Targeting RNA with small molecules: lessons learned from Xist RNA

ELLIOTT B. NICKBARG,[1] KERRIE B. SPENCER,[1] JONATHAN D. MORTISON,[1] and JEANNIE T. LEE[2,3]

[1]Merck & Co., Inc., Boston, Massachusetts 02115, USA

[2]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

[3]Department of Genetics, The Blavatnik Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

ABSTRACT

Although more than 98% of the human genome is noncoding, nearly all drugs on the market target one of about 700 disease-related proteins. However, an increasing number of diseases are now being attributed to noncoding RNA and the ability to target them would vastly expand the chemical space for drug development. We recently devised a screening strategy based upon affinity-selection mass spectrometry and succeeded in identifying bioactive compounds for the noncoding RNA prototype, Xist. One such compound, termed X1, has drug-like properties and binds specifically to the RepA motif of Xist in vitro and in vivo. Small-angle X-ray scattering analysis reveals that X1 changes the conformation of RepA in solution, thereby explaining the displacement of cognate interacting protein factors (PRC2 and SPEN) and inhibition of X-chromosome inactivation. In this Perspective, we discuss lessons learned from these proof-of-concept experiments and suggest that RNA can be systematically targeted by drug-like compounds to disrupt RNA structure and function.

Keywords: RNA targeting; RNA therapeutics; Xist RNA; small molecules

INTRODUCTION

Greater than 98% of our genome does not code for protein and, within the remaining 2% of sequence space, no more than 25,000 proteins are believed to be encoded (The ENCODE Project Consortium 2007). Despite the fact that most of the human transcriptome is noncoding (Kung et al. 2013), nearly all commercially available drugs target one of the ~700 disease-related proteins (Santos et al. 2017). These proteins account for only 0.05% of the genome, but still represent the major focus of R&D efforts in the pharmaceutical industry. However, an increasing number of diseases are now known to originate in noncoding space (Zhang and Lupski 2015; Disney 2019). The sheer size of the noncoding transcriptome and the growing potential to encode disease variants make RNAs appealing new targets for drug discovery. The ability to target RNA directly would therefore vastly increase the chemical space for drug development.

Several reasons may explain the historical reluctance to invest in noncoding space (Connelly et al. 2016; Warner et al. 2018; Ursu et al. 2019): First, RNA has been historically considered to be a passive intermediary between genome and the proteome. More recent work has led to the realization that RNA plays essential regulatory roles at both transcriptional and translational levels, and has functions in RNA modification, chromatin structure modification, as well as other processes (Andergassen and Rinn 2022). Second, until recently, there has been little progress in development of small molecule drugs known to be targeting RNA, and conventional high-throughput screening strategies have proven to have low success (Aboul-ela 2010). Third, RNAs have been considered "undruggable" because pharmacological probing of macromolecules generally requires the drug target to adopt a stable conformation (Santos et al. 2017).

With structural stability being a requirement for target engagement, RNAs have rated poorly because of the capacity to assume many secondary and tertiary structures of similar stabilities. In contrast, proteins are generally conformationally stable and therefore rate more favorably as drug targets. RNA's poor druggability also stems from the fact that determination of their tridimensional structure inside cells remains challenging (Connelly et al. 2016; Warner et al. 2018; Ursu et al. 2019). Few RNA structures

Corresponding authors: lee@molbio.mgh.harvard.edu, elliott.nickbarg@merck.com

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.079523.122. Freely available online through the *RNA* Open Access option.

© 2023 Nickbarg et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

have been solved to date despite recent improvements in chemical probing and imaging methods (Warner et al. 2018). Some clinical success has been achieved by targeting RNA with complementary nucleic acids, such as antisense oligonucleotides (ASOs). However, although easier to design and offering greater selectivity, ASOs are more difficult to deliver than small molecules and cell penetration of these large, highly charged pharmaceuticals can be an obstacle, whereas drug-like small molecules of <500 daltons are generally easier to optimize for delivery and tissue penetration. The potential for oral delivery, greater systemic exposure, and better cell and tissue penetration are among the advantages of small molecule drugs. For these reasons, the quest to find RNA-binding small molecules has intensified in the past several years (Connelly et al. 2016; Rizvi and Smith 2017; Disney et al. 2018; Warner et al. 2018).

Indeed, the number of examples of RNA-binding small molecules and approaches used to identify them have grown in recent years (Childs-Disney et al. 2022). Initial examples emerged from phenotypic screens that were actually not specific to RNA targets. For example, the antibacterial Ribocil selectively targets the flavin mononucleotide riboswitch (Howe et al. 2015, 2016), and the SM1 splice modulators, risdiplam and branaplam, are approved drugs to treat spinal muscular atrophy. These drugs work to bind the SMN2 pre-mRNA to bias splicing patterns (Palacino et al. 2015; Baranello et al. 2021). Both were initially progressed for their phenotypic activity and favorable drug properties, and only later determined to be RNA-binding compounds (Palacino et al. 2015; Sivaramakrishnan et al. 2017). Regardless, these advances have validated the idea that the chemical space for small molecule drug development can be expanded much beyond the proteome to include RNA targets (Howe et al. 2015; Palacino et al. 2015; Sivaramakrishnan et al. 2017; Aguilar et al. 2022).

A more direct approach has been to target structured RNAs identified by experimental structure-based approaches such as X-ray crystallography, NMR spectroscopy, CryoEM, or to analyze RNA sequence. Models are then generated using in silico computational approaches and chemical matter extrapolated from these models to exploit RNA features such as hairpins, pockets, and helices (Childs-Disney et al. 2022). The ability of single-stranded RNA to base-pair with itself and thereby fold into diverse sets of tertiary structures may limit the reliability of such predictive approaches. Relatively few RNAs have high-quality experimental structural data that are useful for drug development, and structural data may not adequately reflect the conformational dynamics of the folded structure in solution or the effects of protein partners or other components that may modulate actual functionality in vivo. It is anticipated that in silico predictions will require extensive follow-up experimental confirmation. Nevertheless, although rational design is still in its early days, the approach may become more feasible in the future as we gain a better understanding of how RNAs fold in vivo and better algorithms become available to predict their folding (Fauman et al. 2011; Disney et al. 2016; Abulwerdi et al. 2019).

