JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Defining the RNA Internal Loops Preferred by Benzimidazole Derivatives via 2D Combinatorial Screening and Computational Analysis

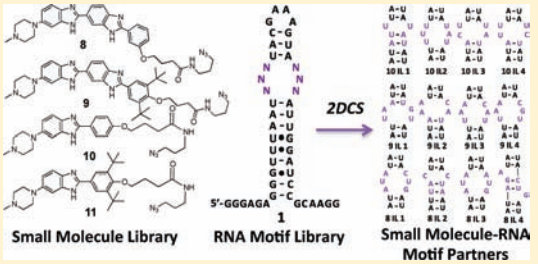
Sai Pradeep Velagapudi,[†,‡] Steven J. Seedhouse,[‡] Jonathan French,[‡] and Matthew D. Disney*[,†]

[‡]Department of Chemistry, The University at Buffalo, The State University of New York, Buffalo, New York 14260, United States

[†]The Kellogg School of Science and Engineering, The Department of Chemistry, The Scripps Research Institute, Scripps Florida, 130 Scripps Way #3A1, Jupiter, Florida 33458, United States

Supporting Information

ABSTRACT: RNA is an important therapeutic target; however, RNA targets are generally underexploited due to a lack of understanding of the small molecules that bind RNA and the RNA motifs that bind small molecules. Herein, we describe the identification of the RNA internal loops derived from a 4096 member 3 × 3 nucleotide loop library that are the most specific and highest affinity binders to a series of four designer, druglike benzimidazoles. These studies establish a potentially general protocol to define the highest affinity and most specific RNA motif targets for heterocyclic small molecules. Such information could be used to target functionally important RNAs in genomic sequence.



INTRODUCTION

RNA plays a number of diverse roles in cellular biology including encoding and translating protein, regulating the amount of protein expressed under different cellular conditions, and many other functions.[1−4] Because of this, RNA is an attractive target for chemical genetics probes or therapeutics. Many studies have identified small molecules that bind RNA;[5−7] however, few RNAs outside of the bacterial ribosome have been exploited for small molecule intervention. One of the major reasons that there are few therapeutics that elicit their effects by modulating RNA function is the limited amount of information that is gathered from traditional small molecule screening. Moreover, the information is sparse compared to the structural diversity of RNA present in genomic sequence, and it is not general in that it does not define the preferred RNA mofits for a small molecule.

Because the motifs that comprise an RNA structure can be accurately predicted from sequence via free energy minimization or phylogenic comparison,[8] strategies to design small molecules that bind RNA could be developed if structure prediction algorithms were used in conjunction with databases of small molecule−RNA interactions. To enable this approach, however, much more information is required than what is currently available on the RNA motifs that bind small molecules and the small molecules that bind RNA motifs.

In an effort to obtain RNA motif−ligand partners, multidimensional combinatorial screens, for example 2D combinatorial screening (2DCS), have been developed to identify the optimal RNA motifs from a library of discrete secondary structures (**1**, Figure 1) that bind small molecules.[9,10] In 2DCS studies, small molecule microarrays are hybridized with RNA libraries in which the randomized regions in the RNA are restricted to small motifs (internal loops or hairpins, for example) that have a high probability of being found in genomic RNAs. The members of the RNA motif library that bind to a small molecule displayed on the array surface are excised and sequenced. The features that the bound RNAs share, and hence those that contribute to binding affinity, are then determined computationally using statistical analysis.[11,12] The results of the statistical analysis are confirmed by in solution binding measurements. Interestingly, statistical significance for particular features in RNA structures that bind ligands can be used to predict affinity and specificity.[11,12] 2DCS and statistical analysis have been previously used to determine the RNA motifs preferred by aminoglycosides. The results have shown that there is unique RNA motif and sequence space for many of the aminoglycosides tested.[9−14]

Although these studies have helped to establish 2DCS as a potentially general technique to identify and annotate RNA motif−ligands interactions, the exclusive use of aminoglycosides as small molecule binders could limit its scope and hence its use in in the development of chemical genetics probes targeting cellular RNAs. For example, aminoglycosides display less than favorable profiles for use in cell culture and animals.[15,16] Therefore, an improved set of ligands would include compounds that are cell permeable, nontoxic, and more charge neutral.[17,18] Herein, the 2DCS approach is utilized to identify the preferred

Received: January 9, 2011
Published: May 23, 2011

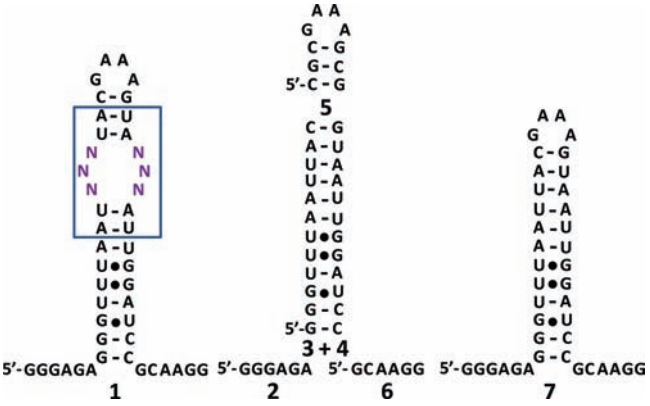


Figure 1. Secondary structures of the RNAs used in this study. Oligonucleotide **1** is the 3 × 3 nucleotide internal loop library. Oligonucleotides **2**−**6** are competitor oligonucleotides used to constrain selected interactions to the randomized region in **1**. Oligonucleotide **7** is the cassette into which the 3 × 3 nucleotide library was inserted.

RNA targets for a series of benzimidazole ligands that are based on the cell permeable Hoechst scaffold. The results show that the combination of 2DCS and computational analysis of the selected sequences can allow for the accurate prediction of small molecule affinity and specificity for each RNA structure contained in an RNA internal loop library.

EXPERIMENTAL SECTION

Instrumentation. Mass spectra were collected on a LCQ Advantage Ion Trap LC/MS equipped with a Surveyor HPLC system. HPLC was completed on a Waters 1525 Binary HPLC Pump equipped with a Waters 2487 Dual Absorbance Detector system. High-resolution mass spectra were collected at the Scripps Florida Proteomics Facility or at UC Riverside.

Chemicals. Hexanes and *N*-*N*-dimethylformamide (DMF) were from EMD; 3-bromopropylamine hydrobromide was from TCI; sodium azide was from Fisher Scientific; 4-hydroxybenzaldehyde, diisopropylethylamine (DIPEA), and ethyl 4-bromobutanoate were from Alfa Aesar; 3,5-di-*tert*-butyl-4-hydroxybenzaldehyde was from Acros Organics. Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) was from Advanced ChemTech. All were used without further purification. The HPLC solvents used were HPLC grade acetonitrile from EMD and water obtained from a Barnstead NANOpure Diamond Water Purification System operating at 18.2 mΩ-cm.

