Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

ACS Chem Biol. Author manuscript; available in PMC 2019 August 26.

Published in final edited form as:

ACS Chem Biol. 2019 February 15; 14(2): 223–235. doi:10.1021/acschembio.8b00807.

Selective Small-Molecule Targeting of a Triple Helix Encoded by the Long Noncoding RNA, *MALAT1*

Fardokht A. Abulwerdi[†,∇], **Wenbo Xu**[‡,§,∇], **Abeer A. Ageeli**[‖,∇], **Michael J. Yonkunas**[‖], **Gayatri Arun**[‡], **Hyeyeon Nam**[⊥], **John S. Schneekloth Jr.**[#], **Theodore Kwaku Dayie**[⊥], **David Spector**[‡], **Nathan Baird**[*,‖], **Stuart F. J. Le Grice**[*,†]

[†]Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702, United States

[‡]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, United States

[§]Stony Brook University, Molecular and Cellular Biology Program, Stony Brook, New York 11794, United States

[‖]University of the Sciences, 600 South 43rd Street, Philadelphia, Pennsylvania 19104, United States

[⊥]Center for Biomolecular Structure and Organization, Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, United States

[#]Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702, United States

Abstract

Metastasis-associated lung adenocarcinoma transcript 1 (*Malat1*/*MALAT1*, mouse/human), a highly conserved long noncoding (lnc) RNA, has been linked with several physiological processes, including the alternative splicing, nuclear organization, and epigenetic modulation of gene expression. *MALAT1* has also been implicated in metastasis and tumor proliferation in multiple cancer types. The 3′ terminal stability element for nuclear expression (ENE) assumes a triple-helical configuration that promotes its nuclear accumulation and persistent function. Utilizing a

[*]Corresponding Authors: legrices@mail.nih.gov. n.baird@usciences.edu.

Author Contributions

F.A.A. designed and performed the SMM screen and dye displacement assay and analyzed data. W.X. and G.A. designed and performed the organoid and RT-qPCR experiments and analyzed data. A.A.A. designed and performed FRET, DS-FRET, and ITC experiments and analyzed data. M.J.Y. performed the molecular docking studies and analyzed data. H.N. performed the STD NMR experiments and analyzed data. J.S.S., Jr. supervised the SMM screen. T.K.D. supervised and directed NMR experiments. N.B. supervised and directed the biophysical and molecular docking studies. D.S. supervised and directed the biological studies. S.F.J.L. supervised and directed the screening and biochemical studies. F.A.A., N.B., and S.F.J.L. prepared the manuscript with input from W.X., G.A., A.A.A, M.J.Y., and D.S.

[∇]These authors contributed equally to this work.

Supporting Information

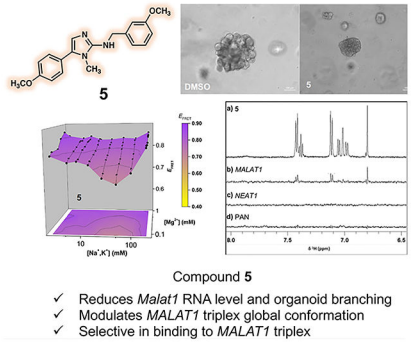
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.8b00807. Figures showing scanned images of all SMM hit compounds along with their chemical structures, branching morphogenesis data for all 28 hits, titration binding curve of **16** by FRET using $M1^{ET}$ construct, conformational landscape of $M1^{AB}$ with **5** and **16**. Triplex thermal stability landscapes of $M1^{ET}$ and $M1^{AB}$ with **5** and **16**, binding affinity curves of **5** by FRET and DS-FRET, additional docking studies with **5** and **16**, dye displacement assay binding curves, Tables with binding affinities of **5** and **16** from all assays, thermodynamic binding parameters by ITC and the range of fit in the ITC parameters for **5** and **16** (PDF)

The authors declare no competing financial interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

novel small molecule microarray strategy, we identified multiple *Malat1* ENE triplex-binding chemotypes, among which compounds 5 and 16 reduced *Malat1* RNA levels and branching morphogenesis in a mammary tumor organoid model. Computational modeling and Förster resonance energy transfer experiments demonstrate distinct binding modes for each chemotype, conferring opposing structural changes to the triplex. Compound 5 modulates *Malat1* downstream genes without affecting *Neat1*, a nuclear lncRNA encoded in the same chromosomal region as *Malat1* with a structurally similar ENE triplex. Supporting this observation, the specificity of compound 5 for *Malat1* over *Neat1* and a virus-coded ENE was demonstrated by nuclear magnetic resonance spectroscopy. Small molecules specifically targeting the *MALAT1* ENE triplex lay the foundation for new classes of anticancer therapeutics and molecular probes for the treatment and investigation of *MALAT1*-driven cancers.

Graphical Abstract



Long noncoding RNAs (lncRNAs) have diverse biological functions associated with various disease states, including cancer.[1] Metastasis-associated lung adenocarcinoma transcript 1 (*Malat1/MALAT1*, mouse/human) is a highly conserved ~8 kb lncRNA linked with several physiological processes, including the alternative splicing, nuclear organization, and epigenetic modulating of gene expression.[2] *Malat1* is also an up-regulated transcript in nonsmall cell lung tumors with a high propensity for metastasis[3] and subsequently associated with poor survival in patients with various tumor types.[4] Depleting *Malat1* with antisense oligonucleotides (ASOs)[4,5] or gene knockdown strategies[6,7] show promising antiproliferative effects. A mammary tumor model was recently developed by breeding *Malat1*-knockout mice with MMTV-PyMT mice expressing a strong oncogene in mammary tissue.[5] The resulting offspring, expressing normal levels of *Malat1*, had poorly differentiated, aggressive mammary carcinomas that were prone to metastasizing to the lung. In contrast, *Malat1*-deficient offspring developed cystic tumors with poor metastatic potential, highlighting a key role for *Malat1* in the progression of breast cancer. Treatment of MMTV-PyMT mice with ASOs targeting *Malat1* resulted in slower tumor growth and significantly reduced metastasis, underscoring its potential for therapeutic intervention. Furthermore, the depletion of *Malat1* was not lethal to cells grown in culture, and its knockout in mice did not impair normal growth and development[8,9] suggesting that *Malat1* inhibition has minimal side effects in normal tissue.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA Pol II-transcribed *MALAT1* is post-transcriptionally processed by nuclear machinery that modifies tRNA but is not exported to the cytoplasm.[10] The RNA is not polyadenylated but instead contains a highly unusual 3′-terminal structural motif designated the stability element for nuclear expression (ENE).[11] This comprises a U-rich hairpin with which a 3′-genomically encoded A-rich tract interacts to form a triple helix (Figure 1a) that both protects *MALAT1* from degradation and promotes nuclear retention.[12–14] Although the mechanism of triplex-mediated protection remains unknown, a recent study demonstrated that the degree of *MALAT1* nuclear accumulation is determined by a balance between ENE triple helix formation and RNA decay.[15] Modulation of triple helix dynamics might therefore be predicted to influence *MALAT1* lncRNA half-life and *MALAT1*-driven metastasis in a variety of cancers.

Small molecules offer an opportunity to target structured RNA motifs[16–20] e.g pseudoknots, bulges, and hairpins, which are often highly conserved and related to RNA function. The pharmacokinetics of small molecules are generally superior to those of oligonucleotides, whose anionic character limits cell permeability and systemic distribution.[20] Several recent examples have been cited where *in vitro* and *in silico* high-throughput screening (HTS) have identified small molecules that selectively bind RNA motifs, with corresponding physiological effects in cell culture or animal models.[21–23] The relationship among triple helix dynamics, *MALAT1* nuclear accumulation, and breast cancer metastasis suggests the ENE triplex as an excellent candidate for small molecule targeting, provided that specificity can be achieved. Recently, Donlic et al. have identified small molecule binders of *MALAT1* ENE and characterized them in *in vitro* assays;[24] however, the biological outcome of this binding has not been studied in cell-based assays.

As an exploratory step in developing a therapeutic strategy, we utilized an HTS approach to identify ligands specific to the mouse *Malat1* ENE triplex, which shares ~90% of its identity with the human *MALAT1* ENE triplex.[25] Herein, we report the ability of two such molecules to reduce *Malat1* expression in culture and branching morphogenesis in a mammary tumor organoid model. These compounds were further demonstrated to bind the ENE triplex *via* orthongonal biophysical assays, while computational molecular docking implies distinct binding modes. The more potent compound 5 modulated *Malat1* downstream genes in a dose-dependent manner without affecting expression of nuclear enriched abundant transcript 1 (*Neat1*), a nuclear lncRNA encoded in the same chromosomal region as *Malat1* with a structurally similar ENE triplex. Compound **5** selectivity for *Malat1* ENE over counterpart elements of *NEAT1* and Kaposi's sarcoma herpesvirus (KSHV) polyadenylated nuclear (PAN) lncRNA[26] was further demonstrated by nuclear magnetic resonance (NMR) spectroscopy. Compounds described here demonstrate selective small-molecule targeting of an RNA triple helix and present novel classes of potential anticancer therapeutics.

RESULTS AND DISCUSSION

SMM Screening of the *Malat1* ENE Triple Helix.

Our small molecule microarray (SMM) approach (Figure 1a) has successfully identified several chemotypes targeting RNA and DNA motifs, including the HIV TAR hairpin,[27,28]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

miR21,[29] Myc, and KRAS G-quadruplexes.[30,31] Here, a minimal 5′-fluorescently labeled *Malat1* (Figure 1a) RNA construct was used to screen a ~ 26 000 compound library. To select the most promising compounds, a pipeline composed of statistical analysis, inspection of pharmacophore properties (*i.e.*, whether compound would be amenable to later medicinal chemistry efforts), selectivity, commercial availability, and cell-based evaluation of the top 28 compounds was developed (Figure 1b). For each compound, a composite Z-score (see the Methods section) was calculated based on increased fluorescence at that location on the array in the presence of *Malat1* ENE triplex (Figure S1). Compounds with a Z-score >3 were further investigated.[32] Visual inspection of array fluorescence signals and elimination of false positive signals yielded 188 molecules, representing a hit rate of 0.7%. From this list, 28 compounds (Figures S1 and S2), representing the most selective binders (defined as hits that failed to bind more than three RNAs including hairpins, riboswitches, and a triple helix encoded by KSHV PAN in our internal database of small molecule RNA binders) were selected for further investigation.

Compounds 5 and 16 and Reduction of *Malat1* Levels and Mammary Tumor Organoid Branching.

Hit compounds were next evaluated in an organoid model of mammary cancer derived from MMTV-PyMT luminal B tumors in mice.[5] The MMTV-PyMT mouse model is a well-studied mammary tumor model generated through expression of the Polyoma Virus middle T antigen oncoprotein under the direction of a mouse mammary tumor virus promoter. The model recapitulates aspects of human Luminal B breast cancer.[33] MMTV-PyMT tumors are undifferentiated, aggressive mammary carcinomas prone to metastasizing to the lungs. Primary mammary tumor organoids grown in 3D culture are a closer mimic of native epithelia than cells grown in 2D culture. This system presents an *ex vivo* model of mammary gland development and branching morphogenesis, which is driven by collective cell migration and proliferation.[34] Previous studies have shown that *Malat1* loss caused a significant reduction in organoid branching morphogenesis.[5] Therefore, we developed mammary tumor organoids from this mouse model and evaluated 27 compounds (compound **17** was cytotoxic) at a final concentration of 1 μM for their effect on lowering *Malat1* RNA levels. The most promising of these are depicted in Figure 1c, with compounds **5** and **16** (Figure 1d) inducing the largest reductions, *i.e.,* 54% and 41%, respectively. HEK293T cells treated with either compounds **5** and **16** displayed a morphology similar to that of untreated and DMSO-treated cells, indicating a lack of toxicity.[35] Freshly derived tumor organoids were next cultured on matrigel for 7 days independently with each compound. DMSO was used as a negative control, and a *Malat1* ASO was used as a positive control. The latter targets a region unrelated to the ENE triple helix and was previously demonstrated to reduce *Malat1* accumulation *in vivo* and *in vivo*.[5] At the end of the incubation period, mock- and DMSO-treated cells underwent extensive branching morphogenesis. In contrast, we observed a 38% and 27% decrease in organoid branching upon treatment with compounds **5** and **16**, respectively (Figure 2). Akin to data obtained with the *Malat1* ASO, cells treated with compound **5** remained as spherical acini (Figure 2a). Following quantification, compounds **5** and **16** were demonstrated to reduce *Malat1* levels by 54% and 41%, respectively. We also observed a 38% and 27% decrease in organoid branching upon

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

treatment with compounds **5** and **16** (Figure 2b,c), respectively. Organoid branching data for all hits is provided in Figure S3.

