# G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme

ALEXANDER HIRSCHI,[1,3] WILLIAM J. MARTIN,[1,3] ZIGMUND LUKA,[1] LIOUDMILA V. LOUKACHEVITCH,[2] and NICHOLAS J. REITER[1]

[1]Department of Biochemistry, Vanderbilt University Medical Center, Nashville, Tennessee 37232-0146, USA
[2]Department of Pharmacology, Vanderbilt University Medical Center, Nashville, Tennessee 37232-6600, USA

## ABSTRACT

**Lysine-specific histone demethylase 1 (LSD1) is an essential epigenetic regulator in metazoans and requires the co-repressor element-1 silencing transcription factor (CoREST) to efficiently catalyze the removal of mono- and dimethyl functional groups from histone 3 at lysine positions 4 and 9 (H3K4/9). LSD1 interacts with over 60 regulatory proteins and also associates with lncRNAs (TERRA, HOTAIR), suggesting a regulatory role for RNA in LSD1 function. We report that a stacked, intramolecular G-quadruplex (GQ) forming TERRA RNA (GG[UUAGGG]$_8$UUA) binds tightly to the functional LSD1–CoREST complex ($K_d \approx 96$ nM), in contrast to a single GQ RNA unit ([UUAGGG]$_4$U), a GQ DNA ([TTAGGG]$_4$T), or an unstructured single-stranded RNA. Stabilization of a parallel-stranded GQ RNA structure by monovalent potassium ions ($K^+$) is required for high affinity binding to the LSD1–CoREST complex. These data indicate that LSD1 can distinguish between RNA and DNA as well as structured versus unstructured nucleotide motifs. Further, cross-linking mass spectrometry identified the primary location of GQ RNA binding within the SWIRM/amine oxidase domain (AOD) of LSD1. An ssRNA binding region adjacent to this GQ binding site was also identified via X-ray crystallography. This RNA binding interface is consistent with kinetic assays, demonstrating that a GQ-forming RNA can serve as a noncompetitive inhibitor of LSD1-catalyzed demethylation. The identification of a GQ RNA binding site coupled with kinetic data suggests that structured RNAs can function as regulatory molecules in LSD1-mediated mechanisms.**

**Keywords: LSD1; TERRA; lncRNA; ncRNA; G-quadruplex; chromatin; enzyme; kinetics; binding; mass spectrometry; structure; RNA–protein interactions**

## 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 (Wang and Chang 2011). Although it is well established that distinct ncRNAs can act as gene regulators, recognize defined targets, and even function as catalysts (ribozymes) (Reiter et al. 2011), key mechanistic questions remain regarding how lncRNAs interact with and recruit chromatin-associated protein complexes to specific regions in the genome.

The lysine specific histone demethylase 1A (LSD1 or KDM1A) is an essential chromatin-remodeling enzyme conserved from yeast to humans and is also known to interact with lncRNAs (Shi et al. 2004; Stavropoulos et al. 2006; Khalil et al. 2009; Amente et al. 2013; Porro et al. 2014b). 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) (Shi et al. 2004; Forneris et al. 2007; Laurent et al. 2015). 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]) (Luka et al. 2011; Kooistra and Helin 2012; Chatr-Aryamontri et al. 2013). Of these, CoREST is the primary interacting partner required for post-transla-

*Abbreviations:* LSD1, lysine-specific histone demethylase 1; CoREST, co-repressor for repressor element 1 silencing transcription factor; diMeK4H3$_{1\text{-}21}$, dimethyl-Lys4 Histone H3 peptide aa1-21; 4-AAP, 4-aminoantipyrine; DCHBS, 3,5-dichloro-2-hydroxybenzenesulfonic acid; HRP, horseradish peroxidase; EDTA, ethylenediaminetetraacetic acid; TB, Tris–Borate; GST, glutathione *S*-transferase; GQ, G-quadruplex; TERRA, telomeric repeat containing RNA; HOTAIR, homeotic transcript antisense RNA; p53, tumor suppressor p53; E2F1, E2F transcription factor 1; DNMT1, DNA cytosine-5-methyltransferase 1; MRE11, meiotic recombination 11; HDAC1/2, histone deacetylase 1/2; nt, nucleotide; lncRNA, long noncoding ribonucleic acid

[3]**These authors contributed equally to this work.**
**Corresponding author: nick.reiter@vanderbilt.edu**

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.057265.116.

tional LSD1 stabilization and is required for H3K4 demethylation during development, hematopoiesis, and stem cell maintenance (Shi et al. 2004; Forneris et al. 2008; Hwang et al. 2011).

While distinct LSD1-containing protein complexes are established to repress or activate gene transcription (Shi et al. 2004; Laurent et al. 2015), it is unknown how lncRNAs bind and modulate these complexes. 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 (Porro et al. 2014a,b; Azzalin and Lingner 2015; Cusanelli and Chartrand 2015; Rippe and Luke 2015) In addition, TERRA has a strong propensity to form intramolecular, parallel-stranded G-quadruplex (GQ) RNA structures due to its repeating UUAGGG sequence (Patel et al. 2007).

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 (Porro et al. 2014b; Azzalin and Lingner 2015). 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 MRE11A (MRE11) to trim the 3′ G overhangs at uncapped telomeres (Fig. 1A; Porro et al. 2014b). Validation of these interactions in vivo and in vitro (Porro et al. 2014b) 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 (Tsai et al. 2010; Engreitz et al. 2014; Porro et al. 2014b; Somarowthu et al. 2015). Taken together, these data support a role for TERRA in recruiting proteins to modulate heterochromatin formation at chromosome ends.

Here we identify and characterize GQ nucleic acid binding on the surface of a functional LSD1–CoREST complex (Fig. 1B). We demonstrate that LSD1–CoREST has a strong preference to bind a stacked GQ-forming RNA and reveal that the primary binding site of the GQ RNA exists within the SWIRM/amine-oxidase domain of LSD1. In addition to cross-linking mass spectrometry and X-ray crystallographic data, we analyzed RNA binding and the influence of distinct nucleic acid structures on the kinetics of LSD1-catalyzed demethylation, demonstrating that a GQ-forming RNA acts as a potential noncompetitive inhibitor of LSD1-catalyzed histone demethylation. These data indicate that structured RNAs can function as important regulators in LSD1-mediated pathways and supports an emerging theme that structure-specific RNA binding can influence the function of chromatin-associated proteins.