A third approach toward identifying RNA-targeting compounds involves affinity-based screening of RNA-binding compounds, such as Affinity-selection mass spectrometry (AS-MS), which could be performed without detailed knowledge of an RNA's structure. In this Perspective, we discuss our recent success in targeting a long noncoding RNA using such a structure-agnostic method. We then discuss lessons learned from the proof-of-concept and how the AS-MS approach can become a generally applicable, scalable, and reproducible method for discovering new chemical matter in the RNA-targeting space.

TARGETING XIST RNA IN A PROOF-OF-CONCEPT

Our proof-of-concept work focused efforts around Xist, as it is an RNA with established function and is arguably one of the best characterized long noncoding RNAs. Xist is a 17 kb transcript produced from one of the two X chromosomes in female mammals to initiate dosage compensation of X-linked genes (Brown et al. 1992; Wutz et al. 2002; Starmer and Magnuson 2009; Disteche 2012; Lee 2012). Among its unusual properties is the ability to spread and localize to the inactive X chromosome (Xi) in *cis* to recruit repressive complexes for XCI (Brown et al. 1992; Zhao et al. 2008). A priori, identifying small molecules based on structure-enabled rational design would have required considerable upfront work lasting (likely) years due to the large size of Xist RNA and the suspicion that the RNA is substantially unstructured despite the presence of conserved modular domains (Smola et al. 2016). Despite its noncoding nature and overall modest sequence conservation among mammalian species, Xist contains various better-conserved motifs, including "Repeat A" (RepA)—a 431-nucleotide domain consisting of 8.5 units of a GC-rich motif (Fig. 1A; Brown et al. 1992; Wutz et al. 2002). RepA is essential for Xist function (Wutz et al. 2002; Hoki et al. 2009; Nesterova et al. 2019; Colognori et al. 2020).

Because reliable assays for RepA's silencing function are established, we chose to screen for compounds that target RepA. Notably, prior efforts to deduce the structure of RepA have relied on SHAPE and DMS-seq methodologies, but these studies have not arrived at a consensus (Maenner et al. 2010; Fang et al. 2015; Smola et al. 2016; Liu et al. 2017; Rivas et al. 2017; Kirk et al. 2018; Jones and Sattler 2019). Furthermore, our recent imaging of RepA using small angle X-ray scattering (SAXS) revealed that RepA RNA in solution can adopt at least 16 different families of conformations in vitro (Aguilar et al. 2022), potentially explaining why methods that rely on population averaging have yielded dramatically different results.

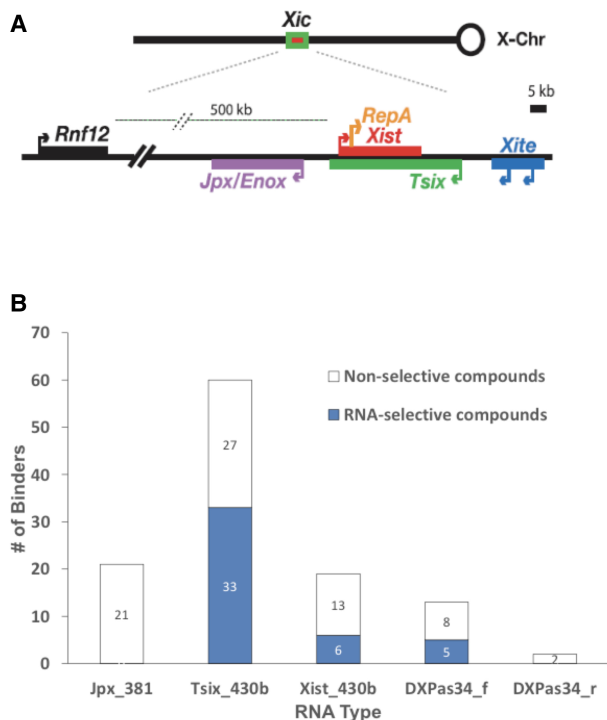

FIGURE 1. Selective and nonselective compounds identified in a screen for compounds targeting noncoding RNAs of the X-inactivation center. (*A*) The X-inactivation center and its noncoding genes (Tian et al. 2010). (*B*) Numbers of hits binding to each X-linked RNA from the original AS-MS screen, showing the proportion of binders that were selective to each RNA versus the total set of 41 other RNAs that were screened (Rizvi et al. 2020).

These limitations accentuate the need for structure-agnostic approaches to small molecule screening for RNA targets. We therefore turned to AS-MS—specifically a high-throughput automated ligand identification system (ALIS)—that initially relied upon ligand binding to RNA directly and then followed through with functional and structural analyses to differentiate and characterize phenotypically relevant ligands that disrupt the silencing function of RepA. In screening 50,000 compounds, we identified the small molecule, X1 (Aguilar et al. 2022). The X1 compound has drug-like properties, with a molecular weight of 416 Da, four hydrogen bond donors, and one hydrogen bond acceptor. We demonstrated that X1 specifically binds to RepA in vitro and in vivo, alters the conformation of RepA RNA, and displaces two known interacting protein factors (PRC2 and SPEN). By interfering with PRC2 and SPEN interactions, X1 blocks histone H3 lysine-27 trimethylation (H3K27me3) and abrogates the initiation of X-chromosome inactivation (Aguilar et al. 2022). The study thereby demonstrates that RNA can be systematically targeted by drug-like compounds that disrupt RNA structure and epigenetic function. Below, we enumerate lessons learned from these proof-of-concept experiments.

LESSONS LEARNED

Advantages of the AS-MS screening method

Affinity-selection mass spectrometry (AS-MS) is a general term that encompasses both direct and indirect target–ligand complex detection techniques (Annis et al. 2007). AS-MS measures binding of small molecules using an affinity-selection stage, in which a potential ligand or mixture of ligands is equilibrated with the target macromolecule to form a target–ligand complex which is then separated from the nonbinding components. The binding ligands are then identified by MS. Direct approaches use the mass spectrometer to separate the target–ligand complex from the unbound components in the gas phase, and indirect techniques, such as ALIS ("automated ligand identification system"), typically use a chromatographic or ultrafiltration-based system to separate the target–ligand complex away from the unbound species in solution, and the mass spectrometer to denature the complex and identify the small molecule ligand after release from the target (Annis et al. 2007).