Compound **5** (Figure 1d) contains a 1-methyl-imadazol-2-amine scaffold and is one of five closely related compounds (compounds **6**–**9**; see Figure S2). Although compounds **6**–**9** differed only in their substitution pattern on the phenyl and benzyl side chains, they were less potent, suggesting the importance of these substitutions and demonstrating an initial structure–activity relationship (SAR) in this class. Also, while compounds **16** and **10** share a benzimidazole scaffold (Figure S2), their substitution pattern around the scaffold differed significantly, and the latter had a significantly lower effect on both *Malat1* levels and organoid branching. No evidence for RNA targeting has been reported for compounds **5** and **16**. Thus, data presented here provide the first evidence of small molecules capable of targeting the ENE triplex and *Malat1* expression levels.

Compound 5 Modulation of *Malat1* Target Genes.

We also evaluated the levels of two previously reported downstream targets of *Malat1* upon treatment with compound **5**. These included *krt16*, which encodes for keratin 16, and was reported to be down regulated upon *Malat1* loss,[5] and *csn2*, which encodes for the β-casein protein, a major milk protein, which increases upon depletion of *Malat1*.[5] In both cases, results with compound **5** were consistent with the effect observed following *Malat1* ASO treatment (Figure 3a,b), further strengthening the notion of the *Malat1* ENE triplex as the biological target.

Distinct Interaction Mechanisms for Compounds 5 and 16 with the ENE Triplex.

To affect gene regulatory responses, natural small molecule-RNA complexes, *e.g.,* those occurring in riboswitches, couple ligand binding and RNA conformational changes. Here, we utilized Förster resonance energy transfer (FRET) to examine whether the interaction between the *Malat1* ENE triplex and compounds **5** and **16** likewise induced conformational changes (Figure 4). Importantly, the FRET-based assay is orthogonal to the SMM screen and can be used to confirm direct interaction of compounds **5** and **16**. We designed a bimolecular triplex construct wherein the ENE motif contains a site-specific Cy3 donor dye, and the Cy5 acceptor dye is attached to the A-rich tail (Figure 4d). This ENE-tail *MALAT1* triplex ($M1^{ET}$) reports on structural changes at the base of the triple helix and adjacent P1 helix. High FRET efficiency (E_{FRET}) indicates proximity of the two dyes, which can be interpreted as a more structured RNA, while lower energy transfer is diagnostic of structural unfolding.

Using FRET, we have demonstrated diverse *MALAT1* triplex conformations are controlled by alterations in the ionic environment.[36] Triplex formation is variable across an arrayed landscape of eight monovalent and eight Mg^{2+} concentrations. In the absence of small molecules, FRET efficiency (E_{FRET}) increases monotonically with increasing $MgCl_2$ but exhibits a chevron-like response to increasing total monovalent salt (Figure 4a). A globally folded RNA ($E_{FRET} \approx 0.8$) is highly populated in 1 mM $MgCl_2$ and 2 mM total monovalent salt. In contrast, a dynamic conformation with disrupted triplex interactions dominates in 0.1 mM $MgCl_2$ and 52.6 mM total monovalent salt ($E_{FRET} \approx 0.55$). Under these conditions, triplex dynamics lead to degradation of the RNA *in vitro* by RNase R.[36] Intermediate E_{FRET}

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

values under all other ionic conditions indicate changes in relative populations between these largely folded and disrupted triplex states.

Because distinct structural ensembles may interact differently with a small molecule, which may aid in elucidating the binding mechanism for compounds **5** and **16**, we evaluated the E_{FRET} response across the same array of ionic conditions. The distribution of triplex conformations is significantly altered by compounds **5** and **16** (Figures 4b,c). Addition of 10 μM compound **5** leads to increased E_{FRET} across the entire landscape of ionic conditions (Figure 4b), with the largest changes observed in monovalent concentrations exceeding 25 mM (Figure 4e). In contrast, compound **16** induces an overall opposite change in E_{FRET}. Decreased E_{FRET} in 50 μM compound **16** indicates compound-induced triplex disruption (Figure 4f). Again, in monovalent concentrations above 25 mM, larger E_{FRET} changes are conferred by compound **16** binding. However, the overall change in E_{FRET} amplitude is small (Figure 4f), indicating that compound **16** may bind to a region of the RNA not immediately affecting the fluorophores. Specific monovalent cations are not expected to mediate binding because neither compound contains negatively charged functional groups requiring charge neutralization. At physiological pH, a positively charged primary amine on compound **16** is expected to interact favorably with the polyanionic triplex. Thus, in the absence of direct cation-mediated small molecule binding, preferential binding in the presence of high monovalent cation concentrations (>25 mM) is likely due to favorable changes in the local triplex structural dynamics and folding such that compound binding is increased.[36]

Qualitatively similar compound-dependent E_{FRET} changes were observed for a second bimolecular *MALAT1* triplex FRET construct, wherein a FRET-reporting pair is located at the apex of the triplex near the 3′ end and the P2 helix ($M1^{AB}$, Figure S5D). The consistent compound-induced structural changes observed in both FRET constructs confirm that compounds **5** and **16** directly bind the ENE triplex with distinct mechanisms of interaction.

To further examine the interaction mechanisms of compounds **5** and **16**, we monitored changes they induced in triplex thermal stability ($T_m^{3°}$) across the same ionic landscape (Figure S6A−C,F–H) using differential scanning FRET (DS-FRET), a high-throughput thermal melting assay.[37] Increases in $M1^{ET}$ and $M1^{AB}$ thermal stability in the presence of 10 μM compound **5** occur primarily in high monovalent salt concentrations (Figure S6D,I), consistent with the changes observed in the E_{FRET} landscapes (Figures 4e and S5E). Interestingly, compound **5** decreases triplex stability in solutions containing <25 mM monovalent salt. Under these same conditions, an increased E_{FRET} is recorded upon compound **5** binding (Figure 4b). The opposing E_{FRET} and stability measurements suggest that higher E_{FRET}, while indicative of a globally folded RNA, does not specifically report on the uniformity of triplex structure. Accommodation of compound **5** within the triplex likely induces local structural rearrangements, to which our FRET probes are insensitive, resulting in differential triplex stabilities. Alterations to triplex structure and dynamics may contribute an important role in the mechanism of inhibition,[36] leading to reduced *Malat1* levels observed in Figure 1c. In contrast to the large stability changes induced by compound **5**, the binding of compound **16** does not produce significant changes in triplex stability for either $M1^{ET}$ or $M1^{AB}$ (Figure S6E,J).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Compounds 5 and 16 and Similar ENE Binding Affinities.

To determine binding affinities for each compound, titrations were monitored by FRET and DS-FRET. Using the $M1^{ET}$ and $M1^{AB}$ *Malat1* constructs, low-micromolar affinities were calculated from titrations under several different ionic conditions (Table S1). Compound-induced changes in E_{FRET} or $T_m^{3°}$ were fit assuming a single binding event. The K_d values for compound **5** ranged from 2.3 ± 1.7−7.7 ± 5.2 μM, depending on the ionic environment, when bound to $M1^{ET}$, and similar K_d values were determined from titrations using $M1^{AB}$ (1.3 ± 0.5−6.7 ± 0.7 μM) (Figures 5a and S7). Titrations of compound **16**, monitored by FRET, provided binding affinities of 2.5 ± 0.8−6.1 ± 1.8 μM (Figures 5b and S4). Therefore, although the two compounds induce opposing structural changes (E_{FRET}, Figure 4), both bind the ENE triplex with comparable affinity.

Binding interactions were further characterized by iso-thermal titration calorimetry (ITC). Direct observation of binding enthalpies by ITC allows the determination of thermodynamic binding properties for these RNA-small molecule interactions. Because ITC is a label-free experiment, binding between the full-length, unimolecular *MALAT1* triple helix ($M1^{TH}$), and each compound can be monitored. The direct observation of binding enthalpies by ITC allows the determination of thermodynamic binding properties for these RNA–small molecule interactions. Binding affinities in the low-micromolar range were determined by ITC, consistent with FRET and DS-FRET observations (Table S1). Importantly, compound **5** binding to tRNA was not observed and a minor ITC binding signal between compound 16 and tRNA indicates an affinity weaker than millimolar (ITC data not shown). Under near-physiologic ionic conditions, compound **5** binds to $M1^{TH}$ with K_d = 2.9 ± 1.6 μM (Figure 5c). The $M1^{TH}$-compound **16** binding interaction was approximately 2-fold weaker (K_d = 6.1 ± 2.1 μM, Figure 5d). The enthalpy of binding is more favorable for compound **16** than compound **5** (−22.5 *versus* −6.1 kJ/mol, respectively; Tables S2 and S3) and may reflect charge–charge and hydrogen bonding interactions between the RNA and functional groups on compound **16** (Figure 1d). While favorable enthalpy of binding for compound **16** implies the formation of intermolecular interactions with several nucleotides on the RNA surface, compound **16** does not produce significant changes in triplex stability for either $M1^{ET}$ or $M1^{AB}$ (Figure S6E,J). This suggests compound **16** disengages from the RNA at temperatures below the triplex melting transition, consistent with binding on the RNA surface with a fast off-rate.

The overall low enthalpy of binding for compound **5** is consistent with its lack of charge and limited hydrogen bonding potential (Figure 1d). In contrast to compound **16**, binding of compound **5** is entropically driven and may reflect release of multiple water molecules or ions upon binding in a deep pocket on the RNA.

Molecular Docking and Confirmation of Distinct Binding Modes of Compounds 5 and 16 on the ENE Triplex.

To assess molecular interactions between the *MALAT1* ENE triplex (Figure 6a, PDBID: 4PLX)[25] and compounds **5** and **16**, we performed docking studies using AutoDock 4.[38] This approach employs a flexible ligand and rigid receptor to calculate independent docking events utilizing a Lamarckian genetic algorithm (300 trials, initial population of 300, 2.7 ×

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

10^4 generations, and 10^6 energy evaluations). A total of 246 and 156 distinct docking poses were identified for compounds **5** and **16**, respectively (Figure 6b,e). To evaluate the molecular interactions more closely, docking poses were parsed into clusters based on their structural localization (Figure 6c,f and S8D). Each individual cluster comprises >5% of the total docking poses for that compound.

The distribution of docking poses for compound **5** suggests preferential binding within the deep and narrow major groove of the triplex (Figure 6b). The calculated electrostatic potential mapped to the surface of the ENE (Figure S8F) reveals a highly electronegative major groove. This high charge density rationalizes the importance of charge neutralization provided by elevated monovalent salt concentrations (>25 mM) to support triplex formation. Indeed, compound **5** binding induced larger conformational and stability changes in >25 mM monovalent salt (Figures 4e, and S6D,I). Analysis of docking poses for compound **5** yields a single cluster representing the most probable docking poses. This cluster is represented by the lowest energy docking pose deep within the triplex major groove (Figure 6c), where the molecule makes several van der Waals interactions with the phosphate and nucleobase groups of the Hoogsteen strand (U10, U11, and C12) (numbering from PDBID: 4PLX[25]) and the Watson–Crick strand (U37, U38, and U39) (Figure 6d). These binding interactions provide insight into an explanation for the low enthalpy of compound **5** binding determined by ITC (Figure 5c). Overall, nonspecific interactions between compound **5** and the ENE triplex rationalize the micromolar (not tighter) binding affinities. To assess the variability of potential interactions within major groove, the individual docking poses within the compound **5** cluster were analyzed. Compound **5** is capable of making nonspecific interactions with other nucleotides within the major groove but only within a very narrow region (Figure S8A−C). It is clear that not only does compound **5** preferentially bind within the major groove of the triplex, but also, there is a strong preference for van der Waals interactions with the two U-rich strands over the 3′ A-rich tail.

