Design and Implementation of an Ribonucleic Acid (RNA) Directed Fragment Library[||]

Khaled Bodoor,[†,#] Vamsi Boyapati,[‡,#] Vikram Gopu,[‡,#] Marietta Boisdore,[‡] Kiran Allam,[§] Janae Miller,[‡] W. Dale Treleaven,[§] Thomas Weldeghiorghis,[§] and Fareed Aboul-ela*[,‡]

Departments of Biological Sciences and Chemistry, Louisiana State University, Baton Rouge, Louisiana 70803, and Department of Physics, University of Jordan, Amman 11942, Jordan

Received January 16, 2009

The design of RNA binding ligands is complicated by issues of specificity, target flexibility, and the tractability of known RNA inhibitors toward chemical derivitization. To address these difficulties, an RNA-directed fragment compound library is presented. We began with an analysis of 120 small molecules with reported RNA-binding activity. Calculated physical and chemical properties for the RNA ligands are comparable to those of ligands for established protein drug targets. To ensure that our library contained RNA-binding functionalities that might not be detected by the above comparisons, 114 fragment compounds were purchased on the basis of similarity to substructures of RNA ligands. Five "hits" were identified for the decoding site from the bacterial ribosome by NMR. These included fragments derived from A-site binding ligands but also compounds not previously identified as A-site binders. Hits generated in this manner can be used to probe the interaction surface of RNA and its conformational plasticity, facilitating structure-based optimization.

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm9000659

Introduction

Strong evidence exists for the importance of RNA as a drug target.[1,2] X-ray structures of clinically used antibiotics bound to ribosomal subunits demonstrated that these drugs contact primarily RNA.[3−5] Structured elements in untranslated regions of viral mRNAs[a,6−8] are needed for viral growth. Genomes of higher organisms contain a surprising amount of noncoding RNA, which is not translated into proteins.[9] The roles so far discovered for these RNAs raise the possibility that small molecule drugs targeted at RNA could be used to treat cancer, neurological, and genetic diseases.[10−12] Small molecule RNA-binding ligands may be useful as "knockdown" probes for RNA function. Ligands that can modify or cleave a target RNA may be even more potent as therapeutics and functional probes.[13,14]

Efforts to target RNA molecules from HIV[15−18] and the bacterial ribosome[19−21] produced a number of novel chemical entities with potent binding. However, existing drugs that target RNA have been identified through serendipity.

The polyanionic and highly solvated character of RNA presents unique challenges for medicinal chemistry.[2] RNA conformational flexibility complicates structure-based drug design,[15] and, particularly, the application of in silico docking to RNA.[22−25]

In recent years, fragment-based strategies combined structural biology with careful library design to develop viable clinical drug candidates targeted against proteins.[26−28] These strategies begin with careful design of a screening "fragment" library consisting of "leadlike" molecules.[29−31] Typically, fragment compounds are of low molecular weight (<300), contain two or more polar groups, and a moderate predicted log *P* (where *P* is the partition coefficient for the compound in a water/octanol mixture). Fragments are screened for weak binding to the molecular target. In contrast to standard in silico docking protocols, a rigid, preformed target conformation is not assumed. As a framework for computational structure-based design, a set of fragment hit/target complex structures is superior to a single apo-structure.

The fragment approach has potential to address the distinct challenges posed by RNA. It can probe conformational flexibility. Cheminformatics tools used to design libraries can be applied to the special medicinal chemistry issues presented by RNA. A handful of studies discovered RNA-binding ligands by screening generic fragment libraries against RNA targets.[32−35] However, none of the reported RNA fragment screens utilized a fragment library specifically designed to enhance the probability of binding to RNA.

Here, we report the design of an RNA-directed fragment library. We assembled a database of 120 small molecule ligands that, according to literature, bind directly to an RNA. We show that known RNA binding ligands, in their calculated physicochemical properties, span a range comparable to those of ligands for kinases and proteases. We therefore chose 114 RNA-focused fragments via direct incorporation of compounds with similarity to substructures of RNA ligands. We screened the library for binding to a 27 nucleotide oligonucleotide incorporating the binding site for aminoglycoside antibiotics[36] and obtained a set of hits, including at least two compounds that had not been previously reported as ribosomal decoding site binders. Our RNA-directed fragment library provides a powerful tool for identifying novel chemical starting points for design of RNA-binding ligands.

Experimental Details

RNA-Binding Ligand Database. The RNA-binding ligand database was assembled by searching published literature for reports of small molecules binding to any RNA molecule. A small molecule

[||] We dedicate this manuscript to the memory of our friend and colleague, Dr. Kiran Allam.

* To whom correspondence should be addressed. Phone: 225-578-2791. Fax: 225-578-7258. E-mail: faboul@lsu.edu.

[†] University of Jordan.

[#] These authors contributed equally to this work.

[‡] Department of Biological Sciences, Louisiana State University.

[§] Department of Chemistry, Louisiana State University.

[a] Abbreviations: NMR, nuclear magnetic resonance, QC, quality control; water-LOGSY, water-ligand observed via gradient spectroscopy; NCI, National Cancer Institute; UPMCLD, University of Pittsburgh Center for Chemical Libraries and Library Development; 1D, one-dimensional; MHz, megahertz; TPSA, topological polar surface area, MOE, Molecular Operating Environment (software); STD, saturation transfer difference spectroscopy; rRNA, ribosomal RNA; mRNA, messenger RNA.

10.1021/jm9000659 CCC: $40.75 © 2009 American Chemical Society
Published on Web 05/15/2009

was defined as a nonpeptidic or peptido-mimetic ligand, although natural products such as thiazole antibiotics, which contain peptidic linkages, were included. Ligands were only included when the reported K_d is less than 50 μM. Where a series of chemically similar compounds was reported to bind to an RNA, the ligand with the strongest reported affinity was chosen as a representative, together with any other member of the series that contained a chemically distinctive substituent, provided that the latter fit the K_d criteria. The full list of compounds together with references is included in the Supporting Information.

Descriptor Comparison of Ligand Sets. Seven ligand databases were analyzed: the RNA-binding ligand database, the NCI/DTP Open Chemical Repository (NCI) diversity set prepared for use with AutoDock, the 327 compound University of Pittsburgh Center for Chemical Libraries and Library Development (UPCMLD) diversity set, the 211 nonredundant entries from the first 300 entries of the Ashgate Drug Index (purchased from CamSoft), a set of 626 protein-binding ligands, a set of 162 kinase binding ligands and 353 protease binding ligands, the latter three identified from the PDB as described (see below). Each ligand set was imported into MOE (Molecular Operating Environment, CCG) as a set of SD files. The NCI and UPCMLD libraries were downloaded as SD files, while the RNA-binding and Ashgate sets were manually drawn using ChemDraw and exported as SD files. Protein binding ligands were downloaded from the PDB Web site as sdf files.

