Targeting RNA structures with small molecules

Jessica L. Childs-Disney\(^1\), Xueyi Yang\(^1\), Quentin M. R. Gibaut\(^{1,7}\), Yuquan Tong\(^1\), Robert T. Batey\(^{2,3,4,5}\) and Matthew D. Disney\(^{1,7}\)

Abstract | RNA adopts 3D structures that confer varied functional roles in human biology and dysfunction in disease. Approaches to therapeutically target RNA structures with small molecules are being actively pursued, aided by key advances in the field including the development of computational tools that predict evolutionarily conserved RNA structures, as well as strategies that expand mode of action and facilitate interactions with cellular machinery. Existing RNA-targeted small molecules use a range of mechanisms including directing splicing — by acting as molecular glues with cellular proteins (such as branaplam and the FDA-approved risdiplam), inhibition of translation of undruggable proteins and deactivation of functional structures in noncoding RNAs. Here, we describe strategies to identify, validate and optimize small molecules that target the functional transcriptome, laying out a roadmap to advance these agents into the next decade.

Antisense oligonucleotides (ASOs). Single-stranded DNA or RNA oligonucleotides, including combinations thereof that are chemically modified, that are complementary to the sequence of target RNA and can induce RNAse H-mediated degradation or sterically block the ribosome.

Riboswitch Region of RNA structure located in the 5′ leader of bacterial RNAs that undergoes conformational switching upon small molecule binding to regulate translation.

The first evidence that RNA could be targeted for therapeutic intervention was provided by the discovery of streptomycin in 1944 (REF. \(^1\)) and the subsequent identification of the bacterial ribosome, a macromolecular RNA–protein complex, as the natural product’s cellular target\(^1\). Further insight was provided 20 years later when the first nucleic acid sequence and RNA structure — a tRNA carrying alanine (tRNA\(_{\text{Ala}}\)) — was reported\(^2\). The key to understanding the role of tRNA in translation was its 3D fold, required for aminoacylation and hence decoding of mRNA sequence into protein. The discovery of catalytic RNAs\(^3,4\) expanded the chemical functions of RNAs far beyond encoding proteins and decoding mRNAs.

The sequence of the human genome further reinforced RNA’s contribution to biology, with far fewer canonical open reading frames (ORFs) observed than expected\(^5\). Indeed, organismal complexity is not correlated with the number of ORFs but instead with the number and diversity of noncoding (nc)RNAs\(^6\) that function in epigenetic regulation\(^6\) and regulate gene expression\(^6\), particularly during development. Complementarily, genome-wide association studies (GWAS) have defined biological pathways that are dysregulated in disease, elucidating potential therapeutic targets, both RNA and protein. Such data can be leveraged to enable a bench-to-bedside paradigm in which small molecules bind to and deactivate structured RNAs: a patient’s genome could be sequenced and compared with GWAS to identify the malfunctioning RNA. The targeted therapy would bind a functional structure within the RNA to short-circuit disease pathways.

Promising therapeutic strategies to target RNA include antisense oligonucleotides (ASOs), CRISPR gene editing and small molecules that recognize RNA structures. ASOs and CRISPR editing have been invaluable to the field of chemical biology. However, translating these technologies to the clinic has been challenging owing to difficulties with delivery and significant adverse reactions\(^6,7\). Small molecules offer an important alternative with potential for oral bioavailability and blood–brain barrier penetration, particularly with the wealth of knowledge from medicinal chemistry whereby physicochemical properties can be systematically optimized to improve pharmacokinetics and potency.

Several small molecules that bind RNA structures have been identified that exhibit various modes of action (MOAs), from simple binding to direct cleavage to recruitment of endogenous nucleases (induced proximity). These molecules have been demonstrated to modulate diverse biological processes, such as inhibiting bacterial and viral translation (ribocil and a riboswitch in Escherichia coli\(^11\), 2-aminobenzimidazole derivatives and the hepatitis C internal ribosome entry site (IRES)\(^12\); inhibiting mRNA translation (synucleozid and α-synuclein\(^13\)); facilitating alternative pre-mRNA splicing (risdiplam\(^14\) and branaplam\(^15\) and survival of motor neuron 2 (SMN2)); and inhibiting microRNA (miRNA) biogenesis (a spermine–amide conjugate and the mir-372 precursor\(^16\)).

This Review describes foundational methods and strategies to: identify structured, functional RNAs; design and discover lead molecules that bind structured regions; and lead-optimize small molecules, including
Defining RNA structures for small-molecule targeting

Determining RNA structure

An accurate model of RNA structure is key to the design or discovery of small molecules that modulate its function. Computational methods can model RNA structure from sequence, including free energy minimization and phylogenetic comparison. Free energy minimization uses experimentally derived thermodynamic parameters to predict RNA structures, outputting the minimum free energy structure, assumed to be the functional structure, as well as suboptimal structures. The total free energy of an RNA structure is calculated by summing the free energy of each substructure in the system, such as base pairs, bulges and loops. Dynamic programming algorithms, the basis of programs such as mFold, RNAfold and RNAstructure, incorporate these experimentally determined thermodynamic parameters, predicting accurately ~70% of the base pairs for RNAs <700 nucleotides long.

Experimental constraints can be integrated into the algorithms to improve the reliability of the prediction. Unpaired nucleotides can be identified both in vitro and in vivo using chemical modification by dimethyl sulfate (DMS) or by selective 2'-hydroxyl acylation analysed by primer extension (SHAPE). Chemical modification either causes reverse transcriptase (RT), causing truncation of the cDNA (an "RT stop"), or induces a mutation, both of which can be read out by high-throughput sequencing. These sites of modification restrain secondary structure predictions in which the extent of modification is correlated with an energetic penalty for a nucleotide to be paired. Transcriptome-wide DMS and SHAPE probing techniques greatly expanded knowledge of the overall RNA structural profiles in cells and enabled investigation of RNA conformational changes under various biological conditions. Although these experimental constraints refine the predicted models, they do not provide information about the 3D fold of the RNA, as provided by X-ray crystallography, NMR spectroscopy and cryo-electron microscopy (cryo-EM).

Phylogenetic comparison also provides insight into RNA structure, where genetic differences that preserve secondary structure (covariations: for example, an AU base pair is replaced with GU or GC) can elucidate structure, and conservation suggests selective pressure to retain a functional structure. Covariation-based structural prediction of many RNA structures is highly accurate, as determined by comparison with crystal structures. However, owing to the large dataset required and the complexity of the analysis, automated covariation-based prediction methods remain challenging. Phylogenetic comparison has been coupled with other prediction methods, either in sequence or co-processed, including homology modelling, free energy minimization and chemical modification to improve accuracy, which relies on the quality of alignment and sequence availability.

The folding of RNA is hierarchical, and RNAs can form tertiary structures. Information about RNA 3D structure is especially valuable in uncovering the functional mechanism of RNAs and identifying druggable targets. Although 3D prediction methods are still in their infancy compared with those for proteins, currently available programs, including FARFAR2 (RNA analogue of ROSETTA for protein prediction), MC-Fold/MC-Sym and iFoldRNA, have shown promising results for predicting 3D structures from sequence. Scoring functions that estimate the accuracy of prediction results in which the native structure is unknown have also been developed, such as Rosetta, RASP and ARES. With the implementation of machine learning techniques, ARES outperforms the other two approaches. Root-mean-square deviation (r.m.s.d) analysis of predicted structures to known crystal structures demonstrated the power of this approach.

Biophysical methods such as NMR spectroscopy, X-ray crystallography and cryo-EM have also been extensively used to determine RNA structure. NMR spectroscopy studies also provide information about structural dynamics. Although cryo-EM is typically used to study the structure of molecules of large molecular weight, considerable effort is being exerted to develop approaches that can access smaller RNA structures, for example, as recently reported for a small riboswitch (<40 kDa). Computationally assisted cryo-EM has also been introduced to determine the global conformation of RNA molecules, although it does not provide atomic-level resolution.

Assessing the quality of RNA structural predictions

Many algorithms have been developed that predict RNA structure. However, without a known structure, it is hard to assess the reliability and accuracy of the prediction, which has been addressed by development of various statistical methods.

Partition function calculations have been incorporated into various structure prediction programs, including RNAfold, Sfold and CONTRAfold. A partition function contains all of the thermodynamic information for a system and quantifies the probability of the predicted structure or substructure therein. Statistics have also been applied to phylogenetic comparisons such as R-scape, which measures the statistical significance of evolutionary covariation, indicative of functionality. Although R-scape improved the annotation of the consensus structure in 5S rRNA from the Rfam database, it did not find significant covariation in the long noncoding (lnc)RNAs Hox antisense intergenic RNA (HOTAIR), steroid receptor RNA activator (SRA) or X-inactive specific transcript (Xist), although plausible structures have been predicted by other computational methods. The lack of functional structures in these lncRNAs did not stem from the lack of variance in the phylogenetic tree, nor does it imply that these lncRNAs do not fold. Instead, they likely form dynamic structures stabilized in the context of other interacting partners, pointing to the importance of knowing the limitations of predictive methods and of experimental validation.
Collectively, these methods demonstrate that predicted structures must be viewed through the lens of statistical power and/or rigour and tempered with in-depth dissection of their biological function. That is, these predictions are hypotheses until further validated.

**Defining functional RNA structures in the transcriptome**

Functional RNA structures can be found throughout a transcript from the 5′ end to the 3′ end, from untranslated regions to ORFs (Fig. 1). Although data suggest that ORFs are less structured than other regions, highly structured coding sequences have been discovered. Functional structures can be identified computationally (below) or experimentally by using ASOs that sterically block a functional structure or by mutational analysis.

Although many RNA structural prediction methods can provide accurate models of RNA secondary structure, they do not predict functionality. Functionality can be predicted by identifying regions of structure that are unusually thermodynamically stable compared with random sequences. The hypothesis is that these stable structures are evolutionarily retained because of selective pressure. ScanFold, a scanning window approach, uses these principles to identify potentially functional structures, accurately predicting known functional viral structures as well as predicting potential new functional structures in viral and human transcriptomes.

Structural conservation and genetics can also indicate function. For the former, structural motifs that are evolutionarily conserved across species likely have a biological function; for the latter, genetic mutations can cause gain or loss of function. Perhaps the best example of evolutionary conservation of structure and function is rRNA. Indeed, highly conserved structures can be found in bacteria (riboswitches), viruses (IRESs) and humans (IRESs, splicing regulatory elements).

The hunt for targetable, functional RNA structures, particularly in the human transcriptome, has only just begun. Thus far, RNA structures that have been effectively targeted with small molecules (binding produces a downstream biological response), participate...
Factors that affect the selectivity of small molecules targeting RNA

Ideally, a small molecule would be completely selective for its RNA target, but in practicality that is likely not required or even achievable. Historically, selective recognition of RNA by small molecules was thought intractable, owing to its perceived lack of structural diversity with only four building blocks, its anionic backbone and the lack of success in high-throughput screening campaigns. Selective recognition is possible as our fundamental understanding of the RNA structures that are targetable and the chemotypes that bind RNA has evolved, alongside advances in RNA structure modelling that provide insight into the functionality of structures.