Direct methods, such as DOLCE-MS (Greig and Robinson 2000) and MASS (Hofstadler and Sannes-Lowery 2006), have the advantage of providing direct observation of the ligand binding to the target. However, preservation of noncovalent target–ligand binding under gas phase conditions may require removal of excess nonvolatile salts, detergents, buffers or other solution components and may affect functional target folding and stability. Generally, high-resolution MS is necessary to detect ligand binding, which may limit the number and type of targets that are amenable to screening, as RNA targets that are too large or heterogeneous may not be compatible. The requirement for assay conditions that are amenable to gas-phase detection of the native state of a target–ligand complex renders direct AS-MS a more challenging and limited screening technique (Bergsdorf and Ottl 2010). As a result, direct AS-MS methods have not been generally adopted for high throughput screening approaches in drug discovery.

In contrast, indirect AS-MS methods such as ALIS equilibrate a target molecule with a mixture of molecules in a solution that is compatible with functional activity. The target–ligand complexes are then rapidly separated from the pool of nonbinding compounds using size-exclusion chromatography (SEC), captured, and then transferred to reverse-phase chromatography to denature the complex and remove potentially interfering salts and other buffer components. The target–ligand complexes are denatured during the reverse-phase step and small molecule ligands are released and identified by mass spectrometry. Target-compound binding is done in solution with no tagging of either the target molecule or the screening collection required, and separation and capture of the target–

ligand complex is done under physiologically relevant conditions using SEC. AS-MS is a binding assay and therefore identifies ligands independently of their functional effect, which potentially allows for ligands acting at novel sites to be identified. There are several types of indirect approaches which have been reviewed previously (Annis et al. 2007; Bergsdorf and Ottl 2010). The applications and use of indirect AS-MS have now become an established drug discovery approach with numerous applications involving a wide target and compound space (Prudent et al. 2021). Although originally developed for screening traditional drug target proteins, the approach has been shown to enable screening of large collections of drug-like small molecules against additional types of RNA-based targets.

Initial application of the ALIS AS-MS platform to screening noncoding RNAs used the FMN bacterial riboswitch as a model system, since this target had known drug-like active binders (Howe et al. 2015; Rizvi et al. 2018). The riboswitches proved to have sufficient stability and were generally of excellent solubility, so that no modifications to the screening system itself and only minor precautions (e.g., use of RNase-free buffers) were necessary to achieve stable screening conditions (Rizvi and Nickbarg 2019). However, screening results revealed that even a relatively compact RNA such as a riboswitch could adopt different conformational states causing different ligands to bind in the same pocket but with different phenotypic consequences (Rizvi et al. 2018). Indeed, recent work with the TPP riboswitch shows that RNAs can exhibit significant cooperativity effects upon binding of ligands, resulting in substantial structural rearrangements (Zeller et al. 2022). A binding-based screening method such as ALIS can sample multiple conformational states and can provide hits across these states, whereas structure-based methods may miss or deprioritize hits based upon such interactions. However, the AS-MS system merely provides binding data and must be linked to appropriate functional and structural follow-up approaches so that hits of the desired properties can be identified. Indeed, some hits within the expanded set also bound RepA, but had no biochemical or biological activity as measured in vitro and in vivo (Aguilar et al. 2022).

Selection of RNA targets and hit rates

Target selection is a critical aspect of every drug discovery effort. Although we succeeded in identifying small molecule binders to RepA of Xist, we initially considered a number of other possible targets, including other X-inactivation-related RNA targets, for the proof-of-concept effort (Rizvi et al. 2020). The original expanded set of 42 RNA targets were selected without knowledge of detailed RNA structure to represent a broad range of RNA types from a variety of disease areas. Disease-relevant noncoding RNA (ncRNA) transcripts containing single-nucleotide polymorphisms (SNPs) were chosen by examination of genome-wide association studies (GWAS) and literature data, and included mammalian lncRNA, ncRNAs known to bind to RNA-binding proteins, G-quadruplexes, RNA repeat elements, noncoding splice variants, segments of mammalian mRNA containing structural elements in the 3′ or 5′ untranslated regions, and some bacterial and viral ncRNA elements (Rizvi et al. 2020). These RNAs were each screened against a diversity library of approximately 50,000 drug-like small molecule compounds and a functionally annotated library of ∼5100 compounds that were drawn from the larger Merck corporate screening collection. Screening of the set of RNA targets gave a total of 1424 hits, of which 944 were RNA-selective (binding to only a single target out of the 42 screened) and 545 were both RNA-selective and not binding to any of the proteins included in the screening comparison data set (Rizvi et al. 2020). It is important to note that the AS-MS technique used here is designed to discriminate against weakly bound ligands (K_d approximately <10 µM) and therefore the sets of hits included binders with affinities that can range from submicromolar to low micromolar. The observed numbers of hits for the various RNA targets in the AS-MS approach were also quite variable. Two of the 42 targets failed to bind any compounds while the G-quadruplex targets exhibited hit rates that were 10-fold higher than other classes of RNAs. We did not functionally characterize the hits to the 42 targets in great detail, however, as most of these RNAs did not lend themselves to easy characterization, due to the lack of established biochemical and functional assays.

On the other hand, as one of the primary purposes of the original AS-MS screening effort was to characterize the chemical properties of the RNA binding compound set, we confirmed hits from the screen of the RNA targets and compared them to small molecule interactors from a comparative protein AS-MS screening data set. Naïve Bayesian models for chemical properties that bias small molecules toward RNA binding were then derived. The resulting chemical features were used to assemble an additional set of ∼3700 compounds that were enriched in compounds containing RNA-binding features, as demonstrated by elevated hit rates in subsequent AS-MS screening. The hit compound set therefore consists of compounds binding with a range of affinities against the included RNA and protein targets. The data showed that there was considerable overlap in the types of compounds binding between the RNA-binding and protein-binding compound sets, but it was possible to identify chemical features that were more prevalent among compounds that were binding selectively to RNA targets (Rizvi et al. 2020).

Interestingly, RepA RNA bound 19 of the compounds in the original 1424 hit set. Of these binders, six were RNA-selective, binding to a single RNA from the set of 42 targets that were screened (Rizvi et al. 2020). This is similar to the average number of compounds binding (23) and average

number of RNA-selective binders (9) binding among the five X-linked RNAs that were screened, including Jpx and Tsix (Fig. 1B; Rizvi et al. 2020). Of the binders detected during screening, only one—X22—proved to have significant activity in the various cellular and biochemical assays used to characterize function (Aguilar et al. 2022). X22 has a molecular weight of 382 Da and exhibits reasonable pharmacological properties and drug-likeness (Lipinski 2004), as well as a reasonable affinity (K_d approximately greater than 25 µM) for a first-pass molecule. From an expansion set around X22, ALIS was further used to confirm binding via competition studies and to determine binding affinities in order to rank order and prioritize candidates among the 20 expanded compounds. Fifteen compounds demonstrated a spectrum of binding affinities to RepA RNA, whereas the remaining five compounds did not bind detectably. Among the 15, X1 emerged as the highest affinity binder for Repeat A, with an affinity (K_d) of 0.4 ± 0.3 µM.

