Published in final edited form as:

ACS Chem Biol. 2016 June 17; 11(6): 1720–1728. doi:10.1021/acscchembio.6b00001.

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[†]

Matthew D. Disney: Disney@scripps.edu

[†]Department of Chemistry, The Scripps Research Institute, 130 Scripps Way, Jupiter, Florida 33458, United States

[‡]Informatics Core, The Scripps Research Institute, 130 Scripps Way, Jupiter, Florida 33458, United States

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 Folding and Function

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

Correspondence to: Matthew D. Disney, Disney@scripps.edu.

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

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

The authors declare no competing financial interest.

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.(1-3) 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 RNAstructure.(4-8) 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.(9, 10) On average, the lowest free energy structures predicted via free energy minimization are ~75% accurate;(8, 11) other folding algorithms are 50–70% accurate.(4) The accuracy of prediction can be greatly improved by including experimental constraints from enzymatic or chemical mapping.(11-14) Other methods of RNA structure determination include phylogenetic comparison,(15, 16) NMR spectroscopy,(17, 18) and X-ray crystallography.(19, 20)

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).(21, 22) 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 competitor 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.(23)

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).(24, 25) 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.(26) 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,(27) RNA Strand,(28) and the Comparative RNA Web Site and Project (CRW).(29) 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 motifs 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 compounds for an RNA of interest (<http://www.scripps.edu/disney/software.html>).

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(30) 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 motif found in the target RNA, the corresponding motif in the database, the lead small molecule(s) (Simplified Molecular Input Line Entry System (SMILES) notation and chemical structure), a link to chemical similarity searching of the lead small molecule, fitness of the RNA motif–small molecule interaction in the database,

dissociation constant or EC50 if measured, and a link to PMID/PMCID that reports the interaction (Figure 2C).

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.⁽³²⁾ 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, heterocycles, 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)cyclohexanol ($p = 0.001$).

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.(33) 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 Informa

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 $r(\text{sequence})^{\text{exp}}$) that cause or contribute to microsatellite disorders including myotonic muscular dystrophy (DM),(34–36) amyotrophic lateral sclerosis and frontotemporal dementia (ALS/FTD),(37) fragile X-associated tremor ataxia syndrome (FXTAS),(38, 39) and fragile X syndrome (FXS).(40) These disorders affect the neuromuscular systems of millions of people worldwide.(41)

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

catalyst for a cellular click reaction, the first demonstration of this powerful chemistry to template an inhibitor in cells.

Interestingly, Inforna also identified overlap between kan hex and the 1×1 nucleotide UU loops that periodically repeat in $r(\text{CUG})^{\text{exp}}$, 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 $r(\text{CUG})^{\text{exp}}$ over $r(\text{CCUG})^{\text{exp}}$ (47) and improves DM-associated defects in cellular and animal models of disease.(35) An additional small molecule binder was identified for $r(\text{CUG})^{\text{exp}}$, the bis-benzimidazole **H**. As observed for kan hex, multivalent **H** compounds improve defects in cellular models of DM, although **H** itself is inactive.(34) A chemical similarity search using **H** as a lead compound afforded **H1**, which is bioactive in a DM mouse model.(48) Thus, there are multiple methods in which lead compounds identified by Inforna can be optimized.

More recently, we identified a compound that binds the expanded RNA repeat that causes ALS/FTD, or $r(\text{G}_4\text{C}_2)^{\text{exp}}$.(37) Like the repeats that cause DM, $r(\text{G}_4\text{C}_2)^{\text{exp}}$ causes nuclear inclusions and produces toxic proteins via RAN translation.(49-53) Our designed small molecule inhibits RAN translation and formation of nuclear foci, providing a chemical tool to probe ALS/FTD biology.(37) Indeed, another small molecule that we designed against $r(\text{CGG})^{\text{exp}}$, linked to both FXTAS and FXS, elucidated an RNA-mediated mechanism of gene silencing.(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 (*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 $r(\text{CGG})^{\text{exp}}$ in the 5' untranslated region of *FMR1* RNA binds to the *FMR1* gene.(40) Our small molecule thermodynamically stabilizes the $r(\text{CGG})^{\text{exp}}$ repeat and blocks its binding to the DNA.(40) This mechanism of gene silencing revealed new targets for chemical biology probes—the d(CCG) repeating DNA in *FMR1* that is bound by $r(\text{CGG})^{\text{exp}}$ 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 Inforna approach, however, is not restricted to expanded repeating RNAs. We applied Inforna to human microRNAs (miRNAs) using a target agnostic approach to study and manipulate miRNA biology.(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 Dicer to generate mature miRNAs. Mature miRNAs bind to the 3' untranslated regions (UTRs) of mRNAs and silence the transcript via several mechanisms.(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.(55-60)

