



Chem Soc Rev

Tutorial Review: Targeting RNA with small molecules: from fundamental principles towards the clinic

Journal:	<i>Chemical Society Reviews</i>
Manuscript ID	CS-TRV-09-2020-001261.R1
Article Type:	Tutorial Review
Date Submitted by the Author:	25-Nov-2020
Complete List of Authors:	Falese, James; Duke University School of Medicine, Biochemistry Donlic, Anita; Princeton University, Chemical and Biological Engineering Hargrove, Amanda; Duke University, Chemistry

SCHOLARONE™
Manuscripts

Tutorial Review: Targeting RNA with small molecules: from fundamental principles towards the clinic

James P. Falese, Anita Donlic, and Amanda E. Hargrove

Key Learning Points

1. The reader will learn about emerging cellular functions mediated by RNA structure and the potential for small molecule-based modulation of those functions.
2. The reader will learn about the unique physicochemical and structural properties of RNA that differentiate it from other biomacromolecules.
3. The reader will learn about the noncovalent interactions and driving forces behind small molecule-RNA recognition.
4. The reader will learn about common methods used to characterize RNA-small molecule interactions as well as associated challenges and emerging solutions in the field.
5. The reader will learn about pioneering and recent successful examples of small molecule-RNA targeting, including risdiplam, which has received clinical approval.

Introduction

Around the world, researchers are continuously searching for new drugs to treat a wide variety of diseases. While this process can involve continuing drug development for established disease targets, it can also involve searching for new disease targets that may be “druggable,” particularly for conditions that show drug resistance or that have limited treatment options. Currently, most therapies target proteins,¹ and while development of protein-targeted drugs has been crucial in modern medicine, limitations in developing protein-targeted therapeutics in many diseases have inspired some scientists to explore targeting other biomolecules.

The search for new drug targets has led scientists to explore targeting RNA, an interest which has been amplified as a result of the so-called “RNA Revolution,” which included the discovery of important RNA functions in multiple disease states.² Antisense oligonucleotides (ASOs) and small interfering RNAs (siRNAs), which act via complementary base pairing with the target of interest, have contributed greatly in the identification of therapeutically tractable target RNAs and have been successful as therapies in a few cases.³ Small molecule drugs, however, remain the avenue of choice for disease treatment when available, owing to generally more favorable pharmacological properties as well as the ease of tunability via medicinal chemistry approaches. Moreover, their ability to modulate RNA structure has made them complementary tools to sequence-recognizing ASOs.³

Further adding to rising interest in RNA targeting are recent discoveries of ubiquitous functional RNA molecules in human cells. For example, the Encyclopedia of DNA Elements (ENCODE) project, a follow-up study to the Human Genome Project, sought to map the functional regions throughout the genome.⁴ This project uncovered that while only 1.5% of our genome is ultimately translated into protein, about 70-90% is transcribed to RNA, showing that the vast majority of the genome yields noncoding (ncRNAs) over protein coding messenger RNAs (mRNAs).⁴ While multiple ncRNA classes were known at the time of the ENCODE project, including microRNAs (miRNAs) and long noncoding RNAs (lncRNAs), this study allowed for mapping of previously unidentified lncRNAs, miRNAs, and other ncRNAs in a variety of cell

types.⁴ In addition, viruses and bacteria have long been known to rely on regulatory RNAs, which have served as additional inspiration to explore similar structural elements in humans.³ Motivated by these discoveries, notable progress has been made in small molecule targeting of RNA, both in understanding the underlying principles of these interactions as well as demonstrating clinical efficacy.

This Tutorial will first overview RNA fundamentals to give the reader insight into the myriad functions of RNA and highlight how these functions are driven by structure. We will next discuss established and emerging principles of small molecule recognition of these structures as well as methods for their identification. We will then explore some of the remaining challenges and emerging solutions in the field, before ending with a selection of “Greatest Hits” that are paving the way towards clinical use or have already received approval. Due to length constraints, our focus will remain on direct small molecule-RNA interactions, rather than drugs that indirectly modulate the transcription or promote the degradation of an RNA of interest, though promising work has been accomplished in those areas as well.⁵ We apologize to the authors of many works that could not be cited due to reference limitations, and we invite the reader to consult the review articles cited herein to learn more about RNA biology and targeting. We hope that this tutorial review serves as an inspiration and resource for both current researchers as well as aspiring scientists and students alike who are unfamiliar with the RNA field to join us in this exciting quest for RNA-targeting small molecule drugs.

Section 1: Overview of RNA Classes and Targetable Functions

There are multiple classes of RNA, with the first classification level often involving their coding potential. While mRNAs are translated into protein, it is important to note that not all regions of a coding RNA undergo translation. For example, pre-mRNAs can contain introns that are removed by splicing and also contain untranslated regions (UTRs) that can have functional roles.² One example of functional regulation via RNA structure in 5'-UTRs, called the riboswitch, is explored below. In contrast, RNAs that do not undergo translation are termed ncRNA. Some of these are essential for the translation process: transfer RNAs (tRNAs) carry the amino acid building blocks for proteins, and ribosomal RNA (rRNA) catalyzes the reactions necessary for polymerizing the amino acids into a peptide chain. Other ncRNA molecules regulate transcription and signaling events and can be further classified by their lengths and functional roles within the cell. Together, this network of coding and ncRNA molecules, while concurrently challenging the traditional view of RNA as a messenger between DNA and proteins, represents an open space for therapeutic exploration and discovery.

This section will review several important examples from RNA classes in both human and microbial cells that represent attractive therapeutic targets for small molecules. Human RNAs are implicated in conditions such as cancer or neurodegenerative diseases, which can be characterized by dysregulation of various, disease-specific ncRNAs. At the same time, therapeutics for various microbial diseases may be attained by targeting unique RNAs within viruses or bacteria. Overall, misexpressed human RNAs may represent an untapped class of potential drug targets, and viral and bacterial RNAs may allow for the development of novel antimicrobial agents.

miRNAs

miRNAs are initially synthesized as larger primary miRNAs (pri-miRNA) that are first processed by the enzyme Drosha in the nucleus, followed by Dicer in the cytosol, to yield a processed, single-stranded miRNA product that is about 17-22 nucleotides in length.² miRNAs

regulate mRNA expression through complementary base pairing, causing mRNA degradation or inhibition of translation. It is important to note that one miRNA can have multiple targets, meaning that dysregulation of one miRNA can result in wide-sweeping changes in gene expression. Indeed, dysregulation of miRNAs is implicated in disease states such as cancer.²

lncRNAs

lncRNAs are defined as RNAs being over 200 nucleotides in length that lack coding function. The aforementioned ENCODE project specifically identified almost 10,000 lncRNA genes,⁴ many of which were subsequently found to have a variety of cellular functions, such as guiding epigenetic modification complexes or functioning as decoys that divert another biomolecule away from its target.² For example, the lncRNA X-inactive Specific Transcript (XIST) is involved in X-inactivation for dosage compensation in female mammals.⁶ Specifically, XIST becomes overexpressed from the X chromosome that will become inactivated, coating the chromosome and recruiting a multitude of proteins including chromatin remodellers such as histone deacetylases and the Polycomb repressive complex 2 (PRC2), thereby promoting gene silencing and chromosome inactivation.⁶ As with miRNAs, lncRNAs can also be dysregulated in disease. For example, lncRNA Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1) plays roles in normal pre-mRNA splicing by altering the localization and phosphorylation status of splicing factors as well as regulating gene expression.⁷ However, the overexpression of MALAT1 appears to facilitate various cancer processes. As explored below, MALAT1 has a unique structure that allows for its overaccumulation,⁷ and it has been successfully modulated by small molecules.^{8,9}

Short tandem repeats and repeat expansions

Multiple RNAs contain 1-6 nucleotide repetitive sequences called short tandem repeats. The exact number of these repeats can vary across individuals, and while naturally-occurring, an excess number of repeats called “expansions” in certain genes results in a variety of genetic conditions.¹⁰ These repeats can sequester important proteins to prevent them from executing their normal functions or, if translated, create toxic polypeptides. For example, myotonic dystrophy type 1 (DM1), discussed in the “Greatest Hits” section, results from such an expansion.¹⁰

Viral and bacterial RNAs

Multiple viruses utilize RNA for their genome instead of DNA, including those that cause HIV (Human Immunodeficiency Virus), influenza, and hepatitis.¹¹ The HIV-1 RNA genome in particular contains several structural elements that have been successfully targeted with small molecules to inhibit viral replication.¹¹ These successes and others serve as inspiration to target a multitude of other and less explored viruses with RNA genomes. Of particular note, coronaviruses, including the pandemic Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) that causes COVID19, are in this category.¹¹

In addition, bacterial systems rely on metabolite interactions with RNA structures called riboswitches to control gene expression.¹² Riboswitches are typically found in the UTR of a gene of interest and are able to detect a related metabolite concentration by direct binding. This event induces a conformational change that allows for transcription or translation of the gene to be regulated. Riboswitches have been identified for multiple metabolites, including amino acids, nucleotides, metals, and vitamins.¹² Targeting of riboswitches with synthetic small molecules may

have antibiotic potential, as they are critical for bacterial survival, and similar structures have not been found in human cells.¹³

Section 2: RNA structure and properties

Primary RNA structure and backbone

RNA structural architecture is often referred to as consisting of primary, secondary and tertiary structure. The primary structure of RNA is the nucleotide sequence comprised of three building blocks as described in Figure 1A.¹⁴ Specifically, the “glycosidic bond” connects the ribose sugar to one of four aromatic, nitrogen-containing nucleobases: adenine (A) and guanine (G) purines, and cytosine (C) and uracil (U) pyrimidines (Figure 1A). The negatively charged sugar-phosphate backbone is comprised of phosphodiester and ribose groups, which, owing to the 2'-hydroxyl moiety, induce distinct sugar “puckering” patterns compared to DNA. The backbone structure is highly dependent on the solvent environment, as RNA folding has to overcome electrostatic repulsion from the phosphate groups. This repulsion can be overcome through association of metal ions, such as a divalent magnesium or monovalent sodium.¹⁴

Base pairing patterns and RNA secondary structure

Secondary structure refers to local folds in the RNA structure created through base pairing, or the interaction between hydrogen bond donor and hydrogen bond acceptor groups on RNA bases. The most common secondary structure in RNA forms through the Watson-Crick-Franklin (WCF) hydrogen bonding pattern, where the canonical base pairs consist of A-U and G-C (Figure 1B).[†] RNA base pairing, similar to DNA, allows for the formation of a double helix or duplex. The conformation of the ribose sugar in the RNA backbone leads to the formation of an A-form duplex, distinct from the B-form structure in DNA, and is stabilized through highly favorable and cooperative stacking interactions formed between the planar and polarizable nucleobase surfaces (Figure 1C, black dashes). At the same time, the increased number of hydrogen bonds results in G-C base pairs having slightly greater stability than A-U.

In addition to duplex regions, nucleotides that do not form hydrogen bonds enable the formation of single-stranded secondary structures unique to RNA. These motifs include apical loops, internal loops, and bulges (Figure 1D).¹⁵ Together, combinations of base paired and single-stranded structural “modules” throughout an RNA provide distinct topologies that may serve as recognition sites by ligands, including proteins and small molecules.

Further adding to the structural complexity of RNA are non-canonical base pairing patterns. The stability of non-canonical base pairs can depend on several factors, including sterics if they require rotation of the glycosidic bond,¹⁶ protonation of the bases, and/or direct binding of metals. For example, uridines or protonated cytosines can engage in base pairing with pyrimidines on the other side of the WCF interface. This hydrogen bonding pattern is referred to as Hoogsteen, and it can result in a base triple when combined with the WCF base pair (Figure 1E). Additionally, uridines can bond with both guanines and protonated adenosines via “wobble” interactions at the WCF face (Figure 1F). Purines can also interact with each other: guanines can base pair with adenosines through WCF-WCF (imino G-A) or Hoogsteen-sugar edge (sheared G-A) conformations (Figure 1G). Together, these non-canonical interactions can provide stabilizing forces for the formation of RNA tertiary structures, which serve as additional recognition sites for natural or synthetic ligands.

RNA tertiary structure

RNA tertiary structures form when secondary structure elements participate in through-space interactions. While these regions may be distant in primary or secondary structure, they come into proximity during RNA folding. The driving force towards formation of these structures is maximization of base stacking, similar to the folding of proteins that is driven by the hydrophobic effect.¹⁶

The folding architecture of RNA tertiary structures is largely dependent on coaxial stacking of adjacent helices (the linear alignment of the central axes of two separate helices), which is influenced by the RNA junction (the central connection of adjacent helices) topology.¹⁶ A prominent early example came from the structural characterization of the yeast phenylalanine transfer RNA by X-ray crystallography (Figure 2A).¹⁷ In addition to coaxial stacking factors, long-range complementarity in nucleotide sequence enables base pairing interactions that drive the folding of structures such as kissing loops and pseudoknots (Figure 2B and C).¹⁵ Furthermore, consecutive base triples have been documented in MALAT1, forming a unique triple helix structure (Figure 2D).⁷ This is believed to confer stability and prevent degradation of the abundant MALAT1 transcript. Lastly, backbone desolvation and charge screening or direct binding by metals can promote RNA tertiary structure formation. For example, guanines are known to form tetrads with each other called G-quartets that are coordinated by potassium ions, which may then stack on top of each other to form highly stable G-quadruplex structures (Figure 2E) that have been implicated in disease.¹⁸ Taken together, these unique tertiary structures offer opportunities for specific recognition by small molecules to enable the interrogation of structure-function relationships and therapeutic potential of these folds.

RNA quaternary structure

Diversity of RNA function is often achieved by RNAs interacting with a wide variety of other biomolecules, including DNA, proteins, and other RNAs.² Such assemblies are referred to as RNA quaternary structures. While all of these interactions can have functional importance, we will focus on RNA:protein interactions, as they represent a class of quaternary structures that are actively being investigated for small molecule regulation.

Understanding the structure of RNA:protein interactions can give further insight into RNA biology and potential targeting of these interfaces with small molecules. For example, many RNA-binding proteins (RBPs) utilize stacking between nucleobases and aromatic amino acids, and/or charge-charge interactions between cationic amino acids and the RNA backbone, to promote binding.¹⁴ Stacking and electrostatic interactions appear to be important for RNA recognition from both a protein and small molecule perspective, as discussed below.

In addition, some RBPs are known to have preferential interactions with particular RNA tertiary structures. For example, the aforementioned PRC2 complex appears to interact with a multitude of RNAs. In an effort to explain this promiscuous binding, Cech, Davidovich, and co-workers found that PRC2 preferentially recognized RNAs at G-quadruplex sites.¹⁹ Understanding the sequence and structure determinants of RNA-protein interactions will facilitate the discovery of small molecules that modulate these interactions, either by direct disruption or allosteric regulation of these complexes.