Molecular docking events for compound **16** indicate a significantly different binding mode, *i.e.,* docking positions are distributed along the narrow minor groove surface of the ENE triplex (Figure 6e). The most probable docking cluster for compound *16* is located on the RNA surface (Figure 6f). Additional clusters for compound *16* were also identified, although they were part of many fewer docking events (Figure S8D). Importantly, two of these clusters also localize to the triple helix surface, and a third cluster found in the major groove is only minimally populated. Superficial binding is consistent with the inferred fast off-rate and the lack of additional thermal stability conferred by compound **16** binding (Figure S6E,J). Inspection of the most probable cluster near the base of the triplex shows that G48-G49 (PDBID: 4PLX numbering)[25] create a “bulge cleft” (Figure S8D,E). The lowest energy pose within this cluster makes favorable hydrogen bonds with backbone atoms of nucleotides A67 and, potentially, the base of U47 (Figure 6g, purple arrows). While more hydrogen bonds can be identified for compound **16**, these are formed primarily with backbone sugar atoms and do not form many specific interactions with nucleobases of the triplex. Therefore, favorable enthalpy of binding is achieved for compound **16** (Figure 5d and Table S2), but a specific base readout of the overall ENE triplex is not improved relative to compound **5**.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To further support the docking studies and eliminate the possibility of **5** and **16** simply intercalating within the triple helix, we performed a dye-displacement assay[39] using SYBR Green II, a known intercalating dye. Incubation of $M1^{TH}$ with SYBR Green II results in high dye fluorescence. Supplementing with high concentrations of ellipticine, which contains an acridine-like moiety supporting intercalation, causes a significant reduction of fluorescence upon displacement of SYBR Green II with a IC_{50} of 35 μM. In contrast, neither compound **5** nor compound **16** induced loss of SYBR Green II fluorescence, strongly suggesting that they do not interclate (Figure S9).

Compounds 5 and 16 and Lack of Effect on *Neat1* lncRNA Levels.

The apparent lack of specific nucleobase read-out by compounds **5** or **16** raises an important question regarding their specificity for the *MALAT1* ENE triplex. Within mammalian genomes, only one other homologous RNA triple helix has been documented. This is located at the 3′ end of the *NEAT1* lncRNA (also known as *MENβ*) and serves a similar polyA-independent protective role.[12,40] Lack of a *NEAT1* triple helix crystal structure precluded comparative docking analysis.

Nonetheless, sequence and secondary structure comparison reveal that the *NEAT1* triplex has an altered “bulge cleft” where compound **16** binds and there are fewer base-triples surrounding the predicted compound **5** binding site (Figure 7a,b). It is not immediately known whether these differences would significantly alter compound binding. We therefore examined *Neat1* RNA levels by RT-qPCR analysis upon treatment with compounds **5** and **16**. Despite high structural homology, we failed to observe any significant change in *Neat1* levels upon treatment with either compound, demonstrating their specificity for *Malat1* lncRNA (Figure 7c).

Saturation Transfer Difference NMR and Confirmation of Selective Recognition of the *MALAT1* ENE Triplex.

We decided to use another biophysical method of saturation transfer difference NMR (STD NMR) to probe the interactions between **5** and *MALAT1* ENE triplex. STD NMR is a ligand-based screening technique that builds on the nuclear Overhauser effect[41] and works for ligands in fast exchange, with K_d values in a range of 10^{-8}–10^{-3} mol L^{-1}. Experimentally, a pair of 1D ^{1}H NMR spectra are recorded: an on-resonance experiment in which the RNA (receptor) proton magnetization is selectively saturated with radio frequency pulses (at 5.9 or 6.5 ppm) and a second experiment in which the sample is selectively saturated off-resonance (reference) of both the receptor and the ligand (50 ppm). In the on-resonance experiment, saturation is transferred throughout the RNA *via* spin-diffusion and to any bound ligand *via* intermolecular ^{1}H–^{1}H cross-relaxation mechanisms. If the ligand and receptor interact within the STD detectable binding affinity range (typically high-nanomolar to low millimolar K_d), subtraction of the on-resonance from the reference off-resonance spectrum results in positive peaks for protons that were saturated (receptor and bound ligand). STD data shown in Figure 8b reveal several well-defined signals from compound **5** when incubated with the *MALAT1* ENE (same construct as Figure 6a). In contrast, equivalent compound **5** signals are absent when it is incubated with the ENE triplexes of *NEAT1* (Figure 8c) and KSHV PAN (Figure 8d), further highlighting compound **5**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

selectivity. Unfortunately, solubility issues precluded performing the equivalent analysis with compound **16**. STD signals from protons on para-methoxy phenyl ring and the methyl-substituted imidazole ring are not only consistent with docking calculations, assisting in orienting compound **5** within the *MALAT1* triple helix (Figure 6d), but also suggest these sites as components of the pharmacophore, which will aid future inhibitor design.

The ENE triple helix located at the 3′ end of the ~8 kb lncRNA *MALAT1* is responsible for evading the cellular degradation machinery and promotes persistent activity of this potential oncogenic lncRNA within many cancer types. We report the use of an unbiased SMM to rapidly screen ~26 000 compounds for binding to a fluorescently labeled *Malat1* triple helix, revealing 188 hit compounds, which were narrowed to 28 hits based on selectivity and commercial availability. We monitored *in cellulo* RNA levels using RT-qPCR after incubation with 27 of these, selecting compounds **5** and **16** as the most promising leads. Using *in vitro* assays orthogonal to the primary screen, including FRET, DS-FRET, and ITC, we confirmed that these small molecules directly interact with the triplex with low-micromolar affinities. Interestingly, compounds **5** and **16** exert control over the triplex structure and function through distinct mechanisms. Because the biological mechanism of triplex-mediated protection remains an open question, we cannot directly correlate either binding mechanism with a biological mode of inhibition.

We uncovered distinct mechanisms of binding for compounds **5** and **16**. Biophysical analyses reveal that compound **5** binds deeply within the triplex, and the binding is entropically driven, while compound **16** binds preferentially to the RNA surface with a large enthalpic contribution stemming from intermolecular interactions with the RNA. Although a detailed description of binding mechanisms cannot *a priori* predict biological efficacy and 3D tumor growth experiments indicate that compound **5** is more effective than compound **16** in reducing *Malat1* abundance and organoid branching. Therefore, deep binding within the triplex, as with compound **5**, is more favorable than surface interactions observed with compound **16**. Mechanistically, it is intuitive that a globally disrupted triplex, induced by compound **16**, would result in the degradation of *MALAT1 in vivo*. It is less straightforward to interpret the FRET responses to compound **5** because a higher E_{FRET} is often interpreted as a more folded RNA. Nonetheless, the biological efficacy of compound **5** indicates that the triplex-mediated protective function is disrupted. We suggest that local structural rearrangements in the ENE triplex accompany compound **5** binding. Such rearrangements may affect subsequent interactions between the triplex and proteins involved in the cellular degradation machinery. In this way, the use of compounds **5** and **16** as chemical probes may help address gaps surrounding the mechanism of *MALAT1* degradation.

Our results demonstrate that multiple small-molecule binding mechanisms can lead to a desired biological inhibition. This expands the possibilities for targeting *MALAT1* with small molecules, either alone or in combination with antisense oligonucleotides. Furthermore, it highlights two important aspects for consideration regarding therapeutic development targeting this ENE triplex. First, the RNA structure itself can be disrupted (*e.g.,* compound **16**) to reduce *MALAT1* levels. Second, small-molecule binding may serve to inhibit other biological cofactors from binding to the triplex, which may explain the mode of action for compound **5**. For example, binding within the major groove may directly inhibit a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

protein-binding cofactor, such as the methyltransferase like protein 16 (METTL16), which was recently identified to interact with the *MALAT1* ENE.[42] However, understanding the inhibitory mechanism of compound **5** is limited by the lack of knowledge surrounding the actual mechanism of triplex-mediated protection. In this way, our computational docking results for compounds **5** and **16** provide an avenue to begin interrogating the mechanism of triplex protection (for example, by altering METTL16 binding). Serving as chemical biology probes, these small molecules may serve to elucidate both the mechanism of triplex-mediated protection of *MALAT1* and potentially effective mechanisms to inhibit this important therapeutic target.

Collectively, our findings provide the first example of small molecules that specifically recognize the *MALAT1* ENE triplex, providing useful chemical tools to enhance understanding of *MALAT1* function in breast cancer progression, and potentially a novel class of anticancer therapeutics. Understanding the sequence and structural features controlling ENE triplex binding specificity (*i.e., MALAT1 versus NEAT1 versus PAN ENE*) and cellular factors whose affinity for the triple helix might be affected by compound binding is the focus of future studies.

METHODS

SMM Preparation and Screen.

SMM screening was carried out as described.[32] Briefly, γ-aminopropyl silane (GAPS) microscope slides (Corning) were functionalized with a short Fmoc-protected amino polyethylene glycol spacer. After deprotection with piperidine, 1,6-diisocyanatohexane was coupled to the surface by urea bond formation to provide functionalized isocyanate-coated microarray slides that react with primary and secondary amines and primary alcohols to form immobilized small molecule libraries. A total of 26 229 small-molecule stock solutions (10 mM in DMSO) from (i) commercial vendors including ChemBridge and ChemDive, (ii) MIPE,[43] and (iii) NCI diversity set V[44] screening collections, as well as dyes and controls, were printed on 7 slides each containing ~3745 distinct molecules printed in duplicate. Slides were exposed to pyridine vapor to facilitate covalent attachment and then incubated with a 1:20 polyethylene glycol/DMF (v/v) solution to quench unreacted isocyanate surface. The deprotected 5′-5Alex647N-*Malat1* minimal triple helix motif:[40]

5′-
AAAGGUUUUUCUUUUCCUGAGAAAUUUCUCAGGUUUUGCUUUUUAAAAAAAA
AGCAAAA-3′, purified by RNase-free high-performance liquid chromatography (HPLC) was purchased from IDT (Coralville, Iowa) and dissolved in RNase-free distilled water. The stock was further diluted to 500 nM with folding buffer (25 mM sodium cacodylate, 50 mM KCl, and 1 mM $MgCl_2$ at pH 6.9) and annealed by heating to 95 °C for 3 min, snap-cooling on ice for 10 min and slow equilibration to room temperature for 1 h. Following incubation, slides were gently washed twice for 2 min in folding buffer with 0.01% Tween-20 and once in folding buffer and dried by centrifugation for 2 min at 4000 rpm. Fluorescence intensity was measured (650 nm excitation and 670 nm emission) on an Innopsys Innoscan 1100 AL Microarray Scanner. The scanned image was aligned with the corresponding GenePix Array List (GAL) file to identify individual features.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SMM Statistical Analysis.

A total of 188 hits were identifed (hit rate of 0.7%), based on signal-to-noise ratio (SNR), defined as [(mean foreground – mean background)/(standard deviation of background)] and Z-score, defined as Z equal to (mean SNR_{635} compound – mean SNR_{635} library)/(standard deviation of SNR_{635} library) with the following criteria:

- SNR > 0;
- Z score > 3;
- coefficient of variance (CV) of replicate spots < 200;
- $[(Z_{RNA\ incubated} - Z_{buffer\ incubated})/Z_{buffer\ incubated}] > 3$; and
- visual inspection and removal of false positive (*e.g.*, dust particulates).

The final hit list was compared to an internal hit list with other labeled nucleic acids including RNA hairpins, riboswitches, and an RNA triple helix from Kaposi sarcoma herpes virus polyadenylated nuclear (PAN) RNA. Any compounds annotated to bind more than three other RNA motifs were removed to providing a final hit list of 28 compounds.

