Published in final edited form as:

ACS Comb Sci. 2018 August 13; 20(8): 482–491. doi:10.1021/acscmbosci.8b00049.

Selective Small Molecule Recognition of RNA Base Pairs

Hafeez S. Haniff, Amanda Graves, and Matthew D. Disney

Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States

Abstract

Many types of RNAs exist in the human transcriptome, yet only the bacterial ribosome has been exploited as a small molecule drug target. Aside from rRNA, other cellular RNAs such as noncoding RNAs have primarily secondary structure and limited tertiary structure. Within these secondary structures of noncanonically paired and unpaired regions, more than 50% are base paired, with most efforts to target these structures focused on looped regions. A void exists in the availability of small molecules capable of targeting RNA base pairs. Using chemoinformatics, an RNA-focused library enriched for nitrogencontaining heterocycles was developed and tested for binding RNA base pairs, leading to the identification of six selective and previously unknown binders. While all binders were derivatives of benzimidazoles, those with expanded aromatic polycycles bound selectively to AU pairs, while those with flexible urea side chains bound selectively to GC pairs. Two of the three selective GC pair binders can distinguish between two different orientations, 5'GG/3'CC and 5'GC/3'CG pairs. Furthermore, all six molecules showed >50-fold selectivity for RNA over DNA. These studies provide foundational knowledge to better exploit RNA as targets for small molecule chemical probes or lead therapeutics by using modules that target RNA base pairs.

Graphical Abstract

*Correspondence: disney@scripps.edu.

Author Contributions

M.D.D. conceived of the study, and H.S.H. and M.D.D. wrote the manuscript with assistance from A.G. A.G. implemented the TO-PRO-1 displacement assay and completed initial high-throughput screening. H.S.H. completed the chemoinformatic analysis, designed the RNA-focused small molecule library, and completed all binding measurements and modeling studies. All authors have given approval to the final version of the manuscript.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscmbosci.8b00049.

Summary of all screening data, energy minimized measured for hits, photophysical data of hit compounds, and MST binding data (PDF).

The authors declare no competing financial interest.

Keywords

RNA; nucleic acids; high-throughput screening; base pairs; small molecules

INTRODUCTION

RNA has recently emerged as a promising drug target¹ as there are various RNA-driven diseases including the myotonic dystrophies, Huntington's disease, Tau-mediated neurodegeneration, and spinal muscular atrophy (SMA).² Most efforts toward developing therapeutics targeting RNAs involve antisense oligonucleotides (ASOs) that function through Watson–Crick base pairing.³ Although ASOs have shown success in a few cases, they also have several limitations, such as causing toxicity in clinical trials and the need for complex delivery systems.⁴ Small molecules are well established as drugs in the proteome; however, their utility in targeting RNA is relatively nascent and wrought with challenges. For example, only the bacterial ribosome and riboswitches have been extensively investigated as drug targets for small molecules.⁵ Compared to most cellular RNAs, ribosomes and riboswitches have significant tertiary structure, akin to proteins, to which small molecules can bind. Coding and noncoding RNAs primarily fold into composites of secondary structural elements dictated by noncanonical base pairing and looped regions. Although these types of RNAs are more common than those with tertiary folds, there are currently few examples of small molecules that target RNAs exhibiting predominantly secondary structural elements.

To address this challenge, our group has developed a sequence-based design strategy to identify small molecules that selectively target RNAs that have defined secondary structures. This approach, dubbed Inforna, annotates a target RNA's secondary structure (canonically and noncanonically paired or loop regions) and cross references it to experimentally determined RNA motif-small molecule interactions.^{6,7} These interactions are identified via two-dimensional combinatorial screening (2DCS), a library-versus-library screen that elucidates small molecule binding preferences.⁸ By using Inforna, several small molecule lead medicines have been developed for various targets and indications. For example, small molecules targeting the expanded repeats that cause myotonic dystrophy types 1⁹ and 2¹⁰

and oncogenic microRNA precursors,^{8b,c,11} have in vitro and in vivo efficacy. Efforts have also been made to target other noncoding RNAs such as HIV TAR RNA.¹²

Although it has been demonstrated that small molecules targeting noncanonically paired motifs such as internal loops, bulges, and hairpins provide a rich source of chemical probes and lead medicines, RNAs are ~50% base paired. Thus, to expand the applicability of Informa, more data defining the small molecules that bind base pairs are required. Previously, a selective modulator of r(AUUCU)^{exp} repeats in spinocerebellar ataxia 10 (SCA10)¹³ was developed. To enable these studies, an AU base pair targeting small molecule was discovered and transformed into a potent dimeric compound that specifically targets the periodic AU pairs formed by r(AUUCU) repeats. Expanding the knowledge base of base pair targeting increases the potential for broader applicability of sequence-based design of small molecules that target RNA.

Therefore, in this study, we created an RNA-focused small molecule library and interrogated its binding capacity to various RNA base paired constructs. These studies identified various chemotypes capable of selective recognition. Many of the compounds not only bind selectively to GC or AU pairs but can also discriminate among different base pair orientations. For example, small molecules can selectively recognize 5'GC/3'CG pairs from 5'GG/3'CC pairs. These data are likely to be broadly applicable to the rational design of small molecules that target base pairs and can enable the development of hybrid base pair- and loop-targeting compounds.

RESULTS AND DISCUSSION

Construction of an RNA-Focused Small Molecule Library.

Previously, RNA-focused small molecule libraries have been constructed by displaying RNA-binding submonomers on peptoid backbones,¹⁴ computational docking of molecules to known RNA structures,¹⁵ and chemoinformatics analysis of known RNA binders.¹⁶ Small molecules that bind RNA contain benzimidazole, 2-aminobenzimidazole, *bis*-benzimidazole, alkylpyridinium, indole, and 2-phenyl indole cores, among others.¹⁷ Relevant physiochemical properties include three to four rotatable bonds, total polar surface area (TPSA) ranging from 60 to 92 Å², and at least two or more hydrogen bond acceptor (HBA) and donor (HBD) moieties per molecule. Compared to typical drug-like compounds, RNA binders often have more hydrogen bond donors and acceptors.¹⁸

To construct a diverse collection of small molecules to probe RNA binding, a physiochemical analysis of known RNA binders contained within Informa was completed. The physiochemical properties of these compounds are summarized in Table 1. Using these properties, we analyzed compounds in the Chembridge Core and Express libraries for molecules with similar properties, affording 3271 molecules in our RNA-focused chemotype library. Although all library members were chosen to fit within these defined physiochemical parameters, they were also selected to be as structurally diverse as possible with at least 20% of the library deviating from normal RNA-binding chemotypes.

Among several compound features, nitrogen-containing heterocyclic small molecules are enriched within the library to engage RNA targets via hydrogen bonding. These compounds also have different partial charges that could stack differentially on RNA bases pairs. Common chemotypes include 2-phenyl-1,3-benzimidazole (a), 1,2-benzimidazole (b), 2-phenyl indole (c), 4-phenyl thiazole (d), and 2-amino quinazolines (e), vide infra. The compound collection was compared to the physiochemical properties of current drugs on the market, as shown in Table 1. These comparisons show that small molecules contained within the RNA-focused chemotype library share similar physiochemical properties to known drugs and known RNA binders, suggesting the potential for identifying therapeutically applicable RNA-binding compounds.