Preparative HPLC. All HPLC purifications were performed at room temperature monitoring at 218 and 254 nm. A SymmetryPrep C18 (7 μm, 19 × 150 mm column) was used with a flow rate of 10 mL/min and a linear gradient of 0% to 90% B in A over 45 min. (A is water +0.1% trifluoroacetic acid (TFA) (v/v) and B is acetonitrile +0.1% TFA (v/v).)

Analytical HPLC. The purity of all compounds was determined by analytical HPLC. A Waters Symmetry C18 (5 μm, 4.6 × 150 mm) column was used with a flow rate of 1 mL/min and a linear gradient of 0% to 100% B in A over 50 min (A is 0.1% TFA in water, B is 0.1% TFA in methanol). Absorbance was monitored at 218 and 254 nm.

*Synthesis of N-(3-Azidopropyl)-4-(4-(6-(4-methylpiperazin-1-yl)-1H-benzo[d]imidazole-2-yl)phenoxy)butanamide (**10**).* A 16.2 mg (0.041 mmol) portion of 4-(4-(6-(4-methylpiperazin-1-yl)-1H-benzo[*d*]imidazol-2-yl)phenoxy)butanoic acid, 31.2 mg (0.06 mmol) of PyBOP, and 15.5 mg (0.12 mmol) of DIPEA were dissolved in dry DMF. The mixture was stirred under argon for 30 min followed by addition of 7 mg (0.06 mmol) of 3-azidopropylamine.[19] After 36 h, the reaction was concentrated under vacuum. The residue was dissolved in water and purified by reverse phase HPLC to afford 14.5 mg of a yellow powder (75% yield); t_R = 16.5 min. MS−ESI(+) LRMS: Calcd.: 476; observed: 477 [M + H$^+$]; HRMS: Calcd.: 477.2721 [M + H$^+$]; observed 477.2720 [M + H$^+$]. ^{1}H NMR (CD_3OD, 500 MHz) δ 1.61 (2H, m, J = 7 Hz), 2.01 (2H, m, J = 7 Hz), 2.28 (2H, t, J = 7.5 Hz), 3.06 (3H, bs), 3.14 (2H, t, J = 6.5 Hz), 3.2 (2H, with solvent), 3.54−79 (4H, bd), 4.02 (2H, t, J = 6.5 Hz), 7.09 (2H, d, J = 9 Hz), 7.14 (1H, s), 7.23 (1H, d, J = 6.5 Hz), 7.54 (1H, d, J = 9 Hz), 7.91 (2H, d, J = 8.5 Hz). ^{13}C NMR (CD_3OD, 75 MHz) δ 26.26, 29.69, 33.27, 37.77, 43.57, 46.00 (with solvent), 50.10, 54.62, 68.93, 101.04, 115.29, 115.99, 117.06, 119.32, 127.63, 130.65, 134.15, 150.46, 150.70, 164.77, 175.35.

*Synthesis of N-(3-Azidopropyl)-4-(2,6-di-tert-butyl-4-(6-(4-methylpiperazin-1-yl)-1H,-2,5′-bibenzo[d]imidazol-2′-yl)phenoxy) butanamide (**11**).* An 82 mg (0.16 mmol) portion of 4-(2,6-di-*tert*-butyl-4-(6-(4-methylpiperazin-1-yl)-1*H*-benzo[*d*]imidazole-2-yl)phenoxy)butanoic acid, 125 mg (0.24 mmol) of PyBOP, and 31 mg (0.24 mmol) of DIPEA were added to dry DMF under argon, and the solution was stirred for 30 min. To this solution, 25 mg (0.24 mmol) of 3-azidopropylamine were added. The solution was stirred for 36 h under argon and then concentrated under vacuum. The residue was dissolved in water and purified by reverse phase HPLC to afford 73 mg of a yellow powder (76% yield); t_R = 27 min. MS−ESI(+) LRMS: Calcd.: 588; observed: 589 [M + H$^+$], 1177 [2M + H$^+$]; HRMS: Calcd.: 589.3973 [M + H$^+$]; observed: 589.3996 [M + H$^+$]. ^{1}H NMR (CD_3OD, 500 MHz) δ 1.48 (18H, s), 1.71 (2H, t, J = 6.5 Hz), 2.20 (2H, m, J = 7 Hz), 2.32 (2H, t, J = 7.5 Hz), 3.00 (3H, s), 3.17 (2H, bd), 3.21 (2H, m, J = 6.5), 3.3 (2H, with solvent), 3.64 (2H, d, J = 11.5 Hz), 3.77 (2H, t, J = 7 Hz) 3.90 (2H, d, 10.5), 7.37 (1H, s), 7.44 (1H, d, J = 8.5 Hz), 7.75 (1H, d, J = 8.5 Hz), 8.13 (2H, s). ^{13}C NMR (CD_3OD, 75 MHz) δ 25.82, 28.74, 31.35, 31.97, 36.36, 36.84, 42.58, 47.52, 49.15, 53.64, 76.50, 100.12, 114.38, 117.55, 118.45, 126.74, 133.31, 146.60, 149.76, 163.14, 174.04.

*Synthesis of N-(3-azidopropyl)-4-(2,6-di-tert-butyl-4-(6-(4-methylpiperazin-1-yl)-1H,-2,5′-bibenzo[d]imidazol-2′-yl)phenoxy) butanamide (**9**).* A 9 mg (0.016 mmol) sample of 4-(2,6-di-*tert*-butyl-4-(6-(4-methylpiperazin-1-yl)-1*H*,-2,5′-bibenzo[*d*]imidazol-2′-yl)phenoxy) butanoic acid, 11 mg (0.023 mmol) of PyBOP, and 3 mg (0.023 mmol) of DIPEA were dissolved in DMF under argon. The solution was stirred for 30 min followed by addition of 3 mg (0.023 mmol) of 3-azidopropylamine. The resulting solution was stirred for 36 h under argon and then concentrated under vacuum. The residue was dissolved in water and purified by reverse phase HPLC to afford 7.2 mg of a yellow powder (72% yield); t_R = 29 min. MS−ESI(+) LRMS: Calcd.: 704; observed: 705 [M + H$^+$], 359 [(M/2)+H$^+$]; HRMS: Calcd.: 705.4275 [M + H$^+$]; observed: 705.4645 [M + H$^+$]. ^{1}H NMR (DMSO-d_6, 500 MHz) δ 1.47 (18H, s), 1.63 (2H, t, J = 7 Hz), 2.08 (2H, t, J = 7 Hz), 2.23 (2H, t, J = 7 Hz), 2.91 (3H, s), 3.10 (4H, m, J = 6.5 Hz), 3.25 (2H, bs), 3.34 (2H, t, J = 7 Hz), 3.59 (2H, d, J = 11.5 Hz), 3.72 (2H, t, J = 7.5), 3.92 (2H, bd, J = 11.5 Hz) 7.27 (1H, s), 7.33 (1H, d, J = 8.5 Hz), 7.73 (1H, d, J = 8.5), 7.98 (2H, d, J = 8.5), 8.10 (1H, d, J = 8 Hz), 8.19 (2H, s), 8.49 (1H, s), 10.19 (1H, bs). ^{13}C NMR (CD_3OD, 75 MHz) δ 26.72, 29.73, 32.43, 33.01, 37.29, 37.81, 43.56, 50.10, 54.62, 77.36, 101.10, 115.64, 115.51, 116.59, 119.52, 120.39, 124.74, 127.93, 128.40, 134.80, 147.01, 150.74, 155.87, 163.53, 175.10.