Binding selectivity and functional selectivity are distinct. Several factors have been identified that drive the cellular (functional) selectivity of ligands that target RNA, including the uniqueness of the structure in the transcriptome, expression of the target compared with off-targets, the relative affinity of the small molecule for on- and off-targets, accessibility of the target site and the functionality of the binding site, where binding to non-functional sites is biologically silent. Regarding target expression, if two or more RNAs have the same binding site, the more highly expressed target will be more occupied by the compound. In the same vein, if two or more RNAs have different structures ligandable by the same compound, the most occupied target will be a composite of relative affinity and expression level. In cases where selectivity is not sufficient, compounds can be designed that target multiple sites within an RNA target simultaneously, thus overcoming the limitations of structural degeneracy (a structure is not unique in the transcriptome).

A few studies have completed transcriptome- and proteome-wide studies that provide insight into the selectivity of RNA-targeting small molecules. These studies have shown that small molecules can exert selective effects across the transcriptome and proteome. Notably, the observed changes and selectivities are similar to those observed for oligonucleotides. An analogue of the FDA-approved drug risidolipam, which binds to an RNA–protein manifold not solely an RNA, is selective, altering the levels of 12 of 11,174 transcripts and altering the alternative splicing of a subset of mRNAs, including the desired target. As observed for small molecules that target proteins, the off-targets for RNA-targeting small molecules could be other RNAs, DNA or proteins. Thus far, it appears that the scaffolds that bind to RNA are different from those that bind to proteins (Box 1), as also indicated by the lack of success in identifying selective RNA binders from small-molecule libraries designed for proteins.

**Box 1 | Commonalities amongst small molecules that bind RNA**

Ideally, small molecules that bind RNA targets would be drug-like and orally bioavailable, as defined by their physicochemical properties. Indeed, analysis of the physicochemical properties of drugs that target proteins has provided guidelines for drug development, such as the Lipinski’s ‘rule of five’. RNA, however, is very different from proteins in terms of its chemical composition, its highly electronegative surface potential and its limited buried surface area. It is therefore reasonable to infer that RNA-binding small molecules will have unique properties that do not necessarily fall into traditional rule of five guidelines. Notably, new modalities such as chimeric compounds that target proteins for degradation have many drug discovery efforts lying ‘outside of the rule of five’. Various studies have identified privileged scaffolds and chemotypes that confer affinity for RNA, for example, indole, 2-phenylindole, 2-phenyl benzimidazole, 2-phenylimidazole, methylpyrimidine-2,4-diamine and others. Further, comparison of the physicochemical properties of RNA-binding molecules with those of FDA-approved drugs available in DrugBank revealed distinct differences. In particular, RNA-binding compounds on average have lower octanol–water partition coefficients (LogPs) than protein-binding compounds (ranges from two separate analyses: 0.16 ± 0.5 versus 2.0 ± 3.5 [REF.195] and 1.02 versus 1.78 [REF.196], respectively), greater topological polar surface areas (156 ± 118 versus 92 ± 144 Å²), more hydrogen bond donors (5.2 ± 2.5 versus 2.0 ± 4) and acceptors (6.6 ± 6 versus 5.0 ± 6), and more heteroatom-containing aromatic rings (2.16 versus 1.25). An analysis by the Hargrove laboratory suggested that the shape of RNA binders is generally rod-like. Given the infancy of RNA targeting, these comparisons will likely change as new molecular interactions and activities are catalogued.

A previous analysis investigated the interplay between small-molecule affinity, target complexity and quantitative estimate of drug-likeness (QED). QED, a metric of compound quality, evaluates how closely each physicochemical property of a small molecule, measured on different scales, reflect the ideal value and combine them into a single score. The analysis suggested that the complexity of the RNA target scaled with the affinity for the targets and that drug-like chemical matter emerged when the biological activity of the complex structure, it does not necessarily indicate a priori that a selective small molecule can be designed or discovered. In other cases, where the RNA is locally structured (a common occurrence for mRNAs, for example), the lack of buried surface area (in comparison with a complex structure) to drive selective interactions could be overcome by binding to multiple structures within an RNA target simultaneously with a single molecule. Although such compounds may be outside of the rule of five, they can be optimized into orally bioavailable medicines.

**Strategies to identify small-molecule RNA binders**

**Target-centric approaches**

Various approaches can be used to find small molecules that bind RNA structures in vitro, including fluorescence-based assays, fluorescence resonance energy transfer (FRET)-based approaches and dynamic combinatorial screening. The methods described below are target centric, that is, the small molecule’s only choice for binding is a single or a few targets. As with any primary screening assay, secondary analyses are required to identify selective binders. For all fluorescence-based assays described below, care must be taken, as many compounds can interact with the dyes themselves or have emission properties that overlap with the fluorophores.

**Affinity mass spectrometry**: Affinity selection mass spectrometry (AS-MS) is a label-free method that allows the direct identification of target–ligand complexes by mass spectrometry after separation from unbound ligands
Methods to identify chemical matter that binds RNA

**a** ALIS

- Size-exclusion chromatography
- RNA target–small molecule complexes
- Complex dissociation and LC-MS

**b** Fluorescence-based assay

- TO-PRO-1 dye displacement
- TO-PRO-1–RNA complex
- 2-AP assay
- Adenine (A)
- Excitation
- Emission
- RNA–protein complex

**c** Microarray-based screening

- Immobilized small molecules
- Labelled RNA targets
- Incubate and wash

**d** Fragment mapping

- Cross-linking
- Purification tag and pull-down

**e** DNA-encoded chemical library

- Encoded library
- Encoded target
- Non-selective binding
- Selective binding
- RFU (582 nm)
- RFU (670 nm)

**f** Informa mining

- RNA target of interest
- Informa
- RNA fold
- Small molecule

**g** Structure-based design

- Docking

*Identify on- and off-targets
*Design selective compounds

by size-exclusion chromatography\(^{72,73}\). Used widely for proteins\(^{74}\), it has only been recently adopted for RNA targets\(^{75,76}\). A variant of AS-MS, named automated ligand identification system (ALIS), uses indirect detection of a target–ligand interaction by dissociating the formed complex before liquid chromatography–mass spectrometry (LC-MS) analysis to identify the bound ligand\(^{74–76}\) (Fig. 2a). In one example of the use of ALIS, synthetic ligands that bind the flavin mononucleotide (FMN) riboswitch were identified\(^{75}\). One challenge associated
**Methods to identify or design small-molecule RNA binders.**

**Fluorescence-based assays.** Another high-throughput assay relies on displacement of a fluorescent dye or compound by a small molecule of interest. Although originally developed to study the binding of DNA structures, it has also been applied to RNA targets by displacement of the fluorescent dye TO-PRO-1 (REFS. [76,79]) (Fig. 2b), as well as others [79-82]. An extension of this dye displacement method is an assay with turn-on fluorescence that uses a known fluorescent or fluorescently labelled compound by a small molecule of interest. Although fragments can be maintained (discussed below).

**DNA-encoded compound libraries.** DNA-encoded compound library (DEL) technology is a powerful method to explore chemical space that binds a target biomolecule either in solution or on the solid phase. A DEL is synthesized on beads by split and pool in an iterative process in which one of many building blocks is conjugated and its properties can be maintained (discussed below).

**Microarray-based screening.** Small-molecule microarrays (SMMs), created by delivery of minute amounts of compounds to glass slides in a spatial array, were initially used to interrogate protein binding [51-54] and later extended to study the binding of aminoglycosides to the rRNA A-site and how binding is affected by aminoglycoside modification by resistance enzymes [55]. SMMs have now been used to screen a wide variety of compounds and RNA targets [56-101] (Fig. 2c). One advantage of SMMs is that only a small amount of the small molecule is needed to complete a screen and many thousands of interactions can be profiled at once. Compounds are typically covalently attached to the array. Notably, binding to a surface can be quite different from binding in solution. Small molecules can also be non-covalently attached to agarose-coated microarray surfaces by adsorption [104]. Although this method can be broadly applied to many compounds [105], not all compounds adhere to surfaces.

**Fragments.** Fragment-based ligand discovery uses libraries of low-molecular-weight compounds to efficiently explore chemical space that might bind the target of interest (Fig. 2d). Although fragments can be screened for binding, such as by NMR spectroscopy (as demonstrated to identify a fragment that binds severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [105]), these interactions are often difficult to detect as they are low affinity and have short residence times. To overcome these limitations, small-molecule fragments have been functionalized with photoaffinity groups (fully functionalized fragments; FFFs) that enable capture and identification of bound targets, first applied to proteins [106] and later to RNA [107,108]. Low-molecular-weight fragments are ideal for a modular assembly approach in which two fragments are tethered together to bind two structural elements in an RNA target simultaneously, as favourable physicochemical properties can be maintained (discussed below).
be converted into a base pair or a different bulge, the closing base pairs could be altered and so on. Beads that bind the desired target but not the off-target decoy can be analysed and sorted by flow cytometry. Deep sequencing of the beads identifies the binding compound. Because of the potential of nonspecific binding of the encoding DNA tag, nucleic acids and their binding proteins have been avoided. Thus far, the DEL technology has been applied only twice to nucleic acid targets: RNA109 and highly structured DNA G-quartets110.

Designing small-molecule RNA binders

Rather than screening for chemical matter, RNA binders have been designed using selection-based methods that define the preferred targets of RNA-binding small molecules and structure-based design.

Defining binding landscapes. A selection-based platform, 2D combinatorial screening (2DCS), defines the binding landscape of an RNA-binding small molecule. Small molecules displayed on a microarray select their
Resistant profiling is applicable when the small molecule exerts enough selective pressure to induce mutations that confer resistance. These mutations, identified by sequencing, reveal the targets of the small molecule. The target validation method Chem-CLIP (chemical cross-linking and isolation by pull-down) generates a covalent bond between a small-molecule probe and its targets, which are isolated and purified by bead pull-down. Bona fide targets are those enriched in the pulled-down fractions, as compared with the starting cell lysate. Upon co-treating increasing concentrations of the lead compound with Chem-CLIP probe, a dose-dependent restoration of the RNA target in the pulled-down fractions would indicate target engagement of the lead compound.

Phenotypic screening. Phenotypic screening is a strategy to identify compounds that affect pathways associated with a specific phenotype and therefore require no knowledge of the MOA or target. These assays design a screen around a biological process, for example, alternative pre-mRNA splicing, inhibition of translation, derepression of protein targets and bacterial growth. Assays completed in mammalian cells typically use generation of either luciferase or a fluorescent protein as a readout. Two small molecules that modulate splicing were discovered from phenotypic screens, the FDA-approved rilpivirine and branaplam (currently in phase II clinical trials) for the treatment of spinal muscular atrophy (SMA).

Although phenotypic screens have been executed successfully, challenges remain, largely because the approach is target agnostic. Despite the fact that target validation and mechanistic studies can prove difficult, phenotypic screens have enabled lead optimization, particularly by using structure-based drug design. The reader is referred to Ref. 127 for a review on the validation of phenotypic screens of RNA targets.

Target validation and selectivity of RNA-targeting small molecules

Direct target engagement is key to defining the MOA of a compound. Below, we describe various target validation methods based on covalent bond formation between the small molecule and the RNA target, cleavage of the RNA target and resistance profiling. Except for resistance profiling, these methods can study target engagement in vitro, in cells or in vivo.

