilibrium between monomeric and dimeric, stacked GQ structures. This result is consistent with previous nuclease digestion studies of TERRA.[168,169] Thus, choice of cation and RNA construct enables us to manipulate the tertiary RNA structure and examine the nucleic acid binding properties of LSD1.

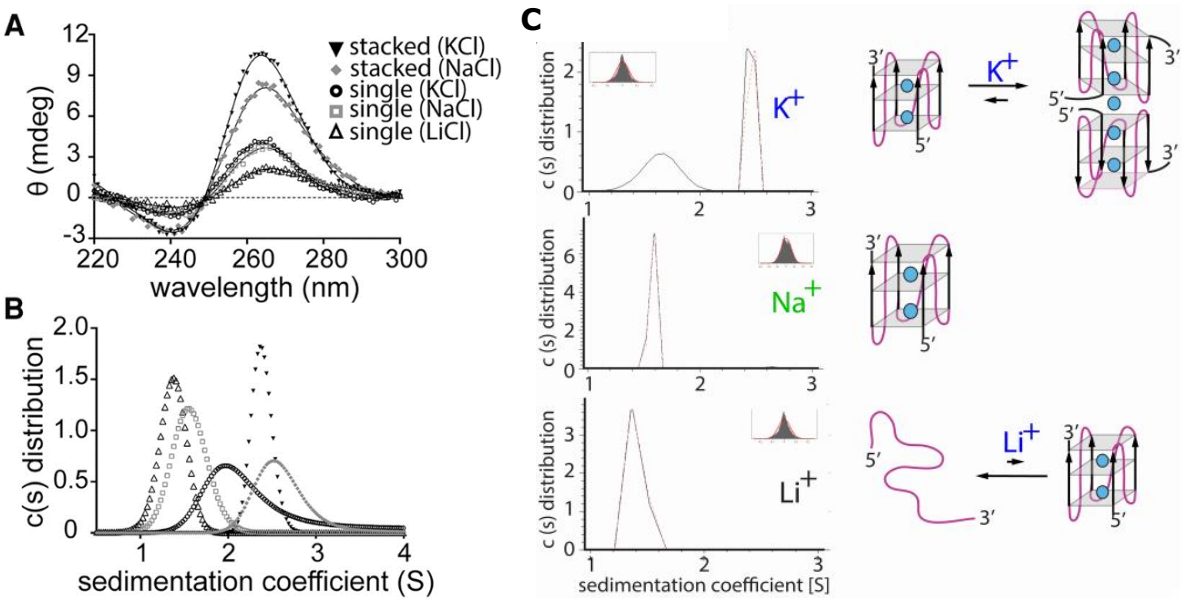


Figure 13: Monovalent ions dramatically alter the structure of GQ-forming RNAs.

(A) Parallel-stranded GQ RNAs are known to have an Θ_{max} of ~263 nm.[167] Circular dichroism spectroscopy demonstrates that parallel-stranded GQ structures form in the presence of potassium (black, solid triangle, and circles) and sodium (gray diamond and box), consistent with previous studies. In contrast, lithium (outlined triangles) destabilizes GQ formation. (*B*) The analysis of analytical ultracentrifugation (AUC) data of $(UUAGGG)_4U$ and $(UUAGGG)_8U$ RNAs in the presence of potassium (K^+), sodium (Na^+), and lithium (Li^+). Figure symbols as in *A*. The plot shows an overlay of the continuous distribution [C(s)] versus the sedimentation distribution coefficient (S). (C) The analysis of deconvoluted analytical ultracentrifugation (AUC) data of the 5′-GG[UUAGGG]4UUA-3′ RNA in the presence of potassium (top panel), sodium (middle panel), and lithium (bottom panel). Each plot shows the continuous distribution (C(s)) versus the sedimentation distribution coefficient (S) and the observed peaks indicate a single quadruplex, a stacked quadruplex, or a broad heterogeneous profile in the case of Li^+. The residual distribution histogram for each plot is shown in the upper left or right corners.

3.3: LSD1 Recognizes a Stacked G-Quadruplex TERRA

In an effort to understand the nucleotide sequence and structural preferences associated with the LSD1/CoREST complex, we measured the LSD1–nucleic acid affinities from a panel of single and stacked GQ-forming oligonucleotides under different monovalent ions. (Figure 14) A TERRA RNA that contains a stacked ($GG[UUAGGG]_8UUA$) RNA, a single G-quadruplex-forming repeat element ($[UUAGGG]_4U$), a cognate DNA ($[TTAGGG]_4T$), and a sequence-unrelated 25-nt RNA (ssRNA) for which there is no predicted structure were incubated with enzymatically active LSD1. Because both LSD1 and CoREST localize at telomeres and LSD1 is optimally stabilized in the presence of CoREST, we have primarily focused on how TERRA interacts with the LSD1/CoREST system.[50,67,155,171]

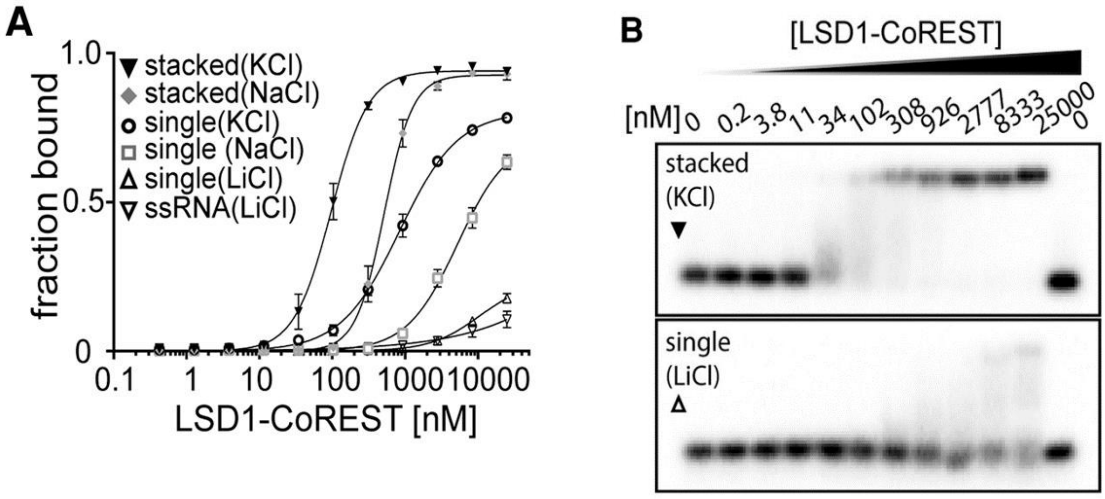


Figure 14: Affinity and specificity of LSD1/CoREST for GQ RNA.

LSD1/CoREST binds to distinct RNA structures as controlled by monovalent ions. (A) Analysis of gel-mobility shift assay binding curves of $(UUAGGG)_8U$, $(UUAGGG)_4U$, and 25-nt ssRNA. Assays were performed in potassium (K^+), sodium (Na^+), and lithium (Li^+) (symbols same as in Figure 13) using LSD1/CoREST (amino acid residues 171–852 and 286–482, respectively) with exogenous protein purification tags removed. LSD1/CoREST strongly prefers to bind stacked GQ-forming RNA structures. The plot shows the fraction of RNA bound at various LSD1/CoREST concentrations (log scale). Error bars for each data point represent the range of three independent experiments.
(B) Representative gels showing that RNA binding activity of LSD1/CoREST is dependent upon the ability of the RNA to form into a GQ conformation. Complexes and free oligonucleotides were resolved on a 0.6% native agarose gel. The concentration of the LSD1/CoREST complex is noted for each lane (nM).

Table 2: Affinity of LSD1/CoREST for GQ RNA Panel.

RNA (monovalent)	**Dissociation Constant (K_d) (nM)**	**Hill Coefficient** *(h)*
5'-GG$(UUAGGG)_8$U-3' (K^+)	96.4±3	1.70±0.08
5'-GG$(UUAGGG)_8$U-3' (Na^+)	516±15	2.24±0.1
5'-$(UUAGGG)_4$U-3' (K^+)	835±35	1.07±0.04
5'-$(UUAGGG)_4$U-3' (Na^+)	5200 ±460	1.25±0.08
5'-$(UUAGGG)_4$U-3' (Li^+)	10,500±1900	1.3±0.1
25-nt ssRNA (Li^+)	N/A	N/A

The mean and standard error from three independent EMSAs is given.

Using electrophoretic mobility shift assays (EMSAs), incubation of these oligonucleotides with LSD1/CoREST revealed vastly different binding affinities. (Table 2) Analysis of the percent nucleic acid-bound fraction demonstrates that an eight-repeat UUAGGG RNA that forms a stacked GQ RNA structure tightly binds the LSD1/CoREST complex (apparent K_d = 96.4 nM), whereas a four repeat UUAGGG RNA and a four-repeat TTAGGG DNA bind with approximately nine- and 14-fold weaker affinities (the apparent K_d values for each oligonucleotide in K^+ are 835 nM and 1.3 μM, respectively). The four-repeat RNA exhibited negligible binding, indicating that the LSD1/CoREST specifically binds folded TERRA. (Figure 14) Mutant, non-GQ TERRA sequences also failed to bind LSD1/CoREST, demonstrating that the loss of binding is GQ-dependent and not purely an effect of the lithium. The Hill coefficient for an eight-repeat UUAGGG RNA–protein interaction ($h \approx$ 1.7 ± 0.08 and 2.24 ± 0.1 in K^+ and Na^+, respectively) indicates substantial positive cooperativity, implying that there may be an extended RNA binding interface on the LSD1/CoREST complex. In contrast, the $([UUAGGG]_4)U$ single GQ-forming RNA does not exhibit cooperativity ($h \approx$ 1.07 ± 0.04) upon binding to LSD1/CoREST, implying that single GQ-forming RNAs may bind the LSD1/CoREST complex as preformed units.

An EMSA analysis of individual LSD1 and CoREST components binding to TERRA confirms a strong LSD1–TERRA interaction (apparent K_d = 404 nM, h = 0.91 ± 0.17) and an interaction with the CoREST fragment that is approximately 10-fold weaker (apparent K_d = 3.4 μM, h = 1.40 ± 0.15).

(Figure 15) This suggests that the high-affinity binding site for RNA resides within LSD1. These EMSA results are consistent with the previously identified RNA–LSD1 binding equilibrium studies, where a GST–LSD1 fragment (aa 1–382), containing the SWIRM domain and a region of the catalytic amine oxidase domain (AOD), were shown to bind a 10-repeat UUAGGG RNA with high affinity (K_d~70 nM).[46]

3.4: An RNA—Binding Domain in the SWIRM Domain of LSD1

Previous studies of fragmented LSD1 constructs concluded that the GQ RNA likely associates with the SWIRM/AOD of LSD1.[46] However, we were unable to replicate binding with the truncated and fragmented domain in our hands, and sought to gain a better resolution picture of the RNA binding site. *In vitro,* 254-nm UV light cross-linked G-quadruplex TERRA RNA, but not a size-matched, non-GQ RNA, to form a covalent complex with LSD1 and CoREST. (Figure 16) To more precisely identify the RNA binding regions of LSD1, a biotinylated GQ RNA (GG[UUAGGG]$_8$UUA) was covalently cross-linked to purified LSD1/CoREST complex and subjected to high-resolution LC–MS/MS mass spectrometry. Two separate and independent XL-MS experiments were performed using purified LSD1/CoREST with and without a 6x-His tag at the N terminus of CoREST. Consistent with previous studies, analyses of both XL-MS data sets indicate a strong GQ RNA cross-link to residues 227–251 of the SWIRM domain (Figure 17).

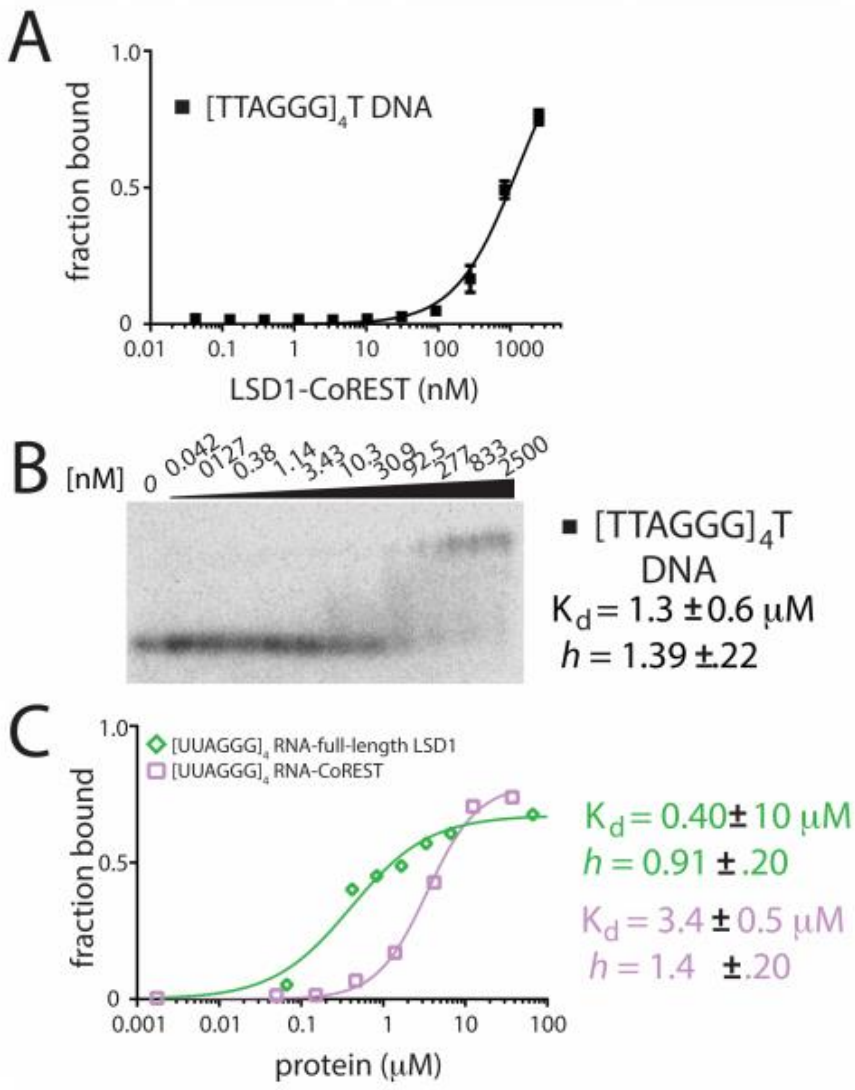


Figure 15: Nucleic acid binding specificity of LSD1.