Design of Base Paired RNAs to Assess Small Molecule Binding.

RNAs that display various base pairs were designed and studied for binding to members of the small molecule library, including $r(\text{AAUU})_3$, $r(\text{AU})_6$, $r(\text{GGCC})_2$, and $r(\text{GC})_4$ (Figure 1A) to contain the most common base pair orientations surrounding looped regions in RNA. Each construct contains a 5'GAAA3' (GNRA) tetraloop sequence to ensure proper folding of the hairpin.¹⁹ To study whether small molecules can recognize differences in base pair orientation, RNAs were designed to have either 5'AA/3'UU or 5'AU/3'UA pairs with similar designs for the GC paired constructs. The $r(\text{AAUU})_3$ and $r(\text{AU})_6$ constructs are eight nucleotides longer than their GC pair counterparts to increase their thermodynamic stability. As predicted by RNA structure,²⁰ constructs of equivalent length, or $r(\text{AAUU})_2$ and $r(\text{AU})_4$ have $G_{37^\circ\text{C}}$ values of -2.5 and -3.7 kcal/mol respectively; addition of eight nucleotides to each construct [$r(\text{AAUU})_3$ and $r(\text{AU})_6$] improves their predicted thermodynamic stability to -6.7 and -8.5 kcal/mol, respectively. The GC paired sequences were size-minimized while ensuring homogeneous folding due to issues in the synthesis of long GC-rich RNAs and their propensity to form more complex structures such as quadruplexes.²¹ The $r(\text{GGCC})_2$ and $r(\text{GC})_4$ have $G_{37^\circ\text{C}}$ values of -19.1 and -17.5 kcal/mol, respectively, and are therefore sufficiently stable for study. This panoply of targets can thus be used to assess selective binding among different types and orientations of RNA base pairs.

Development of the TO-PRO-1 Displacement Assay for High-Throughput Screening.

To develop an assay that allowed for efficient screening of compounds for binding base paired RNA in a high-throughput format, we used a dye displacement assay. This approach has been used broadly with various dyes²² to define compounds that target nucleic acid structures and was pioneered by the Boger group.^{22a,b} We previously determined that TO-PRO-1 is an ideal RNA binding probe due to favorable properties such as a low false positive rate.²³ When TO-PRO-1 binds RNA, its fluorescence increases substantially; when the dye is displaced by a compound, fluorescence decreases (Figure 1B). Using this assay, the $r(\text{AAUU})_3$, $r(\text{AU})_6$, $r(\text{GGCC})_2$, and $r(\text{GC})_4$ constructs were probed for binding to our RNA-focused library.

We first measured the binding affinity of TO-PRO-1 for each RNA construct, which afforded K_d values ranging from 5 to 12 μM (Table S1). For screening, a concentration of RNA which gave a signal greater than 3-fold above background was used (Figure 1C–F). The assay's Z-factors were calculated using mitoxantrone (MTX), a known RNA binder,²⁴ for each RNA

construct. The Z-factors ranged between 0.5 and 0.9 and indicate the assay's robustness for high-throughput screening (Figure 1G).²⁵

Screening of the RNA-Focused Library Provides Selective Binders of GC and AU Paired RNA.

The RNA-focused library was studied for binding each base paired RNA as outlined in Figure 2A. Each library member was first screened at a single dose (100 μ M), and hits were defined as compounds that reduced TO-PRO-1 fluorescence by more than 3 standard deviations (3σ) from the mean, affording 28 compounds (Figures 2B, 2C, S1, and S2).

Compounds that increased signal in the TO-PRO-1 channel (likely due to the inherent fluorescence of the compound) were eliminated from further consideration. Hit rates for r(AAUU)₃, r(AU)₆, r(GGCC)₂, and r(GC)₄ RNAs were 0.7%, 0.6%, 0.7%, and 0.3%, respectively. These rates are similar in magnitude to other high throughput screens. Hit molecules were binned into subclasses based on structural similarity (Figure 3 and Table S2). Notably, many of the hit small molecules are derivatives of benzimidazoles, whose RNA-binding capacity is wellknown.^{17a,26} The RNA-focused library is comprised of ~63% benzimidazoles, while ~93% ($p < 0.001$) of hit compounds contain this core structure. On the basis of the proportion of benzimidazoles in the starting library and their known RNA-binding capacity, such enrichment is expected.^{17a,26} Class 1 comprises small molecules with tricyclic benzimidazoles and oxygen-rich functionalities on the 3, 4, and 5 positions of the phenyl ring (compounds **1–4**). Interestingly, this chemotype only makes up 0.3% of the library, while comprising 14% of the hit compounds, a statistically significant enrichment ($p < 0.001$). These molecules have the highest hydrophobic character of all six subcategories, with LogDs of ~5.0 and TPSA's averaging 60 $\text{\AA}^2$. Class 2 compounds (**5–8**) maintain the 2-phenyl benzimidazole core with strong polar character (average TPSA of ~97 $\text{\AA}^2$) primarily due to their HBA capacity. Compounds **9–14** comprise Class 3, which contains 2-phenyl benzimidazoles bearing amine rich functionalities on the 2-phenyl ring. Physiochemically, Class 3 compounds do not differ much in their properties from the library averages in Table 1. Compounds **15–18** (Class 4) are unique in that they contain a tricyclic system formed by two imidazoles fused to a benzene ring. Similar to Class 1, 0.2% of the library is comprised of this fused ring system, yet 14% ($p < 0.001$) of the hits have this chemotype. Compounds that did not appear as hits contained substituted alkyl chains and multiple methyl groups on the phenyl ring, potentially blocking binding via steric hindrance. The 1,3-benzimidazoles account for 23% of the starting compound library but make up 64% ($p < 0.001$) of the hit compounds, a 3-fold enrichment.

Class 5 hits are the most structurally diverse and unlike all other classes contain 1,2-benzimidazoles. They are also functionally diverse, bearing sulfonamide (**19–21**), spirocycle (**22**), and nitrogen-rich heterocycle (**23** and **24**) substitutions that are rarely observed in other hit compounds. These compounds are also highly polar, with LogDs averaging 0.84 and TPSAs averaging 105 $\text{\AA}^2$. They generally are not specific for a particular base pair (Table S4). The 1,2-benzimidazoles, unlike 1,3-benzimidazoles, show no enrichment between the starting library (~39%) and hit compounds (~21%), suggesting no preferential binding toward RNA for this core structure. The remaining hit compounds (**25–28**; Class 6) do not

share similar patterns of functionalization as other classes; however, quinazoline (1 of 28; $p > 0.01$) and benzothiazole (1 of 28; $p < 0.001$) moieties appear in **25** and **26** that do not appear in any other hits.

Compounds that displaced TO-PRO-1 for both AU paired RNAs exhibit planar, highly aromatic structures such as those in Class 1. The most notable difference is the presence of extended fused ring systems, for example observed in **1** and **4**. Small molecules that bound $r(\text{AAUU})_3$ and $r(\text{AU})_6$ showed no preference for base pair orientation, which is exemplified in the similar hit rates observed and the direct binding of **1**, **3**, and **14**, vide infra. All specific binders to $r(\text{AAUU})_3$ and $r(\text{AU})_6$ RNAs are contained within Classes 1 and 3.