Thus, although current data are limited, it appears that a raw hit rate to a given RNA in the ALIS AS-MS system can be estimated. The average hit rate for non-G-quadruplex RNA targets with our 50,000 compound diversity set is currently estimated to be ∼0.01%, whereas protein targets were 0.05% from the same compound set. Though we obtained only one to two bona fide selective binders (X1, X22) to RepA with functional activity in cells, the low hit rate should not discourage further work on RNA targets. Indeed, we screened only a small diversity set of 50,000 compounds. Typical drug screens cover a library of 1 million or more compounds. The low hit frequency may also reflect the fact that the diversity set of 50,000 was originally curated against protein targets. One approach of future research should therefore be to develop diversity sets of RNA-selective compounds.

RNA is sufficiently stable for AS-MS approaches

RNAs are notoriously prone to degradation and can be more difficult to handle than proteins. However, we note that, by taking proper precautionary measures, all of the tested RNAs were of sufficient stability to withstand short-term storage, preparatory work for the ALIS runs, and the screening runs themselves. In the case of RepA, RNA samples were prepared as close to the time of screen as possible. RepA templates were in vitro transcribed plasmids using the T7 expression system, treated with DNase to remove contaminating plasmid DNA, purified by size exclusion chromatography, and resulting RNA fractions were concentrated using Amicon 100 kDa MWCO devices. Care was taken to preserve the native structure of the freshly transcribed RNAs (Chillon et al. 2015) by storing the RNA at 4°C and performing all screens within 24 h. It is possible that some RNAs will be stable enough to withstand additional purification steps, such as rounds of denaturation and renaturation. The size of the pooled end-products and their integrity were confirmed by denaturing urea 6% polyacrylamide gel electrophoresis. It was furthermore confirmed that handling during the screen through AS-MS did not result in loss of RNA integrity. We therefore conclude that RNAs such as RepA are sufficiently stable for a high-throughput automated system such as ALIS and are viable targets for future drug discovery efforts based on AS-MS.

Effect of RNA size on small molecule binding

Our analysis of the screening data across the set of non-G-quadruplex RNA targets showed no relationship between RNA target size or target type with number of binding compounds. The number of compounds binding did not increase with RNA target length, suggesting that compounds are not binding in a linear RNA-size dependent fashion and binding instead may be mediated by more complex secondary and tertiary structural elements. G-quadruplex RNAs had substantially higher hit rates, but most of these binders could be eliminated by changing screening salt conditions from high [K^+] to high [Na^+]—conditions that destabilize quadruplex structure and favor hairpin formation. It is therefore likely that the differences in hit rates observed across our set of RNA targets are dependent upon the overall folded RNA target structures for the individual targets, rather than RNA size (Rizvi et al. 2020).

Pipeline for hit validation

The screen against the original 42 RNA targets returned a total of 1424 hits after screening a library of 50,000 compounds representing the structural diversity and bioactivity space covered by the Merck total sample collection (Rizvi et al. 2020). In the case of X-linked RNAs including Xist, Tsix, and Jpx, we obtained a total of 115 hits. When faced with so many hits, a systematic approach is needed to separate selective from non-selective binders and functional from nonfunctional compounds in order to determine which are worthy of further research and development, such as depicted in Figure 2. Following up the initial hits was somewhat challenging, as most of the RNA targets were structurally uncharacterized and remain uncharacterized with regard to protein interactors and cellular targets. Most lacked established functional assays.

Thus, we suggest initiating screens only on RNAs with well-characterized functions, ideally with known mechanisms of action. Moreover, although the AS-MS approach is structure-agnostic, some prior knowledge of the RNA's functional domains could be very helpful. In the case of Xist, the 17-kb size precludes in vitro transcription of the full length RNA. AS-MS also does not lend itself to screening with such large transcripts. Thus, knowledge of conserved and functional domains was especially important. Mutational analyses have determined how each