It is also important to note that RNAs can form higher order assemblies containing multiple protein and RNA subunits. For example, the spliceosome, an RNA:protein complex that catalyzes the RNA splicing reaction, has been analyzed structurally. Notable work from Yan, Shi, and co-workers²⁰ yielded important insights into the intermolecular interactions between the various subunits of the spliceosome, including the RNA substrate, protein subunits, and small nuclear

RNAs (snRNAs), a family of approximately 150-nucleotide RNAs involved in the splicing process.²⁰ Another significant RNA-protein assembly is the ribosome, where structural insight has allowed for understanding of the molecular interactions that take place during translation. For example, the bacterial 70S ribosome, consisting of both the small and large ribosomal subunits and mRNA and tRNA subunits, was analyzed via X-ray crystallography by Ramakrishnan and co-workers.²¹ The authors observed extensive RNA-RNA and RNA-protein interactions, with many interactions requiring magnesium coordination and other electrostatic interfaces.²¹

These higher-order assemblies serve as great inspiration for successful RNA targeting in the clinic, as small molecule binders of the bacterial ribosomal RNA and splice sites represent the only FDA-approved RNA-binding drugs to date. Specifically, FDA-approved aminoglycoside antibiotics such as neomycin bind bacterial ribosomal RNA and hinder translation.²² These molecules have found extensive clinical use and have been reviewed elsewhere.²² Risdiplam, the recently approved splicing modifier for treating spinal muscular atrophy (SMA) will be explored below. Together, these advances serve as proof-of-principle for the targeting of RNA-containing complexes.

RNA dynamics

All levels of RNA structure are prone to changes that are regulated by a variety of dynamic cellular cues, such as protein binding and ionic environment. Additionally, post-transcriptional modifications, which are reversible covalent modifications of RNA nucleotides, are also known to modulate RNA structural dynamics.^{23,24} Advances in structure determination methods continue to provide evidence of diverse conformational landscapes found in RNA, leading to a paradigm shift of considering these molecules as containing structural “ensembles,” i.e. a collection of structures, rather than a single static structure. Importantly, distinct ensemble members are known to differentially regulate functional outcomes and can also provide unique sites for ligand binding and consequent modulation of the conformational equilibria.²⁵

Local changes in RNA dynamics can be induced by post-transcriptional modifications and other regulatory processes and lead to major functional consequences. For example, *N*⁶-methyladenosine (m⁶A) is known to disfavor WCF base pairing due to steric clashes and to simultaneously enhance base-stacking in single-stranded regions, thus altering RNA structure and dynamics.²³ The presence of m⁶A in a hairpin loop at the 5'-end of the aforementioned lncRNA MALAT1 was shown to increase dynamics, thereby enhancing heterogeneous nuclear ribonucleoprotein C (hnRNP C) binding (Figure 3A).²⁴ In addition, secondary structure rearrangements can affect protein binding. For example, the HIV Trans-Activation Response (TAR) element RNA can sample both ground (GS) and excited state (ES) base-pairings (Figure 3B), with only the GS able to bind the Cyclin T1-Tat protein complex. The specific biological impact of this sampling is under investigation, but it offers an opportunity for small molecule control of viral transcription and ultimately virion production.²⁶

Collectively, the folding landscapes and dynamic nature of RNA demonstrate its unique structural architecture and properties compared to other biomolecules. While more similar to DNA than proteins in its building blocks, the dynamic nature of RNA enables this biomolecule to access a broader range of conformations compared to DNA.²⁷ Furthermore, the ensemble nature of a single RNA molecule distinguishes it from proteins, which often form one major thermodynamically stable conformation.²⁸ Indeed, the RNA folding landscape is now widely considered to be more “rugged” than that of proteins, as multiple energetically favorable conformations can exist in equilibrium with each other.²⁸ It is therefore possible that the

conformational diversity of RNA molecules compensates for its limited chemical diversity in terms of the four nucleotide building blocks, compared to protein's twenty amino acids, and that this diversity provides opportunities for unique regulation of this biomolecule by small molecule ligands.

Section 3: Principles and methods for studying RNA-small molecule recognition

A unique property of small molecules compared to oligonucleotides is their ability to specifically recognize binding pockets in RNA and thereby modulate a transcript's function.³ Structural and biophysical characterization of these binding pockets has provided insights into the major driving forces that enable recognition. The following section will highlight several noncovalent interactions important in small molecule-RNA recognition as well as discuss common and emerging methods utilized to detect and characterize binding events.

Fundamental recognition principles and noncovalent interactions

With the increase in RNA-small molecule structure characterization efforts, there are opportunities to analyze these data sets and identify general trends in recognition of RNA by small molecules. The identification of such trends is expected to expedite the discovery of general guidelines for selective RNA targeting as well as clarify differences in recognition of RNA versus proteins (the "traditional" drug targets) by small molecules.

A recent analysis of all RNA-small molecule complexes and comparison to protein-small molecule complexes in the Protein Data Bank (PDB, a depository of solved biomolecular structures) by the Hargrove group, for example, found differences in the types of interactions that are enriched in small molecule recognition with each biomacromolecule.²⁹ For example, the major contributors for RNA recognition by small molecules are stacking and hydrogen-bonding interactions. Stacking interactions are enabled by small molecules with aromatic surfaces. For example, ligands containing extended aromatic ring systems are known to intercalate, i.e. insert themselves between stacked base pairs and strengthen the existing π - interactions. A notable example from this analysis is an acridine-based ligand discovered to stabilize a G-quadruplex structure found in a lncRNA Telomeric Repeat-Containing RNA (TERRA, Figure 4A).³⁰ Additionally, specific hydrogen bonding can occur between small molecules and nucleotides. In another example from this analysis, the aminoglycoside paromomycin uses its hydroxyl and ether groups to hydrogen bond with an adenine residue in the HIV-1 dimerization initiation site, another important RNA regulatory element (Figure 4B).³¹

In contrast to proteins, interactions were found to be less driven by the hydrophobic effect in RNA, though these effects still play a critical role.²⁹ Complementing this finding is a recent study by Schneekloth and co-workers, in which ligand-bound binding pockets in RNA and proteins from the PDB were compared.³² While these binding pockets were found to have similar volume and buriedness (lack of solvent exposure) parameters, binding pockets in RNA were measured to be much less hydrophobic, consistent with the more polar nature of this biomolecule and suggesting a potential opportunity for specific RNA targeting.³² It is worth noting that the knowledge gained from these studies is currently limited to RNA and proteins that are amenable to structural characterization, and that these principles may be different for more dynamic and disordered recognition regions in both biomolecules. Nevertheless, the identification of differences in both interaction types and hydrophobicity of the binding pockets between these targets has led some to hypothesize, as explored below, that there may be privileged chemical properties of RNA-

binding small molecules. As such, this research paves the way towards exploiting distinct RNA-binding properties to rationally design selective ligands for these targets.

Biophysical methods and considerations to detect RNA-small molecule interactions

To detect RNA-small molecule binding events, several *in vitro* screening methods can be used. This section will review commonly used methods as well as considerations for future studies, including expanding the repertoire of binding techniques and inclusion of functional assays.

Fluorescence-based methods – Most current methods for identifying small molecule RNA binders rely on the sensitive measure of fluorescence. For example, relative binding affinities can be measured by small molecule titrations with fluorescently-labeled RNA constructs,³³ fluorescence indicator displacement (FID) assays,³⁴ or by monitoring changes in intrinsically fluorescent ligands upon RNA titration.⁸ In FID assays, binding is measured by a small molecule of interest “displacing” an RNA-binding indicator molecule, which causes a fluorescence change in the indicator (Figure 5A).³⁴ To detect binding events in a high-throughput fashion, fluorescently-labelled RNA constructs can be incubated on small molecule microarrays (SMMs), as demonstrated by Schneekloth and co-workers (Figure 5B).³³ In these assays, small molecules are spotted on a surface via covalent linkages, and the fluorescent RNA is retained on a spot if it is interacting with the molecule.

While useful approaches for identifying *de novo* RNA-small molecule interactions, these methods suffer from several disadvantages that call for careful consideration. FID assays can be biased in detecting ligands with similar binding modes and affinity ranges as the indicator used. Similarly, methods that rely on fluorescently labeled RNA or intrinsically fluorescent ligands may fail to detect binding events that do not result in fluorescence changes, and the reliance of SMMs on ligand immobilization may not accurately represent ligand behavior in solution. Moreover, in systems that aim to detect disruption of high-affinity binding events, such as inhibition of RNA-protein interactions, fluorescence-based measurements may need to be replaced with more sensitive measures of radiolabeling. Finally, all of these methods can benefit from follow-up studies using other methods.

To gain a deeper understanding of fundamental principles of RNA-small molecule recognition, detailed biophysical studies of such interactions are needed.³⁵ Evaluating thermodynamic and kinetic parameters of these interactions would provide important insights into differential small molecule-based regulation of RNA structural landscapes, and such parameters cannot be directly measured using the aforementioned techniques.

Isothermal Titration Calorimetry (ITC) – For thermodynamic insights, ITC is most commonly used.³⁵ This method provides a direct measurement of a binding constant (K_d) by reporting the heat change from small molecule addition to the RNA solution, and the resulting curve can be used to calculate all thermodynamic parameters (Figure 5C). As such, this information can be used to determine enthalpic and entropic contributions in small molecule-RNA binding events.

Surface Plasmon Resonance (SPR) – For kinetic insights, SPR experiments are typically conducted.³⁵ By immobilizing the RNA on a glass surface and measuring changes in the refraction angle upon small molecule binding, this method outputs k_{on} and k_{off} parameters (Figure 5D). Understanding if high affinity binding events are driven by fast on-rates or slow off-rates for small molecule-RNA systems, and how these differ for different small molecule and/or RNA classes, will be of significance.

Ideally, binding assays should be complemented with RNA-focused functional assays to fully characterize RNA-small molecule interactions as well as to establish correlations between binding strengths and functional outcomes, which is of rising interest in the field.³ This consideration is also of particular importance given the increase in phenotypic screening campaigns that result in the identification of RNA as targets, as seen recently in studies by Novartis,³⁶ Roche,³⁷ and others. Together, such combined studies of biophysical and functional methods can increase our understanding of requirements for bioactivity of RNA-binding ligands and accelerate the path towards therapeutic drug discovery for these targets.

Structural approaches for analyzing RNA-small molecule interactions

Along with understanding RNA-small molecule interactions through the binding and biophysical assays discussed above, it is also important to gain structural insight into such interactions. Compared to over 100,000 crystal structures and over 10,000 NMR structures for proteins, there are only over 800 crystal structures and over 500 NMR structures of RNAs currently deposited in the PDB. The apparent paucity of known RNA structures is partly the result of several challenges associated with RNA structure determination, including its high negative charge and dynamic nature. Methods for determining RNA structure can be broadly characterized as solid-state or solution-state. This section will focus on four major methods: X-ray crystallography, cryo-electron microscopy (cryo-EM), nuclear magnetic resonance (NMR) spectroscopy, and chemical probing.

Solid-state methods

X-ray crystallography – In X-ray crystallography, crystals containing a molecule of interest are analyzed with X-rays, yielding a diffraction pattern that serves as an electron-density map through which the structure of the biomolecule can be modelled. X-ray crystallography has been a major workhorse for biomolecular structure analysis, with one of the first examples of RNA crystallography in the aforementioned tRNA crystal structure (Figure 2A).¹⁷ Since this work, X-ray crystallography has given crucial insight into RNA apo structures, RNA-RNA interactions, and RNA-protein complexes, most notably the ribosome.²¹ The aforementioned TERRA lncRNA was crystalized with the acridine mimic shown in Figure 2E. We will also see the example of Ribocil in the Greatest Hits section.

X-ray derived structures based on crystals generally capture only one conformation of the RNA. Thus, other conformations sampled by the RNA can be missed in these experiments. In addition, as a solid state method, a crystal structure may not be representative of a biologically-relevant structure, as crystal packing may induce non-native conformational changes.³⁸ Finally, the high dynamics and negative charge on RNA can hinder crystal formation, meaning that crystallography is not attainable for many RNAs, in particular large RNAs.

Cryo-EM – Cryo-EM refers to the rapid incubation of a sample at cryogenic temperatures with subsequent electron microscopy analysis. This rapid cooling results in the formation of vitreous ice, where the water molecules in the solvent do not have time to form a crystal lattice. As a result, the molecule of interest is protected from damage from freezing while maintained in a water-containing environment, and the cold temperatures minimize damage from electron radiation.³⁸ Cryo-EM has allowed for significant understanding of both the ribosome³⁸ and spliceosome.²⁰ Notably, through single-molecule analysis of the electron micrograph, multiple structural “states” can be observed. These states can be used to generate structural ensembles to gain further insight into RNA dynamics, which is not achievable with X-Ray crystallography.³⁸

It is anticipated that further advancements in cryo-EM will facilitate analysis of RNA:ligand complexes even at lower molecular weights. Again, as a solid-state method, it is possible that the solved structures may not necessarily represent the biologically-relevant structures, although circumventing crystal packing may make such artifacts less likely than in crystallography.³⁸

Solution-state methods

NMR spectroscopy – NMR spectroscopy utilizes magnetic fields to analyze a molecule of interest, wherein individual atoms from the molecule produce chemical shift values related to their identity and surrounding environment. As this method is solution-state, RNA is more likely to assume a biologically-relevant conformation, although there are limitations due to buffer conditions that can be used in NMR. In addition, chemical shift overlap can occur due to the similar chemical functionality of the nucleobases. Thus, NMR analysis is typically limited to shorter RNAs (optimally 50 or fewer nucleotides), unless site-specific radiolabeling is accomplished.³⁹ Notably, Telesnitsky, Summers, and co-workers solved an NMR structure for HIV-1 5'-UTR (155 nucleotides), the longest RNA-only complex analyzed by NMR to date, utilizing site-specific labeling methodologies.⁴⁰ NMR analysis can also allow for identification of structural ensembles, as was identified by Al-Hashimi and co-workers in a virtual small molecule screen against an RNA element in HIV-1, discussed below.⁴¹ Apart from structural information, NMR can also yield information on ligand binding. For example, in water-Ligand Observed via Gradient Spectroscopy (waterLOGSY) magnetization is transferred from the water to the ligand in the presence and absence of an RNA target via through-space interactions called an intermolecular Nuclear Overhauser Effect (NOE).³⁹ Due to the size difference between the RNA-ligand and ligand-only complex, binding can be assessed. We will see another example in the Greatest Hits section where NMR was used to solve the structure of an RNA-small molecule interaction in splicing regulation.

Chemical probing – Chemical probing refers to incubating RNA with a reactive chemical that forms adducts on nucleobase or sugar groups in a manner that depends on base-pairing or flexibility. There are multiple probing reagents that modify the RNA at distinct positions. For example, SHAPE reagents (Selective 2'-hydroxyl acylation analyzed by primer extension) modify the 2'-hydroxyl of the ribose sugar of conformationally flexible nucleotides, generally corresponding to single-stranded regions (Figure 6A, top).⁴² Another reagent, dimethylsulfate (DMS) modifies A and C nucleobases at the WCF face, giving insight into base pairing (Figure 6A, bottom).⁴² After chemical probing, the RNA is reverse transcribed into copy DNA (cDNA). Modified sites result in either the truncation in the cDNA, where the reverse transcriptase halts, or a mutation that can be detected by deep sequencing in the more recently developed mutational profiling (MaP) approach (Figure 6B).⁴² However, it is important to note that reactivity data is used as a restraint in an RNA folding algorithm. Thus, the accuracy of a determined structure is dependent on the validity of the folding algorithm.

Schneekloth and co-workers³³ performed chemical probing in the presence of a small molecule identified utilizing the aforementioned SMM approach against HIV-1 TAR. Probing was carried out with the entire 5'-untranslated region (UTR) of HIV-1, which contains the TAR RNA to validate that the small molecule binding site is maintained even in the presence of a larger RNA construct. The authors noticed that the most significant changes in nucleotide reactivity were observed near the TAR structure, allowing for validation that this site is indeed the small molecule binding site (Figure 6C).³³ Thus, chemical probing represents a potential approach for validating small molecule binding sites and selectivity particularly on large RNAs.