In total, 25 molecules displaced TO-PRO-1 from $r(\text{GGCC})_2$, while only 11 bound $r(\text{GC})_4$. That is, unlike the compounds for AU paired RNAs, the screening data suggest that a subset of compounds can distinguish between the orientation of GC base pairs. Structurally, the compounds that are selective for $r(\text{GGCC})_2$ over $r(\text{GC})_4$ vary widely in their degree of functionalization, physiochemical properties, and aromatic nature, making it difficult to pinpoint which features drive specificity. The affinities of hit molecules for each RNA construct were measured for those amenable to direct binding assays; that is, compounds that showed absorption maxima at wavelengths greater than 300 nm and greater than 3-fold emission above background (photophysical data can be found in Table S4; $n = 24$ of 28 compounds) were further tested. Of the 24 molecules studied, six showed selective binding to either AU or GC paired RNAs.

Compounds that Selectively Bind AU Base Pairs.

Compounds **1**, **3**, and **14** bound selectively to both $r(\text{AU})_6$ and $r(\text{AAUU})_3$ with dissociation constants ranging from 30 to 24 000 nM (Figures 4A–C and S4). The $r(\text{AU})_6$ and $r(\text{AAUU})_3$ each have a 9 potential binding sites between AU base pairs; that is, a molecule that binds between AU base pairs would have a stoichiometry of 9:1. Jobs analysis²⁷ of **14** revealed that it bound $r(\text{AAUU})_3$ and $r(\text{AU})_6$ in a 2:1 ratio of compound to RNA. These data suggest that the binding pocket for the compounds encompasses multiple pairs in a row or that the ligand exhibits negative cooperativity upon binding. Such effects have been observed for other compounds. For example, upon binding of ethidium to AU and GC RNA pairs, the base pair interplanar distance expands from 3.0 to 6.7 Å.²⁸ This expansion can influence the ability of binding to adjacent sites and also change the accessibility of distal sites based on alterations in three-dimensional structure.²⁹ Accurate determination of **1**'s and **3**'s binding stoichiometries to $r(\text{AAUU})_3$ and $r(\text{AU})_6$ was not possible due to the low signal observed at low nanomolar concentrations of compound.

Binding analyses of **1** and **3** indicated their affinities were 2 orders of magnitude stronger than **14**. Enhanced affinities of **1** and **3** may be due to the phenanthrene core, which increases their stacking areas by $\sim 220 \text{ \AA}^2$ (Figure S3). Tam et al. revealed similar effects in the study of ethidium, where expansion of ethidium's polycyclic core increased its affinity for AU-rich RNAs by 4-fold.³⁰ Luedtke et al. also reported phenanthrene's selectivity for AU base pairs, demonstrating that ethidium and $[\text{Ru}(\text{bpy})_2\text{eilatin}]^{2+}$ bind 30-fold more tightly to AU paired RNAs than GC pairs.³¹ Although base pair orientation was not studied by Luedtke et al. the data herein indicate that phenanthrenes cannot distinguish between

5'AA/3'UU and 5'AU/3'UA pairs. Indeed, other compound classes that bind AU pairs also cannot distinguish between their orientations. For example, the inability of benzimidazole derivatives to distinguish between 5'AA/3'UU and 5'AU/3'UA pairs is also suggested by Yang et al. where substituted benzimidazoles appear in compound clusters for both r(AAU)₃ and r(AU)₆ binders.

Other previously identified AU binders include diphenylfuran amidines, which were shown to selectively bind AU pairs in r(AUUCU) repeats.¹³ Docking studies of diphenylfuran amidines show binding by threaded intercalation as they contain functionalities capable of hydrogen bonding in the major groove. Such a binding mode is unlikely for **1**, **3**, and **14** as they apparently lack flexible functionalities capable of groove threading.^{31b,c}

Compounds that Selectively Bind GC Base Pairs.

Compounds **9**, **10**, and **17** bound selectively to GC paired RNAs with dissociation constants ranging from 250 to 1200 nM. Compounds **9** and **10** selectively bound r(GC)₄, while **17** was unable to discriminate between r(GC)₄ and r(GGCC)₂ RNAs (Figure 4 and S5). Further analyses revealed that compound **9** and **10** bound r(GC) RNA with a stoichiometry of 5:1 and 4:1 respectively (Figure 4D–G). Compound **17** bound with a stoichiometry of 9:1 to r(GC)₄ and 3:1 to r(GGCC)₂ (Figures 4H–I and S5). These stoichiometries are close to the predicted binding to each GC pair in r(GC)₄ and r(GGCC)₂, suggesting intercalation as their binding mode. The 9:1 stoichiometry observed for **17** to r(GC)₄ suggests a more complex binding interaction. When modeled using an OPLS3 force field, **9** adopts a compact structure (Figure S6A). These studies showed that 75% of the poses for **9** place the saturated heterocycle perpendicular to the aromatic plane, *anti* relative to the urea moiety; the *syn* conformation places the heterocycle above the aromatic plane. This compact structure may support **9**'s ability to bind more sites in r(GC)₄ RNA than **10**. Compound **10** contains a sterically encumbered *N*-methyl urea. Energy minimization, placed this functionality perpendicular to the aromatic plane, with a dihedral angle of –89.0°. Jobs analysis showed that **10** binds with 1 fewer molecule to r(GC)₄ than **9**, which may be a result of this added steric bulk perturbing the RNA's structure. Interestingly, compounds **9** and **10** contain functional groups capable of potentially threading into the major groove and may contribute to their ability to discriminate between r(GC)₄ and r(GGCC)₂ RNAs. In contrast, the vinyl methoxy group in **17** is not positioned to participate in a threading intercalation binding mode. Furthermore, the 3:1 stoichiometry observed for **17** to r(GGCC)₂ suggests that upon binding the increase in the interplanar space between GC base pairs prevents binding to adjacent sites. This may result from accommodation of the vinyl methoxy group upon stacking that is not coplanar to the aromatic system. Compounds **9** and **10** have a urea moiety spacing the bulky heterocycle and cyclopropane, respectively, from the aromatic core allowing more effective stacking and less perturbation of the RNA's structure. Selective GC pair binders have been reported by Beal et al.³² and Zimmerman et al.^{9c,10a} In the former case, amino acid functionalized acridines were shown to selectively bind GC pairs by intercalation of the acridine and threading of the amino acid moieties into the major groove.³² In the latter case, Zimmerman used triazine acridine hybrids to simultaneously bind a UU mismatch and an adjacent GC pair.^{9c,10a} They utilized a similar approach incorporating polyamines that hydrogen bond to the major groove. The molecules studied here lack such

amino acid and polyamine functionalization. Furthermore, **9**, **10**, and **17** show selective binding to RNA versus DNA, vide infra.

Binding Affinities as Determined by Microscale Thermophoresis (MST).