Fortuitously, some methods not only assess engagement of the desired target but also measure selectivity. Here, we note again the difference between binding selectivity and functional selectivity. Binding selectivity, often measured in vitro by introducing point mutations into a model of the binding site of the small molecule, measures relative affinity and hence the extent of target occupancy. Engagement of most structures in a target RNA is biologically silent with no functional consequence. In contrast, functional selectivity measures whether target engagement has a biological effect — the function of the RNA is modulated as assessed by changes in the downstream pathways of the target and by its associated phenotype.

Resistant profiling

Resistant profiling exerts selective pressure to induce mutations in bacterial or viral genomes to confer resistance (Fig. 5a). The genomes of resistant strains are sequenced to identify in which genes mutations have occurred, thus identifying the target of the compound. Such an approach was used to validate the riboswitch target of ribocil (discussed below), roseoflavin and preferred RNA partners from an RNA library displaying a randomized region in a discrete secondary structure pattern. The selection experiment is completed under stringent conditions in the presence of a large excess of competitor oligonucleotides that mimic regions common to all members of the library, restricting binding interactions to the randomized region. Fully paired DNA and RNA oligonucleotide competitors or RNAs are also often used to increase the stringency of the selection. Selected RNAs are analysed by RNA sequencing (RNA-seq), followed by rigorous statistical analysis of the enrichment of an RNA by the small molecule. Statistical significance scales with affinity such that the more significant the enrichment of the RNA, the more tightly it binds to the small molecule. This analysis affords a binding landscape, or molecular fingerprint, for each small molecule, where a selective small molecule has few RNAs with statistically significant enrichment and a promiscuous one has many. These binding landscapes inform the ideal target for a small molecule and potential off-targets. Cellular RNA targets can be computationally compared with these molecular fingerprints to inform small-molecule design, as in the lead identification strategy Inforna (FIG. 2f). Informa outputs the targetable sites present in a cellular RNA and the rank order of potential small-molecule binders. This approach has been implemented in both a target-centric and a target-agnostic fashion.

Structure-based design and docking

First employed for protein targets, structure-based design and docking have enabled the discovery and the optimization of lead compounds for RNA targets. Using well-defined RNA structures, typically elucidated by NMR spectroscopy, small molecules can be designed to fit within binding pockets. Further, NMR studies enable the prediction of dynamic ensembles of RNA conformations computationally, including short-lived, non-functional species. Small-molecule libraries can then be docked into these ensembles to screen small molecules for selective RNA binding in silico. The significant limitations of these approaches are the quality of the RNA structure, as highly accurate structures are difficult to generate by NMR spectroscopy, as well as the docking programs themselves. Interestingly, an NMR method was recently developed to study conformational biases in HIV-1 TAR and Rev responsive element (RRE) RNAs. A series of mutations were made to each RNA, and the effect on conformational equilibria was measured by NMR spectroscopy. These spectral studies can estimate the percentage of the RNA folded into a non-functional fold, confirmed by studying the cellular activity of the RNA mutants.
pyrithiamine\textsuperscript{11,128,129} (TABLE 1). By comparison with covalent- and cleavage-based target validation methods, mutational resistance profiling does not require synthesis of chemical probes and overall has fewer experimental steps. However, it requires the small molecules to exert sufficient evolutionary pressure to induce mutational resistance, frequently used in cancer biology\textsuperscript{10}.

**Covalent bond formation to measure direct target engagement**

Cellular target validation methods for RNA have been recently developed based on covalent bond formation with\textsuperscript{65,131–133}, or cleavage of, the target\textsuperscript{65}. One covalent method, chemical cross-linking and isolation by pull-down (Chem-CLIP; FIG. 3b), developed in 2013 (REF.\textsuperscript{134}), relies on functional modification of small molecules, attaching a cross-linking module (for example, diazirine or chlorambucil) and a purification handle (for example, biotin or an alkyne)\textsuperscript{65}. The RNA-binding module drives target engagement, bringing the cross-linking module into proximity to the RNA such that they react either directly (electrophilic in the case of chlorambucil) or upon irradiation (diazirine). Small molecule–target complexes are captured and purified with beads (streptavidin- or azide-functionalized). Quantitative PCR with reverse transcription (RT–qPCR) or RNA-seq and subsequent statistical analysis are used to analyse the enrichment of RNA targets in the pulled-down fraction, as compared with the starting lysate. Although the exact protocol may vary depending upon the chemistry of cross-linking\textsuperscript{134} and purification modules, the fundamental idea of Chem-CLIP remains unchanged: transformation of dynamic reversible binding to covalent bonds by cross-linking and amplifying the signal by pull-down enrichment. Chem-CLIP experiments are completed side by side with a probe that lacks the RNA-binding module, controlling for nonspecific reaction of the cross-linking module. Indeed, this approach has validated the RNA targets of small molecules\textsuperscript{65}. This method, however, requires modification of the lead compound, which can be hampered when the synthesis is challenging or if molecular recognition has not been sufficiently defined to inform a site within the small molecule not required for molecular recognition.

Chem-CLIP can also be used to study target binding and functional selectivity, when the pulled-down fractions are analysed by RNA-seq. The enrichment of a transcript indicates the extent of its occupancy by the small molecule, affording the binding selectivity of the small molecule (see ‘Quantifying selectivity below’). Occupancy of a target RNA is not sufficient for a biological response; the small molecule must bind a functional site. Functional selectivity can be defined by comparing target occupancy with the effect on target expression by RNA-seq analysis upon treatment with the lead compound. Many RNA targets will be occupied but their depletion upon analysis by RNA-seq, akin to ASOs. If a small molecule binds to the same region, it impedes hybridization of the ASO and thus diminishes target depletion. This strategy was inspired by studies that used ASOs to probe RNA folding, in which regions of RNA that fold quickly into stable structures are less accessible to ASO hybridization\textsuperscript{135,136}. In contrast to Chem-CLIP, ASO-Clip-Map does not require modification of the lead compound, but does require knowledge of the binding site within the cellular RNA and that the small molecule stabilizes the structure of the RNA sufficiently to impede ASO hybridization. ASO-Clip-Map cannot be used to study selectivity transcriptome-wide, as it would require an ASO to every structure in every target. If an off-target is suspected, ASO-Clip-Map could prove or disprove target engagement.

Methods have also been developed in which the small molecule directly cleaves the RNA target or recruits a nuclease to do so, identifying both on- and off-targets by their depletion upon analysis by RNA-seq, akin to ASOs. To directly cleave RNA targets, a method dubbed small molecule nucleic acid profiling by cleavage applied to RNA (RibosNAP), uses a lead molecule conjugated to the natural product bleomycin. Bleomycin induces strand scission of both DNA and RNA through a metal-ion and oxidative process\textsuperscript{137–142}. Bleomycin A5 is typically conjugated through its terminal amine, which drives affinity for DNA; alkylation of the amine reduces DNA damage by the conjugate and directs its activity for the RNA target\textsuperscript{143}. As with Chem-CLIP, side-by-side experiments are completed with a probe that lacks the RNA-binding module to control for non-selective reaction of bleomycin. The cellular targets of the small molecule–bleomycin conjugate and the control probe are identified by their depletion in RNA-seq data where bona fide targets are those cleaved only by the conjugate. Likewise, the lead compound can be studied directly in a competition experiment. As with Chem-CLIP, RibosNAP is a robust method that can measure cellular binding selectivity by analysing the extent to which each transcript is cleaved. Functional selectivity is measured by coupling these
Table 1 | Examples of small molecules that bind to RNAs implicated in infectious disease

| Compound/institution | Associated disease | RNA target | Mode of action | Stage of development | Activity | Refs. |
|----------------------|--------------------|------------|----------------|----------------------|----------|-------|
| Compound 1/National Cancer Institute | Bacterial infection | Bacterial PreQ₁ riboswitch | Modulates riboswitch activity through transcriptional termination | Early discovery | $K_d = 490 \text{ nM}$ (Bs aptamer domain); $K_d = 99 \text{ nM}$ (Tt aptamer domain) | 100 |
| Ribocil/Merck | Bacterial infection | Bacterial FMN riboswitch | Inhibits translation by binding to the FMN riboswitch | Preclinical | $K_d = 16 \text{ nM}$; EC₅₀ (FMN reporter) = 0.2 μM | 11 |
| Roseoflavin/Yale University | Bacterial infection | Bacterial FMN riboswitch | Inhibits translation by binding to the FMN riboswitch | Early discovery | $K_d = 100 \text{ nM}$ | 128 |
| Pyritiamine/Yale University | Bacterial infection | Bacterial FMN riboswitch | Inhibits translation by binding to the FMN riboswitch | Early discovery | $K_d = 6 \mu M$ (Bacillus subtilis) | 129 |
| 5FDQD/Yale University | Bacterial infection | Bacterial FMN riboswitch | Inhibits translation by binding to the FMN riboswitch | Early discovery | $K_d = 7.5 \text{ nM}$; MIC = 0.5–1 μg ml⁻¹; (antibacterial activity in Clostridium difficile) | 293,254 |
| Amiloride/Duke University | Viral infection (HIV-1) | TAR apical loop | Inhibition of the Tat–TAR interaction | Early discovery | CD₅₀ = 4.4 μM (Tat/TAR-displacement FRET) | 160 |
| Compound 13/University of California, San Diego | Viral infection (HCV) | HCV IRES IIA domain | Inhibits translation by inducing conformational destabilization of the binding of IRES to ribosome | Early discovery | $K_d = 720 \text{ nM}$ (29-mer IIA); EC₅₀ = 5.4 μM (HCV-replicon assay) | 158,180 |
| DMA-155/Duke University, Case Western Reserve University | Viral infection (SARS-CoV-2) | 5’UTR of SARS-CoV-2 RNA | Inhibits translation by perturbing the interaction between viral RNA and host protein | Early discovery | IC₅₀ = 10 μM (virus titre assay) | 181 |
### Table 1 (cont.) | Examples of small molecules that bind to RNAs implicated in infectious disease

| Compound/institution | Associated disease | RNA target | Mode of action | Stage of development | Activity | Refs. |
|----------------------|--------------------|------------|----------------|----------------------|----------|-------|
| DMA-135/Duke University, Case Western Reserve University | Viral infection (EV71) | EV71 SLII IRES domain | Inhibits EV71 replication by stabilizing the translation repressive complex SM–SLII–AUF1 | Early discovery | $K_i = 500 \text{nM}$ (EV71 SLII RNA); $CD_50 = 12 \mu\text{M}$ (EV71 SLII RNA by displacement) | 270 |
| Compound 19/Yale University | Fungal infection | Mitochondrial pre-rRNA group II intron | Inhibition of group II intron splicing leading to mitochondria malfunction and fungal growth inhibition | Early discovery | In vitro splicing inhibition constants ($K_i = 0.36 \mu\text{M}$; MIC $2–4 \mu\text{g ml}^{-1}$) | 154 |
| Compound 15/National Cancer Institute | Viral infection (HIV-1) | TAR hairpin | Inhibition of the Tat–TAR interaction | Early discovery | $K_i = 49 \mu\text{M}$ (to 25–2AP TAR); $IC_{50} = 40 \mu\text{M}$ (Tat/TAR–displacement FRET) | 257 |

**Ribonuclease-targeting chimeras (RIBOTACs)**: Small-molecule chimeras that activate and recruit endogenous ribonucleases to an RNA target to trigger its degradation.