The sequences and secondary structures of all known precursor miRNAs in the human transcriptome were downloaded from miRBase(61) and their structures modeled by free energy minimization.(11) Precursor miRNAs fold into small hairpin secondary structures that can be predicted from sequence.(62) The entire set of secondary structures was then parsed by Inforna, 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,(63) 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 approach(24) 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 antagomir 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 Inforna 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 Inforna (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 DNA(66) 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.(36, 47, 69)

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 uptake,(70) 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 permeability(73) 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 DM, (34-36) ALS/FTD,(37) FXTAS,(38, 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.(33) 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,(37, 75) 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. (1)(2)where n1 is the size of population 1 (number of small molecules that bind a certain motif), n2 is the size of population 2 (number of small molecules in the database), p1 is the observed proportion of population 1 (fraction of small molecules that bind a certain motif and have the scaffold of interest), and p2 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 5.11.4.872 (ChemAxon; <http://www.chemaxon.com>).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Boström and L. Knerr for helpful suggestions. This work was funded by the National Institutes of Health grants 5R01GM097455 (to MDD).

References

1. Berg, JM.; Tymoczko, JL.; Stryer, L. Biochemistry. 6th. W.H. Freeman and Company; New York: 2007.
2. Johnson LF, Abelson HT, Penman S, Green H. The relative amounts of the cytoplasmic RNA species in normal, transformed and senescent cultured cell lines. J Cell Physiol. 1977; 90:465–470. DOI: 10.1002/jcp.1040900310 [PubMed: 192738]
3. Johnson LF, Williams JG, Abelson HT, Green H, Penman S. Changes in RNA in relation to growth of the fibroblast. III. Posttranscriptional regulation of mRNA formation in resting and growing cells. Cell. 1975; 4:69–75. DOI: 10.1016/0092-8674(75)90135-X [PubMed: 1078787]