To confirm our fluorescence-based binding analyses, compounds **10**, **14**, and **17** were studied for binding to Cy5-r(GC)₄ and Cy5-r(AU)₆ via MST (Table S6 and Figure S8). MST analysis showed that compound **10** bound r(GC)₄ with a K_d of 12.5(±1.5) nM and exhibited no saturable binding to r(AU)₆. Compound **14** showed a similar trend, with no saturable binding to r(GC)₄ and a K_d of 8400(±640) nM for r(AU)₆, maintaining its selectivity. Lastly, compound **17**, a GC selective molecule, bound r(GC)₄ with a K_d of 43.2(±6.8) nM and did not bind r(AU)₆. Collectively, these data show that both the measured K_d values and selectivities are similar to those measured by fluorescence.

Selectivity for RNA over DNA.

Nucleic acid binders like ethidium, acridines, and triazines reported by Wilson et al.³³ and Zimmerman et al.^{9c} are known to also bind DNA. Given that DNA is a potential target, the dissociation constants for **1**, **3**, **9**, **10**, **14**, and **17** for DNA were measured by direct binding measurements. DNA constructs were the same sequences as the RNAs studied (Figure S7 and Table S5). All six molecules exhibited exquisite selectivity for RNA over DNA base pairs, with no saturable binding observed for the DNA constructs (K_d values > 100 μ M; Table S5). In particular, **9**, **14**, and **17** gave no observable change in fluorescence when incubated with 300 μ M of DNA (Figure S7 and Table S5). Thus, all compounds show at least a 50-fold selectivity for RNA. The selectivities of the molecules described herein can be driven by structural differences in B-form DNA and A form RNA helices. It is known that DNA has a narrower minor groove than the minor groove of RNA, which can affect molecular recognition by a small molecule.³⁴ Neither the DNA minor groove nor its base pairs can accommodate the bulky derivatives due to steric clashes with the backbone and rigidity in its structure, preventing sufficient expansion of the interplanar space required for binding. Compounds **9**, **10**, and **17** also contain sterically encumbering groups that can preclude DNA binding via similar clashes. Side chain functionalities on **9**, **10**, and **17** also lack sufficient hydrogen-bonding capacity for effective groove binding like distamycin and other reported DNA groove binders, further precluding DNA binding.³⁴ Compounds **1**, **3**, and **14** lack side chain functionalities, eliminating groove binding as a possible binding mode to DNA. They do, however, have substituents (ethoxy, methoxy, and methyl groups) that could prevent effective intercalation with DNA because of steric clashes, particularly since studies of similar molecules required expansion of the interplanar space by 2-fold to accommodate binding. Therefore, it is likely that DNA's structural rigidity drives these molecules' binding preference toward RNA.

Summary and Conclusions.

In summary, we developed a novel 3271 member RNA-focused small molecule library using chemoinformatic analysis and chemical similarity searching of known RNA binders within Inforna.^{8a} Using this library, a high-throughput screen was completed via a dye displacement assay to assess binding of each compound to r(AU)₆, r(AAU)₃, r(GC)₄, and

r(GGCC)₂, which mimic the most common base pairs that flank looped regions in cellular RNAs. Screening studies afforded 28 compounds that are novel binders to RNA, with a global hit rate of 0.7%, similar to hit rates previously obtained by our group. Compounds **1**, **3**, and **14** selectively bound AU paired RNAs with high nanomolar to low micromolar affinities. Of the AU pair binders, **14** bound r(AAUU)₃ and r(AU)₆ with a 2:1 stoichiometry, suggesting its binding significantly perturbs the RNA's structure limiting further binding events. Compounds **9**, **10**, and **17** selectively bound GC paired RNA, with **9** and **10** selectively binding r(GC)₄ over r(GGCC)₂. Binding stoichiometries of these compounds also revealed steric factors may affect binding affinity and selectivity. Importantly, all six molecules amenable to direct binding assays are selective for RNA over DNA, likely due to their sterically bulky structures. This study and those previously reported by our group and others sets the foundation for the rational design of small molecules targeting RNA.

EXPERIMENTAL PROCEDURES

General Methods.

All RNA was purchased from Dharmacon (GE Healthcare), deprotected per the manufacturer's protocol, and desalted with a PD-10 column (GE Healthcare). RNAs were quantified by UV/vis spectrometry using the absorbance at 260 nm measured at 85 °C. RNA sequences were r(AAUU)₃: AAUUAUUAAUUGAAAAA- UUAUUAAUU, r(AU)₆: AUAUAUAUAUAUGAAAAU- AUAUAUAUAU, r(GGCC)₃: GGCCGGCCGAAAGGC-CGGCC, r(GC)₄: GCGCGCGCGAAAGCGCGCGC. DNA sequences were obtained from IDT and were used without further purification. Their sequences are as follows: d(AATT)₃, AATTAATTAATTGAAAAATTAATTAATT; d(AT)₆, ATA-TATATATATGAAAATATATATATAT; d(GGCC)₂, GGC- CGGCCGAAAGGCCGGCC; d(GC)₄, GCGCGCGCGAAA- GCGCGCGC; d(AT)₁₁, ATATATATATATATATATATAT; d(GC)₁₁, GCGCGCGCGCGCGCGCGCGCGC. The chemical library for screening was designed based on previous RNA-binding studies and purchased from ChemBridge Corporation preplated as 10 mM stocks dissolved in dimethyl sulfoxide (DMSO).

Affinity of TO-PRO-1 for RNA Constructs.

For binding assays, the RNA of interest was prepared in 1× Binding Buffer (8 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, and 2 mM EDTA), and folded by heating for 3 min at 70 °C and slowly cooled to room temperature on the benchtop. Bovine serum albumin (BSA) and TOPRO-1 were added to final concentrations of 40 µg/mL and 200 nM, respectively. Serial dilutions of the RNA were then completed in 1× Binding Buffer containing 40 µg/mL BSA and 200 nM TO-PRO-1. The solutions were incubated for 15 min at room temperature in the dark and then transferred to a well of a black 384-well plate (Greiner BioOne; catalog #: 784076). Fluorescence intensity was measured on Molecular Devices SpectraMax M5 plate reader (515 and 580 nm excitation and emission wavelengths, respectively). The optimal concentrations of RNA were: 625 nM for AU paired RNAs and 2 µM for GC paired RNAs, affording greater than 3-fold signal above background.

High-Throughput Screening of Compounds.

HTS of compounds to identify RNA binders was completed by folding the RNA as described above in 1× Binding Buffer followed by addition of TO-PRO-1 and BSA. A 10 µL aliquot of this sample was loaded into each well of a 384-well plate. Compounds (100 nL; 100 µM final concentration) were transferred to the plate containing the RNA-TO-PRO-1 mixtures using a Biomek NXP pin transfer tool (Beckman Coulter). After the samples were incubated for 30 min at room temperature, fluorescence intensity was measured as described above. Hits were defined as compounds that reduced TO-PRO-1 emission by three standard deviations from the mean change in fluorescence.

Photophysical Characterization.

The photophysical properties of hit compounds identified from HTS were measured using a DU800 UV-vis spectrophotometer (Beckman Coulter) and a Varian Eclipse spectrofluorimeter.

Calculation of Z-Factor for HTS Assays.

Z-factors were calculated according to eq 1²⁵

$$\text{estimated Z factor} = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|} \quad (1)$$

where σ_p and σ_n are the standard deviations of the positive and negative controls, respectively, and μ_p and μ_n are the means of the positive and negative controls, respectively.