**Gini coefficient**
A statistical measure that quantifies the selectivity of a small molecule, ranging from 0 to 1, where 0 indicates that the compound inhibits every drug target studied equally and 1 indicates a compound selective for one drug target.

---

**Quantifying selectivity**
Quantifying the effect of a compound across the transcriptome, coupled with defining direct targets and pathway analysis (at the transcriptome and/or proteome level) provides insight into small-molecule selectivity. An important metric to quantify small-molecule selectivity such that small molecules can be compared is the Gini coefficient. The targets of a small molecule are rank ordered by their percentage inhibition, and the cumulative fraction and the total cumulative effect for each are calculated. The former is a function of the number of targets studied and the latter is a weighted inhibition. The two values are plotted against each other, and the Gini coefficient is equal to $1 - 2B$ where $B$ is the area under the resulting curve. For a review of Gini coefficients as applied to RNA, the reader is referred to Ref.152.

**Lead optimization of RNA-targeting small molecules**
One advantage of small-molecule therapeutics is that medicinal chemistry efforts can be applied to optimize the initial hit compound, improving both its potency and its physicochemical properties. Below, three broad strategies to lead-optimize RNA-targeting small molecules — including traditional medicinal chemistry as well as structure-guided and modular assembly approaches, which are often integrated synergistically — are discussed.

**Traditional medicinal chemistry approaches**
Traditional medicinal chemistry optimization typically starts with analogue synthesis or purchase to establish the SAR around the hit compound. The bioactivity of an RNA-targeting ligand can be affected by
multiple factors, such as binding affinity, cellular permeability and the number of off-targets. SAR is typically defined by a combination of biophysical and cellular assays to evaluate hit analogues. After acquiring sufficient SAR data, pharmacophore modelling and subsequent chemical similarity searching can then diversify and optimize the starting scaffold. Among all scaffolds, those with desirable or optimizable physicochemical properties are prioritized. Traditional medicinal chemistry can be applied to essentially any small-molecule candidate, tempered by the resources required to synthesize or purchase a large number of analogues. Below, we describe how medicinal chemistry approaches were used to optimize risdiplam, a treatment for SMA, and a pyrimido-indole that directs MAPT alternative splicing. The reader is referred to refs. 154–159 for other notable examples of lead optimization of RNA-targeted small molecules using medicinal chemistry.

**Structure-guided approaches**

Structure-guided approaches for lead optimization rely on sophisticated models of a ligand–RNA complex to define key interactions and identify how positions within the compound's structure can be modified to improve interactions or eliminate those that are unfavourable (Fig. 4b). Such models are typically experimentally generated by NMR spectroscopy or X-ray crystallography. The dynamic nature of RNA must be considered, as no single static structure can truly represent all possible conformations. To address this limitation, experimental models are often coupled with molecular dynamics simulations to generate an ensemble of structures, where the algorithm simulates RNA conformations on the basis of experimentally determined parameters.

Molecular dynamics-based virtual screening has successfully identified ligands that bind various RNAs, typically using models from NMR spectroscopy or X-ray crystallography. As these structures do not capture the dynamic nature of RNA, and molecular dynamics simulations do not accurately recapitulate RNA electrostatics, docking of small-molecule ligands is not as predictive as it could be. To overcome these challenges, a docking method that uses the dynamic ensemble of RNA conformations from combined NMR and molecular dynamics approaches has been developed, an example of which is provided below.26,160 The advantage of structure-based approaches is that they minimize the synthesis efforts by rationally designing analogues,
but the limitation is that not all target structures are readily attainable.

**Modular assembly approaches**

One of the many factors that affect the activity of small molecules that target RNA is the uniqueness of the structural motif throughout the transcriptome. A strategy to optimize small molecules with activity against degenerate functional structures is to exploit adjacent structural elements, that is, a modular assembly approach in which a single small molecule binds to structural elements that are uniquely juxtaposed (FIG. 4c). Indeed, polyvalency — used in various biological processes to enhance avidity and specificity, from cell–cell interactions to gene expression and viral infections — has also been exploited in drug discovery efforts to agonize or antagonize receptors [61, 62]. Such an approach has been taken for RNA targets by modularly assembling two or more RNA-binding modules [63, 64–66], some with in vitro activity [67]. One example is described below in which a modular assembly approach broke the degeneracy of two motifs, affording a specific inhibitor of miR-515 biogenesis. Modular assembly has been employed for RNA repeat expansions (BOX 2), which display periodic arrays of internal loops.

**Examples of small-molecule RNA binders**

Several examples in which small-molecule RNA binders perturb downstream biology in a precise and predicted manner are highlighted below. Notably, the small molecules directly engage a functional site, and simple binding is sufficient to exert a biological effect (BOX 1 and TABLE 1–3).

**Targeting functional RNA structures**

Many RNA classes have functional sites, for example, Drosha and Dicer processing sites within miRNA precursors, IRESs in viral RNAs and some human mRNAs, bacterial riboswitches, splicing enhancers and silencers in pre-mRNAs, and regulatory structures in 5′ and 3′ untranslated regions (UTRs). Below, we describe the discovery of exemplar small molecules that bind to these functional sites.

**Discovery of small molecules that bind viral RNAs.** Apart from the bacterial ribosome, noncoding regions have also been investigated as drug targets to treat viral infections such as those caused by HIV [68, 69] and hepatitis C virus (HCV) [70]. The first viral regulatory elements targeted by small molecules that interfere with the infection process were HIV TAR and RRE RNAs [69, 71]. Structure-guided approaches have been used to identify new scaffolds for TAR RNA that inhibited Tat-mediated transcriptional activation in a cellular model. One such scaffold, amiloride (TABLE 1), was later lead-optimized, affording an analogue with >100-fold enhanced affinity compared with the entry compound, and its structural interactions with the RNA target were well characterized [68, 72].

The HCV mRNA harbours a highly structured and conserved IRES in its 5′UTR, which initiates translation by recruiting and assembling host ribosomes [73–75] and can be inhibited by small-molecule binding [76, 77, 78]. A mass spectrometry-based screen identified 2-aminobenzimidazole derivatives as both binders and inhibitors of HCV translation (TABLE 1). The affinity of the lead scaffold was improved ~100-fold after SAR studies [79, 80].

Another amiloride was discovered that inhibits viral replication of the Betacoronavirus OC43 and SARS-CoV-2 [REF. 81]. A panel of 23 amilorides was screened for inhibition of the infectivity of OC43, affording three lead molecules. Sequence conservation of the 5′-end of betacoronaviruses suggested a functional role, and this region is predicted to fold into six stem–loop structures (SL1–SL6). Fortuitously, one amiloride bound to SL6, as determined by a dye displacement assay and NMR spectral studies [82].

**Discovery of an antifungal that targets a group II intron.**

Group II introns are a class of self-splicing ribozymes found in mitochondria of fungi and yeast but not in mammals [83], making them perhaps ideal targets for the development of antifungals. An in vitro screen of ~10,000 compounds for inhibitors of group II intron splicing afforded six hits that served as starting points [84]. Pharmacophore analysis enabled by SAR studies provided alternative scaffolds with improved physicochemical properties and ultimately a low micromolar
Table 2 | Examples of small molecules that bind to RNAs implicated in cancer

| Compound/institution | Associated disease | RNA target | Mode of action | Stage of development | Activity | Ref. |
|----------------------|--------------------|------------|----------------|----------------------|----------|------|
| PA-1/Université Côte d’Azur | Gastric carcinoma | Pre-miR-372 | Inhibits Dicer processing | Early discovery | $K_d = 150\,\text{nM}$; $IC_{50}$ (Dicer cleavage) = 1.06 $\mu$M, $IC_{50}$ 20 $\mu$M | 193 |
| PA-3/Université Côte d’Azur | Gastric carcinoma | Pre-miR-372 | Inhibits Dicer processing | Early discovery | $K_d = 110\,\text{nM}$; $IC_{50}$ (Dicer cleavage) = 0.58 $\mu$M | 16 |
| Neomycin–nucleobase–amino acid conjugate (compound 6b)/Université Côte d’Azur | Breast cancer | Pre-miR-21 | Inhibits Dicer processing | Early discovery | $K_d = 50.7\,\text{nM}$; $IC_{50}$ (Dicer cleavage) = 1.55 $\mu$M | 193 |
| Targaprimir-515/Scripps Research Institute | Breast cancer | Pre-miR-515 | Dimeric compound that targets Drosha processing site and adjacent site for enhanced selectivity | Early discovery | $K_d = 60\,\text{nM}$; $IC_{50}$ = 0.2 $\mu$M (MCF-7 cells) | 194 |
| Dovitinib–RIBOTAC/Scripps Research Institute | Triple-negative breast cancer and Alport syndrome | Pre-miR-21 | Induces degradation of miR-21 via RNase L recruitment | Early discovery | $K_d = 3\,\text{µM}$ (A bulge 3 $\times$ 2 RNA ILL); reduced levels of miR-21 by ~31% at 0.2 $\mu$M | 150 |
| Rocaglamide/UC Berkeley & RIKEN | Leukaemia | Polypurine sequences in the 5′UTR of a subset of oncogenic mRNAs | Inhibits translation initiation by clamping elf4A to polypurine RNA sequence in the 5′UTR | Early discovery | $K_d = 26\,\text{nM}$ (ternary complex between compound, r(AG)$_{10}$ and elf4A) | 265 |
| Zotatifin (eFT226)/eFFECTOR Therapeutics & Inception Therapeutics | Triple-negative breast cancer and chronic myelogenous leukaemia | Polypurine sequences in the 5′UTR of a subset of oncogenic mRNAs | Inhibits translation initiation by clamping elf4A to polypurine RNA sequence in the 5′UTR | Phase I–II clinical trial: NCT04092673 | $K_d = 21–69\,\text{nM}$ (ternary complex between compound, r(AG)$_{10}$ and elf4A); $IC_{50} = 10.06\,\text{nM}$ (MDA-MB-231 cell proliferation) | 266 |
inhibitor of group II intron splicing with antifungal activity comparable to that of amphotericin B.

**Targeting human miRNAs with small molecules.** miRNAs, small noncoding RNAs of 20–25 nucleotides, are key players in post-transcriptional gene regulation, silencing gene expression by association with the Argonaute RNA-induced silencing complex (RISC) and base pairing to the 3′UTR of complementary mRNAs. Transcribed by RNA polymerase II as primary transcripts (pri-miRNAs), they are processed stepwise, first by the nuclease Drosha into precursor miRNAs (pre-miRNAs); after export to the cytoplasm, the mature (functional) miRNA is generated by the nuclease Dicer. Fortuitously, their structures can be accurately modelled from sequence, and processing (functional) sites can be deduced from deep sequencing of mature miRNAs. Many studies have shown that Drosha or Dicer processing sites within miRNA precursors are targetable functional structures that can be short-circuited to alleviate disease.