(A) Affinity of LSD1- CoREST binding to a 25 nt GQ forming [(TTAGGG)4T] DNA (box). The plot shows the concentration of the LSD1/CoREST complex versus fraction of RNA bound. Binding studies were assayed in triplicate and representative data are shown in B. (B) Nucleic acid binding activity of LSD1/CoREST (amino acid residues 171-852 and 286-482, respectively). Complexes and free oligonucleotides were resolved on a 0.75% native agarose gel. The concentration of the LSD1/CoREST complex is noted for each lane (nM) and the dissociation constant (Kd) and Hill coefficient (h) are shown. (C) Independent LSD1 and CoREST binding to [(UUAGGG)4U] RNA. Analysis of gel-mobility shift assay binding curves of a 25 nt. GQ forming RNA (K^+) with full-length LSD1 only (green), or with CoREST (286-482) (pink). The plot shows the protein concentration of LSD1 or CoREST versus fraction bound. Binding dissociation constant (Kd) and Hill coefficient (h) are shown. Figure 15 EMSAs were performed with a 6x His tag located at the C-terminus of CoREST which may have resulted in some nonspecific binding. Fraction bound for LSD1 alone saturates at approximately 60-65%, which may be due to GQ RNA heterogeneity or instability of the LSD1- GQ RNA complex at high micromolar concentrations.

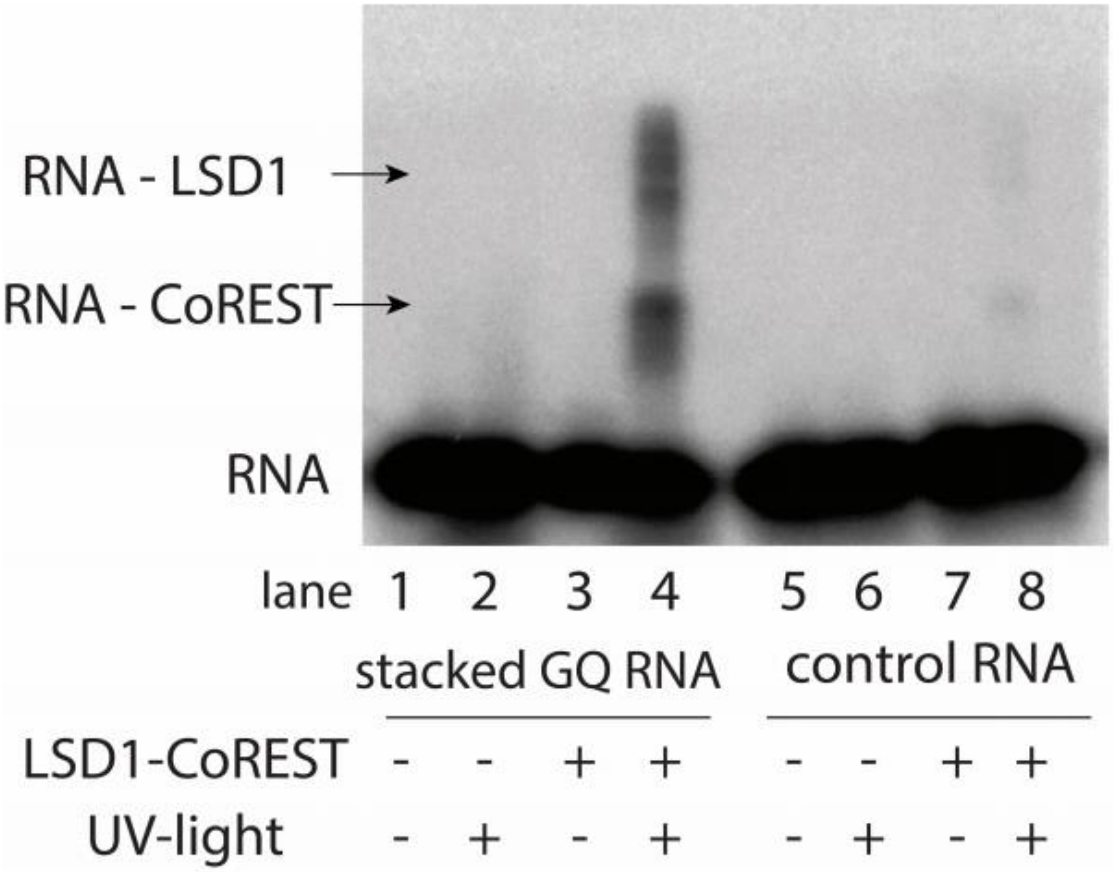


Figure 16: UV Light specifically cross-links LSD1/CoREST to a GQ RNA.

A 5′ ^{32}P-labeled 5′-GG[UUAGGG]8UUA-3′ RNA and a size matched, non-GQ forming RNA were incubated with or without LSD1/CoREST. All RNA concentrations were equivalent. Upon cross-linking at 254 nm, the stacked GQ RNA co-migrates with both LSD1 and CoREST (Lane 4). The higher molecular weight species indicate a degree of oligomerization. Very faint bands can be observed for the equivalent control RNA (Lane 8), likely the result of low levels of non-specific RNA binding. Samples were then run on a 4-20% gradient SDS-PAGE gel and visualized with a phosphor screen.

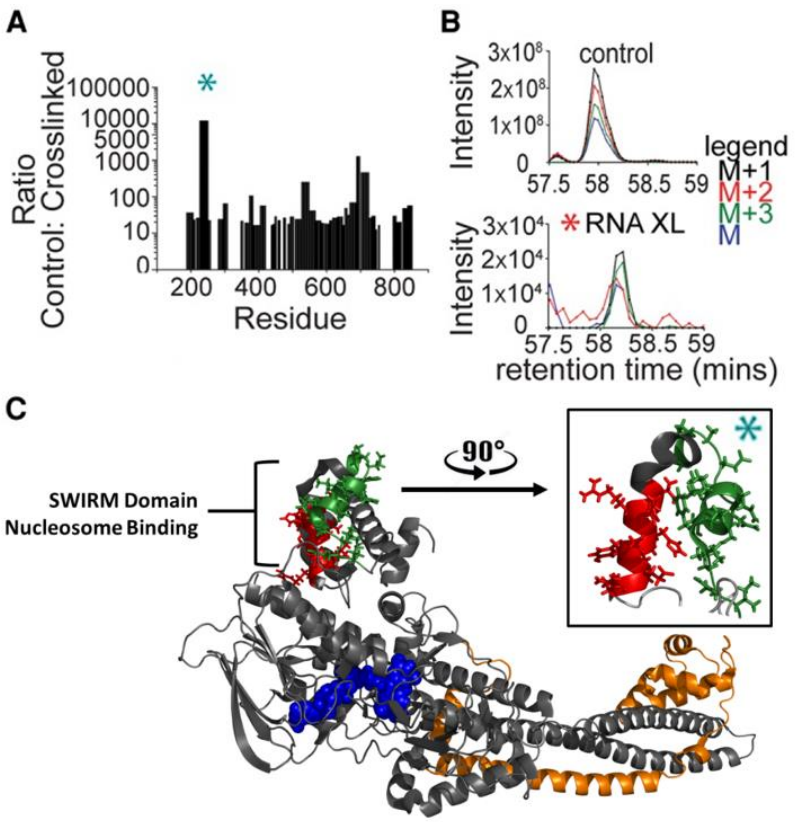


Figure 17: Identification of the LSD1 GQ RNA binding domain via mass spec.

(A) Relative coverage of LSD1 residues upon GQ RNA cross-linking. UV light was used to cross-link biotinylated GQ RNA with LSD1/CoREST and the covalent complex was purified using streptavidin beads. LSD1 was then analyzed with mass spectroscopy alongside a control sample that had been treated with UV light in the absence of RNA. A plot of the signal intensity ratio between the control and GQ RNA cross-linked sample reveals peptide fragments that are strongly depleted in the cross-linked sample, likely due to the change in m/z ratio upon formation of RNA adduct. (B) Analysis of the elution profile reveals a depletion of the peptide signal upon cross-linking for a region in the SWIRM domain (227–251). This peptide is dramatically depleted (~10,000-fold weaker) compared with the control sample. The isotopic distribution (M, M + 1, M + 2, M + 3) confirms the peptide identity, with an isotope dot product (idotp) of 1.00 and 0.97 for the control and cross-linked samples, respectively. (C) The locations of the putative RNA binding regions in the SWIRM domain are mapped onto the structure of LSD1/CoREST (PDB 4XBF). The adjacent *in vitro,* UV_{254} cross-linked peptide is shown in green and denoted by an asterisk, and the 4SU, UV_{365}-cross-linking peptide from the RBR-ID database is highlighted in red. (residues 227–242 and 252-258, respectively) Orange: CoREST; Blue: FAD.

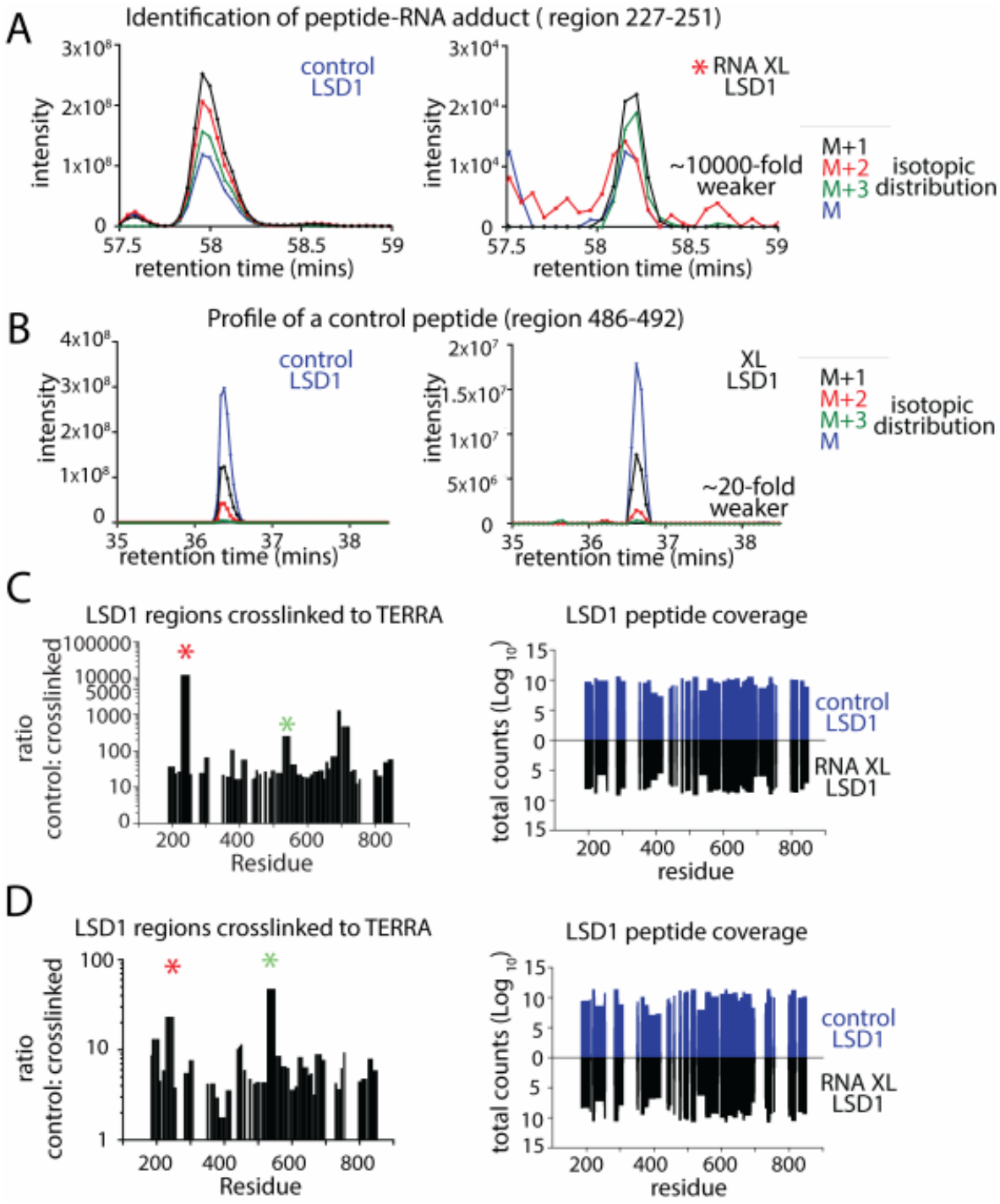


Figure 18: Identification of candidate GQ binding regions on LSD1 via XL-MS.

Samples were prepared as described (Hirschi et al., *RNA*, 2016).[172] (A) The LSD1 peptide fragment (227-251) from the GQ RNA cross-linked sample was ~10,000 fold weaker than in the control, indicating that the mass of this peptide has been altered due to RNA adduction. (B) Representative analysis of an LSD1 peptide fragment (486-492) that does not cross-link to RNA. The change in intensity between the control LSD1 sample and the LSD1-RNA cross-linked sample is modest (20-fold weaker) and

is equivalent to the difference in total mass injected. For A and B, the m/z ratio and calculated isotope dot product values confirm the identity of the peak based on the expected isotopic distribution (M, M+1, M+2, M+3). Peptide analysis and coverage for separate cross-linking-MS experiments (C, D). (C) Plot of the ratio of control to cross-linked sample by LSD1 amino acid residue. Three peptide regions are depleted in the RNA-LSD1 cross-linked sample (LSD1 regions (227-251), (527-550), and (695-720). The significant change in the m/z ratio and elution profiles of the RNA-bound peptide result in the depletion of peptide signal. A plot of the total abundance (Log10) of the peptides in each sample reveals relatively consistent coverage of the control LSD1 (blue bar) and the RNA cross-linked LSD1 (black). This LSD1/CoREST sample contains a 6x His-tag located the C-terminus of CoREST and was cross-linked with a single GQ forming 5'-biotinylated RNA ($GUUU(UUAGGG)_4UUA$). (D) A separate RNA-LSD1 cross-linking MS experiment reveals two peptide regions that are depleted in the RNA-LSD1 cross-linked sample. These two regions were previously observed in C and represent LSD1 regions (227-251) and (527-550). In contrast to the sample analyzed in C, this LSD1/CoREST sample does not contain a 6x His-tag located the C-terminus of CoREST and used a stacked GQ-forming 5'-biotinylated RNA (5'-biotinylated $GGG(UUAGGG)_8UUA$). Analysis of C and D were performed in an identical manner.

As a specificity control, we observed that LSD1/CoREST covalently cross-links with GQ RNA but not a size-matched control RNA (Figure 16). Here, LSD1/CoREST was irradiated with 254 nm UV light either alone (control), in the presence of a 5' ^{32}P-labeled GQ RNA, or in the presence of a size matched 5' ^{32}P-labeled non-GQ-forming RNA. Protein and RNA concentrations were identical for each experiment. These results demonstrate that UV light specifically cross-links LSD1/CoREST to a stacked GQ RNA, consistent with EMSA data in Figure 14.

For XL-MS studies, two samples, the LSD1/CoREST complex alone (control) and in the presence of a 5'-biotinylated GQ RNA, were subjected to 120 mJ of 254 nm UV light (Stratalinker 1800). The cross-linked RNP complex was then isolated using streptavidin beads and both the complex and the irradiated LSD1/CoREST (control) samples were SDS-PAGE purified, in-gel trypsin digested, and analyzed by mass spectrometry. The cross-linked peptides of

LSD1 were deduced by the disappearance of the original peptide signal as the RNA–LSD1 adduct altered the mass-to-charge (*m/z)* ratio. Due to the stability and nuclease-resistant nature of the GQ RNA, we were unable to degrade the RNA into a homogeneous fragment that could be identified on a peptide by mass spectrometry.