Measuring Compound Affinity and Stoichiometry.

All binding assays were completed in 1× Binding Buffer. Each compound was used at a concentration that provided a signal 3-fold greater than background (concentrations and wave-lengths provided in Tables S3 and S4). A 25 µL aliquot of 100 µM RNA/DNA was folded in 1× Binding Buffer as described above and then compound of interest was added. The samples were serially diluted in 1× Binding Buffer containing compound at the same concentration added to the RNA (Table S3). The samples were incubated for 1 h at room temperature, and then fluorescence intensity were measured on either a Molecular Devices SpectraMax M5 or Tecan Safire plate reader at the appropriate wavelength. Isotherms were plotted as percent change in fluorescence as a function of RNA or DNA concentration and fit to eq 2

$$I = I_0 + 0.5 \Delta \varepsilon \left(([FL]_0 + [RNA]_0 + K_t) - \left(([FL]_0 + [RNA]_0 + K_t)^2 - 4[FL]_0[RNA]_0 \right)^{0.5} \right) \quad (2)$$

where I and I_0 are the observed fluorescence and initial fluorescence intensity in the presence and absence of RNA, ϵ is the difference between the fluorescence intensity in the absence and presence of infinite RNA concentration, $[FL]_0$ and $[RNA]_0$ are, respectively, the concentrations of the small molecule and RNA, and K_t is the dissociation constant. Stoichiometry was obtained using the method of continuous variation or Jobs plot as previously described.²⁷

Chemoinformatics Analysis.

All chemoinformatics analyses were completed using Instant JChem (Chem Axon). Structural minimizations were carried out using the Schrödinger computational suite (release no. 4–2017) under a OPLS3 force field under default conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was funded by the National Institutes of Health [R01 GM97455 to M.D.D.]. We thank Dr. Simon Vezina-Dawod, Brendan Dwyer, and Dr. Jessica L. Childs-Disney for editing the manuscript.

REFERENCES

- (1). (a) Alvarez-Erviti L; Seow Y; Schapira AH; Rodriguez-Oroz MC; Obeso JA; Cooper JM Influence of microRNA deregulation on chaperone-mediated autophagy and [alpha]-synuclein pathology in Parkinson's disease. *Cell Death Dis.* 2013, 4, e545. [PubMed: 23492776] (b) Ash PEA; Bieniek KF; Gendron TF; Caulfield T; Lin W-L; DeJesus-Hernandez M; van Blitterswijk MM; Jansen-West K; Paul JW, III; Rademakers R; Boylan KB; Dickson DW; Petrucelli L Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 2013, 77 (4), 639–646. [PubMed: 23415312] (c) Butler AE; Janson J; Bonner-Weir S; Ritzel R; Rizza RA; Butler PC β -Cell deficit and increased β -cell apoptosis in humans with type 2 diabetes. *Diabetes* 2003, 52 (1), 102–110. [PubMed: 12502499] (d) Esteller M Non-coding RNAs in human disease. *Nat. Rev. Genet* 2011, 12 (12), 861–874. [PubMed: 22094949] (e) Greco S; Fasanaro P; Castelvechio S; D'Alessandra Y; Arcelli D; Di Donato M; Malavazos A; Capogrossi MC; Menicanti L; Martelli F MicroRNA dysregulation in diabetic ischemic heart failure patients. *Diabetes* 2012, 61 (6), 1633–41. [PubMed: 22427379]
- (2). (a) Ballatore C; Lee VM; Trojanowski JQ Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat. Rev. Neurosci* 2007, 8, 663–672. [PubMed: 17684513] (b) Andrew SE; Paul Goldberg Y; Kremer B; Telenius H; Theilmann J; Adam S; Starr E; Squitieri F; Lin B; Kalchman MA; Graham RK; Hayden MR The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat. Genet* 1993, 4 (4), 398–403. [PubMed: 8401589] (c) Krauss S; Griesche N; Jastrzebska E; Chen C; Rutschow D; Achmuller C; Dorn S; Boesch SM; Lalowski M; Wanker E; Schneider R; Schweiger S Translation of HTT mRNA with expanded CAG repeats is regulated by the MID1-PP2A protein complex. *Nat. Commun* 2013, 4, 1511. [PubMed: 23443539] (d) Mulders SAM; van den Broek WJAA; Wheeler TM; Croes HJE; van Kuik-Romeijn P; de Kimpe SJ; Furling D; Platenburg GJ; Gourdon G; Thornton CA; Wieringa B; Wansink DG Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. *Proc. Natl. Acad. Sci. U. S. A* 2009, 106 (33), 13915–13920. [PubMed: 19667189]
- (3). (a) Bennett CF; Swayze EE RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol* 2010, 50 (1), 259–293. [PubMed: 20055705] (b) Dean NM; Bennett CF Antisense oligonucleotide-based therapeutics for cancer. *Oncogene* 2003, 22, 9087. [PubMed: 14663487] (c) Mullard A 2016 FDA

drug approvals. *Nat. Rev. Drug Discovery* 2017, 16, 73. [PubMed: 28148938] (d)Southwell AL; Skotte NH; Bennett CF; Hayden MR Antisense oligonucleotide therapeutics for inherited neurodegenerative diseases. *Trends Mol. Med* 2012, 18 (11), 634–643. [PubMed: 23026741] (e)Juliano RL; Alahari S; Yoo H; Kole R; Cho M Antisense pharmacodynamics: critical issues in the transport and delivery of antisense oligonucleotides. *Pharm. Res* 1999, 16 (4), 494–502. [PubMed: 10227702] (f)Kole R; Krainer AR; Altman S RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug Discovery* 2012, 11 (2), 125–140. [PubMed: 22262036]