Compounds that bind functional structures in miRNA precursors and selectively inhibit the biogenesis of oncosgenic miRNAs have been identified. In particular, a spermine–amidine conjugate (PA-1) was discovered using an assay that measures enzymatic inhibition of pre-miR-372 processing (Fig. 5a and Table 2). In vitro, the molecule, without optimization, bound to the pre-miR-372 Dicer site, inhibited its processing and reduced cancer cell proliferation by derepression of its downstream target large tumour suppressor kinase 2 (LATS2). The selectivity of the small molecule was studied miRnome-wide, which revealed that only a small subset of miRNAs were affected. The conjugate also reduced formation of tumour cell spheroids of patient-derived cancer stem cells. PA-1 was later optimized to afford the pre-miR-372–selective small molecule PA-3 (REF. 19). Interestingly, a neomycin-nucleobase-amino acid conjugate that inhibited pre-miR-372 and pre-miR-21 was discovered by the same laboratory; lead optimization by structure-based design afforded a selective inhibitor of pre-miR-21 (REF. 19). Modular assembly has also been applied to target RNA repeat expansions with enhanced potency and specificity (BOX 2).

**Inhibiting undruggable proteins by targeting the encoding RNA:** α-synuclein. Pathogenic proteins are often considered undruggable when they lack a defined structure. One way to drug these intrinsically disordered proteins (IDPs) is to inhibit their translation by targeting the encoding mRNA. Iron-responsive elements (IREs) are small stem–loop structures present in 5′ or 3′ untranslated regions (UTRs) that bind to iron regulatory proteins (IRPs); IRP binding to IREs in the 5′ UTR prevents ribosome docking and blocks RNA translation whereas binding to the 3′ UTR stabilizes the transcript and upregulates translation. The aberrant expression and mutation of the IDP α-synuclein, which harbours an IRE in the encoding SNCA mRNA, is linked to Parkinson disease.

**Table 2 | Examples of small molecules that bind to RNAs implicated in cancer**

| Compound/institution | Associated disease | RNA target | Mode of action | Stage of development | Activity | Ref. |
|----------------------|-------------------|------------|----------------|----------------------|----------|-----|
| Compound 5/National Cancer Institute | Breast cancer | MALAT1 | Binds and modulates element for nuclear expression (ENE) | Early discovery | $K_{d} = 2.3 \mu M$; IC$_{50}$ about 1 μM RNA level in MMTV-PyMT tumours | 97 |

$EC_{50}$, half maximal effective concentration; elf4A, eukaryotic initiation factor 4A; IC$_{50}$, half maximal inhibitory concentration; K$_{d}$, dissociation constant; miR, microRNA; MMTV-PyMT, mouse mammary tumour virus polyoma middle tumour antigen; RNase, ribonuclease; UTR, untranslated region.
strategy. Synucleozid, a small molecule that binds a bulged adenosine residue within the IRE structure and inhibits SNCA translation, was designed by Informa (Fig. 5b and Table 3). Notably, this bulged adenosine is not present in other IREs. Recognition of the SNCA IRE by synucleozid was dependent upon both the bulged nucleotide and its closing base pairs; its MOA, by direct target engagement, was reduction of the number of

| Compound/institution | Associated disease | RNA target | Mode of action | Stage of development | Activity | Refs. |
|----------------------|-------------------|------------|---------------|---------------------|----------|-------|
| Bleomycin A& | Myotonic dystrophy type 1 | r(CUG)& in DMPK mRNA | Displaces MBNL1 from binding to r(CUG)& | Early discovery | $K_d = 365\,\text{nM}$ (for r(CUG)$_3$); $EC_{25} = 840\,\text{nM}$ (rescued splicing defects) | 278 |
| Cugamycin/Scripps Research Institute | Parkinson disease | Iron-responsive element in α-synuclein mRNA | Inhibits ribosomal assembly and polysome loading onto α-synuclein mRNA | Early discovery | $K_d = 1.5\,\mu\text{M}$; $IC_{50}$ (neurons) = 0.5μM | 13 |
| Mitoxantrone/University of Washington, Seattle | Alzheimer disease; frontotemporal dementia with parkinsonism-17 | Exon 10–intron junction in MAPT pre-mRNA | Inhibits binding of splicing machinery at the exon 10–intron junction; also inhibits a fungal group II intron | Early discovery | $K_d < 1\,\mu\text{M}$; $EC_{50} = 700\,\text{nM}$ (competition of binding with 25nM pyrene–neomycin conjugate) | 225,279 |
| Compound 9/Scripps Research Institute | Alzheimer disease; frontotemporal dementia with parkinsonism-17 | Exon 10–intron junction in MAPT pre-mRNA | Inhibits binding of splicing machinery at the exon 10–intron junction | Early discovery | $K_d = 4.9\,\mu\text{M}$; $IC_{50}$ (transfected HeLa cells) = 10μM; decreased 4R/3R ratio by ~50% in primary neurons at 40μM | 70,228 |
| Risdiplam/Roche and PTC Therapeutics | SMA | SMN2 pre-mRNA exon 7–intron junction | Promotes exon inclusion by stabilizing the binding of the splicing machinery | FDA approved | $K_d = 15\,\text{nM}$; $EC_{1.5x}$ (SMA patient-derived fibroblasts) = 7 nM | 14 |
| Branaplam | SMA; HD | SMN2 pre-mRNA exon 7–intron junction | Promotes exon inclusion by stabilizing the binding of the splicing machinery | Phase I/II clinical trial (for SMA: NCT02268552; for HD: NCT05111249) | $EC_{50}$ (SMN ELISA in SMN& mouse myoblasts) = 20 nM | 15 |

$EC_{50}$, half maximal effective concentration; ELISA, enzyme-linked immunosorbent assay; HD, Huntington disease; $IC_{50}$, half maximal inhibitory concentration; $K_d$, dissociation constant; SMA, spinal muscular atrophy; SMN, survival of motor neuron.
polysomes loaded onto SNCA mRNA. Approximately 90 mRNAs have been identified with IREs or IRE-like elements, which vary in their sequences and structures; these studies and those mentioned above lay the foundation for regulating translation of other mRNAs with small molecules.

**a** Small molecule that inhibits miRNA processing

![Small molecule that inhibits miRNA processing](image)

MOA: inhibits Dicer processing of oncogenic pre-miR-372

Identified from fluorescent quenching assay; exhibits anti-tumour activity $K_d = 150$ nM; $IC_{50}$ (Dicer cleavage) = 1.06 $\mu$M

**b** Small molecule that binds SNCA IRE and inhibits translation

![Small molecule that binds SNCA IRE and inhibits translation](image)

MOA: inhibits ribosomal assembly and polypeptide loading onto $\alpha$-synuclein mRNA

Thermodynamic stabilization of the IRE near start codon inhibits translation $K_d = 1.5$ $\mu$M; $IC_{50}$ (neurons) = 0.5 $\mu$M

**c** Splicing modulator that directs MAPT pre-mRNA exon exclusion

![Splicing modulator that directs MAPT pre-mRNA exon exclusion](image)

MOA: inhibits binding of splicing machinery at the exon 10–intron junction in MAPT pre-mRNA to reduce toxic 4R tau

Lowers 4R tau in HTau mouse primary neurons, structurally enabled $K_d = 4.9$ $\mu$M; $IC_{50}$ (transfected HeLa cells) = 10 $\mu$M

**d** Small molecule that binds FMN riboswitch and inhibits translation

![Small molecule that binds FMN riboswitch and inhibits translation](image)

MOA: inhibits translation by binding to the FMN riboswitch and inducing formation of the sequester loop that hides start codon

Selective FMN mimic, inhibits bacterial growth $K_d = 16$ nM; $EC_{50}$ (FMN reporter) = 0.2 $\mu$M

**e** Molecular glue that directs SMN2 pre-mRNA exon inclusion

![Molecular glue that directs SMN2 pre-mRNA exon inclusion](image)

MOA: promotes exon inclusion by stabilizing the binding of the splicing machinery at the exon 7–intron junction

Selective molecular glue $K_d$ of ternary complex = 15 nM; $EC_{50}$ (SMA patient-derived fibroblasts) = 7 nM

**f** Molecular glue that inhibits translation of oncogenic mRNA

![Molecular glue that inhibits translation of oncogenic mRNA](image)

MOA: inhibits translation initiation by clamping eIF4A to polyuridine RNA sequences in the 5′ UTR

Selective molecular glue between protein and RNA $K_d = 26$ nM (eIF4A to (AG)$_{10}$ in the presence of RocA)

Fig. 5 | Small-molecule binding to RNA structures elicits biological effects through various modes of action. a | A small molecule that binds the Dicer site of oncogenic pre-miR-372 inhibits its biogenesis and short-circuits downstream pathways. b | A compound that binds to an iron-responsive element (IRE) in the 5′ untranslated region (UTR) of SNCA mRNA inhibits its translation by mechanically blocking ribosomal assembly and polypeptide loading onto the mRNA. c | A small molecule that binds to a structural element at the exon 10–intron junction of MAPT pre-mRNA directs alternative splicing towards exon exclusion, reducing the amount of toxic 4R tau produced. d | A small molecule inhibits translation by binding to the flavin mononucleotide (FMN) riboswitch and inducing formation of the sequester loop that hides the start codon. e | Another small molecule, later optimized into an FDA-approved drug for the treatment of spinal muscular atrophy, binds to the exon 7–intron junction in SMN2 pre-mRNA and increases SMN2 protein levels by acting as a molecular glue for the RNA and splicing machinery and promoting exon inclusion. f | A compound binds to polyuridine sequences in 5′UTRs of a subset of mRNAs and inhibits translation by acting as a molecular glue for the RNA and eukaryotic initiation factor 4A (eIF4A). $IC_{50}$, half maximal inhibitory concentration; $K_d$, dissociation constant; miR, microRNA; MOA, mode of action.

Alterating protein isoforms by directing exon exclusion with small molecules: microtubule-associated protein tau. Almost every pre-mRNA goes through a series of processing steps, including alternative splicing (inclusion or exclusion of exons) to produce different protein isoforms. Unsurprisingly, splice site mutations can
alter splicing patterns and cause human diseases\(^ {218}\). Studies with oligonucleotides demonstrated that it is indeed possible to rescue splicing defects by binding to and occluding the mutation from the splicingosome machinery\(^ {207,216-222}\), suggesting that small molecules may also be able to direct splicing.

The microtubule-associated protein tau gene (MAPT) produces six tau isoforms by alternative splicing\(^ {223}\). Exons 9–12 encode a microtubule-binding domain (MBD), and exon 10 is alternatively spliced to produce isoforms with three (3R) or four (4R) MBDs, with the latter prone to aggregation\(^ {223}\). In healthy individuals, the ratio between 3R and 4R is approximately equal; however, a genetic mutation in the splicing regulatory element (SRE) present in exon 10 alters the splicing pattern to produce excess 4R, the cause of frontotemporal dementia with parkinsonism-17 (FTDP-17). One such intrinsic mutation, dubbed disinhibition dementia parkinsonism amyotrophy complex (DPAC), converts a GC base pair into a GU wobble pair, destabilizing the structure of the SRE. This destabilization allows U1 small nuclear RNA (snRNA) to more easily bind the SRE and facilitate exon 10 inclusion\(^ {225}\). Stabilization of the structure of the SRE with a small molecule might therefore impede U1 snRNA binding and restore a normal splicing pattern. Various small molecules have been identified that bind to the SRE\(^ {223,226}\) (Figure 5c and Table 3).

Small molecules with activity in primary neurons were lead-optimized by chemical similarity searching, pharmacophore modelling driven by in vitro and cellular screening and analogue synthesis, structure-based design and docking studies based on the 3D structure of the MAPT SRE bound to several lead molecules, and traditional medicinal chemistry approaches\(^ {227}\) (Figure 4b). This synergistic strategy afforded a drug-like molecule that directed MAPT splicing towards the 3R isoform in primary neurons from a human tau transgenic mouse\(^ {228}\).