*Synthesis of **9**-**Fl**.* A 200 nmol sample of 5-(*N*-(2-propyne)-formamide)-fluorescein (**Fl**) was added in methanol to a solution containing 150 nmol of **9**, 200 nmol $CuSO_4$, and 400 nmol freshly dissolved ascorbic acid. The final volume was brought to 700 μL with methanol. The reaction mixture was placed in an Emrys Creator monomode microwave system from Biotage operating at 2.45 GHz frequency with a pulsed microwave irradiation power of 0−300W. The temperature was controlled by using an infrared thermometer perpendicular to the sample vessel. The reaction was maintained at 110 °C with stirring for 4 h. The product was purified by preparative TLC on a Whatman 20 × 20 cm Silica gel TLC plate; a mobile phase of 16:8:1 ethyl acetate/methanol/triethylamine

was applied for approximately 1 h. After 1 h, methanol was added such that the mobile phase was 16:11:1 ethyl acetate/methanol/triethylamine. The product (**9-Fl**) was scraped from plate, extracted into methanol by vortexing for 30 s, and characterized by MS-ESI(+) LRMS: Calcd.: 1117; observed: 1118 [M + H^+], 560 [M+2H^+], 373 [M+3H^+]; MALDI HRMS: Calcd.: 1118.5247 [M + H^+]; observed: 1118.5293 [M + H^+]. The purity of the product was determined by analytical HPLC (t_R = 34 min); 11% yield as determined by absorbance at 496 nm in 1X PBS, pH 7.4 using an extinction coefficient of 45 000 $M^{-1}cm^{-1}$.

*Synthesis of **11-Fl***. A 150 nmol sample of 5-(*N*-(2-propyne)-formamide)-fluorescein in methanol was added to a solution containing 750 nmol of **11**, 200 nmol $CuSO_4$, and 400 nmol freshly prepared ascorbic acid. The final volume was brought to 700 μL with methanol. The reaction mixture was microwaved as described for **9-Fl**. The product, **11-Fl** was purified by TLC as described for **9-Fl** and characterized by MS-ESI(+) LRMS: Calcd.: 1001; observed: 1002 [M + H^+]; HRMS: Calcd.: 1002.4872 [M + H^+]; observed: 1002.4843 [M + H^+]. The purity of **11-Fl** was determined by analytical HPLC (t_R = 35 min); 70% yield as determined by absorbance at 496 nm in 1× PBS, pH 7.4 using an extinction coefficient of 45 000 $M^{-1}cm^{-1}$.[20]

Microarray Construction. Alkyne-functionalized microarrays were constructed as previously described.[10] The azido-functionalized Hoechst-like library (**8**−**11**) was immobilized onto alkyne-functionalized arrays via 1,3 Huisgen dipolar cycloaddition reaction. Five different concentrations (5.0 mM, 3.0 mM, 1.8 mM, 1.1 mM, and 0.65 mM) of **8**−**11** were applied to the array surface in 300 nL of spotting solution [10 mM Tris-HCl (pH 8.5), 500 μM $CuSO_4$, 1 mM ascorbic acid, 100 μM tris-(benzyltriazolylmethyl)amine (TBTA) (dissolved in 4:1 butanol/dimethyl sulfoxide, and 10% glycerol]. The slides were placed in a humidity chamber for 3 h. They were then washed in water for 5 min, followed by washing with 0.5× PBS for 5 min, and then water for 25 min. The slides were allowed to dry at room temperature.

RNA Selection (2DCS). Radioactively labeled internal loop library (**1**, 100 pmol) and competitor oligonucleotides (**2**−**6**, 27.7 nmol each) (Figure 1) were folded separately in 1× Hybridization Buffer (HB1; 20 mM (4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (Hepes), pH 7.5, 150 mM NaCl, and 5 mM KCl) at 60 °C for 5 min and allowed to cool slowly to room temperature. The folded oligonucleotides were then mixed together in a final volume of 400 μL. They were carefully pipetted onto a slide pre-equilibrated with HB1 supplemented with 40 μg/mL bovine serum albumin (BSA) (HB2). The oligonucleotides were evenly spread across the slide surface with a piece of parafilm. The arrays were hybridized at room temperature for 25 min, and the RNA solution was removed. The slides were then washed by submersion in HB2 for 2 × 5 min with gentle agitation and then with NANOpure water.

After drying on the benchtop, the arrays were exposed to a phosphorimager screen and imaged using a BioRad FX phosphorimager. The selected RNAs were mechanically removed by excising the agarose from the surface and RT-PCR amplified as described previously.[9]

Cloning and Sequencing. RT-PCR products were cloned into pUC-19 vector using BamH1 and EcoR1 restriction enzyme sites and transformed into DH5-α *Escherichia coli*. Blue-white screening revealed that the majority of colonies appeared light blue, despite the presence of the RT-PCR products due to a small, in-frame insert. Thus, light-blue colonies were used to inoculate 1 mL of TB medium containing 50 mg/L ampicillin. Plasmids were sequenced by Functional Biosciences, Inc. (Madison, WI).

Determination of the Binding Affinities of 9-Fl and 11-Fl for Nucleic Acids. Dissociation constants for the binding of RNAs to **9-Fl** and **11-Fl** were determined using an in solution, fluorescence-based assay.[21] To prevent nonspecific binding of the ligands to 96 well plates, Corning nonbinding surface 96 well plates were used and exposure of ligands to the plates was always less than 40 min. A selected RNA was folded in Assay Buffer (8 mM Na_2HPO_4, 190 mM NaCl, 1 mM EDTA, and 40 μg/mL BSA; AB1) at 60 °C for 5 min and allowed to slowly cool to room temperature. **9-Fl** or **11-Fl** was then added to final concentration of 50 nM. Serial dilutions (1:2) were then completed in 1X AB1 supplemented with 50 nM **9-Fl** or **11-Fl**. The solutions were incubated for 30 min at room temperature and then transferred to a well of 96 well black plate. Fluorescence intensity was measured on a Bio-Tek FLX-800 plate reader. The change in fluorescence intensity as a function of RNA concentration was fit to eq 1:[22]

$$I = I_0 + 0.5\Delta\varepsilon([FL]_0 - (([FL]_0 + [RNA]_0K_t)^2 - 4[FL]_0[RNA]_0)^{0.5}) \quad (1)$$

where I and I_0 are the observed fluorescence intensity in the presence and absence of RNA respectively, $\Delta\varepsilon$ is the difference between the fluorescence intensity in the absence of RNA and in the presence of infinite RNA concentration, $[FL]_0$ and $[RNA]_0$ are the concentrations of **9-Fl** or **11-FL** and RNA respectively, and K_t is the dissociation constant. Dissociation constants for a DNA containing the preferred Hoechst 33258 binding site were determined analogously.