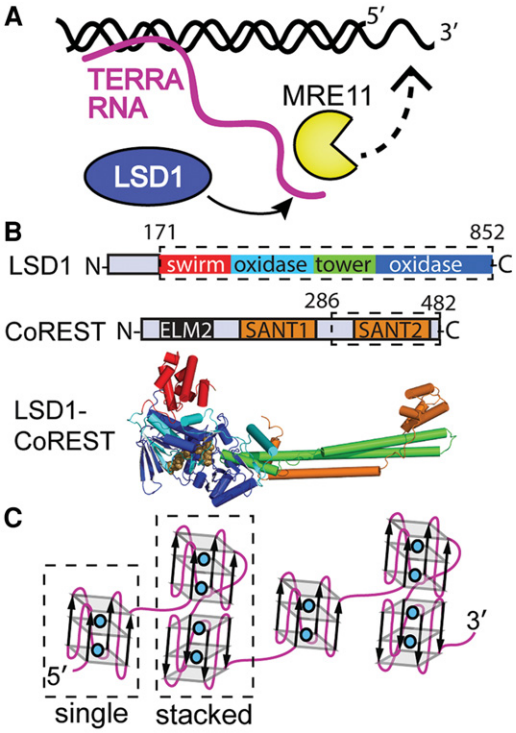


**FIGURE 1.** TERRA RNA recruits LSD1 to deprotected telomeres. (*A*) A functional role for the TERRA RNA in the processing of uncapped telomeres, as previously reported (Porro et al. 2014b). TERRA can serve as a scaffold for LSD1–MRE11 associations. (*B*) Protein constructs used in this study include 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 (Yang et al. 2006). 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 (Xu et al. 2010) and a model of the higher order TERRA RNA architecture has been previously demonstrated (Martadinata and Phan 2013).

## RESULTS

### Monovalent ions dramatically influence TERRA topology

Quadruplex-forming nucleic acid sequences require specific monovalent ions for structural stability. In particular, potassium ions stabilize GQs, while lithium ions destabilize GQs (Balaratnam and Basu 2015). We used circular dichroism (CD) spectroscopy 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 (Martadinata et al. 2011;

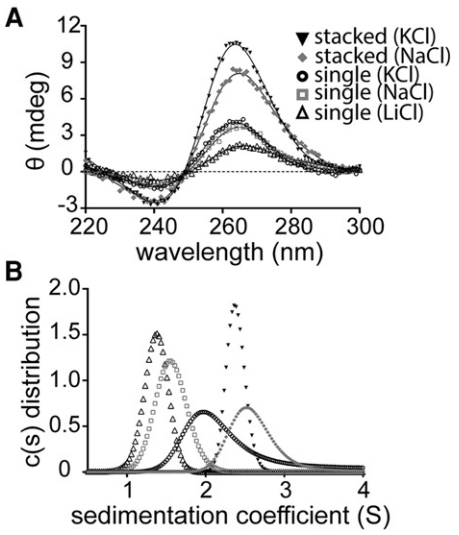


**FIGURE 2.** Monovalent ions dramatically influence the structure of GQ-forming RNAs. (*A*) Parallel-stranded GQ RNAs are known to have an $\Theta_{max}$ of ~263 nm (Balaratnam and Basu 2015). Circular dichroism spectroscopy demonstrates that parallel-stranded GQ structures form in the presence of potassium (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).

Martadinata and Phan 2013; Balaratnam and Basu 2015). In both four and eight UUAGGG repeats (in the presence of $K^+$ and $Na^+$), we observe the formation of stable parallel-stranded GQ structures (Fig. 2A). In contrast, $Li^+$ destabilizes GQ structure and likely promotes a heterogeneous RNA architecture.

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 reveal that a four repeat UUAGGG RNA has completely different tertiary shapes in the presence of $K^+$, $Na^+$, or $Li^+$ (Fig. 2B; Supplemental Fig. 2). 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 stacked GQ structures. This result is consistent with previous nuclease digestion studies of TERRA (Martadinata et al. 2011; Martadinata and Phan 2013). Thus, choice of cation and RNA construct enables us to manipulate the tertiary RNA structure and examine the nucleic acid binding properties of LSD1.

### *Stacked GQ RNA preferentially binds to the LSD1–CoREST complex*

In an effort to understand the nucleotide sequence and structural preferences associated with LSD1 and a stable LSD1–CoREST complex (Supplemental Fig. 1), we measured the LSD1–nucleic acid affinities from a panel of single and stacked GQ-forming oligonucleotides under different monovalent ions (Fig. 3; Supplemental Fig. 3). 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 (Zhang et al. 2011) and LSD1 is optimally stabilized in the presence of CoREST (Shi et al. 2004; Forneris et al. 2005, 2008), we have primarily focused on how TERRA interacts with the LSD1–CoREST system.

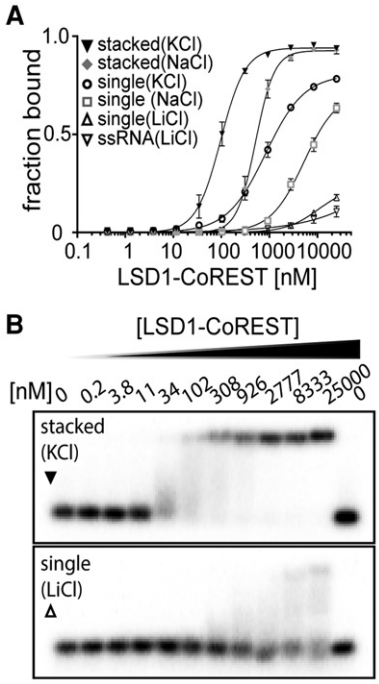


**FIGURE 3.** Affinity and specificity of LSD1–CoREST binding to distinct nucleic acid structures is dependent on monovalent ions. (*A*) Analysis of gel-mobility shift assay binding curves of $(UUAGGG)_8U$, $(UUAGGG)_4U$, and 25-nt ssRNA. Assays were performed in potassium ($K^+$), sodium ($Na^+$), and lithium ($Li^+$) (symbols same as in Fig. 2) using LSD1–CoREST (amino acid residues 171–852 and 286–482, respectively) with exogenous protein purification tags removed. LSD1 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. The dissociation constant ($K_d$) and Hill coefficient ($h$) from this analysis are reported in Table 1. (*B*) Representative gels showing that RNA binding activity of LSD1–CoREST is dependent upon the ability to form a GQ RNA 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 1.** Affinity of LSD1–CoREST binding to GQ-forming RNA