Dye Displacement Assay.

The 94nt *MALAT1* RNA (also termed $M1^{TH}$) was transcribed by T7 RNA polymerase using a *MALAT1* HDV ribozyme plasmid (a generous gift from Prof. Jessica Brown, University of Notre Dame). For this assay, a fixed concentration of 94nt *MALAT1* (500 nM) and SYBRG II (In*vitro*gen) (4×) was used. A total of 5 μL of compound (in DMSO) and 95 μL of RNA/dye complex in assay buffer (5 mM sodium cacodylate at pH 6.5, 50 mM KCl, 1 mM $MgCl_2$, 0.1 mM EDTA, 0.01% Triton-x100) were added to assay plate (black nunc 96-well plate, Fisher Scientific) and incubated at room temperature for 30 min, and fluorescence intensity values were measured (485 ± 5 nm excitation, 525 ± 5 nm emission) using a Tecan plate reader. IC_{50} values were determined by normalizing fluorescence intensity of each well to an average value for the fluorescence intensity of RNA/dye complex and using a nonlinear regression fitting of the competition curves (GraphPad Prism 7.0 software).

RT-qPCR.

cDNA was prepared from 1 μg RNA using random hexamer and 1/50 volume of the cDNA reaction mixture was used in a RT-qPCR reaction with 1× SYBR green mixture (Applied Biosystems) and specific forward and reverse primers. ΔΔCt method was used to calculate RNA fold change,[45] and mouse *Gapdh, Actb* (β-actin), and 18s rRNA were used as reference genes. Primers were: *MALAT1* F(GTTACCAGCCCAAACCTCAA), *MALAT1* R(CTACATTCCCACCCAGCACT), mouse *Gapdh* F(AAGGTCATCCCAGAGCTGAA), mouse *Gapdh* R(CTGCTTCACCTTCTTGA), mouse *Actb* F (AGGGAAATCGTGCGTGACAT), mouse *Actb* R(GAACCGCTCGTTGCCAATAG), 18s rRNA F(GTAACCCGTTGAACCCCATT), and 18s rRNA R(CCATCCAATCGGTAGTAGCG).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Organoid Culture.

Tumors were removed from MMTV-PyMT mice and digested into small clumps of cells by collagenase, which are referred to as organoids. The organoids were cultured in Matrigel as described[5,46] for 7 days with 1 μM compound treatment and 0.1% DMSO as control. Images were obtained using a Zeiss Axio-Observer light microscope. More than 80 organoids from each individual treatment were counted and evaluated for branching morphogenesis.

Preparation of Fluorescently-Labeled RNAs.

Fluorescently labeled synthetic RNAs were purchased from the Keck Foundation Biotechnology Resource Facility (Yale University). Experiments using the $M1^{ET}$ construct contained fluorescently labeled MALAT1-ENE and MALAT-Tail as follows. For MALAT1-ENE, (GAAGGUUUUUCUUUUCCUGAGAAAACAACACGUAUUGUUUUCUCAGGUUUUGCUUUUUGGCCUUU) was synthesized containing a 3′ Cy3 fluorophore. For MALAT-Tail, (AAAAAAAGCAAAA) was synthesized containing a 5′ Cy5 dye. Experiments using the $M1^{AB}$ construct contained fluorescently labeled MALAT1-A and MALAT1-B RNAs as follows. For MALAT1-A, (GGAAGGUUUUUCUUUUCCUGAG) was synthesized containing a 3′ Cy3 fluorphore. For MALAT1-B, (CUCAGGUUUUGCUUUUUGGCCUUUUUCUAGCUUAAAAAAAAAAAAAAGCAAAA) was synthesized containing a 3′ C7 amino linker and subsequently labeled with Cy5 (see below). Synthetic RNAs were purified by denaturing polyacrylamide gel electrophoresis and electroelution (Elutrap, GE). Eluates were concentrated using ethanol precipitation, resuspended in water, and stored at −20 °C until further use.

The 3′ C7 amino linker of MALAT1-B was labeled using Cy5 NHS ester monoreactive dye. One tube of dye (PA25001, GE LifeSciences) was dissolved in 14 μL 100% DMSO and added to 100–500 μg of RNA in 0.1 M sodium carbonate (pH 8.5) to a final volume of 100 μL. The reaction was incubated overnight at RT with shaking. Following ethanol precipitation, a second labeling reaction was performed to achieve higher labeling efficiency. The second labeling reaction was resuspended in 50 mM triethylammonium acetate (TEAA), pH 7, and loaded onto a C-8a reverse-phase HPLC column (XDB-C8, Agilent) equilibrated in TEAA with 5% acetonitrile. RNA was eluted using a 5%–40% acetonitrile gradient in 40 minutes and a flow rate of 4 mL min^{-1}. The labeled RNA fraction was frozen at −80 °C prior to lyophilization. Purified labeled RNA was resuspended in water and stored at −20 °C until further use. Prior to experimentation, bimolecular constructs were annealed in 1:1.5 molar ratio (the $M1^{ET}$ ratio is MALAT1-ENE-Cy3/MALAT1-Tail-Cy5; the $M1^{AB}$ ratio is MALAT1-B-Cy5/MALAT1-A-Cy3) and refolded in a buffer of 20 mM HEPES-KOH, pH 6.9, and 1 mM $MgCl_2$ for ~1 h before loading onto an ENrich SEC70 column (Bio-Rad) pre-equilibrated in the same refolding buffer. Triplex complexes were selected based on overlapping absorbance peaks monitored at 260 nm (RNA), 550 nm (Cy3), and 650 nm (Cy5).

Preparation and Measurement of RT FRET Experiments.

Size exclusion chromatography-purified, fluorescently labeled triplex constructs were diluted to a final concentration of 166 nM and compounds **5** or **16** added to a final

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

concentration of 16 μM. A total of 24 μL of this mixture was dispensed to each well in a grid of eight rows and eight columns on a 384-well microplate. To each column was added 8 μL of a different 5× $MgCl_2$ concentrated stock; each $MgCl_2$ stock solution contained 20 mM HEPES-KOH, pH 6.9, and an individual $MgCl_2$ concentration between 0–5 mM. A total of 8 μL of a 5× monovalent salt concentrated stock was added to each row. Monovalent stocks were prepared similar to $MgCl_2$ stocks but contained equimolar NaCl and KCl at concentrations between 0–1000 mM. The final reactions contained 100 nM RNA, 10 μM or 50 μM compound, 20 mM HEPES-KOH, pH 6.9, 0.1–1 mM $MgCl_2$, and 2.6–202.6 mM monovalent salt including 2.6 mM K^+ from the HEPES buffer. The microplate was covered with tin foil prior to centrifugation at 1000*g* for 2 min at room temperature. The plate was incubated at room temperature for ~1 h before recording fluorescence measurements.

After incubation, fluorescence of the dually labeled RNA was recorded using a SYNERGY H1MF microplate reader (BioTek) with single excitation filter for Cy3 (540 ± 25 nm) and 2 emission filters (590 ± 20 nm) and (680 ± 30 nm) for Cy3 and Cy5, respectively. An apparent FRET efficiency was calculated using $E_{FRET} = I_A/(I_A + I_D)$, where I_A represents fluorescence intensity of the Cy5 acceptor dye and I_D represents the intensity of the Cy3 donor dye. 3D landscape plots were generated using Origin2015 software (Originlab). All experiments using $M1^{ET}$ with compounds **5** or **16** were performed in triplicate (3 independent experiments). Experiments with $M1^{AB}$ and compound **5** were performed in duplicate (2 independent experiments). Only 1 experiment was performed with $M1^{AB}$ and compound **16**.

Thermal FRET Melting Experiments.

Thermal FRET melting experiments were performed on a QuantStudio Flex 7 (Applied Biosystems). A total of 25 μL from the samples used in the RT FRET experiment were transferred to 384-well qPCR plate, which was sealed with an optical qPCR seal prior to centrifugation at 1000*g* for 2 min. Cy3 and Cy5 fluorescence were recorded in a qPCR instrument over the temperature range from 20–85 °C using a ramp rate of 0.015 °C s^{-1}. A single excitation filter was used to excite Cy3 (520 ± 10 nm), and 2 emission filters were used to record Cy3 (586 ± 10 nm) and Cy5 fluorescence (682 ± 14 nm). Raw Cy3 and Cy5 fluorescence intensities were converted to FRET melting curves [$E_{FRET} = I_A/(I_A + I_D)$] using Origin2015 software (OriginLab). Triplex melting temperatures ($T_m{}^{3°}$) were identified from the first derivative of the E_{FRET} curves using a peaks finding algorithm in Origin2015. Subsequently, $T_m{}^{3°}$ landscape plots across the range of $MgCl_2$ and monovalent salts were generated using Origin2015. All experiments using $M1^{ET}$ with compounds **5** or **16** were performed in triplicate (3 independent experiments). Experiments with $M1^{AB}$ and compound 5 were performed in duplicate (2 independent experiments). Only 1 experiment was performed with $M1^{AB}$ and compound **16**.

Compound Titrations and K_d Determination.

A total of 8 different stock concentrations were prepared for each compound, ranging from 0–200 μM. Compound stocks were prepared in identical buffer to the RNA (see below) with a final DMSO concentration ranging from 1–5%. No differences in fluorescence signal were observed within this range of DMSO concentration. SEC-purified RNA was diluted to 200

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

nM in 4 different salt conditions: 1 mM $MgCl_2$ and 2.6 mM monovalent (Condition I), 0.1 mM $MgCl_2$ and 2.6 mM monovalent (Condition II), 0.1 mM $MgCl_2$ and 52.6 mM monovalent salt (Condition III), and 1 mM $MgCl_2$ and 102.6 mM monovalent (Condition IV). A total of 20 μL of RNA was dispensed into in 8 wells in a 384-well microplate, followed by 20 μL from the compound titration stock solutions, yielding final concentrations of 100 nM RNA and 0–100 μM compound. The microplate was sealed and covered with tin foil prior to centrifugation for 2 min at 1000*g* at RT. The sealed plate was incubated ~1 h prior to recording room-temperature FRET as described above. For thermal FRET melting experiments, 25 μL of the titration samples was transferred into a qPCR plate, which was sealed and centrifuged at 1000*g* for 2 min at room temperature. Melting profiles were recorded in the QuantStudio Flex 7 (Applied Biosystems) as described above. All titration experiments were performed in triplicate (three independent experiments), and standard deviation was calculated. Each titration was fit to a single binding event in Origin2015, and the K_d was determined using:

$$\left(S_m + S_x \times \frac{[cmpd]}{K_d}\right) / \left(1 + \frac{[cmpd]}{K_d}\right) \quad (1)$$

where S_m is the signal from the free RNA, S_x is the signal from the RNA-compound complex, [cmpd] is the concentration of the compound added, and K_d is the binding affinity.

ITC Analysis.

Full-length 94 nt *MALAT1* triple helix was transcribed from PCR products. The template strand contained 2′-O-methylated nucleotides at the final 2 positions to minimize nontemplated nucleotide addition at the 3′ end of the transcript.[47] RNA was purified by denaturing gel electrophoresis and electrolution. The eluate was concentrated using ethanol precipitation. RNA was resuspended in buffer containing 20 mM HEPES-KOH, pH 6.9, 1 mM $MgCl_2$, 25 mM NaCl and 25 mM KCl, heated at 95 °C for 2 min, and then snap-cooled on ice for 5 min. The RNA was incubated at room temperature for ~1 h prior to SEC purification. The purified fraction was concentrated by centrifugal ultrafiltration (Millipore). For 3 independent experiments with compound **5**, a sample containing 75 μM RNA in 1 mM $MgCl_2$, 100 mM NaCl, 100 mM KCl, and 1% DMSO in 20 mM HEPES-KOH, pH 6.9 was placed in the ITC cell. The syringe contained 500 μM compound **5** in an identical buffer. For 2 independent experiments with compound **16**, a sample containing 75 μM RNA in 1 mM $MgCl_2$, 25 mM NaCl, 25 mM KCl, and 1% DMSO in 20 mM HEPES-KOH, pH 6.9 was placed in the ITC cell. The syringe contained 500 μM compound **16** in an identical buffer. All experiments were performed using an Affinity ITC (TA Instruments). Data were analyzed using NanoAnalyze (TA Instruments) and exported to Origin2015 (OriginLab) for generating final plots. ITC data represents raw trances of compound titrations into RNA in which their titration into buffer can be approximated using the last ~7 data points after reaching saturation at high concentrations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Flexible Ligand Molecular Docking.

