**Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs**

Matthew D. Disney,* † Audrey M. Winkelsas, † Sai Pradeep Velagapudi, † Mark Southern, ‡ Mohammad Fallahi, ‡* and Jessica L. Childs-Disney ‡

†Department of Chemistry and ‡Informatics Core, The Scripps Research Institute, 130 Scripps Way, Jupiter, Florida 33458, United States

**Supporting Information**

**ABSTRACT:** The development of small molecules that target RNA is challenging yet, if successful, could advance the development of chemical probes to study RNA function or precision therapeutics to treat RNA-mediated disease. Previously, we described Inforna, an approach that can mine motifs (secondary structures) within target RNAs, which is deduced from the RNA sequence, and compare them to a database of known RNA motif—small molecule binding partners. Output generated by Inforna includes the motif found in both the database and the desired RNA target, lead small molecules for that target, and other related meta-data. Lead small molecules can then be tested for binding and affecting cellular (dys)function. Herein, we describe Inforna 2.0, which incorporates all known RNA motif—small molecule binding partners reported in the scientific literature, a chemical similarity searching feature, and an improved user interface and is freely available via an online web server. By incorporation of interactions identified by other laboratories, the database has been doubled, containing 1936 RNA motif—small molecule interactions, including 244 unique small molecules and 1331 motifs. Interestingly, chemotype analysis of the compounds that bind RNA in the database reveals features in small molecule chemotypes that are privileged for binding. Further, this updated database expanded the number of cellular RNAs to which lead compounds can be identified.

RNA is essential for many biological processes including gene regulation, translation, and pre-mRNA splicing, among others. Further, defective, malfunctioning, and aberrantly expressed RNAs cause or contribute to human disease. Thus, there is much interest in the development of small molecules that specifically target RNA. Such small molecules could be used as chemical probes to study RNA function and dysfunction in cells or in vivo or be developed into lead therapeutics. Despite the power of chemical biology probes as demonstrated for various proteins, few have been discovered for RNA. Targeting RNA is inherently challenging due to its polyanionic nature, which can cause nonspecific binding to cationic small molecules. Often chemical probes for proteins are identified by high throughput screening; such small molecule libraries are likely not ideal for RNA targets. That is, since so few compounds have been discovered that bind RNA specifically, it is difficult to predict the proper chemical space to screen for binding to RNA. Compounding these issues, most cellular RNAs are expressed at low levels, with the exception of rRNAs (rRNA), which constitute 80–90% of total cellular RNAs.‡ Thus, even if a small molecule targets a low abundance mRNA with high selectivity, the presence of comparatively high concentrations of rRNA could sequester much of the compound, reducing the overall potency of the small molecule.

Despite these challenges, targeting RNA could be very advantageous. Like proteins, RNA folds into complex three-dimensional structures that form pockets where small molecules could bind. RNA is structurally diverse, forming various secondary structural elements (motifs), including internal loops, hairpins, bulges, multibranch loops, pseudoknots, and paired regions (Figure 1), that could afford highly selective small molecule binding pockets. Selectivity could be further enhanced if one considers that not all small molecule binding sites are functional sites. That is, although a certain motif may be present in multiple transcripts, binding to only one may cause a biological effect. Another major advantage of targeting RNA is that there are reliable tools to gain insight into its secondary structure from sequence. Secondary structure can be predicted via free energy minimization by using programs such as ViennaRNA, mfold, and RNA_structure.‡–‡‡‡ These programs predict the lowest free energy structure (considered the one most likely formed) as well as suboptimal structures within a certain free energy (designated by the user; typically with 10% free energy of the lowest free energy structure). The confidence interval for each predicted base pair can also be calculated using...
a partition function. On average, the lowest free energy structures predicted via free energy minimization are ~75% accurate; other folding algorithms are 50–70% accurate. The accuracy of prediction can be greatly improved by including experimental constraints from enzymatic or chemical mapping. Other methods of RNA structure determination include phylogenetic comparison, NMR spectroscopy, and X-ray crystallography.