- (4). (a)Hagedorn PH; Hansen BR; Koch T; Lindow M Managing the sequence-specificity of antisense oligonucleotides in drug discovery. *Nucleic Acids Res.* 2017, 45 (5), 2262–2282. [PubMed: 28426096] (b)Lambert G; Fattal E; Couvreur P Nanoparticulate systems for the delivery of antisense oligonucleotides. *Adv. Drug Delivery Rev.* 2001, 47 (1), 99–112.(c)Lebedeva I; Benimetskaya L; Stein CA; Vilenchik M Cellular delivery of antisense oligonucleotides. *Eur. J. Pharm. Biopharm* 2000, 50 (1), 101–119. [PubMed: 10840195]
- (5). (a)Andrews WJ; Panova T; Normand C; Gadal O; Tikhonova IG; Panov KI Old drug, new target: ellipticines selectively inhibit RNA polymerase I transcription. *J. Biol. Chem* 2013, 288 (7), 4567–4582. [PubMed: 23293027] (b)Tenson T; Mankin A Antibiotics and the ribosome. *Mol. Microbiol* 2006, 59 (6), 1664–77. [PubMed: 16553874] (c)Blount KF; Breaker RR Riboswitches as antibacterial drug targets. *Nat. Biotechnol* 2006, 24 (12), 1558–64. [PubMed: 17160062] (d)Blount KF; Wang JX; Lim J; Sudarsan N; Breaker RR Antibacterial lysine analogs that target lysine riboswitches. *Nat. Chem. Biol* 2007, 3 (1), 44–9. [PubMed: 17143270] (e)Garst AD; Edwards AL; Batey RT Riboswitches: Structures and Mechanisms. *Cold Spring Harbor Perspect. Biol* 2011, 3 (6), a003533.(f)Serganov A; Nudler E A Decade of Riboswitches. *Cell* 2013, 152 (1–2), 17–24. [PubMed: 23332744]
- (6). (a)Aminova O; Paul DJ; Childs-Disney JL; Disney MD Two-dimensional combinatorial screening identifies specific 6′-acylated kanamycin A- and 6′-acylated neamine-RNA hairpin interactions. *Biochemistry* 2008, 47 (48), 12670–12679. [PubMed: 18991404] (b)Velaga- pudi SP; Disney MD Defining RNA motif–aminoglycoside interactions via two-dimensional combinatorial screening and structure–activity relationships through sequencing. *Bioorg. Med. Chem* 2013, 21 (20), 6132–6138. [PubMed: 23719281]
- (7). Mathews DH; Disney MD; Childs JL; Schroeder SJ; Zuker M; Turner DH Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A* 2004, 101 (19), 7287–7292. [PubMed: 15123812]
- (8). (a)Disney MD; Winkelsas AM; Velagapudi SP; Southern M; Fallahi M; Childs-Disney JL Informa 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chem. Biol* 2016, 11 (6), 1720–8. [PubMed: 27097021] (b)Velagapudi SP; Cameron MD; Haga CL; Rosenberg LH; Lafitte M; Duckett DR; Phinney DG; Disney MD Design of a small molecule against an oncogenic noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A* 2016, 113 (21), 5898–5903. [PubMed: 27170187] (c)Velagapudi SP; Gallo SM; Disney MD Sequencebased design of bioactive small molecules that target precursor microRNAs. *Nat. Chem. Biol* 2014, 10 (4), 291–297. [PubMed: 24509821]
- (9). (a)Parkesh R; Childs-Disney JL; Nakamori M; Kumar A; Wang E; Wang T; Hoskins J; Tran T; Housman D; Thornton CA; Disney MD Design of a bioactive small molecule that targets the myotonic dystrophy type 1 RNA via an RNA motif-ligand database and chemical similarity searching. *J. Am. Chem. Soc* 2012, 134 (10), 4731–42. [PubMed: 22300544] (b)Rzuczek SG; Gao Y; Tang Z-Z; Thornton CA; Kodadek T; Disney MD Features of modularly assembled compounds that impart bioactivity against an RNA target. *ACS Chem. Biol* 2013, 8 (10), 2312–2321. [PubMed: 24032410] (c)Wong C-H; Richardson SL; Ho Y-J; Lucas AMH; Tuccinardi T; Baranger AM; Zimmerman SC Investigating the binding mode of an inhibitor of the MBNL1-RNA complex in myotonic dystrophy type 1 (DM1) leads to the unexpected discovery of a DNA-selective binder. *ChemBioChem* 2012, 13 (17), 2505–2509. [PubMed: 23097190] (d)Childs-Disney JL; Hoskins J; Rzuczek SG; Thornton CA; Disney MD Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. *ACS Chem. Biol* 2012, 7 (5), 856–862. [PubMed: 22332923] (e)Wong C-H; Nguyen L; Peh J; Luu LM; Sanchez JS; Richardson SL; Tuccinardi T; Tsoi H; Chan WY; Chan HYE; Baranger

AM; Hergenrother PJ; Zimmerman SC Targeting toxic RNAs that cause myotonic dystrophy type 1 (DM1) with a bisamidinium inhibitor. *J. Am. Chem. Soc* 2014, 136 (17), 6355–6361. [PubMed: 24702247]

- (10). (a)Wong C-H; Fu Y; Ramisetty SR; Baranger AM; Zimmerman SC Selective inhibition of MBNL1–CCUG interaction by small molecules toward potential therapeutic agents for myotonic dystrophy type 2 (DM2). *Nucleic Acids Res.* 2011, 39 (20), 8881–8890. [PubMed: 21768123] (b)Childs-Disney JL; Yildirim I; Park H; Lohman JR; Guan L; Tran T; Sarkar P; Schatz GC; Disney MD Structure of the myotonic dystrophy type 2 RNA and designed small molecules that reduce toxicity. *ACS Chem. Biol* 2014, 9 (2), 538–550. [PubMed: 24341895] (c)Rzuczek SG; Park H; Disney MD A toxic RNA catalyzes the in cellulo synthesis of its own inhibitor. *Angew. Chem., Int. Ed* 2014, 53 (41), 10956–10959.
- (11). Costales MG; Haga CL; Velagapudi SP; Childs-Disney JL; Phinney DG; Disney MD Small molecule inhibition of microRNA-210 reprograms an oncogenic hypoxic circuit. *J. Am. Chem. Soc* 2017, 139 (9), 3446–3455. [PubMed: 28240549]
- (12). (a)Wang S; Huber PW; Cui M; Czarnik AW; Mei H-Y Binding of neomycin to the TAR element of HIV-1 RNA induces dissociation of Tat protein by an allosteric mechanism. *Biochemistry* 1998, 37 (16), 5549–5557. [PubMed: 9548939] (b)Mei H-Y; Galan AA; Halim NS; Mack DP; Moreland DW; Sanders KB; Hoa NT; Czarnik AW Inhibition of an HIV-1 Tat-derived peptide binding to TAR RNA by aminoglycoside antibiotics. *Bioorg. Med. Chem. Lett* 1995, 5 (22), 2755–2760. (c)Mei H-Y; Cui M; Heldsinger A; Lemrow SM; Loo JA; Sannes-Lowery KA; Sharmeen L; Czarnik AW Inhibitors of protein–RNA complexation that target the RNA: specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules. *Biochemistry* 1998, 37 (40), 14204–14212. [PubMed: 9760258] (d)Patwardhan NN; Ganser LR; Kapral GJ; Eubanks CS; Lee J; Sathyamoorthy B; Al-Hashimi HM; Hargrove AE Amiloride as a new RNA-binding scaffold with activity against HIV-1 TAR. *MedChemComm* 2017, 8 (5), 1022–1036. [PubMed: 28798862] (e)Hamy F; Felder ER; Heizmann G; Lazdins J; Aboul-ela F; Varani G; Karn J; Klimkait T An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc. Natl. Acad. Sci. U. S. A* 1997, 94 (8), 3548–3553. [PubMed: 9108013]
- (13). Yang W-Y; Gao R; Southern M; Sarkar PS; Disney MD Design of a bioactive small molecule that targets r(AUUCU) repeats in spinocerebellar ataxia 10. *Nat. Commun* 2016, 7, 11647. [PubMed: 27248057]
- (14). (a)Labuda LP; Pushechnikov A; Disney MD Small molecule microarrays of RNA-focused peptoids help identify inhibitors of a pathogenic group I intron. *ACS Chem. Biol* 2009, 4 (4), 299–307. [PubMed: 19278238] (b)Seedhouse SJ; Labuda LP; Disney MD The Privileged Chemical Space Predictor (PCSP): A computer program that identifies privileged chemical space from screens of modularly assembled chemical libraries. *Bioorg. Med. Chem. Lett* 2010, 20 (4), 1338–1343. [PubMed: 20097562]
- (15). (a)Gallego J; Varani G Targeting RNA with small-molecule drugs: therapeutic promise and Chemical Challenges. *Acc. Chem. Res* 2001, 34 (10), 836–843. [PubMed: 11601968] (b)Stelzer AC; Frank AT; Kratz JD; Swanson MD; Gonzalez-Hernandez MJ; Lee J; Andricioaei I; Markovitz DM; Al-Hashimi HM Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol* 2011, 7, 553. [PubMed: 21706033] (c)Lang PT; Brozell SR; Mukherjee S; Pettersen EF; Meng EC; Thomas V; Rizzo RC; Case DA; James TL; Kuntz ID DOCK 6: Combining techniques to model RNA–small molecule complexes. *RNA* 2009, 15 (6), 1219–1230. [PubMed: 19369428]
- (16). (a)Velagapudi SP; Pushechnikov A; Labuda LP; French JM; Disney MD Probing a 2aminobenzimidazole library for binding to RNA internal loops via two-dimensional combinatorial screening. *ACS Chem. Biol* 2012, 7, 1902–1909. [PubMed: 22958065] (b)Velagapudi SP; Luo Y; Tran T; Haniff HS; Nakai Y; Fallahi M; Martinez GJ; Childs-Disney JL; Disney MD Defining RNA-small molecule affinity landscapes enables design of a small molecule inhibitor of an oncogenic noncoding RNA. *ACS Cent. Sci* 2017, 3 (3), 205–216. [PubMed: 28386598] (c)Velagapudi SP; Seedhouse SJ; French J; Disney MD Defining the RNA internal loops preferred by benzimidazole derivatives via 2D combinatorial screening and computational analysis. *J. Am. Chem. Soc* 2011, 133 (26), 10111–10118. [PubMed: 21604752] (d)Rzuczek SG; Southern MR; Disney MD Studying a drug-like, RNA-focused small molecule