The structure of the *MALAT1* triple helix core solved using X-ray crystallography was obtained from the Protein Data Bank (PDBID: 4PLX)[25] and edited to contain only a single macromolecular chain (chain A). The AMBER 16 molecular dynamics package was used to add all missing hydrogen atoms, adjust bond lengths, angles, and dihedral angles based on the OL3[48] set of RNA force-field parameters. The LigPrep program[49] was used to generate the three-dimensional structure of compounds **5** and **16**, which were subsequently energy-minimized using the AMBER ff03 force-field.[48]

The AutoDock 4.0 suite of programs[38] was used to prepare and generate all docking poses using the *MALAT1* triple helix structure (receptor) and the structures of compounds **5** and **16** (ligands). The receptor was designated as rigid and all ligands were designated as fully flexible around all rotatable bonds. Atomic charges were assigned to all atoms using the Gastiger charge formalism.[50] The docking search grid was chosen to be the maximum size of the receptor molecule (126 × 62 × 66 *xyz* grid points) with a grid spacing of 0.5 Å. A Lamarckian genetic algorithm using an initial population of 300 conformers, 27 000 generations, 10^6 energy evaluations, and a mutation rate of 0.2 was used to conduct 300 independent dockings for each ligand onto an identical receptor grid. The 247 and 157 unique final docking poses for compounds **5** and **16**, respectively, were histogrammed by energy and further parsed into unique structural clusters representing >5% of total docking poses. Clustering and visualization were done with in-house scripts and the VMD molecular graphics program.[51]

Saturation Transfer Difference NMR.

NMR data were acquired at 283 K on a 600 MHz Bruker spectrometer equipped with a cryoprobe. NMR samples contained 5 μM RNA sample in 10 mM potassium phosphate pH 6.5, 50 mM KCl, 1 mM $MgCl_2$, 20 μM DSS, and 0.02% NaN_3 with 10% D_2O. Compound 5 was added as stock solution in DMSO-d_6 to a final concentration of 150 μM and 4% DMSO-d_6. Final volume of each NMR sample was 300 μL. DSS was used as an internal standard. STD experiments were performed as described previously[52] using a standard Bruker pulse (stddiffesgp.3) with a 2 s saturation time and excitation sculpting for water suppression with 2 ms sinc pulses.[53] The RNA ^1H were saturated with a train of 50 ms Gaussian-shaped pulses. For selective saturation of the RNA, the on-resonance irradiation frequency at 6.46 ppm was used for the *MALAT1* (for sequence see Figure 6d) and *NEAT1* (sequence below) ENEs, and 5.90 ppm was used for the PAN (for sequence see Figure 1D in Mitton-Fry et al.[54]). A total of 128 scans with 16 dummy scans were collected for each. A second control experiment was performed in the absence of RNA to ensure that **5** resonances were not being excited. Spectra were processed in Topspin (Bruker). *NEAT1* sequence below with modification from Brown et al.[12] was used for STD NMR studies. 5′-GGAGGUGUUUCUUUUACUGAGUGCGAAAGCACUCAGGUUUUGCUUUUCACCUUCCCAUCUGUGAAAAAAAAGCAAAA-3′.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ACKNOWLEDGMENTS

F.A.A., S.F.J.L. and J.S.S., Jr. are supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health. D.S. is funded by NCI grant no. 5P01CA013106. G.A. is supported by a Susan G. Komen postdoctoral fellowship. N.B. is supported by a Career Transition Award from the National Heart, Lung, and Blood Institute (grant no. K22-HL121113A). A.A.A. is supported by a graduate fellowship from the Saudi Arabian Cultural Mission and Jazan University. T.K.D. is supported by NSF for NMR instrumentation (grant no. DBI1040158) and the National Institutes of Health (grant no. U54GM103297). We thank S. Yilmaz for assistance with the ITC experiments and Prof. J. Brown (University of Notre Dame) for the generous gift of *MALAT1*-HDV plasmid.

REFERENCES

(1). Arun G, Diermeier SD, and Spector DL (2018) Therapeutic Targeting of Long Non-Coding RNAs in Cancer. Trends Mol. Med 24, 257–277. [PubMed: 29449148]

(2). Tripathi V, Shen Z, Chakraborty A, Giri S, Freier SM, Wu X, Zhang Y, Gorospe M, Prasanth SG, Lal A, and Prasanth KV (2013) Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. PLoS Genet 9, No. e1003368. [PubMed: 23555285]

(3). Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, and Muller-Tidow C (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22, 8031–8041. [PubMed: 12970751]

(4). Gutschner T, Hammerle M, Eissmann M, Hsu J, Kim Y, Hung G, Revenko A, Arun G, Stentrup M, Gross M, Zornig M, MacLeod AR, Spector DL, and Diederichs S (2013) The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res 73, 1180–1189. [PubMed: 23243023]

(5). Arun G, Diermeier S, Akerman M, Chang KC, Wilkinson JE, Hearn S, Kim Y, MacLeod AR, Krainer AR, Norton L, Brogi E, Egeblad M, and Spector DL (2016) Differentiation of mammary tumors and reduction in metastasis upon MALAT1 lncRNA loss. Genes Dev 30, 34–51. [PubMed: 26701265]

(6). Guo F, Li Y, Liu Y, Wang J, Li Y, and Li G (2010) Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. Acta Biochim. Biophys. Sin 42, 224–229. [PubMed: 20213048]

(7). Tano K, Mizuno R, Okada T, Rakwal R, Shibato J, Masuo Y, Ijiri K, and Akimitsu N (2010) MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. FEBS Lett 584, 4575–4580. [PubMed: 20937273]

(8). Nakagawa S, Ip JY, Shioi G, Tripathi V, Zong X, Hirose T, and Prasanth KV (2012) MALAT1 is not an essential component of nuclear speckles in mice. RNA 18, 1487–1499. [PubMed: 22718948]

(9). Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, Xiao X, Booth CJ, Wu J, Zhang C, and Spector DL (2012) The lncRNA MALAT1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell Rep 2, 111–123. [PubMed: 22840402]

(10). Wilusz JE, Freier SM, and Spector DL (2008) 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell 135, 919–932. [PubMed: 19041754]

(11). Tycowski KT, Shu MD, and Steitz JA (2016) Myriad Triple-Helix-Forming Structures in the Transposable Element RNAs of Plants and Fungi. Cell Rep 15, 1266–1276. [PubMed: 27134163]

(12). Brown JA, Valenstein ML, Yario TA, Tycowski KT, and Steitz JA (2012) Formation of triple-helical structures by the 3 ‘-end sequences of MALAT1 and MEN beta noncoding RNAs. Proc. Natl. Acad. Sci. U. S. A 109, 19202–19207. [PubMed: 23129630]

(13). Wilusz JE, JnBaptiste CK, Lu LY, Kuhn CD, Joshua-Tor L, and Sharp PA (2012) A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. Genes Dev 26, 2392–2407. [PubMed: 23073843]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(14). Brown JA, Bulkley D, Wang J, Valenstein ML, Yario TA, Steitz TA, and Steitz JA (2014) Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. Nat. Struct. Mol. Biol 21, 633–640. [PubMed: 24952594]

(15). Brown JA, Kinzig CG, DeGregorio SJ, and Steitz JA (2016) Hoogsteen-position pyrimidines promote the stability and function of the MALAT1 RNA triple helix. RNA 22, 743–749. [PubMed: 26952103]

(16). Thomas JR, and Hergenrother PJ (2008) Targeting RNA with small molecules. Chem. Rev 108, 1171–1224. [PubMed: 18361529]

(17). Disney MD, Yildirim I, and Childs-Disney JL (2014) Methods to enable the design of bioactive small molecules targeting RNA. Org. Biomol. Chem 12, 1029–1039. [PubMed: 24357181]

(18). Shortridge MD, and Varani G (2015) Structure based approaches for targeting non-coding RNAs with small molecules. Curr. Opin. Struct. Biol 30, 79–88. [PubMed: 25687935]

(19). Hermann T (2016) Small molecules targeting viral RNA. Wiley Interdiscip Rev. RNA 7, 726–743. [PubMed: 27307213]

(20). Connelly CM, Moon MH, and Schneekloth JS Jr. (2016) The Emerging Role of RNA as a Therapeutic Target for Small Molecules. Cell Chem. Biol 23, 1077–1090. [PubMed: 27593111]

(21). Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, Andricioaei I, Markovitz DM, and Al-Hashimi HM (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nat. Chem. Biol 7, 553–559. [PubMed: 21706033]

(22). Naryshkin NA, Weetall M, Dakka A, Narasimhan J, Zhao X, Feng Z, Ling KK, Karp GM, Qi H, Woll MG, Chen G, Zhang N, Gabbeta V, Vazirani P, Bhattacharyya A, Furia B, Risher N, Sheedy J, Kong R, Ma J, Turpoff A, Lee CS, Zhang X, Moon YC, Trifillis P, Welch EM, Colacino JM, Babiak J, Almstead NG, Peltz SW, Eng LA, Chen KS, Mull JL, Lynes MS, Rubin LL, Fontoura P, Santarelli L, Haehnke D, McCarthy KD, Schmucki R, Ebeling M, Sivaramakrishnan M, Ko CP, Paushkin SV, Ratni H, Gerlach I, Ghosh A, and Metzger F (2014) Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. Science 345, 688–693. [PubMed: 25104390]

(23). Palacino J, Swalley SE, Song C, Cheung AK, Shu L, Zhang X, Van Hoosear M, Shin Y, Chin DN, Keller CG, Beibel M, Renaud NA, Smith TM, Salcius M, Shi X, Hild M, Servais R, Jain M, Deng L, Bullock C, McLellan M, Schuierer S, Murphy L, Blommers MJ, Blaustein C, Berenshteyn F, Lacoste A, Thomas JR, Roma G, Michaud GA, Tseng BS, Porter JA, Myer VE, Tallarico JA, Hamann LG, Curtis D, Fishman MC, Dietrich WF, Dales NA, and Sivasankaran R (2015) SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. Nat. Chem. Biol 11, 511–517. [PubMed: 26030728]

(24). Donlic A, Morgan BS, Xu JL, Liu A, Roble C Jr., and Hargrove AE (2018) Discovery of Small Molecule Ligands for MALAT1 by Tuning an RNA-Binding Scaffold. Angew. Chem., Int. Ed 57, 13242–13247.