In a few cases, outliers were observed in descriptor values that distorted overall distribution statistics for specific descriptors. The descriptors in question were recalculated with the outlying ligand excluded. This exclusion was applied only when (1) the mean descriptor value for the ligand set was changed by greater than a standard deviation as a result of the value calculated for a single ligand and (2) similar ligands had very different descriptor values, suggesting that the program may have encountered a problem in executing the calculation in question. It was never necessary to exclude more than one ligand from any given database for any given calculation, except for the protein-binding ligand set, for which five SD files that generated nonphysical chemical connectivities had to be excluded.

Fragmenting of Data-Basing Ligands. Each reported RNA-binding ligand was subdivided using ChemDraw. Fragmentation points were chosen at "linker" positions next to rings. The primary criterion for fragmentation corresponded roughly to prioritization rules 2 and 3 for scaffold formation as proposed by Schuffenhauer and colleagues.[37] However, as our purpose was to preserve all potential RNA-binding entities from all fragments, rather than preserving only a scaffold, other prioritization rules were not relevant. Where multiple bonds separated two ring moieties, the linker segment itself was treated as a fragment. In general, we aimed to obtain products that fit the molecular weight criteria for leadlike fragments (MW < 250). In a few cases (e.g., tetracycline) polycyclic rings did not present obvious fragmentation points for products this small. In those instances, the polycyclic ring was broken into components containing, for example, two to three rings each, but the original compound was also included as a fragment. Very large macrocycles, such as those occurring in some macrolides, were retained intact for the clustering analysis.

PDB coordinate sets for 25 of the database ligands bound to their respective RNA targets were examined to determine which ligand fragments were in contact with RNA. Every fragment of every RNA binding ligand examined was found to have contacts within 3 Å heavy atom to heavy atom distance to the RNA. Because of this finding, we proceeded with the assumption that virtually all fragments of RNA binding ligands are likely to make contact with RNA, and no ligand fragments were discarded on the basis of structural contact criteria.

Clustering of Fragments of RNA-Binding Ligands. The descriptors for clustering and related details are available in the Supporting Information.

A representative fragment was chosen from each descriptive cluster. At that stage, because of the overlap between components of the descriptive and fingerprint clusters, some fingerprint clusters were already represented. One fragment was then chosen to represent each of the original 53 fingerprint clusters that was not already represented. Thirteen more fragments were chosen to represent fingerprint subclusters that were at that stage still unrepresented.

Selection, Plating, and Quality Control (QC) of Library Compounds. These are described in Supporting Information.

Screening of Fragments for Binding to the Bacterial Ribosomal A Site RNA. Instrumentation used and buffer conditions were as described for QC in Supporting Information. One-hundred-two compounds were screened in 29 mixtures of 3−4 per tube, containing the same buffer as the QC spectra but with 10 μM A-site RNA. Running the experiment with excess ligand has the advantage of amplifying the bound signal.[38] A competitor with known binding to the A-site RNA, gentamycin, was added to a concentration of 30 μM, and experiments were repeated with identical parameters for each mixture. Spectra from a hit compound are identified from the following characteristics: a positive signal in the water-LOGSY spectrum (320 scans, 1.2 ms mixing period) and a reduction in signal in the T_2 filter experiment (128 scans, 300 ms T_2 filter period) relative to the 1 d experiment with excitation sculpting. We expected that hits that bind specifically to the aminoglycoside binding site show a reduction in positive signal in the water-LOGSY and some recovery of signal in the T_2 filter experiment when gentamycin is added. However, gentamycin is a relatively weak affinity ligand compared to competitor ligands commonly utilized in fragment screening against kinases. Ligands that showed small or ambiguous indications of competition with gentamycin were further tested as described below.

Follow-Up of Fragment Hits. Hit candidates as identified from the library screen were tested directly for binding to the A-site RNA. Buffer conditions were 10 mM NaCl, 10 mM Na phosphate pH 6.7, 0.1 mM EDTA in 700 μL of 90% H_2O, 10% D_2O. Compounds were included singly as shown at 200−300 μM. Water-LOGSY, T_2-filter, and 1D Watergate and excitation sculpting experiments were acquired on a 700 MHz Varian NMR spectrometer with an HCN cold probe. A long acquisition (number of transients = 1024) 1D Watergate spectrum with a large sweep width (24 ppm) was performed to obtain a spectrum of the RNA imino resonances.

Results

RNA-Binding Ligand Database. The first step in designing an RNA-focused fragment library is to determine principles that can enhance the probability for selecting RNA-binding compounds. We assembled a list of compounds that have been reported to bind to any RNA with a K_d in the range of 50 μM or stronger. A few ligands were included for which the reported K_d was not available, if a PDB coordinate set existed showing the ligand bound to an RNA. We incorporated compounds from a variety of sources, including antibiotics, riboswitch effectors, and products of in vitro drug design studies. We intended to limit redundancy in the database while including as much of proven RNA-binding chemical space as possible. Compounds that belonged to similar chemical classes were included if they incorporated distinct chemical moieties that were not present elsewhere in the list. Altogether, 121 compounds were included. A list of the RNA-binding ligands, together with documentation regarding the reported binding target and activity of each, is included in Supporting Information (Table S1).

Comparison of RNA-Binding Ligands to Drug, in Silico Screening, and Protein-Binding Ligands. Screening library design typically includes filters that ensure that library compounds fit criteria for "druglikeness" and/or "leadlikeness". At this "proof-of principle" stage in our investigations we do not wish to be limited by these restrictions, and we consider that any designed RNA-binding ligand provides valuable insights into RNA recognition. Nonetheless, we wish to determine whether there is a correlation between physicochemical proper-

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm9000659

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm900659

ties and RNA-binding and whether such properties may conflict with those required for drug design. We used MOE to calculate the distribution of values for a set of conformation-independent physicochemical descriptors for the RNA-binding ligand set. For comparison, we analyzed a similar sized data set derived from the Ashgate Drug Index (CamSoft). These compounds are in commercial use as therapeutic drugs. In addition, we ran a similar analysis on two publicly available "diversity" (from the National Cancer Institute (NCI) and the University of Pittsburgh Center for Chemical Methodologies and Library Development (UPMCSD)) sets commonly used for virtual screening/in silico docking exercises. We also calculated descriptor values for two ligand sets obtained by searching the protein databank (PDB) for small molecules identified as binding to kinases and proteases. Both these classes of proteins include well established drug targets. Finally, we formulated a ligand set of manageable size that could be taken to represent ligands binding to any protein in a "generic" sense (see methods section).

Comparative analysis of these ligand sets is shown in Figure 1 and, in more detail, in the Supporting Information. For all 35 physicochemical descriptors computed, the average value for the RNA-binding set was within a standard deviation of that for each of the other five ligand sets. In a few cases, discussed in more detail in the Supporting Information, there were small differences in the distribution of values for the four databases.

Average values for most 2D descriptors correlate with the molecular weight. Average molecular weight is largest for RNA and protease binding ligands (near 500), whereas average molecular weight is slightly over 300 for the Ashgate and NCI sets. Average values for total surface charge, fractional positive charge, and the topological polar surface area (TPSA)[39] are higher for the kinase binding ligands than expected on the basis of molecular weight.