Determination of the Binding Affinity of 8 for Nucleic Acids. Dissociation constants for RNAs selected to bind **8** were determined using an in solution assay. A selected RNA was annealed in AB1 at 60 °C for 5 min and allowed to slowly cool to room temperature. The RNA was then titrated into a well of a black 96 well plated containing 500 nM of **8** in AB1. After each addition of RNA, the sample was allowed to equilibrate for 5 min before measuring fluorescence intensity on a SpectralMax M5 plate reader (excitation/emission: 345/460 nm). RNA was added until the fluorescence intensity signal saturated. The change in fluorescence intensity as a function of RNA concentration was fit to eq 1. The dissociation constant for a DNA containing the preferred Hoechst 33258 binding site was determined analogously.

Computational Determination of Features in Selected RNAs that Lead to High Affinity and Selective Binding. The RNAs that were selected to bind to a ligand were analyzed via the *RNA Privileged Space Predictor Program* (*RNA-PSP*, v 2.0).[12] Briefly, *RNA-PSP* extracts the nucleotides derived from the randomized positions in each selected RNA. The extracted sequences are then analyzed for features (a guanine across from an adenine, for example) that appear more often in the selected sequences than in the entire library. By comparing the percentage of selected RNAs that have a feature of interest to the percentage of RNAs in **1** with the same feature, the statistical significance can be calculated as a Z-score:

$$\Phi = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} \quad (2)$$

$$Z_{obs} = \frac{(p_1 - p_2)}{\sqrt{\Phi(1-\Phi)(1/n_1 + 1/n_2)}} \quad (3)$$

where n_1 is the size of Population 1 (the selected mixture), n_2 is the size of Population 2 (**1**, 4096), p_1 is the observed proportion of Population 1 (selected mixture) displaying the feature, and p_2 is the observed proportion for Population 2 (entire library) displaying the feature. The Z-scores are then manually converted to the corresponding two-tailed p-value, which represents the confidence that a feature is preferred by the ligand and did not occur randomly.

A single selected RNA motif has multiple features that are statistically significant (occur with ≥95% confidence). Therefore, the Z-scores for each feature are summed to afford the sum Z-score. The sum Z-score is strongly correlated with the relative affinity of an RNA for a ligand.[12] Thus, a plot of the sum Z-score as a function of affinity allows for a scoring function to be derived to predict the binding affinity of each member of **1** to a ligand.

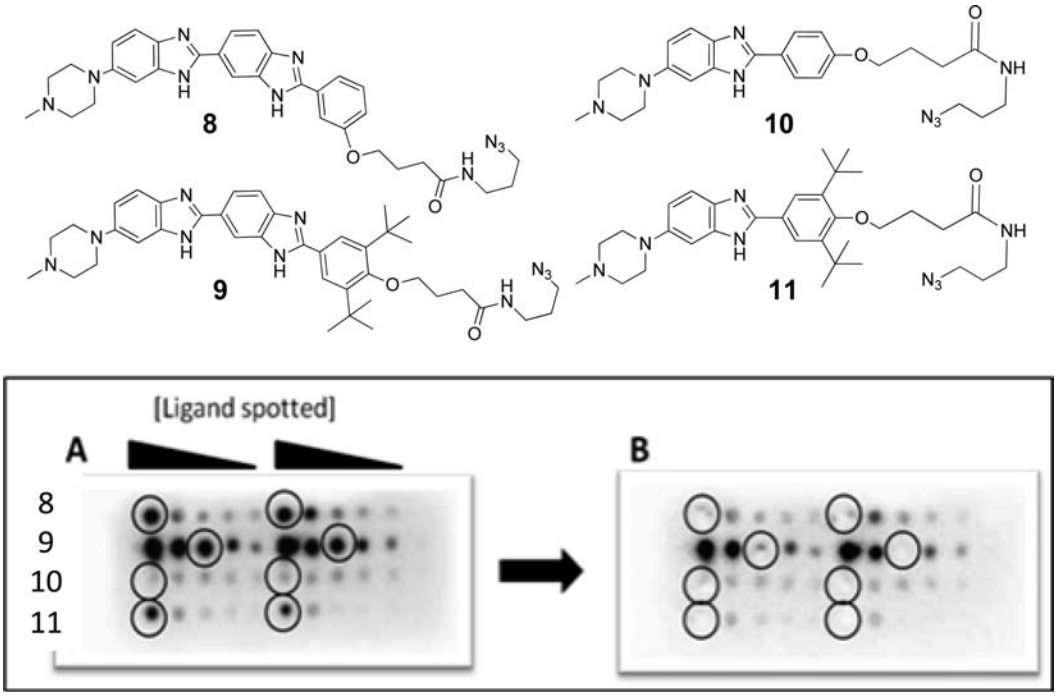


Figure 2. Top, structures of the bis-benzimidazoles and benzimidazoles that were conjugated onto alkyne-agarose microarray surfaces via HDCR. Bottom, microarray surface after hybridization with internally labeled library **1** and competitor oligonucleotides **2**−**6**. Spots of ligand-bound RNAs that were excised are indicated with a circle. The moles of ligand spotted are 1.5, 0.9, 0.5, 0.3, and 0.1 nmoles. A, array prior to excision of bound RNA. B, array after excision of bound RNA.

RESULTS AND DISCUSSION

We previously reported a series of investigations in which the RNAs preferred by aminoglycoside derivatives were determined by screening a small library of aminoglycosides against an RNA library displaying discrete secondary structural elements (Figure 1).[9,10,23] This approach, 2D combinatorial screening (2DCS), coupled with identifying statistically significant, unique features in the selected RNAs,[11,12] afforded high affinity and selective interactions. This information was then used in drug discovery efforts to target triplet-repeating and tetra-repeating RNA transcripts that cause disease.[24−27] The clinical use of aminoglycosides, however, is associated with side effects including ototoxicity and renal disease.[15,16,28,29] Therefore, the 2DCS approach was applied to more druglike ligands (top of Figure 2).