**Development of the Inforna Platform.** As mentioned above, little is known about what comprises the optimal chemical space for RNA. In order to identify chemotypes that engender small molecules for binding to RNA with high affinity and selectivity, we developed a library-vs-library screen named 2-Dimensional Combinatorial Screening (2DCS). Briefly, small molecules are site-specifically immobilized onto an agarose-coated microscope slide to afford small molecule microarrays. The arrays are incubated with a labeled RNA motif library under highly stringent conditions using competitive oligonucleotides that mimic regions common to all library members. Thus, binding events are restricted to the randomized region. The motifs chosen for our RNA libraries include hairpins, symmetric and asymmetric internal loops, and bulges. Each is intentionally kept small, for example five- and six-nucleotide hairpins, 4 × 3 or 4 × 4 nucleotide internal loops (the largest tested thus far), to increase the likelihood of finding biologically relevant binding partners. After incubation, positions where RNAs are bound are harvested by manual excision, RT-PCR amplified, and sequenced. A detailed protocol for 2DCS selections is available.

We also developed a computational method to score the fitness of a given RNA motif–small molecule interaction named Structure-Activity Relationships Through Sequencing (StARTS). Fitness is a measure of the affinity and selectivity of a given interaction and can be extended to reflect the selectivity of an RNA motif among many small molecules. StARTS compares the rate of occurrence of a structural feature within RNA motifs selected by 2DCS (such as a 5′GC step or an adenosine opposite an adenosine) or the motif in its entirety to the rate of occurrence of the same feature/motif in the starting library. The confidence that the feature indeed represents a true preference by the small molecule is calculated by using a pooled population comparison, affording a Z_obs value and the corresponding two-tailed p value. Z_obs values can be positive or negative, indicating that a given feature contributes positively or negatively, respectively, to the fitness of an RNA motif–small molecule interaction. Importantly, we have found that by summing the Z_obs values for all features within a motif (ΣZ-score), that ΣZ-score correlates with affinity and selectivity. Normalization of ΣZ-scores to the score for the most fit RNA affords a fitness score, a metric of affinity and selectivity. By comparing the fitness scores for an RNA motif for various small molecules, we can gain insight into selectivity. For example, an RNA motif that has a fitness score of 100 for small molecule A, 20 for small molecule B, and 5 for small molecule C is more selective than an RNA motif that has fitness scores of 100, 80, and 40.

Inforna integrates advances in RNA structure determination and prediction, 2DCS selection data, and the fitness of interactions identified by 2DCS (via StARTS analysis) to provide lead small molecules for an RNA target of interest. Because Inforna uses RNA motif–small molecule interactions as a basis set for lead small molecule identification, it can be considered a “bottom-up” approach to drug design. The only input required by Inforna is a .CT (connectivity table) file that describes the RNA’s secondary structure (Figure 2A). The file is a simple text file that is automatically generated by structure prediction programs. Various databases that house RNA sequence and structure information also provide .CT files, including RNase P database, RNA Strand, and the Comparative RNA Web Site and Project (CRW). Alternatively, a .CT file can be created manually (see: http://mfold.rna.albany.edu/doc/mfold-manual/node11.php#CT). Inforna parses the .CT file into its composite motifs, compares those motifs to the RNA motif–small molecule database, and outputs lead compounds (Figure 2B,C).

Herein, we describe our updated Inforna platform that incorporates RNA motif–small molecule binding partners reported in the scientific literature, the ability to complete chemical similarity searching, and an improved user interface. Further, we analyzed the small molecules chemoinformatically and RNA motif bioinformatically to provide a global view of the types of structures that can be presently targeted and chemotypes that impart affinity for RNA. The data are housed on a server that is freely available on the web to generate lead
RESULTS AND DISCUSSION

Updating Inforna with RNA–Small Molecule Interactions Reported in the Scientific Literature. After an extensive search of the literature, we identified all known small molecules that bind RNA. These 44 reports include data from comprehensive reviews on small molecules that target RNA by Thomas and Hergenrother and Chow and Bogdan.31 A list of the small molecules and brief descriptions are provided in the Supporting Information. In total, 233 small molecules and 1130 RNA motif–small molecule interactions were added to the database, which previously contained 11 small molecules and 806 RNA motif–small molecule interactions (Table 1). A user guide is provided in the Supporting Information. Notably, the previous version of the database was employed in a target agnostic approach to target human microRNA precursors, affording a hit rate of 44%.26

Organization of the Searchable Inforna Database. A schema of the database is shown in Figure S-1. The database contains a list of all RNA motif–ligand interactions identified by 2DCS or by other methods in MySQL. Each entry is assigned: (i) a unique small molecule identifier, (ii) the motif type (bulge, internal loop, hairpin, multibranch loop), (iii) a unique motif identifier, (iv) the motif sequence, (v) the closing base pair(s), and (vi) other related meta-data. Note: the functional form for internal loops is “A×B,” where A indicates the number of 5′ unpaired nucleotides and B indicates the number of 3′ unpaired nucleotides. Examples of functional forms for multibranch loops include “A×B×C” and “A×B×C×D,” which indicate a three- or four-way junction, respectively.