Our analysis only included peptide regions that could be detected by mass spectroscopy, with coverage limited to ~75%–80% of the truncated LSD1 amino acids (171–852). We observed two distinct regions (227–251 and 527–550) that are strongly depleted in the RNA-containing sample, indicating the formation of a cross-link with GQ RNA. Amino acids 227–251 comprise a helix–loop–helix region of the conserved SWIRM domain and likely represent a primary GQ RNA binding interface. Residues 527–550 are located adjacent to the N terminus of CoREST and near the LSD1 active site and may represent a secondary nucleic acid binding site (Figure 18 C, D). The XL-MS experiment was replicated using modified conditions that include a stacked GQ RNA and a CoREST construct with a cleaved N-terminal 6x-His tag. In both XL-MS experiments, depletion of the identical peptide regions spanning residues 227–251 and 527–550 were observed. A third cross-linked region (residues 689–726) was identified on the tower domain in one experiment; however, this region had very low coverage in the control and was not reproducible.

Validation in Cell Culture of XL-MS Identified RNA-LSD1 Interaction

A subsequent study from Dr. Roberto Bonasio's group used a high-throughput approach to identify RNA-binding peptides.[173] Mouse embryonic stem cells were treated with or without 4-thiouridine, cross-linked with 365 nm UV light, and peptides were analyzed with mass spectroscopy. Peptides depleted by 4-thiouridine cross-linking were then computationally identified as likely RNA binding sites with their custom metric, the RBR-ID score. The results were summarized on an online database (http://rbrid.bonasiolab.org/rbrid/). Previously unknown RNA binding proteins were especially enriched with proteins associated with chromatin binding and chromatin organization, implying that RNA may play a much larger role in the regulation of chromatin associated proteins than is currently understood. The top peptide hit in LSD1 is located adjacent to the TERRA GQ-cross-linking candidate peptide, providing validation of the region as a true RNA binding region. (Figure 17C) Thus, the novel SWIRM-domain RNA binding region is not only involved in the binding of G-quadruplexes *in vitro* but also interacts with proteins in cells. This corroboration also suggests that 4-thiouridine can be used to cross-link LSD1 with associated RNAs in cell culture, a technique which will be explored in the following chapter.

3.5: Discussion

TERRA is an integral component of telomeric heterochromatin which forms distinct G-quadruplex structures *in vitro* and *in vivo,* and has been

demonstrated to interact with critical epigenetic regulators that include LSD1 demethylase and SUV39H1 methyltransferase enzymes.[46,157,158,170,174–178] Despite a correlation between histone H3K4/K9 methylation patterns and gene silencing at telomeres, it is unclear how these enzymes are recruited to telomeres and how they might function to mediate telomere structure and composition.[46,157,179,180] TERRA is thought to act as a molecular decoy to sequester telomerase in a cell cycle-dependent manner and also likely scaffolds with epigenetic regulators during the early stages of DNA damage response activation.[46,87,157,180,181] The unique topology of TERRA suggests that its structure may enable the organization of higher-order RNP complexes at telomeres. To better understand how lncRNAs may recruit enzymes and how TERRA might regulate proteins at the ends of chromosomes, we determined the mode of GQ RNA binding to LSD1.

Monovalent ions dramatically influence the topology of GQ RNAs and knowledge of specific TERRA sub-structures enabled us to manipulate the GQ RNA architecture (Figure 13). The formation of these diverse RNA structures correlates well with observed nucleic acid–LSD1 binding affinities (Figure 14), revealing that LSD1 strongly prefers a stacked GQ (GG-$[UUAGGG]_8$UUA) RNA. Previous TERRA RNA studies combining RNase T1 digestion and molecular dynamics (MD) simulations demonstrate that TERRA (containing up to 96 UUAGGG repeats) primarily assembles into four- and eight-repeat UUAGGG units.[168,169] It appears that a single ($[UUAGGG]_4$U) GQ unit, a stacked ($[UUAGGG]_8$U) GQ unit, and an ssRNA spacer region between GQ units

represent the three unique topologies adopted by TERRA. In addition, biophysical studies of TERRA demonstrate that the 2′–OH functional groups in the RNA G-quadruplex participate in organizing hydration and in the hydrogen-bonding network.[168,182] This may contribute additional stability to the parallel-stranded quadruplex conformation and account for why LSD1 prefers to bind GQ RNA over a cognate GQ DNA (Figure 15), which is known to contain a mixture of heterogeneous G-quadruplex topologies.[183] Our data confirm that single and stacked GQ RNA structural units serve as building blocks of the extended TERRA lncRNA and that the propensity for stacking in GQ RNA molecules may be a key feature in the recruitment of protein–protein complexes at telomeres.

For the recognition of TERRA by LSD1, the RNA structure and shape play a central role in protein–RNA binding specificity. Although GQ RNA shape-based recognition is not well established, it is known that other protein domains can also preferentially recognize GQ-forming RNAs, including the arginine–glycine–glycine repeat (RGG) domain of Fragile-X mental retardation protein (FMRP), and the N-terminal domain of the DEAH-box ATP-dependent helicase 36 (DHX36).[64,184,185] The PRC2 histone methyltransferase complex was also recently shown to specifically bind GQ RNAs, although the binding domain is unknown.[186]

With the development of GQ-specific antibodies that have unambiguously identified GQ TERRA RNA structures in living cells, more biophysical studies are

needed to probe how protein domains and multidomain protein complexes recognize the molecular architecture of GQ RNAs.[54,170,187] Here, the coupling of biophysical studies (CD/AUC), EMSAs, and XL-MS methods has better defined how the secondary and tertiary structural motifs of TERRA can serve as recognition elements for LSD1 interactions.

Implications of RNA Binding in LSD1 Functions

LSD1 is associated with CoREST and together they act as an allosteric clamp on nucleosomes to catalyze specific histone H3K4/K9 demethylation.[67,92,148,150,188,189] Motions of the SWIRM domain and rotation of the AOD of LSD1 must occur when the substrate enters the active site pocket and it has been demonstrated that both substrate binding and protein–protein interactions modulate LSD1 conformational dynamics and activity.[93,150,189] Interestingly, the structural integrity and flexibility of the SWIRM-AOD is essential for recognition of nucleosomal DNA.[92,148,150] GQ RNA binding at the SWIRM domain may therefore alter LSD1 mechanics.

Whereas the biological consequences of LSD1 activity are known, examples of RNA-mediated LSD1 recruitment and the modulation of LSD1 activity by lncRNAs lack in mechanistic detail.[46,110,112,113] The regulation of LSD1 complexes is likely to be multilayered and may be differentially influenced by distinct classes of RNA molecules. For example, RNA helicases are known to actively remodel TERRA RNA at telomeres.[159,160,190] It is possible that RNA

helicases provide an added level of epigenetic regulation by unwinding GQ RNAs, resulting in release of LSD1 binding. Such a molecular mechanism has yet to be demonstrated, though it should be noted that several groups have proposed a link between helicases and the regulation of the GQ-specific PRC2 chromatin modifying complex by ncRNAs.[191–195]

In conclusion, this work identified a primary GQ RNA binding site within the conserved SWIRM/AOD interface of LSD1. Future structural studies will be required to demonstrate how a GQ-forming RNA alters LSD1 structure and how TERRA influences the function of LSD1 at telomeres. Defining the structural interactions of TERRA with LSD1 will provide insight into the diversity of GQ RNA–protein recognition and serves as an important model system to explore lncRNA–protein recruitment mechanisms.

3.6: Materials and Methods

Protein/RNA Design, Expression, and Purification

The plasmids for the N-terminal truncated LSD1 (aa 171–852) and CoREST (aa 286–482 plus His-tag sequence) were a generous gift of Dr. Cole (Johns Hopkins University) and the full-size LSD1 plasmid (aa 1–852) was a generous gift of Dr. Shi (Harvard University).[67,196] (Figure 19) The full size and truncated LSD1 were expressed in *Escherichia coli* BL21(DE3) as previously reported.[67,196,197] The full-size LSD1 was purified by using ammonium sulfate fractionation and anion-exchange chromatography.[154] The truncated LSD1 was purified using GSH-agarose chromatography, with the glutathione S

transferase tag removed through PreScision protease (GE Healthcare) digestion and anion-exchange chromatography.

CoREST is a known stabilizing element of LSD1 and was expressed in *E. coli* BL21(DE3) in Luria Broth (LB) (kanamycin) and purified using Ni-NTA agarose as previously reported.[93,154] For EMSA assays, truncated LSD1 (171–852) was coexpressed with CoREST (286–482) in Rosetta (DE3) pLysS competent cells. Complexes were purified via glutathione and nickel affinity and size-exclusion (Superdex 200) chromatography. The concentration of protein samples was determined by the BCA method (BCA Protein Assay kit, Pierce) with bovine serum albumin as a standard and by UV-VIS-spectroscopy for LSD1 preparation using an extinction coefficient for FAD at 450 nm as 11300 $M^{-1}cm^{-1}$. Protein spectra were recorded on Shimadzu 2401-PC Spectrophotometer. Protein purity was determined by SDS electrophoresis with Coomassie staining (Figure 19).

Stacked GQ RNAs ($GG[UUAGGG]_8UUA$) were transcribed with T7 RNA polymerase and purified over 6% TBE-polyacrylamide gels supplemented with 8 M urea according to standard methods. Dried RNA pellets were resuspended in RNase-free water and diluted to 5 µM in 10 mM Tris–HCl or HEPES-KOH pH 7.4, 1 mM TCEP and 100 mM KCl, NaCl, or LiCl as indicated in the main text. Alternatively, a monomeric GQ-forming RNA ($[UUAGGG]_4U$), a 25-nt single-stranded RNA (5’-UUAGCGUUAACCUUACCAUUACGUU-3’) (ssRNA) were purchased from Thermo-Fisher (Dharmacon) and deprotected according to

manufacturer's instructions. The identical DNA ($[TTAGGG]_4T$) (GQ DNA) and the 5′ biotinylated RNAs for cross-linking ($GUUU[UUAGGG]_4UUA$ and $GGG[UUAGGG]_8UUA$) were purchased as synthetic oligonucleotides (Integrated DNA Technologies, Inc.). All oligonucleotides were ethanol precipitated, desalted on a 20-mL G-25 gel filtration column, concentrated using a centrifuge vacuum manifold, and stored at −20°C. Purity was assessed by denaturing PAGE (8 M Urea) and all solutions were suspended in a 10 mM KPO_4 (pH 6.5), 100 mM KCl buffer prior to use.

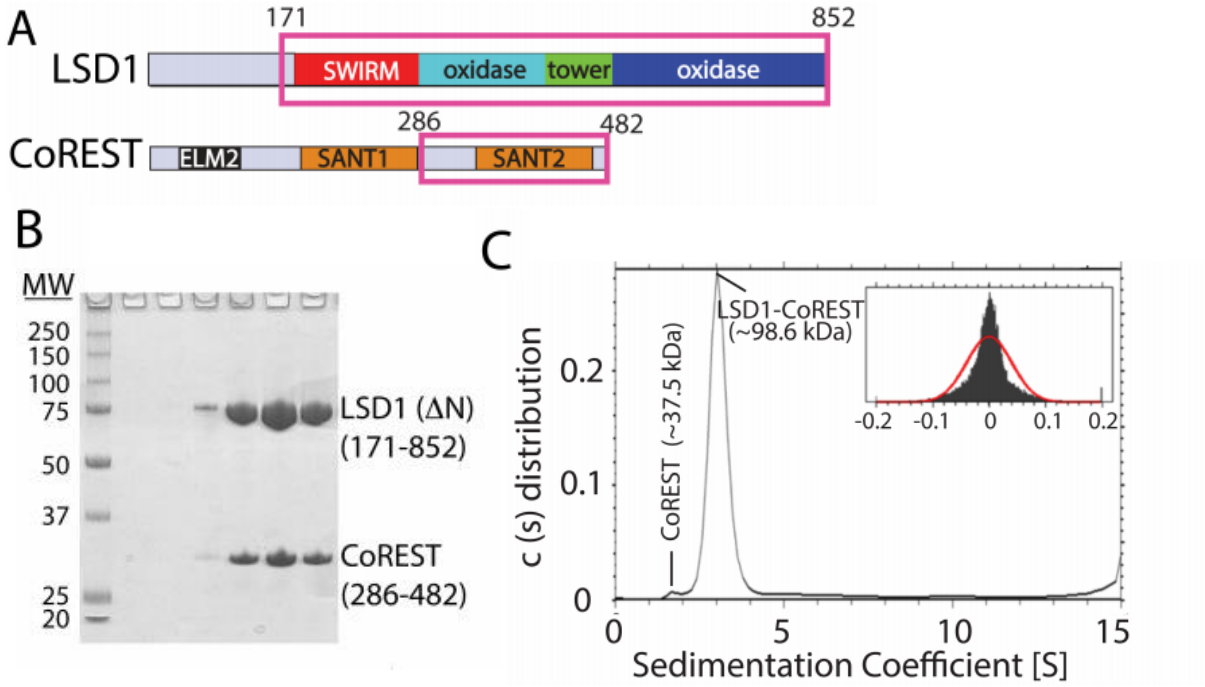


Figure 19: Purity of LSD1/CoREST constructs prepared for this study. (A) Schematic of LSD1 (top) and CoREST (bottom). LSD1 contains a SWIRM domain, an intertwined oxidase domain (cyan, blue) that performs the lysine specific demethylase catalytic reaction, and a tower domain (green) that forms a key coiled-coil interaction with CoREST. CoREST contains an ELM2 domain and two SANT domains. All EMSA and cross-linking studies were performed with truncated LSD1/CoREST (pink boxes). (B) SDS-PAGE gel of the purified LSD1/CoREST complex using size exclusion chromatography (Superdex 200 10/300 GL). (C) The analysis of analytical ultracentrifugation (AUC) data of the LSD1/CoREST complex. The plot shows the continuous distribution C(s) (y-axis) versus the sedimentation distribution coefficient (S) (x-axis), revealing the presence of a stable, intact, stoichiometric 1:1 complex. The residual distribution histogram from the AU analysis is shown in the upper right corner.

Circular dichroism (CD) spectroscopy

RNAs (GG[UUAGGG]$_{4/8}$UUA) were transcribed with T7 RNA polymerase and purified on 6% TBE-polyacrylamide gels supplemented with 8 M urea according to standard methods. Dried RNA pellets were dissolved in RNase-free water and diluted to 5 μM in 10 mM Tris–HCl pH 7.4, 1 mM TCEP and 100 mM KCl, NaCl, or LiCl as indicated in the main text. RNAs were folded using a standard 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). All CD spectra were recorded at room temperature on a Jasco J-810 spectropolarimeter with a 1 mm cell, scan speed of 50 nm/min, and a response time of 1 sec. Spectra from 300–220 nm were averaged over three scans, and background from a matched buffer-only sample was subtracted.