4. Eddy SR. How do RNA folding algorithms work? *Nat Biotechnol.* 2004; 22:1457–1458. DOI: 10.1038/nbt1104-1457 [PubMed: 15529172]
5. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31:3406–3415. DOI: 10.1093/nar/gkg595 [PubMed: 12824337]
6. Zuker M, Sankoff D. RNA secondary structures and their prediction *Bull. Math Biol.* 1984; 46:591–621. DOI: 10.1007/BF02459506
7. Mathews DH. RNA secondary structure analysis using RNAstructure. *Curr Protoc Bioinformatics.* 2014; 46:12.16.11–12.16.25. DOI: 10.1002/0471250953.bi1206s46
8. Lorenz R, Bernhart SH, Honer Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. *Algorithms Mol Biol.* 2011; 6:26.doi: 10.1186/1748-7188-6-26 [PubMed: 22115189]
9. Markham NR, Zuker M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol.* 2008; 453:3–31. DOI: 10.1007/978-1-60327-429-6_1 [PubMed: 18712296]
10. Mathews DH. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA.* 2004; 10:1178–1190. DOI: 10.1261/rna.7650904 [PubMed: 15272118]
11. 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; 101:7287–7292. DOI: 10.1073/pnas.0401799101 [PubMed: 15123812]
12. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature.* 2013; 505:696–700. DOI: 10.1038/nature12756 [PubMed: 24270811]
13. Wildauer M, Zemora G, Liebeg A, Heisig V, Waldsich C. Chemical probing of RNA in living cells. *Methods Mol Biol.* 2014; 1086:159–176. DOI: 10.1007/978-1-62703-667-2_9 [PubMed: 24136603]
14. Wells SE, Hughes JM, Haller Igel A, Ares M Jr. Use of dimethyl sulfate to probe RNA structure in vivo. *Methods Enzymol.* 2000; 318:479–493. DOI: 10.1016/S0076-6879(00)18071-1 [PubMed: 10890007]
15. Han K, Kim HJ. Prediction of common folding structures of homologous RNAs. *Nucleic Acids Res.* 1993; 21:1251–1257. DOI: 10.1093/nar/21.5.1251 [PubMed: 7681944]
16. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A.* 1985; 82:6955–6959. DOI: 10.1073/pnas.82.20.6955 [PubMed: 2413450]
17. Hennig M, Williamson JR, Brodsky AS, Battiste JL. Recent advances in RNA structure determination by NMR. *Curr Protoc Nucleic Acid Chem.* 2001; Unit 7.7. doi: 10.1002/0471142700.nc0707s02
18. Scott LG, Hennig M. RNA structure determination by NMR. *Methods Mol Biol.* 2008; 452:29–61. DOI: 10.1007/978-1-60327-159-2_2 [PubMed: 18563368]
19. Ferre-D'Amare AR. Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods.* 2010; 52:159–167. DOI: 10.1016/j.jymeth.2010.06.008 [PubMed: 20554048]
20. Ferre-D'Amare AR, Zhou K, Doudna JA. A general module for RNA crystallization. *J Mol Biol.* 1998; 279:621–631. DOI: 10.1006/jmbi.1998.1789 [PubMed: 9641982]
21. Disney MD, Childs-Disney JL. Using selection to identify and chemical microarray to study the RNA internal loops recognized by 6'-N-acylated kanamycin A. *ChemBioChem.* 2007; 8:649–656. DOI: 10.1002/cbic.200600569 [PubMed: 17394189]
22. Disney MD, Labuda LP, Paul DJ, Poplawski SG, Pushechnikov A, Tran T, Velagapudi SP, Wu M, Childs-Disney JL. Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. *J Am Chem Soc.* 2008; 130:11185–11194. DOI: 10.1021/ja803234t [PubMed: 18652457]
23. Aminova O, Disney MD. A microarray-based method to perform nucleic acid selections. *Methods Mol Biol.* 2010; 669:209–224. DOI: 10.1007/978-1-60761-845-4_17 [PubMed: 20857369]