**Targeting RNA structures for degradation**

If a functional site has not yet been discovered, various strategies may be applied to modulate RNA function by facilitating its degradation (Figure 6). Two of these strategies are discussed below; both rely on chemically induced proximity — direct small-molecule cleavage of the target RNA (Box 2) or target degradation achieved by nuclease recruitment (Figure 6a). Although these examples target known functional structures, the cleavage-based methods eliminate the RNA transcript and thus their MOA does not require binding to a functional site. The advantage of small molecule degraders, like ASOs, is that they are simultaneously target-validation methods. Small molecules can also facilitate degradation of an RNA target by increasing its accessibility to endogenous decay pathways such as the exosome and directing splicing such that the mature mRNA encodes a premature stop codon (Figure 6b).

**Nuclease recruitment to cleave RNA targets: ribonuclease-targeting chimeras.** Targeted degradation was first demonstrated for proteins, or proteolysis-targeting chimeras (PROTACs)\(^ {229}\). PROTACs are chimeric molecules comprising a protein-binding module and an E3 ubiquitin ligase-recognition module, which tags the targeted protein for selective degradation by the proteasome\(^ {230-234}\). PROTACs have been developed for targeted degradation of RNA\(^ {144}\), composed of an RNA-binding module and a RNase-recruiting module that selectively mediates RNA decay (Figure 6a). PROTACs recruit the ubiquitously expressed cellular endoribonuclease RNase L, which functions in the viral immune response (see above)\(^ {146,149}\). In its first two iterations, an RNA-binding module that selectively recognizes the Drosha site of pri-miR-96 or the Dicer site of pre-miR-210 was coupled to 2′-5′A\(_{4}\), inducing its selective cleavage in cells\(^ {144,235}\). As 2′-5′A\(_{4}\) reduces drug-likeness, a small-molecule RNase L recruiter was developed\(^ {143}\) based on a previously reported small molecule\(^ {236}\). A RIBOTAC with this new recruiter was developed to target pre-miR-21 for degradation by recognition of its Dicer processing site\(^ {144}\).

Notably, the RIBOTAC was more potent than the binder from which it was derived, possessed a prolonged duration of effect and inhibited breast cancer metastasis in a mouse model. Transcriptome-wide studies showed that the RIBOTAC was also more selective than the binder\(^ {144}\), as quantified by a Gini coefficient\(^ {149}\) and did not elicit an immune response. This enhanced selectivity is a composite of the specificity of the RNA-binding small molecule, the inherent substrate specificity of RNase L and whether the target has an RNase L substrate adjacent to the site at which the RNA-binding small molecule binds, the distance dictated by the linker that tethers the two components of the chimera.

These studies led to the hypothesis that RIBOTACs may allow reprogramming of known drugs for RNA targets. DCS selection of the Repurposing, Focused Rescue, and Accelerated Medchem (ReFRAME) library\(^ {237}\) indicated that the receptor tyrosine kinase (RTK) inhibitor dovitinib binds the Dicer processing site of pre-miR-21 and therefore might inhibit its cellular processing, albeit at higher concentrations than those required to inhibit RTKs\(^ {150}\). Converting the binding molecule dovitinib into a RIBOTAC enhanced its inherent RNA-targeting activity in cells and concomitantly decreased potency against canonical RTK protein targets, shifting selectivity for pre-miR-21 by 2,500-fold (Table 2). Further, the chimera alleviated disease progression in two mouse models caused by miR-21 overexpression: triple-negative breast cancer and Alport syndrome\(^ {150}\).

RNA function can indeed be modulated with small-molecule binders or small-molecule degraders, expanding the MOA of RNA-targeting compounds and likely the number of targetable RNAs. Chimeric small molecules that target RNAs set the foundation for new drug discoveries akin to the revolution of PROTACs, enabling inhibition of RNA circuits when the functional site is unknown or absent. It will be exciting to see whether other RNA-modifying enzymes (editing, splicing machinery and so on) can be selectively recruited to RNA targets.

**Targeting RNA-associated pathways**

Several small molecules that target RNA-associated pathways have been derived from phenotypic screening and are described below.
Ribocil, an antibacterial that targets the FMN riboswitch. Riboswitches are structured noncoding sequences in the 5’ leader of bacterial mRNAs that control gene expression of a downstream ORF\textsuperscript{11,238-242} (FIG. 5d). Widely distributed across all known phylogenetic groups of bacteria, riboswitches form highly specific binding pockets for small-molecule metabolites, second messengers and inorganic ions\textsuperscript{243-247}. Binding of the small molecule to the receptor (aptamer) domain directs formation of alternative secondary structures in the adjacent regulatory domain (expression platform)\textsuperscript{248-250} that modulates transcription or translation of the message\textsuperscript{248-250}. The antibacterial ribocil was identified by a phenotypic screen for inhibitors of the riboflavin biosynthetic pathway in \textit{E. coli}\textsuperscript{11} (FIG. 5d and TABLE 1). Resistance mutations mapped to the FMN riboswitch immediately upstream of the \textit{ribB} gene\textsuperscript{11,251}, validating the RNA target and elucidating compound MOA.

\textbf{a} Degraders elicit biological functions even if bound to non-functional sites

Degraders cleave RNA targets sub-stoichiometrically

Degraders (RNase recruiters) can be optimized to improve target selectivity

\textbf{b} Nuclear RNA exosome-mediated degradation

Direct oxidative cleavage

Induced proximity and activation of a ribonuclease

Modulating pre-mRNA splicing to include premature termination codon

\textbf{Fig. 6 | Small-molecule RNA degraders and their mechanisms of action.} \textbf{a} | Degraders can elicit biological functions even if bound to non-functional sites within the RNA target as they cleave the transcript. Simply binding to non-functional sites is in principle biologically silent. Notably, degraders cleave RNA targets sub-stoichiometrically, as the same degrader molecule can cleave more than one RNA transcript by substrate turnover and can be optimized to improve target selectivity, including linker length, substrate preference and cellular localization of the target. \textbf{b} | Mechanisms of small-molecule-facilitated degradation. From left to right: small molecules can bind to intronic RNA repeat expansions that are harboured as retained introns. This causes excision of the intron, which is decayed by the RNA exosome\textsuperscript{280}; RNA-binding compounds can be appended to natural products such as bleomycin that can cause oxidative cleavage of RNA targets selectively; ribonuclease-targeting chimeras induce the proximity of ribonucleases to unnaturally target an RNA for destruction by native quality control pathways; and small molecules can affect pre-mRNA splicing to create a mature mRNA with included exons that contain premature termination codons that trigger decay via nonsense-mediated decay. RNase, ribonuclease.
A crystal structure of the riboswitch–ribocil complex revealed that the compound competitively binds to the FMN-binding pocket, using a similar, but not identical, binding mode.

A structure-guided approach to target the FMN riboswitch was also fruitful, yielding the compound 5FDQD (TABLe 1). Inspection of the crystal structures of the FMN-bound252 and apo-riboswitch253 revealed conformational changes in the RNA that occur upon FMN recognition. An iterative structure-based design strategy was pursued in which, first, structures were analysed for regions that could potentially accommodate chemical changes to FMN; second, a set of structure-guided derivatives was synthesized; third, productive binding was tested using chemical probing and in vitro transcription assays; and fourth, crystal structures of new lead compounds that emerged were determined254,255. The compound that resulted from these efforts, 5FDQD, is an analogue of FMN that binds to its RNA target with activity equivalent to that of the natural effector254,255. The bactericidal activity of this compound is highly selective for *Clostridium difficile* while having little effect upon diverse other bacteria commonly found in the gut microbiota24. Importantly, in mice, 5FDQD prevented lethal antibiotic-induced *C. difficile* infection, validating the use of a structure-guided approach to yield potent RNA-targeting therapeutics.

**Risdiplam and branaplam, small molecules that target an RNA–protein manifold to direct alternative pre-mRNA splicing.** SMA is a genetic disease caused by the deletion or mutation of the survival motor neuron 1 (*SMN1*) gene, which encodes a catalytic component of a complex responsible for assembly of small nuclear ribonucleoproteins (snRNPs) and hence the spliceosome. Its loss of function in SMA ultimately leads to the degradation of spinal motor neurons, muscle weakness, muscle atrophy and respiratory complications66,256. Fortuitously, humans encode an *SMN1* parologue, *SMN2*, with the two genes differing by only two nucleotides, one in exon 7 and the other in exon 8. The single-nucleotide polymorphism in exon 7 disrupts a splicing enhancer, resulting in exclusion of *SMN2* exon 7 and reduction of the half-life of the encoded protein. If *SMN2* exon 7 alternative splicing could be directed towards inclusion, then *SMN2* could substitute functionally for loss of *SMN1* and hence as a treatment for SMA (FIG. 5e).

A phenotypic screen to discover small molecules that direct splicing of *SMN2* such that exon 7 is included identified an orally bioavailable compound, SMN-C3. Subsequent studies showed that SMN-C3 directed endogenous alternative splicing of *SMN2* and provided therapeutic benefits in an SMA mouse model, with limited off-target effects. A medicinal chemistry campaign around SMN-C3 generated risdiplam (Evrysdi; FIG. 5e and TABLe 3), the first small-molecule FDA-approved drug to treat SMA14.

The lead optimization process that afforded risdiplam highlights a traditional medicinal chemistry approach (FIG. 5e). After chemical optimization of SMN-C3 to avoid mutagenicity and to confer favourable pharmacokinetic profiles, the preclinical candidate RG7800 was selected for advancement262,263. As RG7800 caused retinal degeneration4, a medicinal chemistry campaign began around key elements of the chemical structure, particularly with the goal to lower basicity, enhance potency, improve influx into the central nervous system and reduce compound metabolism. A virtual study of a library containing a pyridoprimidinone central core and a right-hand-side imidazopyridazine fragment was carried out14. Risdiplam emerged from these compounds after a battery of preclinical tests showed high potency for directing SMN2 splicing in vitro and in vivo, reduced basicity, no phototoxicity risk and no formation of active metabolites14.

The challenge was to identify the target of risdiplam and hence its MOA, later determined after a series of studies on a derivative dubbed SMN-C5. NMR spectroscopy revealed that SMN-C5 stabilized an adenosine bulge at the exon 7–intron junction15,256,259, acting as a molecular glue for the ternary complex formed with U1 snRNP, revealed by Chem-CLIP260.

The small molecule branaplam was also identified from a similar phenotypic screen of *SMN2* exon 7 alternative splicing, using a mini-gene reporter of a breast cancer 1 (BRCA1) exon 18 mutant that induces exon skipping as a counter-screen261. A series of studies using a related molecule, NVS-SM2, pinpointed that the molecule interacted with 21 nucleotides of the *SMN2* 5′ splice site, in particular a GA sequence found at the end of exon 7, suggesting involvement of U1 snRNP. NVS-SM2 acts as a molecular glue, as U1 snRNP only bound to SMN2 exon 7 when the small molecule was present262. Branaplam was later discovered to reduce mutant huntingtin (mHTT) protein levels by facilitating pseudo-exon inclusion in the *HTT* mRNA263. It is currently in clinical trial for the treatment of both SMA and Huntington disease.