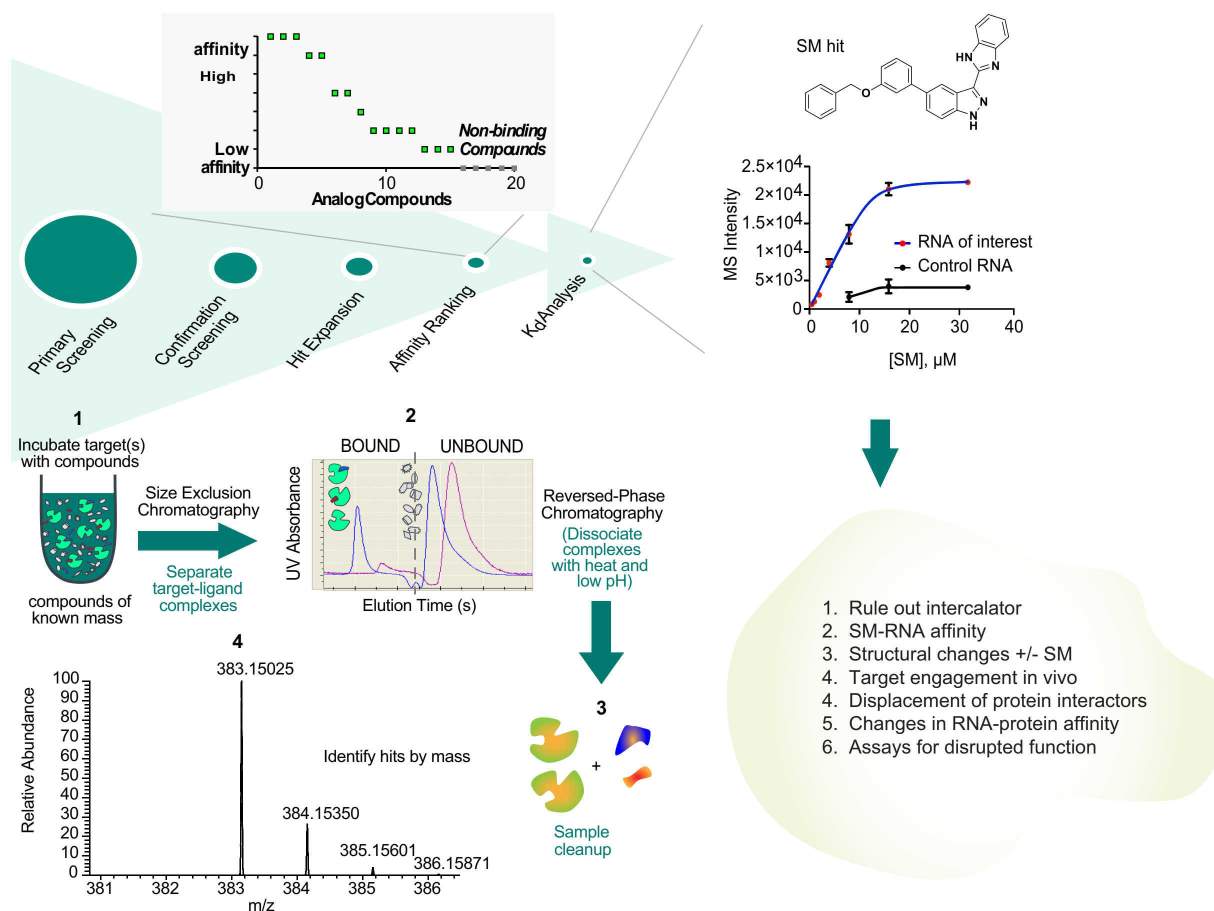

FIGURE 2. The AS-MS pipeline for identifying RNA-targeting small molecules. Shown is our ALIS screening of 50,000 diverse small molecule compounds against RNA, as was performed in Aguilar et al. (2022). In the primary screen (step 1), the compound mixture is equilibrated with target RNA, and the bound RNA–ligand complex is separated from unbound ligands by size exclusion chromatography (step 2). During reverse-phase chromatography, ligand and RNA are separated (step 3) prior to mass spectrometry to identify the hit compound (step 4). After the primary screen, the hits are confirmed, a hit expansion is conducted around confirmed hits, and MS-based affinity ranking performed to identify the top binders. In the depicted hypothetical graph, the *y*-axis represents the relative binding affinity of each compound to the target RNA, as determined by ALIS following the MS step. The top SM hits are then subjected to a series of assays (#1–6) to identify a functional disruptor. The diagram is adapted from Aguilar et al. (2022) (with permission from Nature Publishing Group).

domain contributes crucially to various aspects of XCI and what protein partners may be bound to them (Wutz et al. 2002; Colognori et al. 2019; Lee et al. 2019; Colognori et al. 2020; Dixon-McDougall and Brown 2022). RepA was an obvious target because multiple studies have confirmed a requirement for initiating XCI silencing, and biochemical and cellular studies have identified a large suite of interacting protein partners, including PRC2 and SPEN (Zhao et al. 2008; McHugh et al. 2015; Monfort et al. 2015). A number of assays are also available for testing biological effects and confirming mechanisms of action (Fig. 2).

After identifying X1 as a hit in the AS-MS screen (Aguilar et al. 2022), we first ruled out X1 as a nonspecific intercalator of nucleic acids, as evidenced by a low percent effect of −11.7 at 40 µM in a Thiazole Orange displacement assay. We used mass spec to obtain a preliminary affinity for X1's interaction with RepA and revealed that it has high specificity for RepA relative to 40 other RNA species. We tested X1 against a panel of kinases to rule out kinase activity, as kinase inhibitors are well-represented in the Merck diversity set. A small set of analogs was synthesized to confirm activity and establish some initial structure–activity relationships. Inactive analogs are also critical for mechanistic studies, so we identified both close analogs and structurally distinct compounds to serve as inactive controls. One issue we identified with X1 and its analogs was poor solubility above 50 µM in aqueous media, thus making the determination of affinities somewhat challenging. For X1, we would propose future work to include synthesizing a set of compounds aimed at addressing the solubility challenge with predicted solubility as a prioritization metric. Improved solubility could enable additional characterization such as SPR binding measurements.

To determine whether X1 has biological activity related to X-chromosome inactivation, we performed a set of biochemical tests. First, we asked if X1 specifically displaces two of Xist's required interacting protein partners, Polycomb repressive complex 2 (PRC2) and SPEN. RNA electrophoretic mobility assays (EMSA) revealed that the presence of the X1 compound substantially reduces the affinity of RepA for PRC2 from a K_d of 47 to 420 nM (IC_{50} of 30 µM) and SPEN, and for SPEN from 76 to 717 nM (IC_{50} of 48 µM). In vivo, RNA immunoprecipitation (RIP) using anti-PRC2 and anti-SPEN antibodies also showed that X1 disrupts the interaction between RepA and PRC2. Validating that X1 physically engaged RepA RNA in vivo was naturally more difficult. We designed a proxy measurement by creating a tritiated version of X1 and examining the number of H^3 counts recovered in RNA pulldown experiments from cells expressing wild-type Xist versus cells expressing an RNA version lacking the RepA domain. Indeed, the recovered counts were significantly reduced in the mutant Xist RNA, strongly suggesting that the RepA domain is the binding site for X1. These observations raised our confidence in X1 as a compound that not only binds Xist but also affects its biological activity inside cells.

Using small angle X-ray scattering studies, we found that X1 binding changes RepA's RNA conformation and proposed that the conformational change explains X1's ability to displace PRC2 and SPEN. Consistent with this effect, ChIP-seq demonstrated that X1 blocks histone H3 lysine-27 trimethylation (H3K27me3) specifically on the inactive X chromosome (Xi), and RNA-seq revealed that drug treatment inhibits the initiation of whole-chromosome gene silencing. The epigenomic analyses showed that the effects were specific to the Xi and that the active X (Xa) was not affected. We also applied cellular studies and found that retardation of growth and differentiation are female-specific, as would be expected if X1 targeted Xist to block a phenomenon that only occurs in females. Importantly, X1 did not inhibit the histone methyltransferase activity of PRC2 directly; rather it bound RepA and prevented the action of RepA in recruiting SPEN and PRC2. Thus, the panoply of available functional tests in the RepA study enabled us to determine that X1 was a selective inhibitor of Xist function.