For these experiments, four benzimidazoles were synthesized (top of Figure 2) including **8**, which is an azide-functionalized derivative of Hoechst 33258,[26] and **10**, which is a benzimidazole version of **8**. Compounds **9** and **11** were synthesized to contain steric bulk at the phenolic side chain of Hoechst 33258. Previous modeling studies have shown that the addition of steric bulk to Hoechst derivatives decreases their affinities for AT-rich DNA structures.[30] Binding assays were completed and confirm this finding, showing that there was no detectable binding of **9** and **11** (K_d's ≫ 50 000 nM) to a DNA hairpin with the preferred Hoechst 33258 binding site (a 5′AATT sequence). Therefore, **9** and **11** could be specific for RNA. Binding assays with **10** also show that it is a weak DNA binder to that same sequence (K_d ≫ 50 000 nM). In contrast, the binding of **8** to this DNA is higher affinity with a K_d of 250 nM.[26]

Identifying Unique RNA Motif Space for Each Ligand. Ligands **8**−**11** were arrayed onto alkyne-functionalized agarose microarrays and conjugated to the surface using a Cu-catalyzed Huisgen dipolar cycloaddition reaction (HDCR).[31,32] Arrays were then hybridized with radioactively labeled **1** in the presence of excess competitor oligonucleotides **2**−**6** (Figure 1). A dose response for the ligands binding to **1** was observed (Figure 2). The RNAs captured by the ligands at various concentrations were excised and amplified. The RNAs bound to ligand at the lowest loading that could be amplified over background were cloned and sequenced because they are the highest affinity binders.[9] For the cases of **8** and **11**, bound RNAs where 1.5 nmoles of material were delivered to the array surface were analyzed, whereas the bound RNAs where 0.5 nmoles of **9** were delivered to the array surface were analyzed.

As can be observed from Figure 2, compounds **8**, **9**, and **11** gave signals that were much higher than background, whereas **10** did not yield appreciable signal above background. RNAs bound to **10** could not be amplified from the array due to the low signal. Most interestingly, compounds that had greater steric bulk (**9** and **11**) gave higher signals on the array than the corresponding parent structures, **8** and **10**, respectively (Figure 2). This is interesting because **9** and **11** do not bind DNA (K_d ≫ 50 000 μM).

The sequences of the RNA loops selected to bind **8**, **9**, and **11** were determined by RT-PCR amplification, cloning, and sequencing. Statistically significant features (a guanine across from an adenine, for example) within the selected RNAs were then determined by using *RNA-PSP*.[11,12] *RNA-PSP* statistically analyzes the selected sequences and structures to identify RNA motif space that is privileged to bind a particular ligand, represented as a Z-score. That is, the proportion of the selected RNAs that display a feature of interest is compared to the population of the entire library (**1**) that displays the same feature (eqs 2 and 3). *RNA-PSP* reports the features that are over-represented with at least a 95% confidence interval. Importantly, it also scores each

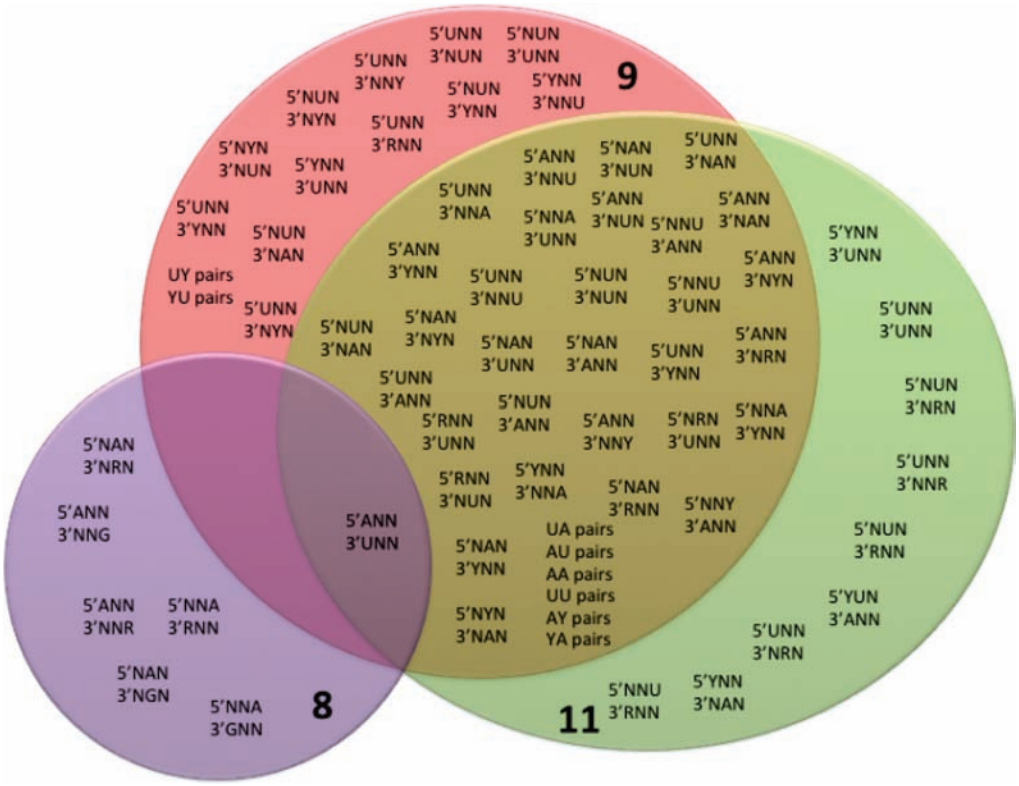


Figure 3. Venn diagram that illustrates the types of unique and overlapping RNA sequence spaces selected to bind **8**, **9**, and **11**. The Venn diagram analysis was enabled by using the *RNA-PSP* v 2.0 program.[12] Data for **8** are shown at the 95% confidence interval and data for **9** and **11** are shown at the 99% confidence interval. Structures that are listed as pairs refer to two nucleotides that are across from each other in the selected structures.

library member in **1** for its fitness to bind to a specific ligand (relative to the entire library) by summing the Z-score for each privileged feature in that RNA, or a sum Z-score.[12]

A Venn diagram was constructed by using the *RNA-PSP* analysis to compare the RNA motif spaces that bind each ligand and to identify overlap (Figure 3). Previous analyses have shown that unique space indicates a specific RNA motif−small molecule interaction, whereas overlapping space (a feature that belongs to more than one ligand) indicates that an RNA motif is not unique to one small molecule.[11] Qualitatively, a Venn diagram gives insights into specificity. A large amount of overlapping space is observed between ligands **9** and **11**, and these two ligands capture more types of RNA motif space than **8**. In general, the features unique to **8** contains purines on both sides of the loop, for example (5′*NNA*)/(3′*GNN*) and (5′*NAN*)/(3′*NGN*). In contrast, **9** generally prefers pyrimidines on both sides of the loop such as (5′*UNN*)/(3′*NUN*) and (5′*NUN*)/(3′*NYN*). Much of the RNA space unique to **11** contains a purine on one side of the loop and a pyrimidine on the other (5′*NNU*)/(3′*RNN*) and (5′*UNN*)/(3′*NRN*).

The affinities of a subset of the selected RNAs were determined for the corresponding ligands using in solution binding measurements (Figure 4). Compound **8** was used directly to determine dissociation constants as it is fluorescent, and a large change in fluorescence intensity is observed in the presence of RNA. In contrast, the dynamic range in fluorescence intensity upon binding of **9** and **11** is in general too small to accurately determine dissociation constants as previously observed for DNA aptamers.[30] Thus, these compounds were conjugated to fluorescein to afford **9-Fl** and **11-Fl**.