| RNA (monovalent) | Dissociation constant ($K_d$) (nM) | Hill coefficient ($h$) |
|---|---|---|
| 5′-GG(UUAGGG)$_8$U-3′ ($K^+$) | 96.4 ± 3.0 | 1.70 ± 0.08 |
| 5′-GG(UUAGGG)$_8$U-3′ ($Na^+$) | 515.6 ± 14.6 | 2.24 ± 0.10 |
| 5′-(UUAGGG)$_4$U-3′ ($K^+$) | 835.0 ± 34.7 | 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 ± 0.14 |
| 25-nt ssRNA ($Li^+$) | N/A | N/A |

The $K_d$ (in nM) and Hill coefficient ($h$) of the RNA–protein complexes are given. Experimental standard errors of the mean (±) incorporate a range of three independent electrophoretic mobility shift assays (EMSAs).

Using electrophoretic mobility shift assays (EMSAs), incubation of these oligonucleotides with LSD1–CoREST revealed vastly different binding affinities (Table 1; Fig. 3; Supplemental Fig. 3A,B). Analysis of the percent nucleic acid fraction bound 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 (apparent $K_d$ for each oligonucleotide in $K^+$ are 835 nM and 1.3 µM, respectively). The Hill coefficient for an eight repeat UUAGGG RNA–protein interaction ($h \approx 1.7 \pm 0.08$ and $2.24 \pm 0.1$ 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 ($h \approx 1.07 \pm 0.04$) 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, $h = 0.91 \pm 0.17$) and an interaction with the CoREST fragment that is approximately 10-fold weaker (apparent $K_d = 3.4$ µM, $h = 1.40 \pm 0.15$) (Supplemental Fig. 3C). 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 1–385), containing the SWIRM domain and a region of the amine oxidase domain (AOD), was shown to bind a 10 repeat UUAGGG RNA with high affinity ($K_d$~70 nM) (Porro et al. 2014b).

### *GQ RNA can act as an inhibitor of LSD1-catalyzed demethylation*

A G-quadruplex RNA preferentially inhibits LSD1-catalyzed demethylation activity over other nucleic acid structures and this inhibition trend qualitatively corresponds to the apparent equilibrium dissociation constants derived from EMSA data in Figure 3. To examine this inhibition preference, (UUAGGG)$_4$U, the cognate TERRA DNA (TTAGGG)$_4$T, an unstructured 25-nt RNA, and a short 6-nt RNA (UUAGGG) were incubated with the full length LSD1 without CoREST in the presence of a 21 amino acid K4 dimethylated peptide that mimics the H3K4 substrate. Initial velocity measurements were performed using an LSD1-peroxidase-coupled assay (Culhane et al. 2010) and the inhibition profiles were measured for each nucleic acid, enabling the extraction of apparent half maximal inhibitory concentration ($IC_{50}$) values (Supplemental Fig. 4A).

Kinetic measurements of the full-length LSD1-catalyzed demethylation reaction with and without RNA reveal that the (UUAGGG)$_4$U RNA is a reversible, noncompetitive inhibitor of LSD1 activity (Supplemental Fig. 4B). Increasing amounts of the (UUAGGG)$_4$U RNA (0, 1.5, 2.5 µM) show an overall decrease in maximal rate of the chemical reaction ($V_{max}$) but with no or nominal changes in the apparent binding affinity of the catalyst for the substrate ($K_M = 8.0 \pm 0.9$, $6.7 \pm 0.9$, $8.8 \pm 1.5$ µM), respectively. These data suggest that RNA does not compete for substrate binding at the LSD1 active site.

### *Cross-linking mass spectrometry (XL-MS) identifies LSD1–GQ RNA binding interactions*

Previous studies concluded that a GQ RNA likely associates with the SWIRM/AOD of LSD1 (Porro et al. 2014b). 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 cross-link MS data sets indicate a strong GQ RNA cross-link to residues 227–251 of the SWIRM domain (Fig. 4).

As a specificity control, we observed that LSD1–CoREST covalently cross-links with GQ RNA but not a size-matched control RNA (Supplemental Fig. 6). Here, LSD1–CoREST was irradiated with 254 nm UV light either alone (control), in the presence of a 5′ $^{32}$P GQ RNA, or in the presence of a size matched 5′ $^{32}$P 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 3.

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

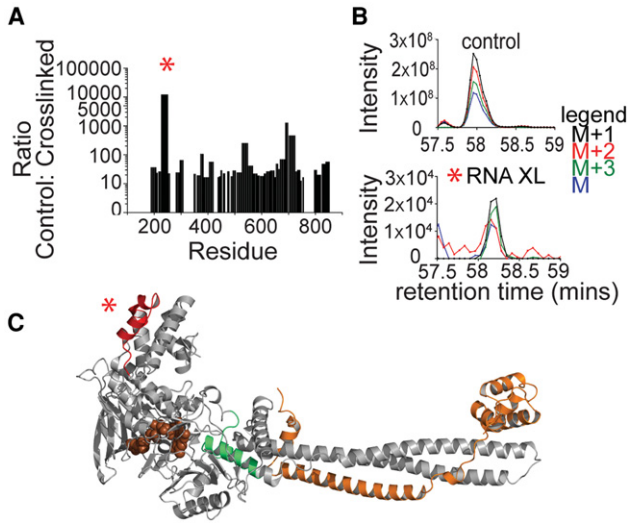


**FIGURE 4.** Identification of the G-quadruplex RNA binding domain of LSD1. The location of the LSD1 region that cross-links with the GQ RNA was detected via mass spectrometry. (*A*) Relative coverage of LSD1 residues upon GQ RNA cross-linking. UV light was used to cross-link biotinylated GQ RNA with 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 GQ RNA binding regions are mapped onto the structure of LSD1–CoREST (PDB 4XBF). Two distinct regions of LSD1 appear to cross-link with GQ RNA. The primary GQ RNA cross-link is located within the SWIRM domain (red) (residues 227–251, 210–216) and a minor RNA–LSD1 adduct region exists adjacent to the active site and FAD (brown spheres) and close to the C-terminal domain of CoREST (green) (residues 527–550). Additional GQ RNA–LSD1 cross-link locations are noted (Results and Supplemental Material) but were not consistent between the two separate and independent cross-link-MS experiments.