24. Velagapudi SP, Seedhouse SJ, Disney MD. Structure-activity relationships through sequencing (STARTS) defines optimal and suboptimal RNA motif targets for small molecules. *Angew Chem, Int Ed.* 2010; 49:3816–3818. DOI: 10.1002/anie.200907257
25. 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; 133:10111–10118. DOI: 10.1021/ja200212b [PubMed: 21604752]
26. Velagapudi SP, Gallo SM, Disney MD. Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nat Chem Biol.* 2014; 10:291–297. DOI: 10.1038/nchembio.1452 [PubMed: 24509821]
27. Brown JW. The Ribonuclease P Database. *Nucleic Acids Res.* 1999; 27:314.doi: 10.1093/nar/27.1.314 [PubMed: 9847214]
28. Andronescu M, Bereg V, Hoos HH, Condon A. RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC Bioinf.* 2008; 9:340.doi: 10.1186/1471-2105-9-340
29. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinf.* 2002; 3:2.doi: 10.1186/1471-2105-3-2
30. Thomas JR, Hergenrother PJ. Targeting RNA with small molecules. *Chem Rev.* 2008; 108:1171–1224. DOI: 10.1021/cr0681546 [PubMed: 18361529]
31. Chow CS, Bogdan FM. A structural basis for RNA–ligand interactions. *Chem Rev.* 1997; 97:1489–1514. DOI: 10.1021/cr960415w [PubMed: 11851457]
32. Liu B, Childs-Disney JL, Znosko BM, Wang D, Fallahi M, Gallo SM, Disney MD. Analysis of secondary structural elements in human miRNA hairpin precursors. *BMC Bioinf.* 2015; 17:112.doi: 10.1186/s12859-016-0960-6
33. Tran T, Disney MD. Identifying the preferred RNA motifs and chemotypes that interact by probing millions of combinations. *Nat Commun.* 2012; 3:1125.doi: 10.1038/ncomms2119 [PubMed: 23047683]
34. 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 potentially bioactive. *ACS Chem Biol.* 2012; 7:856–862. DOI: 10.1021/cb200408a [PubMed: 22332923]
35. Childs-Disney JL, Parkesh R, Nakamori M, Thornton CA, Disney MD. Rational design of bioactive, modularly assembled aminoglycosides targeting the RNA that causes myotonic dystrophy type 1. *ACS Chem Biol.* 2012; 7:1984–1993. DOI: 10.1021/cb3001606 [PubMed: 23130637]
36. Rzuczek SG, Gao Y, Tang ZZ, Thornton CA, Kodadek T, Disney MD. Features of modularly assembled compounds that impart bioactivity against an RNA target. *ACS Chem Biol.* 2013; 8:2312–2321. DOI: 10.1021/cb400265y [PubMed: 24032410]
37. Su Z, Zhang Y, Gendron TF, Bauer PO, Chew J, Yang WY, Fostvedt E, Jansen-West K, Belzil VV, Desaro P, Johnston A, Overstreet K, Oh SY, Todd PK, Berry JD, Cudkowicz ME, Boeve BF, Dickson D, Floeter MK, Traynor BJ, Morelli C, Ratti A, Silani V, Rademakers R, Brown RH, Rothstein JD, Boylan KB, Petrucelli L, Disney MD. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron.* 2014; 83:1043–1050. DOI: 10.1016/j.neuron.2014.07.041 [PubMed: 25132468]
38. Disney MD, Liu B, Yang W, Sellier C, Tran T, Charlet-Berguerand N, Childs-Disney JL. A small molecule that targets r(CGG)^{exp} and improves defects in fragile X-associated tremor ataxia syndrome. *ACS Chem Biol.* 2012; 7:1711–1718. DOI: 10.1021/cb300135h [PubMed: 22948243]
39. Tran T, Childs-Disney JL, Liu B, Guan L, Rzuczek S, Disney MD. Targeting the r(CGG) repeats that cause FXTAS with modularly assembled small molecules and oligonucleotides. *ACS Chem Biol.* 2014; 9:904–912. DOI: 10.1021/cb400875u [PubMed: 24506227]
40. Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, Disney MD, Jaffrey SR. Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science.* 2014; 343:1002–1005. DOI: 10.1126/science.1245831 [PubMed: 24578575]
41. Brouwer JR, Willemsen R, Oostra BA. Microsatellite repeat instability and neurological disease. *BioEssays.* 2009; 31:71–83. DOI: 10.1002/bies.080122 [PubMed: 19154005]