We had set our threshold for RNA binding activity at a relatively weak level ($K_d < 50\ \mu M$) in order to ensure a thorough exploration of RNA-binding chemical space for our fragment library design below. In order to probe the possibility that by doing so we may be somehow diluting the distinctiveness of our RNA-binding ligands, we repeated the analysis with a subset of 67 ligands with reported K_d of less than 3 μM (Table 1). Again, the average value for each descriptor was within a standard deviation of the average for the comparison ligand sets. The average molecular weight is slightly higher for the set of submicromolar RNA ligands than it is for the complete RNA ligand set, as are the average values for the correlated 2D descriptors. This effect, together with a small increase in TPSA, suggests that the large, polar, and chemically complex ligands pick up affinity through a larger surface area of interaction.

Choice of Fragment Library Ligands. Our analysis convinced us that development of RNA binding ligands is not intrinsically incompatible with drug development, but it gave us no obvious "rules" for choosing compounds with propensity to bind RNA. This result does not preclude the likelihood that a finite set of chemical moieties constitute a "privileged"[30] RNA binding set. The special properties of these functionalities may be too subtle or complex to detect using standard descriptors.

We proceeded with a strategy (Figure 2) to build our library as directly as possible from fragments of compounds with proven RNA-binding propensity. Data base ligand structures were "cleaved" in silico (see Experimental Details). Altogether approximately 250 fragments were identified. Many of these fragments are chemically similar. Therefore, we clustered the fragments, first, based upon chemical descriptors and, second, based upon molecular fingerprints using MOE

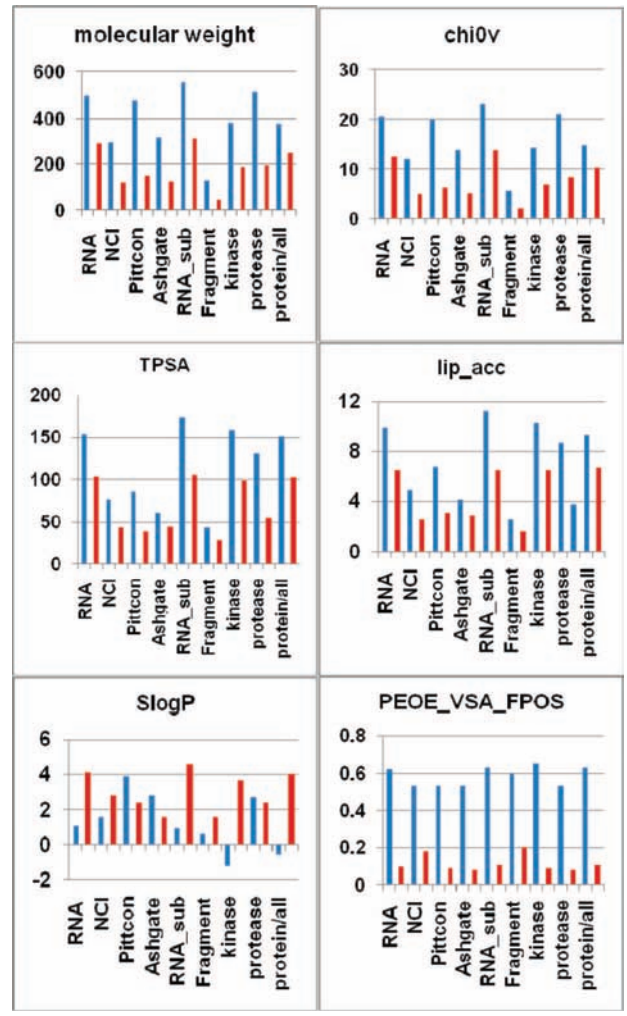


Figure 1. Average values and standard deviations for physicochemical descriptors for nine ligand sets. For each ligand set, two bars are shown: the mean value is shown on the left and standard deviation on the right. RNA: 125 ligands reported to bind RNA with $K_d < 50\ \mu M$. NCI: 1944 NCI diversity set compounds. Pittcon: 327 UPMCSD virtual screening set. Ashgate: 211 Ashgate Drug Index compounds. RNA_sub: 67 compound subset of RNA ligands with reported $K_d < 3\ \mu M$. Fragments: 109 compounds that underwent QC for RNA-directed fragment library. Kinase: 162 kinase-binding ligands identified from PDB. Protease: 282 protease-binding ligands identified from PDB. Protein: 628 protein binding ligands identified from PDB. The following plots are shown (from left to right): (a) molecular weight, (b) Chi0v (zero order atomic valence connectivity index)[53], (c) topological polar surface area[39], (d) hydrogen bond acceptors as defined by Lipinski[54], (e) simulated log of the partition coefficient in a water/octanol mixture, (f) positively charged fraction of the summed partial charge.

Table 1. Clustering Statistics

parameter	value
no. of descriptive clusters	55
no. of fingerprint clusters	53
subclusters of fingerprint cluster 3	71
sublusters of fingerprint cluster 74	19
total RNA-binding ligand fragments	260
total clusters + subclusters	198
compounds purchased	114
fragments passing QC	100

(Figure 3). Altogether this process produced 55 descriptive clusters and 53 fingerprint clusters. We wished to represent each cluster in our library, but we noted that a disproportionate fraction of the original fragments were represented in two fingerprint clusters (Table 1, Figure 3, Table S3). We