SDS-PAGE purified, in-gel trypsin digested, and analyzed by mass spectrometry as described in the Materials and Methods section. The cross-linked peptides of LSD1 were deduced by the disappearance of the original peptide signal as the RNA–LSD1 adduct altered the *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.

Our analysis included peptide regions that could be detected, 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 (Fig. 4). 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.

### *The structure of LSD1–CoREST bound to UUAGG identifies an ssRNA binding site adjacent to the primary GQ RNA–LSD1 cross-link location*

To examine how RNA may inhibit LSD1 activity and identify an RNA binding location on LSD1, cocrystallization screening and extensive soaking experiments were performed with the LSD1–CoREST complex and different RNA molecules. The 53- and 25-nt GQ RNA and a series of short oligonucleotides (UUAGG, UUUUU, CCCUAA, UUAGGAG, AAAAAGCAAA, and GGUUUUUCUUUU) were soaked into crystals at various concentrations (0.1–5.0 mM) and incubation periods (4–24 h). Cocrystallization and crystal soaking trials of the GQ RNAs with LSD1–CoREST were unsuccessful and only crystal soaking experiments with UUAGG and UUAGGAG ssRNA fragments showed difference maps ($|F_o|-|F_c|$) corresponding to oligonucleotide electron density (Fig. 5). From the UUAGG ssRNA soaked LSD1–CoREST crystals, the difference Fourier map ($|F_o|-|F_c|$) identified an RNA binding location on LSD1 and the directionality of the RNA. A 2.8 Å resolution map allowed us to accurately model build and define the ssRNA structure using RCrane and Coot (Emsley and Cowtan 2004; Keating and Pyle 2012) ($R_{work}$ = 20.1% $R_{free}$ = 22.7%, PDB 4XBF, Supplemental Table 1). In addition to the location of an ssRNA binding site on LSD1, the quality of the difference map ($|F_o|-|F_c|$) revealed well-ordered regions that correspond to FAD, sulfate ions, and a previously identified glycerol molecule (Fig. 5; Supplemental Fig. 7; Luka et al. 2011). The structure of LSD1–CoREST bound to UUAGG does not overlap with the GQ RNA binding location in the XL-MS experiments; rather, it is positioned adjacent to the GQ RNA binding region.

### *Nucleobase RNA recognition by the conserved amine oxidase domain (AOD) of LSD1*

The structure of the LSD1–CoREST complex with ssRNA reveals that UUAGG RNA binds to the AOD of LSD1, which is located opposite of the histone 3 peptide binding cleft (Fig. 5). This RNA binding site is positioned at a conserved and "intertwined" region of the AOD, containing both N- and C-terminal regions (residues 278–280 and 615–619).

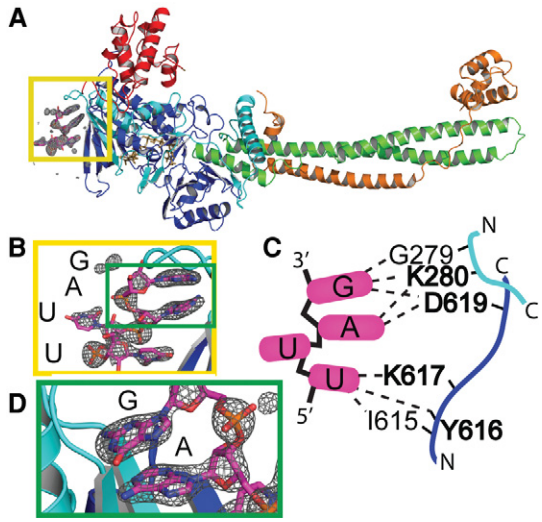


**FIGURE 5.** Location of an ssRNA binding region on LSD1. (*A*) Analysis of crystal soaks of a 5′-UUAGG-3′ RNA ligand into LSD1–CoREST crystals shows clear difference electron density ($|F_o|-|F_c|$) (dark gray mesh contoured at 5.5 r.m.s.d). RNA–LSD1 interactions occur along β-sheet and loop regions of LSD1 (yellow box). LSD1 and CoREST as colored in Figure 1B. (*B*) The 2.8 Å resolution map reveals the unambiguous identity and directionality of the RNA (pink), shown in stick representation. The difference electron density ($|F_o|-|F_c|$) is noted (dark gray mesh contoured at 5.5 r.m.s.d) and nucleobase density was observed for UUAG nucleotides. (*C*) Schematic of noncovalent interactions between the ssRNA fragment and LSD1. Dotted lines indicate RNA–LSD1 residues within 3.0 Å (as summarized in Supplemental Table 2), with conserved LSD1 residues (bold) spanning the monoamine oxidase domain (cyan/blue). (*D*) A close-up view of the difference density ($|F_o|-|F_c|$) shows the unambiguous assignment of purine (A, G) nucleobases.

Interestingly, the UUAGG RNA does not electrostatically bind to the various electropositive grooves that exist across LSD1 which are surface accessible within the crystal lattice environment. Rather, the RNA nucleobases are oriented toward a cleft that is comprised of the β-sheet interface and three loop regions (Lys276-Lys280, Ala597-Asn599, Pro782-Arg795). Unfortunately, point mutations of Lys280, Tyr616, Lys617, and Asp619 to alanine resulted in a loss of recombinant protein expression and protein stability, making it difficult to assess the effect of RNA binding activity within this region of LSD1. Taken together, the RNA soaking studies identify how LSD1 may interact with the single-stranded linker regions of TERRA, supporting an extended TERRA binding interface that spans the SWIRM and AOD of LSD1.

The identification of an ssRNA binding region and the orientation of the RNA also demonstrate that LSD1 can interact with nucleic acids through the formation of sequence specific interactions. The most extensive RNA–protein contacts occur between the AG dinucleotide region and LSD1 residues that are part of a preformed binding pocket (Fig. 5C). This single-stranded oligonucleotide binding region of LSD1 may correlate with a genome-wide analysis of nucleotide binding elements of the LSD1 complex, which originally proposed that an AG dinucleotide region is part of an enriched motif in LSD1 binding (Tsai et al. 2010). Together with our kinetic assays (Supplemental Fig. 4), the structure provides additional support that RNA can serve as a noncompetitive allosteric inhibitor of LSD1 function.