42. Jiang H, Mankodi A, Swanson MS, Moxley RT, Thornton CA. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum Mol Genet.* 2004; 13:3079–3088. DOI: 10.1093/hmg/ddh327 [PubMed: 15496431]
43. Timchenko NA, Cai ZJ, Welm AL, Reddy S, Ashizawa T, Timchenko LT. RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J Biol Chem.* 2001; 276:7820–7826. DOI: 10.1074/jbc.M005960200 [PubMed: 11124939]
44. Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MA, Nan Z, Forster C, Low WC, Schoser B, Somia NV, Clark HB, Schmechel S, Bitterman PB, Gourdon G, Swanson MS, Moseley M, Ranum LP. Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A.* 2011; 108:260–265. DOI: 10.1073/pnas.1013343108 [PubMed: 21173221]
45. 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; 9:538–550. DOI: 10.1021/cb4007387 [PubMed: 24341895]
46. Rzuczek SG, Park H, Disney MD. A toxic RNA catalyzes the in cellulo synthesis of its own inhibitor. *Angew Chem, Int Ed.* 2014; 53:10956–10959. DOI: 10.1002/anie.201406465
47. Lee MM, Childs-Disney JL, Pushechnikov A, French JM, Sobczak K, Thornton CA, Disney MD. Controlling the specificity of modularly assembled small molecules for RNA via ligand module spacing: targeting the RNAs that cause myotonic muscular dystrophy. *J Am Chem Soc.* 2009; 131:17464–17472. DOI: 10.1021/ja906877y [PubMed: 19904940]
48. Parkesh R, Childs-Disney JL, Nakamori M, Kumar A, Wang E, Wang T, Hoskins J, Tran T, Housman DE, Thornton CA, Disney MD. Design of a bioactive small molecule that targets the myotonic dystrophy type 1 RNA via an RNA motif-ligand database & chemical similarity searching. *J Am Chem Soc.* 2012; 134:4731–4742. DOI: 10.1021/ja210088v [PubMed: 22300544]
49. Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, DeJesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW 3rd, Rademakers R, Boylan KB, Dickson DW, Petrucelli L. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron.* 2013; 77:639–646. DOI: 10.1016/j.neuron.2013.02.004 [PubMed: 23415312]
50. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung GY, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, Feldman H, Knopman DS, Petersen RC, Miller BL, Dickson DW, Boylan KB, Graff-Radford NR, Rademakers R. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011; 72:245–256. DOI: 10.1016/j.neuron.2011.09.011 [PubMed: 21944778]
51. Lee YB, Chen HJ, Peres JN, Gomez-Deza J, Attig J, Stalekar M, Troakes C, Nishimura AL, Scotter EL, Vance C, Adachi Y, Sardone V, Miller JW, Smith BN, Gallo JM, Ule J, Hirth F, Rogelj B, Houart C, Shaw CE. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.* 2013; 5:1178–1186. DOI: 10.1016/j.celrep.2013.10.049 [PubMed: 24290757]
52. Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretschmar HA, Cruts M, Van Broeckhoven C, Haass C, Edbauer D. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science.* 2013; 339:1335–1338. DOI: 10.1126/science.1232927 [PubMed: 23393093]
53. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Rytan M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A, Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita VM, Kaivorinne AL, Holtta-Vuori M, Ikonen E, Sulkava R, Benatar M, Wu J, Chio A, Restagno G, Borghero G, Sabatelli M, Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A,

- Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ. A hexanucleotide repeat expansion in *C9ORF72* is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*. 2011; 72:257–268. DOI: 10.1016/j.neuron.2011.09.010 [PubMed: 21944779]
54. Sharp PA. RNAi and double-strand RNA. *Genes Dev*. 1999; 13:139–141. DOI: 10.1101/gad.13.2.139 [PubMed: 9925636]
55. Calin GA, Croce CM. MicroRNA-cancer connection: the beginning of a new tale. *Cancer Res*. 2006; 66:7390–7394. DOI: 10.1158/0008-5472.CAN-06-0800 [PubMed: 16885332]
56. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer*. 2006; 6:259–269. DOI: 10.1038/nrc1840 [PubMed: 16557279]
57. Gregory RI, Shiekhattar R. MicroRNA biogenesis and cancer. *Cancer Res*. 2005; 65:3509–3512. DOI: 10.1158/0008-5472.CAN-05-0298 [PubMed: 15867338]
58. Hammond SM. MicroRNAs as oncogenes. *Curr Opin Genet Dev*. 2006; 16:4–9. DOI: 10.1016/j.gde.2005.12.005 [PubMed: 16361094]
59. Garofalo M, Condorelli G, Croce CM. MicroRNAs in diseases and drug response. *Curr Opin Pharmacol*. 2008; 8:661–667. DOI: 10.1016/j.coph.2008.06.005 [PubMed: 18619557]
60. Farazi TA, Spitzer JJ, Morozov P, Tuschl T. MiRNAs in human cancer. *J Pathol*. 2011; 223:102–115. DOI: 10.1002/path.2806 [PubMed: 21125669]
61. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 2008; 36:D154–D158. DOI: 10.1093/nar/gkm952 [PubMed: 17991681]
62. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T. A uniform system for microRNA annotation. *RNA*. 2003; 9:277–279. DOI: 10.1261/rna.2183803 [PubMed: 12592000]
63. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116:281–297. DOI: 10.1016/S0092-8674(04)00045-5 [PubMed: 14744438]
64. Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D. MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. *J Biol Chem*. 2007; 282:25053–25066. DOI: 10.1074/jbc.M700501200 [PubMed: 17597072]
65. Guttilla IK, White BA. Coordinate regulation of *FOXO1* by miR-27a, miR-96, and miR-182 in breast cancer cells. *J Biol Chem*. 2009; 284:23204–23216. DOI: 10.1074/jbc.M109.031427 [PubMed: 19574223]
66. Pilch DS, Poklar N, Gelfand CA, Law SM, Breslauer KJ, Baird EE, Dervan PB. Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. *Proc Natl Acad Sci U S A*. 1996; 93:8306–8311. DOI: 10.1073/pnas.93.16.8306 [PubMed: 8710866]
67. Gestwicki JE, Cairo CW, Strong LE, Oetjen KA, Kiessling LL. Influencing receptor-ligand binding mechanisms with multivalent ligand architecture. *J Am Chem Soc*. 2002; 124:14922–14933. DOI: 10.1021/ja027184x [PubMed: 12475334]
68. Kitov PI, Sadowska JM, Mulvey G, Armstrong GD, Ling H, Pannu NS, Read RJ, Bundle DR. Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. *Nature*. 2000; 403:669–672. DOI: 10.1038/35001095 [PubMed: 10688205]
69. Pushechnikov A, Lee MM, Childs-Disney JL, Sobczak K, French JM, Thornton CA, Disney MD. Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J Am Chem Soc*. 2009; 131:9767–9779. DOI: 10.1021/ja9020149 [PubMed: 19552411]
70. Yang NJ, Hinner MJ. Getting across the cell membrane: an overview for small molecules, peptides, and proteins. *Methods Mol Biol*. 2015; 1266:29–53. DOI: 10.1007/978-1-4939-2272-7_3 [PubMed: 25560066]
71. Wender PA, Mitchell DJ, Pattabiraman K, Pelkey ET, Steinman L, Rothbard JB. The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. *Proc Natl Acad Sci U S A*. 2000; 97:13003–13008. DOI: 10.1073/pnas.97.24.13003 [PubMed: 11087855]
72. Zheng N, Tsai HN, Zhang X, Rosania GR. The subcellular distribution of small molecules: from pharmacokinetics to synthetic biology. *Mol Pharmaceutics*. 2011; 8:1619–1628. DOI: 10.1021/mp200092v