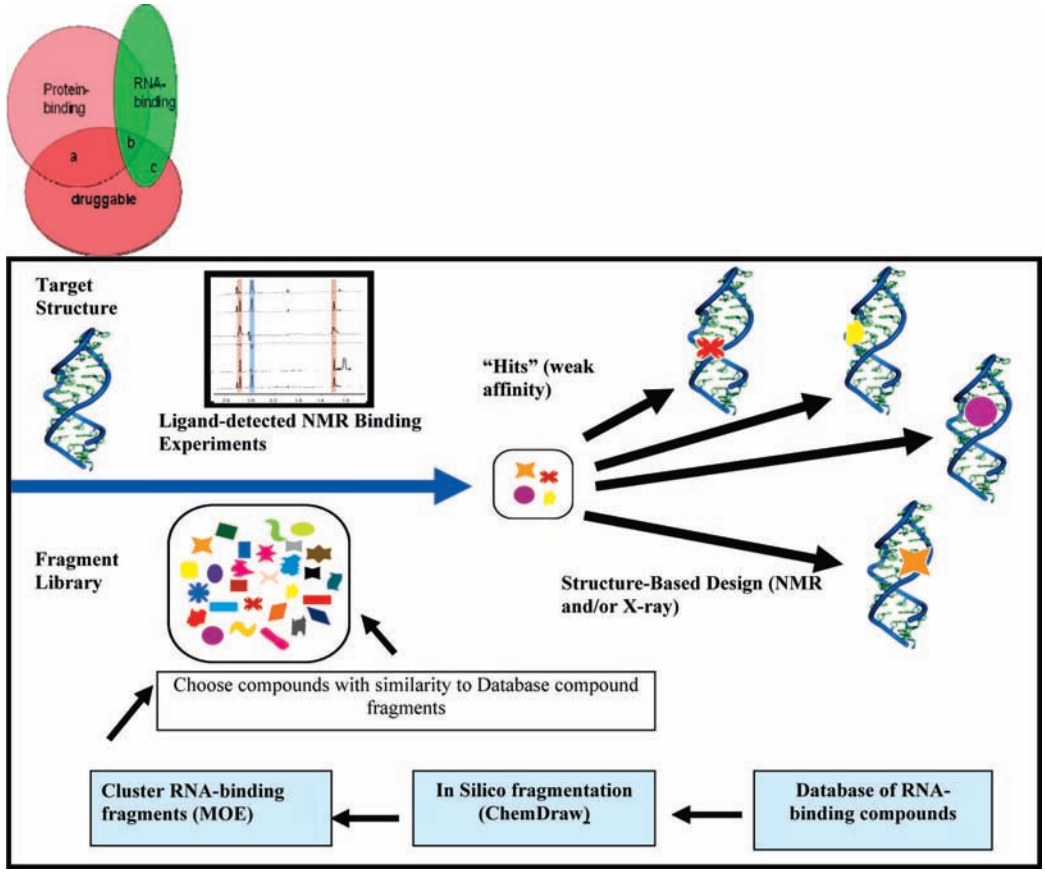


Figure 2. (a) Schematic illustration of the need for selecting RNA-directed compounds for screening for binding to RNA. While we can assume that there is some overlap between the regions of "chemical space" with the propensity to bind to RNA or to proteins, it is likely that the probabilities for binding to the two classes of target will differ in any given region. Generic and commercial screening libraries, designed to cover the overlap between protein-binding and "druglike" chemical spaces, yield low hit rates for RNA targets. Note simplifications: for example, regions of RNA-binding space (and other subregions) are not necessarily contiguous. (b) Schematic of overall strategy for selection of an RNA-directed fragment library and for screening the library against RNA. Fragment library compounds are derived from the database of RNA-binding ligands. NMR is used to screen the library for compounds that bind with weak affinity to the RNA target. The complexes of these compounds with the RNA go into structural studies and provide a starting point for structure-based design of more potent ligands.

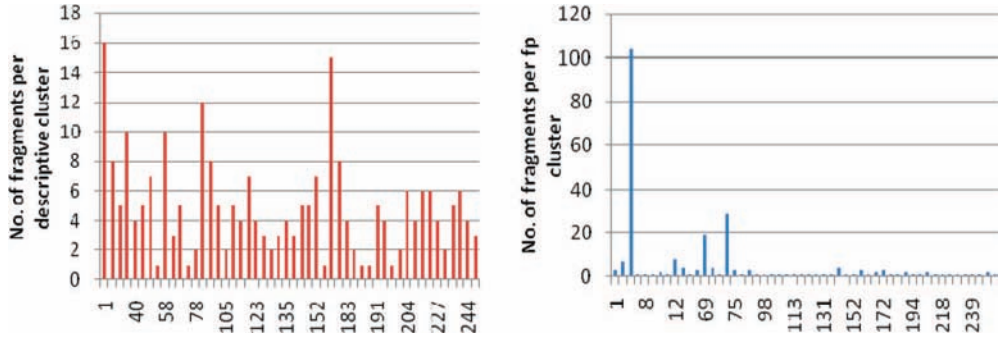


Figure 3. (a) Distribution of number of fragments per cluster derived from fragmenting RNA-binding ligand structures in silico and clustering the resulting fragments according to physicochemical descriptors. (b) Distribution of number of fragments per cluster derived from fragmenting RNA-binding ligand structures in silico and clustering the resulting fragments according to chemical fingerprints. Over half of the fragments were clustered together in either cluster 3 or 71. These two clusters were further subclustered (see text).

judged that these clusters may represent a region of chemical space with particularly favorable RNA-binding propensity. We therefore further clustered these ligand subsets with the aim of achieving a "denser" representation of the corresponding chemical space regions.

We searched for at least one purchasable compound that could represent each cluster, and some subclusters, which had been identified through the above process. Figure 4 shows an example of a descriptive cluster together with the original ligand fragments composing the cluster and the purchasable fragment(s) chosen to represent the cluster. A more exhaustive catalogue of clusters and purchased fragments for the library is included in the Supporting Information (Table S3).

Altogether, 102 compounds were initially selected, purchased, survived quality control, and were plated as described in Experimental Details.

Cheminformatics Analysis and Quality Control of the RNA-Directed Fragment Library. We subjected the compounds that we had chosen for the RNA directed fragment library to the same descriptor analysis that was described

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm900659

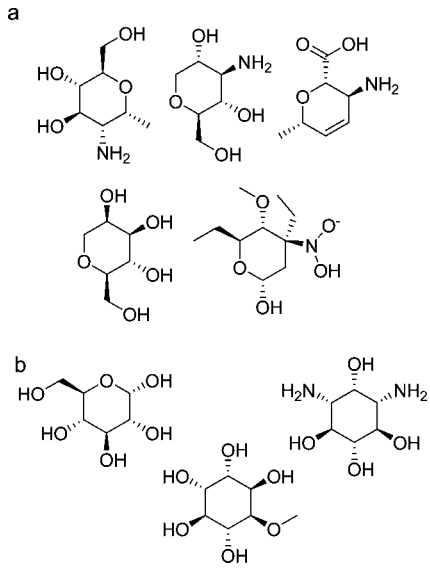


Figure 4. (a) Fragments of RNA binding ligands grouped together as "descriptive cluster 40". (b) Library compounds that represent chemical space surrounding that covered by cluster 40. The fragment on the left has been chosen to represent the given cluster. The other two fragments, representing other descriptive or fingerprint clusters, nonetheless have similarity to a consensus structure represented by cluster 40.

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm9000659

above for the ligand, drug, and diversity sets. Being of smaller molecular weight (on average) than the other sets, the fragment compounds show a more "leadlike"[31] (though still not rigorously leadlike) distribution of descriptor values than any of the ligand sets. This result illustrates one of the important advantages of the fragment strategy. Our fragments are able to span a chemical space derived from bulkier, chemically more complex compounds.

Before being added to RNA for screening, each fragment compound was subjected to a quality control process (described below). Briefly, the integrity and solubility of each compound were tested by inspection of a one-dimensional NMR spectrum. Each compound was also subjected to the full suite of NMR experiments to detect ligand binding. The latter spectra are stored in a database and provide "control" spectra for comparison to screening data.

Screen of Fragment Library for Binding to the Ribosomal Decoding Site ("A-Site") RNA. We used NMR to screen the fragment library for weak affinity binding to a 27-residue oligonucleotide containing the internal loop sequence from the decoding site of the *E. coli* 16S rRNA (the "A-site" RNA). The sequence of the oligonucleotide was identical to that used in previous NMR studies of the A-site structure and its interaction with antibiotics.[36] We chose this target for our first fragment library screen because the same ribosomal site was the target for several of the known RNA-binding ligands that provided the source of our fragment structures.