The highest affinity interactions were observed for the RNA structures selected to bind **8** with K_d's that range from about 100 to 1200 nM. Lower affinity interactions are observed with RNAs selected to bind **9** and **11**. The highest affinity interactions are in the low micromolar regime, with **9** having K_d's between 1000 and 3000 nM and **11** having K_d's between 1000 and 20 000 nM.

Analysis of the microarray signals, the affinity data, and the data from the Venn diagram allows one to determine if higher signal for related ligands at a given loading indicates interactions are higher affinity or that more types of RNAs are bound. In the case of this selection, it appears that that the higher signals on the array for compound **9** relative to **8** are due to **9** binding more types of RNA structures.

Prediction of RNA−Ligand Affinity Using Statistical Analysis. A quantitative relationship between statistical analysis and affinity can be derived using an approach we call structure-activity relationship through sequencing (StARTS).[12] In this approach, the statistical score (Z-score at 95% confidence) for each feature in a selected RNA calculated by *RNA-PSP* is summed to afford a sum Z-score. Measured affinities are then plotted as a function of the sum Z-score. These data can be fit to a pseudofirst-order equation (Figure 5), which can be used to predict the affinities of all members of **1**.

StARTS plots for **8**, **9**, and **11** show that there is a strong correlation between affinity and the sum Z-score, in good agreement with a previous study (Figure 5).[12] For all three ligands, the RNA motifs with the highest sum Z-scores have the highest affinities. Likewise, selected loops with lower sum Z-scores have lower affinities. Thus, this computational approach is apparently general and allows the accurate prediction of the

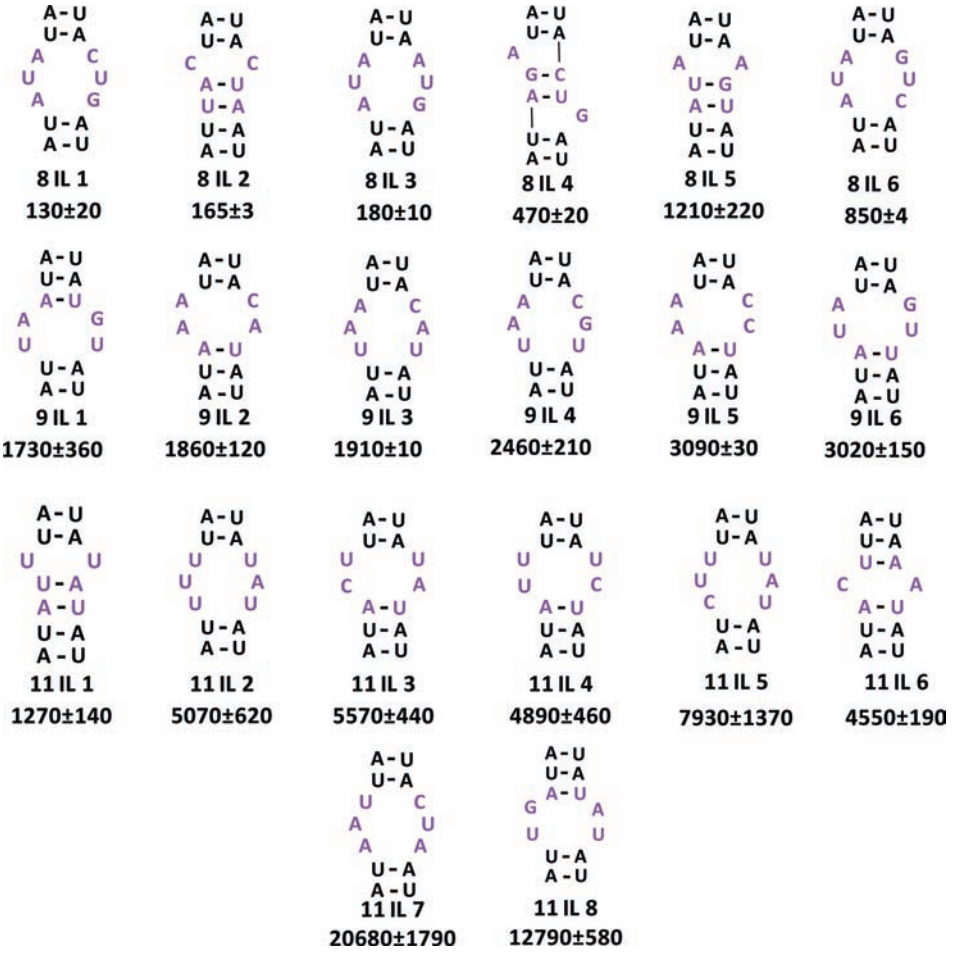


Figure 4. Secondary structures of a subset of RNA loops selected to bind **8**, **9**, and **11** via 2DCS. Secondary structures were predicted by the RNA structure program.[8] The nomenclature for the loops refers to the ligand that the loops were selected to bind followed by an internal loop (IL) number. Values below the loop identifier are the K_d's for the RNA—ligand complex (nM). None of the compounds binds to cassette **7** or library **1**, indicating that binding is specific to the selected randomized regions.

highest affinity RNA binders from a population of structures via statistical analysis.

Instead of the qualitative prediction of selectivity afforded by the Venn diagram, the StARTS approach can provide more quantitative predictions. Because affinity scales with sum Z-score values (via a pseudofirst-order equation), the affinity of a loop for a different ligand can be estimated from the corresponding StARTS plot. For example, the affinity of **8 IL4** (a loop selected to bind **8**) for **11** can be estimated by calculating **8 IL4**'s sum Z-score for **11** and using the pseudofirst-order equation that describes the relationship between sum Z-score and affinity for **11**.

The ability of the StARTS analysis to predict selectivity was validated by studying the affinities of **8 IL4**, **9 IL5**, and **11 IL5** for the other arrayed ligands. The sum Z-scores for the loops and the ligands they were selected to bind are in top 0.4% for all members of library **1**. In contrast, their sum Z-scores for the other ligands range from the top 68% to the top 10% indicating that they should bind more weakly (Figure 6). Indeed, no saturable binding was observed at 30 μM RNA when **11 IL5** or **8 IL4** were tested for binding **9** and **11**, and no saturable binding was observed when up to 4.5 μM **9 IL5** or **11 IL5** was added to **8**. One would expect that **11** would bind more weakly to RNA relative to **8** and **9** due to its dimished hydrophobic surface area to stack with RNA and the decreased number of hydrogen bond acceptors and donors. The statistical analysis and StARTS of the 2DCS selection identifies RNAs that bind selectively to **11** over **8** and **9** despite its less favorable chemical properties. The selectivity data are graphically represented in Figure 6, which contains a plot for each ligand and the sum, Z-scores of all RNAs. Selected RNAs for each ligand are represented as different colors, and RNAs whose affinities were measured are labeled. Thus, the relative sum Z-score for selected internal loops can allow for predictions of the selectivity of small molecule—RNA motif interactions.