Section 4: Challenges and Emerging Solutions

While the established and emerging methods for studying RNA-small molecule interactions are paving the way towards understanding factors contributing to selective recognition and ultimately clinical advances, many unanswered questions remain in this flourishing field. Both the RNA and small molecule perspectives will be discussed in the following sections.

RNA perspective

Identification of suitable RNA targets – From the RNA perspective, one major challenge is identifying RNA molecules that are amenable for selective small molecule targeting. Contributing to this challenge is the relative lack of 3D structures for RNA, as highlighted in the previous section. At the same time, a small molecule must identify an RNA of interest in the cell where a large majority (>90%) of RNA consists of rRNA and tRNA. As drug-development efforts extend to capitalizing on RNA's potential for therapeutic targeting with small molecules, several criteria can be considered when evaluating the suitability of a given RNA.³ One criterion is RNA abundance, as a higher amount of the RNA target in diseased tissue is likely to decrease off-target effects, especially considering that RNAs can have tissue- or disease-specific expression. Additionally, Weeks and co-workers proposed that ideal small molecule recognition sites on RNA would consist of a binding pocket analogous to protein targeting and also have high “information content,” which is defined as structural complexity that makes that particular site unique (e.g. 1 in a 1,000,000,000 from other RNA structures).¹

Leveraging RNA conformational landscapes – RNA dynamics complicates our understanding of small molecule recognition. Dynamics presents challenges for structure-based drug design, as RNA cannot be described as a static structure and a small molecule ligand may only bind one of multiple ensemble members. An emerging strategy to exploit these dynamics is identifying and trapping inactive or non-functional RNA conformations with small molecules. In this case, small molecule-based regulation of RNA function occurs by altering its conformational landscape, as already demonstrated in nature with riboswitch-metabolite binding. Importantly, this strategy is shifting the view that RNA dynamics is an obstacle for selective targeting to leveraging this property towards specific recognition.

In one example, Al-Hashimi and co-workers combined NMR studies and computational simulations to generate an ensemble of representative TAR conformations.⁴¹ Small molecules were computationally docked to this set of structures, rather than just one structure, to more fully account for the role of RNA dynamics in small molecule binding. This study uncovered a novel compound, dimethylamiloride (DMA), that showed moderate affinity for TAR.⁴¹ In follow-up collaborative studies, the Al-Hashimi and Hargrove groups ultimately found that DMA-169 can stabilize the aforementioned ES conformation of TAR RNA (Figure 7A), as can argininamide (Figure 7B). Interestingly, other TAR-binding ligands, such as neomycin (Figure 7B) do not stabilize this conformation.⁴³ Since the ES conformation is poorly recognized by the Tat protein complex,²⁶ these results provide evidence that a strategy of modulating structural ensembles with synthetic ligands may indeed be used to inhibit RNA function.

Small molecule perspective

From the small molecule perspective, many have suggested that we are hindered by a limited understanding of the general driving principles behind preferential and selective RNA recognition. Several unanswered questions remain. For example: what is the interplay between

affinity and selectivity necessary for favorable binding *and* functionally-relevant events? What are the physical and chemical properties of molecules that lead to such events?

Insights into this area would aid in busting the myth of RNA being perceived as ‘undruggable’, a notion that largely emerged as a result of repeated identification of promiscuous chemical scaffolds and low hit rates in high-throughput screening campaigns targeted against RNA. One hypothesis is that these initial failures may be the result of using small molecule screening libraries that are enriched in protein-binding chemical scaffolds, as they represent the traditional target for small molecule ligands. As discussed in previous sections, the forces and noncovalent interactions that drive small molecule recognition of RNA and proteins are different, providing support for this hypothesis.

Exploration of RNA-privileged chemical space – To address the possibility that screening libraries are biased towards proteins, an understanding and validation of an RNA privileged “chemical space” is necessary. In cheminformatics research, chemical space is a concept referring to a multi-dimensional space defined by properties analyzed for a given set of molecules. Molecules that adhere to similar properties will be located in the same region of that space, which enables comparisons of small molecule groups. By quantifying and comparing twenty physicochemical properties of RNA- and protein-binding molecules, as represented by the curated RNA-Targeted Bioactive ligaNd Database (R-BIND) and FDA libraries, respectively, the Hargrove group indeed found that RNA-binding small molecules have different properties from those of protein-binding small molecules (Figure 8A).⁴⁴ Similar distinguishing properties for hit RNA-binding ligands from a large Astra Zeneca library were later identified by Disney and co-workers.⁴⁵ While similar in properties, the hit ligands contained previously unexplored chemical matter in RNA targeting, supporting the notion of chemical space analyses as a tool to identify novel RNA-privileged chemotypes.

The finding of a putative “RNA-privileged” chemical space further corroborates the view of differential driving forces for RNA recognition by small molecules and offers the possibility of enriching screening libraries with the newly discovered properties to increase hit rates. Indeed, recent work by Merck demonstrated that building an RNA-focused set of small molecules upon identifying the physicochemical properties that prevent protein binding ultimately increased propensity towards binding an array of RNA targets.⁴⁶ To make the curation of RNA-privileged screening libraries accessible to the broader scientific community, Hargrove and co-workers recently reported a website interface with a “Nearest Neighbor” algorithm, which identifies how similar a molecule of interest is to R-BIND database members (Figure 8B).⁴⁷ Ultimately, these and similar approaches are expected to provide tools for scientists for efficient discovery of novel RNA-binding small molecules.

Towards Specific Sequence and Motif Targeting – Another area of interest has been in identifying properties of small molecules that enable preferential recognition of a specific RNA sequence or structural motif over another. Towards this goal, Disney and co-workers conducted 2-dimensional combinatorial screening (2DCS) by testing a range of secondary structure motifs for binding to a large set of small molecules⁵ and then curated a database of these interactions called infoRNA to serve as a discovery platform (Figure 8C).⁴⁸ Analyses of the database have also led to the identification of privileged chemotypes for targeting specific secondary structure motifs (Figure 8D). In another example, the aforementioned PDB analysis of RNA-small molecule complexes found an enrichment of cation- π and stacking interactions in purine-rich sites, offering opportunities for rational design.²⁹ Finally, advances in machine learning technologies can enable additional insights. Both the recent finding of enriched non-canonical base pairing interactions in

selective small molecule-RNA recognition events,⁴⁹ as well as the application of pattern recognition methodologies to characterize differential small molecule-based recognition of RNA structures,⁵⁰ may be exploited for rational design in the future.

Section 5: Selection of “Greatest Hits” to Date

In this section, we will cover the recent advances and promising biological examples of small molecule regulation of RNA. We will first explore molecules that have been validated in cellular systems before moving to molecules with evaluation in animal models. We will conclude this section with a recently FDA-approved drug, showcasing the promise and clinical utility of RNA-targeting small molecules.

MALAT1/RNA triple helices

As discussed above, the MALAT1 lncRNA is overexpressed in multiple cancers and drives cancer proliferation. Notably, a triple helix structure appears to facilitate its cellular abundance and overexpression in cancer and other diseases, thereby posing this element as an attractive target for small molecules.⁷ The Hargrove group generated derivatives of a known nucleic acid scaffold called furamide and identified DPF (diphenylfuran) p8 (Figure 9A) as the first high affinity and selective molecule for the triple helix.⁸ Since this work, Baird, Le Grice, and co-workers identified two MALAT1 triple helix binders, SM5 (Figure 9B) and SM16, through the aforementioned SMM platform. These molecules were shown to modulate triple helix dynamics *in vitro* as well as reduce MALAT1 transcript abundance and branching in breast cancer organoids (Figure 9C).⁹ More recent work by the Hargrove group found that DPFp8 derivatives were triple helix stabilizers and can prevent *in vitro* exonucleolytic degradation, making these molecules complementary to SM5/SM16 ligands.⁵¹ Together, these probes have strong potential for future mechanistic studies of the triple helix structure, as well as spatiotemporal regulation of MALAT1 transcript abundance.⁵¹

RNA repeat expansions

Repeat expansions are promising drug targets as they have a large number of potential identical binding sites for a small molecule. In a notable example, Disney and co-workers¹⁰ targeted a CUG expansions in myotonic dystrophy type 1 (DM1). This condition results from hundreds to thousands of CUG expansions in the dystrophin protein kinase (DMPK) 3'-UTR, whereas a healthy individual typically has less than 30 expansions. These repeats sequester the Muscleblind-like 1 (MBNL1) splicing regulator, which in turn causes mRNA splicing abnormalities.¹⁰ Using the aforementioned infoRNA database, the authors generated a multivalent molecule, 2H-K4NMeS (Figure 10A), which rescued splicing defects (Figure 10B) and reduced the number of CUG:MBNL1 foci within the cell (Figure 10C).¹⁰

Enterovirus 71 internal ribosomal entry site

As noted above, there is growing interest in small molecule targeting of viral RNAs.¹¹ A recent contribution by Brewer, Hargrove, Tolbert, and co-workers targets the internal ribosomal entry site (IRES) of Enterovirus 71 (EV-71), one of the main agents of hand, foot, and mouth disease.⁵² The EV-71 IRES relies on a structure, called stem-loop 2 (SLII), that interacts with human RBPs, including hnRNP A1, which promotes viral translation, and AUF1, which suppresses translation. By screening an amiloride-based library, the authors discovered one molecule, DMA-135 (Figure 11A), that exhibited strong binding for SLII and caused a significant perturbation in the SLII structure (Figure 11B). DMA-135 significantly decreased viral translation

in a dose dependent manner (Figure 11C), significantly increased the interaction between SLII and AUF1 in cells (Figure 11D), and significantly reduced viral titer in a dose-dependent manner.⁵² This study supports the use of small molecules for allosteric regulation of RNA structure and stabilization of RNA-protein interactions.

RNA G-quadruplexes in cancer and neurological disorders

G-quadruplexes are found in the mRNAs of several oncogenes and can reduce translation. As a result, the stabilization of these structures via small molecules is an attractive therapeutic goal for the treatment of cancer.⁵³ Multiple examples of small molecules targeting G-quadruplexes have been documented, including for the mRNA for *NRAS* and *KRAS*,⁵³ which are implicated in multiple cancers but have not been successfully inhibited in the clinic.³

Targeting G-quadruplexes has also been explored in neurological disorders.³ For example, a GGGGCC repeat expansion in *C9orf72* is implicated in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD).⁵⁴ These repeats can generate nuclear RNA foci that bind and antagonize the healthy function of other RNA-binding proteins by sequestering them as well as be translated into toxic, cytoplasmic dipeptide repeats. Both species have been hypothesized to contribute to the pathology of ALS and FTD. Neidle, Patani, Fratta, Isaacs, and co-workers identified three ligands that selectively stabilized RNA G-quadruplexes over those in DNA. Through favoring G-quadruplex formation, these ligands were hypothesized to reduce dipeptide translation and were shown to significantly reduce RNA foci in patient-derived cells.⁵⁴ One molecule, DB1273 (Figure 12A), significantly reduced RNA foci (Figure 12B) and dipeptide levels both in cells and *Drosophila* models, where it also increased fly survival.⁵⁴

The riboflavin riboswitch

As discussed above, riboswitches are found in bacterial systems and are a naturally-occurring example of RNA-small molecule recognition. As a result, targeting these systems may facilitate antimicrobial development. Merck identified Ribocil, which binds to the bacterial riboflavin riboswitch, thereby deactivating riboflavin synthesis genes and resulting in bacterial death.¹³ The authors evaluated binding affinities for two stereoisomers of Ribocil, termed Ribocil-A and Ribocil-B. Ribocil-B (Figure 13A) was found to have significantly a higher binding affinity than Ribocil-A, with a K_d of 6.6 nM and over 10,000 nM, respectively. These changes in binding affinity were also reflected in biological assays, where much lower concentrations of Ribocil-B were needed to inhibit bacterial growth as compared to Ribocil-A.

Mechanistically, Ribocil-B functions as a riboflavin mimic, causing an equivalent conformational change in the riboswitch to deactivate the riboflavin synthesis pathway by occupying the same binding site. Interestingly, Ribocil is not structurally similar to riboflavin, yet is able to occupy this site through similar noncovalent interactions, including stacking interactions and hydrogen bonds, that mediate this recognition (Figure 13B).¹³ Although Ribocil induced rapid development of resistant strains through mutation of the riboflavin aptamer, another derivative, Ribocil-C, successfully reduced systemic *E. coli* infection in a mouse model.¹³ Combined, insights from these studies showcase RNA's ability to discriminate between minute changes in small molecule structure, in this case stereochemistry, despite more limited chemical functionality. This work also highlights the potential for targeting this biomolecule with small molecules that are structurally distinct from the natural ligand.

RNA splicing in spinal muscular atrophy

As RNAs can have multiple splice isoforms, one avenue within RNA-based drug design is to identify molecules that alter splicing to mitigate disease states.⁵⁵ For example, spinal muscular atrophy (SMA) is the result of loss-of-function in a gene called survival of motor neuron 1 (SMN1), resulting in loss of muscle function and early death.³⁶ Humans have a paralog gene called SMN2, but the resulting transcript from this gene lacks the seventh exon, critical for protein stability, and thus does not rescue loss of SMN1 function. Drugs that could favor Exon 7 inclusion in SMN2 are a promising treatment option (Figure 14A).³⁶

Two promising molecules, NVS-SM1 by Novartis³⁶ (Figure 14B, left) and SMN-C5 by Roche³⁷ (Figure 14B, right), were identified through phenotypic screens to favor Exon 7 inclusion through interacting with the RNA components of the splicing machinery. In particular, SMN-C5 was found to interact with the U1 snRNP complex, and an NMR structure for the apo- and SMN-C5-bound U1 snRNA and Exon 7 splice site was solved by Campagne, Allain, and co-workers via NMR (Figure 14C).³⁷ It was found that SMN-C5 binds the major groove of the splice site, and electrostatic interactions between SMN-C5 and a U1 snRNA phosphate group facilitates binding. Stabilization of an adenine residue via hydrogen bonding maintains the adenine within the duplex, as compared to being flipped out as seen in the apo structure. As a result, this increased stabilization enhances Exon 7 stability and promotes splicing. In addition, this splice site structure is unique from other splice sites, thereby facilitating selectivity.³⁷

Excitingly, an SMN-C5 derivative called risdiplam (Evrysdi™) by Roche/Genentech in collaboration with PTC Therapeutics recently received FDA approval for the treatment of SMA.⁵⁶ Risdiplam improved patient outcomes in both infantile and later-onset SMA in two separate clinical trials and represents the first orally-available drug for SMA.⁵⁶ The approval of risdiplam showcases the promise of small molecule targeting of RNA and higher-order RNA assemblies, and also how structural biology studies can give greater insight into the mode of action of these drugs. It is certain that the success of risdiplam will spur interest into RNA targeting and modulation of splicing in a variety of diseases.

Outlook

The field of RNA-targeted drug discovery is rapidly progressing. Daily discoveries of RNA functions and structures, as well as privileged small molecule scaffolds, are accelerating the development of RNA-based drugs into the clinic. Small molecule approaches have great potential for therapeutic development as well as functioning in complement with ASOs and siRNAs for laboratory research. Identifying small molecule-RNA recognition principles, from both the RNA perspective and the small molecule perspective, will continue to facilitate such development. The importance of understanding fundamental recognition for these processes is elegantly illustrated with risdiplam, where the elucidation of its mechanism of action on a molecular level serves to explain its clinical success. In a larger sense, it is clear that drug development is moving past a protein-only mindset and that RNA's rich structure and function make it an excellent therapeutic target to be further explored.