CONCLUSIONS

The RepA study provided proof-of-concept that RNA can be drugged by a small molecule for epigenetic effects. Our experience with screening a much larger set of RNAs led us to conclude that drug discovery efforts best focus on RNA targets with known functions and a well-established set of downstream functional readouts. Once such targets are identified, we believe that RNA-binding compounds can be discovered for them using a generally applicable, scalable, and reproducible method to find novel chemical matter in the RNA-targeting space. For X1, further modifications of the tool compound through medicinal chemistry could increase potency and specificity, possibly leading to clinical candidates for treatment of X-linked diseases. Xist RNA has emerged as a key therapeutic target for Rett syndrome and other X-linked diseases through an X-reactivation strategy that would unsilence the wild-type allele of the causal gene on the Xi (Bhatnagar et al. 2014; Carrette et al. 2017; Sripathy et al. 2017; Grimm and Lee 2022), although such an approach would not be MECP2-specific (X-reactivation could affect other genes on the inactive X chromosome [Xi]). However, only a small fraction of Xi genes is increased in expression and the relative contribution of reactivation to the overall expression of other X-linked genes would be small in comparison to that of MECP2. Mouse experiments have also demonstrated the safety of X-reactivation in the brain (Carrette et al. 2017). With regard to developing X1 as a possible therapeutic, we emphasize that the same molecule may not work on human XIST RNA and X-chromosome inactivation, as mouse Xist and human XIST are only partially conserved in the RepA domain. Developing X1 would require synthesizing additional derivatives to explore structure–activity relationships in human cells and to improve compound solubility without impairing potency. Additional screening would also be advisable to identify new leads with better profiles.

Finally, unlike previously reported RNA-binding drugs, our hits did not arise from phenotypic screens (and retrospectively shown to bind RNA) (Howe et al. 2015; Sivaramakrishnan et al. 2017), but from a direct RNA-binding screen. A major advantage of ALIS is that it is agnostic to mechanism of action and function, and is therefore potentially applicable to any RNA, including those for which detailed structural knowledge is missing or where rational design proves challenging. Even as interest in identifying RNA-targeting compounds has increased over time, most efforts still focus on sequence- or structure-based design (Park et al. 2011; Stelzer et al. 2011; Nguyen et al. 2015; Barros et al. 2016; Disney et al. 2016; Luu et al. 2016; Abulwerdi et al. 2019). However, drug-binding pockets or tridimensional features are not available for the vast majority of transcripts, including Xist. In the case of Xist, the large size, the unstructured regions, and the many coexisting structural subpopulations have precluded derivation of a consensus structure (Maenner et al. 2010; Fang et al. 2015; Smola et al. 2016; Liu et al. 2017; Rivas et al. 2017; Kirk et al. 2018; Jones and Sattler 2019; Aguilar et al. 2022), rendering Xist an overall unfavorable candidate for rational design of small molecules. The ALIS method circumvented these prerequisites. Using ALIS, the overall number of hits to RNA was lower than typically observed for protein targets (Rizvi et al. 2020), but nevertheless resulted in a tool compound after conducting a hit expansion screen. The reduced hit frequency may

reflect the fact that the diversity set of 50,000 compounds was originally designed to hit protein targets. Additional future research will expand upon the diversity of compounds with RNA-binding properties and enhance the identification of RNA-targeting drugs. Indeed, recent reports from other laboratories have already begun to expand the knowledge base of properties of RNA-binding small molecules (Donlic et al. 2022).

COMPETING INTEREST STATEMENT

E.N., K.S., and J.D.M. are current employees of Merck & Co., Inc. and have stock or other financial interests in Merck & Co. J.T.L. is a cofounder of Fulcrum Therapeutics and is also a scientific advisor to Skyhawk Therapeutics.

ACKNOWLEDGMENTS

We acknowledge the Merck MINt Award and funding from the National Institutes of Health (R01-HD097665) to J.T.L.

REFERENCES

Aboul-ela F. 2010. Strategies for the design of RNA-binding small molecules. *Future Med Chem* **2:** 93–119. doi:10.4155/fmc.09.149

Abulwerdi FA, Xu W, Ageeli AA, Yonkunas MJ, Arun G, Nam H, Schneekloth JS Jr, Dayie TK, Spector D, Baird N, et al. 2019. Selective small-molecule targeting of a triple helix encoded by the long noncoding RNA, *MALAT1*. *ACS Chem Biol* **14:** 223–235. doi:10.1021/acschembio.8b00807

Aguilar R, Spencer KB, Kesner B, Rizvi NF, Badmalia MD, Mrozowich T, Mortison JD, Rivera C, Smith GF, Burchard J, et al. 2022. Targeting *Xist* with compounds that disrupt RNA structure and X inactivation. *Nature* **604:** 160–166. doi:10.1038/s41586-022-04537-z

Andergassen D, Rinn JL. 2022. From genotype to phenotype: genetics of mammalian long non-coding RNAs *in vivo*. *Nat Rev Genet* **23:** 229–243. doi:10.1038/s41576-021-00427-8

Annis DA, Nickbarg E, Yang X, Ziebell MR, Whitehurst CE. 2007. Affinity selection-mass spectrometry screening techniques for small molecule drug discovery. *Curr Opin Chem Biol* **11:** 518–526. doi:10.1016/j.cbpa.2007.07.011

Baranello G, Darras BT, Day JW, Deconinck N, Klein A, Masson R, Mercuri E, Rose K, El-Khairi M, Gerber M, et al. 2021. Risdiplam in type 1 spinal muscular atrophy. *N Engl J Med* **384:** 915–923. doi:10.1056/NEJMoa2009965

Barros SA, Yoon I, Chenoweth DM. 2016. Modulation of the *E. coli rpoH* temperature sensor with triptycene-based small molecules. *Angew Chem Int Ed Engl* **55:** 8258–8261. doi:10.1002/anie.201601626

Bergsdorf C, Ottl J. 2010. Affinity-based screening techniques: their impact and benefit to increase the number of high quality leads. *Expert Opin Drug Discov* **5:** 1095–1107. doi:10.1517/17460441.2010.524641

Bhatnagar S, Zhu X, Ou J, Lin L, Chamberlain L, Zhu LJ, Wajapeyee N, Green MR. 2014. Genetic and pharmacological reactivation of the mammalian inactive X chromosome. *Proc Natl Acad Sci* **111:** 12591–12598. doi:10.1073/pnas.1413620111