Analytical ultracentrifugation (AUC)

RNA samples were run for 16 h in an Optima XLI ultracentrifuge equipped with a four-hole An-60 Ti rotor at 48,000 rpm at 4°C. Samples and buffer-matched blanks were loaded into double-sector cells (path length of 1.2 cm) with charcoal-filled Epon centerpieces and sapphire windows. Data were fit to a continuous c(s) distribution model using SedFit, with a partial specific volume of 0.73, buffer density of 1.005, buffer viscosity of 0.0102, and a frictional ratio of 1.4. Bayesian modeling operations included in the Sedfit software enabled us to deconvolute asymmetric peaks.

Electrophoretic Mobility Shift Assay (EMSA)

All oligonucleotides were 5’-end labeled with T4 Polynucleotide Kinase (NEB) and [γ-^{32}P] ATP (6000 Ci/mmol, 10 mCi/mL, PerkinElmer). Unincorporated radiolabel was removed by application to Micro Bio-Spin columns packed with Bio-Gel P6 in 10 mM Tris–HCl pH 7.4, 0.02% sodium azide (Bio-Rad) according to manufacturer's instructions. Immediately prior to binding, radiolabeled oligonucleotide stocks (typically 1–2 μM) were diluted to 20 nM in EMSA buffer (25 mM HEPES pH 7.4, 1 mM TCEP, 10% glycerol, 0.02% bromophenol blue, 1 U/μL RNasin (Promega) and 100 mM KCl, NaCl, or LiCl, depending on reaction conditions) and folded as described (CD spectroscopy, Materials and Methods section). To initiate binding reactions (10 μL final volume) threefold serial dilutions of 50 μM LSD1/CoREST in EMSA buffer were mixed 1:1 with 20 nM oligonucleotide stocks and incubated at room temperature for 20 min. Reactions were placed on ice and chilled for 5 min before loading into 0.75% THE (34 mM Tris base, 66 mM HEPES free acid, 0.1 mM EDTA, pH 7.4) agarose gels supplemented with 10 mM potassium acetate, sodium acetate, or lithium sulfate depending on reaction conditions. Gels were run for 45 min at 6 V/cm in THE running buffer supplemented with appropriate salt (10 mM), with constant buffer recirculation at 4 °C. Gels were exposed to an Imaging-Screen K (Kodak) and images were collected with a Pharos FX Plus Molecular Imager (Bio-Rad). All binding reaction profiles were quantified using the Quantity One 4.6.9 (Bio-Rad) software package. Only signal corresponding to fully bound or unbound positions was analyzed; smears due to complex dissociation were not included. The integrated volume for each signal was determined by measuring

the identical area that encompasses the probe-only control with minimal background.

Results of oligonucleotide binding assays were expressed as the fraction of oligonucleotide bound and plotted as a function of protein concentration using Prism 6.0 (GraphPad Software, Inc., http://www.graphpad.com). Data were fit to a one-site hyperbolic binding function including a Hill coefficient ($Y = B_{max} \times X^h/\{K_d^h + X^h\}$), where Y is the fraction bound and X is the protein concentration (nM), and h is the Hill coefficient. An average of the Hill coefficient (h) was determined by finding the slope of a straight line fitted to points from a plot of log $[\theta/(1-\theta)]$ versus log of the protein concentration, where θ is the fraction of bound oligonucleotide.

Cross-linking mass spectrometry

To verify that LSD1/CoREST cross-links to GQ RNA in a specific manner, an RNA oligo of TERRA GQ repeats ($GG[UUAGGG]_8UUA$) and a control non-GQ-forming, size-matched RNA were 5′ ^{32}P-labeled. The RNAs were folded as described (CD spectroscopy, Materials and Methods section). The RNA was diluted to 1.4 μM in folding buffer plus 0.3 U/μL RNase Inhibitor (Promega, N2111) and incubated with or without 1.6 μM LSD1/CoREST at room temperature for 10 min. The reaction was then placed on ice and exposed to 240 mJ/cm^2 of 254 nm UV light with a Stratalinker 1800 and separated on a 4%–20% Mini-Protean TGX Precast Gel (Bio-Rad, 4561094). The gel was

exposed using an Imaging Screen-K (Bio-Rad, 1707841) and visualized on a PharosFX imager using the Quantity One software system (Figure 16).

The same general protocol was followed to identify the cross-linked peptides, with a few important distinctions. 5′-biotinylated RNA was purchased from IDT and folded as described above. 20 μM GQ RNA and 40 μM LSD1/CoREST were incubated together for 10 min at room temperature before being placed on ice and UV-cross-linked alongside a negative-control sample of LSD1/CoREST without RNA. The covalently cross-linked LSD1–RNA complex was enriched using Sera-Mag magnetic streptavidin-coated beads, medium binding (Genesee Scientific, 85–592) and a biotinylated RNA pull-down kit according to the manufacturer's instructions (Pierce, 20164). 0.5% SDS and 1% SDS were added to the RNA capture buffer and RNA wash buffers, respectively, as these conditions removed non-cross-linked LSD1 without disrupting the biotin–streptavidin interaction (data not shown).

The complex was eluted by boiling in 1xSDS loading buffer and gel purified via SDS-PAGE. The cross-linked LSD1 (control) and LSD1–RNA (sample) were visualized using colloidal Coomassie blue stain, cut out, and subjected to in-gel trypsin digestion overnight. Two separate cross-linking LC–MS/MS experiments were performed on two freshly purified LSD1/CoREST samples with a single GQ RNA and with a stacked GQ RNA. All samples were injected onto an LTQ Orbitrap high-resolution LC–MS/MS system (Vanderbilt Proteomics Core

Facility). The isotopic distribution was used to confirm the identity of peptide peaks covering ~75%–80% of LSD1 residues using SkyLine 3.5.[198]

CHAPTER 4

Identification of RNAs Bound by LSD1 and CoREST in Cells

4.1: Introduction

The specificity of LSD1/CoREST for the TERRA G-quadruplex and the emerging role of RNAs in the recruitment of chromatin regulators raise the question of the role of GQ motifs in LSD1 recruitment beyond the telomere. A number of studies have identified specific lncRNAs which recruit LSD1 to genes which are then regulated through H3K4 or H3K9 demethylation.[114–117,199,200] However, the determinants that dictate which RNAs are targeted by LSD1 complexes are unclear.

The existing studies of LSD1-bound RNAs have limitations: one looked at CoREST-associated RNAs using the less sensitive microarray detection, and the other identified LSD1-bound RNAs with a novel formaldehyde-RIP method (fRIP).[112,113] The latter method pulls down both RNA and DNA, utilizes an LSD1-specific antibody which may bind off-targets, and has the potential to misidentify RNAs bound to other members of LSD1-containing complexes through formaldehyde cross-linking. In order to better characterize the RNAs bound by LSD1, a photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP) study was commenced based on an established protocol.[201]

The PAR-CLIP technique has been applied to study the RNAs bound by numerous proteins, including the GQ-binding FMRP protein and histone

acetyltransferase CREB-binding protein (CRB), yielding insight to the RNA regulation and binding specificity of the proteins.[202,203] The FMRP system was shown to regulate specific mRNAs involved in particular pathways, while CRB was found bound to particular enhancer RNAs which led to histone acetylation at target genes.

PAR-CLIP is a powerful technique which incorporates 4-thiouridine or 6-thioguanosine nucleotide analogs into the RNA transcripts of cells. The analogs are specifically cross-linked with bound proteins by relatively low-energy 365-nm UV light under native conditions. The cross-linking reaction occurs between 4-thiouridine and the aromatic amino acids phenylalanine, tyrosine, and tryptophan as well as lysine and cysteine.[204]

A key benefit of PAR-CLIP is that cross-linked 4-thiouridine is interpreted as a cytosine by the reverse transcriptase during the generation of the cDNA library. This mutation allows better resolution of the binding site and a reduction in signal from non-cross-linked background RNA. (Figure 20) The PARalyzer analysis program developed by Dr. Uwe Ohler's group uses the overlapped reads containing T-to-C conversions to identify clusters which represent likely protein binding sites.[205]

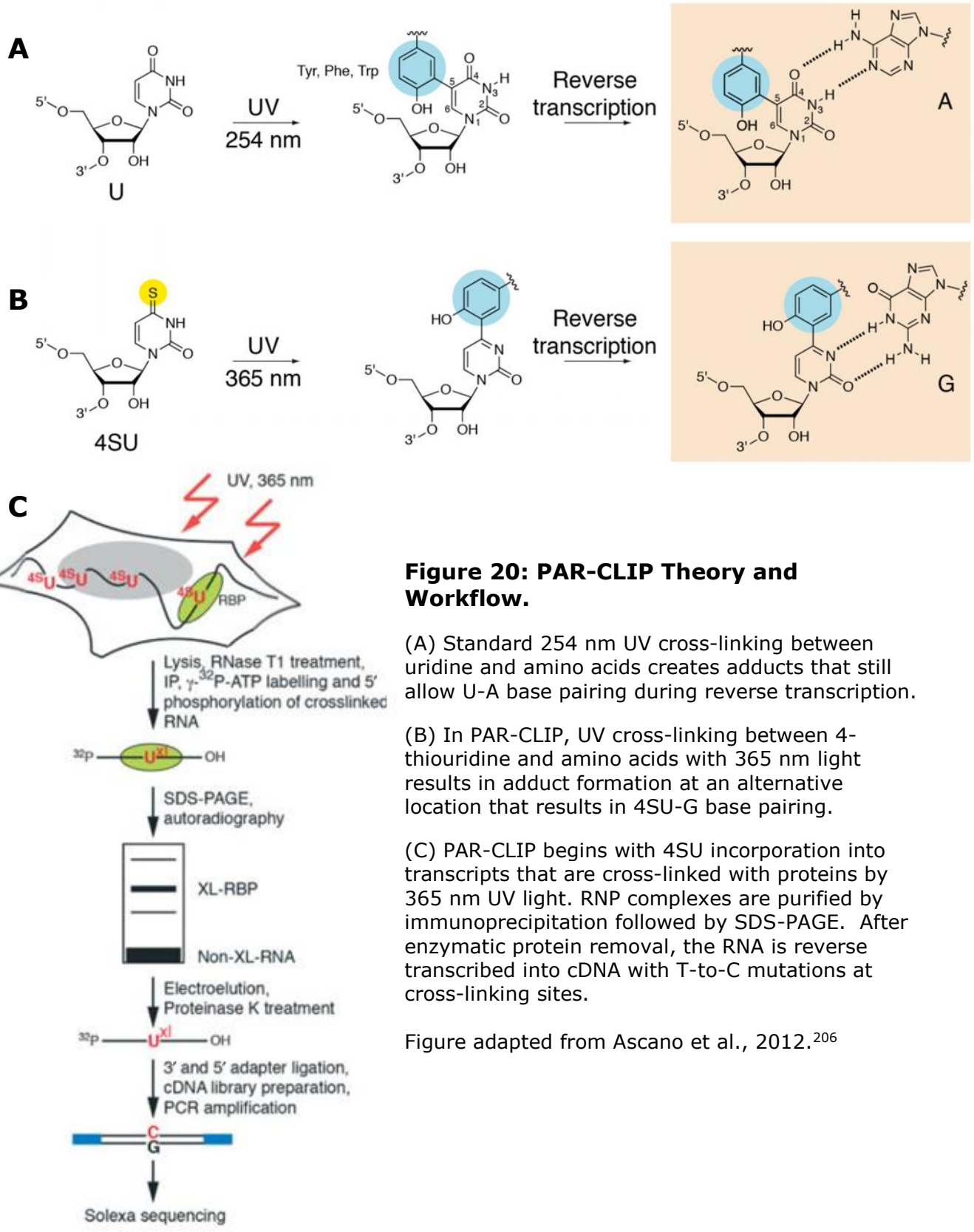


Figure 20: PAR-CLIP Theory and Workflow.

(A) Standard 254 nm UV cross-linking between uridine and amino acids creates adducts that still allow U-A base pairing during reverse transcription.

(B) In PAR-CLIP, UV cross-linking between 4-thiouridine and amino acids with 365 nm light results in adduct formation at an alternative location that results in 4SU-G base pairing.

(C) PAR-CLIP begins with 4SU incorporation into transcripts that are cross-linked with proteins by 365 nm UV light. RNP complexes are purified by immunoprecipitation followed by SDS-PAGE. After enzymatic protein removal, the RNA is reverse transcribed into cDNA with T-to-C mutations at cross-linking sites.

Figure adapted from Ascano et al., 2012.[206]

4.2: PAR-CLIP Results

Two separate batches of Flp-In HEK293 cells expressing FLAG-HA tagged LSD1 from a single genomic locus were harvested to create independent indexed libraries for sequencing. Combined, the sequencing runs produced 172 million LSD1 reads and 189 million CoREST reads. The LSD1 and CoREST datasets were merged and analyzed with the PARalyzer pipeline.[205] The pipeline aligned the reads to the human transcriptome, allowing for one T-to-C transition per read. After the pipeline was complete, it generated 59,335 nonredundant reads from the LSD1 dataset that correspond to the human transcriptome and 54,340 nonredundant CoREST reads. These sequences were organized into 1,433 and 1,068 clusters, respectively. Clusters are formed from overlapping reads, at least a subset of which contain T-to-C mutations.