(25). Brown JA, Bulkley D, Wang J, Valenstein ML, Yario TA, Steitz TA, and Steitz JA (2014) Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. Nat. Struct. Mol. Biol 21, 633–640. [PubMed: 24952594]

(26). Conrad NK (2016) New insights into the expression and functions of the Kaposi's sarcoma-associated herpesvirus long noncoding PAN RNA. Virus Res 212, 53–63. [PubMed: 26103097]

(27). Sztuba-Solinska J, Shenoy SR, Gareiss P, Krumpe LR, Le Grice SF, O'Keefe BR, and Schneekloth JS Jr. (2014) Identification of biologically active, HIV TAR RNA-binding small molecules using small molecule microarrays. J. Am. Chem. Soc 136, 8402–8410. [PubMed: 24820959]

(28). Abulwerdi FA, Shortridge MD, Sztuba-Solinska J, Wilson R, Le Grice SF, Varani G, and Schneekloth JS Jr. (2016) Development of Small Molecules with a Noncanonical Binding Mode to HIV-1 Trans Activation Response (TAR) RNA. J. Med. Chem 59, 11148–11160. [PubMed: 28002966]

(29). Connelly CM, Boer RE, Moon MH, Gareiss P, and Schneekloth JS Jr. (2017) Discovery of Inhibitors of MicroRNA-21 Processing Using Small Molecule Microarrays. ACS Chem. Biol 12, 435–443. [PubMed: 27959491]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(30). Felsenstein KM, Saunders LB, Simmons JK, Leon E, Calabrese DR, Zhang S, Michalowski A, Gareiss P, Mock BA, and Schneekloth JS Jr. (2016) Small Molecule Microarrays Enable the Identification of a Selective, Quadruplex-Binding Inhibitor of MYC Expression. ACS Chem. Biol 11, 139–148. [PubMed: 26462961]

(31). Calabrese DR, Zlotkowski K, Alden S, Hewitt WM, Connelly CM, Wilson RM, Gaikwad S, Chen L, Guha R, Thomas CJ, Mock BA, and Schneekloth JS (2018) Characterization of clinically used oral antiseptics as quadruplex-binding ligands. Nucleic Acids Res 46, 2722. [PubMed: 29481610]

(32). Connelly CM, Abulwerdi FA, and Schneekloth JS Jr. (2017) Discovery of RNA Binding Small Molecules Using Small Molecule Microarrays. Methods Mol. Biol 1518, 157–175. [PubMed: 27873206]

(33). Guy CT, Cardiff RD, and Muller WJ (1992) Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. Mol. Cell. Biol 12, 954–961. [PubMed: 1312220]

(34). Diermeier SD, and Spector DL (2017) Antisense Oligonucleotide-mediated Knockdown in Mammary Tumor Organoids. Bio Protoc 7, 1 DOI: 10.21769/BioProtoc.2511.

(35). Brown JA (11 2018) Department of Chemistry and Biochemistry, University of Notre Dame, Personal communication. Notre Dame, IN

(36). Ageeli AA, McGovern-Gooch KR, Kaminska MM, and Baird NJ (2018) Finely tuned conformational dynamics regulate the protective function of the lncRNA MALAT1 triple helix. Nucleic Acids Res, 1 DOI: 10.1093/nar/gky1171.

(37). Baird NJ, Inglese J, and Ferre-D'Amare AR (2015) Rapid RNA-ligand interaction analysis through high-information content conformational and stability landscapes. Nat. Commun 6, 8898. [PubMed: 26638992]

(38). Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, and Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem 30, 2785–2791. [PubMed: 19399780]

(39). Blakeley BD, DePorter SM, Mohan U, Burai R, Tolbert BS, and McNaughton BR (2012) Methods for identifying and characterizing interactions involving RNA. Tetrahedron 68, 8837–8855.

(40). Wilusz JE (2016) Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. Biochim. Biophys. Acta, Gene Regul. Mech 1859, 128–138.

(41). Becker W, Bhattiprolu KC, Gubensak N, and Zangger K (2018) Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. ChemPhysChem 19, 895–906. [PubMed: 29314603]

(42). Brown JA, Kinzig CG, DeGregorio SJ, and Steitz JA (2016) Methyltransferase-like protein 16 binds the 3′-terminal triple helix of MALAT1 long noncoding RNA. Proc. Natl. Acad. Sci. U. S. A 113, 14013–14018. [PubMed: 27872311]

(43). Mathews Griner LA, Guha R, Shinn P, Young RM, Keller JM, Liu D, Goldlust IS, Yasgar A, McKnight C, Boxer MB, Duveau DY, Jiang JK, Michael S, Mierzwa T, Huang W, Walsh MJ, Mott BT, Patel P, Leister W, Maloney DJ, Leclair CA, Rai G, Jadhav A, Peyser BD, Austin CP, Martin SE, Simeonov A, Ferrer M, Staudt LM, and Thomas CJ (2014) High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. Proc. Natl. Acad. Sci. U. S. A 111, 2349–2354. [PubMed: 24469833]

(44). Moody MR, Morris MJ, Young VM, Moye LA, Schimpff SC, and Wiernik PH (1978) Effect of two cancer chemotherapeutic agents on the antibacterial activity of three antimicrobial agents. Antimicrob. Agents Chemother 14, 737–742. [PubMed: 103494]

(45). Livak KJ, and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408. [PubMed: 11846609]

(46). Ewald AJ (2013) Isolation of mouse mammary organoids for long-term time-lapse imaging. Cold Spring Harb Protoc 2013, 130–133. [PubMed: 23378653]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(47). Kao C, Zheng M, and Rudisser S (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. RNA 5, 1268–1272. [PubMed: 10496227]

(48). Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr., Onufriev A, Simmerling C, Wang B, and Woods RJ (2005) The Amber biomolecular simulation programs. J. Comput. Chem 26, 1668–1688. [PubMed: 16200636]

(49). Schrödinger. (2012) LigPrep, 2.5 ed, Schrödinger, LLC, New York, NY.

(50). Gasteiger J, and Marsili M (1980) Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges. Tetrahedron 36, 3219–3228.

(51). Humphrey W, Dalke A, and Schulten K (1996) VMD: Visual molecular dynamics. J. Mol. Graphics 14, 33–38.

(52). Mayer M, and Meyer B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J. Am. Chem. Soc 123, 6108–6117. [PubMed: 11414845]

(53). Hwang TL, and Shaka AJ (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. J. Magn. Reson., Ser. A 112, 275–279.

(54). Mitton-Fry RM, DeGregorio SJ, Wang J, Steitz TA, and Steitz JA (2010) Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. Science 330, 1244–1247. [PubMed: 21109672]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

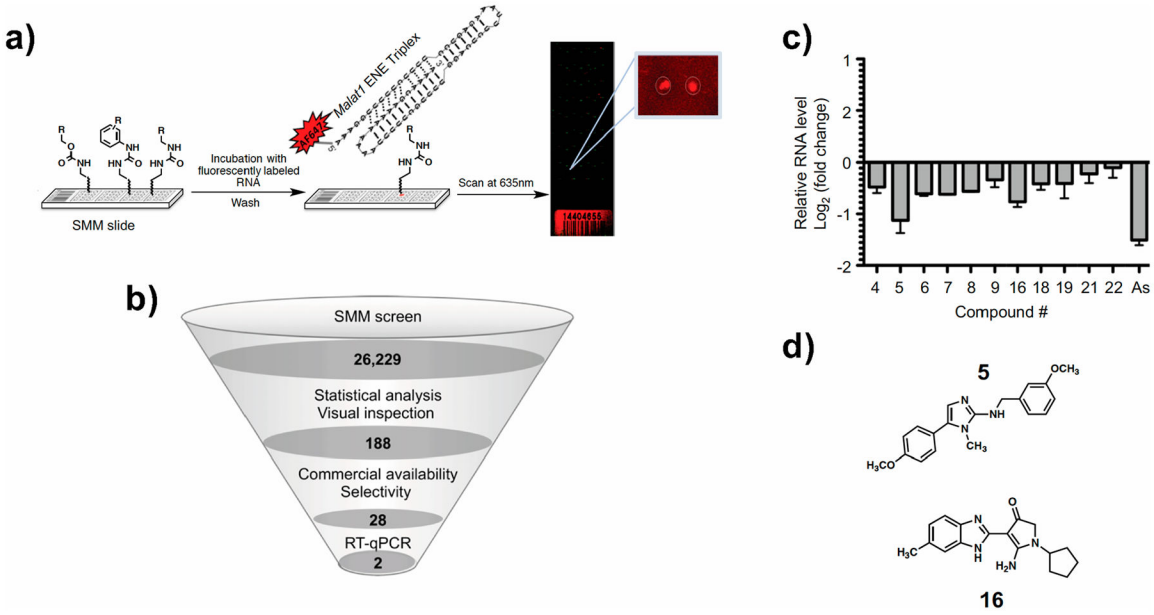


Figure 1.

HTS identification of compounds **5** and **16** as *MALAT1* ENE triplex-binding ligands. (a) Schematic of SMM screening. Libraries of small molecules are spatially arrayed and covalently linked to a functionalized glass surface. For screening, fluorescently labeled *Malat1* triple helix was incubated with the slides. After incubation, slides were washed to remove unbound oligonucleotide and imaged at 635 nm. (b) Pipeline developed to identify and validate compounds **5** and **16**. (c) RT-qPCR of relative *Malat1* levels in MMTV-PyMT tumors organoids with treatments of mock, DMSO, and indicated compounds (final concentration of 1 μM) and *Malat1* antisense oligo (As) (final concentration of 200 nM). n = 3 biological replicates; bars represent plus or minus the standard error of the mean (SEM). (d) The structures of compounds **5** and **16**.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

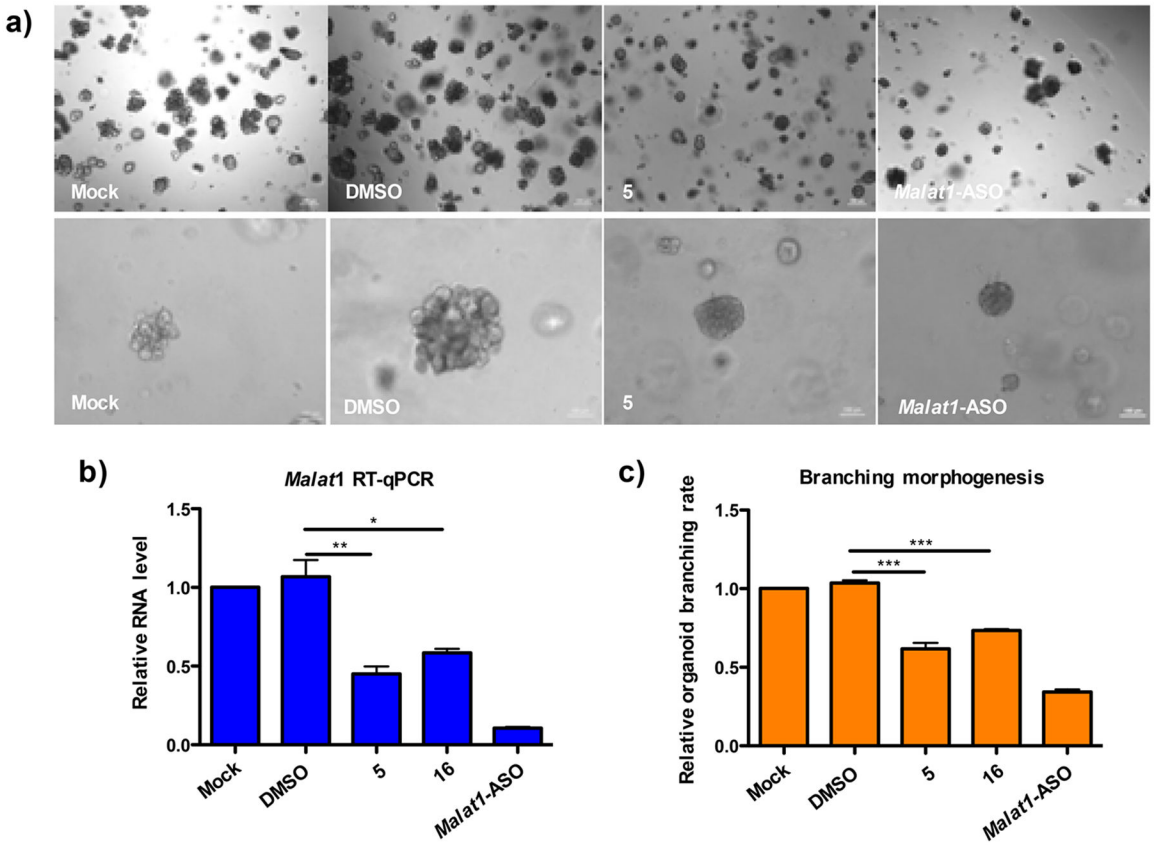


Figure 2.