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. 1992. The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71:** 527–542. doi:10.1016/0092-8674(92)90520-M

Carrette LLG, Wang CY, Wei C, Press W, Ma W, Kelleher RJ III, Lee JT III. 2017. A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. *Proc Natl Acad Sci* **115:** E668–E675. doi:10.1073/pnas.1715124115

Childs-Disney JL, Yang X, Gibaut QMR, Tong Y, Batey RT, Disney MD. 2022. Targeting RNA structures with small molecules. *Nat Rev Drug Discov* **21:** 736–762. doi:10.1038/s41573-022-00521-4

Chillon I, Marcia M, Legiewicz M, Liu F, Somarowthu S, Pyle AM. 2015. Native purification and analysis of long RNAs. *Methods Enzymol* **558:** 3–37. doi:10.1016/bs.mie.2015.01.008

Colognori D, Sunwoo H, Kriz AJ, Wang CY, Lee JT. 2019. *Xist* deletional analysis reveals an interdependency between Xist RNA and polycomb complexes for spreading along the inactive X. *Mol Cell* **74:** 101–117.e110. doi:10.1016/j.molcel.2019.01.015

Colognori D, Sunwoo H, Wang D, Wang CY, Lee JT. 2020. Xist repeats A and B account for two distinct phases of X inactivation establishment. *Dev Cell* **54:** 21–32.e25. doi:10.1016/j.devcel.2020.05.021

Connelly CM, Moon MH, Schneekloth JS Jr. 2016. The emerging role of RNA as a therapeutic target for small molecules. *Cell Chem Biol* **23:** 1077–1090. doi:10.1016/j.chembiol.2016.05.021

Disney MD. 2019. Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. *J Am Chem Soc* **141:** 6776–6790. doi:10.1021/jacs.8b13419

Disney MD, Winkelsas AM, Velagapudi SP, Southern M, Fallahi M, Childs-Disney JL. 2016. Inforna 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chem Biol* **11:** 1720–1728. doi:10.1021/acschembio.6b00001

Disney MD, Dwyer BG, Childs-Disney JL. 2018. Drugging the RNA world. *Cold Spring Harb Perspect Biol* **10:** a034769. doi:10.1101/cshperspect.a034769

Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annu Rev Genet* **46:** 537–560. doi:10.1146/annurev-genet-110711-155454

Dixon-McDougall T, Brown CJ. 2022. Multiple distinct domains of human XIST are required to coordinate gene silencing and subsequent heterochromatin formation. *Epigenetics Chromatin* **15:** 6. doi:10.1186/s13072-022-00438-7

Donlic A, Swanson EG, Chiu LY, Wicks SL, Juru AU, Cai Z, Kassam K, Laudeman C, Sanaba BG, Sugarman A, et al. 2022. R-BIND 2.0: an updated database of bioactive RNA-targeting small molecules and associated RNA secondary structures. *ACS Chem Biol* **17:** 1556–1566. doi:10.1021/acschembio.2c00224

The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816. doi:10.1038/nature05874

Fang R, Moss WN, Rutenberg-Schoenberg M, Simon MD. 2015. Probing Xist RNA structure in cells using targeted structure-seq. *PLoS Genet* **11:** e1005668. doi:10.1371/journal.pgen.1005668

Fauman EB, Rai BK, Huang ES. 2011. Structure-based druggability assessment—identifying suitable targets for small molecule therapeutics. *Curr Opin Chem Biol* **15:** 463–468. doi:10.1016/j.cbpa.2011.05.020

Greig MJ, Robinson JM. 2000. Detection of oligonucleotide-ligand complexes by ESI-MS (DOLCE-MS) as a component of high throughput screening. *J Biomol Screen* **5:** 441–454. doi:10.1177/108705710000500607

Grimm NB, Lee JT. 2022. Selective Xi reactivation and alternative methods to restore MECP2 function in Rett syndrome. *Trends Genet* **38:** 920–943. doi:10.1016/j.tig.2022.01.007

Hofstadler SA, Sannes-Lowery KA. 2006. Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes. *Nat Rev Drug Discov* **5:** 585–595. doi:10.1038/nrd2083

Hoki Y, Kimura N, Kanbayashi M, Amakawa Y, Ohhata T, Sasaki H, Sado T. 2009. A proximal conserved repeat in the *Xist* gene is essential as a genomic element for X-inactivation in mouse. *Development* **136:** 139–146. doi:10.1242/dev.026427

Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, Galgoci AM, Malinverni JC, Mayhood T, Villafania A, Nahvi A, et al. 2015. Selective small-molecule inhibition of an RNA structural element. *Nature* **526:** 672–677. doi:10.1038/nature15542

Howe JA, Xiao L, Fischmann TO, Wang H, Tang H, Villafania A, Zhang R, Barbieri CM, Roemer T. 2016. Atomic resolution mechanistic studies of ribocil: a highly selective unnatural ligand mimic of the *E. coli* FMN riboswitch. *RNA Biol* **13:** 946–954. doi:10.1080/15476286.2016.1216304

Jones AN, Sattler M. 2019. Challenges and perspectives for structural biology of lncRNAs—the example of the Xist lncRNA A-repeats. *J Mol Cell Biol* **11:** 845–859. doi:10.1093/jmcb/mjz086

Kirk JM, Kim SO, Inoue K, Smola MJ, Lee DM, Schertzer MD, Wooten JS, Baker AR, Sprague D, Collins DW, et al. 2018. Functional classification of long non-coding RNAs by *k*-mer content. *Nat Genet* **50:** 1474–1482. doi:10.1038/s41588-018-0207-8

Kung JT, Colognori D, Lee JT. 2013. Long noncoding RNAs: past, present, and future. *Genetics* **193:** 651–669. doi:10.1534/genetics.112.146704

Lee JT. 2012. Epigenetic regulation by long noncoding RNAs. *Science* **338:** 1435–1439. doi:10.1126/science.1231776

Lee HJ, Gopalappa R, Sunwoo H, Choi SW, Ramakrishna S, Lee JT, Kim HH, Nam JW. 2019. En bloc and segmental deletions of human XIST reveal X chromosome inactivation-involving RNA elements. *Nucleic Acids Res* **47:** 3875–3887.