Several additional lines of research can be envisioned to progress the field. For example, more work is needed to parse the sequence and structure dependencies of small molecule-RNA interactions. The possibility of using molecules to selectively interact with unique RNA structure or sequences of interest, such as protein-binding motifs, is an exciting avenue for future work. As seen in the EV-71 example above, this work may also include searching for modulators of RNA-protein interactions via allosteric regulation as opposed to competition for a given binding site. In addition, research is needed to understand to what extent currently approved protein-targeting

therapeutics may bind RNA, either as part of their mechanism of action or as part of off-target effects. Similarly, natural products are currently underexplored for RNA recognition, and the often complex structures of many natural products may demonstrate novel selectivity or biological activity.

Given the unique chemical properties and modes of recognition that appear to be enriched in RNA-binding small molecules, it is possible that these structures may impact cell uptake and bioavailability. Perhaps, for example, certain transporters preferentially interact with RNA-binding small molecules on account of these unique structures or propensity for cationic chemotypes. Future work should explore these uptake mechanisms to aid in generating molecules with high bioavailability. This may be particularly relevant for targeting RNA that are enriched in certain cellular compartments.

Finally, it is most clear that continued basic understanding of RNA biology will be critical towards developing these novel therapeutics. RNA has been found to be core to a wide variety of cellular processes. Every day, new discoveries in chemistry and RNA biology are overturning what we thought we knew and allowing for ever-growing progress.² One particularly exciting area of expanding research is the study of RNA's role in liquid-liquid phase separation that drives the formation of membraneless cellular compartments, several of which have important disease implications.⁵⁷ The possibility of modulating this dynamic cellular process with small molecules is a fascinating direction that will require cross-talk between the phase separation and RNA-small molecule targeting field.

If even just the last few years are anything to guide us, it is clear that the future in the RNA-targeting field will bring incredible advances. New research, coming out almost daily, is unlocking the clues of how to capitalize on RNA to change the course of human health. We greatly look forward to the upcoming years in RNA-targeting and all the treatments and promise it will bring.

Conflict of Interests

The authors declare no conflicts of interest.

Acknowledgements

The authors would like to thank the members of the Hargrove Lab for their invaluable feedback on this manuscript. The authors acknowledge financial support from Duke University, the National Science Foundation (CAREER 1750375), the U.S. National Institutes of General Medicine (R35 GM124785) and the Prostate Cancer Foundation. JPF was further supported by the Duke University Department of Biochemistry fellowship and Duke University School of Medicine scholarship funding.

References

1. K. D. Warner, C. E. Hajdin and K. M. Weeks, *Nat. Rev. Drug Discov.*, 2018, **17**, 547-558.
2. T. R. Cech and J. A. Steitz, *Cell*, 2014, **157**, 77-94.
3. A. Donlic and A. E. Hargrove, *Wiley Interdiscip. Rev. RNA*, 2018, **9**, e1477.
4. The ENCODE Project Consortium, *Nature*, 2012, **489**, 57-74.
5. M. G. Costales, J. L. Childs-Disney, H. S. Haniff and M. D. Disney, *J. Med. Chem.*, 2020, **63**, 8880-8900.
6. C. A. McHugh, C. K. Chen, A. Chow, C. F. Surka, C. Tran, P. McDonel, A. Pandya-Jones, M. Blanco, C. Burghard, A. Moradian, M. J. Sweredoski, A. A. Shishkin, J. Su, E. S. Lander, S. Hess, K. Plath and M. Guttman, *Nature*, 2015, **521**, 232-236 and references therein.
7. J. A. Brown, D. Bulkley, J. Wang, M. L. Valenstein, T. A. Yario, T. A. Steitz and J. A. Steitz, *Nat. Struct. Mol. Biol.*, 2014, **21**, 633-640 and references therein.
8. A. Donlic, B. S. Morgan, J. L. Xu, A. Liu, C. Roble Jr. and A. E. Hargrove, *Angew. Chem. Int. Ed. Engl.*, 2018, **57**, 13242-13247.
9. F. A. Abulwerdi, W. Xu, A. A. Ageeli, M. J. Yonkunas, G. Arun, H. Nam, J. S. Schneekloth Jr., T. K. Dayie, D. Spector, N. Baird and S. F. J. Le Grice, *ACS Chem. Biol.*, 2019, **14**, 223-235.
10. S. G. Rzuczek, L. A. Colgan, Y. Nakai, M. D. Cameron, D. Furling, R. Yasuda and M. D. Disney, *Nat. Chem. Biol.*, 2017, **13**, 188-193 and references therein.
11. T. Hermann, *Wiley Interdiscip. Rev. RNA*, 2016, **7**, 726-743.
12. M. Mandal and R. R. Breaker, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 451-463.
13. J. A. Howe, H. Wang, T. O. Fischmann, C. J. Balibar, L. Xiao, A. M. Galgoci, J. C. Malinverni, T. Mayhood, A. Villafania, A. Nahvi, N. Murgolo, C. M. Barbieri, P. A. Mann, D. Carr, E. Xia, P. Zuck, D. Riley, R. E. Painter, S. S. Walker, B. Sherborne, R. de Jesus, W. Pan, M. A. Plotkin, J. Wu, D. Rindgen, J. Cummings, C. G. Garlisi, R. Zhang, P. R. Sheth, C. J. Gill, H. Tang and T. Roemer, *Nature*, 2015, **526**, 672-677.
14. B. S. Morgan and A. E. Hargrove, in *Synthetic Receptors for Biomolecules: Design Principles and Applications*, The Royal Society of Chemistry, 2015, 253-325.
15. A. Blythe, A. Fox and C. Bond, *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 2015, **1859**, 46-58.
16. S. E. Butcher and A. M. Pyle, *Acc. Chem. Res.*, 2011, **44**, 1302-1311.
17. J. L. Sussman, S. R. Holbrook, R. W. Warrant, G. M. Church and S. H. Kim, *J. Mol. Biol.*, 1978, **123**, 607-630.
18. G. W. Collie, S. M. Haider, S. Neidle and G. N. Parkinson, *Nucleic Acids Res.*, 2010, **38**, 5569-5580.
19. X. Wang, K. J. Goodrich, A. R. Gooding, H. Naem, S. Archer, R. D. Paucek, D. T. Youmans, T. R. Cech and C. Davidovich, *Mol. Cell*, 2017, **65**, 1056-1067 and references therein.
20. X. Zhang, C. Yan, J. Hang, L. I. Finci, J. Lei and Y. Shi, *Cell*, 2017, **169**, 918-929.
21. M. Selmer, C. M. Dunham, F. V. Murphy IV, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir and V. Ramakrishnan, *Science*, 2006, **313**, 1935-1942.
22. D. P. Arya, *Aminoglycoside antibiotics: From chemical biology to drug discovery*, John Wiley and Sons, Hoboken, New Jersey, 2007.

23. C. Roost, S. R. Lynch, P. J. Batista, K. Qu, H. Y. Chang and E. T. Kool, *J. Am. Chem. Soc.*, 2015, **137**, 2107-2115.
24. N. Liu, Q. Dai, G. Zheng, C. He, M. Parisien and T. Pan, *Nature*, 2015, **518**, 560-564.
25. L. R. Ganser, M. L. Kelly, D. Herschlag and H. M. Al-Hashimi, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 474-489.
26. E. A. Dethoff, K. Petzold, J. Chugh, A. Casiano-Negroni and H. M. Al-Hashimi, *Nature*, 2012, **491**, 724-728.
27. P. Belmont, J.-F. Constant and M. Demeunynck, *Chem. Soc. Rev.*, 2001, **30**, 70-81.
28. S.-J. Chen and K. A. Dill, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 646-651.
29. G. Padroni, N. N. Patwardhan, M. Schapira and A. E. Hargrove, *RSC Med. Chem.*, 2020, **11**, 802-813.
30. G. W. Collie, S. Sparapani, G. N. Parkinson and S. Neidle, *J. Am. Chem. Soc.*, 2011, **133**, 2721-2728.
31. Q. Vicens and E. Westhof, *Structure*, 2001, **9**, 647-658.
32. W. M. Hewitt, D. R. Calabrese and J. S. Schneekloth Jr., *Bioorg. Med. Chem.*, 2019, **27**, 2253-2260.
33. J. Sztuba-Solinska, S. R. Shenoy, P. Gareiss, L. R. H. Krumpke, S. F. J. Le Grice, B. R. O'Keefe and J. S. Schneekloth Jr., *J. Am. Chem. Soc.*, 2014, **136**, 8402-8410.
34. S. L. Wicks and A. E. Hargrove, *Methods*, 2019, **167**, 3-14.
35. A. U. Juru, N. N. Patwardhan and A. E. Hargrove, *ACS Chem. Biol.*, 2019, **14**, 824-838.
36. J. Palacino, S. E. Swalley, C. Song, A. K. Cheung, L. Shu, X. Zhang, M. Van Hoosear, Y. Shin, D. N. Chin, C. G. Keller, M. Beibel, N. A. Renaud, T. M. Smith, M. Salcius, X. Shi, M. Hild, R. Servais, M. Jain, L. Deng, C. Bullock, M. McLellan, S. Schuierer, L. Murphy, M. J. Blommers, C. Blaustein, F. Berenshteyn, A. Lacoste, J. R. Thomas, G. Roma, G. A. Michaud, B. S. Tseng, J. A. Porter, V. E. Myer, J. A. Tallarico, L. G. Hamann, D. Curtis, M. C. Fishman, W. F. Dietrich, N. A. Dales and R. Sivasankaran, *Nat. Chem. Biol.*, 2015, **11**, 511-517.
37. S. Campagne, S. Boigner, S. Rüdissler, A. Moursy, L. Gillioz, A. Knörlein, J. Hall, H. Ratni, A. Cléry and F. H. Allain, *Nat. Chem. Biol.*, 2019, **15**, 1191-1198 and references therein.
38. A. Brown and S. Shao, *Curr. Opin. Struct. Biol.*, 2018, **52**, 1-7.
39. R. D. Thompson, J. T. Baisden and Q. Zhang, *Methods*, 2019, **167**, 66-77.
40. S. C. Keane, X. Heng, K. Lu, S. Kharytonchyk, V. Ramakrishnan, G. Carter, S. Barton, A. Hosic, A. Florwick, J. Santos, N. C. Bolden, S. McCowin, D. A. Case, B. Johnson, M. Salemi, A. Telesnitsky and M. F. Summers, *Science*, 2015, **348**, 917-921.
41. A. C. Stelzer, A. T. Frank, J. D. Kratz, M. D. Swanson, M. J. Gonzalez-Hernandez, J. Lee, I. Andricioaei, D. M. Markovitz and H. M. Al-Hashimi, *Nat. Chem. Biol.*, 2011, **7**, 553-559.
42. N. A. Siegfried, S. Busan, G. M. Rice, J. A. Nelson and K. M. Weeks, *Nat. Methods*, 2014, **11**, 959-965 and references therein.
43. L. R. Ganser, M. L. Kelly, N. N. Patwardhan, A. E. Hargrove and H. M. Al-Hashimi, *J. Mol. Biol.*, 2020, **432**, 1297-1304.
44. B. S. Morgan, J. E. Forte, R. N. Culver, Y. Zhang and A. E. Hargrove, *Angew. Chem. Int. Ed. Engl.*, 2017, **56**, 13498-13502.

45. H. S. Haniff, L. Knerr, X. Liu, G. Crynen, J. Boström, D. Abegg, A. Adibekian, E. Lekah, K. W. Wang, M. D. Cameron, I. Yildirim, M. Lemurell and M. Disney, *Nat. Chem.*, 2020.
46. N. F. Rizvi, J. P. Santa Maria Jr., A. Nahvi, J. Klappenbach, D. J. Klein, P. J. Curran, M. P. Richards, C. Chamberlin, P. Saradjian, J. Burchard, R. Aguilar, J. T. Lee, P. J. Dandliker, G. F. Smith, P. Kutchukian and E. B. Nickbarg, *SLAS Discov.*, 2019, 384-396.
47. B. S. Morgan, B. G. Sanaba, A. Donlic, D. B. Karloff, J. E. Forte, Y. Zhang and A. E. Hargrove, *ACS Chem. Biol.*, 2019, **14**, 2691-2700.
48. M. D. Disney, A. M. Winkelsas, S. P. Velagapudi, M. Southern, M. Fallahi and J. L. Childs-Disney, *ACS Chem. Biol.*, 2016, **11**, 1720-1728.
49. C. Oliver, V. Mallett, R. S. Gendron, V. Reinharz, W. L. Hamilton, N. Moitessier and J. Waldispühl, *Nucleic Acids Res.*, 2020, **48**, 7690-7699.
50. C. S. Eubanks, J. E. Forte, G. J. Kapral and A. E. Hargrove, *J. Am. Chem. Soc.*, 2017, **139**, 409-416.
51. A. Donlic, M. Zafferani, G. Padroni, M. Puri and A. E. Hargrove, *Nucleic Acids Res.*, 2020, **48**, 7653-7664.
52. J. Davila-Calderon, N. Patwardhan, L. Chiu, A. Sugarman, Z. Cai, S.R. Penutmutchu, M. Li, G. Brewer, A. E. Hargrove and B. S. Tolbert, *Nat. Commun.*, 2020, **11**.
53. A. Bugaut, R. Rodriguez, S. Kumari, S. T. D. Hsu and S. Balasubramanian, *Org. Biomol. Chem.*, 2010, **8**, 2771-2776.
54. R. Simone, R. Balendra, T. G. Moens, E. Preza, K. M. Wilson, A. Heslegrave, N. S. Woodling, T. Niccoli, J. Gilbert-Jaramillo, S. Abdelkarim, E. L. Clayton, M. Clarke, M. T. Konrad, A. J. Nicoll, J. S. Mitchell, A. Calvo, A. Chio, H. Houlden, J. M. Polke, M. A. Ismail, C. E. Stephens, T. Vo, A. A. Farahat, W. D. Wilson, D. W. Boykin, H. Zetterberg, L. Partridge, S. Wray, G. Parkinson, S. Neidle, R. Patani, P. Fratta and A. M. Isaacs, *EMBO Mol. Med.*, 2018, **10**, 22-31 and references therein.
55. M. A. Havens, D. M. Duelli and M. L. Hastings, *Wiley Interdiscip. Rev. RNA*, 2013, **4**, 247-266.
56. FDA.gov (2020) FDA Approves Oral Treatment for Spinal Muscular Atrophy. Retrieved from <https://www.fda.gov/news-events/press-announcements/fda-approves-oral-treatment-spinal-muscular-atrophy>.
57. Y. Shin and C. P. Brangwynne, *Science*, 2017, **357**.

Footnotes

† We note that these canonical base-pairing patterns have been historically referred to as Watson-Crick and that our usage of the term Watson-Crick-Franklin acknowledges the critical yet traditionally underrecognized contributions of Dr. Rosalind Franklin in X-Ray crystallography to solving the structure of DNA. It is our hope to further the acceptance of this term and encourage its broader adaptation by using it in this tutorial.