The discoveries of risdiplam and branaplam are perhaps the best examples of successful phenotypic drug discovery efforts, as their development did not rely on a specific target but instead on a desired activity.

**Rocaglamide, a molecular glue that inhibits translation of polypurine-containing transcripts.** Molecular glues that affect the translation of specific mRNAs have also been identified by phenotypic screening, particularly for oncogenic mRNAs that promote proliferation264. Notable amongst several interesting compounds is rocaglamide, a member of the flavagline family, a class of bioactive natural products264 (FIG. 5f and TABLe 2). Rocaglamide has potent anti-tumour activity, specifically inhibiting the translation of a subset of transcripts with polypurine sequences in their 5′UTRs. The compound stabilizes the interaction between eukaryotic initiation factor 4A (eIF4A), an ATP-dependent DEAD-box RNA helicase, and mRNAs that depend on eIF4A binding to unwind them for translation265. Recent studies have afforded rocaglamide analogues with improved potency and physicochemical properties266,267. Thus, translation can be affected not only by small-molecule binding to an mRNA, but also by pharmacological stabilization of transient interactions between transcripts.
and translation factors, providing specific functional outcomes. Other compounds have been shown to have similar activities, suggesting that medicinal optimization of these compounds could provide small-molecule therapeutics264.

The future of small-molecule targeting of RNA
First viewed as simply an intermediate between DNA and protein, a renaissance in RNA structure and function beginning in the 1980s revealed the complex roles of RNA in homeostasis and disease183,269–273. Although considered a challenging target, modulation of RNA function by structure-binding small molecules is becoming increasingly well established as therapeutically practical. However, tremendous gaps in knowledge remain (TABLE 4 and BOX 3).

In the protein world, there are target classes that are considered both druggable and undruggable, generally classified by the formation of ordered structures. There is currently not enough knowledge in the RNA world to be able to classify RNA targets; however, most, if not all, bioactive small molecules target stable, functional structures. By identifying regions with unusually stable structures that are evolutionarily conserved, we can gain insight into potential functional structures and expand the druggable transcriptome. Key to the discovery of small molecules that bind these sites is to hypothesize function and hence potential compound MOA. Once the hypothesis is formulated, it must be verified by gain- and loss-of-function studies, particularly the effect on phenotype. Fortuitously, if binding alone is insufficient to modulate RNA function, the binding small molecule can be

| Development step | Tools | Challenges |
|------------------|-------|------------|
| **Identify druggable RNA targets** | Identify regions in an RNA target that are stably folded by using evolution and prediction restrained by structure mapping experiments | Validating a predicted RNA structure in cells and identifying factors to prioritize targets that are more easily drugged |
| | Hypothesize a MOA for a binding compound | |
| | Convert a binder into a degrader if binding alone is not functional | |
| **Identify small molecule binders** | Complete sequence-based design and identify on- and off-targets and routes for optimization | Defining chemotypes in compounds that bind RNA selectively |
| | Screen compound collections to find binders, considering screening on- and off-targets | Performing quantitative structure–activity relationships for RNA targets |
| **Evaluate bioactivity of small molecules in cells** | Study the effect of RNA target-associated phenotypes and validate by using gain- or loss-of-function | Correlating binding to function is not straightforward as not all binders will be biologically functional |
| | Use Chem-CLIP to define cellular on- and off-targets and evaluate the selectivity on the transcriptome and the proteome; correlate Chem-CLIP with -omics studies to define optimization plan | |
| **Structure-based lead design and optimization** | Solve structure of complex by NMR spectroscopy and X-ray crystallography, etc. | Structure-based drug design and docking for RNA is different than for protein, i.e., electrostatic interactions are a significant driver |
| | Study pharmacophore model to define features in binding compounds and expand chemical space for target by using chemical similarity searching of pharmacophore and structure-based drug design | Force fields need development, as there are limited examples of driving potency and specificity using these approaches |
| **Evaluate efficacy and PK in animal models** | Study disease defect in model organisms, including transgenic models | Physicochemical properties that potently and specifically affect RNA are unknown; need to be open minded and empirical with compounds, especially those outside rule of five space |
| | In vitro and in vivo PK assays are well established | Studies in model organisms can be challenging especially for noncoding RNAs owing to differences in the human sequence and structure |
| | | Varied modes of action can complicate PK and pharmacodynamic profiles |
| **Preclinical and clinical studies** | Study compound activity in patient tissues, available through many repositories | Should the pipeline of tests used for oligonucleotide-based medicines be altered? Can this process be streamlined by using RNA-seq and other tools? Differences in sequence and structure in human versus mouse or other species should be considered in efficacy and toxicity studies |
| | RNA-seq can be incorporated into clinical development pipeline to study efficacy, toxicity, potential off-targets | |

MOA, mode of action; PK, pharmacokinetic(s); RNA-seq, RNA sequencing.
converted into a degrader. As more data are collected around functional structures, we will be able to tease out factors that contribute to druggability.

Complementary to identifying high-priority RNA targets is defining physicochemical properties and chemotypes that confer affinity for RNA, which will inform design of RNA-focused small-molecule libraries (Box 1). Such libraries can be used in two ways: first, sequence-based design by defining binding landscapes, which defines both on- and off-targets; and second, screening against defined targets, in which an appropriate counter-screen must be completed to improve the likelihood of selective binding. The resultant datasets refine hypotheses around physicochemical properties and chemotypes that are ideal for RNA targets and hence RNA-focused small-molecule libraries. Notably, these screening collections must maintain chemical diversity to be useful.

Advances in structure-based design and lead optimization for RNA-targeted small molecules are also needed. Force fields that have been developed to enable structure-based design for proteins have not yet been fully customized for RNA, despite recent advances in computational biology. Molecular recognition of RNA by small molecules is driven differently than for protein, specifically the importance of aromatic ring stacking interactions that will require modification of electrostatic parameters. Accurate modelling of electrostatic parameters is key to understanding how small-molecule cores bind to and interact with an RNA target and to guide positioning of functionalities outside the core. Further, small-scale conformational dynamics are important for recognition of RNA targets. In its native state, RNA can have fast exchange between conformations, which are recognized by ligands differently. When the energetic difference between these states is low, there is likely only a modest effect on affinity. When the energy difference between conformations is large, however, the difference in affinity could be dramatic. Such knowledge is key for structure-based design.

Small-molecule binding is often not sufficient for bioactivity. Determining whether binding occurred to the cellular target but produced a biologically silent interaction is now possible with recently developed tools. Correlating target occupancy studies, using methods such as Chem-CLIP, with transcriptome- and proteome-wide studies is a powerful strategy that defines on- and off-targets and when binding elicits a biological response that is not desired. Such data are key to inform lead optimization. Intriguingly, fragment mapping and Chem-CLIP, in combination with RNA structure prediction programs such as ScanFold, can provide insight into ligandable cellular RNA structures; biologically silent interactions inform cellular RNAs amenable to targeting by small-molecule-induced degradation.

Translating bioactive small molecules from cells to animals and then to the clinic is a significant hurdle for RNA-targeted small molecules. Of importance will be incorporation of transcriptome-wide analyses in patient tissues that assess efficacy and toxicity into clinical development pipelines, which should also be implemented for oligonucleotide-based modalities. Such studies also inform the range of selectivity that will be acceptable for RNA targets and may be quite different from that of proteins. For in vivo studies, significant differences in the sequence and structure of human and mouse RNAs, particularly noncoding RNAs, can change the activity or selectivity of a compound. As very few RNA-targeted compounds have advanced to animal studies and the clinic, more data will be needed to define pharmacokinetic and pharmacodynamic profiles ideal for preclinical and clinical candidates. As PROTACs and other protein-targeted medicines have changed the view of the rule of five, we will need to be open to the fact that the physicochemical properties of small molecules that modulate RNA function may be outside the traditional drug-like space.

**Box 3 | Accelerating development of small molecules targeting RNA**

Key to advancing the RNA-targeting and chemical biology fields is an established pipeline for target validation and MOA studies, a longstanding process for proteins. Compounds with ill-defined mechanism are of less utility in chemical probe and lead medicine development than those that have an established MOA. A pathway for target validation and MOA studies has been previously presented. The first step in this process is to confirm that the RNA target of interest is indeed driving phenotype, if not previously validated from in vitro or in vivo studies. For miRNAs, forced expression of the miRNA precursor in cells should induce or enhance the observed phenotype, while treatment with an antagonist should alleviate it. Validation for other RNA classes can be completed analogously by forced expression or knockout with ASOs. Small molecules that bind to the target can then be designed or discovered as described above.

Compounds that bind selectively to an RNA target in vitro can be carried forward to cellular studies to investigate whether they have the desired effect on the RNA and its downstream pathways, including phenotype. The effect of the small molecule on the abundance of the target RNA will depend on the target and the MOA of the compound. For example, small molecules that inhibit miRNA biogenesis by simply binding should boost levels of the precursor that it binds, leading to reduction of mature miRNA abundance of the target rNA will depend on the target and the MOA of the compound. For example, small molecules that inhibit miRNA biogenesis by simply binding should boost levels of the precursor that it binds, leading to reduction of mature miRNA abundance.

Small molecules that degrade the RNA target, whether directly or by recruiting nuclease, should reduce abundance. Further investigation is therefore required when an unexpected effect on RNA abundance is observed to define the compound’s MOA and rule out other effects such as inhibition of transcription. Notably, small molecule binding can alert RNA surveillance and quality control pathways that could induce degradation, and thus an unexpected effect on abundance does not necessarily eliminate an RNA-centric MOA. Instead, cellular target engagement should be confirmed using the methods described above, as should be the case for all RNA-targeting endeavours.

To fully define the effect of an RNA-targeting small molecule on phenotype, gain-of-function studies should be completed. For the former, forced expression of the RNA target is expected to reduce compound efficacy, as less of the target can be occupied at the same dose. Loss-of-function studies is a multistep process that is MOA-dependent. The first step, regardless of RNA target or MOA, is to express the wild-type RNA and an RNA with point mutations that ablate the binding site of the small molecule in cells with low or no endogenous expression. Here, the small molecule should only alleviate molecular defects and phenotype in wild-type RNA-expressing cells. The second step, applied to all RNA targets, is to confirm that the small molecule has no effect on cells that do not (aberrantly) express the target, typically cells derived from healthy patients. The remaining steps in loss-of-function studies are tailored to compound MOA. For example, bioactive small molecules that modulate miRNA function derepress a downstream protein target that alleviates phenotype. The downstream target can be knocked out chemically or with a small interfering RNA. The small molecule is no longer able to boost abundance of the downstream target and hence alleviate phenotype.
druggable by small molecules. Further, RNA-targeted small molecules can be designed and lead optimized, the latter using strategies developed for protein targets with important modifications and considerations. As both academia and industry push the frontiers of this field forward, we believe that many more RNA-targeted medicines will reach the clinic in the years to come.

Published online 8 August 2022
longevity in mice with spinal muscular atrophy. Science 345, 688–693 (2014).

67. Parsons, J. et al. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. Nat. Biotechnol. 28, 825–830 (2010).

68. Carnevali, M., Parsons, J., Wyles, D. L. & Hermann, T. A modular approach to synthetic RNA binders of the hepatitis C virus internal ribosome entry site. ChemBioChem 11, 1364–1367 (2010).