Inforna Algorithm. Once a .CT file has been submitted for lead identification, Inforna parses the file into its composite motifs and compares it to the database. The user can specify: (i) search loop nucleotides with closing base pairs (stringent), that is, the motif in the target RNA must match a motif in the database exactly, or (ii) search loop nucleotides without closing base pairs (only loop nucleotides must match an entry in the database; less stringent; Figure 2B). Additionally, Inforna utilizes RNA ambiguity codes (N) to enable searching of positional variations within the query motif sequence.

Inforna Output. The results of the search can be viewed within a web browser or exported as an MS Excel spreadsheet. The output includes the structure of the lead small molecule, the corresponding loop that the compound is predicted to bind, a fitness score, a link to a chemical similarity searching tool, and the PMID/PMCID that reports the interaction (Figure 2C).
Table 1. Comparison of the Inforna Database before and after Updating, Including a Summary of the Number of RNA Motifs, Small Molecules, and RNA Motif-Small Molecule Interactions Added

<table>
<thead>
<tr>
<th></th>
<th>Inforna 1.0</th>
<th>Inforna 2.0</th>
<th>fold increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 × 1 internal loop</td>
<td>112</td>
<td>156</td>
<td>1.4</td>
</tr>
<tr>
<td>2 × 2 internal loop</td>
<td>106</td>
<td>216</td>
<td>2.0</td>
</tr>
<tr>
<td>3 × 3 internal loop</td>
<td>144</td>
<td>284</td>
<td>2.0</td>
</tr>
<tr>
<td>4 × 4 internal loop</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1 × 2 internal loop</td>
<td>19</td>
<td>21</td>
<td>1.1</td>
</tr>
<tr>
<td>3 × 2 internal loop</td>
<td>37</td>
<td>41</td>
<td>1.1</td>
</tr>
<tr>
<td>4 × 3 internal loop</td>
<td>45</td>
<td>46</td>
<td>1.0</td>
</tr>
<tr>
<td>large asymmetric internal loop</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>3-nt. hairpin</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4-nt. hairpin</td>
<td>63</td>
<td>155</td>
<td>2.5</td>
</tr>
<tr>
<td>5-nt. hairpin</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6-nt. hairpin</td>
<td>101</td>
<td>271</td>
<td>2.7</td>
</tr>
<tr>
<td>large hairpins</td>
<td>0</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>1-nt. bulge</td>
<td>41</td>
<td>87</td>
<td>2.1</td>
</tr>
<tr>
<td>2-nt. bulge</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3-nt. bulge</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>large bulges</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>base pairs</td>
<td>6</td>
<td>15</td>
<td>2.5</td>
</tr>
<tr>
<td>total motifs</td>
<td>686</td>
<td>1331</td>
<td>1.9</td>
</tr>
<tr>
<td>small molecules</td>
<td>11</td>
<td>244</td>
<td>22.2</td>
</tr>
<tr>
<td>total RNA motif–small molecule interactions</td>
<td>806</td>
<td>1936</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Each small molecule in the Inforna database was queried with the PubChem structure search tool found at http://pubchem.ncbi.nlm.nih.gov/search/. This searches the PubChem depository and identifies chemically similar molecules as determined by the Tanimoto equation and a binary substructure fingerprint, which measure the three-dimensional structural similarity between two molecules. The input format is SMILES, and the Tanimoto threshold percent was set to ≥95. Similarity searching can be used to identify compounds that are structurally similar to the lead compound predicted by Inforna that may have improved properties. Thus, a cohort of compounds can be quickly identified and tested for modulation of RNA (dys)function.

Analysis of RNA Motifs in the Inforna Database. A summary of the motifs contained in the database before and after our update is presented in Table 1. Overall, the number of motifs increased from 686 to 1331, while the number of RNA motif–small molecule interactions increased from 806 to 1936. Importantly, new types of motifs were added including two- and three-nucleotide bulges, three- and five-nucleotide hairpins, large asymmetric loops, and large hairpins. RNA motifs that bind various small molecules are provided in Table S-1 and Figure S-2.