Figure Captions

Figure 1. Overview of RNA primary and secondary structure and base pairing patterns. (A) Nucleotide building blocks of RNA primary structure. Each consists of a ribose sugar, containing a 2' hydroxyl unique to RNA over DNA; phosphodiester groups, which connect the 5' and 3' hydroxyl groups of the ribose sugars; and adenine, guanine, uracil and cytosine nucleobases, which define the sequence of an RNA molecule. (B) Canonical Watson-Crick-Franklin

hydrogen bonding patterns. Dashed lines represent hydrogen bonds. (C) An A-form RNA duplex secondary structure (PDB ID: 1LNT). Red dots represent metal ions. Yellow dashes indicate electrostatic interactions, including hydrogen bonds. Black dashes indicate stacking interactions. (D) Example of canonical RNA secondary structures as observed in 5'-end of long non-coding RNA (lncRNA) B2-SINE. Single and double solid lines represent A-U and G-C Watson-Crick-Franklin base pairs, respectively, while circles represent wobble base pairs. Adapted from reference 15 with permission from Elsevier B.V., Copyright 2015. (E) Higher-order base triple interactions. (F) Non-canonical wobble interactions. (G) Purine-purine interactions, including imino and sheared hydrogen bonding patterns.

Figure 2. Examples of RNA tertiary structures. (A) Yeast phenylalanine t-RNA (PDB ID: 6TNA). The individual helices arising from the multi-way junction (arrow) that form the cloverleaf-shaped secondary structure (left) were found to stack on each other to determine the L-shaped three-dimensional fold (right). Adapted from reference 16 with permission from American Chemical Society, Copyright 2011. (B) Example of a kissing loop tertiary interaction formed between substrate and catalytic domain stem-loops of the *Neurospora* Varkud satellite ribozyme (PDB ID: 2MIO). Dashed lines indicate hydrogen bonding. Wobble and non-WCF base pairs are shown as unfilled and filled circles, respectively. Adapted from reference 15 with permission from Elsevier B.V., Copyright 2015. (C) Example of a pseudoknot tertiary interaction in the turnip yellow mosaic virus RNA (PDB ID: 1A60). Non-WCF base pair is shown as a filled circle. Adapted from reference 15 with permission from Elsevier B.V., Copyright 2015. (D) 3'-triple helix in lncRNA MALAT1 (PDB ID: 4PLX). Uridine-rich stem loop (green) form WCF (simple lines) and Hoogsteen (dot and square lines) base pairs with the A-rich tail (purple). Adapted from reference 7 with permission from Springer Nature, Copyright 2014. (E) G-quartets are stabilized by potassium cations through interactions with the guanine's oxygens (left), and stack on top of each other to form G-quadruplex structures (right), as seen in the human telomeric (TERRA) RNA (PDB ID: 3IBK). Dashed lines represent hydrogen bonds, dots represent lone pairs of electrons, and purple circles represent the potassium ion. Adapted from reference 16 with permission from American Chemical Society, Copyright 2011.

Figure 3. Examples of RNA dynamic regulation and associated functional consequences. (A) N-methyladenosine (m^6A) modification disfavors WCF base pairing (top), promoting single-stranded hairpin formation in MALAT1 and consequent enhanced binding of the heterogeneous nuclear ribonucleoprotein C (hnRNP C, bottom). Adapted from reference 24 with permission from Springer Nature, Copyright 2015. (B) The ground state (GS) of HIV-1 TAR RNA is more strongly bound by the Cyclin T1-Tat protein complex than the less energetically stable excited state. Adapted from reference 26 with permission from Springer Nature, Copyright 2012.

Figure 4. Examples of common noncovalent interactions in RNA-small molecule recognition. Adapted from reference 29 with permission from The Royal Society of Chemistry, Copyright 2020. (A) Binding of two acridine-based ligands to the Telomeric Repeat-Containing RNA (TERRA) via stacking interactions (PDB ID: 3MIJ). Ball denotes a K^+ ion. (B) Hydrogen bonding interactions (dashed lines) between paromomycin and HIV-1 Dimerization Initiation Site (DIS) (PDB ID: 3C44).

Figure 5. Common screening methods utilized for detecting RNA-small molecule binding. (A) Fluorescence Indicator Displacement (FID) assays detect changes in fluorescence intensity of an indicator as a result of small molecule binding. This method enables detection of binding events if a single small molecule concentration is used, and relative affinities or Competitive Displacement (CD_{50}) values if multiple concentrations are used. Adapted from reference 34 with permission from Elsevier B.V., Copyright 2019. (B) Small Molecule Microarrays (SMM) enable detection of binding as a result of detecting fluorescent wells upon irradiating a microarray plate. The plate contains immobilized small molecule ligands and is incubated with a fluorescently labeled RNA construct prior to irradiation. (C) Isothermal Titration Calorimetry (ITC) functions through directly measuring changes in heat as a result of small molecule (titrant) binding to the RNA construct (sample cell) relative to buffer (reference cell). As such, this method enables determination of binding affinity (K_d), stoichiometry, as well as Gibbs free energy (ΔG), enthalpy (ΔH), and entropy (ΔS) values. Adapted from reference 35 with permission from American Chemical Society, Copyright 2019. (D) Surface Plasmon Resonance (SPR) directly measures changes in resonance angle or wavelength of a glass surface containing the immobilized RNA of interest upon addition of a small molecule. This method can be utilized to obtain both the association (K_{on}) and dissociation (K_{off}) constants that comprise binding affinities (K_d), providing insights into the kinetics of a binding event. Adapted from reference 35 with permission from American Chemical Society, Copyright 2019.

Figure 6. Chemical probing for interrogation of RNA structure. A) Mechanism for covalent modification by SHAPE reagent 1-methyl-7-nitroisatoic anhydride (1M7, top) and dimethylsulfate (DMS, bottom). In SHAPE probing, deprotonation of the 2'-hydroxyl by a general base facilitates its nucleophilic attack of the anhydride moiety, resulting in the ring opening of 1M7. While this attack is an equilibrium process, the subsequent decarboxylation, generating the final adduct as depicted, is irreversible. In DMS probing, the nitrogen lone pair on A (depicted) or C attack the electrophilic methoxy position, resulting in irreversible covalent modification of the base. B) Workflow for integration of chemical probing information into a structure model. Reagent modifications are incorporated as mutations during reverse transcription in the presence of manganese, and mutations are detected via sequencing. Reactivities for each nucleotide are incorporated as restraints in a structure model. Adapted from reference 42 with permission from Springer Nature, Copyright 2014. C) Chemical probing with HIV TAR following small molecule treatment reveals changes in reactivity for small molecule binding site. Larger scale changes in secondary structure were not observed. Scale shows changes in reactivity between control and small molecule treatment. PBS: primer binding site. Adapted from reference 33 with permission American Chemical Society, Copyright 2014.

Figure 7. Structural dynamics of HIV-1 TAR are modulated with small molecules. A) TAR oscillates between a ground state (GS) and excited state (ES) characterized by base flipping and angular rearrangement of the helix. Several ligands, including argininamide and DMA-169, were shown to alter excited state accessibility, whereas other ligands such as neomycin were not. Adapted from reference 43 with permission from Elsevier, Copyright 2020. B) The structures of argininamide, DMA-169, and neomycin.

Figure 8. Approaches towards discovering privileged properties and chemotypes of RNA-binding small molecules. (A) Comparison of 20 physicochemical properties of bioactive, RNA-binding small molecules (R-BIND) and FDA ligands (mostly protein-targeted). R-BIND members were found to have increased nitrogen atom count, decreased number of sp^3 -hybridized carbons, among other properties. Statistically significant differences determined by Mann Whitney U test are indicated by P values of < 0.05 in light blue and highly significant differences as < 0.001 in dark blue. MW = Molecular Weight, HBA = Hydrogen Bond Acceptors, HBD = Hydrogen Bond Donors, LogP = n-Octanol/Water Partition Coefficient, RotB = Number of Rotatable Bonds, tPSA = Topological Polar Surface Area, logD = n-Octanol/Water Partition Coefficient, N = Number of Nitrogen Atoms, O = Number of Oxygen Atoms, Rings = Number of Rings, ArRings = Number of Aromatic Rings, HetRings = Number of Heteroatom-Containing Rings, SysRings = Number of Ring Systems, SysRR = Ring Complexity, Fsp³ = Fraction of sp^3 -Hybridized Carbons, nStereo = Number of Sterocenters, ASA = Accessible Surface Area, RelPSA = Relative Polar Surface Area, TC = Total Charge, VWSA = Van der Waals Surface Area. Adapted from reference 44 with permission from John Wiley & Sons, Copyright 2017. (B) Nearest Neighbor algorithm utilized to select R-BIND-like molecules as developed by Hargrove and co-workers. Each cheminformatic parameter from (A) is defined as a “dimension”, quantified along its dimension, and the Euclidian distance (smallest distance in space) between each library member is measured and averaged (purple line). The user's input molecule is considered “R-BIND-like” (turquoise dots) or enriched in RNA-binding properties if it is found to be within the measured average distance to an R-BIND member (blue dots). Adapted from reference 47 with permission from American Chemical Society, Copyright 2019. (C) Summary of infoRNA platform utilized to select potential binders of an RNA target. By inputting an RNA sequence of interest, infoRNA is mined for similar RNA motifs to output lead small molecules that can be tested and optimized for specific targeting. Adapted from reference 48 with permission from American Chemical Society, Copyright 2016. (D) Examples of privileged chemotypes for binding of RNA hairpin or bulge secondary structures as discovered by infoRNA mining. Adapted from reference 48 with permission from American Chemical Society, Copyright 2016.

Figure 9. Small molecule targeting of the MALAT1 triple helix. A) Structures of DPFp8 and SM5, which selectively bind the MALAT1 triple helix over controls. B) Treatment of breast cancer organoids with SM5 significantly reduces organoid size (upper panel and lower panel differ by magnification). Adapted from reference 9 with permission from American Chemical Society, Copyright 2019. C) SM5 and SM16 (not shown) significantly reduce branching morphogenesis of breast cancer organoids. Adapted from reference 9 with permission from American Chemical Society, Copyright 2019. ASO: antisense oligonucleotide.

Figure 10: Targeting CUG repeat expansions with a multivalent ligand in DM1. A) Structure of 2H-K4NMeS. B) 2H-K4NMeS treatment reduces CUG:MBNL1 foci. RNA was detected with Fluorescence *In Situ* Hybridization (FISH), where fluorescently-labeled oligonucleotides complementary to an RNA of interest are incubated in cells to show their subcellular localization. Nuclei are labeled with DAPI (4',6-diamidino-2-phenylindole). Adapted from reference 10

with permission from Springer Nature, Copyright 2016. MBNL1 was detected using immunofluorescence, where a fluorescently-labeled antibody for MBNL1 was incubated in cells to show its subcellular localization. White regions in the “Overlay” show where FISH, MBNL1, and DAPI signals overlap. C) Treatment with 2H-K4NMeS shows dose-dependent rescue of MBNL1 splicing defect. Adapted from reference 10 with permission from Springer Nature, Copyright 2016.

Figure 11. Small molecule allosteric modulation of EV71 SLII. A) The structure of DMA-135. B) DMA-135 results in a 77° interhelical shift in SLII as shown by NMR. Adapted from reference 52 with permission from Springer Nature, Copyright 2020. C) A luciferase-based reporter assay shows reduction of FLuc translation from 5' UTR upon DMA-135 treatment as compared to RLuc control. Adapted from reference 52 with permission from Springer Nature, Copyright 2020. D) AUF1 protein was isolated via immunoprecipitation and bound RNA was detected via qPCR. Dose-dependent increase in bound SL2 RNA was observed upon DMA-135 treatment as compared to GAPDH control RNA. N.I. non-immune antibody. Adapted from reference 52 from Springer Nature, 2020 under Creative Commons CC-BY 4.0 License.

Figure 12. Small molecule disruption of foci containing G-quadruplex GGGGCC repeat expansions in ALS/FTD. A) Structure of DB1273. B) DB1273 significantly reduces GGGGCC RNA foci (red) in cellular nuclei (blue). Adapted from reference 54, Copyright 2017, John Wiley and Sons in accordance to Creative Commons CC BY 4.0 License.

Figure 13. Small molecule targeting of the riboflavin riboswitch. A) The structures of Ribocil-B and Riboflavin. B) Crystal structure of Ribocil-B (blue sticks) in complex with riboflavin riboswitch. Ribocil binding is stabilized by stacking interactions with A48 and A85 and hydrogen bonding between the oxygen and A48 and A99. Additional stacking interactions and a methyl hydrogen bond are observed on the other face of the molecule. Adapted from reference 13 with permission from Springer Nature, Copyright 2015.

Figure 14. Small molecule modulation of splicing in SMA. A) Schematic of SMN2 splicing. Inclusion of Exon 7 results in a functional SMN2 protein. B) Structures of two documented SMN2 splicing modulators. C) SMN-C5 (yellow) stabilizes an adenine residue (red) between Exon 7 and a spliceosomal U1 small nuclear ribonucleoprotein (snRNP), facilitating Exon 7 inclusion. SS: splice site. Adapted from reference 37 with permission from Springer Nature, Copyright 2019.

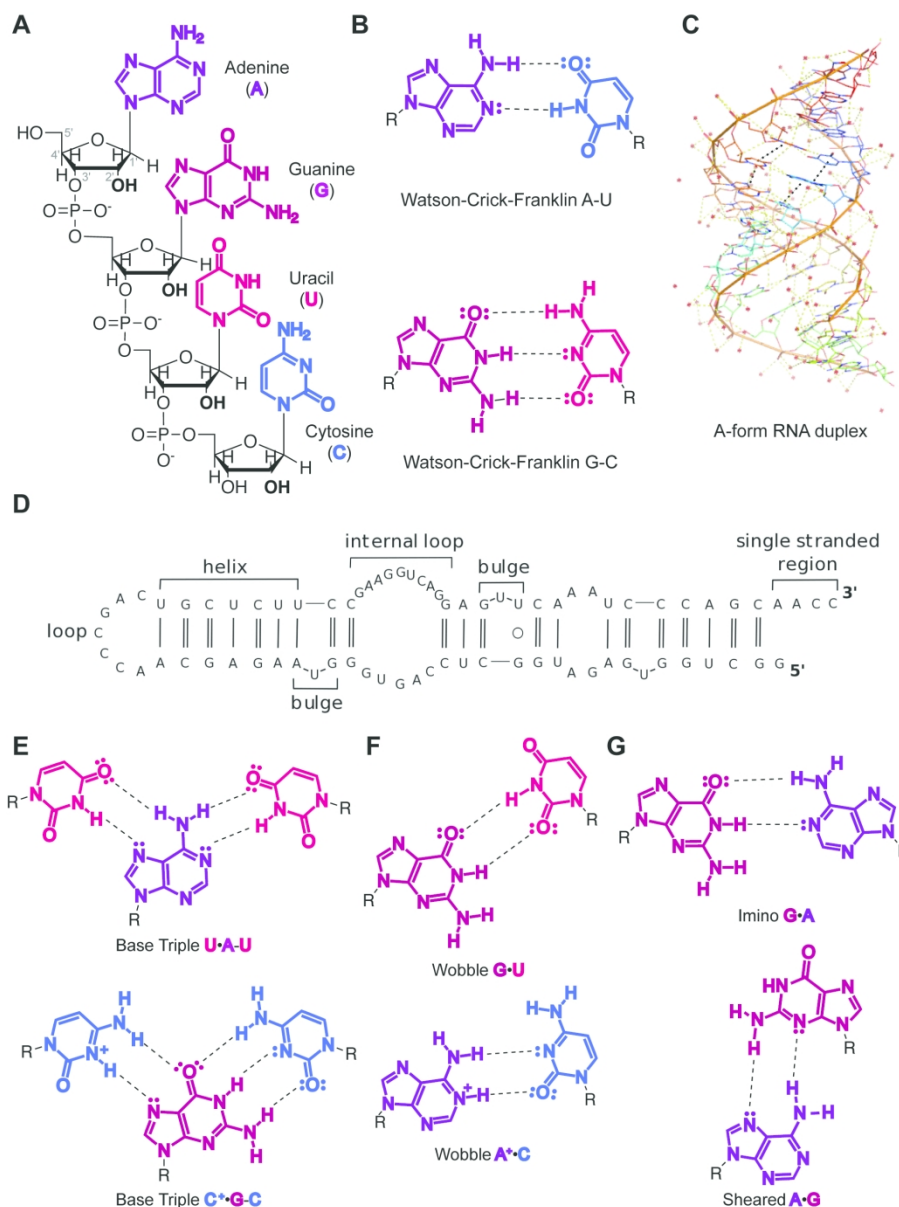


Figure 1. Overview of RNA primary and secondary structure and base pairing patterns. (A) Nucleotide building blocks of RNA primary structure. Each consists of a ribose sugar, containing a 2' hydroxyl unique to RNA over DNA; phosphodiester groups, which connect the 5' and 3' hydroxyl groups of the ribose sugars; and adenine, guanine, uracil and cytosine nucleobases, which define the sequence of an RNA molecule. (B) Canonical Watson-Crick-Franklin hydrogen bonding patterns. Dashed lines represent hydrogen bonds. (C) An A-form RNA duplex secondary structure (PDB ID: 1LNT). Red dots represent metal ions. Yellow dashes indicate electrostatic interactions, including hydrogen bonds. Black dashes indicate stacking interactions. (D) Example of canonical RNA secondary structures as observed in 5'-end of long non-coding RNA (lncRNA) B2-SINE. Single and double solid lines represent A-U and G-C Watson-Crick-Franklin base pairs, respectively, while circles represent wobble base pairs. Adapted from reference 15 with permission from Elsevier B.V., Copyright 2015. (E) Higher-order base triple interactions. (F) Non-canonical wobble patterns. (G) Purine-purine interactions, including imino and sheared hydrogen bonding patterns.