Reduction of *Malat1* levels and organoid branching by compounds **5** and **16**. (a) Image of organoids from MMTV-PyMT tumors after 7 days of culturing. The data for compound 5 are shown. (b) Relative *Malat1* levels in MMTV-PyMT tumor organoids with treatments of Mock, DMSO, compounds **5** and **16** (final concentration of 1 μM), and *Malat1*-ASO (final concentration of 200 nM). $n = 3$ biological replicates; bars represent plus or minus SEM. (c) Relative organoid branching rate of MMTV-PyMT tumor organoids with treatment of mock, DMSO, compounds **5** and **16** (final concentration of 1 μM), and *Malat1*-ASO (final concentration of 200 nM). $n = >80$ organoids from 3 biological replicates; bars represent plus or minus the SEM. Single asterisks indicate $P < 0.05$, double asterisks indicate $P < 0.01$, and triple asterisks indicate $P < 0.001$ by Student's *t* test.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

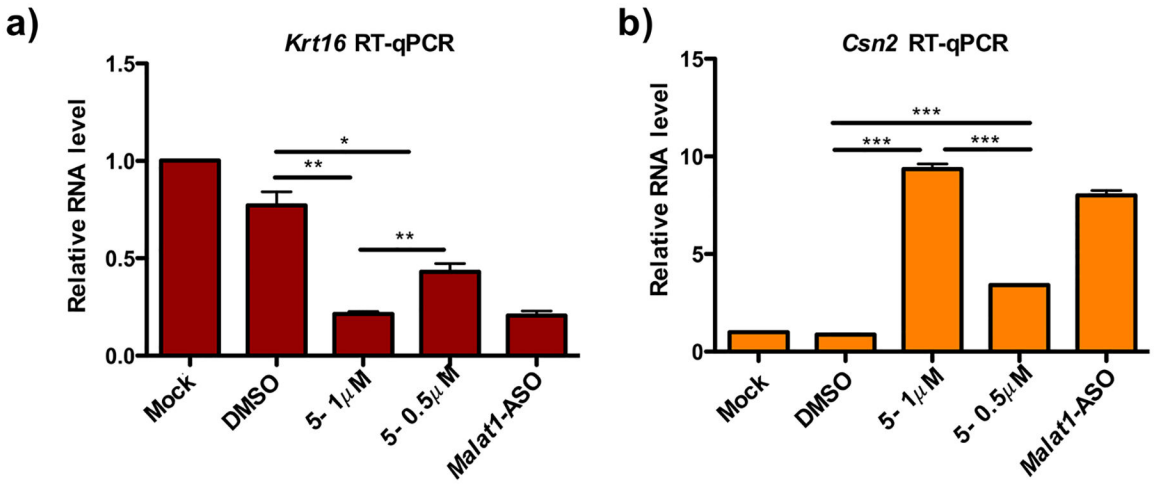


Figure 3.

Modulation of *Malat1* downstream genes by compound **5**. (a) Relative RNA level of *Krt16* in MMTV-PyMT tumor organoids with treatment of mock, DMSO, 1 and 0.5 μM of **5**, and *Malat1*-ASO (final concentration of 200 nM). $n = 3$ biological replicates; bars represent plus or minus SEM. Single asterisks indicate $P < 0.05$ and double asterisks indicate $P < 0.01$ by Student's *t* test. (b) Relative RNA level of *Csn2* in MMTV-PyMT tumor organoids with treatment of mock, DMSO, 1 and 0.5 μM of **5**, and *Malat1*-ASO (final concentration of 200 nM). $n = 3$ biological replicates; bars represent plus or minus SEM. Triple asterisks indicate $P < 0.001$ by Student's *t* test.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

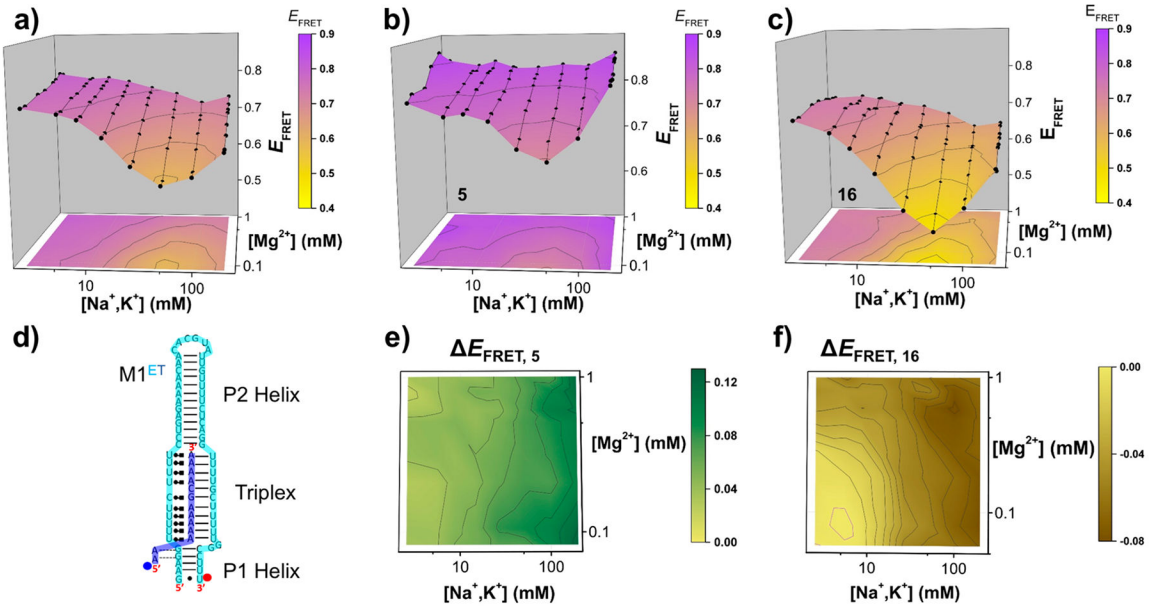


Figure 4.

Induction of opposing structural responses in the ENE triple helix by compounds **5** and **16**. (a) EFRET signal of $M1^{ET}$ in the presence of vehicle (DMSO) alone reveals the conformational landscape of this triplex across multiplexed ionic conditions. The folded triplex has a higher E_{FRET} than partially unfolded triplex. The triplex conformational landscape is raised in the presence of (b) 10 μM compound **5**, and primarily lowered in the presence of (c) 50 μM compound **16**. In panels a–c, E_{FRET} was evaluated as a function of total monovalent concentrations (equimolar KCl and NaCl) from 2.6–202.6 mM and magnesium concentrations from 0.1–1 mM. All experiments were performed in triplicate (3 independent experiments), but a single set of 8 × 8 experiments is represented in each panel. (d) Secondary structure of $M1^{ET}$ (ENE in cyan and Tail in blue) with the Cy3 and Cy5 positions illustrated by red and blue circles, respectively. Difference plots of E_{FRET} in the presence and absence of (e) 10 μM compound **5** and (f) 50 μM **16**. Difference plots were calculated by subtracting E_{FRET} values in panel a from the values in panels b and c, respectively. An average of difference plots performed in triplicate (three independent experiments) is plotted in panels e and f. Compound **5** addition increases E_{FRET} (green shading), while compound **16** decreases E_{FRET} (brown shading) under all salt conditions.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

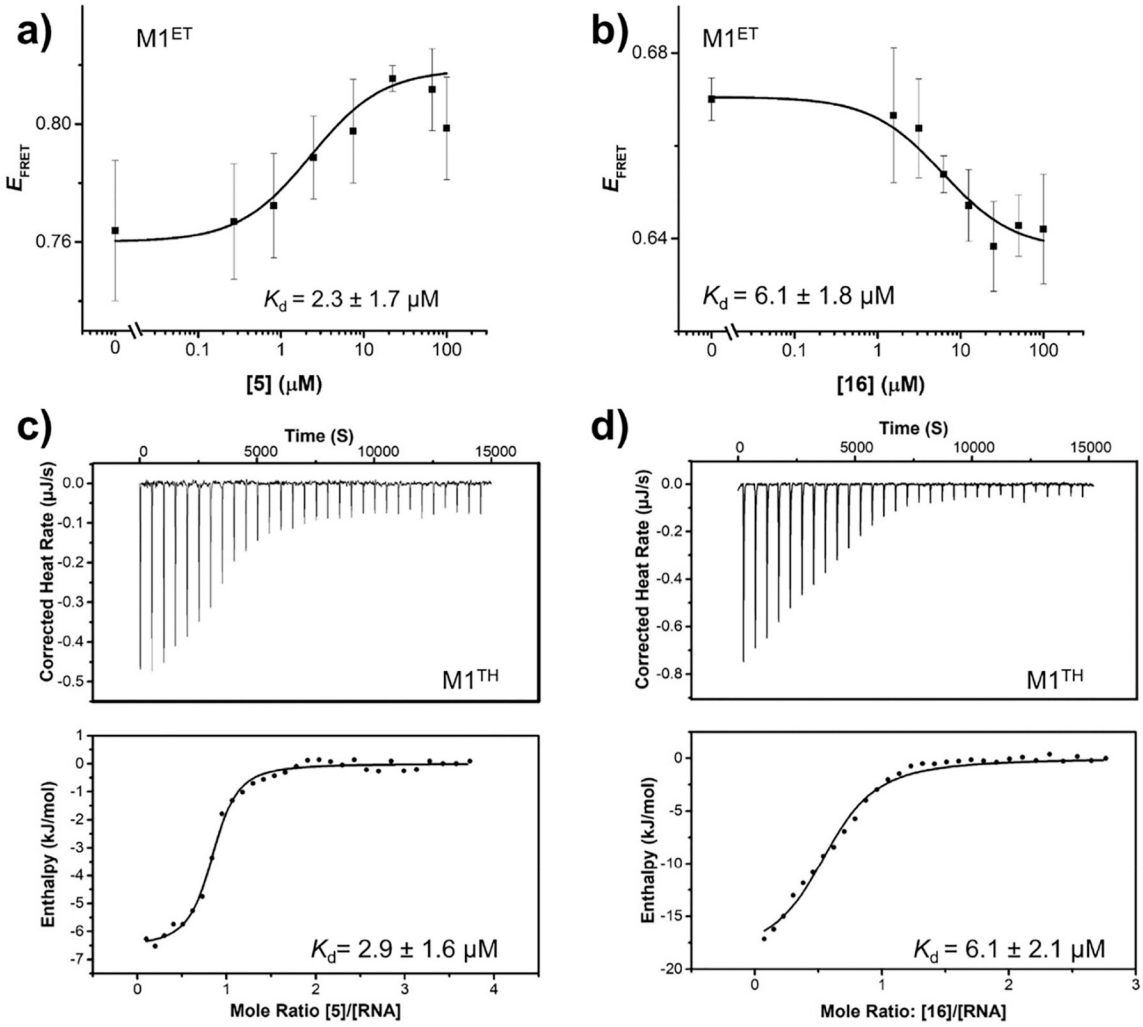


Figure 5.