We recently completed an analysis of the secondary structures of all human miRNA precursors, other human RNAs with known structures, and RNAs with known structures from other organisms. Interestingly, small bulges, internal loops, and hairpins prevail in all three data sets. We compared the distribution of motifs in these RNAs to the motifs in our Inforna database. At present, Inforna effectively covers small internal loops and hairpins. However, additional data that define RNA motif–small molecule interactions with bulges (n = 3 to n = 7), larger internal loop (4 nucleotides or more on the 5′ or 3′ side), and larger hairpins (n = 7 to n = 10) will be required to effectively target cellular RNAs.

Analysis of Small Molecules in the Inforna Database. Prior to our large update, Inforna contained 11 small molecules of the aminoglycoside or benzimidazole class. The update increased the number of small molecules to 244, which were broadly categorized as aminoglycosides, aminoglycoside derivatives, heterocyclics, amino acids, peptides, cofactors, metal complexes, quinones, and miscellaneous. These broad classes were further divided into subclasses, such as acridine derivatives and benzimidazoles under the umbrella class heterocycles (Supporting Information). These broad classes were not biased for binding to a single RNA motif type. Therefore, the molecules were resorted, this time by the motif to which they bind (bulge, internal loop, hairpin, etc.) as opposed to by their chemical structures. The molecules of each motif category were analyzed for common scaffolds using the program Scaffold Hopper. A pooled population comparison was then completed to identify common scaffolds that occur more frequently in each motif type as compared to their occurrence in all small molecules housed in the database. No statistically significant scaffolds were identified that are privileged for binding internal loops. In contrast, statistically significant scaffolds were identified for bulges and hairpins (Figure 3A). For bulges, these scaffold include naphthalene (p = 0.021), 1-benzyl-cyclopenta[b]naphthalene (p = 0.017), indene (p = 0.008), 1-benzyl-indene (p = 0.012), 2-(cyclohexyloxy)oxane (p = 0.041), and oxazolidin-2-one (p = 0.035). Statistically significant, privileged scaffolds for hairpins are 2-(cyclohexyloxy)-oxane (p = 0.005) and 2-((tetrahydrofuran-2-yl)oxy)-cylohexanol (p = 0.001).

Next, dissimilarity scores were calculated for the small molecules that bind each motif type. This analysis compares the shape-based similarity scores of two molecules. Matrices were constructed to compare each compound to every other compound that binds a particular motif according to the Tanimoto dissimilarity function. From these scores, heat maps were constructed (Figure 3B). The average Tanimoto dissimilarity score for each motif is internal loops, 0.6934; hairpins, 0.6825; bulges, 0.7550; and base pairs, 0.7492, indicating various degrees of diversity that bind different RNA motifs. Taken together, these data suggest that the small molecules that recognize various RNA motifs can be structurally diverse.

Using Privileged Scaffolds to Develop RNA-Focused Libraries. Designing compounds that target RNA is difficult, in large part because little is known about chemotypes that impart affinity and selectivity. The RNA–small molecule interactions identified in our laboratory and reported in the literature can begin to provide insight into features of small molecules that afford a productive, selective interaction with RNA. In addition to the chemotypes mentioned above, a previous study identified that 2-phenyl indole, 2-phenyl benzimidazole, and pyridinium chemotypes are privileged for binding RNA. As more privileged scaffolds are identified, they can be used to refine current small molecule screening decks to create RNA-focused small molecule libraries.

Applications of Inforna. We have reported our bottom-up approach to modulate dysfunction of RNAs that are linked with human disease. Most of our studies have aimed at expanded repeating RNAs (denoted r(sequence)) that cause or contribute to microsatellite disorders including myotonic muscular dystrophy (DM), amyotrophic lateral sclerosis and frontotemporal dementia (ALS/FTD), fragile X-associated tremor ataxia syndrome (FXTAS), and fragile X syndrome.
These disorders affect the neuromuscular systems of millions of people worldwide.\(^\text{41}\)