The dataset of background PAR-CLIP clusters compiled by the Keene group was subtracted from the PAR-CLIP results.[207] The same high-occurrence reads were shown to reproducibly occur across different PAR-CLIP datasets. As these reads dilute the true binding sites, the removal of background clusters enhances the signal of true binding motifs and preferences. As further evidence that these clusters did not represent specific binding sites, the reads that make up LSD1 clusters also found in background had a slightly lower T-to-C conversion rates than other clusters, 30.1% vs. 34.9%. In total, 15% of PAR-CLIP identified genes were eliminated by the removal of clusters overlapping with the background datasets. Furthermore, 32% of genes found

in both the LSD1 and CoREST-bound datasets corresponded with background clusters. (Figure 21)

Enrichment of RNA Classes and G-Quadruplex Motifs

The lack of a true input dataset complicates the quantification of enrichment. In order to identify which transcripts are preferentially bound by LSD1 and CoREST, the composition of the mapped RNA reads was compared with the composition of RNA clusters. 3'-UTR transcripts are enriched over two-fold in both CoREST and LSD1 datasets. lncRNA transcripts and microRNAs are also enriched among the LSD1 transcripts but less so in the CoREST data. (Table 3)

A Perl script was modified and used to search for potential G-quadruplex forming motifs in the extended clusters dataset.[186] G-quadruplexes were conservatively defined as conforming to the formula, $(G_3N_{1-5})k$, where N can be A, C, G, or T and $k \geq 4$. G-quadruplexes require four groups of guanosines separated by loop sequences to form the four corners of the G-quadruplex. Motifs with a k value below 4 were used as a negative control to differentiate between general enrichment of guanosine-rich sequences and G-quadruplexes. While the LSD1 clusters were not enriched in G-quadruplex motifs, CoREST clusters exhibited a 2-fold enrichment for G-quadruplexes after background removal. (Table 4)

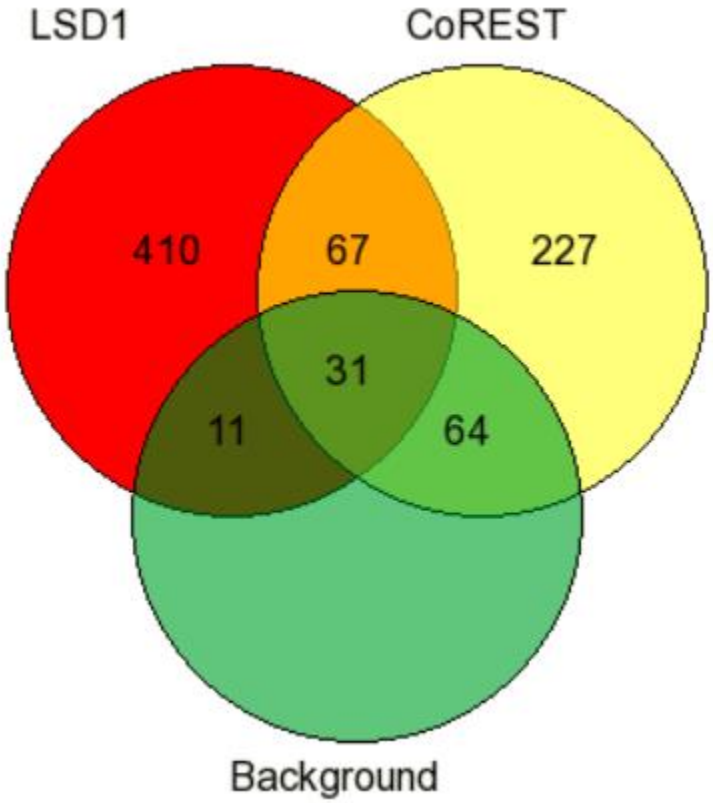


Figure 21: Overlap between genes containing PAR-CLIP clusters.

Table 3: Enrichment of Different RNAs in Gene Clusters vs. Reads.

LSD1 Transcripts	Enrichment	CoREST Transcripts	Enrichment
lncRNA	7.92	3' UTR	2.01
3'utr	2.25	Processed Transcript	1.50
miRNA	2.14	NMD	1.47
Intronic	1.24	Unannotated	1.43
Unannotated	1.15	Intronic	1.42
Repeat	1.13	lncRNA	1.25
rRNA	0.72	Stop Codon	1.22
Processed Transcript	0.38	Repeat	1.06
Stop codon	0.22	snRNA	0.77
tRNA	0.16	miRNA	0.76
Mt_tRNA	0.06	Mt_tRNA	0.73
5'utr	0	5' UTR	0.47
Mt_rRNA	0	Coding	0.33
NMD	0	rRNA	0.31
scRNA	0	Mt_rRNA	0.22
Start Codon	0	Start Codon	0.21
		tRNA	0.13

Enrichment of various classes of transcripts in LSD1 and CoREST PAR-CLIP results when comparing the percentage each class contributes to the total number of aligned reads versus the PARalyzer identified clusters that represent likely binding sites. Annotation performed by Ensembl/HAVANA.

Table 4: G-Quadruplex Enrichment.

Enrichment of G-Quad. Elements		
Dataset	**Non-GQ k=1-3**	**GQ Forming k=4-20**
CoREST	1.06	1.80
CoREST Minus Bckgd	1.06	2.11
LSD1	0.96	0.84
LSD1 Minus Bckgd	0.95	0.98

Enrichment of G-quadruplexes in LSD1 and CoREST datasets following background removal. In G-quadruplexes over non-G-quadruplex forming motifs according to the formula $(G_3N_{1-5})k$. G-quadruplexes only occur in elements where $k \geq 4$.

4.3: Validation of LSD1/CoREST Bound RNAs

The two-fold enrichment of G-quadruplex motifs among the CoREST-bound transcripts suggests that the LSD1/CoREST complex globally binds GQs regardless of the loop sequence. In order to verify this hypothesis, *in vitro* binding assays were conducted with several putative GQ-forming RNAs. Potential GQs in the LSD1 and CoREST datasets were largely located in mRNA introns and at intron/exon boundaries. Three GQ-forming motifs identified at LSD1/CoREST binding sites were chosen for *in vitro* verification. The motifs were found in the intronic regions of the FAM57B, MYO1B, and EPHB1 gene transcripts. The FAM57B and MYO1B G-quadruplex clusters were found in the CoREST dataset, while the EPHB1 cluster was associated with LSD1.

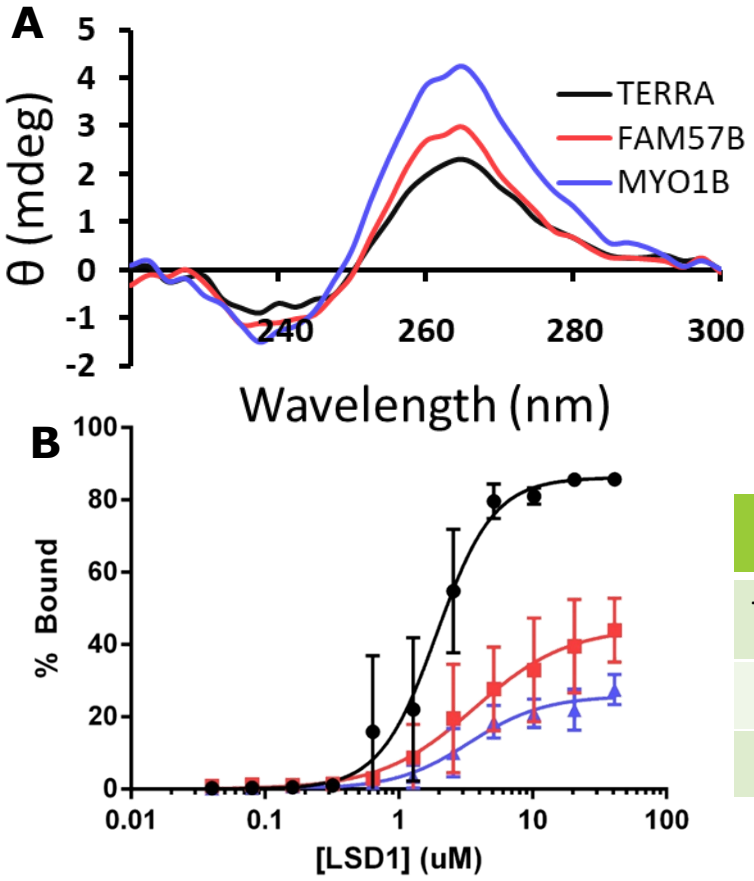
The RNAs were transcribed off of DNA oligos with T7 RNA polymerase and purified as described in the Materials and Methods section. CD was used to verify GQ formation. (Figure 22) FAM57B and MYO1B displayed characteristic GQ formation. Subsequent EMSA experiments showed that the LSD1/CoREST complex has a comparable affinity for FAM57B and MYO1B as for the TERRA RNA consisting of a single GQ motif, demonstrating that the protein complex does not specifically bind the TERRA RNA but instead recognizes the RNA GQ structure. The Bmax of the PAR-CLIP identified RNAs were significantly lower than for TERRA. A portion of the population of these RNAs may be folding into a conformation incompatible with LSD1/CoREST binding or the protein/GQ complex formed by these RNAs is more dynamic and unstable. The third GQ RNA, EPBH1, was tested in a separate set of experiments, but displayed less of

a characteristic CD profile and no significant binding with LSD1/CoREST (data not shown). This suggests that EPBH1 does not form a GQ-like conformation, though it is possible that there were heterogeneity issues associated with the *in vitro* transcription and sample preparation.

4.4: Implications of PAR-CLIP Study

lncRNAs are thought to recruit LSD1 to specific genes where LSD1 catalytically demethylates histones or serves as a scaffold for additional regulatory proteins.[114–117,199,200] Surprisingly, a previous fRIP study found that LSD1 primarily binds mRNAs.[113] mRNAs were much more common than lncRNAs in the PAR-CLIP datasets, although unannotated reads made up 40% of mapped reads and may contain additional lncRNAs which have yet to be unannotated. This may represent LSD1 binding at nascent mRNA transcripts or some novel function of LSD1.

When overlapping reads containing T-to-C mutations were grouped into clusters representing potential binding sites, lncRNA clusters made up a greater percentage of the total clusters than the total mapped reads, although the low total number of starting lncRNA reads is still low. Future studies will be needed to test whether these lncRNAs in fact recruit LSD1 to chromatin resulting in the regulation of neighboring genes.



	K_D (µM)	Bmax
TERRA	1.9±0.18	86.2
FAM57B	3.7±1.2	44.8
MYO1B	3.4±0.6	25.8

Figure 22: LSD1-CoREST Binds PAR-CLIP Identified GQ RNAs.

(A) Putative GQ-forming RNAs from the PAR-CLIP datasets were tested for GQ formation with circular dichroism. TERRA, FAM57B, and MYO1B exhibited the characteristic GQ signature. RNA sequences are described in the accompanying materials and methods section. (B) Summary of three EMSA experiments. The three RNAs exhibited similar affinity but variable maximum binding.

While there is precedent for LSD1 regulation by lncRNAs, the consequences of mRNA binding are unknown. 3' UTR clusters were enriched in both LSD1 and CoREST datasets compared with the total mapped reads and are a much larger percentage of the total reads and clusters. (Table 3) G-quadruplexes are known to be enriched in 3' UTRs where they have been shown to regulate microRNA binding and polyadenylation; numerous LSD1 and CoREST-bound G-quadruplexes were located in the 3'UTR region of our dataset.[208,209] Further work is required to delineate the significance of LSD1 binding at these sites and whether they represent a novel function of LSD1, possibly stabilizing GQ structures or serving as scaffolds for additional proteins.

4.5: Materials and Methods

Generation and Optimization of PAR-CLIP System

HEK293 cells efficiently incorporate 4-thiouridine and are commonly used in PAR-CLIP studies, allowing for relatively consistent protocols and the ability to better account for background reads.[207] The HEK293 cell line is derived from human embryonic cells and is commonly used in PAR-CLIP studies.[207] The Flp-In T-REx-293 cell line genome contains a single site for stable insertion of exogenous genes and expresses a tetracycline repressor protein, optimizing it for transient, induced gene expression. The Gateway BP/LR Clonase system was used to insert a FLAG-HA tag at the N-terminus of the full-length human LSD1 gene in between recombination sites, according to the manufacturer's directions. The entry vector was transfected into Flp-In T-REx-293 cells with Lipofectamine 2000 and cells with stably inserted FLAG-HA LSD1 were selected

using hygromycin B selection. To verify viable gene insertion, clones were treated with 3 ng/μL doxycycline for 24 hours, harvested, lysed, and Western blotted using anti-HA and anti-LSD1 antibodies. (Figure 23) Cell lines with stable inducible expression of FLAG-HA LSD1 were amplified and frozen to make cell line stocks.

PAR-CLIP experiments were carried out according to a previously described protocol.[201] The RNase T1 digestion steps were optimized using pilot experiments to obtain RNAs in the 19-24 nucleotide range for sequencing. RNase and DNase treatment ensured that RNA only and not DNA was being purified (Figure 24).

Briefly, cells were grown to 80% confluency and then treated with 100 μM 4-thiouridine and 1 ng/μL doxycycline 14 and 24 hours prior to cross-linking, respectively. Cross-linking was performed on ice in a Stratalinker 2400 using 0.15 J cm^{-2} of 365 nm light. Upon harvesting, the cell lysate was treated with 2 U/μL RNase T1 for 15 minutes at 22 °C. Lysate was then immunoprecipitated using anti-FLAG antibody before on-bead digestion using 0.3 U/ μL RNase T1. Samples were ^{32}P-radiolabeled using PNK kinase, run on as SDS-PAGE gel, and visualized with a phosphor screen. (Figure 23) Radioactive bands were observed corresponding with both LSD1 and CoREST and the presence of both proteins in the IP was confirmed with Western blotting, verifying that CoREST co-immunoprecipitates with LSD1 and both cross-link with RNA. The LSD1 and CoREST bands were extracted, proteinase K treated, and reverse transcribed into indexed cDNA libraries using the

NextFlex v3 Small RNA Preparation Kit from Illumina and purified on a 3% agarose gel using a BluePippin instrument.[210] The cDNA libraries were sequenced at the Vanderbilt Sequencing Core using NextSeq sequencing.

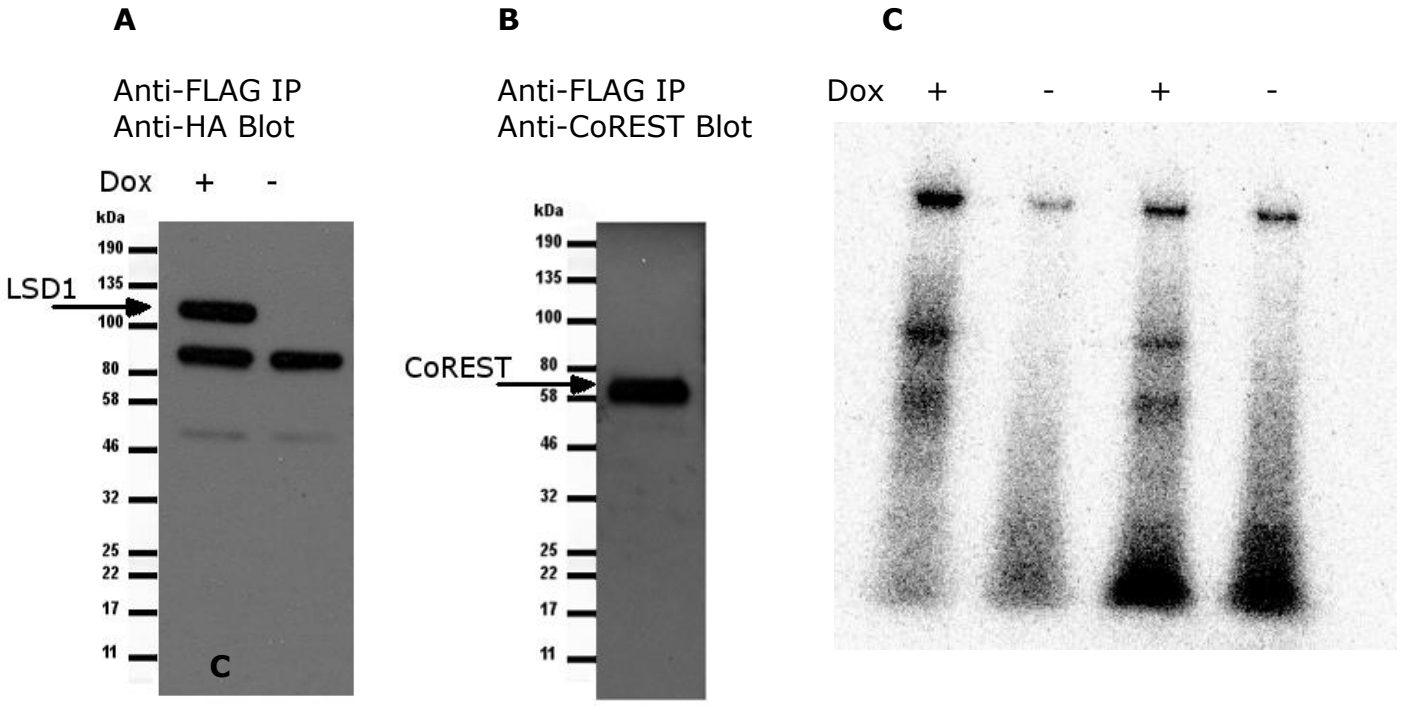


Figure 23: Induced expression of FLAG-HA LSD1 and Co-IP of CoREST.