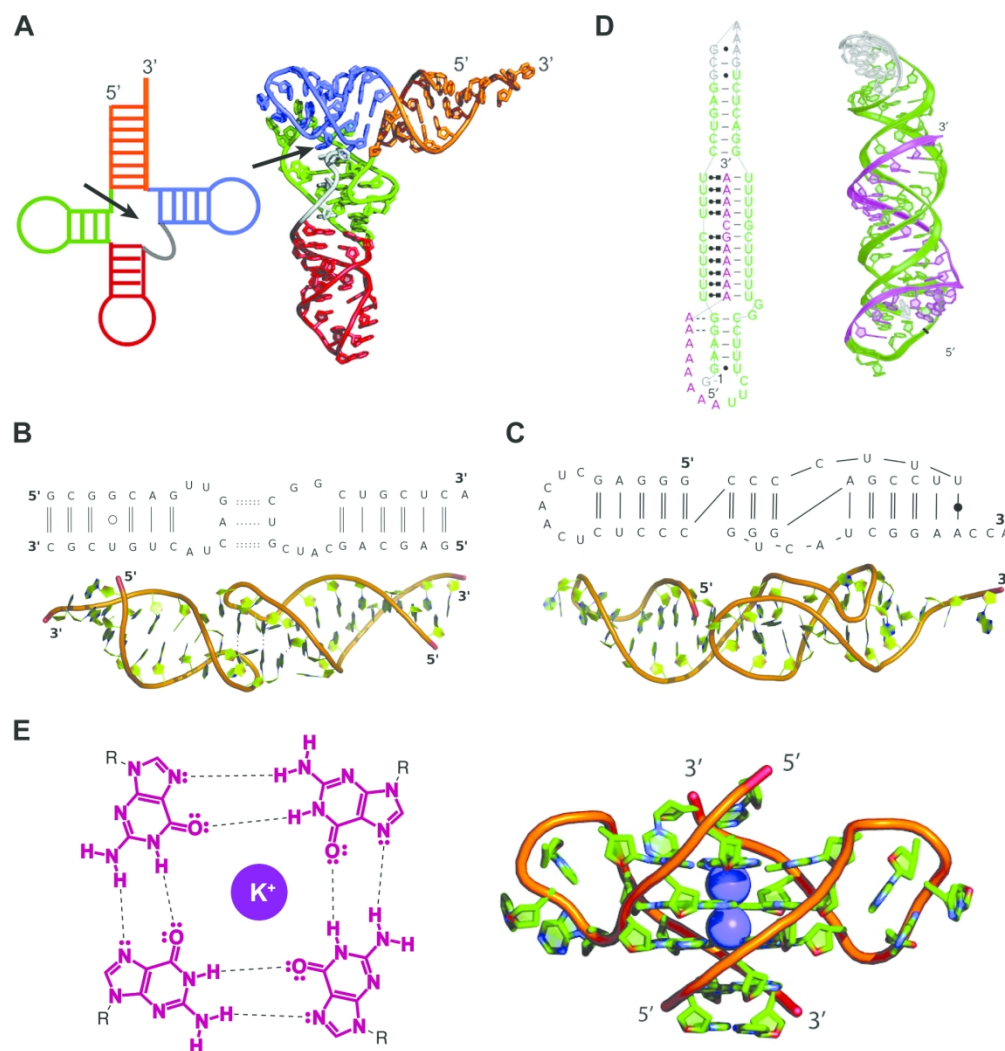


Figure 2. Examples of RNA tertiary structures. (A) Yeast phenylalanine t-RNA (PDB ID: 6TNA). The individual helices arising from the multi-way junction (arrow) that form the cloverleaf-shaped secondary structure (left) were found to stack on each other to determine the L-shaped three-dimensional fold (right). Adapted from reference 16 with permission from American Chemical Society, Copyright 2011. (B) Example of a kissing loop tertiary interaction formed between substrate and catalytic domain stem-loops of the *Neurospora Varkud* satellite ribozyme (PDB ID: 2MIO). Dashed lines indicate hydrogen bonding. Wobble and non-WCF base pairs are shown as unfilled and filled circles, respectively. Adapted from reference 15 with permission from Elsevier B.V., Copyright 2015. (C) Example of a pseudoknot tertiary interaction in the turnip yellow mosaic virus RNA (PDB ID: 1A60). Non-WCF base pair is shown as a filled circle. Adapted from reference 15 with permission from Elsevier B.V., Copyright 2015. (D) 3'-triple helix in lncRNA MALAT1 (PDB ID: 4PLX). Uridine-rich stem loop (green) form WCF (simple lines) and Hoogsteen (dot and square lines) base pairs with the A-rich tail (purple). Adapted from reference 7 with permission from Springer Nature, Copyright 2014. (E) G-quartets are stabilized by potassium cations through interactions with the guanine's oxygens (left), and stack on top of each other to form G-quadruplex structures (right), as seen in the human telomeric (TERRA) RNA (PDB ID: 3IBK). Dashed lines represent hydrogen bonds, dots represent lone pairs of electrons, and purple circles represent the potassium ion. Adapted from reference 16 with permission from American Chemical Society, Copyright 2011.

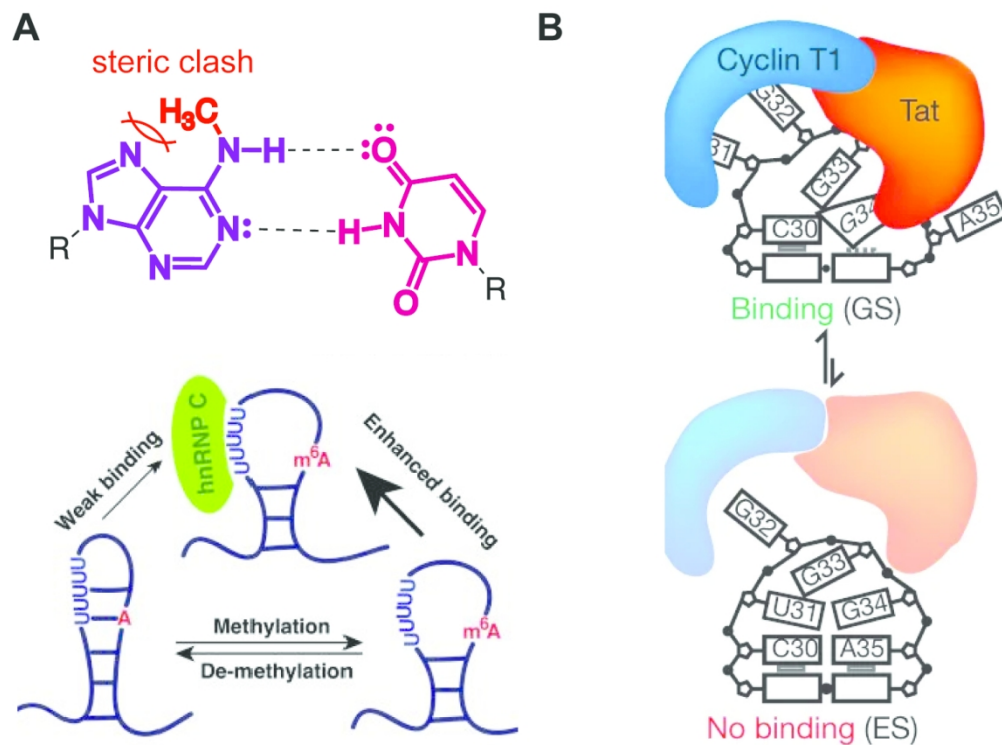


Figure 3. Examples of RNA dynamic regulation and associated functional consequences. (A) N-methyladenosine (m^6A) modification disfavors WCF base pairing (top), promoting single-stranded hairpin formation in MALAT1 and consequent enhanced binding of the heterogeneous nuclear ribonucleoprotein C (hnRNP C, bottom). Adapted from reference 24 with permission from Springer Nature, Copyright 2015. (B) The ground state (GS) of HIV-1 TAR RNA is more strongly bound by the Cyclin T1-Tat protein complex than the less energetically stable excited state. Adapted from reference 26 with permission from Springer Nature, Copyright 2012.

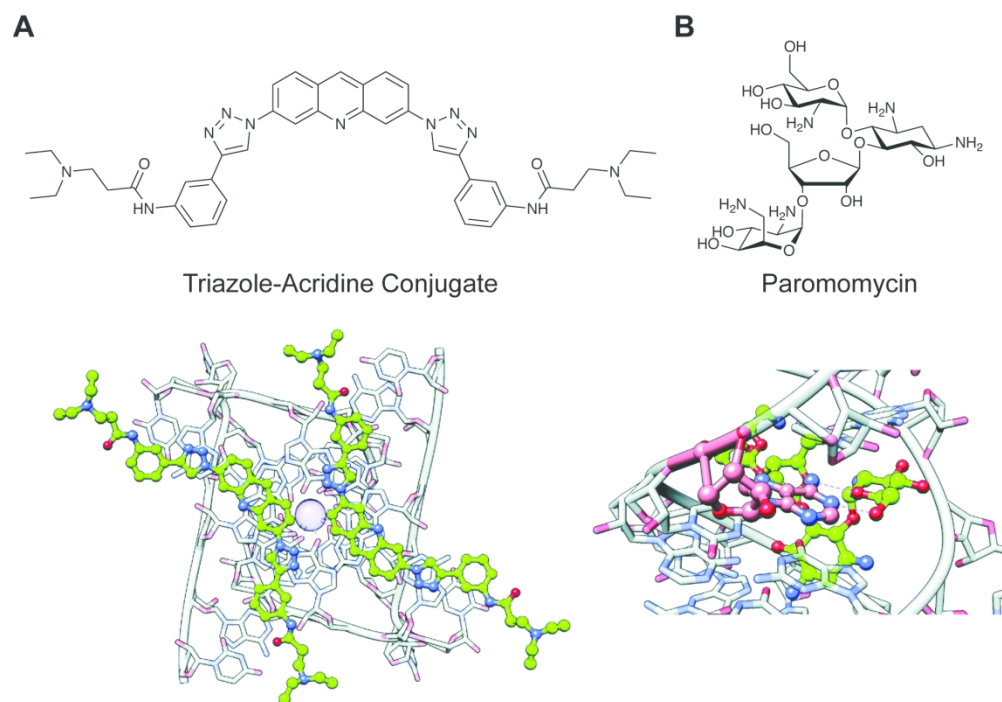


Figure 4. Examples of common noncovalent interactions in RNA-small molecule recognition. Adapted from reference 29 with permission from The Royal Society of Chemistry, Copyright 2020. (A) Binding of two acridine-based ligands to the Telomeric Repeat-Containing RNA (TERRA) via stacking interactions (PDB ID: 3MIJ). Ball denotes a K^+ ion. (B) Hydrogen bonding interactions (dashed lines) between paromomycin and HIV-1 Dimerization Initiation Site (DIS) (PDB ID: 3C44).

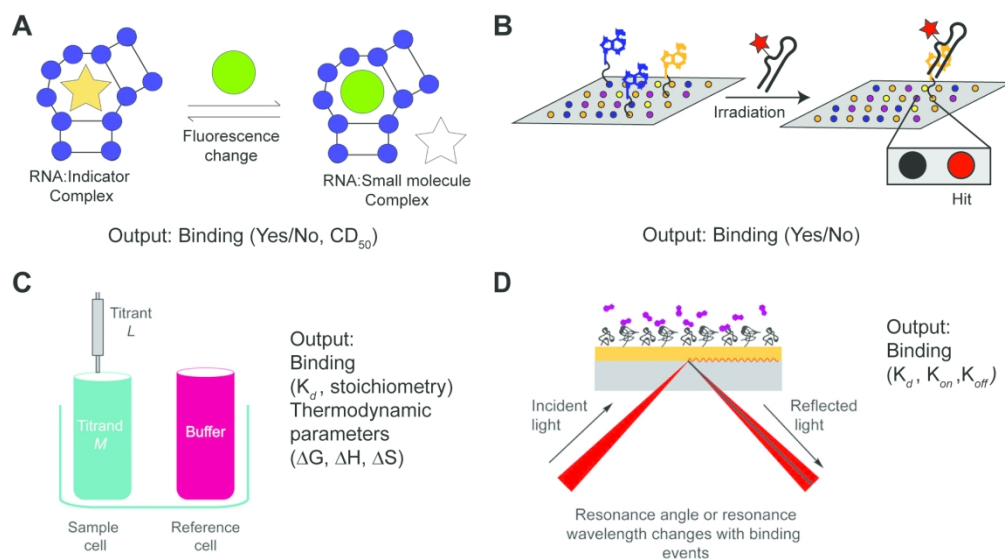


Figure 5. Common screening methods utilized for detecting RNA-small molecule binding. (A) Fluorescence Indicator Displacement (FID) assays detect changes in fluorescence intensity of an indicator as a result of small molecule binding. This method enables detection of binding events if a single small molecule concentration is used, and relative affinities or Competitive Displacement (CD_{50}) values if multiple concentrations are used. Adapted from reference 34 with permission from Elsevier B.V., Copyright 2019. (B) Small Molecule Microarrays (SMM) enable detection of binding as a result of detecting fluorescent wells upon irradiating a microarray plate. The plate contains immobilized small molecule ligands and is incubated with a fluorescently labeled RNA construct prior to irradiation. (C) Isothermal Titration Calorimetry (ITC) functions through directly measuring changes in heat as a result of small molecule (titrant) binding to the RNA construct (sample cell) relative to buffer (reference cell). As such, this method enables determination of binding affinity (K_d), stoichiometry, as well as Gibbs free energy (ΔG), enthalpy (ΔH), and entropy (ΔS) values. Adapted from reference 35 with permission from American Chemical Society, Copyright 2019. (D) Surface Plasmon Resonance (SPR) directly measures changes in resonance angle or wavelength of a glass surface containing the immobilized RNA of interest upon addition of a small molecule. This method can be utilized to obtain both the association (K_{on}) and dissociation (K_{off}) constants that comprise binding affinities (K_d), providing insights into the kinetics of a binding event. Adapted from reference 35 with permission from American Chemical Society, Copyright 2019.