DM is caused by \(\text{r(CUG)}^{\text{exp}}\) and \(\text{r(CCUG)}^{\text{exp}}\) and causes toxicity via binding and sequestering key RNA-binding proteins as well as producing toxic proteins via repeat-associated non-ATG (RAN) translation.\(^\text{42-44}\) One of our very first selections identified that a kanamycin A derivative (\(6^\prime\)-N-5-hexynoate kanamycin A; kan hex) prefers to bind 2 \(\times\) 2 nucleotide, pyrimidine-rich internal loops like those found in \(\text{r(CCUG)}^{\text{exp}}.\)\(^\text{21}\) Indeed, kan hex and multivalent compounds thereof that target multiple loops simultaneously improve defects associated with \(\text{r(CCUG)}^{\text{exp}}\) in cellular models, the first small molecules identified to do so.\(^\text{45}\) Using a model of the kanamycin derivative bound to \(\text{r(CCUG)}^{\text{exp}}\) in cellular models, the first small molecules identified to do so.\(^\text{45}\) Using a model of the kanamycin derivative bound to \(\text{r(CCUG)}^{\text{exp}}\) in cells using a Huisgen dipolar cycloaddition reaction, or click chemistry.\(^\text{46}\) That is, \(\text{r(CCUG)}^{\text{exp}}\) is the catalyst for a cellular click reaction, the first demonstration of this powerful chemistry to template an inhibitor in cells.

Interestingly, Informa also identified overlap between kan hex and the 1 \(\times 1\) nucleotide UU loops that periodically repeat in \(\text{r(CUG)}^{\text{exp}},\) albeit with a lower fitness score. By altering the distance between kanamycin RNA-binding modules in multivalent compounds, we were able to design a compound that is selective for \(\text{r(CUG)}^{\text{exp}}\) over \(\text{r(CCUG)}^{\text{exp}}\) and improves DM-associated defects in cellular and animal models of disease.\(^\text{35}\) An additional small molecule binder was identified for \(\text{r(CCUG)}^{\text{exp}},\) the bis-benzimidazole H. As observed for kan hex, multivalent H compounds improve defects in cellular models of DM, although H itself is inactive.\(^\text{34}\) A chemical similarity search using H as a lead compound afforded H1, which is bioactive in a DM mouse model.\(^\text{48}\) Thus, there are multiple methods in which lead compounds identified by Informa can be optimized.

More recently, we identified a compound that binds the expanded RNA repeat that causes ALS/FTD, or \(\text{r}(\text{G}_{4}\text{C}_{2})^{\text{exp}}.\)\(^\text{37}\) Like the repeats that cause DM, \(\text{r}(\text{G}_{4}\text{C}_{2})^{\text{exp}}\) causes nuclear inclusions and produces toxic proteins via RAN translation.\(^\text{39-53}\)

Our designed small molecule inhibits RAN translation and formation of nuclear foci, providing a chemical tool to probe ALS/FTD biology.\(^\text{37}\) Indeed, another small molecule that we designed against \(\text{r(CGG)}^{\text{exp}},\) linked to both FXTAS and FXS, elucidated an RNA-mediated mechanism of gene silencing.\(^\text{38,40}\) FXS is caused by a loss of fragile X mental retardation protein (FMRP), which is encoded by the fragile X mental retardation 1 (\(\text{FMR1}\)) gene. The exact mechanism by which FMRP loss occurs is of the utmost importance for understanding FXS disease pathology and for developing therapeutic agents. Our designed small molecule allowed interrogation into disease pathology, proving that FMRP loss is caused by gene silencing/chromatin remodeling induced by an RNA:DNA hybrid formed when \(\text{r(CGG)}^{\text{exp}}\) in the \(5^\prime\) untranslated region of \(\text{FMR1}\) RNA binds to the \(\text{FMR1}\) gene.\(^\text{41}\) Our small molecule thermodynamically stabilizes the \(\text{r(CGG)}^{\text{exp}}\) repeat and blocks its binding to the DNA.\(^\text{41}\) This mechanism of gene silencing revealed new targets for chemical biology probes—the d(CCG) repeating DNA in \(\text{FMR1}\) that is bound by \(\text{r(CGG)}^{\text{exp}}\) and the RNA–DNA hybrid complex that causes gene silencing. By specifically targeting each with a small molecule, their roles in gene silencing could be further elucidated.

Our Informa approach, however, is not restricted to expanded repeating RNAs. We applied Informa to human microRNAs (miRNAs) using a target agnostic approach to study and manipulate miRNA biology.\(^\text{26}\) That is, the small molecules dictate the RNA target as determined by 2DCS and STARTS.