Evaluating compounds **5** and **16** binding to the *MALAT1* ENE by FRET and ITC. (a) Compound **5** titration monitored by changes in E_{FRET} in 0.1 mM $MgCl_2$ and 2.6 mM total monovalent salt. The fit yields K_d = 2.3 ± 1.7 μM. (b) Compound **16** titration monitored by changes in E_{FRET} in 0.1 mM [Mg^{2+}] and 52.6 mM total monovalent salt. The fit yields K_d = 6.1 ± 1.8 μM. (c) ITC analysis using compound **5** (500 μM) and wild-type $M1^{TH}$ (75 μM) in a 1 mM $MgCl_2$, 200 mM monovalent salt, and 1% DMSO in 20 mM HEPES-KOH; pH of 6.9 indicates K_d = 2.9 ± 1.6 μM. (d) ITC analysis using compound **16** (500 μM) and wild-type $M1^{TH}$ (75 μM) in a buffer containing 1 mM $MgCl_2$, 50 mM monovalent salt, and 1% DMSO in 20 mM HEPES-KOH; pH of 6.9 indicates K_d = 6.1 ± 2.1 μM. Error bars in all panels represent standard deviation of three independent experiments for all experiments except in panel d, which represents the standard error of the mean for two experiments.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

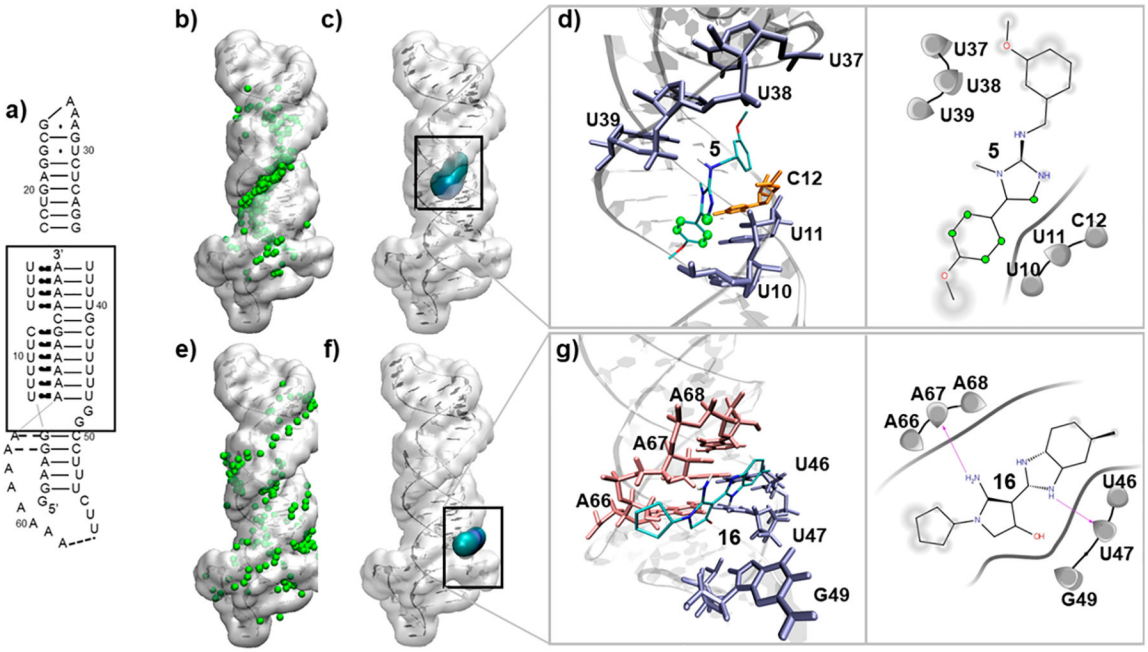


Figure 6.

Molecular docking of compounds **5** and **16** to the *MALAT1* ENE triplex core crystal structure. Three-dimensional structures are shown in a transparent surface and gray ribbon representation (PDBID: 4PLX). (a) *MALAT1* sequence, within which the ENE is boxed. (b) The distribution of 246 independent dockings, shown as green van der Waals spheres, for compound **5** reveals a preference for major groove binding. (c) Highest populated docking cluster (and lowest energy) for compound **5** shown in surface representation buried within the triplex. (d) The interactions with the backbone phosphates and bases of nucleotides U10, U11, and C12 of the triplex Hoogsteen strand. Longer-range interactions are made with backbone phosphates of U37, U38, and U39 on the triplex Watson–Crick strand. The C11, C13, C14, and C15 atoms of compound **5** are shown as green CPK spheres interacting explicitly with U10, U11, and C12 of the triplex Hoogsteen strand, consistent with saturation transfer difference (STD) NMR experiments (see Figure 8). (e) Distribution of 156 dockings of compound **16** indicates primarily surface or minor groove interactions. (f) Cluster distribution analysis for compound **16** shows the highest populated cluster (44% of dockings) is located in a superficial pocket formed by the G48−G49 bulge. (g) Within this “bulge cleft”, compound **16** is positioned to make backbone phosphate hydrogen bond interactions with A67 and carbonyl nucleobase hydrogen bond interactions with U47 (purple arrows). The solvent accessibilities of compounds **5** and **16** are shown as gray densities in the 2D interaction plots.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

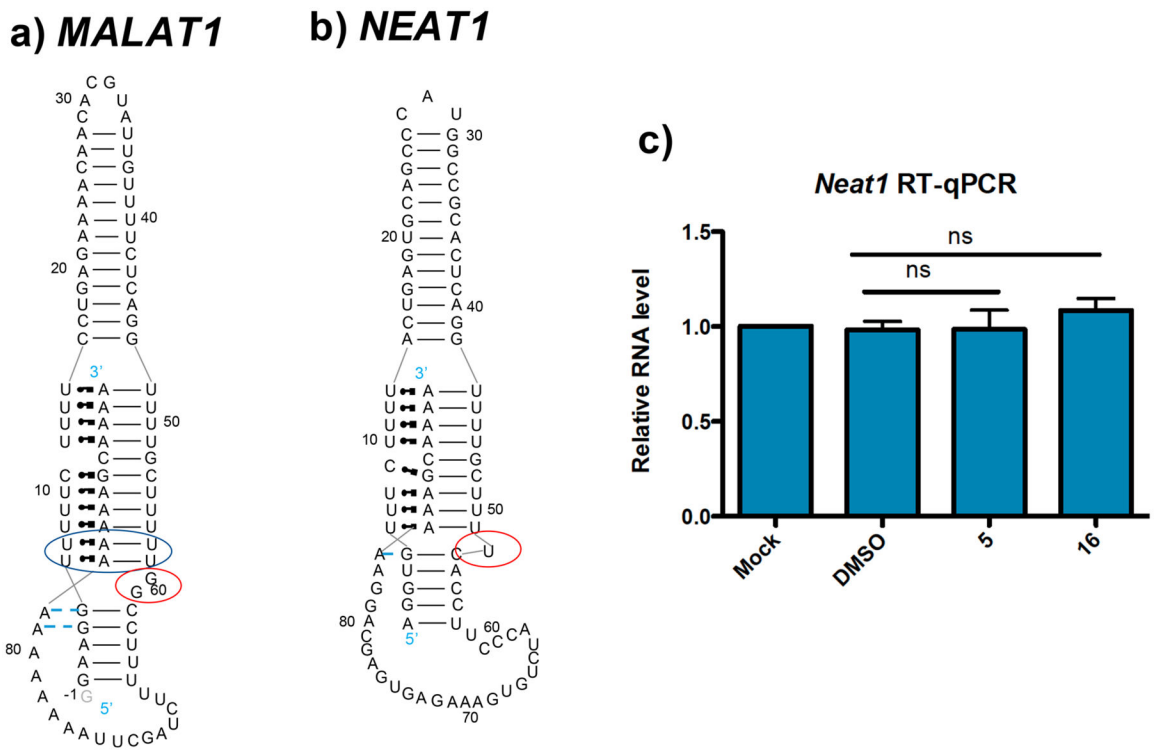


Figure 7.

No effect on *Neat1* levels in organoids treated with compounds **5** and **16**. (a, b) Secondary structures of *MALAT1* ENE and lncRNA *NEAT1* triple helices, respectively. The pair of regions highlighted are significantly different between, namely (i) the bulge nucleotide region (red circle) and (ii) the number of U·AU base triples (blue circle). It unclear from the secondary structure if a cleft (binding site for compound 16 on the *MALAT1* ENE triplex) is present in the *NEAT1* structure and a crystal structure of the *NEAT1* triplex is unavailable. Additionally, the deletion of two U·AU base triples within the triple helix of *NEAT1* render a much shorter potential U·AU region (binding site for compound **5** on *MALAT1* ENE triplex). (c) RT-qPCR of relative *Neat1* levels in MMTV-PyMT tumors organoids with treatments of mock, DMSO, and compounds **5** and **16** (final concentration of 1 μM). $n = 3$ biological replicates; bars represent plus or minus SEM. Ns refers to no statistically significant differences in RNA levels.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

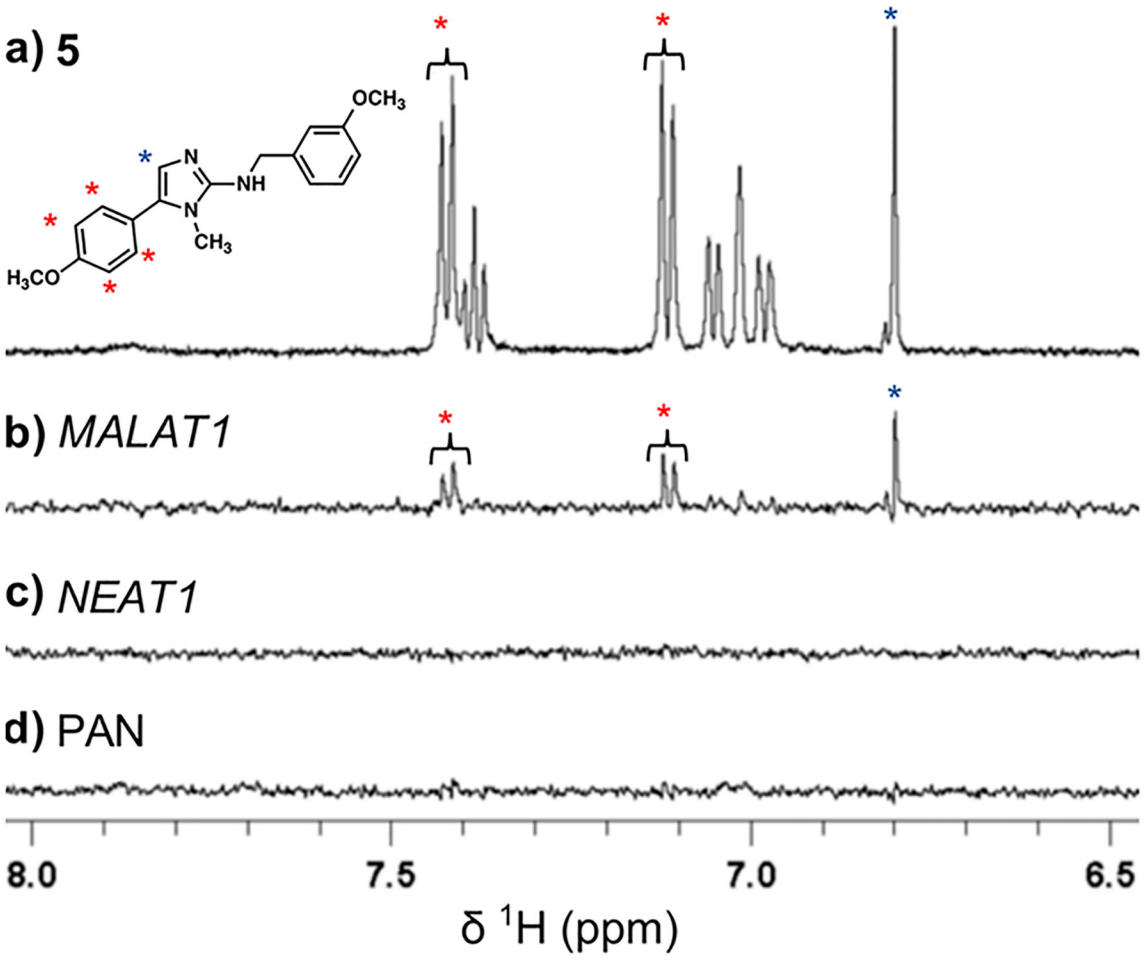


Figure 8.

Saturation transfer difference (STD) NMR confirmation of compound **5** specificity for the *MALAT1* ENE triple helix. (a) 1D ^{1}H NMR spectrum for compound 5 alone. (b–d) STD spectra for compound **5** following incubation with the ENE triplexes of *MALAT1*, *NEAT1*, and KSHV PAN, respectively. The protons on the *para*-methoxy aryl ring (highlighted in red) appear as two doublets at 7.12 and 7.42 ppm, while the proton on methyl substituted imidazole (highlighted in blue) appears as a singlet at 6.8 ppm. These peaks show an STD signal in the presence of *MALAT1* triplex, suggesting that this part of the molecule is interacting with the RNA.