Comparison of the Binding of Benzimidazoles to RNA and to DNA. The RNA motifs selected to bind **9** and **11** are much higher affinity than binding of the ligands to AT-rich DNA. This is likely due to different binding modes. Hoechst 33258 interacts with the DNA minor groove.[33] Introduction of steric bulk onto the phenolic side chain, affording **9** and **11**, would preclude DNA binding because the bulky *tert*-butyl groups would sterically clash with the DNA minor groove.[30] There are many potential binding modes of **9** and **11** to the selected RNA structures in which the bulky side chains do not occlude binding. For example, the ligands could interact directly with the bases of selected internal loop nucleotides or with the RNA major or minor grooves, which have very different shapes than DNA. Moreover, RNA grooves

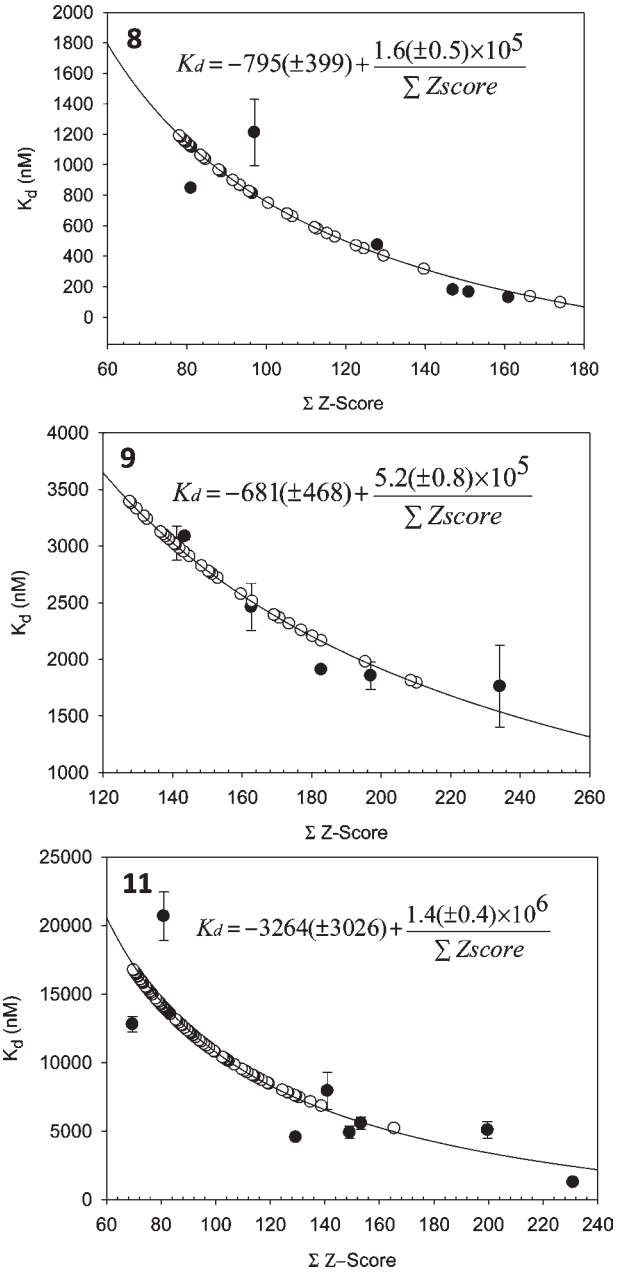


Figure 5. StARTS[12] plots for **8**, **9**, and **11**. Filled circles indicate structures to which affinities were measured (Figure 4). Open circles indicate predicted affinities for members of **1** with sum Z-score values in the top ~10%.

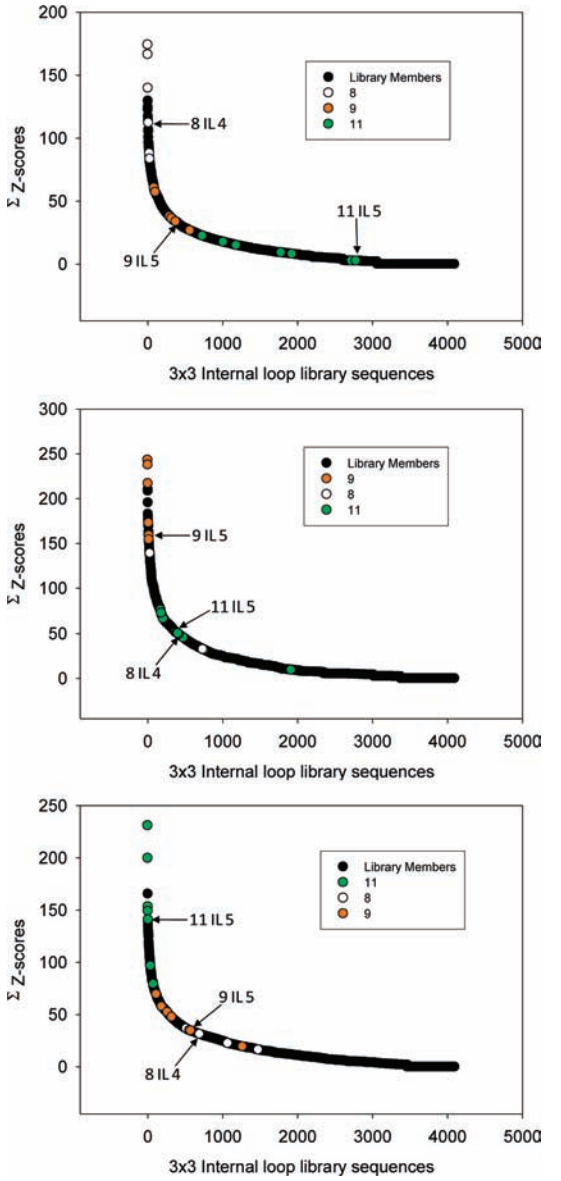


Figure 6. StARTS plots containing all library members for each ligand. Each plot depicts the sum Z-scores for each member of **1** as a function of affinity. Top, selection for **8**; middle, selection for **9**; and bottom, selection for **11**. The open or colored circles are the RNA motif–ligand partners shown in Figure 5. A library member refers to the calculated sum Z-score for the entire 4096 member library **1** based on the analysis of selected structures for binding a given ligand. As noted in the text, compounds that have lower relative sum Z-scores are weaker binders. Thus, by comparing the position of an RNA in each StARTS plots, the specificity of the RNA–ligand interactions can be estimated.

could expand or contract around loop nucleotides,[34–39] a degree of flexibility that a fully paired DNA likely does not have.

Comparisons to Previous Studies. Previous studies using 2DCS have exclusively investigated the binding of aminoglycoside ligands to RNA structures.[9,10,13,14] These studies found that selected interactions can occur with nanomolar K_d's to internal loops[9,10,13] and low micromolar K_d's to hairpin loops.[11,14] Aminoglycosides have a large number of amines and are therefore highly cationic ligands. Herein, we identified RNA–ligand interactions with affinities that are comparable or somewhat weaker than aminoglycosides but with more druglike ligands that have significantly diminished charge.