MiRNAs are transcribed as precursors, pri-miRNAs, which are processed by the nuclear nuclease Drosha to produce pre-miRNAs. Once processed, the pre-miRNAs are translocated to the cytoplasm and are then processed by the nuclease Drosha to generate mature miRNAs. Mature miRNAs bind to the 3′ untranslated regions (UTRs) of mRNAs and silence the transcript via several mechanisms.\(^\text{54}\) Properly functioning miRNAs regulate essential cellular processes including differentiation,
programmed cell death, metabolism, etc. Thus, aberrant expression or dysregulation of miRNAs can cause disease, in particular cancer.11−13

The sequences and secondary structures of all known precursor miRNAs in the human transcriptome were downloaded from miRBase61 and their structures modeled by free energy minimization.11 Precursor miRNAs fold into small hairpin secondary structures that can be predicted from sequence.26 The entire set of secondary structures was then parsed by Informa, where the output generated includes the targetable motifs in each pri- and pre-miRNA and the corresponding lead small molecules that bind them. Next, we refined the lead interactions based on the following criteria: (i) the targetable motif must be in a Drosha or Dicer processing site, which are cleaved to produce pre-miRNAs and mature miRNAs, respectively,35 and (ii) the miRNA must be causative of disease. In summary, lead small molecules (26 total) were identified for 22 different miRNA precursors.26

The activities of all lead small molecules were assessed in cell lines by measuring mature miRNA expression by quantitative real-time RT-PCR. A significant reduction in expression of the desired miRNA target was observed for 11 of the 26 lead compounds, i.e., a hit rate of 44%.26 Importantly, these data establish that bioactive small molecules that target RNA can be designed in a transcriptome-wide manner without target bias (i.e., this approach is target agnostic). Among 26 lead interactions, our STARTS statistical approach36 predicted that the most fit interaction is between a benzimidazole (1) and miR-96.

Indeed, we showed that compound 1 selectively blocks the biogenesis of miR-96, up-regulates its downstream protein target (FOXO1), and induces apoptosis in MCF7 breast cancer cells.26 Importantly, 1 allowed us to probe several aspects of miRNA biology.26 First, miR-96, -182, and -183 precursors are transcribed from a single pri-miRNA transcript.64 We found that 1 inhibits Drosha processing of miR-96 but not that of miR-182 or -183. Thus, there is no cooperativity between Drosha processing sites. Second, these analyses established that inhibition of miR-96 biogenesis by 1 selectively affects downstream targets and pathways. One target of miR-96 is FOXO1, a pro-apoptotic transcription factor. Although FOXO1 is targeted by several miRNAs (miR-96, -182, and -27a),65 we showed that inhibiting miR-96 is sufficient to selectively upregulate FOXO1 and promote apoptosis of cancer cells.26 Further, 1 selectively modulates apoptosis via the miR-96−FOXO1 pathway as siRNA directed against FOXO1 mRNA ablated the apoptotic effect of 1.26

Finally, we established the selectivity of 1 by profiling 149 other miRNAs in MCF7 cells. Compound 1 only significantly affects miR-96 biogenesis, providing an unparalleled level of selectivity. Moreover, 1 is more selective than an oligonucleotide antagonist that targets sequence.26 Collectively, these studies illustrate that small molecules targeting RNAs can be extremely selective and affect downstream processes such as apoptosis.

Troubleshooting. We envision that Informa can be employed to identify lead small molecules for any RNA target of interest. Below, we discuss potential avenues if small molecule leads do not bind, bind without requisite affinity and selectivity, or do not modulate the intended target in cells due to localization to an undesired compartment or due to poor permeability.

If lead compounds predicted by Informa (or otherwise identified) do not bind avidly or do not possess the desired level of selectivity, they can be optimized in various ways. For example, a traditional medicinal chemistry approach can be employed in which derivatives of the lead compound are studied to define structure−activity relationships. In the same vein, chemical similarity searching, which has been incorporated into Informa 2.0, could identify compounds with similar structural features and three-dimensional shapes. Focusing on small molecules that contain privileged chemotypes (described above) could facilitate both medicinal chemistry and chemical similarity approaches.