Altogether, 102 compounds were screened for A-site binding in 29 mixtures of three to four compounds, each at 200 μM concentration, through the acquisition of four 1D NMR experiments on each mixture. Each sample contained 10 μM RNA. The four experiments were chosen to include one that detects signal from bound ligands (water-LOGSY), one that filters out signals from bound ligands (T_2-filter), and two routine spectra acquired with alternative solvent suppression schemes. The screening experiments work on the principle of detecting ligand signals, through the transfer of magnetization properties from bound to free signal via chemical exchange. Figure 5a shows an example of the screening results for one mixture.

Five hit compounds were verified through follow-up experiments performed at higher sensitivity on a 700 MHz spectrometer with a cold probe and a long acquisition time in order to detect imino signals from the RNA. The latter are diagnostic of RNA base pairing. Chemical shift change and/or line broadening in this region indicates ligand contact with the RNA. Figure 5 shows examples of follow-up screening experiments (Figure 5b) and imino spectra (Figure 5c) for hit compounds mixed with A site RNA. The set of hits includes compounds that have been, or that are closely related to compounds that have been, previously reported as A-site binders, and at least two fragments that, to our knowledge, have not been reported as binding to the RNA in question.

The initial screen was followed by a "competition screen" with the known A-site binding ligand, gentamycin, present. The competition experiment indicates which, if any compounds, bind specifically to the known "active site". Some of the hit compounds showed reduced evidence of binding (Table 2 and Figure 5). However, some hits, such as 2-aminoquinoline (**2**), which has been reported as a component of an A-site active site binder, did not show reduced evidence of binding upon addition of competitor. This result may indicate that the competitor ligand does not bind tightly enough to fully displace the fragment hit. Alternatively, the fragment may be capable of binding in the same region, or other parts of the RNA, simultaneously with the competitor. In the case of compound **2**, the result is complicated by competition with acridine (**4**), which happened to be present in the same initial mixture.

Discussion

Properties of RNA-Binding Compounds. The average values for molecular weight, the correlated 2D descriptors, TPSA, sum of partial charges, and fractional positive charge for the RNA ligand set are relatively high. These averages reflect the component of the ligand set that includes bulky, chemically complex natural product antibiotics, including aminoglycosides and macrolides. Though many of these ligands are clinically useful drugs, they do not fit the classical paradigm of druglikeness. TPSA, which is correlated to the number of hydrogen bond donors and acceptors, is exceptionally large for aminoglycosides and macrolides. It is sometimes considered to be a predictor of oral absorption.[40]

Our cheminformatics analysis of RNA-binding ligands did not reveal distinct properties that can be characterized using standard physical and chemical descriptors. Nonetheless, the hypothesis that the properties of RNA-binding ligands are dependent upon the properties of RNA, and that therefore their distribution in chemical space will differ from that of the set of protein-binding ligands, is not disproved. Clearly there is overlap between these regions of chemical space, as many small molecules are known to bind both. Yet medicinal chemists are well aware of chemical functionalities that are commonly found in nucleic acid binding ligands and often select to avoid them. The chemically complex ribosome-binding antibiotics in particular not only violate standard druglikeness rules but also present a huge synthetic challenge. Screening of "generic" fragment libraries against RNA targets is therefore likely to miss some of the most important RNA-binding moieties. Our library is designed to "rescue" these functionalities from their presumably undesirable natural

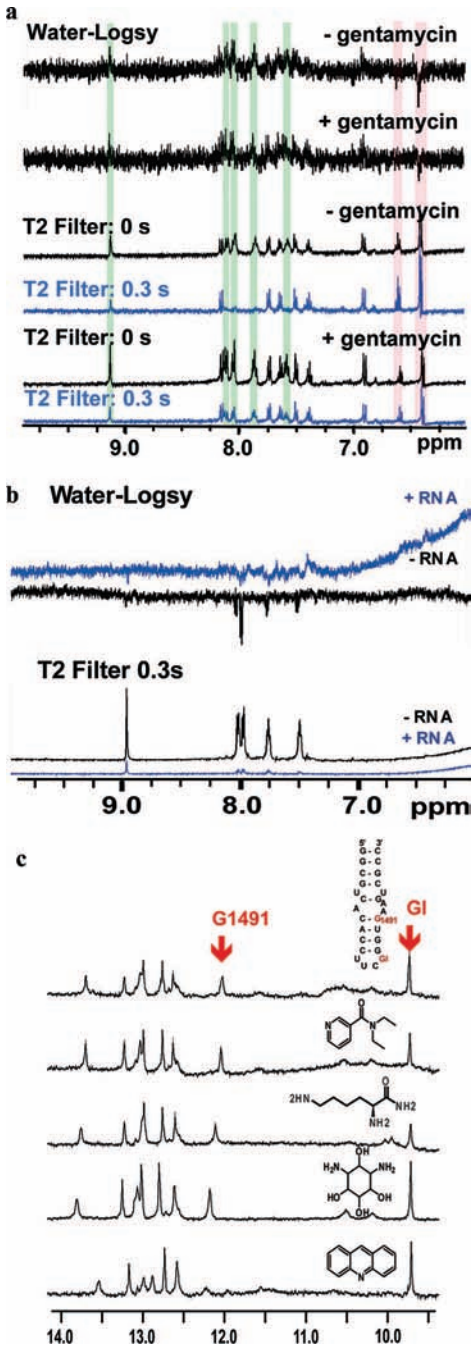
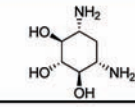
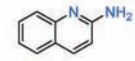
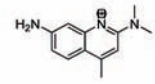
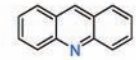
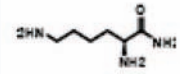


Figure 5. (a) Screening data. The sample contained a mixture including compounds **2** and **4** and two nonbinding compounds, together with 27 nucleotide A-site RNA. From top to bottom: water-LOGSY, without and with gentamycin competitor, 1D spectra with excitation sculpting (T_2-filter period of 0 s) and with (0.3 s) T_2 filter, without gentamycin, and repeat of the excitation sculpting 1D with and without T_2-filter with gentamycin present. Competition of **4** for the gentamycin site is evident from the recovery of signal in the T_2-filter experiment, although little competition is observed in the water-LOGSY. Binding of compound **2** is indicated by signals near 6.9 and 7.3−8.1 pm in the water-LOGSY. A nonbinding compound, a furan, gives negative signals between 6.5 and 6.9 ppm. (b) Follow-up experiment confirming binding of **4** to A-site RNA. Water-LOGSY, T_2-filter, and 1D data are shown with (blue) and without (black) RNA. (c) RNA imino resonances from samples containing 10:1 excess of the respective compounds. Compounds **1**, **4**, and **5** all interact with the RNA, as indicated by shifts in imino resonances, notably in the signal for the G1491 residue (near 12 ppm), indicating that all bind near the active site. The spectrum in the second panel from the top, shows no change from the "free RNA" control, indicating that diethyl nicotinamide does not bind A-site RNA.