library identifies compounds that inhibit RNA toxicity in myotonic dystrophy. *ACS Chem. Biol* 2015, 10 (12), 2706–2715. [PubMed: 26414664]

- (17). (a)Thomas JR; Hergenrother PJ Targeting RNA with small molecules. *Chem. Rev* 2008, 108 (4), 1171–1224. [PubMed: 18361529] (b)Angel- bello AJ; Chen JL; Childs-Disney JL; Zhang P; Wang Z-F; Disney MD Using genome sequence to enable the design of medicines and chemical probes. *Chem. Rev* 2018, 118 (4), 1599–1663. [PubMed: 29322778]
- (18). Childs-Disney JL; Disney MD Approaches to validate and manipulate RNA targets with small molecules in cells. *Annu. Rev. Pharmacol. Toxicol* 2016, 56 (1), 123–140. [PubMed: 26514201]
- (19). (a)Antao VP; Tinoco I Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.* 1992, 20 (4), 819–824. [PubMed: 1371866] (b)Murphy FL; Cech TR GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol* 1994, 236 (1), 49–63. [PubMed: 8107125]
- (20). Mathews DH Using the RNAstructure software package to predict conserved RNA structures. *Curr. Protoc. Bioinformatics* 2014, 12.4.1–12.4.22.
- (21). (a)Bugaut A; Balasubramanian S 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* 2012, 40 (11), 4727–4741. [PubMed: 22351747] (b)Millevoi S; Moine H; Vagner S G- quadruplexes in RNA biology. *Wiley Interdiscip. Rev. RNA* 2012, 3 (4), 495–507. [PubMed: 22488917]
- (22). (a)Boger DL; Fink BE; Brunette SR; Tse WC; Hedrick MP A simple, high-resolution method for establishing DNA binding affinity and sequence selectivity. *J. Am. Chem. Soc* 2001, 123 (25), 5878–91. [PubMed: 11414820] (b)Tse WC; Boger DL A fluorescent intercalator displacement assay for establishing DNA binding selectivity and affinity. *Acc. Chem. Res* 2004, 37 (1), 61–69. [PubMed: 14730995] (c)Asare-Okai PN; Chow CS A modified fluorescent intercalator displacement assay for RNA ligand discovery. *Anal. Biochem* 2011, 408 (2), 269–276. [PubMed: 20863807] (d)Zhang J; Umemoto S; Nakatani K Fluorescent indicator displacement assay for ligand-RNA interactions. *J. Am. Chem. Soc* 2010, 132 (11), 3660–1. [PubMed: 20196592] (e)Krishnamurthy M; Schirle NT; Beal PA Screening helix-threading peptides for RNA binding using a thiazole orange displacement assay. *Bioorg. Med. Chem* 2008, 16 (19), 8914–8921. [PubMed: 18789700]
- (23). (a)Asare-Okai PN; Chow CS A modified fluorescent intercalator displacement assay for RNA ligand discovery. *Anal. Biochem* 2011, 408 (2), 269–276. [PubMed: 20863807] (b)Tran T; Disney MD Identifying the preferred RNA motifs and chemotypes that interact by probing millions of combinations. *Nat. Commun* 2012, 3, 1125. [PubMed: 23047683]
- (24). (a)Liu Y; Peacey E; Dickson J; Donahue CP; Zheng S; Varani G; Wolfe MS Mitoxantrone Analogues as Ligands for a Stem-Loop Structure of Tau Pre-mRNA. *J. Med. Chem* 2009, 52 (21), 6523–6526. [PubMed: 19839622] (b)Zheng S; Chen Y; Donahue CP; Wolfe MS; Varani G Structural basis for stabilization of the tau premRNA splicing regulatory element by novantrone (mitoxantrone). *Chem. Biol* 2009, 16 (5), 557–566. [PubMed: 19477420]
- (25). Zhang J-H; Thomas DY; Kevin RO A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screening* 1999, 4 (2), 67–73.
- (26). Guan L; Disney MD Recent advances in developing small molecules targeting RNA. *ACS Chem. Biol* 2012, 7 (1), 73–86. [PubMed: 22185671]
- (27). Renny JS; Tomasevich LL; Tallmadge EH; Collum DB Method of continuous variations: applications of job plots to the study of molecular associations in organometallic chemistry. *Angew. Chem., Int. Ed* 2013, 52 (46), 11998–2013.(b)Huang CY Determination of binding stoichiometry by the continuous variation method: The job plot. *Methods Enzymol.* 1982, 87, 509–525. [PubMed: 7176926]
- (28). (a)Jain SC; Sobell HM Visualization of drug-nucleic acid interactions at atomic resolution. VII. structure of an ethidium/ dinucleoside monophosphate crystalline complex, ethidium: uridylyl-(3'-5')adenosine. *J. Biomol. Struct. Dyn* 1984, 1 (5), 1161–1177. [PubMed: 6400816] (b)Jain SC; Sobell HM Visualization of drug-nucleic acid interactions at atomic resolution. VIII. structures of two ethidium/dinucleoside monophosphate crystalline complexes containing ethidium: cytidylyl(3'-5') guanosine. *J. Biomol. Struct. Dyn* 1984, 1 (5), 1179–1194. [PubMed: 6400817]