Other studies have disclosed RNA–ligand interactions using ligands related to **8**–**11**. For example, Hoechst 33258, a derivative of **8**, has low nanomolar[26,40] to micromolar[41] affinities for a variety of RNA targets. Compounds similar to **10** and **11** were found as hits from an 180 000 member chemical library that bind a loop in the Hepatitis C virus internal ribosomal entry site. The initial hit bound with a K_d of ~100 000 nM. It was optimized to afford ligands with K_d's ranging from 700 to 17 000 nM.[42] Thus, our ability to identify and characterize RNA structures that bind small molecules with K_d's from the low nanomolar to the

micromolar range and predict ligand selectivity via 2DCS bodes well for this approach identifying RNA motif−druglike ligand interactions using larger and more diverse chemical libraries.

■ ASSOCIATED CONTENT

Supporting Information. Details of chemical synthesis, RNA binding assays, and computational analysis of the selected structures for binding each arrayed ligand. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
Disney@scripps.edu

■ ACKNOWLEDGMENT

This work was funded by the National Institutes of Health (1R01GM079235-01A2) and the Research Corporation. M.D.D. is a Dreyfus New Faculty Awardee, a Dreyfus Teacher-Scholar, and a Research Corporation Cottrell Scholar.

■ REFERENCES

(1) Blount, K. F.; Breaker, R. R. *Nat. Biotechnol.* **2006**, *24*, 1558.
(2) Calin, G. A.; Croce, C. M. *Oncogene* **2006**, *25*, 6202.
(3) Poehlsgaard, J.; Douthwaite, S. *Nat. Rev. Microbiol.* **2005**, *3*, 870.
(4) Fedor, M. J.; Williamson, J. R. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 399.
(5) Thomas, J. R.; Hergenrother, P. J. *Chem. Rev.* **2008**, *108*, 1171.
(6) Chow, C. S.; Bogdan, F. M. *Chem. Rev.* **1997**, *97*, 1489.
(7) Gallego, J.; Varani, G. *Acc. Chem. Res.* **2001**, *34*, 836.
(8) Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7287.
(9) Childs-Disney, J. L.; Wu, M.; Pushechnikov, A.; Aminova, O.; Disney, M. D. *ACS Chem. Biol.* **2007**, *2*, 745.
(10) 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. *J. Am. Chem. Soc.* **2008**, *130*, 11185.
(11) Paul, D. J.; Seedhouse, S. J.; Disney, M. D. *Nucleic Acids Res.* **2009**, *37*, 5894.
(12) Velagapudi, S. P.; Seedhouse, S. J.; Disney, M. D. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 3816.
(13) Tran, T.; Disney, M. D. *Biochemistry* **2010**, *49*, 1833.
(14) Aminova, O.; Paul, D. J.; Childs-Disney, J. L.; Disney, M. D. *Biochemistry* **2008**, *47*, 12670.
(15) Silva, J. G.; Carvalho, I. *Curr. Med. Chem.* **2007**, *14*, 1101.
(16) Mingeot-Leclercq, M. P.; Brasseur, R.; Schanck, A. *J. Toxicol. Environ. Health* **1995**, *44*, 263.
(17) Herzog, A.; Schutze, H. R. *Dtsch. Tierarztl. Wochenschr.* **1968**, *75*, 476.
(18) Disney, M. D.; Stephenson, R.; Wright, T. W.; Haidaris, C. G.; Turner, D. H.; Gigliotti, F. *Antimicrob. Agents Chemother.* **2005**, *49*, 1326.
(19) Lampkins, A. J.; O'Neil, E. J.; Smith, B. D. *J. Org. Chem.* **2008**, *73*, 6053.
(20) Durroux, T.; Peter, M.; Turcatti, G.; Chollet, A.; Balestre, M.-N.; Barberis, C.; Seyer, R. *J. Med. Chem.* **1999**, *42*, 1312.
(21) Childs-Disney, J. L.; Wu, M.; Pushechnikov, A.; Aminova, O.; Disney, M. D. *ACS Chem. Biol.* **2007**, *2*, 745.
(22) Wang, Y.; Rando, R. R. *Chem. Biol.* **1995**, *2*, 281.
(23) Disney, M. D.; Childs-Disney, J. L. *Chembiochem* **2007**, *8*, 649.
(24) Lee, M. M.; Childs-Disney, J. L.; Pushechnikov, A.; French, J. M.; Sobczak, K.; Thornton, C. A.; Disney, M. D. *J. Am. Chem. Soc.* **2009**, *131*, 17464.
(25) Lee, M. M.; Pushechnikov, A.; Disney, M. D. *ACS Chem. Biol.* **2009**, *4*, 345.
(26) Pushechnikov, A.; Lee, M. M.; Childs-Disney, J. L.; Sobczak, K.; French, J. M.; Thornton, C. A.; Disney, M. D. *J. Am. Chem. Soc.* **2009**.
(27) Disney, M. D.; Lee, M. M.; Pushechnikov, A.; Childs-Disney, J. L. *Chembiochem* **2010**.
(28) Wright, G. D.; Berghuis, A. M.; Mobashery, S. *Adv. Exp. Med. Biol.* **1998**, *456*, 27.
(29) Hutchin, T.; Haworth, I.; Higashi, K.; Fischel-Ghodsian, N.; Stoneking, M.; Saha, N.; Arnos, C.; Cortopassi, G. *Nucleic Acids Res.* **1993**, *21*, 4174.
(30) Sando, S.; Narita, A.; Aoyama, Y. *Chembiochem* **2007**, *8*, 1795.
(31) Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. *Org. Lett.* **2004**, *6*, 2853.
(32) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 2004.
(33) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215.
(34) SantaLucia, J., Jr.; Turner, D. H. *Biochemistry* **1993**, *32*, 12612.
(35) Wu, M.; Turner, D. H. *Biochemistry* **1996**, *35*, 9677.
(36) Burkard, M. E.; Turner, D. H. *Biochemistry* **2000**, *39*, 11748.
(37) Batey, R. T.; Rambo, R. P.; Doudna, J. A. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2326.
(38) Doudna, J. A. *Nat. Struct. Biol.* **2000**, *7* (Suppl), 954.
(39) Weeks, K. M.; Crothers, D. M. *Science* **1993**, *261*, 1574.
(40) Cho, J.; Rando, R. R. *Nucleic Acids Res.* **2000**, *28*, 2158.
(41) Dassonneville, L.; Hamy, F.; Colson, P.; Houssier, C.; Bailly, C. *Nucleic Acids Res.* **1997**, *25*, 4487.
(42) Seth, P. P.; Miyaji, A.; Jefferson, E. A.; Sannes-Lowery, K. A.; Osgood, S. A.; Propp, S. S.; Ranken, R.; Massire, C.; Sampath, R.; Ecker, D. J.; Swayze, E. E.; Griffey, R. H. *J. Med. Chem.* **2005**, *48*, 7099.