Table 2. A-Site Fragment Hits

ID	Structure	Source target	Water Logsy/ competition with gentamycin	T_2 Filter/ competition with gentamycin
1		A-site	+/comp-etitive	+/ comp-etitive
2		A-site	+/non-competitive	+/ non-comp-etitive
3		A-site	Shifted, broad + signal/ competitive	+/ comp-etitive
4		RNA bulge site	+/non-competitive	+/ comp-etitive
5		Ribo-switch	Shifted signal/ competitive	+/ comp-etitive

product context and to present them to RNA targets in a more "leadlike" form. One important qualification to this statement is that at this stage our library is not restricted to compounds that fit classical criteria for leadlikeness. For example, sugar moieties that are not generally considered as good lead candidates are included because of their proven role in RNA recognition. Our fragments are more leadlike than the original ligand sets.

Test Screen of the RNA-Directed Fragment Library. We chose the A-site RNA as our first target to enhance the chances of obtaining hits from "control" fragments. For example, 2-deoxystreptamine[41] **1** and **2**[35] had already been reported in the literature as A-site binding ligands. The latter was discovered from a generic fragment library screen and incorporated into a stronger affinity derivative, which was then included in our original database. Thus, as we anticipated, we have recapitulated interactions discovered in the literature. On the other hand, we also obtained and confirmed A-site RNA binding by **4** and lysinamide **5**. Though the A-site RNA has been subject to intense ligand binding studies and drug design efforts, including a screen against 10 000 compounds in the above-cited study,[35] we could not find a report of binding by these two common compounds. That the strategy depicted in Figure 2, with the RNA-directed fragment library, was able to identify novel ligands for such a well-characterized RNA bodes well for the prospects of identifying novel chemical starting points for other RNA targets.

Screening the RNA Fragment Library against Other RNA Targets. During the course of this study, we learned lessons regarding the process of screening fragments for weak binding to RNA using NMR. Some binding fragments do not produce positive signals in the water-LOGSY (see Figure 5). In these cases, the original screen gave an ambiguous result. Only in the follow-up screen, in which the spectrum of the excess ligand mixed with RNA is compared to that without RNA, does the binding become clear. In that case, the compound

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm900659

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm900659

mixture in the water-LOGSY in the absence of RNA produces negative signals that are reduced when RNA is added. Signals from some ligands become broad in the presence of RNA, leading to poor sensitivity. Ligands containing exchangeable protons give stronger water-LOGSY signals than those that do not contain protons that are capable of exchange with water. The hit rate reported in this study should be seen as a lower limit of that achievable with the library, since potential hits may have been overlooked in the initial screen.

An alternative NMR screening method used by the Abbot group[35] used a 1D imino spectrum, with and without ligand present, as a primary screen for ligand binding. As Figure 5c shows, this can be an effective strategy, but it requires larger amounts of RNA and/or instrumentation of greater sensitivity than the ligand-detecting experiments. A good compromise may be to add an initial ligand-detected screen with compound mixture alone before adding RNA and to follow-up putative hits via the 1D imino measurement. A recent study of in silico hits from a "virtual screen" utilized a saturation transfer difference (STD) experiment to identify RNA binding ligands.[42] In our hands the STD experiment is less sensitive for RNA than water-mediated transferred NOE experiments, as has been reported by others.[32]

Hit rates may be increased by screening against larger RNA targets, such as riboswitches. Larger RNAs form tertiary folds that in turn produce suitable binding pockets for the small ligand. Ligand-detected NMR methods rely on differences in size between free species and bound complex and therefore become more sensitive for large molecular weight targets.

Lessons have also been learned in choice of fragment compounds. We did not impose a lower molecular weight limit. We noted that pyrimidines of molecular weight less than 100, often used as a lower limit in the selection of fragments, have been previously been identified as ribosomal A-site binders. However, in our NMR screen we were unable to reproduce a clear indication of binding to our A-site RNA for these ligands.

Optimization of Fragment Hits. A second stage of ligand design can link hit fragments or synthesize/select larger compounds that contain the initial hits as substructures. A number of quantitative assays, such as fluorescence-based assays,[15,43,44] mass spectrometry,[34] or surface plasmon resonance (SPR),[45] are available to screen the second round of compounds for RNA binding. Ligands that have been identified through assays of ribozyme cleavage activity[46] have not been included in our database, since these measurements do not directly provide a K_d for binding. Nonetheless, these assays may detect relatively weak binding. Moreover, ligands identified in this way clearly bind to RNA and provide additional material for expanding the library.

Conclusion

This study represents a first step toward the development of an RNA-directed cheminformatics. We hope that such a development will be realized in the coming years, exploiting the ligand databases that we and others are assembling. At this time, we have not uncovered a set of clearly definable criteria that can be applied across the range of RNA targets for enhancing the possibility of RNA binding. Nor have we discovered a set of negative criteria that would preclude RNA binding. Therefore, we have chosen our fragments to mimic chemical functionalities already proven to bind to an RNA in one context, in order to test for binding in other contexts.

Screening fragments of the RNA-binding ligands provide an important advantage over screening the original ligands directly, since the fragments will contain smaller polar and hydrophobic surface areas than the original compounds. In addition, starting the drug design process from a fragment hit allows for the design of ligands with specificity for a given RNA target. This approach contrasts with those studies that have worked predominantly with relatively bulky RNA-binding drugs, such as aminoglycosides, to detect binding to other RNA targets.[47−50] The latter approach tends to select for formation of complexes driven by relatively nonspecific electrostatic interactions involving large, polar surface contacts.

Another advantage of the fragments approach is that no prior assumption is made regarding the target structure. By the probing of target flexibility through multiple conformations observed with multiple hit fragments, fragments will complement in silico screening methods recently designed to incorporate RNA induced fit.[25,42]

A drawback to our library is that it is limited to known regions of RNA-binding chemical space. However, though this report concentrates on the first stages of fragment-based ligand discovery, it is important to remember that later stages will "grow" the compound, based on substructure searches and virtual screening from larger purchasable catalogues. These derivatives will sample functionalities not previously known to bind to RNA, when attached to fragment hits that will help to direct the compound toward the RNA binding pocket. A recent study applied template-based de novo design toward the design of TAR RNA binding ligands.[51] The de novo approach and other such methods can be used to derive novel chemical entities from fragment hits. Another study has applied microarray technology to simultaneously explore ligand and RNA chemical space to identify binding partners.[52] By use of fragment hits as starting points for ligand libraries, such methods may identify more druglike ligands.

The RNA fragment library will be pruned and appended, benefiting from the experience of each screen against each new target. Additions to the library based upon pharmacophores derived from the RNA binding ligand database and derived fragment clusters, as well as from analysis of the surfaces of RNA-binding motifs, will lead to the inclusion of new chemical entities to the screening library. Later iterations will also split the library into more and less drug leadlike sections, for the sake of screens that are directed toward purely therapeutic goals or for use as functional probes.