- (29). Regulski EE; Breaker RR In-line probing analysis of riboswitches In Post-Transcriptional Gene Regulation; Wilusz J, Ed.; Humana Press: Totowa, NJ, 2008; pp 53–67.
- (30). Tam VK; Liu Q; Tor Y Extended ethidium bromide analogue as a triple helix intercalator: synthesis, photophysical properties and nucleic acids binding. *Chem. Commun* 2006, 25, 2684–2686.
- (31). (a)Luedtke NW; Hwang JS; Nava E; Gut D; Kol M; Tor Y The DNA and RNA specificity of eilatin Ru(II) complexes as compared to eilatin and ethidium bromide. *Nucleic Acids Res.* 2003, 31 (19), 5732–5740. [PubMed: 14500837] (b)Luedtke NW; Liu Q; Tor Y On the electronic structure of ethidium. *Chem. Eur. J* 2005, 11 (2), 495–508. [PubMed: 15549769] (c)Luedtke NW; Liu Q; Tor Y Synthesis, photophysical properties, and nucleic acid binding of phenanthridinium derivatives based on ethidium. *Bioorg. Med. Chem* 2003, 11 (23), 5235–5247. [PubMed: 14604688]
- (32). Carlson CB; Vuyisich M; Gooch BD; Beal PA Preferred RNA binding sites for a threading intercalator revealed by in vitro evolution. *Chem. Biol* 2003, 10, 663. [PubMed: 12890540]
- (33). Wilson WD; Ratmeyer L; Zhao M; Strekowski L; Boykin D The search for structure-specific nucleic acid-interactive drugs: Effects of compound structure on RNA versus DNA interaction strength. *Biochemistry* 1993, 32 (15), 4098–4104. [PubMed: 7682441]
- (34). (a)Dervan P Design of sequence-specific DNA-binding molecules. *Science* 1986, 232 (4749), 464–471. [PubMed: 2421408] (b)Dervan PB Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem* 2001, 9 (9), 2215–2235. [PubMed: 11553460]

Figure 1.

Schematic of the TO-PRO-1 displacement assay to identify small molecules that bind RNA base pairs. (A) Structures of RNAs screened for binding to the RNA-focused compound library. (B) Schematic of TO-PRO-1 displacement assay methodology. When not bound to RNA, TO-PRO-1 (blue sphere) has a low fluorescence signal; binding to RNA enhances TO-PRO-1 emission. Small molecule binding to the TO-PRO-1/RNA complex displaces TO-PRO-1, reducing its emission. (C) TO-PRO-1 direct binding to 625 nM r(AAUU)₃ gives 7-fold enhancement in signal-to-noise ratio relative to TO-PRO-1 alone. (D) TO-PRO-1 direct binding to 625 nM r(AU)₆ affords 6-fold enhancement in signal-to-noise ratio. (E) Direct binding of TO-PRO-1 to 2 μM r(GGCC)₂ affords a 17-fold enhancement in signal-to-noise ratio. (F) Direct binding of TO-PRO-1 to 2 μM r(GC)₄ affords a 20-fold enhancement in signal-to-noise ratio. Concentrations of 625 nM and 2 μM of RNA were used to screen AU and GC paired RNAs, respectively. (G) plot of Z factor for each RNA tested.

Figure 2.

Screening cascade for the RNA-focused small molecule library. (A) An RNA-focused chemotype library containing (a) 2-phenyl-1,3-benzimidazoles, (b) 1,2-benzimidazoles, (c) 2-phenyl indoles, (d) 4-phenyl thiazoles, and (e) 2-amino quinazolines screened for binding to RNA base pairs via TO-PRO-1 displacement. The pool of identified binders were then characterized for their fluorescence properties and assessed by direct binding assay in a dose response to obtain K_d values. (B & C) Representative results from the screening data obtained via TO-PRO-1 displacement assay for $r(\text{GC})_4$ and $r(\text{AU})_6$, respectively. Hits were selected based on reduction of TO-PRO-1 signal by >3 standard deviations from the mean (3σ). Compounds that enhanced TO-PRO-1 emission (signal > 0) were not considered.

Figure 3.

Structures and classification of hit small molecules obtained from TO-PRO-1 displacement screen. A total of 28 hits were obtained from HTS of the RNA- focused small molecule library. Five distinct structural classes were obtained with only four compounds not sharing significant structural similarity to enable categorization. Globally, benzimidazoles, whether 1,2- or 1,3-, are present in 93% of hits compared to 63% of the starting library ($p < 0.001$). Class 1 benzimidazoles are enriched 50-fold, 0.3% to 14% ($p < 0.001$), from the starting library to the final hit pool. Classes 2, 3, and 4 contain derivatives of 2-phenyl benzimidazoles, bearing oxygen, nitrogen and *bis*-imidazole fused ring systems, respectively. Among the hits, 2-phenyl benzimidazoles (classes 1–4) make up 63% of the hit pool, a 3-fold enrichment compared to the starting library ($p < 0.001$). Class 5 molecules contain a 1,2-benzimidazole core, and when compared to the entire compound library show a decrease in representation, 39% to 21%, however this change is not significant ($p > 0.05$). Class 6 molecules (gray) are not functionally similar to hits in other classes, exhibiting functionalities, such as benzothiazoles, that when compared to their representation in the starting library (0.2%) are enriched 20-fold (0.04%; $p < 0.001$). Quinazoline derivatives show no significant enrichment when compared to the starting library.

Figure 4.

Binding isotherms and Jobs analysis for binders. (A) Binding isotherms of **14** and base paired RNAs. (B and C) Jobs analysis of **14** and AU paired RNA shows binding stoichiometry of ~2:1 compound to RNA. (D) Binding isotherm of **9** and base paired RNA shows selective binding to r(GC)₄. (E) Jobs analysis of **9** and r(GC)₄ shows binding stoichiometry of 5:1. (F) Binding isotherms of **10** and base paired RNAs show selective binding to r(GC)₄. (G) Jobs analysis of **10** and r(GC)₄ RNA shows a stoichiometry of 4:1 compound to RNA. (H) Binding isotherms of **17** and base paired RNA show selective binding to GC paired RNAs only. (I) Jobs analysis of **17** binding to r(GGCC)₂ shows a stoichiometry of 3:1, indicating expansion of the interplanar space to accommodate **17**.

Table 1.

Summary of Physiochemical Properties Used to Design the RNA-Focused Library

library	molecular weight (Da)	cLogP	TPSA (Å)	H-bond donor	H-bond acceptor	rotatable bonds
Inforna ^a	468.6 ± 213.3	-0.3 ± 4.7	165.9 ± 124.2	9.0 ± 6.7	5.7 ± 4.8	7.6 ± 5.7
Drugs in DrugBank ^{a,b}	369 ± 263	2.0 ± 2.3	102.0 ± 106.7	5.0 ± 5.5	3.0 ± 3.6	6.0 ± 7.6
RNA-focused Library ^a	369.0 ± 59.3	2.81 ± 1.81	71.3 ± 20.2	3.9 ± 1.2	1.4 ± 0.9	4.2 ± 1.5

^aPhysiochemical properties were calculated for RNA binders within Inforna and the RNA-focused library using Instant J Chem (Chemaxon). Averages are reported with standard deviations for all parameters calculated.

^bData obtained from a physiochemical analysis of 8719 cataloged drugs in DrugBank.