73. Yu P, Liu B, Kodadek T. A high-throughput assay for assessing the cell permeability of combinatorial libraries. *Nat Biotechnol.* 2005; 23:746–751. DOI: 10.1038/nbt1099 [PubMed: 15908941]
74. Lee MM, French JM, Disney MD. Influencing uptake and localization of aminoglycoside-functionalized peptoids. *Mol BioSyst.* 2011; 7:2441–2451. DOI: 10.1039/c1mb05074e [PubMed: 21611644]
75. Guan L, Disney MD. Covalent small molecule-RNA complex formation enables cellular profiling of small molecule-RNA interactions. *Angew Chem Int Ed.* 2013; 52:10010–10013. DOI: 10.1002/anie.201301639

Figure 1.

Design of small molecules to target RNA from sequence aided by Infora. A user identifies an RNA of interest and generates a secondary structure whether by prediction, phylogenetic comparison, or experiment. Infora then extracts motifs from the target RNA's secondary structure and compares them to the database, outputting lead small molecules.

Figure 2.

Input required for generation of lead small molecules by Inforna and the program's user interface. (A) Example of a .CT file that is used as input. (B) A user can generate lead small molecules by searching for exact matches in the database, which includes loop nucleotides and closing base pairs, or by searching only for loop nucleotides. (C) Image of the user interface and Inforna output. 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 for the article in which the RNA motif–small molecule interaction was identified.

Figure 3.

Privileged scaffolds for RNA motifs and the diversity of the small molecules in Inforna. (A) Common scaffolds identified for hairpins and bulges. (B) Tanimoto dissimilarity scores between small molecules. Compounds were compared with all other small molecules that bind the same motif, i.e., internal loops, hairpins, bulges, and base pairs. Green indicates highest similarity, while red indicates lowest similarity. The average Tanimoto dissimilarity score for each motif is internal loops, 0.6934; hairpins, 0.6825; bulges, 0.7550; and base pairs, 0.7492.

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

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