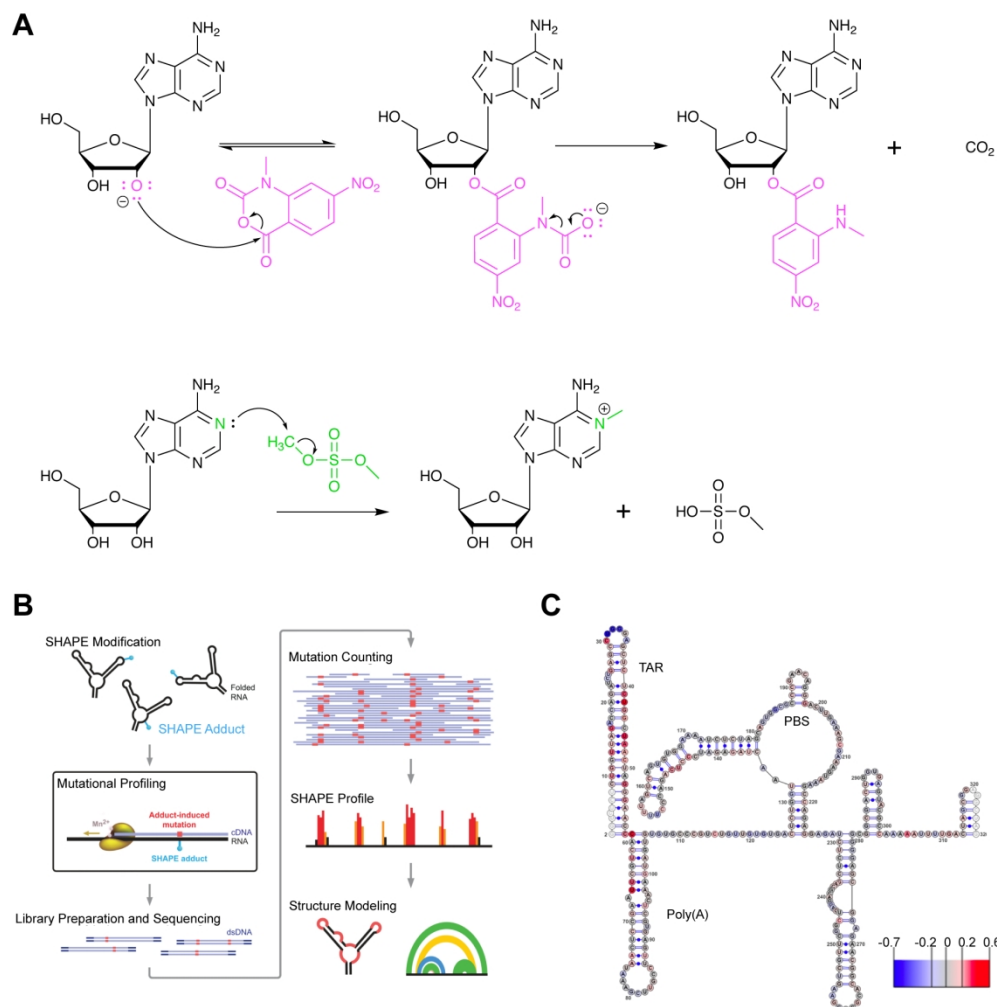


Figure 6. Chemical probing for interrogation of RNA structure. A) Mechanism for covalent modification by SHAPE reagent 1-methyl-7-nitroisoatoic anhydride (1M7, top) and dimethylsulfate (DMS, bottom). In SHAPE probing, deprotonation of the 2'-hydroxyl by a general base facilitates its nucleophilic attack of the anhydride moiety, resulting in the ring opening of 1M7. While this attack is an equilibrium process, the subsequent decarboxylation, generating the final adduct as depicted, is irreversible. In DMS probing, the nitrogen lone pair on A (depicted) or C attack the electrophilic methoxy position, resulting in irreversible covalent modification of the base. B) Workflow for integration of chemical probing information into a structure model. Reagent modifications are incorporated as mutations during reverse transcription in the presence of manganese, and mutations are detected via sequencing. Reactivities for each nucleotide are incorporated as restraints in a structure model. Adapted from reference 42 with permission from Springer Nature, Copyright 2014. C) Chemical probing with HIV TAR following small molecule treatment reveals changes in reactivity for small molecule binding site. Larger scale changes in secondary structure were not observed. Scale shows changes in reactivity between control and small molecule treatment. PBS: primer binding site. Adapted from reference 33 with permission American Chemical Society, Copyright 2014.

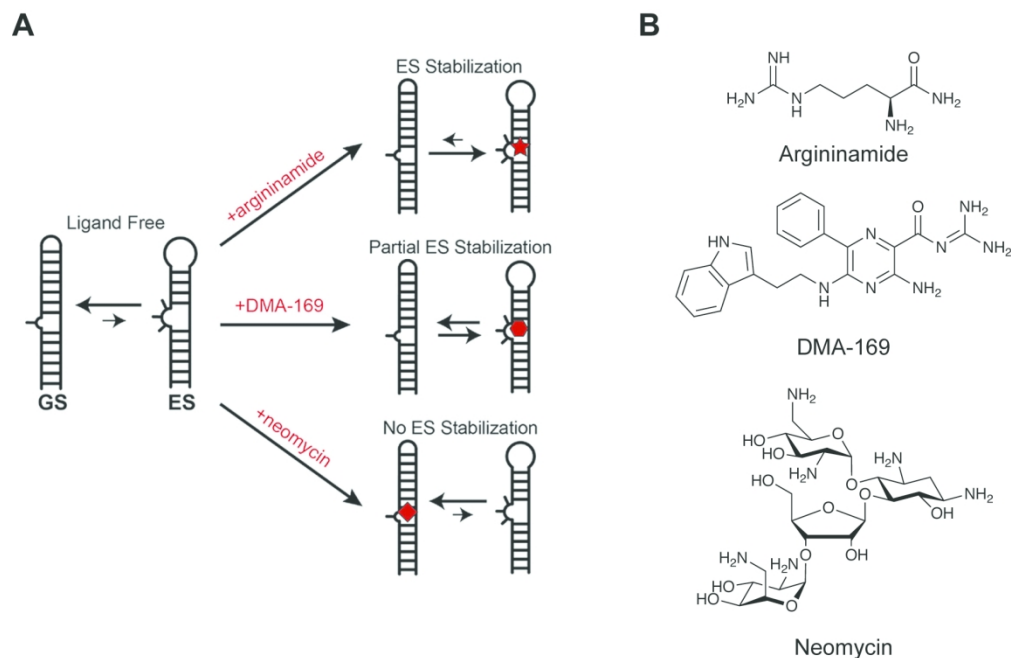


Figure 7. Structural dynamics of HIV-1 TAR are modulated with small molecules. A) TAR oscillates between a ground state (GS) and excited state (ES) characterized by base flipping and angular rearrangement of the helix. Several ligands, including argininamide and DMA-169, were shown to alter excited state accessibility, whereas other ligands such as neomycin were not. Adapted from reference 43 with permission from Elsevier, Copyright 2020. B) The structures of argininamide, DMA-169, and neomycin.

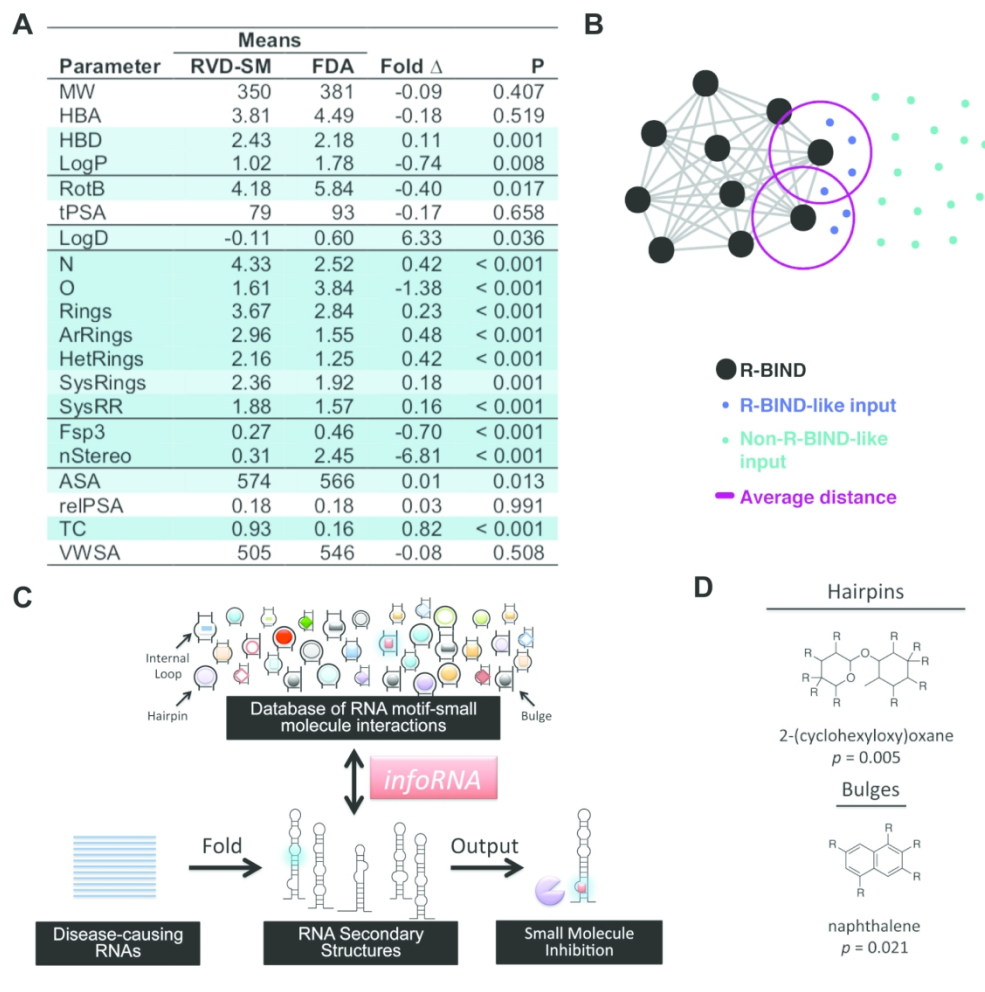


Figure 8. Approaches towards discovering privileged properties and chemotypes of RNA-binding small molecules. (A) Comparison of 20 physicochemical properties of bioactive, RNA-binding small molecules (R-BIND) and FDA ligands (mostly protein-targeted). R-BIND members were found to have increased nitrogen atom count, decreased number of sp³-hybridized carbons, among other properties. Statistically significant differences determined by Mann Whitney U test are indicated by P values of < 0.05 in light blue and highly significant differences as < 0.001 in dark blue. MW = Molecular Weight, HBA = Hydrogen Bond Acceptors, HBD = Hydrogen Bond Donors, LogP = n-Octanol/Water Partition Coefficient, RotB = Number of Rotatable Bonds, tPSA = Topological Polar Surface Area, logD = n-Octanol/Water Partition Coefficient, N = Number of Nitrogen Atoms, O = Number of Oxygen Atoms, Rings = Number of Rings, ArRings = Number of Aromatic Rings, HetRings = Number of Heteroatom-Containing Rings, SysRings = Number of Ring Systems, SysRR = Ring Complexity, Fsp3 = Fraction of sp³-Hybridized Carbons, nStereo = Number of Sterocenters, ASA = Accessible Surface Area, RelPSA = Relative Polar Surface Area, TC = Total Charge, VWSA = Van der Waals Surface Area. Adapted from reference 44 with permission from John Wiley & Sons, Copyright 2017. (B) Nearest Neighbor algorithm utilized to select R-BIND-like molecules as developed by Hargrove and co-workers. Each cheminformatic parameter from (A) is defined as a “dimension”, quantified along its dimension, and the Euclidian distance (smallest distance in space) between each library member is measured and averaged (purple line). The user’s input molecule is considered “R-BIND-like” (turquoise dots) or enriched in RNA-binding properties if it is found to be within the measured average distance to an R-BIND member (blue dots). Adapted from reference 47 with permission from American Chemical Society, Copyright 2019. (C) Summary of infoRNA platform utilized to select potential binders of an RNA target. By inputting an RNA sequence of interest, infoRNA is mined for similar RNA motifs to output lead small molecules that can be tested and optimized for specific targeting. Adapted from reference 48 with permission from American Chemical Society, Copyright 2016. (D) Examples of privileged chemotypes for binding of RNA hairpin or

bulge secondary structures as discovered by infoRNA mining. Adapted from reference 48 with permission from American Chemical Society, Copyright 2016.

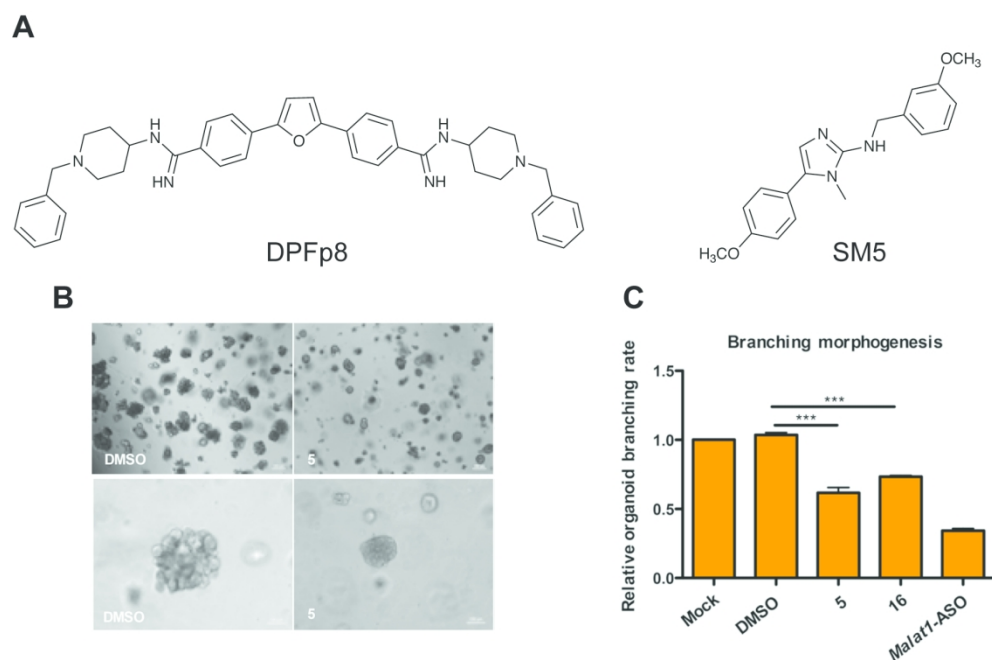


Figure 9. Small molecule targeting of the MALAT1 triple helix. A) Structures of DPFp8 and SM5, which selectively bind the MALAT1 triple helix over controls. B) Treatment of breast cancer organoids with SM5 significantly reduces organoid size (upper panel and lower panel differ by magnification). Adapted from reference 9 with permission from American Chemical Society, Copyright 2019. C) SM5 and SM16 (not shown) significantly reduce branching morphogenesis of breast cancer organoids. Adapted from reference 9 with permission from American Chemical Society, Copyright 2019. ASO: antisense oligonucleotide.