## DISCUSSION

TERRA is an integral component of telomeric heterochromatin, 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 (Luke et al. 2008; Schoeftner and Blasco 2008; Deng et al. 2009; Xu et al. 2010; Iglesias et al. 2011; Lopez de Silanes et al. 2014; Porro et al. 2014a,b; Azzalin and Lingner 2015). Despite a correlation between histone H3K4/K9 methylation patterns and gene silencing at telomeres (Krogan et al. 2002; Vaquero-Sedas et al. 2012; Porro et al. 2014b), it is unclear how these enzymes are recruited to telomeres and how they might function to mediate telomere structure and composition. TERRA is thought to act as a molecular decoy to sequester telomerase in a cell cycle-dependent manner and also likely serves as a scaffold with epigenetic regulators during the early stages of DNA damage response activation (Redon et al. 2010; Porro et al. 2014a,b). The unique topology of TERRA (Fig. 1C) suggests that its structure may enable the organization of higher-order RNP complexes at telomeres. To better understand how lncRNAs may recruit protein enzymes and how TERRA might regulate protein–protein interactions 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 (Fig. 2; Supplemental Fig. 2). The formation of these diverse RNA structures correlates well with observed nucleic acid–LSD1 binding affinities (Fig. 3; Supplemental Fig. 3), 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 UUAGGG repeat units (Martadinata et al. 2011; Martadinata and Phan 2013). It appears that a single ($[UUAGGG]_4U$) GQ unit, a stacked ($[UUAGGG]_8U$) GQ unit, and a ssRNA spacer region between GQ units represent the three unique topologies adopted by TERRA (Martadinata et al. 2011; Martadinata and Phan 2013). In addition, biophysical studies of TERRA demonstrate that the 2′ –OH functional groups in the RNA G-quadruplex participate in organizing water hydration and in the hydrogen-bonding network (Haider et al. 2011; Martadinata et al. 2011). 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 (Supplemental Fig. 3), which is known to contain a mixture of heterogeneous

G-quadruplex topologies (Patel et al. 2007). Our data confirm that single and stacked GQ RNA structural units serve as building blocks of the extended TERRA 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), the glycine-arginine-rich (GAR) domain of TRF2, and the N-terminal domain of the DEAH-box ATP-dependent helicase 36 (DHX36) (Deng et al. 2009; Meier et al. 2013; Chen et al. 2015; Vasilyev et al. 2015). With the development of GQ-specific antibodies that have unambiguously identified GQ TERRA RNA structures in living cells (Xu et al. 2010; Di Antonio et al. 2012; Biffi et al. 2013, 2014), more biophysical studies are needed that probe how protein domains and multidomain protein complexes recognize the molecular architecture of GQ RNAs. Our data have defined how TERRA's secondary and tertiary structural motifs can serve as recognition elements for LSD1 interactions by coupling biophysical studies (CD/AUC), EMSA, enzyme kinetics, XL-MS, and X-ray crystallographic methods. Results from our studies suggest that TERRA could serve as an allosteric effector of LSD1 function at telomeres, as previously proposed (Porro et al. 2014b).

## Functional implications for allostery in LSD1–RNA interactions

LSD1 is associated with CoREST and together they act as an allosteric clamp on nucleosomes to catalyze specific histone H3K4/K9 demethylation (Shi et al. 2004; Chen et al. 2006; Stavropoulos et al. 2006; Yang et al. 2006; Forneris et al. 2007; Baron and Vellore 2012). Motions of the SWIRM domain and rotation of the AOD of LSD1 must occur when the substrate enters the active site pocket (Baron and Vellore 2012) and it has been demonstrated that both substrate binding and protein–protein interactions modulate LSD1 conformational dynamics and activity (Shi et al. 2005; Forneris et al. 2007).

Our data suggest that GQ-forming RNAs can provide an additional layer of allosteric regulation in LSD1 function. Using XL-MS, we identified a GQ RNA binding site distal to the H3 binding cleft. X-ray crystallographic soaking experiments of LSD1–CoREST with various ssRNAs suggest that the RNA binding interface spans the amine-oxidase domain (Fig. 6). While it is possible that there may be multiple regions of LSD1–CoREST that contact nucleic acids, the identified primary RNA binding interface likely extends from the SWIRM domain to the AOD of LSD1. Interestingly, the structural integrity and flexibility of the SWIRM-AOD is essential for recognition of nucleosomal DNA (Stavropoulos et al. 2006; Yang et al. 2006; Forneris et al. 2007). In fact, our identified GQ RNA binding region within the SWIRM domain (amino acids 227–251) overlaps with the putative nucleosomal DNA binding surface (Yang et al. 2006; Pilotto et al. 2015), suggesting that both nucleosomal DNA and TERRA RNA bind a similar SWIRM/AOD recognition interface. A putative mechanism whereby TERRA alters LSD1 demethylation activity and modulates nucleosome binding or DNA damage response activation at telomeres is consistent with our RNA–LSD1 structural and enzyme inhibition results.

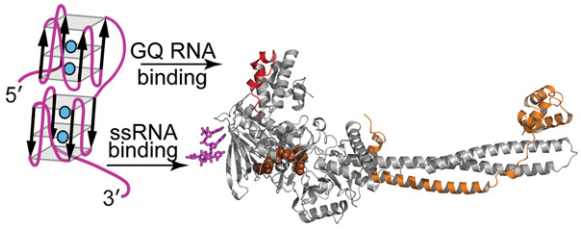


**FIGURE 6.** Model of the TERRA RNA–LSD1 binding interface. Cross-linking mass spectrometry data support a model whereby LSD1 residues (red) in the SWIRM domain interact with a stacked GQ RNA. In addition, X-ray crystallography data identifies an ssRNA binding location (pink), which resides adjacent to the GQ RNA binding site. Binding locations are consistent with a GQ RNA that binds LSD1 in a noncompetitive manner.