Acknowledgment. We thank Dr. Ron Crouch (Varian, Inc.) for help with implementing automation on the Varian Innova 500 spectrometer. We thank Deepthi Ipparthi for helping to assemble the RNA ligand database. M.B. was supported by Bridge to the Doctorate Program. We thank Dr. Graca Vicente (on behalf of K.A.) for helpful discussions, and Drs. Azeem Hasan and Dan Pu for ESI analysis of hit compounds. This work was supported by startup funds from the Department of Biological Sciences of Louisiana State University, a grant from the Binational Fulbright Commission in Jordan (K.B.), and a Faculty Research Grant (Louisiana State University Office of Research).

Note Added after ASAP Publication. This paper was published on the web May 15, 2009 with a typographical error in the title. The revised version was published on June 3, 2009.

Supporting Information Available: Extended discussion of cheminformatics; extended description of experimental procedures for descriptor comparisons, selection, plating, and quality control of fragments, fragment clustering, RNA binding ligands database (Table S1); average and standard deviations of MOE calculated

descriptors of eight ligand sets (Table S2); RNA directed fragment library compounds (Table S3); list of the 20 RNA-binding ligands with the largest calculated TPSA value; distribution of calculated TPSA for eight ligand sets (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lange, R. P.; Locher, H. H.; Wyss, P. C.; Then, R. L. The targets of currently used antibacterial agents: lessons for drug discovery. *Curr. Pharm. Des.* **2007**, *13*, 3140–3154.

(2) Foloppe, N.; Matassova, N.; Aboul-ela, F. Towards the discovery of drug-like RNA ligands. *Drug Discovery Today* **2006**, *11* (21/22), 1019–1027.

(3) Yonath, A. Antibiotics targeting ribosomes: resistance, selectivity, synergism, and cellular regulation. *Annu. Rev. Biochem.* **2005**, *74* (1), 649–679.

(4) Steitz, T. A. On the structural basis of peptide-bond formation and antibiotic resistance from atomic structures of the large ribosomal subunit. *FEBS Lett.* **2005**, *579*, 955–958.

(5) Ogle, J. M.; Ramakrishnan, V. Structural insights into translational fidelity. *Annu. Rev. Biochem.* **2005**, (74), 129–177.

(6) Karn, J.; Gait, M. J.; Churcher, M. J.; Mann, D. A.; Mikaelian, I. P. C. Control of Human Immunodeficiency Virus Gene Expression by the RNA Binding Proteins Tat and Rev. In *RNA−Protein Interactions*; Nagai, K., Mattai, I. W., Eds.; IRL Press: Oxford, U.K., 1995.

(7) Lemon, S. M.; Honda, M. Internal ribosome entry sites within the RNA genomes of hepatitis C virus and other flaviviruses. *Semin. Virol.* **1997**, *8*, 274–288.

(8) Blackwell, J. L.; Brinton, M. A. BHK cell proteins that bind to the 3′ stem-loop structure of the West Nile virus genome RNA. *J. Virol.* **1995**, *69* (9), 5650–5658.

(9) Amaral, P. P.; Dinger, M. E.; Mercer, T. R.; Mattick, J. S. The eukaryotic genome as an RNA machine. *Science* **2008**, *319* (5871), 1787–1789.

(10) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **2007**, *3* (4), 218–221.

(11) Hagerman, R. J.; Rivera, S. M.; Hagerman, P. J. The fragile X family of disorders: a model for autism and targeted treatments. *Curr. Pediatr. Rev.* **2008**, *4*, 40–52.

(12) Shay, J. W.; Wright, W. E. Telomerase therapeutics for cancer: challenges and new directions. *Nat. Rev. Drug Discovery* **2006**, *5* (7), 577–584.

(13) Burgstaller, P.; Famulok, M. Flavin-dependent photocleavage of RNA at G*U base pairs. *J. Am. Chem. Soc.* **1997**, *119* (5), 1137–1138.

(14) Burgstaller, P.; Hermann, T.; Huber, C.; Westhof, E.; Famulok, M. Isoalloxazine derivatives promote photocleavage of natural RNAs at G.U base pairs embedded within helices. *Nucleic Acids Res.* **1997**, *25* (20), 4018–4027.

(15) Murchie, A. I. H.; Davis, B.; Isel, C.; Afshar, M.; Drysdale, M. J.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T. M.; Mirza, S.; Prescott, C. D.; Vaglio, P.; Aboul-ela, F.; Karn, J. Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J. Mol. Biol.* **2004**, *336* (3), 625–638.

(16) Davis, B.; Afshar, M.; Varani, G.; Murchie, A. I. H.; Karn, J.; Lentzen, G.; Drysdale, M.; Bower, J.; Potter, A. J.; Starkey, I. D.; Swarbrick, T.; Aboul-ela, F. Rational design of inhibtors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J. Mol. Biol.* **2004**, *336* (2), 343–356.

(17) Du, Z.; Lind, K. E.; James, T. L. Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. *Chem. Biol.* **2002**, *9*, 707–712.

(18) Hamy, F.; Brondani, V.; Florsheimer, A.; Stark, W.; Blommers, M. J. J.; Klimkait, T. A new class of HIV-1 Tat antagonist acting through Tat-TAR inhibition. *Biochemistry* **1998**, *37*, 5086–5095.

(19) Zhou, Y.; Gregor, V. E.; Ayida, B. K.; Winters, G. C.; Sun, Z.; Murphy, D.; Haley, G.; Bailey, D.; Froelich, J. M.; Fish, S.; Webber, S. E.; Hermann, T.; Wall, D. Synthesis and SAR of 3,5-diamino-piperidine derivatives: novel antibacterial translation inhibitors as aminoglycoside mimetics. *Bioorg. Med. Chem. Lett.* **2007**, *17* (5), 1206–1210.

(20) Franceschi, F.; Duffy, E. M. Structure-based drug design meets the ribosome. *Biochem. Pharmacol.* **2006**, *71* (7), 1016–1025.

(21) Kung, P. P.; Casper, M. D.; Cook, K. L.; Wilson-Lingardo, L.; Risen, L. M.; Vickers, T. A.; Ranken, R.; Blyn, L. B.; Wyatt, J. R.; Cook, P. D.; Ecker, D. J. Structure−activity relationships of novel 2-substituted quinazoline antibacterial agents. *J. Med. Chem.* **1999**, *42* (22), 4705–4713.

(22) Morley, S. D.; Afshar, M. Validation of an empirical RNA−ligand scoring function for fast flexible docking using RiboDock. *J. Comput.-Aided Mol. Des.* **2003**, *18* (3), 189–208.

(23) Lind, K. E.; Du, Z.; Fujinaga, K.; Peterlin, B. M.; James, T. L. Structure-based computational database screening, in vitro assay, and NMR assessment of compounds that target RNA. *Chem. Biol.* **2002**, *9*, 185–193.