LSD1 cells were treated with or without 1 ng/μL doxycycline for 24 hours and 4-thiouridine for 14 hours before cross-linking with 365 nm UV light. Lysate was immunoprecipitated with M2 anti-FLAG antibody and blotted with 12CA5 anti-HA antibody (A) or CoREST specific antibody (B). The band beneath LSD1 is from a cross-reaction with the IP antibody. (C) Stable FLAG-HA expressing cells were induced, treated with 4-thiouridine, cross-linked, immunoprecipitated with M2 anti-FLAG antibody, ^{32}P-radiolabeled, and run on a 4-20% gradient SDS-PAGE gel. The gel was then exposed with a K-screen phosphorimager screen. The bands correspond with the migration of LSD1 and CoREST.

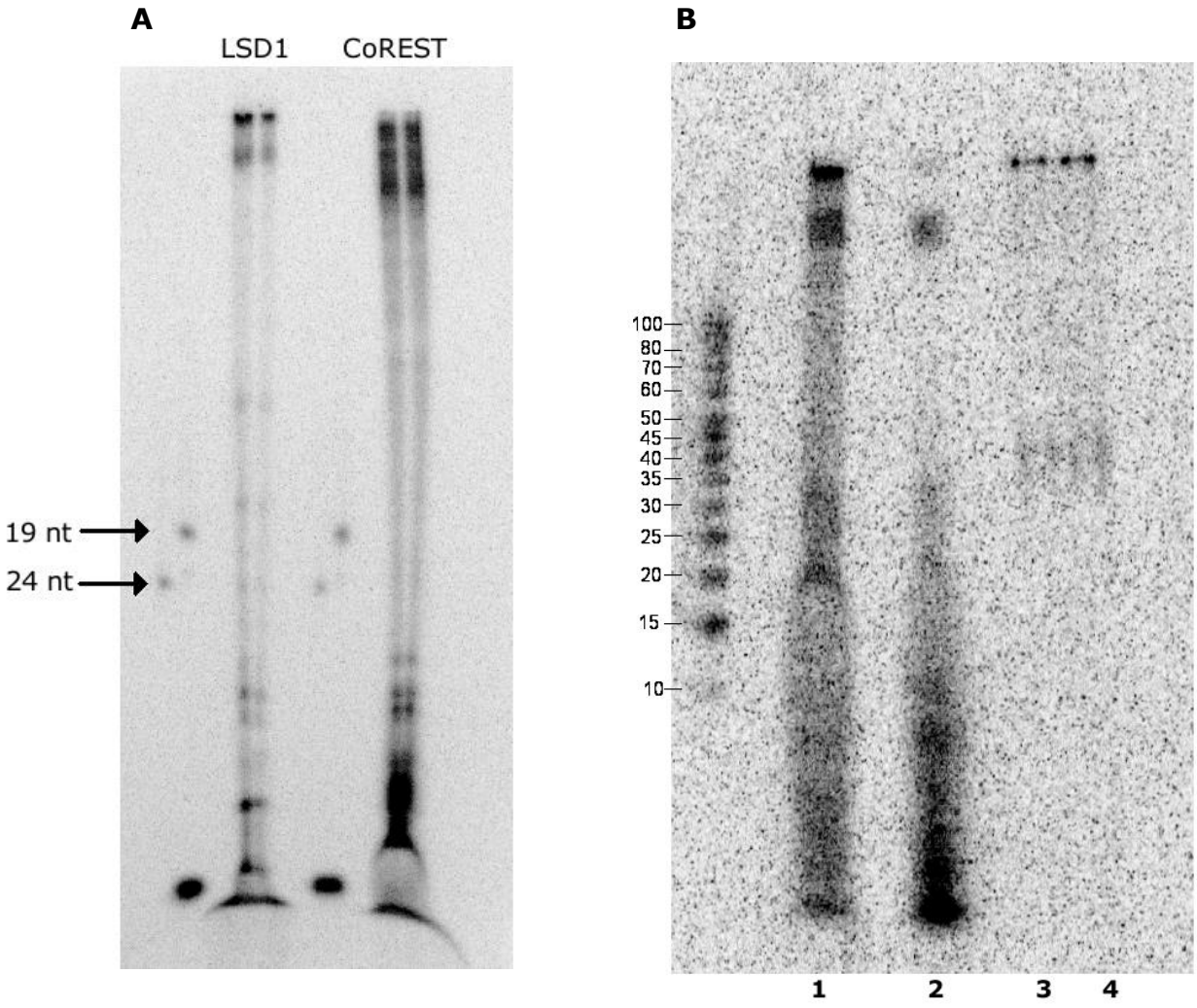


Figure 24: RNA binding by LSD1 and CoREST after Cross-link and IP.

(A) Both LSD1 and CoREST bands have associated RNA signal. Bands from previous figure were extracted, proteinase K treated, and run on a 16% polyacrylamide 8M urea denaturing gel. Radiolabeled 19 nt and 24 nt markers were run alongside the samples to denote the desired range for cDNA library generation. (B) An LSD1 band was extraction as in A and divided into aliquots that were treated with proteinase K and RNase I (230 U/μL, 20 minutes at room temperature). The samples were run out on a 16% PA-8M Urea-1xTBE gel. Lane assignments: 1) Proteinase K treated; 2) Proteinase K+RNase treated; 3,4) No treatment. The results indicate that the sample is RNase sensitive and not associated DNA.

Identification and Biophysical Testing of GQ RNAs

RNAs were identified by adapting a search script from Wang et al., 2017, to search for GQ motifs in the PAR-CLIP defined clusters.[186] Cluster sequences were extended by 20 nucleotides in either direction to account avoid eliminating GQs due to prematurely cutting off motifs at the cluster boundary. Circular dichroism and EMSA assays were conducted as described in Chapter 3 materials and methods section in K^+ conditions. The RNA sequences were as follows: TERRA=UUAGGGUUAGGGUUAGGGUUAGGGUUAU

FAM57B= GGUGGAGGGUGGGAGGGUGUACGGGGAAG

MYO1B= GGGUGUGGGUAGGGAAGAGGGGAACACUGGG

EPHB1= GGGGGUGGGGUUCAGGGACAGAGGGGUGGGAAAU

RNAs were prepared by T7 RNA polymerase transcription followed by denaturing gel purification.

CHAPTER 5

Discussion and Conclusions

ncRNAs represent a large, diverse, and biologically important class of biomolecules, but our understanding of the mechanisms underlying RNA interactions remains relatively primitive. At the time of this writing, RNA-containing structures represent only 2.7% of all PDB-deposited structures and RNAs make up a sliver of drug targets.[211,212] (https://www.rcsb.org) In order to more fully understand the biology and eventually target RNAs and RNA-protein complexes with novel therapeutics, it is critical to better understand these interactions both structurally and biochemically. This study advances that aim by studying the recognition of structured RNAs by two enzymes, PRORP and LSD1.

Summary of Structural and Biochemical PRORP Studies

PRORP separates the long polycistronic mitochondrial transcript by catalyzing the hydrolysis of the phosphodiester backbone on the 5' end of mitochondrial pre-tRNAs. These pre-tRNA genes are genetic hotspots of disease-associated SNPs which are linked to an accumulation of precursor tRNAs.[129–131] The RNAs are also processed by additional PPR domain proteins which may utilize similar mechanisms of recognition to PRORP.[139] Although recent studies have greatly advanced our models for how PRORP binds and cleaves its substrate, a crystal structure of the complex would shed light on precisely how PRORP and other mitochondrial PPR proteins recognize pre-tRNAs.

Towards this end, variants of the *A. thaliana* PRORP3 enzyme were cloned, expressed, and purified. Truncations of the N-terminus removed disordered regions of the protein and variants were cloned with inactivating mutations at the active site. Minimal pre-tRNA substrates were designed which contained only the necessary T-domain and acceptor stem and were engineered with kissing loops and tetraloop elements to promote crystallization. These pre-tRNAs were produced via *in vitro* transcription and verified as substrates with binding studies and activity assays. Various crystal screens were performed, but with limited success; only crystals of the PRORP enzyme alone were obtained, and the crystal was not readily reproducible.

Future Prospects

Future studies can build on these results. Additional PRORP variants can be created, guided by rational approaches as surface entropy reduction to promote crystallization and the optimization of construct properties.[213,214] More thermostable substrates such as engineered *T. thermophila* pre-tRNAs may yield better results. Ultimately, however, recent advances make cryo-electron microscopy an attractive method for this system. While obtaining a structure nearing the atomic-level resolution of X-ray crystallography would be very technically challenging for the small *A. thaliana* system, a cryo-EM study of the human mitochondrial complex is a realistic prospect. This approach has the added potential of readily obtaining structures for multiple different states, including both substrate and product bound complexes. Such structural

studies would aid in the understanding of mitochondrial biology and the pathogenesis of tRNA-related diseases.

LSD1-RNA Project Overview

RNA plays a number of roles in chromatin regulation and is capable of both coordinating large-scale chromosomal architecture and regulating the recruitment of protein complexes to chromatin.[215,216] The structural basis of these interactions has largely remained obscure. The histone demethylase LSD1 relies on forming complexes with additional factors for genomic localization, and in 2010 the lncRNA HOTAIR was demonstrated to functionally recruit LSD1 to specific genes.[110] However, the features that led to the specific LSD1-HOTAIR interaction remain unknown.

Characterization of the LSD1-TERRA Interaction

The subsequent discovery that the GQ-forming lncRNA TERRA also recruits LSD1 provided the opportunity to investigate LSD1 binding with a well-structured substrate.[46] Binding studies established that the LSD1-TERRA interaction is dependent on the tertiary GQ RNA fold and not the result of sequence specificity. It was recently reported that the other HOTAIR-binding chromatin regulator, the methyltransferase PRC2, also contains intrinsic GQ specificity.[186] However, it is not yet clear whether these GQ RNA recognition properties are involved with the HOTAIR interactions.

Cocrystallization studies with the LSD1/CoREST complex and telomeric GQ

RNAs were unsuccessful, with the components crystallizing independently but not as a complex. An alternative cross-linking/mass spectrometry approach was devised and optimized, and the results indicated that the binding site is located on the regulatory SWIRM domain of LSD1. Subsequent data from a high-throughput cross-link/proteomics study in mouse stem cells corroborated this noncanonical RNA binding site.[173] Future work with hydrogen/deuterium exchange and potentially electron microscopy could better characterize the binding interface in a cross-linking independent manner and identify potential conformational changes. Additional enzymatic studies with nucleosome substrates would provide insight into the functional consequences of GQ binding. Finally, preliminary negative-stain EM studies have been promising, with particles exhibiting the native conformation and GST-tagging LSD1/CoREST inducing dimerization to achieve the necessary molecular size. (Data not shown) Gold labeling of RNA allows EM visualization, though cross-linking likely needs to be employed to obtain analysis of complexes.

Identification of LSD1/CoREST Bound Transcripts

PAR-CLIP was employed to better understand LSD1 regulation and function by identifying associated transcripts in human cells. Endogenous CoREST co-immunoprecipitated with FLAG-HA tagged LSD1. The CoREST and LSD1 associated RNAs were analyzed and found to largely bind independent transcripts.

Potential GQ motifs were identified in the PAR-CLIP results, and several

potential motifs were validated *in vitro* with CD and shown to bind LSD1/CoREST with similar affinity to the TERRA GQ. This demonstrates that the LSD1/CoREST complex contains specificity for the RNA GQ and is not significantly affected by changes in the identity or length of the loop regions. However, unlike TERRA, these GQ elements were typically only capable of forming a single GQ structure and could not fold into the stacked dimeric GQs that are formed by longer TERRA constructs and preferred by LSD1. Therefore, it is possible that these are poorer substrates in cells.

LSD1 and CoREST largely bound intronic regions of mRNAs, a result that is consistent with previous studies.[113] It is unknown at this time whether these represent nascent, chromatin-associated pre-mRNA transcripts that could target LSD1/CoREST to chromatin or if the interaction serves some other undefined function. lncRNAs occurred with more frequency in the identified protein binding sites than in annotated reads as a whole, indicating that they represent a valid subset of LSD1-bound RNAs, but were present at low levels overall. A large proportion of mapped transcripts were located in unannotated regions, a pool which may contain additional unverified lncRNAs.

There are examples of lncRNAs recruiting LSD1 complexes with either H3K4 or H3K9 demethylation activity, suggesting that various LSD1-containing complexes are associated with different transcripts.[114–117] Future ChIP and co-immunoprecipitation studies will be needed to determine what macromolecular LSD1/CoREST complexes are bound with these various RNAs and to quantify

the resulting epigenetic effects.

In conclusion, this body of results underscores the role of structured RNA motifs in RNA/protein interactions. Combining *in vitro* and cellular assays allows for dissection of the determinants of binding while ensuring the applicability of the system to biological processes. Future studies will hopefully lead to a deeper understanding of how RNA transcripts regulate chromatin through lncRNA/protein complexes and how RNA processing proteins such as PRORP distinguish and regulate target RNAs, potentially resulting in better techniques to target RNA/protein complexes in the treatment of clinical disorders.

WORKS CITED

1. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38,** 381–393 (1968).
2. Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **38,** 367–379 (1968).
3. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5088–5090 (1977).
4. Butcher, S. E. & Pyle, A. M. The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. *Acc. Chem. Res.* **44,** 1302–1311 (2011).
5. Reiter, N. J., Chan, C. W. & Mondragón, A. Emerging structural themes in large RNA molecules. *Curr. Opin. Struct. Biol.* **21,** 319–326 (2011).
6. Reiter, N. J. *et al.* Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468,** 784–789 (2010).
7. Galej, W. P. *et al.* Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537,** 197–201 (2016).
8. Feigon, J., Chan, H. & Jiang, J. Integrative structural biology of *Tetrahymena* telomerase - insights into catalytic mechanism and interaction at telomeres. *FEBS J.* **283,** 2044–2050 (2016).
9. Martin, W. J. & Reiter, N. J. Structural Roles of Noncoding RNAs in the Heart of Enzymatic Complexes. *Biochemistry* **56,** 3–13 (2017).
10. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–857 (1983).
11. Kazantsev, A. V. & Pace, N. R. Bacterial RNase P: a new view of an ancient enzyme. *Nat. Rev. Microbiol.* **4,** 729–740 (2006).
12. Drainas, D. & Denis. Antibiotics and RNase P. *Antibiotics* **5,** 15 (2016).
13. Mondragón, A. Structural Studies of RNase P. *Annu. Rev. Biophys.* **42,** 537–557 (2013).
14. Yuan, Y. & Altman, S. Substrate recognition by human RNase P: identification of small, model substrates for the enzyme. *EMBO J.* **14,** 159–168 (1995).
15. Wilusz, J. E., Sunwoo, H. & Spector, D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23,** 1494–1504 (2009).
16. Forster, A. C. & Altman, S. External guide sequences for an RNA enzyme. *Science* **249,** 783–786 (1990).
17. Stein, C. A. *et al.* Intracellular mRNA cleavage induced through activation of RNase P bynuclease-resistant external guide sequences. *Nat. Biotechnol.* **18,** 58–61 (2000).
18. Kim, K. & Liu, F. Inhibition of gene expression in human cells using RNase P-derived ribozymes and external guide sequences. *Biochim. Biophys. Acta* **1769,**

603–612 (2007).