Lipinski CA. 2004. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* **1:** 337–341. doi:10.1016/j.ddtec.2004.11.007

Liu F, Somarowthu S, Pyle AM. 2017. Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat Chem Biol* **13:** 282–289. doi:10.1038/nchembio.2272

Luu LM, Nguyen L, Peng S, Lee J, Lee HY, Wong CH, Hergenrother PJ, Chan HY, Zimmerman SC. 2016. A potent inhibitor of protein sequestration by expanded triplet (CUG) repeats that shows phenotypic improvements in a *Drosophila* model of myotonic dystrophy. *ChemMedChem* **11:** 1428–1435. doi:10.1002/cmdc.201600081

Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, Dubois A, Sanglier-Cianferani S, Van Dorsselaer A, Clerc P, Avner P, et al. 2010. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol* **8:** e1000276. doi:10.1371/journal.pbio.1000276

McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, et al. 2015. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521:** 232–236. doi:10.1038/nature14443

Monfort A, Di Minin G, Postlmayr A, Freimann R, Arieti F, Thore S, Wutz A. 2015. Identification of *Spen* as a crucial factor for *Xist* function through forward genetic screening in haploid embryonic stem cells. *Cell Rep* **12:** 554–561. doi:10.1016/j.celrep.2015.06.067

Nesterova TB, Wei G, Coker H, Pintacuda G, Bowness JS, Zhang T, Almeida M, Bloechl B, Moindrot B, Carter EJ, et al. 2019. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat Commun* **10:** 3129. doi:10.1038/s41467-019-11171-3

Nguyen L, Luu LM, Peng S, Serrano JF, Chan HY, Zimmerman SC. 2015. Rationally designed small molecules that target both the DNA and RNA causing myotonic dystrophy type 1. *J Am Chem Soc* **137:** 14180–14189. doi:10.1021/jacs.5b09266

Palacino J, Swalley SE, Song C, Cheung AK, Shu L, Zhang X, Van Hoosear M, Shin Y, Chin DN, Keller CG, et al. 2015. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat Chem Biol* **11:** 511–517. doi:10.1038/nchembio.1837

Park SJ, Kim YG, Park HJ. 2011. Identification of RNA pseudoknot-binding ligand that inhibits the −1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J Am Chem Soc* **133:** 10094–10100. doi:10.1021/ja1098325

Prudent R, Annis DA, Dandliker PJ, Ortholand JY, Roche D. 2021. Exploring new targets and chemical space with affinity selection-mass spectrometry. *Nat Rev Chem* **5:** 62–71. doi:10.1038/s41570-020-00229-2

Rivas E, Clements J, Eddy SR. 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* **14:** 45–48. doi:10.1038/nmeth.4066

Rizvi NF, Nickbarg EB. 2019. RNA-ALIS: methodology for screening soluble RNAs as small molecule targets using ALIS affinity-selection mass spectrometry. *Methods* **167:** 28–38. doi:10.1016/j.ymeth.2019.04.024

Rizvi NF, Smith GF. 2017. RNA as a small molecule druggable target. *Bioorg Med Chem Lett* **27:** 5083–5088. doi:10.1016/j.bmcl.2017.10.052

Rizvi NF, Howe JA, Nahvi A, Klein DJ, Fischmann TO, Kim HY, McCoy MA, Walker SS, Hruza A, Richards MP, et al. 2018. Discovery of selective RNA-binding small molecules by affinity-selection mass spectrometry. *ACS Chem Biol* **13:** 820–831. doi:10.1021/acschembio.7b01013

Rizvi NF, Santa Maria JP, Nahvi A, Klappenbach J, Klein DJ, Curran PJ, Richards MP, Chamberlin C, Saradjian P, Burchard J, et al. 2020. Targeting RNA with small molecules: identification of selective, RNA-binding small molecules occupying drug-like chemical space. *SLAS Discov* **25:** 384–396. doi:10.1177/2472555219885373

Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, et al. 2017. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* **16:** 19–34. doi:10.1038/nrd.2016.230

Sivaramakrishnan M, McCarthy KD, Campagne S, Huber S, Meier S, Augustin A, Heckel T, Meistermann H, Hug MN, Birrer P, et al. 2017. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat Commun* **8:** 1476. doi:10.1038/s41467-017-01559-4

Smola MJ, Christy TW, Inoue K, Nicholson CO, Friedersdorf M, Keene JD, Lee DM, Calabrese JM, Weeks KM. 2016. SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the *Xist* lncRNA in living cells. *Proc Natl Acad Sci* **113:** 10322–10327. doi:10.1073/pnas.1600008113

Sripathy S, Leko V, Adrianse RL, Loe T, Foss EJ, Dalrymple E, Lao U, Gatbonton-Schwager T, Carter KT, Payer B, et al. 2017. Screen for reactivation of MeCP2 on the inactive X chromosome identifies the BMP/TGF-β superfamily as a regulator of XIST expression. *Proc Natl Acad Sci* **114:** 1619–1624. doi:10.1073/pnas.1621356114

Starmer J, Magnuson T. 2009. A new model for random X chromosome inactivation. *Development* **136:** 1–10. doi:10.1242/dev.025908

Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, Andricioaei I, Markovitz DM, Al-Hashimi HM. 2011. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat Chem Biol* **7:** 553–559. doi:10.1038/nchembio.596

Tian D, Sun S, Lee JT. 2010. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143:** 390–403. doi:10.1016/j.cell.2010.09.049

Ursu A, Vezina-Dawod S, Disney MD. 2019. Methods to identify and optimize small molecules interacting with RNA (SMIRNAs). *Drug Discov Today* **24:** 2002–2016. doi:10.1016/j.drudis.2019.06.019

Warner KD, Hajdin CE, Weeks KM. 2018. Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* **17:** 547–558. doi:10.1038/nrd.2018.93

Wutz A, Rasmussen TP, Jaenisch R. 2002. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat Genet* **30:** 167–174. doi:10.1038/ng820

Zeller MJ, Nuthanakanti A, Li K, Aube J, Serganov A, Weeks KM. 2022. Subsite ligand recognition and cooperativity in the TPP riboswitch: implications for fragment-linking in RNA ligand discovery. *ACS Chem Biol* **17:** 438–448. doi:10.1021/acschembio.1c00880

Zhang F, Lupski JR. 2015. Non-coding genetic variants in human disease. *Hum Mol Genet* **24:** R102–R110. doi:10.1093/hmg/ddv259

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322:** 750–756. doi:10.1126/science.1163045