69. Chen, C. Z. et al. Two high-throughput screening assays for abasic site repair by mismatch repair in myotonic dystrophy type 1. Anal. Bioanal. Chem. 402, 1889–1898 (2012).

70. Chen, J. L. et al. Design, optimization, and study of small molecule inhibitors that target pre-mRNA and affect splicing. J. Am. Chem. Soc. 142, 8706–8727 (2020).

71. Corbett, P. T. et al. Dynamic combinatorial chemistry. Chem. Soc. Rev. 39, 1271–1286 (2010).

72. Motoyoi, T. Revolution of small molecule drug discovery by affinity selection–mass spectrometry technology. Chem. Pharm. Bull. 60, 191–195 (2012).

73. Kiernan, U. A., Nedekov, D., Niedekker, E. O., Tubbs, K. A. & Nelson, R. W. High-throughput affinity mass spectrometry. Methods Mol. Biol. 328, 141–150 (2006).

74. Anns, D. A. et al. Affinity selection–mass spectrometry methodology for the identification of small molecule ligands from DNA-encoded combinatorial libraries. Int. J. Mass Spectrom. 238, 73–83 (2004).

75. Rizvi, N. F. et al. Discovery of selective RNA-binding small molecules using a affinity selection–mass spectrometry. ACS Chem. Biol. 13, 820–831 (2018).

76. Rizvi, N. F. & Nickberg, E. B. RNA-ALIS: methodology for small molecule target selection using ALIS affinity-selection mass spectrometry. Methods 167, 28–38 (2019).

77. Stover, J. S., Shi, J., Jin, W., Vogt, P. K. & Boger, D. L. Discovery of inhibitors of aberrant gene transcription from libraries of DNA binding molecules: inhibition of LIF-1 mediated gene transcription and oncogenic transformation. J. Am. Chem. Soc. 131, 3542–3548 (2009).

78. Tran, T. & Disney, M. D. Identifying the preferred RNA molecules that interact with target RNAs by probing millions of combinations. Nat. Commun. 3, 1125 (2012).

79. Asare-Oka, P. N. & Chow, C. S. A modified fluorescent intercalator displacement assay for RNA ligand discovery. Anal. Biochem. 408, 269–276 (2011). Demonstrates the successful application of ALIS to RNA targets.

80. Wang, Z. F. et al. The hairpin form of r(G,C)m in c9SALS/FTD is repeat-associated non-ATG translated containing transcripts. Proc. Natl Acad. Sci. USA 98, 496–505 (2001).

81. Sham, J. M., Haller, A., Rieder, R. & Micura, R. A powerful approach for the selection of 2-amino purine ligands for RNA targets. J. Am. Chem. Soc. 133, 16611–16617 (2011).

82. Soulière, M. F. & Micura, R. Use of SHAPE to select 2AP-substituted sites to investigate RNA folding. Methods Mol. Biol. 1103, 227–259 (2015).

83. Rizvi, N. F. & Nickberg, E. B. RNA-ALIS: methodology for small molecule target selection using ALIS affinity-selection mass spectrometry. Methods 167, 28–38 (2019).

84. Disney, M. D. & Seeberger, P. H. Aminoglycoside microarrays to explore interactions of antibiotics with RNAs and proteins. Chemistry 10, 3508–3514 (2004). Demonstrates the successful application of ALIS to RNA targets.

85. Disney, M. D. & Barrett, O. J. An aminoglycoside microarray platform for directly monitoring and studying antibiotic resistance. Biochemistry 46, 11225–11230 (2007).

86. Abravira, F. A. et al. Selective small-molecule targeting of a triple helix by the homologous RNA. PLoS ONE 7, e37313 (2012). Demonstrates the successful application of ALIS to RNA targets.

87. Wang, Z. F. et al. The hairpin form of r(G,C)m in c9SALS/FTD is repeat-associated non-ATG translated containing transcripts. Proc. Natl Acad. Sci. USA 98, 496–505 (2001).

88. Soulière, M. F. & Micura, R. A powerful approach for the selection of 2-amino purine ligands for RNA targets. J. Am. Chem. Soc. 133, 16611–16617 (2011).

89. Soulière, M. F. & Micura, R. Use of SHAPE to select 2AP-substituted sites to investigate RNA folding. Methods Mol. Biol. 1103, 227–259 (2015).

90. Sham, J. M., Haller, A., Rieder, R. & Micura, R. A powerful approach for the selection of 2-amino purine ligands for RNA targets. J. Am. Chem. Soc. 133, 16611–16617 (2011).

91. MacBeath, G., Koehler, A. N. & Schierle, S. L. Printing small molecules as microarrays and detecting protein–ligand interactions en masse. J. Am. Chem. Soc. 121, 7967–7968 (1999).

92. Fazou, F., Breyer, S. O., Poulsen, J. C. & Wong, C. H. Synthesis of sugar arrays in microtiter plate. J. Am. Chem. Soc. 124, 14597–14402 (2002).

93. Peng, B., Thomas, J. L., Shih, Y. T., Schulze, H. & Yao, S. G. Small molecule microarray based discovery of PARP14 inhibitors. Angew. Chem. Int. Ed. Engl. 56, 425–430 (2017).

94. Wang, Z. et al. Microarray based screening of peptide nano probes for HER2 positive tumor. Anal. Chem. 87, 8637–8672 (2015).

95. Disney, M. D. & Berger, P. H. Aminoglycoside microarrays to explore interactions of antibiotics with RNAs and proteins. Chemistry 10, 3508–3514 (2004). Demonstrates the successful application of ALIS to RNA targets.

96. Disney, M. D. & Barrett, O. J. An aminoglycoside microarray platform for directly monitoring and studying antibiotic resistance. Biochemistry 46, 11225–11230 (2007).

97. Abravira, F. A. et al. Selective small-molecule targeting of a triple helix by the homologous RNA. PLoS ONE 7, e37313 (2012). Demonstrates the successful application of ALIS to RNA targets.
Iwatani- Yoshihara, M. et al. Discovery of allosteric
Boger, D. L. & Cai, H. Bleomycin: synthetic and
Reviews
Anastasopoulou, P. et al. Synthesis of triazole-
Li, X. L. et al. RNase- L-dependent destabilization of
Costales, M. G. et al. Small-molecule targeted
Carter, B. J. et al. Site-specific cleavage of RNA
159
160
158
177
. Iwatani- Yoshihara, M. et al. Discovery of allosteric
Boger, D. L. & Cai, H. Bleomycin: synthetic and
Reviews
Anastasopoulou, P. et al. Synthesis of triazole-
Li, X. L. et al. RNase- L-dependent destabilization of
Costales, M. G. et al. Small-molecule targeted
Carter, B. J. et al. Site-specific cleavage of RNA
159
160
158
177
. Iwatani- Yoshihara, M. et al. Discovery of allosteric
Boger, D. L. & Cai, H. Bleomycin: synthetic and
Reviews
Anastasopoulou, P. et al. Synthesis of triazole-
Li, X. L. et al. RNase- L-dependent destabilization of
Costales, M. G. et al. Small-molecule targeted
Carter, B. J. et al. Site-specific cleavage of RNA
159
160
158
177
. Iwatani- Yoshihara, M. et al. Discovery of allosteric
Boger, D. L. & Cai, H. Bleomycin: synthetic and
Reviews
Anastasopoulou, P. et al. Synthesis of triazole-
Li, X. L. et al. RNase- L-dependent destabilization of
Costales, M. G. et al. Small-molecule targeted
Carter, B. J. et al. Site-specific cleavage of RNA
280. Angelbello, A. J. et al. A small molecule that binds an RNA repeat expansion stimulates its decay via the exosome complex. Cell Chem. Biol. 28, 34–45 (2021). Demonstrates that RNA-targeting small molecules can interface the target with natural decay pathways, in particular the exosome.

281. Murche, A. I. et al. Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. J. Mol. Biol. 356, 625–638 (2004).

282. Fopoppe, N. et al. A structure-based strategy to identify new molecular scaffolds targeting the bacterial ribosomal A-site. Bioorg. Med. Chem. 12, 935–947 (2004).

283. Morgan, B. S., Forte, J. E., Culver, R. N., Zhang, Y. & Hargrove, A. E. Discovery of key physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands. Angew. Chem. Int. Ed. Engl. 56, 13498–13502 (2017).

284. Childs-Disney, J. L. et al. A massively parallel selection of small molecule–RNA motif binding partners informs design of an antiviral from sequence. Chem 4, 2384–2404 (2018).

285.ickerton, G. R., Paulins, G. V., Bensnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. Nat. Chem. 4, 90–98 (2012).

286. Bhullar, K. S. et al. Kinase-targeted cancer therapies: progress, challenges and future directions. Mol. Cancer 17, 48 (2018).

287. Paulson, H. Repeat expansion diseases. Handb. Clin. Neurol. 147, 105–123 (2018).

288. Bernat, V. & Disney, M. D. RNA structures as mediators of neurological diseases and as drug targets. Neuron 87, 28–46 (2015).

289. de Mezer, M., Wojciechowska, M., Najpierza, M., Sobczak, K. & Krzyzosiak, W. J. Mutant CAG repeats of huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. Nucleic Acids Res. 39, 3852–3863 (2011).

290. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. & Krzyzosiak, W. J. RNA structure of trinucleotide repeats associated with human neurological diseases. Nucleic Acids Res. 31, 5469–5482 (2003).

291. Tian, B. et al. Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. RNA 6, 79–87 (2000).

292. Jog, S. P. et al. RNA splicing is responsive to MBNL1 dose. PLoS ONE 7, e48825 (2012).

293. Rzuczek, S. C. et al. Precise small-molecule recognition of a toxic CUG RNA repeat expansion. Nat. Chem. Biol. 13, 188–193 (2017).

294. Wagner-Griffin, S. et al. A druglike small molecule that targets r(CU)G repeats in myotonic dystrophy type 2 facilitates degradation by RNA quality control pathways. J. Med. Chem. 64, 8474–8485 (2021).

295. Bush, J. A. et al. Ribonuclease recruitment using a small molecule reduced c9ALS/FTD r(G4C2) repeat expansion in vitro and in vivo ALS models. Sci. Transl. Med. 13, eaba5991 (2021).

296. Angelbello, A. J. & Disney, M. D. A toxic RNA templates the synthesis of its own fluorogenic inhibitor by using a bio-orthogonal tetrazine ligation in cells and tissues. ACS Chem. Biol. 15, 1820–1825 (2020).

297. Burma, S., Chen, B. P., Murphy, M., Kurimasa, A. & Chen, D. J. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. J. Biol. Chem. 276, 42462–42467 (2001).

298. Li, Y. & Disney, M. D. Precise small molecule degradation of a noncoding RNA identifies cellular binding sites and modulates an oncogenic phenotype. ACS Chem. Biol. 13, 5065–5071 (2018).

299. Liu, X. et al. Targeted degradation of the oncogenic microRNA 17-92 cluster by structure-targeting ligands. J. Am. Chem. Soc. 142, 6970–6982 (2020).

300. Bertramou, R. I. et al. Structure-specific cleavage of an RNA repeat expansion with a dimeric small molecule is advantageous over sequence-specific recognition by an oligonucleotide. ACS Chem. Biol. 15, 485–495 (2020).