Whereas the biological consequences of LSD1 activity are known, mechanistic examples of RNA-mediated LSD1 recruitment (Khalil et al. 2009; Tsai et al. 2010; Porro et al. 2014b; Hendrickson et al. 2016) and the modulation of LSD1 activity by noncoding RNAs remain unclear. 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 (Cusanelli and Chartrand 2015; Flynn et al. 2015; Rippe and Luke 2015). It is possible that RNA helicases provide an added level of epigenetic regulation by unwinding GQ RNAs, resulting in a loss of LSD1 binding and effectively relieving RNA-mediated LSD1 inhibition. 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 chromatin modifying enzymes by ncRNAs (Cifuentes-Rojas et al. 2014; Kaneko et al. 2014; Sarma et al. 2014; Davidovich et al. 2015; Cloutier et al. 2016).

In conclusion, we have identified a primary GQ RNA binding site within the conserved SWIRM/AOD interface of LSD1 and demonstrated that a GQ-forming TERRA can function as a noncompetitive inhibitor of LSD1-catalyzed demethylation. 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 at the atomic level.

## MATERIALS AND METHODS

### LSD1, CoREST, and RNA preparation

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) (Supplemental Fig. 1A; Shi et al. 2004; Culhane et al. 2006). The full size and truncated LSD1 were expressed in *Escherichia coli* BL21 (DE3) as previously reported (Shi et al. 2004; Culhane et al. 2006, 2010). The full-size LSD1 was purified by using ammonium sulfate fractionation and anion-exchange chromatography (Luka et al. 2011). 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 (Luka et al. 2011). CoREST is a known stabilizing element of LSD1 (Shi et al. 2005) and was expressed in *E. coli* BL21 (DE3) in Luria Broth (LB) (kanamycin) and purified using Ni-NTA agarose as previously reported (Luka et al. 2011). For EMSA assays and crystallization trials, 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 with 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 (Supplemental Fig. 1B–D).

Stacked GQ RNAs (GG[UUAGGG]$_8$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 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]$_4$U), a 25-nt single-stranded RNA (5′-UUAGCGUUAACCUUACCAUU ACGUU-3′) (ssRNA), and all RNAs for crystallography, including a 5′-UUAGG-3′ RNA ligand, were purchased from Thermo-Fisher (Dharmacon) and deprotected according to manufacturer's instructions. The identical DNA ([TTAGGG]$_4$T) (GQ DNA) and the 5′ biotinylated RNAs for cross-linking (GUUU[UUAGGG]$_4$ UUA and GGG[UUAGGG]$_8$UUA) 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.

### 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. Standard 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 [θ/(1−θ)] versus log of the protein concentration, where θ is the fraction of bound oligonucleotide.

### *Activity assay*

LSD1 activity assays were performed using a peroxidase-coupled assay under aerobic conditions as described previously (Forneris et al. 2005). A 150 µL reaction mixture contained 50 mM HEPES(Na), pH 7.5, 0.2–0.3 µM LSD1, 1 µg of HRP, 0.1 mM 4-aminoantipyrine, 1.0 mM 3,5-dichloro-2-hydroxybenzenesulfonic acid and substrate concentrations as indicated in the main text. Enzyme reactions were initiated by the addition of substrate (dimethyl-Lys4 Histone H3 peptide aa 1–21 [diMeK4H3$_{1-21}$] [Sigma]) and the time course of the reaction was monitored using a Shimadzu 2401-PC spectrophotometer (λ = 515 nm) at 25°C in a thermostated quartz cell chamber. Initial velocity values were measured using an extinction coefficient of 26,000 $M^{-1}$ $cm^{-1}$ (Forneris et al. 2005). Kinetic data were fitted to a Michaelis–Menten equation using GraphPad Prism 6.0 (GraphPad Software, Inc., http://www.graphpad.com).

### *Inhibition studies*

The $IC_{50}$ values for different oligonucleotides were obtained via the LSD1 activity assay in the presence of different concentrations of RNA/DNA as detailed (Supplemental Fig. 4). Control experiments showed that the peroxidase components of this coupled assay were not influenced or altered by the addition of nucleic acids. The mechanism of inhibition was examined with velocity/substrate concentration curves at different concentrations of RNA in the reaction mixture as detailed in the figure legend (Supplemental Fig. 4).

### *Cross-linking mass spectrometry*

To verify that LSD1–CoREST cross-links to GQ RNA in a specific manner, an (GG[UUAGGG]$_8$UUA) RNA 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 (Supplemental Fig. 6).

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 (Schilling et al. 2012).

### *Crystallization of LSD1–CoREST complex*

The LSD1–CoREST complex was crystallized as previously described (Yang et al. 2006; Luka et al. 2011). Briefly, the LSD1–CoREST complex was prepared by mixing the LSD1 and CoREST stock solutions in a 1:1.5 molar ratio of LSD to CoREST and removal of excess CoREST by using an Amicon Ultra 50K Centrifugal Filters after 1-h incubation.

The LSD1–CoREST complex was crystallized by the hanging drop or sitting drop method at room temperature as previously described (Luka et al. 2011). The LSD1–CoREST complex (10–12 mg/mL concentration in 25 mM HEPES-Na, pH 7.4, 100 mM NaCl, 5 mM DTT, 1 mM PMSF) was mixed with the reservoir solution [0.60 M $Li_2SO_4$, 0.63 M $(NH_4)_2SO_4$, 0.25 M NaCl, 100 mM Na-citrate, pH 5.6, 10 mM DTT]. The crystals belong to the orthorhombic *I222* space group ($a$ = 123.86 Å, $b$ = 179.37 Å, $c$ = 235.05 Å).

A 5′-UUAGG-3′ RNA ligand was introduced into the crystals by the soaking method (Hassell et al. 2007; Reiter et al. 2010). Other oligonucleotides were tested in crystal soaking trials and include: the 53- and 25-nt GQ RNA, as well as UUAGG, UUUUU, CCCUAA, UUAGGAG, AAAAAGCAAA, and GGUUUUUCUUUU RNAs. The 53- and 25-nt GQ RNAs proved to be unsuccessful in crystal soaking trials due to the size of the RNAs and because the reservoir solution contained high salt or high concentrations of $Li^+$ ions that inhibit GQ formation (Hardin et al. 2000). Aside from UUAGG and UUAGGAG soaked crystals, no other single-stranded RNAs yielded any noticeable difference electron density upon careful analysis of the Fourier map ($|F_o|-|F_c|$). The LSD1–CoREST complex crystals were incubated for 3 h or overnight in the cryoprotectant containing UUAGG RNA [0.76 M $Li_2SO_4$, 0.74 M $(NH_4)_2SO_4$, 0.35 M NaCl, 100 mM Na-citrate, pH 5.6, 23% (v/v) glycerol and 2–5 mM RNA]. After soaking, crystals were harvested and flash-cooled in liquid nitrogen.