19. Wang, M. J., Davis, N. W. & Gegenheimer, P. Novel mechanisms for maturation of chloroplast transfer RNA precursors. *EMBO J.* **7,** 1567–1574 (1988).

20. Rossmanith, W., Tullo, A., Potuschak, T., Karwan, R. & Sbisà, E. Human mitochondrial tRNA processing. *J. Biol. Chem.* **270,** 12885–12891 (1995).

21. Holzmann, J. *et al.* RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135,** 462–474 (2008).

22. Thomas, B. C., Li, X. & Gegenheimer, P. Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism. *RNA* **6,** 545–553 (2000).

23. Rossmanith, W. & Potuschak, T. Difference between mitochondrial RNase P and nuclear RNase P. *Mol. Cell. Biol.* **21,** 8236–8237 (2001).

24. Gobert, A. *et al.* A single Arabidopsis organellar protein has RNase P activity. *Nat. Struct. Mol. Biol.* **17,** 740–744 (2010).

25. Rackham, O. *et al.* Hierarchical RNA Processing Is Required for Mitochondrial Ribosome Assembly. *Cell Rep.* **16,** 1874–1890 (2016).

26. Weber, C., Hartig, A., Hartmann, R. K. & Rossmanith, W. Playing RNase P Evolution: Swapping the RNA Catalyst for a Protein Reveals Functional Uniformity of Highly Divergent Enzyme Forms. *PLoS Genet.* **10,** e1004506 (2014).

27. Gößringer, M. *et al.* Protein-only RNase P function in Escherichia coli: viability, processing defects and differences between PRORP isoenzymes. *Nucleic Acids Res.* **45,** 7441–7454 (2017).

28. Howard, M. J., Lim, W. H., Fierke, C. A. & Koutmos, M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 16149–16154 (2012).

29. Pavlova, L. V. *et al.* tRNA Processing by Protein-Only versus RNA-Based RNase P: Kinetic Analysis Reveals Mechanistic Differences. *ChemBioChem* **13,** 2270–2276 (2012).

30. Brillante, N. *et al.* Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res.* **44,** 2323–2336 (2016).

31. Gobert, A. *et al.* Structural insights into protein-only RNase P complexed with tRNA. *Nat. Commun.* **4,** 1353 (2013).

32. Imai, T. *et al.* Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in Arabidopsis thaliana play a crucial role in recognition of nucleotide bases at TψC loop in precursor tRNAs. *Biochem. Biophys. Res. Commun.* **450,** 1541–1546 (2014).

33. Pinker, F. *et al.* Biophysical analysis of *Arabidopsis* protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding. *J. Biol. Chem.* **292,** 13904–13913 (2017).

34. Liu, X., Chen, Y. & Fierke, C. A. Inner-sphere Coordination of Divalent Metal Ion with Nucleobase in Catalytic RNA. *J. Am. Chem. Soc.* jacs.7b08755 (2017).

35. Blakely, E. L. *et al.* Pathogenic Mitochondrial tRNA Point Mutations: Nine Novel Mutations Affirm Their Importance as a Cause of Mitochondrial Disease. *Hum. Mutat.* **34,** 1260–1268 (2013).

36. Shen, C. *et al.* Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat. Commun.* **7,** 11285 (2016).

37. Yin, P. *et al.* Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504,** 168–171 (2013).

38. Ke, J. *et al.* Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat. Struct. Mol. Biol.* **20,** 1377–1382 (2013).

39. Longhese, M. P. DNA damage response at functional and dysfunctional telomeres. *Genes Dev.* **22,** 125–140 (2008).

40. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59,** 871–80 (1989).

41. Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342,** 825–829 (1989).

42. Meyne, J., Ratliff, R. L. & Moyzis, R. K. Conservation of the Human Telomere Sequence (TTAGGG)n among Vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **86,** 7049–7053

43. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric Repeat Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science.* **318,** 798–801 (2007).

44. Porro, A., Feuerhahn, S., Reichenbach, P. & Lingner, J. Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways. *Mol. Cell. Biol.* **30,** 4808–48017 (2010).

45. Montero, J. J., López de Silanes, I., Graña, O. & Blasco, M. A. Telomeric RNAs are essential to maintain telomeres. *Nat. Commun.* **7,** 12534 (2016).

46. Porro, A., Feuerhahn, S. & Lingner, J. TERRA-Reinforced Association of LSD1 with MRE11 Promotes Processing of Uncapped Telomeres. *Cell Rep.* **6,** 765–776 (2014).

47. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DMA structures. *Nature* **350,** 718–720 (1991).

48. GELLERT, M., LIPSETT, M. N. & DAVIES, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.* **48,** 2013–2018 (1962).

49. Pandey, S., Agarwala, P. & Maiti, S. Effect of Loops and G-Quartets on the Stability of RNA G-Quadruplexes. *J. Phys. Chem. B* **117,** 6896–6905 (2013).

50. Zhang, A. Y. Q., Bugaut, A. & Balasubramanian, S. A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* **50,** 7251–7258 (2011).

51. Collie, G. W., Haider, S. M., Neidle, S. & Parkinson, G. N. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* **38,** 5569–5580 (2010).

52. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353,** (2016).

53. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33,** 877–881 (2015).

54. 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. *Nat. Chem.* **6,** 75–80 (2013).

55. Morris, M. J., Negishi, Y., Pazsint, C., Schonhoft, J. D. & Basu, S. An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *J. Am. Chem. Soc.* **132,** 17831–17839 (2010).

56. Bonnal, S. *et al.* A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.* **278,** 39330–39336 (2003).

57. Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40,** 4727–4741 (2012).

58. Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3,** 218–221 (2007).

59. Brázda, V., Hároníková, L., Liao, J. C. C. & Fojta, M. DNA and RNA quadruplex-binding proteins. *Int. J. Mol. Sci.* **15,** 17493–17517 (2014).

60. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146,** 247–261 (2011).

61. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506,** 179–184 (2014).

62. Castets, M. *et al.* FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts. *Hum. Mol. Genet.* **14,** 835–844 (2005).

63. Schaeffer, C. *et al.* The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **20,** 4803–4813 (2001).

64. Vasilyev, N. *et al.* Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. U.S.A.* **112,** E5391-5400 (2015).

65. von Hacht, A. *et al.* Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res.* **42,** 6630–6644 (2014).

66. González, V., Guo, K., Hurley, L. & Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* **284,** 23622–23635 (2009).

67. Shi, Y. *et al.* Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **119,** 941–953 (2004).

68. Wang, J. *et al.* Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* **446,** 882–887 (2007).

69. Klemm, B. P. *et al.* Molecular recognition of pre-tRNA by *Arabidopsis* protein-only Ribonuclease P. *RNA* **23,** 1860–1873 (2017).

70. Hayami, S. *et al.* Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. *Int. J. Cancer* **128,** 574–586

(2011).

71. Lv, T. *et al.* Over-Expression of LSD1 Promotes Proliferation, Migration and Invasion in Non-Small Cell Lung Cancer. *PLoS ONE* **7,** e35065 (2012).

72. Harris, W. J. *et al.* The Histone Demethylase KDM1A Sustains the Oncogenic Potential of MLL-AF9 Leukemia Stem Cells. *Cancer Cell* **21,** 473–487 (2012).

73. Højfeldt, J. W., Agger, K. & Helin, K. Histone lysine demethylases as targets for anticancer therapy. *Nat. Rev. Drug Discov.* **12,** 917–930 (2013).

74. Gardner, K. E., Allis, C. D. & Strahl, B. D. Operating on chromatin, a colorful language where context matters. *J. Mol. Biol.* **409,** 36–46 (2011).

75. Adamo, A. *et al.* LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat. Cell Biol.* **13,** 652–660 (2011).

76. Gu, B. & Lee, M. G. Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells. *Cell Biosci.* **3,** 39 (2013).

77. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).

78. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res.* **21,** 564–578 (2011).

79. Vermeulen, M. *et al.* Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131,** 58–69 (2007).

80. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–124 (2001).

81. Canzio, D. *et al.* Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol. Cell* **41,** 67–81 (2011).

82. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7,** 715–727 (2006).

83. Wang, J. *et al.* The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41,** 125–129 (2009).

84. Cho, H.-S. *et al.* Demethylation of RB Regulator MYPT1 by Histone Demethylase LSD1 Promotes Cell Cycle Progression in Cancer Cells. *Cancer Res.* **71,** 655–660 (2011).

85. Lysine Methylation Regulates E2F1-Induced Cell Death. *Mol. Cell* **39,** 152–160 (2010).

86. Huang, J. *et al.* p53 is regulated by the lysine demethylase LSD1. *Nature* **449,** 105–108 (2007).

87. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439,** 811–816 (2006).

88. Fang, R. *et al.* Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. *Mol. Cell* **39,** 222–233 (2010).

89. Karytinos, A. *et al.* A Novel Mammalian Flavin-dependent Histone Demethylase. *J. Biol. Chem.* **284,** 17775–17782 (2009).

90. van Essen, D., Zhu, Y. & Saccani, S. A Feed-Forward Circuit Controlling Inducible NF-κB Target Gene Activation by Promoter Histone Demethylation. *Mol. Cell* **39,** 750–760 (2010).

91. Yang, Y., Yin, X., Yang, H. & Xu, Y. Histone Demethylase LSD2 Acts as an E3 Ubiquitin Ligase and Inhibits Cancer Cell Growth through Promoting Proteasomal Degradation of OGT. *Mol. Cell* **58,** 47–59 (2015).

92. Yang, M. *et al.* Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol. Cell* **23,** 377–387 (2006).

93. Shi, Y.-J. *et al.* Short Article Regulation of LSD1 Histone Demethylase Activity by Its Associated Factors. *Mol. Cell* **19,** 857–864 (2005).

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

95. Yang, P. *et al.* RCOR2 Is a Subunit of the LSD1 Complex That Regulates ESC Property and Substitutes for SOX2 in Reprogramming Somatic Cells to Pluripotency. *Stem Cells* **29,** 791–801 (2011).

96. Barrios, A. P. *et al.* Differential Properties of Transcriptional Complexes Formed by the CoREST Family. *Mol. Cell. Biol.* **34,** 2760–2770 (2014).

97. Kuppuswamy, M. *et al.* Role of the PLDLS-binding cleft region of CtBP1 in recruitment of core and auxiliary components of the corepressor complex. *Mol. Cell. Biol.* **28,** 269–281 (2008).

98. Lee, M. G. *et al.* Functional interplay between histone demethylase and deacetylase enzymes. *Mol. Cell. Biol.* **26,** 6395–6402 (2006).

99. Jurkin, J. *et al.* Distinct and redundant functions of histone deacetylases HDAC1 and HDAC2 in proliferation and tumorigenesis. *Cell Cycle* **10,** 406–412 (2011).

100. Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422,** 735–738 (2003).

101. Li, L. *et al.* ZNF516 suppresses EGFR by targeting the CtBP/LSD1/CoREST complex to chromatin. *Nat. Commun.* **8,** 691 (2017).

102. Cowger, J. J. M., Zhao, Q., Isovic, M. & Torchia, J. Biochemical characterization of the zinc-finger protein 217 transcriptional repressor complex: identification of a ZNF217 consensus recognition sequence. *Oncogene* **26,** 3378–3386 (2007).

103. Boxer, L. D., Barajas, B., Tao, S., Zhang, J. & Khavari, P. A. ZNF750 interacts with KLF4 and RCOR1, KDM1A, and CTBP1/2 chromatin regulators to repress epidermal progenitor genes and induce differentiation genes. *Genes Dev.* **28,** 2013–2026 (2014).

104. Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437,** 436 (2005).

105. Bolton, E. C. *et al.* Cell- and gene-specific regulation of primary target genes by the androgen receptor. *Genes Dev.* **21,** 2005–2017 (2007).

106. Carnesecchi, J. *et al.* ERRα induces H3K9 demethylation by LSD1 to promote cell invasion. *Proc. Natl. Acad. Sci. U.S.A.* **114,** 3909–3914 (2017).

107. Ray, S. K., Li, H. J., Metzger, E., Schule, R. & Leiter, A. B. CtBP and Associated

LSD1 Are Required for Transcriptional Activation by NeuroD1 in Gastrointestinal Endocrine Cells. *Mol. Cell. Biol.* **34,** 2308–2317 (2014).

108. Hamamoto, R., Saloura, V. & Nakamura, Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat. Rev. Cancer* **15,** 110–124 (2015).

109. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322,** 750–756 (2008).

110. Tsai, M.-C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329,** 689–93 (2010).

111. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

112. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 11667–11672 (2009).

113. G Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17,** 28 (2016).

114. Huang, M. *et al.* Long Noncoding RNA LINC00673 Is Activated by SP1 and Exerts Oncogenic Properties by Interacting with LSD1 and EZH2 in Gastric Cancer. *Mol. Ther.* **25,** 1014–1026 (2017).

115. Li, Y. *et al.* HBXIP and LSD1 Scaffolded by lncRNA Hotair Mediate Transcriptional Activation by c-Myc. *Cancer Res.* **76,** 293–304 (2016).

116. Lim, S. *et al.* Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. *Carcinogenesis* **31,** 512–520 (2010).

117. Liu, Y.-W. *et al.* LincRNAFEZF1-AS1 represses p21 expression to promote gastric cancer proliferation through LSD1-Mediated H3K4me2 demethylation. *Mol. Cancer* **16,** 39 (2017).

118. O'Connor, J. P. & Peebles, C. L. In vivo pre-tRNA processing in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **11,** 425–439 (1991).

119. Nickel, A. I. *et al.* Minimal and RNA-free RNase P in Aquifex aeolicus. *Proc. Natl. Acad. Sci. U.S.A.* **114,** 11121–11126 (2017).

120. Randau, L., Schröder, I. & Söll, D. Life without RNase P. *Nature* **453,** 120–123 (2008).

121. Holzmann, J. & Rossmanith, W. tRNA recognition, processing, and disease: Hypotheses around an unorthodox type of RNase P in human mitochondria. *Mitochondrion* **9,** 284–288 (2009).