(24) Detering, C.; Varani, G. Validation of automated docking programs for docking and database screening against RNA drug targets. *J. Med. Chem.* **2004**, *47* (17), 4188–4201.

(25) Guilbert, C.; James, T. L. Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J. Chem. Inf. Model.* **2008**, *48* (6), 1257–1268.

(26) Hubbard, R. E.; Davis, B.; Chen, I.; Drysdale, M. J. The SeeDs approach: integrating fragments into drug discovery. *Curr. Top. Med. Chem.* **2007**, *7* (16), 1568–1581.

(27) Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discovery* **2007**, *5* (3), 211–219.

(28) Jhoti, H.; Cleasby, A.; Verdonk, M.; Williams, G. Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Curr. Opin. Chem. Biol.* **2007**, *11* (5), 485–493.

(29) Fejzo, J.; Lepre, C. A.; Peng, J. W.; Bemis, G. W.; Ajay; Murcko, M. A.; Moore, J. M. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* **1999**, *6* (10), 755–769.

(30) Hajduk, P. J.; Bures, M.; Praestgaard, J.; Fesik, S. W. Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.* **2000**, *43* (18), 3443–3447.

(31) Baurin, N.; Aboul-ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. Design and characterization of molecular fragments for use in NMR screening against protein tagets. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2157–2166.

(32) Johnson, E. C.; Feher, V. A.; Peng, J. W.; Moore, J. M.; Williamson, J. R. Application of NMR SHAPES screening to an RNA target. *J. Am. Chem. Soc.* **2003**, *125* (51), 15724–15725.

(33) Swayze, E. E.; Jefferson, E. A.; Sannes-Lowery, K. A.; Blyn, L. B.; Risen, L. M.; Arakawa, S.; Osgood, S. A.; Hofstadler, S. A.; Griffey, R. H. SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J. Med. Chem.* **2002**, *45* (18), 3816–3819.

(34) Seth, P. P.; Miyaji, A.; Jefferson, E. A.; Sannes-Lowry, K. A.; Osgood, S. A.; Propp, S. S.; Ranken, R.; Massire, C.; Sampath, R.; Ecker, D. J.; Swayze, E. E.; Griffey, R. H. SAR by MS: Discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA domain. *J. Am. Chem. Soc.* **2005**, *48*, 7099–7102.

(35) Yu, L.; Oost, T. K.; Schkeryantz, J. M.; Yang, J.; Janowick, D.; Fesik, S. W. Discovery of aminoglycoside mimetics by NMR-based screening of *Escherichia coli* A-site RNA. *J. Am. Chem. Soc.* **2003**, *125* (15), 4444–4450.

(36) Fourmy, D. R., M. I.; Blanchard, S. C.; Puglis, J. D. Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **1996**, *274*, 1367–1371.

(37) Schuffenhauer, A.; Ertl, P.; Roggo, S.; Wetzel, S.; Koch, M. A.; Waldmann, H. The scaffold tree. Visualization of the scaffold universe by hierarchical scaffold classification. *J. Chem. Inf. Model.* **2007**, *47* (1), 47–58.

(38) Stockman, B. J.; Dalvit, C. NMR screening techniques in drug discovery and drug design. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *41* (3−4), 187–231.

(39) Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43* (20), 3714–3717.

(40) Hou, T.; Wang, J.; Zhang, W.; Xu, X. ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. *J. Chem. Inf. Model.* **2007**, *47* (1), 208–218.

(41) Griffey, R. H.; Sannes-Lowery, K. A.; Drader, J. J.; Mohan, V.; Swayze, E. E.; Hofstadler, S. A. Characterization of low-affinity complexes between RNA and small molecules using electrospray ionization mass spectrometry. *J. Am. Chem. Soc.* **2000**, *122* (41), 9933–9938.

(42) Pinto, I. G.; Guilbert, C.; Ulyanov, N. B.; Stearns, J.; James, T. L. Discovery of ligands for a novel target, the human telomerase RNA, based on flexible-target virtual screening and NMR. *J. Med. Chem.* **2008**, *51*, 7205–7215.

(43) Jenne, A.; Hartig, J. S.; Piganeau, N.; Tauer, A.; Samarsky, D. A.; Green, M. R.; Davies, J.; Famulok, M. Rapid identification and characterization of hammerhead-ribozyme inhibitors using fluorescence-based technology. *Nat. Biotechnol.* **2001**, *19* (1), 56–61.

(44) Blount, K.; Puskarz, I.; Penchovsky, R.; Breaker, R. Development and application of a high-throughput assay for glmS riboswitch activators. *RNA Biol.* **2006**, *3* (2), 20–24.

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm900659

(45) Wong, C.-H.; Hendrix, M.; Priestley, E. S.; Greenberg, W. A. Specificity of aminoglycoside antibiotics for the A-site of the decoding region of ribosomal RNA. *Chem. Biol.* **1998**, *5*, 397–406.
(46) Mayer, G.; Famulok, M. High-throughput-compatible assay for glmS riboswitch metabolite dependence. *ChemBioChem* **2006**, *7* (4), 602–604.
(47) Faber, C. S., H.; Schweimer, K.; Rosch, P. Structural rearrangements of HIV-Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* **2000**, *275* (27), 20660–20666.
(48) Ennifar, E.; Paillart, J.-C.; Bodlenner, A.; Walter, P.; Weibel, J.-M.; Aubertin, A.-M.; Pale, P.; Dumas, P.; Marquet, R. Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell. *Nucleic Acids Res.* **2006**, *34* (8), 2328–2339.
(49) Staple, D. W.; Venditti, V.; Niccolai, N.; Elson-Schwab, L.; Tor, Y.; Butcher, S. E. Guanidinoneomycin B recognition of an HIV-1 RNA helix. *ChemBioChem* **2008**, *9* (1), 93–102.
(50) Anupam, R.; Denapoli, L.; Muchenditsi, A.; Hines, J. V. Identification of neomycin B-binding site in T box antiterminator model RNA. *Bioorg. Med. Chem.* **2008**, *16* (8), 4466–4470.
(51) Schüller, A.; Suhartono, M.; Fechner, U.; Tanrikulu, Y.; Breitung, S.; Scheffer, U.; Göbel, M. W.; Schneider, G. The concept of template-based de novo design from drug-derived molecular fragments and its application to TAR RNA. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 59–68.
(52) Disney, M. D.; Labuda, L. P.; Paul, D. J.; Poplawski, S. G.; Pushechnikov, A.; Tran, T.; Velagapudi, S. P.; Wu, M.; Childs-Disney, J. L. Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. *J. Am. Chem. Soc.* **2008**, *130* (33), 11185–11194.
(53) Hall, L. H.; Kier, L. B. The molecular connectivity chi indexes and kappa shape indexes in structure-property modeling. *Reviews in Computational Chemistry* **1991**, *2*, 367–422.
(54) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **1997**, *46* (1−3), 3–26.

JM9000659

Downloaded by Jonathan Berry on September 11, 2009 | http://pubs.acs.org
Publication Date (Web): May 15, 2009 | doi: 10.1021/jm9000659