### *X-ray data collection*

Diffraction data were collected at 100 K at LS-CAT beamline 21 (F hutch), Advanced Photon Source, Argonne National Laboratory, using a MarMosaic 225 CCD detector. Data were processed and scaled using the HKL2000 package (Otwinowski and Minor 1997). Data collection and data processing statistics are summarized in Supplemental Table 1.

### *Crystal structure determination*

Molecular replacement was applied to locate a solution using PHENIX and a previously determined structure of the LSD1–CoREST complex (PDB 2IW5) (Afonine et al. 2012). The difference Fourier map ($|F_o|-|F_c|$) revealed the presence of RNA and clear difference density enabled us to unambiguously determine the directionality of four nucleotides of the RNA (Fig. 5). While a 3′ phosphate of G5 is observed, the 3′ G nucleobase is not observed in the density. Coot was used for model building throughout the refinement and the RNA was built using RCrane (Emsley and Cowtan 2004; Keating and Pyle 2012). The final model consists of residues 171–836 of LSD1, residues 308–440 of CoREST, one FAD molecule, one RNA molecule, 5 sulfate ions, 1 glycerol molecule, and 23 water molecules. Refinement statistics are listed in Supplemental Table 1 (PDB ID 4XBF).

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## ACKNOWLEDGMENTS

We thank Conrad Wagner for enthusiastic discussions, M. Ascano, M. Egli, F.P. Guengerich, and C. Sanders for critical reading of the manuscript, J. Harp for crystallographic assistance, H. MacDonald for mass spectrometry assistance, the Vanderbilt University Center for Structural Biology, and the Vanderbilt Department of Biochemistry. N.J.R. acknowledges ACS-IRG (58-0009-54) and AHA (14GRNT20380334) for support. Z.L. acknowledges the National Institutes of Health for support (DK015289, PI C. Wagner). W.J.M. is a Molecular Biophysics trainee (2T32GM008320-26). We thank Philip A. Cole and Yang Shi for the generous gifts of plasmids for LSD1 and CoREST expression. Use of APS and LSCAT-21 was supported by the US Department of Energy (DE-AC02-06CH11357), the Michigan Economic Development Corporation, and the Michigan Technology Tri-Corridor (grant 085P1000817).

*Author contributions*: A.H., Z.L., W.J.M., and N.J.R. conceived the project. A.H. and W.J.M. performed RNA biophysical studies. A.H. and N.J.R. performed and designed EMSA experiments. Z.L. and N.J.R. performed all enzyme assays. W.J.M. performed all cross-linking studies and analyzed mass spectrometry data. A.H., Z.L., L.V.L., and N.J.R performed the X-ray crystallographic studies. All authors contributed to writing the manuscript.

*Received May 2, 2016; accepted May 5, 2016.*

## REFERENCES

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biological Crystallogr* **68:** 352–367.

Amente S, Lania L, Majello B. 2013. The histone LSD1 demethylase in stemness and cancer transcription programs. *Biochim Biophys Acta* **1829:** 981–986.

Azzalin CM, Lingner J. 2015. Telomere functions grounding on TERRA firma. *Trends Cell Biol* **25:** 29–36.

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

Baron R, Vellore NA. 2012. LSD1/CoREST is an allosteric nanoscale clamp regulated by H3-histone-tail molecular recognition. *Proc Natl Acad Sci* **109:** 12509–12514.

Biffi G, Tannahill D, McCafferty J, Balasubramanian S. 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **5:** 182–186.

Biffi G, Di Antonio M, Tannahill D, Balasubramanian S. 2014. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat Chem* **6:** 75–80.

Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, et al. 2013. The BioGRID interaction database: 2013 update. *Nucleic Acids Res* **41:** D816–D823.

Chen Y, Yang Y, Wang F, Wan K, Yamane K, Zhang Y, Lei M. 2006. Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc Natl Acad Sci* **103:** 13956–13961.

Chen MC, Murat P, Abecassis K, Ferre-D'Amare AR, Balasubramanian S. 2015. Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res* **43:** 2223–2231.

Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT. 2014. Regulatory interactions between RNA and polycomb repressive complex 2. *Mol Cell* **55:** 171–185.

Cloutier SC, Wang S, Ma WK, Al Husini N, Dhoondia Z, Ansari A, Pascuzzi PE, Tran EJ. 2016. Regulated formation of lncRNA-DNA hybrids enables faster transcriptional induction and environmental adaptation. *Mol Cell* **61:** 393–404.

Culhane JC, Szewczuk LM, Liu X, Da G, Marmorstein R, Cole PA. 2006. A mechanism-based inactivator for histone demethylase LSD1. *J Am Chem Soc* **128:** 4536–4537.

Culhane JC, Wang D, Yen PM, Cole PA. 2010. Comparative analysis of small molecules and histone substrate analogues as LSD1 lysine demethylase inhibitors. *J Am Chem Soc* **132:** 3164–3176.

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

Davidovich C, Wang X, Cifuentes-Rojas C, Goodrich KJ, Gooding AR, Lee JT, Cech TR. 2015. Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol Cell* **57:** 552–558.

Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. 2009. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell* **35:** 403–413.

Di Antonio M, Biffi G, Mariani A, Raiber EA, Rodriguez R, Balasubramanian S. 2012. Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. *Angew Chem Int Ed Engl* **51:** 11073–11078.

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr* **D60:** 2126–2132.

Engreitz JM, Sirokman K, McDonel P, Shishkin AA, Surka C, Russell P, Grossman SR, Chow AY, Guttman M, Lander ES. 2014. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **159:** 188–199.

Flynn RL, Cox KE, Jeitany M, Wakimoto H, Bryll AR, Ganem NJ, Bersani F, Pineda JR, Suva ML, Benes CH, et al. 2015. Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science* **347:** 273–277.