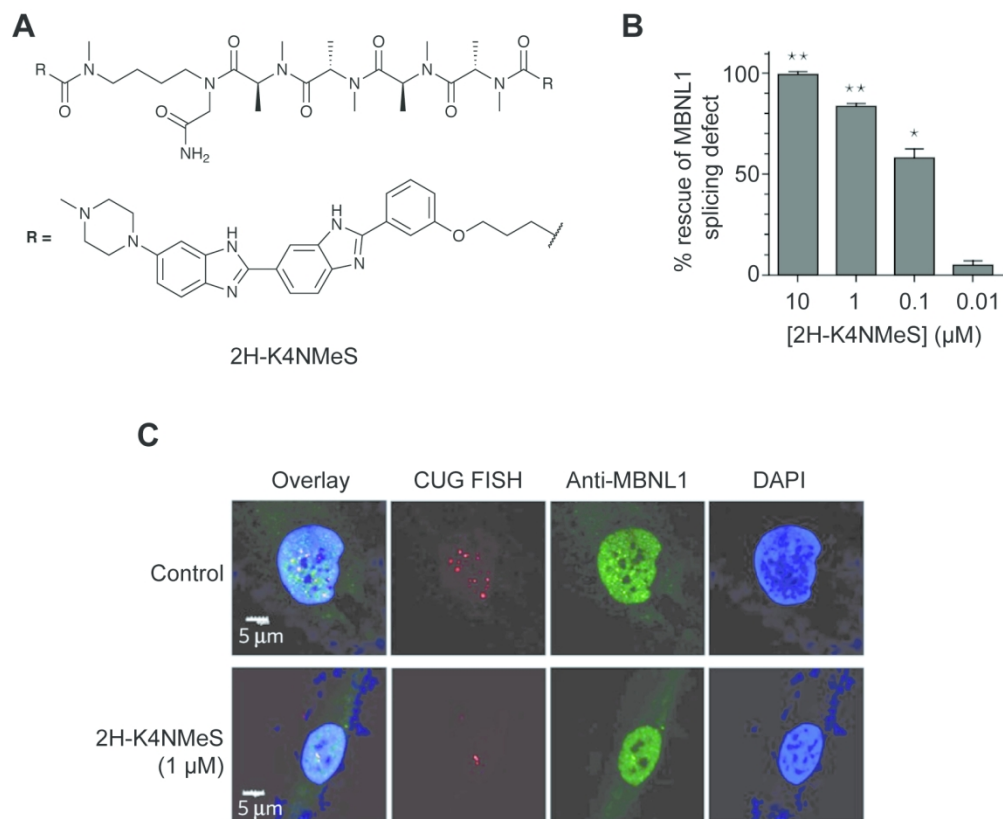


Figure 10: Targeting CUG repeat expansions with a multivalent ligand in DM1. A) Structure of 2H-K4NMeS.

B) 2H-K4NMeS treatment reduces CUG:MBNL1 foci. RNA was detected with Fluorescence In Situ Hybridization (FISH), where fluorescently-labeled oligonucleotides complementary to an RNA of interest are incubated in cells to show their subcellular localization. Nuclei are labeled with DAPI (4',6-diamidino-2-phenylindole). Adapted from reference 10 with permission from Springer Nature, Copyright 2016. MBNL1 was detected using immunofluorescence, where a fluorescently-labeled antibody for MBNL1 was incubated in cells to show its subcellular localization. White regions in the "Overlay" show where FISH, MBNL1, and DAPI signals overlap. C) Treatment with 2H-K4NMeS shows dose-dependent rescue of MBNL1 splicing defect.

Adapted from reference 10 with permission from Springer Nature, Copyright 2016.

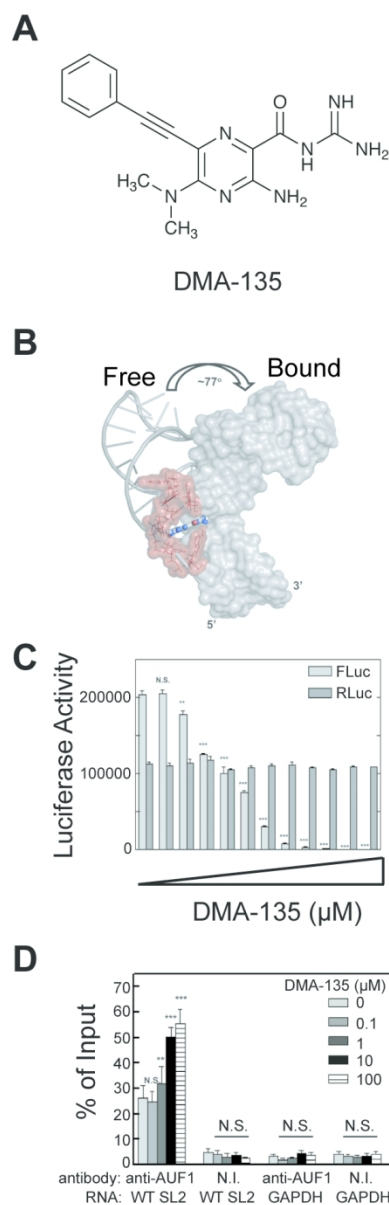


Figure 11. Small molecule allosteric modulation of EV71 SLII. A) The structure of DMA-135. B) DMA-135 results in a 77° interhelical shift in SLII as shown by NMR. Adapted from reference 52 with permission from Springer Nature, Copyright 2020. C) A luciferase-based reporter assay shows reduction of FLuc translation from 5' UTR upon DMA-135 treatment as compared to RLuc control. Adapted from reference 52 with permission from Springer Nature, Copyright 2020. D) AUF1 protein was isolated via immunoprecipitation and bound RNA was detected via qPCR. Dose-dependent increase in bound SL2 RNA was observed upon DMA-135 treatment as compared to GAPDH control RNA. N.I. non-immune antibody. Adapted from reference 52 from Springer Nature, 2020 under Creative Commons CC-BY 4.0 License.

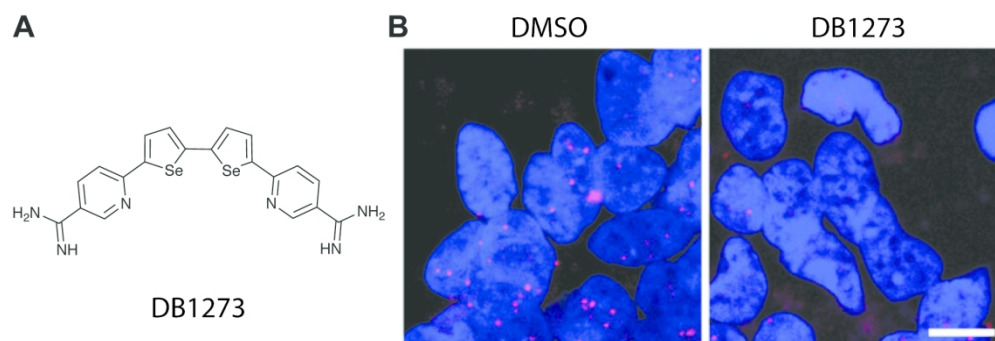


Figure 12. Small molecule disruption of foci containing G-quadruplex GGGGCC repeat expansions in ALS/FTD. A) Structure of DB1273. B) DB1273 significantly reduces GGGGCC RNA foci (red) in cellular nuclei (blue). Adapted from reference 54, Copyright 2017, John Wiley and Sons in accordance to Creative Commons CC BY 4.0 License.

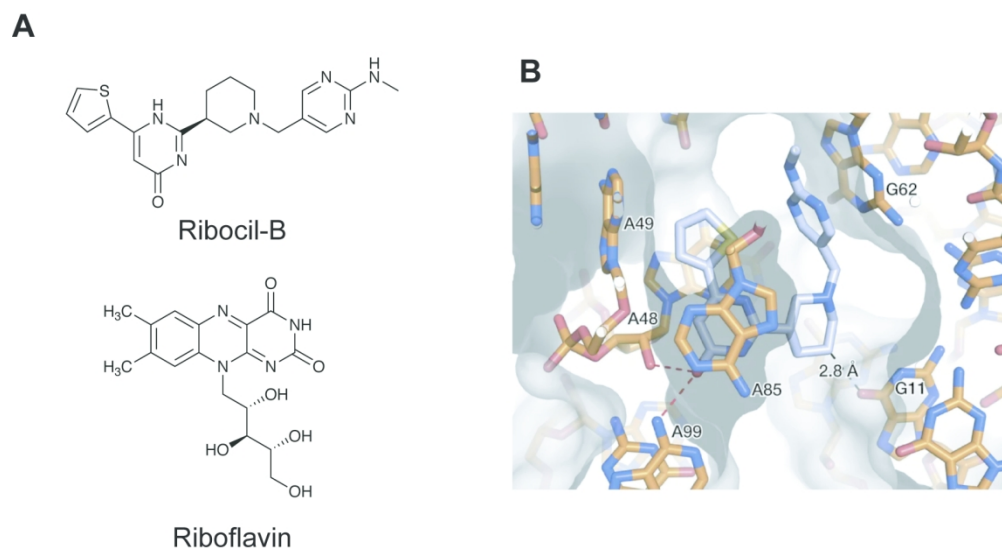


Figure 13. Small molecule targeting of the riboflavin riboswitch. A) The structures of Ribocil-B and Riboflavin. B) Crystal structure of Ribocil-B (blue sticks) in complex with riboflavin riboswitch. Ribocil binding is stabilized by stacking interactions with A48 and A85 and hydrogen bonding between the oxygen and A48 and A99. Additional stacking interactions and a methyl hydrogen bond are observed on the other face of the molecule. Adapted from reference 13 with permission from Springer Nature, Copyright 2015.

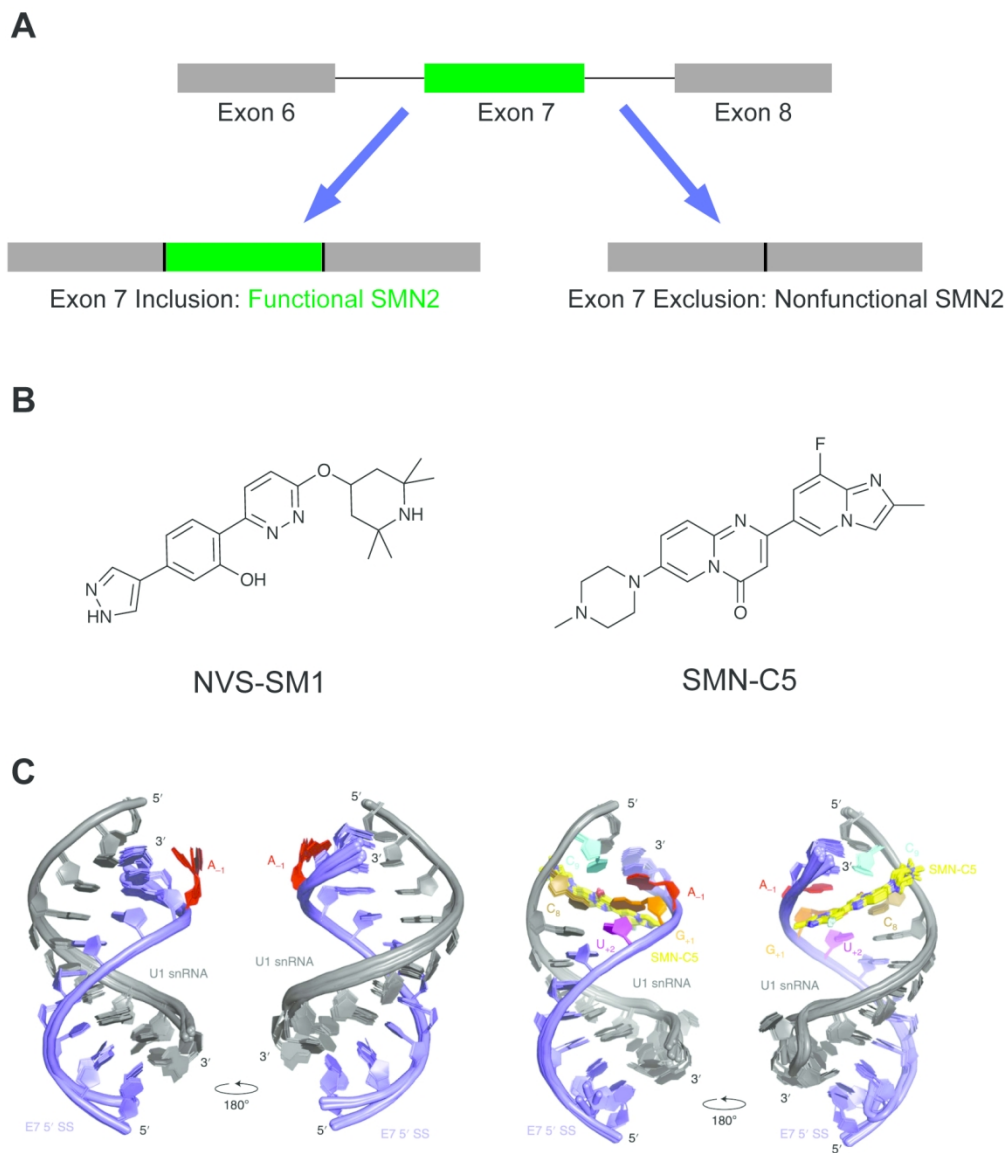
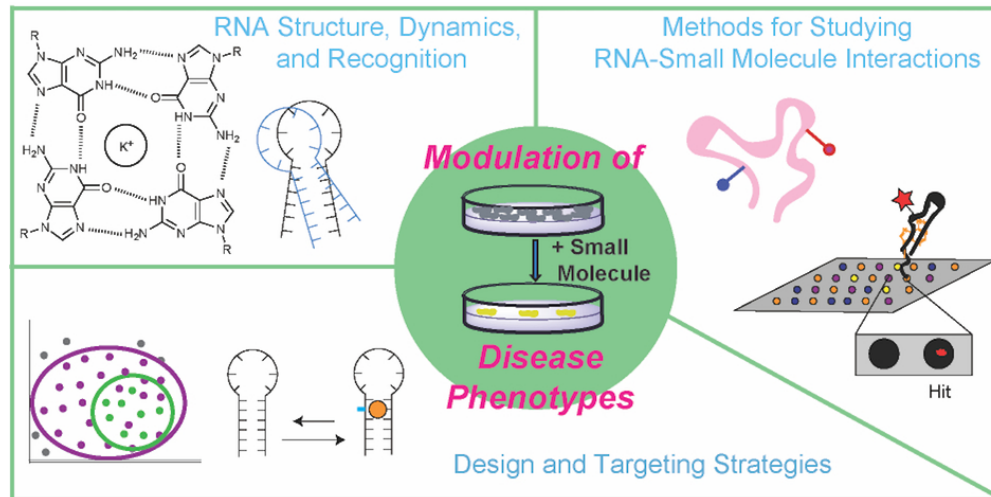


Figure 14. Small molecule modulation of splicing in SMA. A) Schematic of SMN2 splicing. Inclusion of Exon 7 results in a functional SMN2 protein. B) Structures of two documented SMN2 splicing modulators. C) SMN-C5 (yellow) stabilizes an adenine residue (red) between Exon 7 and a spliceosomal U1 small nuclear ribonucleoprotein (snRNP), facilitating Exon 7 inclusion. SS: splice site. Adapted from reference 37 with permission from Springer Nature, Copyright 2019.



80x40mm (300 x 300 DPI)