122. Dang, Y. L. & Martin, N. C. Yeast mitochondrial RNase P. Sequence of the RPM2 gene and demonstration that its product is a protein subunit of the enzyme. *J. Biol. Chem.* **268,** 19791–19796 (1993).

123. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290,** 457–465 (1981).

124. Sissler, M., Pütz, J., Fasiolo, F. & Florentz, C. Mitochondrial Aminoacyl-tRNA Synthetases. Madame Curie Bioscience Database, Landes Bioscience. Austin, TX. (2013).

125. Brzezniak, L. K., Bijata, M., Szczesny, R. J. & Stepien, P. P. Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. *RNA Biol.* **8,** 616–626 (2011).

126. Doersen, C. J., Guerrier-Takada, C., Altman, S. & Attardi, G. Characterization of an RNase P activity from HeLa cell mitochondria. Comparison with the cytosol RNase P activity. *J. Biol. Chem.* **260,** 5942–5949 (1985).

127. Ojala, D., Montoya, J. & Attardi, G. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290,** 470–474 (1981).

128. Xu, F. *et al.* Disruption of a mitochondrial RNA-binding protein gene results in decreased cytochrome b expression and a marked reduction in ubiquinol-cytochrome c reductase activity in mouse heart mitochondria. *Biochem. J.* **416,** 15–26 (2008).

129. Tuppen, H. A. L., Blakely, E. L., Turnbull, D. M. & Taylor, R. W. Mitochondrial DNA mutations and human disease. *Biochim. Biophys. Acta - Bioenerg.* **1797,** 113–128 (2010).

130. Lu, Y. *et al.* Mitochondrial tRNA genes are hotspots for mutations in a cohort of patients with exercise intolerance and mitochondrial myopathy. *J. Neurol. Sci.* **379,** 137–143 (2017).

131. Giordano, C. *et al.* Cardiomyopathies due to homoplasmic mitochondrial tRNA mutations: morphologic and molecular features. *Hum. Pathol.* **44,** 1262–1270 (2013).

132. Goto, Y., Nonaka, I. & Horai, S. A mutation in the tRNALeu(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* **348,** 651–653 (1990).

133. Reiter, N. J., Osterman, A. K. & Mondragón, A. The bacterial ribonuclease P holoenzyme requires specific, conserved residues for efficient catalysis and substrate positioning. *Nucleic Acids Res.* **40,** 10384–10393 (2012).

134. Karasik, A., Shanmuganathan, A., Howard, M. J., Fierke, C. A. & Koutmos, M. Nuclear Protein-Only Ribonuclease P2 Structure and Biochemical Characterization Provide Insight into the Conserved Properties of tRNA 5' End Processing Enzymes. *J. Mol. Biol.* **428,** 26–40 (2016).

135. Gutmann, B., Gobert, A. & Giege, P. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.* **26,** 1022–1027 (2012).

136. Barkan, A. & Small, I. Pentatricopeptide Repeat Proteins in Plants. *Annu. Rev. Plant Biol.* **65,** 415–442 (2014).

137. Li, F., Liu, X., Zhou, W., Yang, X. & Shen, Y. Auto-inhibitory Mechanism of the Human Mitochondrial RNase P Protein Complex. *Sci. Rep.* **5,** 9878 (2015).

138. Yang, J. *et al.* Molecular basis for TPR domain-mediated regulation of protein phosphatase 5. *EMBO J.* **24,** 1–10 (2005).

139. Lightowlers, R. N. & Chrzanowska-Lightowlers, Z. M. A. Human

pentatricopeptide proteins: only a few and what do they do? *RNA Biol.* **10,** 1433–1438 (2013).

140. Barkan, A. *et al.* A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. *PLoS Genet.* **8,** e1002910 (2012).

141. Howard, M. J., Klemm, B. P. & Fierke, C. A. Mechanistic Studies Reveal Similar Catalytic Strategies for Phosphodiester Bond Hydrolysis by Protein-only and RNA-dependent Ribonuclease P. *J. Biol. Chem.* **290,** 13454–13464 (2015).

142. Vera, L., Czarny, B., Georgiadis, D., Dive, V. & Stura, E. A. Practical Use of Glycerol in Protein Crystallization. *Cryst. Growth Des.* **11,** 2755–2762 (2011).

143. Zhang, J. & Ferré-D'Amaré, A. R. New molecular engineering approaches for crystallographic studies of large RNAs. *Curr. Opin. Struct. Biol.* **26,** 9–15 (2014).

144. Ferré-D'Amaré, A. R. Use of the U1A Protein to Facilitate Crystallization and Structure Determination of Large RNAs. *Methods Mol. Biol.* **1320,** 67–76 (2016).

145. Howard, M. J. *et al.* Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. *RNA* **22,** 782–792 (2016).

146. Ferré-D'Amaré, A. R. & Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295,** 541–556 (2000).

147. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell* **43,** 904–914 (2011).

148. Stavropoulos, P., Blobel, G. & Hoelz, A. Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat. Struct. Mol. Biol.* **13,** 626–632 (2006).

149. Amente, S., Lania, L. & Majello, B. The histone LSD1 demethylase in stemness and cancer transcription programs. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829,** 981–986 (2013).

150. Forneris, F., Binda, C., Adamo, A., Battaglioli, E. & Mattevi, A. Structural Basis of LSD1-CoREST Selectivity in Histone H3 Recognition. *J. Biol. Chem.* **282,** 20070–20074 (2007).

151. Laurent, B. *et al.* A Specific LSD1/KDM1A Isoform Regulates Neuronal Differentiation through H3K9 Demethylation. *Mol. Cell* **57,** 957–970 (2015).

152. Chatr-aryamontri, A. *et al.* The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* **41,** D816–D823 (2012).

153. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13,** 297–311 (2012).

154. Luka, Z., Moss, F., Loukachevitch, L. V., Bornhop, D. J. & Wagner, C. Histone Demethylase LSD1 Is a Folate-Binding Protein. *Biochemistry* **50,** 4750–4756 (2011).

155. Forneris, F., Binda, C., Battaglioli, E. & Mattevi, A. LSD1: oxidative chemistry for multifaceted functions in chromatin regulation. *Trends Biochem. Sci.* **33,** 181–189 (2008).

156. Hwang, S., Schmitt, A. A., Luteran, A. E., Toone, E. J. & McCafferty, D. G.

Thermodynamic Characterization of the Binding Interaction between the Histone Demethylase LSD1/KDM1 and CoREST. *Biochemistry* **50,** 546–557 (2011).

157. Porro, A. *et al.* Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat. Commun.* **5,** 5379 (2014).

158. Azzalin, C. M. & Lingner, J. Telomere functions grounding on TERRA firma. *Trends Cell Biol.* **25,** 29–36 (2015).

159. Cusanelli, E. & Chartrand, P. Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. *Front. Genet.* **6,** 143 (2015).

160. Rippe, K. & Luke, B. TERRA and the state of the telomere. *Nat. Struct. Mol. Biol.* **22,** 853–858 (2015).

161. Lamarche, B. J., Orazio, N. I. & Weitzman, M. D. The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett.* **584,** 3682–3695 (2010).

162. Peng, B. *et al.* Modulation of LSD1 phosphorylation by CK2/WIP1 regulates RNF168-dependent 53BP1 recruitment in response to DNA damage. *Nucleic Acids Res.* **43,** 5936–5947 (2015).

163. Mosammaparast, N. *et al.* The histone demethylase LSD1/KDM1A promotes the DNA damage response. *J. Cell Biol.* **203,** 457–470 (2013).

164. Costa, R. *et al.* The lysine-specific demethylase 1 is a novel substrate of protein kinase CK2. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 722–729 (2014).

165. Engreitz, J. M. *et al.* RNA-RNA Interactions Enable Specific Targeting of Noncoding RNAs to Nascent Pre-mRNAs and Chromatin Sites. *Cell* **159,** 188–199 (2014).

166. Somarowthu, S. *et al.* HOTAIR Forms an Intricate and Modular Secondary Structure. *Mol. Cell* **58,** 353–361 (2015).

167. Balaratnam, S. & Basu, S. Divalent cation-aided identification of physico-chemical properties of metal ions that stabilize RNA g-quadruplexes. *Biopolymers* **103,** 376–386 (2015).

168. Martadinata, H., Heddi, B., Lim, K. W. & Phan, A. T. Structure of Long Human Telomeric RNA (TERRA): G-Quadruplexes Formed by Four and Eight UUAGGG Repeats Are Stable Building Blocks. *Biochemistry* **50,** 6455–6461 (2011).

169. Martadinata, H. & Phan, A. T. Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K^+ Solution. *Biochemistry* **52,** 2176–2183 (2013).

170. Xu, Y., Suzuki, Y., Ito, K. & Komiyama, M. Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 14579–14584 (2010).

171. Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E. & Mattevi, A. Human Histone Demethylase LSD1 Reads the Histone Code. *J. Biol. Chem.* **280,** 41360–41365 (2005).

172. Hirschi, A., Martin, W. J., Luka, Z., Loukachevitch, L. V. & Reiter, N. J. G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme. *RNA* **22,** 1250–1260 (2016).

173. He, C. *et al.* High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol. Cell* **64,** 416–430 (2016).

174. Luke, B. *et al.* The Rat1p 5′ to 3′ Exonuclease Degrades Telomeric Repeat-Containing RNA and Promotes Telomere Elongation in Saccharomyces cerevisiae. *Mol. Cell* **32,** 465–477 (2008).

175. Schoeftner, S. & Blasco, M. A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* **10,** 228–236 (2008).

176. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Mol. Cell* **35,** 403–413 (2009).

177. Iglesias, N. *et al.* Subtelomeric repetitive elements determine TERRA regulation by Rap1/Rif and Rap1/Sir complexes in yeast. *EMBO Rep.* **12,** 587–593 (2011).

178. de Silanes, I. L. *et al.* Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nat. Commun.* **5,** 4723 (2014).

179. Vaquero-Sedas, M. I., Luo, C. & Vega-Palas, M. A. Analysis of the epigenetic status of telomeres by using ChIP-seq data. *Nucleic Acids Res.* **40,** e163–e163 (2012).

180. Krogan, N. J. *et al.* COMPASS, a Histone H3 (Lysine 4) Methyltransferase Required for Telomeric Silencing of Gene Expression. *J. Biol. Chem.* **277,** 10753–10755 (2002).

181. Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **38,** 5797–5806 (2010).

182. Haider, S. M., Neidle, S. & Parkinson, G. N. A structural analysis of G-quadruplex/ligand interactions. *Biochimie* **93,** 1239–1251 (2011).

183. Patel, D. J., Phan, A. T. & Kuryavyi, V. Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **35,** 7429–7455 (2007).

184. Chen, M. C., Murat, P., Abecassis, K., Ferré-D'Amaré, A. R. & Balasubramanian, S. Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res.* **43,** 2223–2231 (2015).

185. Meier, M. *et al.* Binding of G-quadruplexes to the N-terminal Recognition Domain of the RNA Helicase Associated with AU-rich Element (RHAU). *J. Biol. Chem.* **288,** 35014–35027 (2013).

186. Wang, X. *et al.* Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Mol. Cell* **65,** 1056–1067. (2017).

187. Di Antonio, M. *et al.* Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry. *Angew. Chemie Int. Ed.* **51,** 11073–11078 (2012).

188. Chen, Y. *et al.* Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc. Natl. Acad. Sci. U.S.A.* **103,** 13956–13961 (2006).

189. Baron, R. & Vellore, N. A. LSD1/CoREST is an allosteric nanoscale clamp regulated by H3-histone-tail molecular recognition. *Proc. Natl. Acad. Sci. U.S.A.*

109, 12509–12514 (2012).

190. Flynn, R. L. *et al.* Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science.* **347,** 273–277 (2015).

191. Cifuentes-Rojas, C., Hernandez, A. J., Sarma, K. & Lee, J. T. Regulatory Interactions between RNA and Polycomb Repressive Complex 2. *Mol. Cell* **55,** 171–185 (2014).

192. Kaneko, S., Son, J., Bonasio, R., Shen, S. S. & Reinberg, D. Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes Dev.* **28,** 1983–1988 (2014).

193. Sarma, K. *et al.* ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets. *Cell* **159,** 869–883 (2014).

194. Davidovich, C. *et al.* Toward a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA. *Mol. Cell* **57,** 552–558 (2015).

195. Cloutier, S. C. *et al.* Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. *Mol. Cell* **61,** 393–404 (2016).

196. Culhane, J. C. *et al.* A Mechanism-Based Inactivator for Histone Demethylase LSD1. *J. Am. Chem. Soc.* **128,** 4536–4537 (2006).

197. Culhane, J. C., Wang, D., Yen, P. M. & Cole, P. A. Comparative Analysis of Small Molecules and Histone Substrate Analogues as LSD1 Lysine Demethylase Inhibitors. *J. Am. Chem. Soc.* **132,** 3164–3176 (2010).

198. Schilling, B. *et al.* Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline. *Mol. Cell. Proteomics* **11,** 202–214 (2012).

199. He, R., Zhang, F. hu & Shen, N. LncRNA FEZF1-AS1 enhances epithelial-mesenchymal transition (EMT) through suppressing E-cadherin and regulating WNT pathway in non-small cell lung cancer (NSCLC). *Biomed. Pharmacother.* **95,** 331–338 (2017).

200. Ranzani, V. *et al.* The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat. Immunol.* **16,** 318–325 (2015).

201. Spitzer, J. *et al.* PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods Enzymol.* **539,** 113–161 (2014).

202. Ascano, M. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492,** 382–386 (2012).

203. Bose, D. A. *et al.* RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell* **168,** 135–149. (2017).

204. Hafner, M. *et al.* Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141,** 129–141 (2010).

205. Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12,** R79 (2011).

206. Ascano, M., Hafner, M., Cekan, P., Gerstberger, S. & Tuschl, T. Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip. Rev. RNA* **3,** 159–177 (2012).

207. Friedersdorf, M. B. & Keene, J. D. Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol.* **15,** R2 (2014).

208. Beaudoin, J.-D. & Perreault, J.-P. Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* **41,** 5898–5911 (2013).

209. Rouleau, S., Glouzon, J.-P. S., Brumwell, A., Bisaillon, M. & Perreault, J.-P. 3' UTR G-quadruplexes regulate miRNA binding. *RNA* **23,** 1172–1179 (2017).

210. Morris, A., Allen, K., Eun, S. H. & Toloue, M. Advances in Small RNA Library Preparation Allow Combination of Bias Reduction with Gel-free or Low Input Protocols. Bioo Scientific, www.biooscientific.com/Portals/0/Posters/Bioo-Scientific-Small-RNA-Poster.pdf.

211. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16,** 19–34 (2016).

212. Matsui, M. & Corey, D. R. Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.* **16,** 167–179 (2016).

213. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci.* **16,** 1569–15 76 (2007).

214. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23,** 3403–3405 (2007).

215. Hacisuleyman, E. *et al.* Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21,** 198–206 (2014).

216. Rinn, J. L. lncRNAs: linking RNA to chromatin. *Cold Spring Harb. Perspect. Biol.* **6,** a018614 (2014).