We have successfully employed a modular assembly approach to increase the affinity and selectivity of RNA-binding modules. In this approach, two targetable motifs within close proximity are bound simultaneously by a single compound that displays the corresponding RNA-binding modules. That is, small molecules that bind adjacent motifs in the target RNA are tethered together. Multivalency has been used to increase affinity and selectivity of compounds targeting other biomolecules, including DNA66 and proteins.67,68 Indeed, the affinity and selectivity of a multivalent compound that targets RNA is controlled by the nature of the scaffold that displays the RNA-binding modules, valency, and the distance between RNA-binding modules.73

In some cases, a high affinity, selective binder will not modulate the desired RNA target in cells. Various factors could cause a lack of bioactivity including targeting of a nonfunctional site (binding has no downstream effect), inadequate cellular permeability, and compound localization to an undesired compartment. Chemotypes have been identified that improve cellular uptake70 including conjugation of polyarginine,71 as well as those that influence localization.72 Such data could be incorporated into a traditional medicinal chemistry approach or chemical similarity searching. Interestingly, the permeability73 and localization of multivalent compounds can be improved by changing the scaffold, in particular the module used to separate RNA-binding modules, and by the RNA-binding module itself.74

Summary and Outlook. Informa is a “bottom-up” strategy for lead identification to target RNA. Indeed, it has enabled the design of small molecules and modularly assembled compounds that modulate dysfunction of RNAs that are linked with human disease. Most of our studies have aimed at expanded repeating RNAs that cause or contribute to microsatellite disorders including DM34−36 ALS/FTD37 FXTAS38,39 and FXS.40 We also applied Informa to human miRNAs to study and manipulate miRNA biology.26 That is, the small molecules dictate the RNA target as determined by 2DCS and STARTS. Importantly, these data establish that bioactive small molecules that target RNA can be designed in a transcriptome-wide manner without target bias (i.e., this approach is target agnostic).

As our recent update has doubled the size of the database, diversifying the small molecules as well as the RNA motifs, we envision that a larger subset of cellular RNAs will be able to be targeted using Informa. In the near future, Informa will be autopopulated with 2DCS selection data and allow submission of interactions from other laboratories and creation of data silos for the secure housing of searchable data from individual users.

We also previously developed a method that combines high throughput screening methods with 2DCS, which will increase the number of known interactions that comprise our RNA motif−small molecule database.53 We are studying the binding of RNA motif libraries and diverse chemotypes to expand the database and provide the best lead small molecules possible. In addition, small molecule binding sites in cells could be
identified by using Chemical Cross-Linking and Isolation by Pull-Down (ChemCLIP) in which small molecules are appended with a reactive moiety that forms a covalent cross-link with cellular targets that it binds, providing a metric of cellular selectivity and relative affinity.

**METHODS**

**Common Scaffolds Analysis.** Common scaffolds were identified for each motif by using NCGC Scaffold Hopper Version 1.0 (http://tripod.nih.gov). A pooled population comparison was then completed to determine if a scaffold was privileged for binding a certain motif. That is, the frequency of a given scaffold within small molecules that bind a motif was compared to the frequency of the same scaffold in all small molecules in the database. Equations 1 and 2 below afford Zobs a measure of statistical significance that was then converted into a two-tailed p value.

\[
\phi = \frac{(n_1 p_1 + n_2 p_2)}{n_1 + n_2}
\]

\[
Z_{obs} = \frac{(p_1 - p_2)}{\sqrt{\phi(1-\phi)(\frac{1}{n_1} + \frac{1}{n_2})}}
\]

where \(n_1\) is the size of population 1 (number of small molecules that bind a certain motif), \(n_2\) is the size of population 2 (number of small molecules in the database), \(p_1\) is the observed proportion of population 1 (fraction of small molecules that bind a certain motif and have the scaffold of interest), and \(p_2\) is the observed proportion for population 2 (fraction of small molecules in the database that have the scaffold of interest).

**Dissimilarity Scores.** Tanimoto dissimilarity scores were calculated by using the JCDissimilarityCFTanimoto function in JChem for Office (Excel), version S11.4.872 (ChemAxon; http://www.chemaxon.com).

**ASSOCIATED CONTENT**

**Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.6b00001.

User manual, description of small molecules in the database including references and their categorization, and supplementary figures (PDF)

**AUTHOR INFORMATION**

**Corresponding Author**

*E-mail: Disney@scripps.edu.*

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We thank J. Bostrom and L. Knerr for helpful suggestions. This work was funded by the National Institutes of Health grants SR01GM097455 (to MDD).

**REFERENCES**


is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72, 257–268.