Forneris F, Binda C, Vanoni MA, Battaglioli E, Mattevi A. 2005. Human histone demethylase LSD1 reads the histone code. *J Biol Chem* **280:** 41360–41365.

Forneris F, Binda C, Adamo A, Battaglioli E, Mattevi A. 2007. Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J Biol Chem* **282:** 20070–20074.

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

Haider SM, Neidle S, Parkinson GN. 2011. A structural analysis of G-quadruplex/ligand interactions. *Biochimie* **93:** 1239–1251.

Hardin CC, Perry AG, White K. 2000. Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* **56:** 147–194.

Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL III, Deng SJ, Gampe RT, Grisard TE, Madauss KP, Nolte RT, et al. 2007. Crystallization of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **63:** 72–79.

Hendrickson DG, Kelley DR, Tennen D, Bernstein B, Rinn JL. 2016. Widespread RNA binding by chromatin-associated proteins. *Genome Biol* **17:** 28.

Hwang S, Schmitt AA, Luteran AE, Toone EJ, McCafferty DG. 2011. Thermodynamic characterization of the binding interaction between the histone demethylase LSD1/KDM1 and CoREST. *Biochemistry* **50:** 546–557.

Iglesias N, Redon S, Pfeiffer V, Dees M, Lingner J, Luke B. 2011. Subtelomeric repetitive elements determine TERRA regulation by Rap1/Rif and Rap1/Sir complexes in yeast. *EMBO Rep* **12:** 587–593.

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

Keating KS, Pyle AM. 2012. RCrane: semi-automated RNA model building. *Acta Crystallogr D Biol Crystallogr* **68:** 985–995.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci* **106:** 11667–11672.

Kooistra SM, Helin K. 2012. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol* **13:** 297–311.

Krogan NJ, Dover J, Khorrami S, Greenblatt JF, Schneider J, Johnston M, Shilatifard A. 2002. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* **277:** 10753–10755.

Laurent B, Ruitu L, Murn J, Hempel K, Ferrao R, Xiang Y, Liu S, Garcia BA, Wu H, Wu F, et al. 2015. A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. *Mol Cell* **57:** 957–970.

Lopez de Silanes I, Grana O, De Bonis ML, Dominguez O, Pisano DG, Blasco MA. 2014. Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nat Commun* **5:** 4723.

Luka Z, Moss F, Loukachevitch LV, Bornhop DJ, Wagner C. 2011. Histone demethylase LSD1 is a folate-binding protein. *Biochemistry* **50:** 4750–4756.

Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J. 2008. The Rat1p 5′ to 3′ exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. *Mol Cell* **32:** 465–477.

Martadinata H, Phan AT. 2013. Structure of human telomeric RNA (TERRA): stacking of two G-quadruplex blocks in K(+) solution. *Biochemistry* **52:** 2176–2183.

Martadinata H, Heddi B, Lim KW, Phan AT. 2011. Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. *Biochemistry* **50:** 6455–6461.

Meier M, Patel TR, Booy EP, Marushchak O, Okun N, Deo S, Howard R, McEleney K, Harding SE, Stetefeld J, et al. 2013. 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.

Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. In *Methods in enzymology* (ed. Carter CW, Sweet RM), Vol. 276, pp. 307–326. Academic Press, New York.

Patel DJ, Phan AT, Kuryavyi V. 2007. 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.

Pilotto S, Speranzini V, Tortorici M, Durand D, Fish A, Valente S, Forneris F, Mai A, Sixma TK, Vachette P, et al. 2015. Interplay among nucleosomal DNA, histone tails, and corepressor CoREST underlies LSD1-mediated H3 demethylation. *Proc Natl Acad Sci* **112:** 2752–2757.

Porro A, Feuerhahn S, Delafontaine J, Riethman H, Rougemont J, Lingner J. 2014a. Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat Commun* **5:** 5379.

Porro A, Feuerhahn S, Lingner J. 2014b. TERRA-reinforced association of LSD1 with MRE11 promotes processing of uncapped telomeres. *Cell Rep* **6:** 765–776.

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

Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragon A. 2010. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468:** 784–789.

Reiter NJ, Chan CW, Mondragon A. 2011. Emerging structural themes in large RNA molecules. *Curr Opin Struct Biol* **21:** 319–326.

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

Sarma K, Cifuentes-Rojas C, Ergun A, Del Rosario A, Jeon Y, White F, Sadreyev R, Lee JT. 2014. ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell* **159:** 869–883.

Schilling B, Rardin MJ, MacLean BX, Zawadzka AM, Frewen BE, Cusack MP, Sorensen DJ, Bereman MS, Jing E, Wu CC, et al. 2012. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol Cell Proteomics* **11:** 202–214.

Schoeftner S, Blasco MA. 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* **10:** 228–236.

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119:** 941–953.

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

Somarowthu S, Legiewicz M, Chillon I, Marcia M, Liu F, Pyle AM. 2015. HOTAIR forms an intricate and modular secondary structure. *Mol Cell* **58:** 353–361.

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

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329:** 689–693.

Vaquero-Sedas MI, Luo C, Vega-Palas MA. 2012. Analysis of the epigenetic status of telomeres by using ChIP-seq data. *Nucleic Acids Res* **40:** e163.

Vasilyev N, Polonskaia A, Darnell JC, Darnell RB, Patel DJ, Serganov A. 2015. Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP. *Proc Natl Acad Sci* **112:** E5391–E5400.

Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43:** 904–914.

Xu Y, Suzuki Y, Ito K, Komiyama M. 2010. Telomeric repeat-containing RNA structure in living cells. *Proc Natl Acad Sci* **107:** 14579–14584.

Yang M, Gocke CB, Luo X, Borek D, Tomchick DR, Machius M, Otwinowski Z, Yu H. 2006. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol Cell* **23:** 377–387.

Zhang P, Casaday-Potts R, Precht P, Jiang H, Liu Y, Pazin MJ, Mattson MP. 2011. Nontelomeric splice variant of telomere repeat-binding factor 2 maintains neuronal traits by sequestering repressor element 1-silencing transcription factor. *Proc Natl Acad Sci* **108:** 16434–16439.


# G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme

Alexander Hirschi, William J. Martin, Zigmund Luka, et al.

*RNA* 2016 22: 1250-1260 originally published online June 8, 2016
Access the most recent version at doi:10.1261/rna.057265.116